CONTRIBUTIONS TO PARAMETRIC STATISTICAL THEORY
AND PRACTICE

JOHN AITCHISON

Volume 2

D.Sc.
University of Edinburgh
1980
CONTENTS OF VOLUME 2

AITCHISON, J. and BROWN, J.A.C. (1957)  
The Lognormal Distribution  
Cambridge University Press

AITCHISON, J. and DUNSMORE, I.R. (1975)  
Statistical Prediction Analysis  
Cambridge University Press
THE LOGNORMAL DISTRIBUTION
with special reference to its uses in economics

BY
J. AITCHISON
Professor of Statistics
University of Glasgow

AND
J. A. C. BROWN
Professor of Applied Economics
at the University of Oxford
and Fellow of Merton College

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON · NEW YORK · MELBOURNE
The very instant that I saw you did
My heart fly to your service; there resides,
To make me slave to it; and for your sake
Am I this patient log-man.

_The Tempest_
CONTENTS

List of tables
List of figures
Preface

Chapter 1. INTRODUCTION
1.1 Purpose of the study
1.2 History of the lognormal distribution
1.3 Review of problems considered
1.4 Notation

Chapter 2. GENERAL PROPERTIES OF LOGNORMAL DISTRIBUTIONS
2.1 Introduction
2.2 The two-parameter distribution: definition
2.3 Moments and other characteristics
2.4 Reproductive properties
2.5 Moment distributions: Gini's coefficient of mean difference
2.6 Central limit theorems
2.7 The three-parameter distribution: definition
2.8 Moments, other characteristics, and moment distributions
2.9 Negatively skew distributions
2.10 A four-parameter distribution
2.11 Generalizations by series representations

Chapter 3. THE GENESIS OF LOGNORMAL DISTRIBUTIONS
3.1 Reasons for considering the genesis of a distribution
3.2 Kapteyn's system of translation and Pearson's criticism
3.3 The theory of proportionate effect
3.4 Kapteyn's analogue machine
3.5 Extensions and criticisms of the theory of genesis
3.6 The theory of breakage
## CONTENTS

**Chapter 4. Artificial Lognormal Samples and Tests of Lognormality**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 The construction of artificial samples from a lognormal population</td>
<td>28</td>
</tr>
<tr>
<td>4.2 The purpose of constructing artificial samples</td>
<td>29</td>
</tr>
<tr>
<td>4.3 The 65 samples constructed</td>
<td>29</td>
</tr>
<tr>
<td>4.4 Testing for lognormality: introductory remarks</td>
<td>30</td>
</tr>
<tr>
<td>4.5 Logarithmic probability paper</td>
<td>31</td>
</tr>
<tr>
<td>4.6 Tests of lognormality: Geary and Pearson tests; ( \chi^2 )</td>
<td>34</td>
</tr>
<tr>
<td>4.7 Tests applied to the 65 samples</td>
<td>35</td>
</tr>
</tbody>
</table>

**Chapter 5. Estimation Problems: I**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 General remarks on estimation</td>
<td>37</td>
</tr>
<tr>
<td>5.2 Estimation of ( \mu ) and ( \sigma^2 ) in a two-parameter distribution</td>
<td>38</td>
</tr>
<tr>
<td>5.2.1 The method of maximum likelihood, p. 39.</td>
<td></td>
</tr>
<tr>
<td>5.2.2 The method of moments, p. 39.</td>
<td></td>
</tr>
<tr>
<td>5.2.3 The method of quantiles, p. 40.</td>
<td></td>
</tr>
<tr>
<td>5.2.4 The graphical method, p. 42</td>
<td></td>
</tr>
<tr>
<td>5.3 Experimental results (I): the methods of §5.2 applied to the 65 samples</td>
<td>43</td>
</tr>
<tr>
<td>5.4 Estimation of ( \alpha ) and ( \beta^2 ) in a two-parameter distribution</td>
<td>44</td>
</tr>
<tr>
<td>5.4.1 The method of maximum likelihood, p. 45.</td>
<td></td>
</tr>
<tr>
<td>5.4.2 The method of moments, p. 46.</td>
<td></td>
</tr>
<tr>
<td>5.4.3 The method of quantiles, p. 46.</td>
<td></td>
</tr>
<tr>
<td>5.4.4 The graphical method, p. 48</td>
<td></td>
</tr>
<tr>
<td>5.5 Experimental results (II): the methods of §5.4 applied to the 65 samples</td>
<td>48</td>
</tr>
<tr>
<td>5.6 Confidence intervals</td>
<td>50</td>
</tr>
<tr>
<td>5.6.1 Confidence intervals for ( \mu ) and ( \sigma^2 ), p. 50.</td>
<td></td>
</tr>
<tr>
<td>5.6.2 Confidence intervals for ( \alpha ) and ( \beta^2 ), p. 50</td>
<td></td>
</tr>
<tr>
<td>5.7 The grouping of observations</td>
<td>51</td>
</tr>
<tr>
<td>5.7.1 Grouping for ( \mu ) and ( \sigma^2 ), p. 51.</td>
<td></td>
</tr>
<tr>
<td>5.7.2 Grouping for ( \alpha ) and ( \beta^2 ), p. 52</td>
<td></td>
</tr>
<tr>
<td>5.8 Some special devices of estimation</td>
<td>52</td>
</tr>
<tr>
<td>5.9 Summary of estimation procedures for the two-parameter distribution</td>
<td>53</td>
</tr>
<tr>
<td>5.9.1 Ungrouped data: estimation of ( \mu ) and ( \sigma^2 ), p. 53.</td>
<td></td>
</tr>
<tr>
<td>5.9.2 Grouped data: estimation of ( \mu ) and ( \sigma^2 ), p. 53.</td>
<td></td>
</tr>
<tr>
<td>5.9.3 Ungrouped data: estimation of ( \alpha ) and ( \beta^2 ), p. 54.</td>
<td></td>
</tr>
<tr>
<td>5.9.4 Grouped data: estimation of ( \alpha ) and ( \beta^2 ), p. 54.</td>
<td></td>
</tr>
</tbody>
</table>
CONTENTS

Chapter 6. Estimation Problems: II

6.1 Introduction... page 55

6.2 Estimation of the parameters of the three-parameter distribution

6.2.1 The method of maximum likelihood, p. 55. 6.2.2 Cohen's least sample value method, p. 56. 6.2.3 The method of moments, p. 57. 6.2.4 The method of quantiles, p. 58. 6.2.5 Kemsley's method, p. 59. 6.2.6 The graphical method, p. 60. 6.2.7 Special devices, p. 60

6.3 Experimental results: comparison of Kemsley's method and the methods of moments and quantiles 61

6.4 Summary of conclusions for the three-parameter distribution 63

6.5 Estimation of the parameters of the four-parameter distribution 65

Chapter 7. The Lognormal Distribution and Probit Analysis

7.1 Introduction 66

7.2 Quantal and quantitative responses to stimuli 66

7.3 Notation 68

7.4 The estimation of quantal response 68

7.5 A comparison with the logistic curve 72

7.6 Convergence problems in quantal response computations 75

7.7 The estimation of quantitative response: the homoscedastic case 78

7.8 The estimation of quantitative response: a special case 81

7.9 The estimation of quantitative response: the heteroscedastic case 82

7.10 Extensions of the theory 84

Chapter 8. Comparisons of Lognormal Populations

8.1 Use of t- and F-tests for \( \mu \) and \( \sigma^2 \) 85

8.2 Comparison of estimates for \( \alpha \) and \( \beta^2 \) 85

8.3 The three-parameter distribution 86

8.4 The analysis of variance 86
## Contents

### Chapter 9. Truncated and Censored Distributions and the Treatment of Zero Observations

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Definitions of truncation and censorship</td>
<td>87</td>
</tr>
<tr>
<td>9.2</td>
<td>Estimation of the parameters of the truncated distribution</td>
<td>88</td>
</tr>
<tr>
<td>9.3</td>
<td>Estimation of the parameters of the censored distribution</td>
<td>90</td>
</tr>
<tr>
<td>9.4</td>
<td>An application to the artificial samples</td>
<td>91</td>
</tr>
<tr>
<td>9.5</td>
<td>Distributions of counts</td>
<td>92</td>
</tr>
<tr>
<td>9.6</td>
<td>Grouped truncated and grouped censored distributions</td>
<td>94</td>
</tr>
<tr>
<td>9.7</td>
<td>The treatment of zero observations</td>
<td>94</td>
</tr>
<tr>
<td>9.8</td>
<td>A statistical model: the $\Delta$-distribution</td>
<td>95</td>
</tr>
<tr>
<td>9.9</td>
<td>Estimation for the $\Delta$-distribution</td>
<td>96</td>
</tr>
</tbody>
</table>

### Chapter 10. Examples of Lognormal Distributions

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1</td>
<td>Introduction</td>
<td>100</td>
</tr>
<tr>
<td>10.2</td>
<td>Small-particle statistics</td>
<td>100</td>
</tr>
<tr>
<td>10.3</td>
<td>Economies and sociology</td>
<td>101</td>
</tr>
<tr>
<td>10.4</td>
<td>Biology</td>
<td>102</td>
</tr>
<tr>
<td>10.5</td>
<td>Anthropometry</td>
<td>103</td>
</tr>
<tr>
<td>10.6</td>
<td>The abundance of species and discrete counts</td>
<td>103</td>
</tr>
<tr>
<td>10.7</td>
<td>Household size</td>
<td>104</td>
</tr>
<tr>
<td>10.8</td>
<td>Physical and industrial processes</td>
<td>104</td>
</tr>
<tr>
<td>10.9</td>
<td>Astronomy</td>
<td>105</td>
</tr>
<tr>
<td>10.10</td>
<td>Ages at first marriage: a difficulty with the threshold parameter</td>
<td>105</td>
</tr>
<tr>
<td>10.11</td>
<td>Philology: classification problems: time as a logarithmic stimulus</td>
<td>105</td>
</tr>
</tbody>
</table>

### Chapter 11. The Distribution of Incomes

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1</td>
<td>Conceptual problems in describing the size-distribution of incomes</td>
<td>107</td>
</tr>
<tr>
<td>11.2</td>
<td>Criteria for statistical descriptions</td>
<td>108</td>
</tr>
<tr>
<td>11.3</td>
<td>Models of generation of income distributions</td>
<td>108</td>
</tr>
<tr>
<td>11.4</td>
<td>Statistical analysis of data</td>
<td>111</td>
</tr>
</tbody>
</table>
## CONTENTS

11.5 Interpretation of the parameters of the lognormal distribution  
   page 111
11.6 The test of practice 116
11.7 Summary and further suggestions 119

**Chapter 12. Applications of Lognormal Theory in the Analysis of Consumers' Behaviour**

12.1 Introduction 121
12.2 The transition from micro-models to macro-models 121
12.3 The primacy of the Engel curve in empirical work 124
12.4 The derivation of an Engel curve for an indivisible good 124
12.5 The extension of the theory to a divisible good 126
12.6 The problem of inferior goods 127
12.7 The consolidation of commodities and of groups of consumers 129
12.8 The economic interpretation of the parameters of the Engel curve 130
12.9 The problem of additivity 131
12.10 The influence of prices 134
12.11 Changes in preferences 136
12.12 The effects of household composition 137
12.13 Aggregation and the analysis of time series 138
12.14 Concluding remarks 139

**Chapter 13. Computation Problems**

13.1 The use of an electronic computer 141
13.2 Description of the EDSAC 141
13.3 The processing of the 65 samples 142
13.4 The programme for quantal probit analysis 143
13.5 The programme for quantitative probit analysis 144
13.6 The convergence programme 145
13.7 The construction of tables for $\lambda_n(t)$ and $\chi_n(t)$ 145

References 146

Appendix A. Tables of Functions 154

Appendix B. The Results of Applying the Different Methods of Estimation to the 65 Artificial Samples 166

Index 175
LIST OF TABLES

4.1 Number of artificial samples constructed for each value of \( n \) and \( \sigma \) ... page 30
4.2 Grouped cumulative frequencies for an artificial sample: \( n=64; \mu=0; \sigma=0.6 \) ... 32
4.3 Cumulative frequency data for five artificial samples ... 34
4.4 Artificial samples giving significant values of \( g_1(3) \) ... 35
4.5 Artificial samples giving significant values of \( g_2(1) \) or \( g_4(4) \) ... 35
4.6 Values of \( \mu \) and \( \sigma \) for the artificial samples of Table 4.3 ... 36
5.1 Values of \( \Delta(m_i) \) for a grouping by sample size ... 43
5.2 Values of \( \Delta(m_i) \) for a grouping by size of \( \sigma \) ... 43
5.3 Values of \( \Delta(s) \) for a grouping by sample size ... 44
5.4 Values of \( \Delta(s) \) for a grouping by size of \( \sigma \) ... 44
5.5 Values of \( \Delta(a_i) \) for a grouping by sample size ... 49
5.6 Values of \( \Delta(a_i) \) for a grouping by size of \( \sigma \) ... 49
5.7 Values of \( \Delta(t_i) \) for a grouping by sample size ... 49
5.8 Values of \( \Delta(t_i) \) for a grouping by size of \( \sigma \) ... 49
6.1 Cases for which the quantile and Kemsley’s methods were impossible, classified by sample size ... 62
6.2 Cases for which the quantile and Kemsley’s methods were impossible, classified by size of \( \sigma \) ... 62
6.3 Values of \( \Delta(t_i) \) for a grouping by sample size ... 62
6.4 Values of \( \Delta(t_i) \) for a grouping by size of \( \sigma \) ... 62
6.5 Values of \( \Delta(m_i) \) for a grouping by sample size ... 63
6.6 Values of \( \Delta(m_i) \) for a grouping by size of \( \sigma \) ... 63
6.7 Values of \( \Delta(s) \) for a grouping by sample size ... 63
6.8 Values of \( \Delta(s) \) for a grouping by size of \( \sigma \) ... 63
6.9 Quantile methods compared for the two- and three-parameter distributions: by sample size ... 64
6.10 Quantile methods compared for the two- and three-parameter distributions: by size of \( \sigma \) ... 64
7.1 Toxicity of Rotenone to Macrocephaliella sanborni ... 72
7.2 Iterations for Rotenone data, starting from different initial values ... 72
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.3</td>
<td>Comparison of quantiles given by probit and logistic curves for data of Table 7.1</td>
<td>74</td>
</tr>
<tr>
<td>7.4</td>
<td>The four samples used in the study of convergence</td>
<td>76</td>
</tr>
<tr>
<td>7.5</td>
<td>Analysis of variance of quantitative responses</td>
<td>79</td>
</tr>
<tr>
<td>9.1</td>
<td>Maximum-likelihood estimates of $\mu$ from full, truncated and censored samples of size 128</td>
<td>92</td>
</tr>
<tr>
<td>9.2</td>
<td>Maximum-likelihood estimates of $\sigma^2$ from full, truncated and censored samples of size 128</td>
<td>93</td>
</tr>
</tbody>
</table>

### TABLES IN APPENDICES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Characteristics of lognormal distributions</td>
<td>154</td>
</tr>
<tr>
<td>A2</td>
<td>The function $\psi_n(t)$</td>
<td>156</td>
</tr>
<tr>
<td>A3</td>
<td>The function $\chi_n(t)$</td>
<td>159</td>
</tr>
<tr>
<td>A4</td>
<td>The solution of the equation $u^2 + 3u = k$</td>
<td>161</td>
</tr>
<tr>
<td>A5</td>
<td>The normal integral $P$, the ordinate $Z$ and the elasticity $Z/P$</td>
<td>162</td>
</tr>
<tr>
<td>A6</td>
<td>Factors for a quantal probit analysis</td>
<td>163</td>
</tr>
<tr>
<td>A7</td>
<td>Factors for a homoscedastic quantitative probit analysis</td>
<td>164</td>
</tr>
<tr>
<td>A8</td>
<td>Factors for a heteroscedastic quantitative probit analysis</td>
<td>165</td>
</tr>
<tr>
<td>B1</td>
<td>Estimates $m$ of $\mu$ for $\Lambda(\mu, \sigma^2)$</td>
<td>166</td>
</tr>
<tr>
<td>B2</td>
<td>Estimates $s^2$ of $\sigma^2$ for $\Lambda(\mu, \sigma^2)$</td>
<td>168</td>
</tr>
<tr>
<td>B3</td>
<td>Estimates $a$ of $\alpha$ for $\Lambda(\mu, \sigma^2)$</td>
<td>170</td>
</tr>
<tr>
<td>B4</td>
<td>Estimates $b^2$ of $\beta^2$ for $\Lambda(\mu, \sigma^2)$</td>
<td>172</td>
</tr>
<tr>
<td>B5</td>
<td>Estimates $t$, $m$ and $s^2$ of $\tau$, $\mu$ and $\sigma^2$ by the method of quantiles for $\Lambda(\tau, \mu, \sigma^2)$</td>
<td>174</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

2.1 Frequency curves of the normal and lognormal distributions page 9
2.2 Frequency curves of the lognormal distribution for three values of $\sigma^2$ 10
2.3 Frequency curves of the lognormal distribution for three values of $\mu$ 10
2.4 Frequency curve of a negatively-skew lognormal distribution 16
2.5 Frequency curve of a four-parameter lognormal distribution 18
3.1 Kapteyn's analogue machine for generating a skew frequency curve facing page 23
4.1 Logarithmic probability graph for the data of Table 4.2 page 33
5.1 Efficiency of method of moments in estimation of $\mu$ and $\sigma^2$ 41
5.2 Efficiency of method of quantiles in estimation of $\mu$ and $\sigma^2$ 42
5.3 Efficiency of method of moments in estimation of $\alpha$ and $\beta^2$ 47
5.4 Efficiency of method of quantiles in estimation of $\alpha$ and $\beta^2$ 48
6.1 The functions $\theta(t)$ and $\phi(t)$ for a sample of size 64 57
6.2 Graph of $f(x)$ for Kemsgley's method of estimation 59
6.3 Estimation of $\tau$ from the probability graph 61
7.1 Toxicity of Rotenone to Macrosiphoniella sanborni 73
7.2 Logarithmic probability graph for Rotenone data 73
7.3 Region of convergence for Rotenone data 77
7.4 Standardized convergence regions for the four samples of Table 7.4 78
9.1 Efficiency of estimation of $\kappa$ and $\rho^2$ in a $\Delta$-distribution for two values of $\delta$ 98
11.1 The Lorenz diagram for a two-parameter distribution of incomes 112
11.2 The Lorenz diagram for a truncated two-parameter distribution of incomes 114
11.3 The Lorenz diagram for a three-parameter distribution of incomes 115
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td>Distribution of personal incomes in the United States: logarithmic probability graphs</td>
<td>117</td>
</tr>
<tr>
<td>11.5</td>
<td>Distribution of incomes in different types of household in the United States in 1947: logarithmic probability graphs</td>
<td>118</td>
</tr>
<tr>
<td>11.6</td>
<td>Earnings in nine male agricultural occupations in Great Britain, 1950: logarithmic probability graphs</td>
<td>119</td>
</tr>
<tr>
<td>12.1</td>
<td>Individual tolerance scheme for expenditure on a divisible commodity</td>
<td>126</td>
</tr>
<tr>
<td>12.2</td>
<td>Community tolerance scheme for expenditure on a divisible commodity</td>
<td>127</td>
</tr>
<tr>
<td>12.3</td>
<td>Marginal propensities to spend on six commodity groups: British industrial working-class households, 1937–8</td>
<td>128</td>
</tr>
<tr>
<td>12.4</td>
<td>Individual tolerance scheme for expenditure on an inferior good</td>
<td>128</td>
</tr>
<tr>
<td>12.5</td>
<td>Income elasticity plotted against saturation ratio $\phi/\kappa$</td>
<td>132</td>
</tr>
<tr>
<td>12.6</td>
<td>Engel curves fitted to British industrial working-class data for 1937–8</td>
<td>133</td>
</tr>
<tr>
<td>12.7</td>
<td>The sigmoid Engel curve and the problem of additivity</td>
<td>134</td>
</tr>
<tr>
<td>12.8</td>
<td>Engel curves for families of different composition</td>
<td>138</td>
</tr>
</tbody>
</table>
PREFACE

In economic data skew frequency curves are the rule rather than the exception. This is by no means an original observation, as we have made clear in our introductory chapter; but it remains true that comparatively little attention is paid to this point in most courses given to students of economic statistics. The result is that each generation must learn anew the lessons taught by Galton, Kapteyn, Gibrat, and many other investigators of the properties of the lognormal distribution. The present authors were led to make their own study as a result of their work on household survey material in the Department of Applied Economics, and from this point of view the most important aspect of the monograph is the discussion of the use of the probit transformation in economic contexts. Again, there seems to be an increasing interest in the study of economic variables which can be considered as realizations of a stochastic process; and we have therefore spent some time discussing the way in which such processes can lead to a lognormal distribution of the variate considered. As an important representative of this class we have treated the size distribution of personal incomes, but work published after the completion of this monograph by Hart and Prais on business concentration† and by Lane and Andrew on labour turnover‡ provides further evidence of the importance of the class and of the possibility of more advanced analysis in particular directions. It is our hope, then, that this monograph may assist in clearing the ground for further advances by a uniform and self-contained treatment of the more general problems of lognormal theory; and that, by our discussion of estimation procedure, more research workers will in future be encouraged to use them.

It is the authors' pleasure to record in this preface the generous help they received from many people at all stages in the preparation of this monograph. Our thanks are due first to Richard Stone who, as then Director of the Department of Applied Economics, encouraged us to pursue this study though we were led at times some distance from our own field. His help was always available and freely given. Next we would acknowledge a general debt of gratitude to our friends among past and present members of the Department and to the tradition of quantitative research in economics they have built there; on this tradition, and on the accumulated understanding of economic processes which grows with it, we have drawn freely. In particular we would thank S. J. Prais, whose interest in the study has been a stimulus;

a number of his valuable suggestions are incorporated in the present
text. J. A. C. Brown would like especially to thank his friend J. L.
Williams, for first interesting him, some seven years ago, in applications
of the lognormal distribution to survey material, and for many sub-
sequent discussions; and both authors join in thanking him for generously
making available his own notes and references.

For the large volume of computing work we are indebted to the
computing staff of the Department, especially to R. A. Arnould and to
Miss G. Smith, who together processed most of the artificial samples
and drew the figures. Again we wish to thank Dr M. V. Wilkes, F.R.S.,
Director of the Mathematical Laboratory of the University of Cam-
bridge, for extending the facilities of the EDSAC. We thank the secre-
tarial staff of the Department for the typing of the manuscript.

For collaboration in our experiment with graphical methods we thank
R. A. Arnould, E. B. Butler, G. D. Garton, J. L. Williams, Miss V. C.
Chapman, Miss G. Smith and Miss M. Potter. We further thank
E. B. Butler for drawing our attention to the data on agricultural
earnings, given in Chapter 11.

The text was read in page proof by F. E. A. Briggs, J. S. Cramer,
S. J. Prais, S. D. Silvey and R. Stone, each of whom made useful
corrections and suggestions. Miss E. M. A. Cracknell checked the tables
and references, and Miss D. G. Finding assisted in the preparation of
the page index.

For permission to reproduce some figures and results previously
published, the authors thank the editors of Biometrika, The Proceedings
of the Nutrition Society, Metroeconomica and The Review of Economic Studies. And
for their help with the references we thank the Librarians of the Institute
of Actuaries and of the Royal Statistical Society. Although some of the
tables in Appendix A contain functions which have been tabulated in a
number of other places, the authors have in fact calculated all the tables
ab initio. The single exception is Table A8, and the authors are indebted
to the editors of Biometrika for permission to reproduce this.

Finally we wish to state that the study has been in the fullest sense
a joint venture by the two authors, who together bear the responsibility
for its final realization.

J. AITCHISON
J. A. C. BROWN
CHAPTER I

INTRODUCTION

ANTIPHOLUS. Here comes the almanack of my true date.

What now? How chance thou art return'd so soon?

DRMIO. Return'd so soon! rather approach'd too late.

_The Comedy of Errors_

1.1. PURPOSE OF THE STUDY

The lognormal distribution in its simplest form may be defined as the distribution of a variate whose logarithm obeys the normal law of probability. Under a variety of names—the Galton-McAlister, Kapteyn or Gibrat distribution, the logarithmico-normal or simply the lognormal distribution—it has had a long though at times precarious career in the theory and application of statistics. Although not of so great an age as its sister distributions, the normal and the binomial, its origin may nevertheless be traced as far as 1879. Since then, knowledge of its theory and wide application has greatly increased; though in many ways it has remained the Cinderella of distributions, the interest of writers in the learned journals being curiously sporadic and that of the authors of statistical text-books but faintly aroused. Its literature is large but diffuse, and our first intention was merely to undertake a collation which was long overdue. It soon appeared, however, that there remained some unexplored problems and that some valuable properties of the distribution had been either overlooked or inadequately exploited. For those who are already familiar with the distribution, we hope that this monograph will prove a useful work of reference. We also hope that our efforts may establish the distribution as a powerful tool for those less acquainted with its possibilities.

This work may be regarded as an experiment in exposition in so far as it is devoted to the discussion of a single probability distribution. Some critics will ask if anything can be contained in such a study other than the well-known properties of the normal distribution and the logarithmic function. We hope to make it clear as the work proceeds that there is an adequate rejoinder to this criticism. Although it is in the nature of things that many of the properties of the lognormal may immediately be derived from those of the normal distribution, there are certain features of the former which differ from anything arising in normal theory. As examples of these we may cite here the concept of the moment distribution, the introduction of extra parameters and the particular difficulties which arise in regard to estimation procedures.

We may, indeed, go further and state our belief that the lognormal is as fundamental a distribution in statistics as is the normal, despite the stigma of the derivative nature of its name. It arises from a theory of elementary errors combined by a multiplicative process, just as the
normal distribution arises from a theory of elementary errors combined by addition. There are, as Galton long since pointed out, many situations in nature where it is more reasonable to suggest that the process underlying change or growth is multiplicative rather than additive. The problem here is formally similar to that of the choice of the geometric or the arithmetic mean as the more appropriate measure of location. Man has found addition an easier operation than multiplication, and so it is not surprising that an additive law of errors was the first to be formulated. Had man been more adept at multiplication the 'exponential-lognormal', or normal, might then have been the derivative distribution.

Many examples of the lognormal distribution have been noted in nature from a variety of fields ranging from sedimentary petrology to the analysis of literary style. Although the authors' interest in the subject derives mainly from its use in the analysis of economic data we discuss in Chapter 10 a number of other applications as they are recorded in the literature. The discussion of problems of more particular interest to the economist is contained in Chapter 11, where we deal with the statistical analysis of income distributions and measures of the concentration of income; and in Chapter 12, where we consider applications to econometric models of consumer demand. In the remaining chapters which are concerned with general properties and estimation procedures a bias towards any particular branch of science is avoided. It is hoped that the bibliography which is appended is sufficiently complete to provide a point of departure for the reader interested in a particular field.

1.2. History of the Lognormal Distribution

McAlister[142] appears to have been the first person to set down explicitly and in some detail a theory of the lognormal distribution. In his memoir presented to the Royal Society of London in 1879 he gave expressions for the mean, median, mode and the second moment of the distribution, together with the quartiles and octiles. In addition, he described a possible model of the genesis of the distribution from the chance combination of elementary errors and briefly demonstrated its properties of reproduction. McAlister's memoir was presented to the Royal Society by Francis Galton, to whom credit must go for suggesting the study. In his introductory remarks[78] Galton put forward the view that in certain cases the geometric mean is to be preferred to the arithmetic mean as a measure of location and significantly justified his point with a criticism of the normal theory of errors:

My purpose is to show that an assumption which lies at the basis of the well-known law of 'Frequency of Error' (commonly expressed by the formula $y = e^{-x^2}$) is incorrect in many groups of vital and social phenomena, although that law has been applied to them by statisticians with partial
success and corresponding convenience. Next, I will point out the correct hypothesis upon which a Law of Error suitable to these cases ought to be calculated....

Galton himself had derived his ideas from a consideration of the Weber-Fechner law relating responses to stimuli, the law which asserts that response is proportional to the logarithm of the stimulus. Both Weber[199] and Fechner[64] had, nearly half a century earlier, used the geometric mean in their practical investigations. Other writers, among them Scidell[177], Thiele[183] and Bruns[29], had also recommended its use in a variety of fields. Galton’s real contribution was to show that a preference for the geometric rather than the arithmetic mean must rest on a new assumption in the theory of errors. This argument was reiterated by Keynes[124] in 1911 and later in his book[125] in 1921.

To modern eyes McAlister’s treatment must inevitably seem naive and inadequate. Since his day the whole science of statistics has undergone a considerable transformation, and with this development knowledge of lognormal theory has been correspondingly enlarged. It is consistent with the precocious history of the distribution that it has been rediscovered with many of its properties several times.

The first advance† after McAlister’s initial paper was the contribution of Kapteyn[115], who in 1903 established more clearly the genesis of the distribution and indeed described a machine for generating samples from a lognormal population similar to that of Galton for binomial or normal populations. Kapteyn was joined by Van Uven[118] in 1916, and a method of estimation using quantiles was added to lognormal theory in their joint revision of Kapteyn’s earlier book; Van Uven continued and developed their work in 1917 in two papers[192,193]. About this period there was a heated correspondence between Kapteyn[116] and Pearson[159,160] on the uses of various frequency curves; the lognormal curve received severe criticism from Pearson who based his objections on a general mistrust of the technique of transformation. S. D. Wicksell[203] independently developed in 1917 a theory of genesis similar to that of Kapteyn and used the method of moments for estimation purposes; shortly afterwards Nydell[149], at the instigation of Wicksell, calculated the large-sample standard errors of the estimators obtained in this way.

Interest seems to have waned for a period after this, except for contributions by Davies[50,51] in 1925 and 1929 (when he dealt mainly with methods of estimation by quantiles), but revived again in 1930 when Gibrat presented his theory of the law‡ of proportionate effect, first in his paper[87] and then in greater detail in his book[88] published in

† Arne Fisher[68,69] mistakenly gives the credit for this to Jorgensen[113], whose contribution was not made until 1916.
‡ Gibrat’s law was foreshadowed by Weber in his theorem: *in observando discriminium rerum inter se comparatorum non differentiam rerum, sed rationem differentiae ad magnitudinem rerum inter se comparatorum percipimus.*
THE LOGNORMAL DISTRIBUTION

1931. His work was commented on and developed by D'Addario[1,2,3] and other Italian investigators of the distribution of incomes.

About the same time Gaddum[75], Bliss[20] and other workers, who were developing the probit method for the analysis of biological assays, became interested in the logarithmic transformation which they found effective in normalizing the distribution of levels of tolerance to the action of drugs on living organisms. The probit method and the logarithmic transformation in bioassay stem directly from the original researches of Weber and Fechner, as papers published in the early 1930's by Clark[38], Hemmingsen[104] and Bliss[21] testify. Also, in 1933, Yuan[216], who was interested in the study of anthropometric data, in particular the relation between height and weight, introduced the notions of a bivariate lognormal distribution and of a semi-logarithmic correlation surface.† There matters seem to have rested until Finney[65] in 1941 presented an efficient method of estimating the mean and variance of the simple lognormal distribution. The method of moments was again investigated by Quinsey[170] in 1945 and a censored lognormal distribution was discussed for the first time. Meanwhile Cochran[39] had suggested the use of the logarithmic transformation for its stabilization properties in the analysis of variance; a satisfactory theoretical foundation for this technique was provided by Curtiss[47] in 1943.

The extension of the simple two-parameter distribution to meet the case where a simple displacement of the variate rather than the variate itself is lognormally distributed was originally made by Wicksell[202] when dealing with the distribution of ages at first marriage. The estimation of the third parameter thus introduced, which, following the physiologists, we term the threshold‡ of the distribution, is one of the weaker points of existing lognormal theory. Most writers have been content with a priori methods, or, where prior information is not available, the method of moments, though in 1951 Cohen[42] considered the method of maximum likelihood and suggested an alternative. Wicksell suggested the possibility of introducing a fourth parameter, which, analogously with the threshold value, would fix an upper bound to the distribution; but it was not until 1949 that this extension was again considered by Johnson[111,112], who also presented methods of estimation.

A descriptive article by Gaddum[76,77] in 1945, which by virtue of its clarity renewed the interest of biologists, also deserves mention in this brief history. In this connexion Spiller[179] pointed out the applicability of truncated or censored lognormal distributions, and methods for handling these have been developed by Stevens[180] and by Thompson[184,185], who applied his theory to biological data which appeared in the form of discrete counts.

† Skew correlation had been treated earlier in 1926 by Van Vuren[194].
‡ Weber[199] in his chapter, 'de minus primum impressum aequi auit et factus cognoscendi' describes experimental attempts to determine the threshold of sensation, and refers to earlier work by Delezenne[57] on the least perceptible difference between two musical notes.
An independent line of development began in the late 1920's in small-particle statistics, and the lognormal distribution was investigated in this connexion by Hatch and Choate[102] and later by Krumbein[131]. In particular, Hatch[101], in an attempt to derive statistics of particle size from a knowledge of the proportions by weight of samples of dust passed through graded sieves, discovered the properties of the moment distributions of the lognormal for this special case. Later, Kolmogoroff[128], Epstein[61] and Halmos[98] explained the genesis of small-particle distributions by a breakage or grinding process.

1.3. Review of Problems Considered

This brief account of the history of lognormal theory in the last seventy-five years will suffice to show that most of its aspects have at one time or another been under review. Yet a closer study of the literature reveals a number of unresolved difficulties which, together with a lack of continuity in development, may account for the want of enthusiasm displayed by the authors of standard text-books. Of these difficulties the most important are to be found in connexion with estimation procedure; and these are both theoretical and practical in character. It is towards an appraisal of these problems that a substantial part of this monograph is directed.

For the two-parameter distribution the main issue in estimation theory would seem to be the relative merits of alternatives to the method of maximum likelihood. In the case of ungrouped data the method of maximum likelihood involves the taking of logarithms of the individual observations before the moments are computed, and this may be unduly laborious compared with the requirements of other methods; for grouped data an even more laborious iterative procedure is necessary. Methods using sample moments or quantiles, and more especially graphical methods, are much simpler to apply but necessarily result in a loss of information. When we have to consider a three- or four-parameter distribution or when additional complications arise, such as truncation, censorship or the occurrence of zero values, the difficulties of assessing the most appropriate method of estimation necessarily increase.

Since doubt about the efficiency of the simpler methods of estimation and the need to avoid heavy computing programmes must have caused many, who were aware of the desirable theoretical properties of the lognormal distribution, to avoid its use, we have approached the problem of estimation from a number of different points of view. Wherever possible we give explicit expressions for the efficiencies of the different methods, so that the practising statistician may choose between them with a knowledge of the quantity of information he may be sacrificing. In many cases it is possible to ease the burden of computation with the help of tables; a number of these are included at the end of this volume. We may also mention here that an experiment, using artificial samples
drawn from lognormal populations with known parameters, has been undertaken. This has served the purpose of providing some practical experience on which to base a critique of the different methods of estimation, in particular of those, such as the graphical, for which the experimental method alone can provide any measure of statistical efficiency. The results of the experiment are given in Chapters 5, 6 and 9.

With the advent of automatic high-speed computing machines it seems likely that the attitude of practical workers towards more sophisticated methods of analysis will be considerably changed in the coming years; for example, iterative procedures which are often troublesome when only desk-calculators are available become powerful tools of numerical analysis when programmed on an automatic computer. It seemed worth while therefore to describe in Chapter 13 the use we have made of one of these machines in processing the artificial samples and in calculating the tables, and to indicate the development of further programmes to facilitate the application of lognormal theory. The remaining chapters which deal with statistical methodology are Chapter 7, which describes the techniques of probit analysis, Chapter 8, where the comparison of population parameters is discussed, and Chapter 9, which covers the special difficulties of truncation, censorship and zero values. Chapters 10, 11 and 12 all deal with practical applications; the last two mentioned show in some detail the uses of lognormal theory in specifically economic contexts and no new statistical theory is introduced. The characteristics of the main lognormal distributions are set out in Chapter 2.

1.4. Notation

We conclude this introductory chapter with a few remarks on the principles on which we have based our notation. We follow the usual convention of distinguishing clearly between population functions or characteristics on the one hand and sample functions or estimators on the other. Greek letters invariably indicate the former while the corresponding Roman letters are used for the latter. Variates are denoted by the capital letters \( X, Y \) and \( Z \), while the corresponding lower case letters \( x, y \) and \( z \) stand for particular values or realizations, that is, sample values, of the variates; the expectation and variance of a variate \( X \) are written \( E\{X\} \) and \( D^2\{X\} \) respectively. Finally, \( P\{A\} \) denotes the probability of the event \( A \).

Our one departure from these rules is in our discussion of probit analysis. Here the use of the letters \( P \) and \( Z \) to denote the normal integral and ordinate respectively is so well established that we follow the same convention without, we hope, causing any confusion. Natural logarithms are used throughout: this is a mathematical convenience and avoids the introduction of scale factors.

† Other than estimators which are denoted by lower-case letters; for example \( \hat{\mu} \) is an estimator of \( \mu \).
CHAPTER 2
GENERAL PROPERTIES OF LOGNORMAL DISTRIBUTIONS

Quince. In the meantime I will draw a bill of properties, such as our play wants.
A Midsummer Night's Dream

2.1. INTRODUCTION

In this chapter we set out systematically the definitions and general statistical properties of various types of lognormal distributions. We begin with a discussion of a variate whose logarithm is distributed according to the normal law; this is the simplest case, and its study consists largely of an interplay of the mathematical properties of the logarithmic function and the statistical properties of the normal distribution. Even so, certain properties emerge which have no analogues in normal theory; the variate is essentially positive and its moment distributions, useful in many contexts, may then be defined and investigated. We next establish a multiplicative form of the central-limit theorem which will provide the basis for our discussion of the genesis of the distribution in Chapter 3.

As the scope of the definition is widened to admit the possibility of three and even four parameters it is to be expected that there will be a corresponding loss in simplicity. We conclude the chapter with some brief remarks on series representations of frequency functions which are approximately lognormal.

2.2. THE TWO-PARAMETER DISTRIBUTION: DEFINITION

Consider an essentially positive variate \( X \) \((0 < x < \infty)\) such that \( Y = \log X \) is normally distributed with mean \( \mu \) and variance \( \sigma^2 \). We then say that \( X \) is lognormally distributed or that \( X \) is a \( \Lambda \)-variate and write: \( X \) is \( \Lambda(\mu, \sigma^2) \) and correspondingly \( Y \) is \( N(\mu, \sigma^2) \). The distribution of \( X \) is completely specified by the two parameters \( \mu \) and \( \sigma^2 \), and this seems to be the simplest natural specification. It may be emphasized here that \( X \) cannot assume zero values, since the transformation \( Y = \log X \) is not defined for \( X = 0 \); this is a point of some importance and will be reconsidered in detail in Chapter 9. We shall use \( \Lambda(x \mid \mu, \sigma^2) \) and \( N(y \mid \mu, \sigma^2) \) to denote the distribution functions of \( X \) and \( Y \) respectively, so that

\[
\Lambda(x \mid \mu, \sigma^2) = P\{X \leq x\} \quad (2.1)
\]

and

\[
N(y \mid \mu, \sigma^2) = P\{Y \leq y\}. \quad (2.2)
\]

When there is no possibility of confusion we shall use the abbreviated forms \( \Lambda(x) \) and \( N(y) \) for the distribution functions.
2.3. MOMENTS AND OTHER CHARACTERISTICS

Since $X$ and $Y$ are connected by the relation $Y = \log X$ the distribution functions of $X$ and $Y$ are related by

$$\Lambda(x) = N(\log x) \quad (x > 0);$$  \hspace{1cm} (2.3)

hence

$$\Lambda(x) = 0 \quad (x \leq 0),$$  \hspace{1cm} (2.4)

and

$$d\Lambda(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\} dx \quad (x > 0),$$  \hspace{1cm} (2.5)

describes the frequency curve with a single mode at $x = e^{\mu - \sigma^2}$. The distribution possesses moments of any order; the $j$th moment about the origin is denoted by $\lambda_j$. Then

$$\lambda_j = \int_0^\infty x^j d\Lambda(x)$$

$$= \int_{-\infty}^\infty e^{\mu y} dN(y)$$

$$= e^{(\mu + j)e^\mu},$$  \hspace{1cm} (2.6)

from the properties of the moment-generating function of the normal distribution. The mean $\mu$ and variance $\sigma^2$ are therefore given by

$$\mu = e^{\mu + \frac{1}{2}\sigma^2}$$  \hspace{1cm} (2.7)

and

$$\sigma^2 = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$$  \hspace{1cm} (2.8)

where

$$\eta^2 = e^{\sigma^2} - 1.$$  \hspace{1cm} (2.9)

From (2.8) $\eta$ is seen to be the coefficient of variation† of the distribution. If two distributions have equal coefficients of variation they have equal values of the parameter $\sigma^2$ and conversely. The moments about the mean, denoted by $\lambda_j$, may readily be found from the $\lambda_j$. In particular, the third and fourth moments about the mean are

$$\lambda_3 = \alpha^3(\eta^6 + 3\eta^4)$$  \hspace{1cm} (2.10)

and

$$\lambda_4 = \alpha^4(\eta^8 + 6\eta^{10} + 15\eta^8 + 16\eta^6 + 3\eta^4),$$  \hspace{1cm} (2.11)

respectively, so that the measures of departure from normality, namely, the coefficient of skewness $\gamma_1(X)$ and the coefficient of kurtosis $\gamma_2(X)$, are given by

$$\gamma_1(X) = \frac{\lambda_3}{\sigma^3} = \eta^3 + 3\eta$$  \hspace{1cm} (2.12)

and

$$\gamma_2(X) = \frac{\lambda_4}{\sigma^4} - 3 = \eta^8 + 6\eta^4 + 15\eta^4 + 16\eta^2.$$  \hspace{1cm} (2.13)

† A tabulation of $\eta$ against $\sigma$ is given in Appendix Table A1.
It is clear that the distribution is positively skew and that the greater the value of $\sigma^2$ the greater is the skewness. Also the distribution has positive kurtosis; again the kurtosis increases as $\sigma^2$ increases.

The median of the distribution is at $x = e^\mu$. The relative positions of the mean, median and mode, namely, at $x = e^{\mu + \sigma^2}$, $e^\mu$ and $e^{\mu - \sigma^2}$ respectively, again emphasize the positive skewness of the distribution. A simple relation obtains between the quantiles of $\Lambda(\mu, \sigma^2)$ and the corresponding quantiles of $N(0, 1)$. For, if $\xi_q$ and $\nu_q$ are the quantiles of order $q$ of $\Lambda(\mu, \sigma^2)$ and of $N(0, 1)$ respectively, then

$$\xi_q = e^{\mu + \nu_q}.$$  \hfill (2.14)

For example, the lower and upper quartiles are $x = e^\mu - 0.675\sigma$ and $x = e^\mu + 0.675\sigma$ respectively.

Fig. 2.1 gives a comparison of the frequency curves of the $N(0, 0.5)$ and $\Lambda(0, 0.5)$ distributions showing the relative positions of the mean, median and mode for the $\Lambda$-distribution. Fig. 2.2 shows the frequency curves for $\Lambda(0, 0.5)$, $\Lambda(0, 0.5)$ and $\Lambda(0, 2.0)$, from which an idea of the flexibility of the distribution may be obtained. And finally Fig. 2.3 compares the frequency curve for $\Lambda(0, 0.5)$ with that for $\Lambda(0.5, 0.5)$ and that for $\Lambda(1.0, 0.5)$.

### 2.4. Reproductive Properties

The two-parameter lognormal distribution possesses a number of interesting reproductive properties, most of which are immediate

† Appendix Table A1 tabulates the coefficients of skewness and kurtosis, and the ratios of mean to median and mean to mode against $\sigma$. 

† Appendix Table A1 tabulates the coefficients of skewness and kurtosis, and the ratios of mean to median and mean to mode against $\sigma$. 

23
consequences of those for the normal distribution. Since the latter has
additive reproductive properties it is to be expected, from the character-
istic property of the logarithmic function \( \log X_1 + \log X_2 = \log X_1 X_2 \),

\[
\sigma^2 = 0.1
\]

\[
\sigma^2 = 0.5
\]

\[
\sigma^2 = 2
\]

0 1 2 3 4

Fig. 2.2. Frequency curves of the lognormal distribution for three values of \( \sigma^2 \).

that the lognormal distribution will have multiplicative reproductive
properties. This is in fact the case.

Before investigating these properties we note here the simple result:
if \( X \) is \( \Lambda(\mu, \sigma^2) \) then \( 1/X \) is \( \Lambda(-\mu, \sigma^2) \). More generally, there is the theorem:
THEOREM 2.1

If \(X \sim \Lambda(\mu, \sigma^2)\) and \(b\) and \(c\) are constants, where \(c > 0\) (say \(c = e^c\)), then \(cX^b\) is \(\Lambda(a + b\mu, b^2\sigma^2)\).

The simple reproductive property is then contained in the following theorem:

THEOREM 2.2

If \(X_1\) and \(X_2\) are independent \(\Lambda\)-variates, then the product \(X_1X_2\) is also a \(\Lambda\)-variate.

More specifically, if \(X_1 \sim \Lambda(\mu_1, \sigma_1^2)\) and \(X_2 \sim \Lambda(\mu_2, \sigma_2^2)\), then \(X_1X_2 \sim \Lambda(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)\). This result may be expressed in terms of distribution functions, giving a convolution property for the lognormal integral corresponding to that for the normal integral (cf. Cramer [46] p. 190):

COROLLARY 2.2a

\[
\int_0^\infty \Lambda(a | \mu_1, \sigma_1^2) d\Lambda(x | \mu_2, \sigma_2^2) = \Lambda(a | \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).
\]

A slightly more general form of this relation which will be used in Chapters 11 and 12 is

COROLLARY 2.2b

\[
\int_0^\infty \Lambda(ax^b | \mu_1, \sigma_1^2) d\Lambda(x | \mu_2, \sigma_2^2) = \Lambda(a | \mu_1 - b\mu_2, \sigma_1^2 + b^2\sigma_2^2).
\]

The reproductive property clearly extends to any finite set of independent \(\Lambda\)-variates and also to an infinite sequence provided that some conditions of convergence are fulfilled. By combining this extension with Theorem 2.1 the general result follows:

THEOREM 2.3

If \(\{X_j\}\) is a sequence of independent \(\Lambda\)-variates, where \(X_j \sim \Lambda(\mu_j, \sigma_j^2)\), \(\{b_j\}\) a sequence of constants and \(c = e^c\) a positive constant, then provided \(\sum b_j\mu_j, \sum b_j^2\sigma_j^2\) both converge the product \(c \prod_j X_j\) is \(\Lambda(a + \sum b_j\mu_j, \sum b_j^2\sigma_j^2)\).

In particular, the ratio \(X_1/X_2\) is \(\Lambda(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)\) and we have the important corollary:

COROLLARY 2.3

If \(X_j\) (\(j = 1, \ldots, n\)) are independent \(\Lambda\)-variates with the same parameters \(\mu\) and \(\sigma^2\), their geometric mean \(\left(\prod_{j=1}^n X_j\right)^{1/n}\) is \(\Lambda(\mu, \sigma^2/n)\).

A similar result holds if the \(X_j\) (\(j = 1, \ldots, n\)) are not distributed independently but have a multivariate lognormal distribution. The random column vector \(X = \{X_1 \ldots X_n\}\), has a multivariate \(\Lambda\)-distribution if the transformed vector \(Y = \{Y_1 \ldots Y_n\}\), where \(Y_j = \log X_j\), has a multivariate normal distribution; let us say with \(E(Y) = \mu\) and the variance matrix of \(Y\) equal to \(V\).
THEOREM 2.4

If X is multivariate lognormal and b is a (column) vector of constants with transpose b', then the product \( c = \prod_{j=1}^{n} X_j \) is \( \Lambda(a + b' \mu, b' \Sigma b) \), where \( c = e^a \) is a positive constant.

To Cramér's theorem on the normal distribution there corresponds the following:

THEOREM 2.5

If \( X_1 \) and \( X_2 \) are two independent positive variates such that their product \( X_1 X_2 \) is a \( \Lambda \)-variate, then both \( X_1 \) and \( X_2 \) are \( \Lambda \)-variates (or, as a special case, one of the variates may be constant and the other a \( \Lambda \)-variate).

This is a converse of the reproductive property of Theorem 2.2 and may be extended to the case of a finite number of independent positive variates, but not to an infinite sequence as is evident from a consideration of § 2.6. Levy's corollary to Cramér's theorem may also be reframed to apply to \( \Lambda \)-distributions.

2.5. MOMENT DISTRIBUTIONS: GINI'S COEFFICIENT OF MEAN DIFFERENCE

The property now to be discussed has no analogue in normal theory since it involves the concept of moment distributions which may be defined meaningfully for positive variates only. This concept will be found important in many practical applications and will be discussed at greater length in Chapters 11 and 12. The \( j \)-th moment distribution function of \( \Lambda(\mu, \sigma^2) \) is defined by

\[
\Lambda_j(x \mid \mu, \sigma^2) = \frac{1}{\Lambda_j} \int_0^x u^j d\Lambda(u \mid \mu, \sigma^2),
\]

and the fundamental theorem of the moment distributions is

THEOREM 2.6

The \( j \)-th moment distribution of a \( \Lambda \)-distribution with parameters \( \mu \) and \( \sigma^2 \) is also a \( \Lambda \)-distribution with parameters \( \mu + j \sigma^2 \) and \( \sigma^2 \) respectively.

Proof

\[
\Lambda_j(x \mid \mu, \sigma^2) = \frac{1}{\Lambda_j} \int_0^x u^j d\Lambda(u \mid \mu, \sigma^2)
= e^{-j \sigma^2 \mu} \int_0^x e^{j \mu u} \frac{1}{u \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (\log u - \mu)^2 \right\} du
= \int_0^x \frac{1}{u \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (\log u - \mu - j \sigma^2)^2 \right\} du
= \Lambda(x \mid \mu + j \sigma^2, \sigma^2),
\]

using (2.5) and (2.6).

This simple result allows us to obtain an explicit expression for Gini's coefficient of mean difference:
The coefficient of mean difference of $A(u, q^2)$ is $G = 2N\left(\frac{\sigma}{\sqrt{2}}, 0, 1\right) - 1$.

Proof

$$G = \int_0^\infty \int_0^\infty |u - v| dA(u) dA(v)$$

$$= \int_0^\infty \int_0^u (u - v) dA(u) dA(v) + \int_0^\infty \int_v^\infty (v - u) dA(u) dA(v)$$

$$= 2 \int_0^\infty uA(u) dA(u) - 2\int_0^\infty v dA(u) dA(v)$$

$$= 2a \int_0^\infty A(u) dA(u)$$

where $a$ is defined by (2.7).

Using Theorem 2.6 we have then

$$G = 2a \int_0^\infty A(u |\mu, \sigma^2) dA(u |\mu + \sigma^2, \sigma^2)$$

$$- 2a \int_0^\infty A(1 |\sigma^2, 2\sigma^2) - 1$$

from Corollary 2.2b.

Thus

$$G = 2a \left[ 2a \left( - \sigma^2, 2\sigma^2 - 1 \right) \right]$$

$$= 2a \left( 2N\left(\frac{\sigma}{\sqrt{2}}, 0, 1\right) - 1 \right)$$

2.6 CENTRAL LIMIT THEOREMS

We shall confine ourselves in this section to a statement of the multiplicative analogues first for the Lindeberg-Levy[140, 136] and secondly for the Liapounoff[138, 139] forms of the additive central limit theorem.

THEOREM 2.8

If $\{X_j\}$ is a sequence of independent, positive variates having the same probability distribution and such that

$$E[\log X_j] = \mu$$

and

$$D^2[\log X_j] = \sigma^2$$

both exist, then the product $\prod_{j=1}^\infty X_j$ is asymptotically distributed as $A(\sqrt{n} \mu, \sqrt{n} \sigma^2)$.

This implies that the geometric mean $\left(\prod_{j=1}^\infty X_j\right)^{1/n}$ is asymptotically
distributed as \( \Lambda(\mu, \sigma^2/n) \). For example, suppose that each \( X_i \) is distributed rectangularly in the interval \( 0 < x_i < 1 \); then \( E(\log X_i) = -1 \) and \( D^2(\log X_i) = 1 \) so that the geometric mean is asymptotically \( \Lambda(-1, 1/n) \).

**Theorem 2.9**

Let \( \{X_i\} \) be a sequence of independent, positive variates such that

\[
E(\log X_i) = \mu_i,
\]

\[
D^2(\log X_i) = \sigma_i^2,
\]

and

\[
E(\| \log X_i - \mu_i \|^2) = \omega_i^2
\]

all exist for every \( i \). Then if

\[
\mu_{(n)} = \sum_{j=1}^{n} \mu_i,
\]

\[
\sigma_i^2 = \sum_{j=1}^{n} \sigma_i^2
\]

and

\[
\omega_{(n)} = \sum_{j=1}^{n} \omega_i^2
\]

the product \( \prod_{j=1}^{n} X_i \) is asymptotically distributed as \( \Lambda(\mu_{(n)}, \sigma_{(n)}^2) \), provided that \( \omega_{(n)}/\sigma_{(n)} \to 0 \) as \( n \to \infty \).

### 2.7. The Three-Parameter Distribution: Definition

In this section the definition and scope of the distribution are extended by the introduction of a third parameter. We are here concerned with a variate \( X \) such that a simple displacement of \( X \), say \( X' = X - \tau \), and not the variate itself, is \( \Lambda(\mu, \sigma^2) \). The range of \( X \) is thus \( \tau < x < \infty \), and we write \( X \) is \( \Lambda(\tau, \mu, \sigma^2) \). The two-parameter distribution is then the special case for which \( \tau = 0 \) and no confusion results from using the contracted notation \( \Lambda(\mu, \sigma^2) \) for \( \Lambda(0, \mu, \sigma^2) \). Since the parameter defines a lower bound to the range of values of the variate \( X \) it will be termed the threshold of the distribution and the distribution function will be denoted by \( \Lambda(x | \tau, \mu, \sigma^2) \).

In certain circumstances the value of the threshold may be determined on \textit{a priori} grounds, and so is not to be regarded as an unknown parameter which requires to be estimated. If this is so the variate \( X' = X - \tau \) may be considered in place of \( X \); when a value of \( X \) is given, the corresponding value of \( X' \) is immediately known. The variate \( X' \) has all the properties of the two-parameter variate and no new theory arises.

On the other hand, \( \tau \) may be an unknown parameter of the distribution of \( X \) to be estimated from sample data; although \( X - \tau \) still has the properties of the two-parameter variate, \( \tau \) is not known exactly and estimation procedures developed for the two-parameter case are not directly applicable to the distribution of \( X - \tau \). We are therefore forced
2.8. Moments, other Characteristics, and Moment Distributions

The distribution function is given by

$$\Lambda(x | \tau, \mu, \sigma^2) = 0 \quad (x \leq \tau) \quad (2.16)$$

and

$$\Lambda(x | \tau, \mu, \sigma^2) = \Lambda(x - \tau | \mu, \sigma^2) \quad (x > \tau), \quad (2.17)$$

so that the frequency curve is that of the $\Lambda(\mu, \sigma^2)$ distribution displaced by $\tau$. The location characteristics are therefore each increased by $\tau$: the mean being at $x = \tau + \alpha$ where $\alpha$ is defined by (2.7) as before; the median at $x = \tau + e^\alpha$; and the mode at $x = \tau + e^{\alpha - 1}$. The quantiles are displaced from $\xi_q$ to $\tau + \xi_q$. The moments about $\tau$ are

$$E[(X - \tau)^j] = e^{j\mu + j\alpha \sigma^2}, \quad (2.18)$$

so that the moments about the mean and hence the measures of departure from normality remain unchanged. The coefficient of variation $\eta'$ is found to be

$$\eta' = \frac{\eta}{1 + \tau/\alpha}. \quad (2.19)$$

For $\tau > 0$ moment distributions may be defined in a way similar to that of §2.5. A relatively simple result is found for the first moment distribution only, which is, however, the most important.

**Theorem 2.10**

The first moment distribution of $\Lambda(\tau, \mu, \sigma^2)$, where $\tau > 0$, has distribution function $\Lambda_1(x | \tau, \mu, \sigma^2)$ given by

$$\Lambda_1(x | \tau, \mu, \sigma^2) = \tau \Lambda(x | \tau, \mu, \sigma^2) + \alpha \Lambda(x - \tau | \mu, \sigma^2) \quad (\tau + \alpha).$$

Proof

$$\Lambda_1(x | \tau, \mu, \sigma^2) = \frac{1}{\tau + \alpha} \int_\tau^{\infty} u d\Lambda(u | \tau, \mu, \sigma^2)$$

$$= \frac{1}{\tau + \alpha} \int_{\tau}^{\infty} (\tau + u) d\Lambda(u | \mu, \sigma^2)$$

$$= \frac{1}{\tau + \alpha} \left\{ \tau \Lambda(x - \tau | \mu, \sigma^2) + \alpha \Lambda_1(x - \tau | \mu, \sigma^2) \right\},$$

from (2.15). Hence

$$\Lambda_1(x | \tau, \mu, \sigma^2) = \frac{\tau \Lambda(x | \tau, \mu, \sigma^2) + \alpha \Lambda(x | \tau, \mu + \sigma^2, \sigma^2)}{\tau + \alpha} \quad (2.17)$$

and Theorem 2.6.

Because of the simple displacement of the frequency curve it follows immediately that the coefficient of mean difference for the three-
parameter distribution is given by the same formula as that for the two-parameter distribution. This result holds for all values of \( \tau \).

**Theorem 2.11**

The coefficient of mean difference of \( \Lambda(\tau, \mu, \sigma^2) \) is

\[
G = 2 \alpha \left\{ 2N \left( \frac{\sigma}{\sqrt{2}}, 0, 1 \right) - 1 \right\}.
\]

### 2.9. Negatively Skew Distributions

The lognormal distributions already discussed have been positively skew, but a slight modification of the three-parameter case provides the possibility of treating negatively skew distributions\(^{169}\). These have an upper bound \( \theta \) and the variate is restricted to the range \(-\infty < x < \theta\); indeed, we consider a variate \( X \) such that \( \theta - X \) is \( \Lambda(\mu, \sigma^2) \). The frequency curve of this distribution is the mirror image about the line \( x = \theta \) of the frequency curve of \( \Lambda(0, \mu, \sigma^2) \). The fact that the range is unbounded below may appear to be a drawback in circumstances where the variate should clearly be confined to positive values,\(^\dagger\) but it may well be a justifiable approximation if the fitted distribution is such that \( P(X < 0) \) is negligible. The same kind of conceptual error is committed when the range \( 0 < X < \infty \) is postulated though large values of \( X \) are inconceivable; here again we take refuge in the fact that for the fitted distribution the probability of these values is negligible. In either case the drawback may be overcome by imposing lower and upper bounds to the range, giving a four-parameter distribution.

In Fig. 2.4 we give the frequency curve for the negatively skew distribution for which \( \mu = 0, \sigma = 0.5 \) and \( \theta = 3 \) and for comparison the frequency curve of \( \Lambda(0, 0.5) \).

\(^{169}\) See also the remarks of Bernstein and Weatherall\(^{19}\).

\(^\dagger\) See also the remarks of Bernstein and Weatherall\(^{19}\).
2.10. A Four-parameter Distribution

As indicated in §2.9 it is possible to introduce an extension of the lognormal distribution to allow for both a lower and an upper bound to the possible values of the variate. Johnson [111,112] has recently suggested the use of such a distribution originally proposed by Wicksell [203]. The variate \( X \) is now confined to the range \( r < x < \theta \) and we suppose \( X' = (X - \tau)/\theta - X \) is \( \Lambda(\mu, \sigma^2) \). The transformation admits the treatment, with certain limitations specified by Johnson, of both positively and negatively skew curves. Johnson has obtained an explicit though complicated expression for the first moment of \( X \), and gives a method of calculating higher moments for given values of the parameters. The quantile \( \xi_q \) of order \( q \) may be readily obtained by equating \( (\xi_q - \tau)/(\theta - \xi_q) \) to the corresponding quantile of \( \Lambda(\mu, \sigma^2) \), namely, \( e^{\mu + q\sigma} \). The fact that the range is bounded both below and above is theoretically attractive because there may be strong a priori grounds for believing in the existence of these bounds; even in such cases, however, it may be advantageous to discard one of the bounds and use a three-parameter distribution in the hope that what one will lose in degree of approximation one will
THE LOGNORMAL DISTRIBUTION

gain in more tractable analysis. If both $\tau$ and $\theta$ are known from the conditions of the experiment the theory of the two-parameter distribution is again directly applicable to $X'$. A brief discussion of estimation is contained in §6.5.

As an example of a four-parameter lognormal distribution we give in Fig. 2.5 the frequency curve for the case $\mu=0$, $\sigma^2=0.5$, $\tau=0.5$, $\theta=5.0$ and compare it with the frequency curve of $X(0,0.5)$.

2.11. GENERALIZATIONS BY SERIES REPRESENTATIONS

In the preceding sections generalizations of the simple lognormal distribution have been considered by the introduction of extra parameters; in this section we discuss briefly two other possible generalizations applicable to distributions which approximate to the two-parameter case. In the first a series representation of the frequency function is given in terms of the orthogonal polynomials associated with the lognormal distribution; in the second we treat the logarithm of the variate as approximately normal and investigate the consequences of representing the frequency function of the transformed variate by a Gram-Charlier A-series or an Edgeworth series, that is in terms of Hermite polynomials. These methods have been popular with Scandinavian writers[34,36,69,113,170].

It is not difficult[46,182] to determine the orthogonal polynomials $p_j(x)$ of any degree $j$ associated with the lognormal distribution $X(\mu, \sigma^2)$. Then

\[ \int_0^\infty p_i(x)p_j(x)d\Lambda(x) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases} \tag{2.20} \]

If $X_1 (0 < x_1 < \infty)$ is approximately $X(\mu, \sigma^2)$ with distribution function $X^* (x)$ and we write $X_1 = \log X_1$ is approximately $X(\mu, \sigma^2)$ by expressing the frequency function of $X_1$ as a series representation involving Hermite polynomials. If $h_j(y)$ is the Hermite polynomial of degree $j$ and $N^*(y)$ is the distribution function of $Y_1$ we write

\[ dN^*(y) = \sum_{j=0}^\infty c_j h_j(y) dN(y), \tag{2.23} \]

where now

\[ c_j = \int_0^\infty h_j(y) dN^*(y) \tag{2.24} \]
(this is the Gram-Charlier A-series for \(N^*(y)\)). Again the \(c_j\) may be expressed in terms of the moments of \(Y_1\). The approach as it stands provides us neither with a means of estimating \(\mu, \sigma^2\) and the \(c_j\) nor with a means of deciding how many terms of the series are needed. This may be more clearly seen from the fact that the moments of \(X_1\) about the origin are given by
\[
E[X_1^j] = \int_0^\infty x^j dN^*(x) = e^{\int \mu + \frac{1}{2} \sigma^2 x^2 + \sum_{i=0}^\infty c_i (\sigma_j)^i / \sqrt{i!}},
\]
so that all moments depend on all the \(c_i\), making estimation difficult.

These difficulties may be largely overcome by considering an Edgeworth rather than a Gram-Charlier expansion. Let the cumulants of \(N^*(y)\) be \(\kappa_1, \kappa_2, \ldots\), so that
\[
\int_0^\infty e^{jy} dN^*(y) = \exp \left\{ \sum_{i=1}^\infty \kappa_i \frac{j^i}{i!} \right\},
\]
then
\[
E[X_1^j] = \int_0^\infty x^j dN^*(x) = \exp \left\{ \sum_{i=1}^\infty \kappa_i \frac{j^i}{i!} \right\},
\]
which is formally equivalent to (2.25). The fact that
\[
\Delta_j = \sum_{i=1}^j (-1)^{i+1} \binom{j}{i} \log E[X_1^j]
\]
depends only on \(\kappa_j\) and higher cumulants is useful for deciding how many terms of the series should be retained.\(^1\)

\(^1\) The Hermite polynomials \(h_i(y)\) associated with \(N(\mu, \sigma^2)\) are defined by
\[
h_i(y) = \frac{1}{\sqrt{i!}} H_i \left( \frac{y-\mu}{\sigma} \right)
\]
where
\[
\frac{d^i}{dy^i} e^{-y^2} = (-1)^i H_i(y) e^{-y^2}.
\]
The orthogonal relation is then
\[
\int_{-\infty}^\infty h_i(y) h_j(y) dN(y) = 0 (i \neq j)
\]
\[
= 1 (i = j).
\]
CHAPTER 3

THE GENESIS OF LOGNORMAL DISTRIBUTIONS

ArgoEon. Yet, that the world may witness that my end
Was wrought by nature, not by vile offence.
_The Comedy of Errors_

3.1. REASONS FOR CONSIDERING THE GENESIS OF A DISTRIBUTION

It has occasionally been argued that the success with which any particular frequency curve graduates empirical data is a sufficient criterion of its worth. There are, indeed, many occasions on which such an argument may be satisfactory. Yet there are at least two important reasons for seeking a more fundamental basis in the theory of probability for any system of frequency curves to which we attach a more than transient importance. First by providing such a basis we may obtain a clearer insight into underlying natural or sociological processes; this in its turn will often suggest a wider application of the system. Secondly, a knowledge of the elementary assumptions from which the law of frequency may be derived will enable us more easily to modify the law to meet new circumstances. For these two reasons alone it may often be more satisfactory to use a system of frequency curves for which there is a plausible basis than one which is more successful in graduating the sample observations immediately to hand.

It will be recalled that the normal curve was originally derived by Laplace from the mathematical theory of probability. Nevertheless, the discussion of alternative models of generation for this case continues. Less attention has as yet been given to the genesis of lognormal distributions. In the following sections we give an account of the most important derivations so far put forward, and suggest certain new lines of development.

3.2. KAPTEYN’S SYSTEM OF TRANSLATION AND PEARSON’S CRITICISM

In his book, _Skew Frequency Curves in Biology and Statistics_, Kapteyn[115] in 1903 laid the foundations of a theory for the generation of an extensive system of frequency curves and made the claim:

The main advantages of the theory are considered to be the following:

(a) It assigns the connection between the form of the curves and the action of the causes to which the form is due. The knowledge of the connection may lead us, at least in some cases, to precious indications about the nature of true causes.
It enables us to reduce the consideration of any skew curve to that of a normal curve.

The extreme simplicity of the application which, in most cases, makes the derivation of the constants of the curve from the observations hardly more difficult than in the case of the normal curve.

The lognormal curve arises as a special case of this theory and is, in Kapteyn's words, 'one of the most important classes occurring in nature'.

Earlier, in 1895, Pearson[158] had evolved his system of curves and, in 1906, Charlier[34] suggested his method of series representations of frequency functions. Neither of these writers made any strong claim for their systems as derivations from possible underlying natural causes, although Pearson showed that his own types of curves might arise as limiting forms of the hypergeometric distribution; rather was the emphasis laid on the curve-fitting properties of the systems.

Kapteyn[116] and Pearson[159, 160] engaged in a lively, and at times bitter, controversy on the relative merits of their approaches. First of all, Pearson rejected the lognormal distribution proposed by Galton and McAlister on the grounds that it did not occur in nature as Galton had suggested; a few experiments, not fully specified, over a limited type of observation seemed to him ample evidence for this rejection. The wider system of Kapteyn was then refuted on several grounds. First, Pearson claimed, it was so general as to lead to a mere tautology; secondly, it was not suitable for graduation purposes; thirdly, the lognormal curve, used by Kapteyn in some examples, was of limited skewness, this criticism being based on a miscalculation of the possible values of Pearsonian skewness which the curve can assume. In the context of our present discussion on the generation of frequency curves Pearson's most fundamental objection is contained in the following remarks:

What is \( x \) [the transformed variate] of which the observed character \( X \) is a function? Is it, as in the explanatory illustrations cited by Kapteyn, another characteristic of the organism? If so we ought in some cases to be able to determine it. What is the character which obeys the normal law?... Supposing, as in English female crania, nasal breadth is asymmetrical, what is the quantity which is symmetrically distributed of which nasal breadth is a function? It has no reality in the organism at all....

This argument clearly reveals that Pearson misunderstood the purpose of Kapteyn's theory, and his insistent demand for a physical correlate of the transformed variate is founded on a naïve conception of the role played by the normal distribution in the same theory.†

In reply Kapteyn substantially maintained his original position. It is not appropriate here to pursue this controversy further, since no new arguments of importance were added on either side. Kapteyn's formulation has antecedents in the Method of Translation treated by

† There are apparently still some adherents of Pearson's point of view! We have recently heard of an American candidate whose thesis was objected to on the grounds that the examiners were not interested in the logarithm of income.
Edgeworth [26] in 1898 and in the original work of Galton and McAlister. It was again put forward independently by Wicksell [203] in 1917; and later Gibrat [87] in 1930 discussed the lognormal case in relation to the law of proportionate effect and used an argument of Kapteyn for arriving at the normal distribution, unaware of the writer’s extension of his argument. More recently the method of translation has been reconsidered in great detail by Johnson [111], and by Draper [59].

The central-limit theorems given in the previous chapter aim to establish the conditions under which a variate defined as the product of a number of elementary variates itself tends to be lognormally distributed, and are the groundwork on which all existing theories of the genesis of lognormal distributions have been erected. We now turn to a more detailed discussion of Kapteyn’s theory and to a description of the analogue machine constructed to his design in the Botanical Laboratory of the University of Groningen.

3.3. THE THEORY OF PROPORTIONATE EFFECT

We shall consider here a positive variate which is the outcome of a discrete random process. Most writers have conceived the process as taking place at successive points of time, as may be the case with a process of biological growth. Thus Cramér [46] states:

If our random variable is the size of some specified organ that we are observing, the actual size of this organ in a particular individual may often be regarded as the joint effect of a large number of mutually independent causes, acting in an ordered sequence during the time of growth of the individual.

In our view, the ordering of the sequence through time is not an essential feature of the process: this is a point to which we return in §3.5.

Meanwhile suppose that the variate is initially $X_0$ and that after the $j$th step in the process it is $X_j$, reaching its final value $X_n$ after $n$ steps. The general case considered by Kapteyn is the following. At the $j$th step the change in the variate is a random proportion of a function of the value already attained; thus

$$X_j - X_{j-1} = c_j \phi(X_{j-1}),$$

where the set $\{c_j\}$ is mutually independent and also independent of the set $\{X_j\}$. The change at any step is therefore not independent of the value attained except for the case $\phi(X) = 1$ or for the trivial case $\phi(X) = 0$.

We are interested here in the important special case $\phi(X) = X$, that is to say where the change in the variate is a random proportion of the momentary value of the variate. The law of proportionate effect then is:

A variate subject to a process of change is said to obey the law of proportionate effect if the change in the variate at any step of the process is a random proportion of the previous value of the variate.

† Another approach is that of Haldane [96], who shows that, on certain assumptions, the $n$th power of a normal variate tends to a lognormal variate as $n$ increases.
Fig. 3.4. Kaperon's analogue machine for generating a skew frequency curve.

(facing p. 22)
For this case (3.1) reduces to

\[ X_j - X_{j-1} = \epsilon_j X_{j-1}. \]  

(3.2)

The importance of the law is embodied in Theorem 3.1, but before this theorem is stated in a rigorous form the following heuristic treatment may show the link with the additive form of the central limit theorem. We may rewrite (3.2) as

\[ \frac{X_j - X_{j-1}}{X_{j-1}} = \epsilon_j, \]  

(3.3)

so that

\[ \sum_{j=1}^{n} \frac{X_j - X_{j-1}}{X_{j-1}} = \sum_{j=1}^{n} \epsilon_j. \]  

(3.4)

Now, supposing the effect at each step to be small,

\[ \log X \sim J, \]  

(3.5)

giving

\[ \log X = \log X_0 + \epsilon_1 + \epsilon_2 + \ldots + \epsilon_n. \]  

(3.6)

By the additive form of the central-limit theorem \( \log X \) is asymptotically normally distributed and hence \( X \) is asymptotically lognormally distributed in a two-parameter form.

**Theorem 3.1**

A variate subject to the law of proportionate effect tends, for large \( n \), to be distributed as a two-parameter \( \Lambda \)-variate, provided that the sequence \( X_0, 1 + \epsilon_1, 1 + \epsilon_2, \ldots \) satisfies the conditions of Theorem 2.8 or of Theorem 2.9.

**Proof**

From (3.2)

\[ X_j = (1 + \epsilon_j) X_{j-1}, \]  

so that

\[ X_n = X_0 (1 + \epsilon_1) \ldots (1 + \epsilon_n), \]  

and the result follows from Theorem 2.8 or from Theorem 2.9.

### 3.4. Kapteyn's Analogue Machine

With the intention of convincing sceptics that skew-frequency curves could arise from natural causes Kapteyn had already built, before he published his 1903 paper, an analogue machine on the lines of Galton's apparatus for demonstrating the binomial and normal distributions. Kapteyn's machine is based on the generating model described in §3.3, namely,

\[ X_j = X_{j-1} (1 + \epsilon_j) \ (j = 1, \ldots, n), \]  

(3.7)

where the random variable \( \epsilon_j \) is simply specified by

\[ P(\epsilon_j = a) = \frac{1}{2} \]  

(3.8)

and

\[ P(\epsilon_j = -a) = \frac{1}{2}, \]  

(3.9)

for all \( j \), where \( a \) is a positive constant. The machine, of which we show a photograph in Fig. 3.1, consists of nine rows of \( \bigwedge \)-shaped wedges attached to a wood and glass frame 10.4 cm. high. The wedges are of
varying breadth, the breadth being proportioned to the distance of the vertex of the wedge from the left-hand side of the frame; if \( X_{j-1} \) denotes the distance of a vertex from the left-hand side of the frame, the breadth of the wedge is \( 2aX_{j-1} \).

Sand is poured into a funnel situated at the top of the frame directly above the middle wedge in the top row; on arriving at the point \( X_{j-1} \) the sand is divided into two equal parts by the wedge and is displaced either to \( X_{j-1}(1+a) \) or to \( X_{j-1}(1-a) \), in both cases arriving at the vertex of a wedge in the next lower row. The sand finally arriving in the receptacles placed at the bottom of the machine therefore forms a skew histogram approximating to that given by a two-parameter lognormal distribution.

Kapteyn’s machine is still to be seen in the laboratory Huize de Wolf adjacent to the Genetics Laboratory of the University of Groningen, and we have to thank the Director of the Laboratory for providing us with the photograph which we reproduce in Fig. 3.1. Should we wish to construct a similar analogue device today it would most likely take the form of a computing programme for an automatic high-speed computer. Such computers may be readily programmed to produce a pseudo-random sequence of binary digits; from these, elementary variates \( e_i \) conforming to any given probability distribution can be constructed, and, by applying a recursive equation of the form of (3.2), the emergence of approximate lognormal distributions may be studied.

### 3.5. Extensions and Criticisms of the Theory of Genesis

In the discussion of the law of proportionate effect in § 3.3 we noted that the working of the law has usually been conceived as an ordered sequence of events in time, and we expressed our own view that other conceptions are possible. Indeed, an emphasis placed on the time sequence may lead to certain difficulties which are not easily resolved. A reconsideration of Theorem 3.1 shows that the greater the number of steps in the sequence, that is, the longer the law of proportionate effect is in operation, the greater the value of the \( \sigma^2 \) parameter associated with \( X_n \) becomes. This implication of the theorem is harmless in a number of cases, such as are found in biology, where the law is assumed to operate only during the period of growth to maturity of an organ or organism. But in other fields, for example, in the study of the size distribution of incomes, it has been objected that if the law operates at all, it must operate continually; and the implication that the inequality of incomes (measured by the \( \sigma^2 \) parameter) must continually increase is contrary to the evidence.

Kalecki [114] has suggested a method of dealing with this deficiency by abandoning the assumptions of the process (3.2). Developing his argument in an economic context, Kalecki postulates that variations in the inequality of incomes are to a great extent determined by economic
GENESIS OF LOGNORMAL DISTRIBUTIONS

forces, and studies first the special case where the inequality of incomes remains constant through time. Formally, then, the assumption is that the variance of $Y_J$, where $Y_J = \log X_J$, remains constant. From (3.2) this implies a negative correlation between $Y_{j-1}$ and $\log (1 + \epsilon_j)$. On the further assumption that the regression of $\log (1 + \epsilon_j)$ on $Y_{j-1}$ is linear

$$\log (1 + \epsilon_j) = -\alpha_j Y_{j-1} + \eta_j,$$

(3.10)

where $\eta_j$ is independent of $Y_{j-1}$, the new generating equation becomes

$$X_j = X_{j-1} e^{\eta_j}.$$

(3.11)

Under fairly general conditions the final distribution of $X_j$ is again approximately lognormal. The operation of the negative correlation implied by (3.10) may be regarded as a stabilizing influence and is by no means unreasonable in the context considered by Kalecki. In the same paper the writer extends the above arguments to admit systematic changes in the $\sigma^2$ parameter and also changes which are partly systematic and partly induced by random events.

There is another possible way of interpreting model (3.2). We may suppose that at any point of time the existing distribution of the variate arises from a large number of causes which operate simultaneously. For example, in attempting to explain the distribution of incomes, we may first think of a completely homogeneous group of wage earners each with a claim to an equal share. We then take into account the fact that the group is not homogeneous, each carrier possessing to a different extent attributes and talents which influence the magnitude of his claim. The outcome of these many different effects, acting in accordance with model (3.2), is again to produce a lognormal distribution of incomes. At other points of time the distribution of incomes may be thought to arise in a similar way; the reasons for the stability of the $\sigma^2$ parameter are then to be sought in the distribution of the attributes and talents in relation to the evaluation of these by the contemporary society. Secular changes in this evaluation may lead to a drift in the value of $\sigma^2$.

An explanation of the genesis of the three-parameter population is also possible. For, if in the law of change the increment is a random proportion of the amount by which the attained value exceeds some fixed value $\tau$, then

$$X_j - X_{j-1} = \epsilon_j (X_{j-1} - \tau)$$

(3.12)

replaces (3.2). This may be rewritten as

$$(X_j - \tau) - (X_{j-1} - \tau) = \epsilon_j (X_{j-1} - \tau),$$

(3.13)

from which it is clear that $X_n$, for large $n$, is distributed as a three-parameter $\Lambda$-variate.

This type of model is appropriate where there are prior grounds for postulating the existence of a fixed $\tau$. For example, in British agriculture there is a statutory minimum wage for hired farm workers; by established practice the contract wage for which an individual worker is hired is often arrived at by the negotiation of an agreed 'premium',

40
which is the amount by which the contract wage exceeds the statutory minimum. It may then be expected that the ‘premium’ will obey the two-parameter law, whereas the distribution of earnings will have a threshold value approximately equal to the statutory minimum. We present some confirmatory evidence of this hypothesis in Chapter 11.

On the other hand, it is not realistic to argue in this manner for a number of cases which are adequately described by the three-parameter distribution: we have particularly in mind the measurement of human body weight and related variates. What is required here is a generating model which will include an explanation of the threshold parameter and suggest the factors which control its value.

It would be possible to adapt our arguments to the negatively skew and to the four-parameter distributions though the resulting growth models would be of more restricted application.

3.6. The Theory of Breakage

We conclude this chapter with a discussion of a theory of breakage which has recently aroused interest among workers in particle-size statistics[128,61,106]. Although the application is novel, the formal theory is essentially a restatement of the theory of proportionate effect in terms of distribution functions rather than variate values. The present theory thus stands in relation to the theory of proportionate effect as the convolution property of distribution functions (corollaries 2.2a and 2.2b) stands in relation to the reproductive property of Theorem 2.2. The aim of the model outlined below, which Kolmogoroff[128] put forward in a discussion of empirical results obtained by Rasumovsky[172], is to explain the occurrence of two-parameter lognormal distributions in ores which have been crushed by natural or artificial processes.

Suppose there is a set of objects or elements with each of which is associated a positive measure, the dimension of the element. Let the initial distribution of the elements be $F_0(x)$, that is to say, the proportion of elements with dimension $\leq x$ is $F_0(x)$. The elements are then subjected to a sequence of independent breakage operations. If, at the $j$th breakage, $G_j(x | u)$ describes the distribution of elements arising from elements of dimension $u$ prior to the breakage, then the law of proportionate effect is equivalent to the statement that $G_j(x | u)$ depends only on the ratio $x/u$; we may write

$$G_j(x | u) = H_j\left(\frac{x}{u}\right). \quad (3.14)$$

Then

$$F_j(x) = \int_u H_j\left(\frac{x}{u}\right) dF_{j-1}(u). \quad (3.15)$$

If $X_j$ and $T_j$ are the variates associated with the distribution functions $F_j(x)$ and $H_j(t)$, then (3.15) implies that

$$X_j = T_j X_{j-1}, \quad (3.16)$$

so that

$$X_n = X_0 T_1 \cdots T_n. \quad (3.17)$$
and the result (that the final distribution tends to the lognormal) follows from Theorem 2.8 or 2.9.

The central idea of the theory of breakage may be carried over into a theory of classification. It is a curious fact that when a large number of items is classified on some homogeneity principle, the variate defined as the number of items in a class is often approximately lognormal. Examples of this phenomena we have noted are the number of persons in a census occupation class, the number of Sino-Japanese characters in a lexicographical group, and the outlay by households on classes of commodities. At first sight it may appear that, in classification problems of this type, the classifier is free to produce any distribution he chooses. But in practice, for a meaningful classification, some principle of homogeneity must be followed, and we suggest that the application of such a principle may lead to a process closely analogous to the breakage process described above. Thus a detailed list of occupations, collected from census returns, may first be divided into manual and non-manual, then each of these into skilled and unskilled, and so on.
CHAPTER 4

ARTIFICIAL LOGNORMAL SAMPLES AND TESTS OF LOGNORMALITY

ANGELO. Now, good my lord,
Let there be some more test made of my metal,
Before so noble and so great a figure
Be stamp'd upon it. Measure for Measure

4.1. THE CONSTRUCTION OF ARTIFICIAL SAMPLES FROM A LOGNORMAL POPULATION

Samples from a specified lognormal population may readily be constructed artificially. In principle it would be possible to use an electronic analogue device of the type described in § 3.4, though the method would not be very efficient. A better method, if an electronic computer were used, would be to generate pseudo-random numbers from a rectangular population and to transform the rectangular variates using the lognormal distribution function. Since, however, the greater part of this task has already been performed by Mahalanobis[143] and Wold[212], who have compiled extensive tables of random normal deviates (that is, random values from a $N(0, 1)$ population), it is more convenient to take these as a starting point. There is, too, the important consideration that although the deviates constructed by Mahalanobis and Wold are strictly pseudo-random (they may retain traces of bias due to their mode of formation) they have in fact been submitted to, and have passed, exhaustive tests of randomness, without which procedure any newly generated samples must be regarded with some suspicion.

Suppose then that we wish to construct a sample of size $n$ from a $\Lambda(\mu, \sigma^2)$ population, using a table of random normal deviates. If we denote by $u_i$ ($i = 1, \ldots, n$), consecutive values from the table, then the transformation $x_i = e^{\mu + \sigma u_i}$ gives a sample of size $n$ from $\Lambda(\mu, \sigma^2)$. In practice we need only consider the case $\mu = 0$, since all other cases may be derived from this in a simple way. For example, if $w_i = e^{\mu u_i}$, then

$$x_i = e^{\mu u_i}$$
$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
$$= e^{\mu} \frac{1}{n} \sum_{i=1}^{n} w_i$$
$$= e^{\mu} \bar{w}$$
$$\nu^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
$$= e^{2\mu} \frac{1}{n - 1} \sum_{i=1}^{n} (w_i - \bar{w})^2$$
$$= e^{2\mu} \nu_0^2$$

43
and so on; thus the sample functions of a $\Lambda(\mu, \sigma^2)$ sample are easily obtained from those of the $\Lambda(0, \sigma^2)$ sample. Equally, artificial samples of the above type may be transformed, by the addition of $\tau$ to each sample value, into samples from a three-parameter distribution. But the adjustment of sample values to allow for a change in $\sigma^2$ is not simple, being as complicated as the original derivation. For this reason tables of variates drawn from a $\Lambda(0, 1)$ population would not have the wide uses of tables of random normal deviates.

4.2. The Purpose of Constructing Artificial Samples

The main object in constructing artificial samples is to provide a test of various methods of estimation. It is not always possible to obtain theoretical criteria for judging the efficiencies of different methods, and it is useful to have artificial samples, drawn from populations whose parameters are known, as a basis for empirical assessment. For example, the use of logarithmic probability paper for estimating $\mu$ and $\sigma^2$ is attractively simple, but only by carrying out some fairly extensive experiment can we decide whether this method is substantially less reliable than a more refined method which involves considerable computation.

A secondary reason is to supply a general comparison with sample distributions found in nature. It may often be more convincing to have side by side the frequency distribution that has arisen in practice with one constructed by some artificial means. This is particularly the case with very skew distributions, since even small samples may contain one or two very high values which might otherwise be suspected.

An extension of the use of artificial samples to judge the efficiency of estimation procedures is to attempt, from a large number of samples, to find an approximation to the distribution of sample functions. This is the basis of the so-called 'Monte Carlo' approach, which would be especially useful in the field of small-sample theory, where analytical methods are seldom available; no use, however, has been made of the technique by the writers.

Finally, in the field of econometrics where structural equations describing economic relationships contain random disturbance terms it would often be helpful to test suggested estimation procedures on artificially constructed models. Often in such models the error is multiplicative and may reasonably be assumed lognormal.

4.3. The 65 Samples Constructed

The samples actually constructed for our present purposes have been derived as indicated in §4.1 with the use of Wold's tables of random normal deviates. These numbers were read into an automatic computer (the EDSAC) where the necessary transformations were carried out and the sample functions calculated; the sample values existed only inside the machine and at no time were they recorded. Checks were applied
to ensure that the computing programmes† were correct and that the machine was functioning correctly during the calculation.

In all, 65 samples, comprising a total of 7616 variate values, were constructed from $\Lambda(0, \sigma^2)$ populations for a selection of values of $\sigma = 0.2(0.1)1.0$, which includes the majority of values of $\sigma$ occurring in practice. The sample sizes $n$ were 32, 64, 128, 256 and 512; the reason for using these particular sample sizes was that while it provides a good selection it also facilitated computation (especially division) in a binary machine such as the EDSAC.

In Table 4.1 we show the number of samples constructed for each $n$ and $\sigma$, and, for identification purposes, the corresponding serial numbers of Wold's random normal deviates.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Serial numbers of corresponding random normal deviates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1153-1728</td>
</tr>
<tr>
<td>0.3</td>
<td>1-1152</td>
</tr>
<tr>
<td>0.4</td>
<td>1728-4032</td>
</tr>
<tr>
<td>0.5</td>
<td>4033-6580</td>
</tr>
<tr>
<td>0.6</td>
<td>6581-7616</td>
</tr>
<tr>
<td>0.8</td>
<td>1-739</td>
</tr>
<tr>
<td>0.9</td>
<td>739-1578</td>
</tr>
<tr>
<td>1.0</td>
<td>1579-4072</td>
</tr>
</tbody>
</table>

For some time writers have been much concerned with the effects that non-normality of a population may have on the distribution of those widely used test statistics which are based on the assumption that the population is normal[10, 49, 80, 81, 82, 83, 154, 155, 156]. The importance of submitting samples to tests for normality of the population before using such statistics has been particularly stressed by Geary[83]. Since, when applying lognormal theory, we must draw heavily on established normal tests, Geary's remarks apply equally well to testing for lognormality. We are thus led to consider the possibility of such tests as will answer the question: given a random sample $x_1, \ldots, x_n$, are we justified in assuming that the population from which the sample is drawn is lognormal? Evidently any test of normality may be adapted, by using transformed sample values, as a test of lognormality (for the two-parameter case).

In the remainder of this chapter, after discussing the use of logarithmic probability paper, we recall the familiar tests of goodness of fit, skewness and kurtosis. Finally the application of tests to the artificial samples is reported.

Before proceeding, however, we suggest that not infrequently the emphasis to be laid on testing for lognormality may be qualified. Often the statistician is presented not with an isolated sample but with a collect-

† The computing programmes are described in greater detail in Chapter 13.
tion of samples from closely similar populations which together indicate the common form of statistical description to be used. When in addition there are grounds for presuming the operation of some kind of generative process, such as described in Chapter 3, there is less compulsion to test each sample separately. For example, if the artificial samples are not to be considered as a composite set, then, as we shall see, too much attention to the results of tests on the individual samples may lead to a somewhat barren study of the set as a whole.

4.5. Logarithmic Probability Paper

It is usually worth while to submit data to some kind of graphical scrutiny as a preliminary to any more detailed analysis. By so doing we may eventually save much time and labour and even have suggested what form the more elaborate analysis should take; moreover we may obtain, for those measures in which we are interested, provisional estimates which will both serve our purpose until more accurate values may be obtained and also provide a check on subsequent calculations. For the lognormal distribution we are fortunate in having a quick and, with experience, fairly accurate graphical method of analysis; this method is facilitated by the use of a special type of graph paper—logarithmic probability paper.†

It is difficult to assign credit for the introduction of this type of paper; it is an obvious development of arithmetic probability paper, which, though hinted at by early writers such as Galton, was first used by Hazen [103] in 1914. The theory underlying its use is derived only from relation (2.14) which connects the quantile of order \( q \) of \( A(\mu, \sigma^2) \) and the corresponding quantile of \( N(0, 1) \); this may be rewritten as

\[
\log \xi_q = \sigma \psi_q + \mu, \tag{4.1}
\]

so that the locus of \((\psi_q, \log \xi_q)\) is a straight line. Suppose now that \( L(x) \) denotes the sample distribution function so that \( L(x) \) is the proportion of sample values \( \leq x \). If we write

\[
g_i = L(x_i) \quad \text{and} \quad y_i = \log x_i \tag{4.2}
\]

for \( i = 1, \ldots, n \), then we should expect the points \((\psi_q, y_i)\) to lie approximately on the straight line

\[
y = \sigma \psi + \mu. \tag{4.3}
\]

The same array of points is obtained if we plot the points \( \{L(x_i), x_i\} \) with \( L(x) \) on a normal probability scale and \( x \) on a logarithmic scale; the purpose of logarithmic probability paper is thus to facilitate the plotting of the points \((\nu_q, y_i)\) by providing these appropriate scales so that only \( \{L(x_i), x_i\} \) need be calculated.

Logarithmic probability paper has its scale of ordinates \( x \) graduated logarithmically while, on the abscissa scale, proportions \( L(x) \) are plotted

† Logarithmic probability paper may be obtained from most retailers of technical stationery. Its use is further described by Herdan [105].
as their equivalent normal deviates. The paper is obtainable in various cycle numbers: one-cycle paper has the logarithmic scale so graduated that it may be conveniently used for increases in $x$ up to tenfold; two-cycle, for increases up to one hundredfold, and so on.

The form of the data required for the application of this method is a grouped cumulative frequency table and this is usually easily formed. Table 4.2 gives the data in this form for one of the artificial samples of size 64 and with $\mu = 0$ and $\sigma = 0.6$. The corresponding array of points on logarithmic probability paper is shown in Fig. 4.1 together with the theoretical line $y = \sigma v + \mu$.

**Table 4.2. Grouped Cumulative Frequencies for an Artificial Sample: $n = 64; \mu = 0; \sigma = 0.6$**

<table>
<thead>
<tr>
<th>$x$</th>
<th>$L(x)$, $n_{o}$ of sample values $\leq x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>1.6</td>
</tr>
<tr>
<td>0.5</td>
<td>20.3</td>
</tr>
<tr>
<td>0.9</td>
<td>48.4</td>
</tr>
<tr>
<td>1.2</td>
<td>63.0</td>
</tr>
<tr>
<td>1.5</td>
<td>70.3</td>
</tr>
<tr>
<td>1.8</td>
<td>81.2</td>
</tr>
<tr>
<td>2.1</td>
<td>87.5</td>
</tr>
<tr>
<td>2.4</td>
<td>92.2</td>
</tr>
<tr>
<td>2.7</td>
<td>97.8</td>
</tr>
<tr>
<td>3.0</td>
<td>99.4</td>
</tr>
<tr>
<td>3.3</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Although the use of the paper can hardly be regarded as a rigorous statistical test of lognormality it nevertheless provides a quick method of judging whether the population may feasibly be lognormal. Moreover, the parameters $\mu$ and $\sigma^2$ may be estimated from a straight line fitted by eye to the points. The method we would advocate is the following: from (2.14) the population quantiles of order 16, 50 and 84% are given by

$$\xi_{16\%} = e^{\mu - \sigma},$$

$$\xi_{50\%} = e^{\mu},$$

$$\xi_{84\%} = e^{\mu + \sigma};$$

so that

$$\mu = \log \xi_{50\%}$$

and

$$\sigma = \log \left[ \frac{1}{2} \left( \frac{\xi_{84\%}}{\xi_{16\%}} \right) \right].$$

If we read from the straight-line graph the values of $x$ corresponding to the 16, 50 and 84% points and substitute these for the $\xi$ values in (4.4) and (4.5), we obtain estimates $m$ and $s$ of $\mu$ and $\sigma$ respectively. For example, one student given the data of Table 4.2 fitted a straight line by eye and obtained

$$x_{16\%} = 0.558,$$

$$x_{50\%} = 1.005,$$

$$x_{84\%} = 1.812,$$
and derived

\[ m = \log 1.005 \]
\[ = 0.005 \]

and

\[ s = \log 1.8 \]
\[ = 0.588 \]

Two advantages of the graphical method are worth noting. It does not require the transformation of the sample values. Data are often presented in a grouped form and the method does not require, as more

Fig. 4.1. Logarithmic probability graph for the data of Table 4.2.

elaborate estimation procedures do, any adjustment for the effect of grouping, whether or not the group intervals are equal (compare §5.7).

The authors' experience has shown that practice helps in judging the relative importance to be attached to points at different percentile levels. For the reader who wishes to acquire such practice we give in Table 4.3 the necessary data for five artificial samples. The theoretical values of \( \mu \) and \( \sigma \) are to be found in Table 4.6 in the note at the end of this chapter. The result of an experimental comparison of the graphical with other methods of estimation is presented in Chapter 5.
### 4.6. Tests of Lognormality: Geary and Pearson Tests; \( \chi^2 \)

As a test of lognormality in the two-parameter case we may apply the skewness and kurtosis tests of normality on the transformed sample values. In his paper [83] Geary treats a series of tests for skewness and kurtosis based on the statistics \( g_1(p) \) and \( g_2(p) \) defined by

\[
g_1(p) = \frac{S'(p) - S''(p)}{(S'(2))^{1/2}}
\]

and

\[
g_2(p) = \frac{S(p)}{(S'(2))^{1/2}}
\]

where

\[
S(p) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}|^p,
\]

\[
S'(p) = \frac{1}{n} \sum_{y_i < \bar{y}} |y_i - \bar{y}|^p,
\]

\[
S''(p) = \frac{1}{n} \sum_{y_i > \bar{y}} |y_i - \bar{y}|^p
\]

and \( p \geq 0 \). He concludes that, for large samples and a wide field of alternative hypotheses regarding the nature of the population, \( g_1(3) \) and \( g_2(4) \) are the most efficient test statistics; also that, for samples of moderate size, \( g_2(1) \) is probably as efficient as \( g_2(4) \). A good account of these tests, together with tables and charts of the 1 and 5% points of \( g_1(3) \) and the 1, 5 and 10% points of \( g_2(1) \) and \( g_2(4) \), is to be found in Geary and Pearson [84].

A method that may be applied to all lognormal distributions is, of
course, the \( \chi^2 \) test of goodness of fit. This test is likely to be less sensitive than the Geary tests since it ignores the sign and pattern of the differences between observed and expected group frequencies and often requires additional grouping at the extremes of the range.

4.7. Tests Applied to the 65 Samples

The \( g_1(3) \) test was applied to all the artificial samples; the samples for which significant skewness of the transformed distribution was obtained are listed in Table 4.4.

**Table 4.4. Artificial Samples Giving Significant Values of \( g_1(3) \)**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( \sigma )</th>
<th>Serial numbers of corresponding random normal deviates</th>
<th>( g_1(3) )</th>
<th>Significance level (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.2</td>
<td>1185-1216</td>
<td>-0.868</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>1217-1248</td>
<td>-0.712</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>1633-1663</td>
<td>-0.700</td>
<td>5</td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>2125-1068</td>
<td>-0.577</td>
<td>5</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>3619-3776</td>
<td>-0.712</td>
<td>5</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>6045-6598</td>
<td>-0.680</td>
<td>5</td>
</tr>
<tr>
<td>512</td>
<td>0.5</td>
<td>6081-6598</td>
<td>-0.680</td>
<td>5</td>
</tr>
</tbody>
</table>

For the samples of size 32, 64 and 128 the \( g_2(1) \) test was applied and for the samples of size 256 and 512 the \( g_2(4) \) test was carried out. The results are given in Table 4.5 for those samples showing significant values.

**Table 4.5. Artificial Samples Giving Significant Values of \( g_2(1) \) or \( g_2(4) \)**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( \sigma )</th>
<th>Serial numbers of corresponding random normal deviates</th>
<th>( g_2(1) ) or ( g_2(4) )</th>
<th>Significance level (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.2</td>
<td>1185-1216</td>
<td>0.734</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>1217-1248</td>
<td>0.855</td>
<td>10</td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>1313-1344</td>
<td>0.756</td>
<td>10</td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>1569-1600</td>
<td>0.734</td>
<td>5</td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>1601-1632</td>
<td>0.660</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>513- 576</td>
<td>0.839</td>
<td>10</td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>769- 832</td>
<td>0.830</td>
<td>10</td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>1025-1086</td>
<td>0.847</td>
<td>5</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>2369-2416</td>
<td>0.829</td>
<td>10</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>4033-4288</td>
<td>3.672</td>
<td>5</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>4545-4800</td>
<td>3.651</td>
<td>5</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>5009-5042</td>
<td>2.430</td>
<td>5</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>5825-6080</td>
<td>2.573</td>
<td>5</td>
</tr>
</tbody>
</table>
THE LOGNORMAL DISTRIBUTION

Note. The theoretical values of \( \mu \) and \( \sigma \) for the artificial samples of Table 4.3 are given in Table 4.6. We also give the maximum-likelihood estimates; for the method of calculation of these, see §5.21.

### TABLE 4.6. VALUES OF \( \mu \) AND \( \sigma \) FOR THE ARTIFICIAL SAMPLES OF TABLE 4.3

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \mu )</th>
<th>( \sigma )</th>
<th>Theoretical values</th>
<th>Maximum-likelihood estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>1.31</td>
<td>0.7</td>
<td>( \mu )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>64</td>
<td>1.13</td>
<td>0.4</td>
<td>( \mu )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>128</td>
<td>0.90</td>
<td>0.6</td>
<td>( \mu )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>256</td>
<td>0.34</td>
<td>0.8</td>
<td>( \mu )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>512</td>
<td>0.17</td>
<td>0.7</td>
<td>( \mu )</td>
<td>( \sigma )</td>
</tr>
</tbody>
</table>
CHAPTER 5

ESTIMATION PROBLEMS: I

COUNTESS. Many likelihoods informed me of this before, which hung so tottering in the balance that I could neither believe nor misdoubt.

All's Well That Ends Well

5.1. GENERAL REMARKS ON ESTIMATION

At some time or another most of the methods of estimation so far devised by statisticians have been applied to the various types of lognormal populations. In this chapter we survey the application of these methods to the two-parameter distribution and attempt to assess their relative merits; in the next chapter we continue the discussion for the more difficult problems that arise where more than two parameters are involved.

Before a decision is reached on which of a number of alternative estimation procedures to adopt, the question that must first be answered is: what desirable properties is a good estimator expected to possess? The three main criteria usually suggested are the following, of which the first two are theoretical, and the third is practical in nature:

(i) The estimator should be unbiased, or, when only large samples are in question, asymptotically unbiased (consistent).

(ii) The variance, or some similar measure, of the estimator should be as small as possible.

(iii) The calculations involved should be reasonable and within the capabilities of the available computing machinery.

An estimator which satisfies the first two criteria will be termed a minimum variance unbiased (or consistent) estimator. Seldom does an estimator possess all the desirable qualities listed above; for lognormal distributions estimators which are satisfactory with respect to the theoretical criteria tend to require much calculation. The decision to adopt a particular method must then be based on a careful compromise between what is theoretically desirable and what is computationally feasible. We have therefore applied all the available methods to the sixty-five artificial samples described in Chapter 4 in order to gain wider experience on this fine balance. The need of some basis for empirical judgement of this kind is especially obvious for methods such as the graphical (described in §4.5), where no theoretical evaluation of the method is possible; for this case we have conducted an extensive experiment with the artificial samples. The results of the experiment are reported later in this chapter.

† To these three criteria there is sometimes added a fourth: how closely do probabilities calculated from the estimated distribution follow the corresponding theoretical probabilities? We take no account of this criterion in this chapter principally because it is difficult to manipulate mathematically.

‡ The term best unbiased estimator is used in the same context by Rao[171], who gives an excellent account of estimation theory.
Of the methods available for point estimation we distinguish five types:

(i) the method of maximum likelihood,
(ii) the method of moments,
(iii) the method of quantiles,
(iv) the graphical method, and
(v) mixed methods.

The last group is intended to cover all those methods which are hybrids of the other four types; for example, one method which has been used for the three-parameter distribution employs the mean and two quantiles.

The discussion of the methods begins with the problem of estimating \( \mu \) and \( \sigma^2 \), and proceeds to that of estimating \( \alpha \) and \( \beta^2 \), for \( \Lambda(\mu, \sigma^2) \). Besides point estimation, the question of confidence intervals is also studied. Towards the end of the chapter grouped data are considered and some devices given which are useful in special circumstances. A summary of conclusions is presented in §5.9.

5.2. Estimation of \( \mu \) and \( \sigma^2 \) in a Two-parameter Distribution

Suppose there is given a sample \( S_n \) of size \( n \) from \( \Lambda(\mu, \sigma^2) \) consisting of the observations \( x_1, x_2, ..., x_n \); the problem is to find sample functions which are suitable estimators of \( \mu \) and \( \sigma^2 \). We shall first establish the notation and following sections. The jth sample moments about the origin and about the sample mean are denoted by \( l'_j \) and \( l_j \) respectively; thus

\[
l'_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j \tag{5.1}
\]
\[
l_j = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^j \tag{5.2}
\]

We also write

\[
\bar{x} = l'_1 \tag{5.3}
\]
\[
\bar{v}_2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n}{n-1} l_2 \tag{5.4}
\]
\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{5.5}
\]
\[
\bar{v}_3 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{5.6}
\]

where \( y_i = \log x_i \). The sample quantile of order \( q \) is denoted by \( x_q \); thus the proportion of sample values \( \leq x_q \) is \( \geq q \) and \( x_q \) is the least sample value with this property; or, in the notation of §4.5, \( x_q \) is the least sample value such that \( L(x_q) \geq q \). For the estimators of \( \mu \) and \( \sigma^2 \) we use \( m_i \) and \( s_i^2 \) respectively, the particular value of the suffix \( i \) indicating the method of estimation employed.
5.21. THE METHOD OF MAXIMUM LIKELIHOOD

The likelihood function of the sample is

$$\frac{1}{\sigma^n(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log x_i - \mu)^2 \right\}, \quad (5.7)$$

and the maximum-likelihood estimators \(m_1\) and \(s_1^2\) of \(\mu\) and \(\sigma^2\) are found to be

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

$$= \bar{y}, \quad (5.8)$$

and

$$s_1^2 = \frac{1}{n} \sum_{i=1}^{n} (\log x_i - m_1)^2$$

$$= \frac{n-1}{n} s_1, \quad (5.9)$$

This is, as is to be expected, equivalent to the method of maximum likelihood applied to the transformed sample. The estimator \(s_1^2\) is biased but consistent; if, however, equation (5.9) is replaced by

$$s_1^2 = \frac{n}{n-1} s_1^2 \quad (5.10)$$

then \(m_1\) and \(s_1^2\) are minimum variance unbiased estimators of \(\mu\) and \(\sigma^2\). The variances of \(m_1\) and \(s_1^2\), required in determining the large-sample efficiencies of other estimators, are readily obtained from normal theory as

$$D^2(m_1) = \frac{\sigma^2}{n} \quad (5.11)$$

and

$$D^2(s_1^2) = \frac{2\sigma^4}{n-1} \quad \sim \frac{2\sigma^4}{n} \quad (5.12)$$

To assist in the calculation for this case Jenkins[110] has tabulated the common logarithms and squares of the logarithms of the first fifty integers; the method of calculating the moments, which he suggests is frequently applicable, is to reduce the given data by a change of scale such that his table and the correction formulae he gives may be used.

5.22. THE METHOD OF MOMENTS

The estimators \(m_2\) and \(s_2^2\) of \(\mu\) and \(\sigma^2\) are here obtained by equating the first two sample moments \(\hat{l}'_1\) and \(\hat{l}'_2\) to the expressions given by substituting \(m_2\) and \(s_2^2\) for \(\mu\) and \(\sigma^2\) in (2.6) with \(j = 1\) and \(2\); so

$$l'_1 = \exp \{m_2 + \frac{1}{2}s_2^2\}, \quad (5.13)$$

and

$$l'_2 = \exp \{2m_2 + 2s_2^2\}; \quad (5.14)$$

whence

$$m_2 = 2 \log l'_2 - \frac{1}{4} \log l'_1, \quad (5.15)$$

and

$$s_2^2 = \log l'_2 - 2 \log l'_1. \quad (5.16)$$
THE LOGNORMAL DISTRIBUTION

The estimators are both consistent. Their large-sample variances, obtained by the variational method, are

\[ D^2[m_2] = \frac{1}{4n} (\eta^8 + 4\eta^6 - 2\eta^4 + 4\eta^2), \quad (5.17) \]

and

\[ D^2[s^2] = \frac{1}{n} (\eta^8 + 4\eta^6 + 2\eta^4); \quad (5.18) \]

where \( \eta^2 = e^{\sigma^2} - 1 \) as before; the large-sample efficiencies are therefore given by

\[ \text{eff.} \{m_2\} = \frac{D^2[m_1]}{D^2[m_2]} = \frac{4\sigma^2}{\eta^8 + 4\eta^6 - 2\eta^4 + 4\eta^2}; \quad (5.19) \]

and

\[ \text{eff.} \{s^2\} = \frac{D^2[s_2]}{D^2[s^2]} = \frac{2\sigma^4}{\eta^8 + 4\eta^6 + 2\eta^4}. \quad (5.20) \]

Fig. 5.1 shows the graphs of \( \text{eff.} \{m_2\} \) and \( \text{eff.} \{s^2\} \) against \( \sigma^2 \); as \( \sigma^2 \) increases the efficiency of the method rapidly declines, especially for estimation of \( \sigma^2 \). Values of \( \sigma^2 \) that arise in practice are frequently in the neighbourhood of 0.5; at this value, while the efficiency of estimating \( \mu \) is 79\%, that of estimating \( \sigma^2 \) is as low as 31\%.

5.23. THE METHOD OF QUANTILES

To obtain quantile estimators the sample quantiles of order \( q_1 \) and \( q_2 \) \((q_1 < q_2)\) are set equal to the expressions obtained by replacing \( \mu \) and \( \sigma \) by \( m_3 \) and \( s_3 \) in (2.14); hence

\[ x_{q_1} = \exp \{m_3 + \nu_{q_1} s_3\}, \quad (5.21) \]

and

\[ x_{q_2} = \exp \{m_3 + \nu_{q_2} s_3\}; \quad (5.22) \]

so that

\[ m_3 = \frac{\nu_{q_1} \log x_{q_1} - \nu_{q_2} \log x_{q_2}}{\nu_{q_1} - \nu_{q_2}} \]

and

\[ s_3 = \frac{\log x_{q_1} - \log x_{q_2}}{\nu_{q_1} - \nu_{q_2}}. \quad (5.24) \]

There are many slight variations of the method, most of which make use of adjusted sample quantiles.†

It can be shown that the maximum efficiency attainable by the method is when the quantiles are symmetrically placed; attention may therefore be confined to the case where the quantiles are of order \( q \) and \( 1 - q \) \((q < \frac{1}{2})\). Then

\[ \nu_{1-q} = -\nu_q \]

\[ = \nu_q, \quad (5.25) \]

† See, for example, G. R. Davies[50, 51] and Davies and Smith[53].
say, so that (5.23) and (5.24) reduce to
\[ m_3 = \frac{1}{2} (\log x_{1-q} + \log x_q), \]
and
\[ s_3 = \frac{1}{2 \nu} (\log x_{1-q} - \log x_q). \]
The large-sample variances\(^\dagger\) of \( m_3 \) and \( s_3 \) are given by
\[ D^2(m_3) = \frac{n \sigma^2 q e^q}{n}, \]
and \( D^2(s_3) = \frac{4 \pi \nu q (1 - 2q) e^q}{n \nu^2} \).

\[^\dagger\text{These may be easily found by formula (9.27) of Kendall[123], which gives the large-sample covariance between two sample quantiles; or by the formulae of section 28.5 of Cramér[46].}\]
These efficiencies are independent of $\sigma^2$; their graphs against $q$ are shown in Fig. 5.2. The maximum efficiencies attainable are not at the same value of $q$; quantiles of order 27 and 73% estimate $\mu$ with 81% efficiency and quantiles of order 7 and 93% estimate $\sigma^2$ with an efficiency of 65%.

![Efficiency graph](image)

Fig. 5.2. Efficiency of method of quantiles in estimation of $\mu$ and $\sigma^2$.

5.24. THE GRAPHICAL METHOD

A graphical method of estimating $\mu$ and $\sigma^2$ has already been described in the discussion on logarithmic probability paper in §4.5. It remains to describe our experiment in graphical analysis with the sixty-five artificial samples. For each of the samples a grouped cumulative frequency table (such as that of Table 4.2) had been obtained in the course of processing the samples (see Chapter 13 for further details). For each sample three different persons were asked to perform a graphical analysis; these were:

(i) an experienced computer having little previous acquaintance with the use of logarithmic probability paper who performed the analysis on all the samples;

(ii) a junior computer, straight from school, who also used all the samples; and

(iii) a miscellaneous group of five persons, all with experience in handling statistical data, but only one with any experience in the method.

Each subject was given a minimum of instruction, enough only to allow him to plot the points on logarithmic probability paper. For each
of his selected samples (presented to him in a random order) he was asked to plot the points, choosing his own scale, and to draw what he considered a straight line through the array. The estimates of \( \mu \) and \( \sigma^2 \) were then calculated by the computing staff of the Department. The results of this experiment are compared below with those of the other estimation procedures.

5.3. Experimental Results (I): The Methods of §5.2 Applied to the 65 Samples

The estimates of \( \mu \) and \( \sigma^2 \) obtained by applying the different methods to the artificial samples are set out in detail in Appendix Tables B1 and B2. Here we are interested only in the light that they throw on the value of the methods of estimation. As empirical measures of their efficiencies the values of

\[
\Delta(m_i) = \sqrt{\frac{1}{N} \sum (m_i - \mu)^2},
\]

and

\[
\Delta(s_i^2) = \sqrt{\frac{1}{N} \sum (s_i^2 - \sigma^2)^2},
\]

were calculated for each method and for different groupings of the samples; \( N \) denotes the number of samples in a group. The grouping is by sample size: 32, 64, 128, 256 and 512; and by range of \( \sigma \): 0.2–0.4, 0.5–0.7 and 0.8–1.0.

**Table 5.1. Values of \( \Delta(m_i) \) for a Grouping by Sample Size**

<table>
<thead>
<tr>
<th>Sample size ( n )</th>
<th>Method of maximum likelihood</th>
<th>Method of moments</th>
<th>Method of quantiles (27, 73%)</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.0542</td>
<td>0.1683</td>
<td>0.1558</td>
<td>0.1934</td>
<td>0.1953</td>
<td>0.1953</td>
</tr>
<tr>
<td>64</td>
<td>0.0737</td>
<td>0.0858</td>
<td>0.0672</td>
<td>0.0692</td>
<td>0.0625</td>
<td>0.0625</td>
</tr>
<tr>
<td>128</td>
<td>0.0909</td>
<td>0.0744</td>
<td>0.0714</td>
<td>0.0692</td>
<td>0.0645</td>
<td>0.0645</td>
</tr>
<tr>
<td>256</td>
<td>0.0916</td>
<td>0.0680</td>
<td>0.0691</td>
<td>0.1181</td>
<td>0.0660</td>
<td>0.0660</td>
</tr>
<tr>
<td>512</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All samples</td>
<td>0.0964</td>
<td>0.1100</td>
<td>0.0961</td>
<td>0.1226</td>
<td>0.1197</td>
<td>0.1197</td>
</tr>
</tbody>
</table>

**Table 5.2. Values of \( \Delta(s_i^2) \) for a Grouping by Size of \( \sigma \)**

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>Method of maximum likelihood</th>
<th>Method of moments</th>
<th>Method of quantiles (27, 73%)</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2–0.4</td>
<td>0.0935</td>
<td>0.0403</td>
<td>0.0397</td>
<td>0.0399</td>
<td>0.0428</td>
<td>0.0428</td>
</tr>
<tr>
<td>0.5–0.7</td>
<td>0.0925</td>
<td>0.1018</td>
<td>0.0843</td>
<td>0.0972</td>
<td>0.1115</td>
<td>0.1115</td>
</tr>
<tr>
<td>0.8–1.0</td>
<td>0.1292</td>
<td>0.1512</td>
<td>0.1340</td>
<td>0.1792</td>
<td>0.1644</td>
<td>0.1644</td>
</tr>
<tr>
<td>All samples</td>
<td>0.0964</td>
<td>0.1100</td>
<td>0.0961</td>
<td>0.1226</td>
<td>0.1197</td>
<td>0.1197</td>
</tr>
</tbody>
</table>

The values of \( \Delta(m_i) \) are given, for sample size groups, in Table 5.1 and, for \( \sigma \) groups, in Table 5.2. The corresponding values of \( \Delta(s_i^2) \) are shown in Tables 5.3 and 5.4. The results of these tables supplement the theoretical measures of efficiency already worked out for the first three
THE LOGNORMAL DISTRIBUTION

### Table 5.3. Values of $\Delta(s^2)$ for a Grouping by Sample Size

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Method of maximum likelihood</th>
<th>Method of moments</th>
<th>Method of quantiles ($7.93%$)</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.048</td>
<td>0.025</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>16</td>
<td>0.069</td>
<td>0.017</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>32</td>
<td>0.076</td>
<td>0.016</td>
<td>0.018</td>
<td>0.018</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>512</td>
<td>0.0123</td>
<td>0.017</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>All samples</td>
<td>0.0940</td>
<td>0.017</td>
<td>0.022</td>
<td>0.022</td>
<td>0.022</td>
<td>0.022</td>
</tr>
</tbody>
</table>

### Table 5.4. Values of $\Delta(s^2)$ for a Grouping by Size of $\sigma$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Method of maximum likelihood</th>
<th>Method of moments</th>
<th>Method of quantiles ($7.93%$)</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2-0.4</td>
<td>0.0192</td>
<td>0.0213</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>0.5-0.7</td>
<td>0.0531</td>
<td>0.0416</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>0.8-1.0</td>
<td>0.0182</td>
<td>0.0199</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>All samples</td>
<td>0.0940</td>
<td>0.017</td>
<td>0.022</td>
<td>0.022</td>
<td>0.022</td>
<td>0.022</td>
</tr>
</tbody>
</table>

methods, and give some indication of the effectiveness of the graphical method. This last method is clearly not as reliable as the numerical methods; it should be remembered, however, that the subjects chosen for the experiment had little experience in its application. Our own view, though we have no objective evidence to support it, is that experience can improve the skill of estimating in this way. A preliminary graphical analysis is, in any case, always advisable and helps to give one a feel for the data; if a more reliable estimate is needed one of the more sophisticated methods should be employed. Of these the choice must lie between the maximum-likelihood and the quantile methods; the method of moments has little to recommend it either computationally or theoretically (especially when $\sigma$ is large). In the experiment the method of quantiles obtains very good results. Although, in this case, the maximum-likelihood estimators are also sufficient, and so cannot be bettered even for small samples, the method of quantiles takes a respectable second place. Our recommendation would then be that for small samples the maximum-likelihood estimators should be used; for large samples, however, the method of quantiles should be used, since it is so easy to apply. It should be remembered that the data are assumed not to be grouped, and the method of maximum likelihood requires the transformation of each individual variate value. The case of grouped data is considered later.

### 5.4. Estimation of $\alpha$ and $\beta^2$ in a Two-Parameter Distribution

We now discuss the application of the different methods to the estimation of the mean $\alpha$ and the variance $\beta^2$ of a $\Lambda(\mu, \sigma^2)$ distribution.
5.41. THE METHOD OF MAXIMUM LIKELIHOOD

Although for this case the maximum-likelihood method leads to intractable equations there is an elegant method due to Finney which is equivalent. This depends on a useful property of jointly sufficient estimators: any function of jointly sufficient estimators is a minimum variance unbiased estimator of its expectation (under certain general regularity assumptions; compare Rao). Now \( \bar{y} \) and \( v_2 \) are jointly sufficient estimators of \( \mu \) and \( \sigma^2 \) and

\[
E(e^{\bar{y}}) = \exp \left\{ \mu + \frac{\sigma^2}{2n} \right\}
= \alpha \exp \left\{ -\frac{n-1}{2n} \sigma^2 \right\}.
\]

(5.34) If a function of \( v_2 \) can be found, say \( f(v_{2j}) \), such that

\[
E(f(v_{2j})) = \exp \left\{ \frac{n-1}{2n} \sigma^2 \right\}
= \sum_{j=0}^{\infty} \left( \frac{n-1}{2n} \right)^j \sigma^{2j} j^j,
\]

(5.35) then, since \( \bar{y} \) and \( v_2 \) are distributed independently, the estimator \( e^{\bar{y}} f(v_{2j}) \) will be a minimum variance unbiased estimator of \( \alpha \). This function is readily found, for

\[
E(v_{2j}) = \frac{(n-1)(n+1) \ldots (n-3+2j)}{(n-1)^j} \sigma^{2j} \quad (j = 1, 2, \ldots),
\]

(5.36) so that

\[
E \left( \frac{(n-1)^{2j}}{(n+1) \ldots (n-3+2j)} \frac{(v_{2j})^j}{j!} \right) = \left( \frac{n-1}{2n} \right)^j \frac{\sigma^{2j}}{j^j} \quad (j = 1, 2, \ldots).
\]

(5.37) If a new function \( \psi_n(t) \) is defined by

\[
\psi_n(t) = 1 + \frac{n-1}{n} t + \frac{(n-1)^3}{n^2 (n+1)} \frac{t^2}{2!} + \frac{(n-1)^3}{n^3 (n+1)} \frac{t^3}{3!} + \ldots,
\]

(5.38) then

\[
E(\psi_n(v_{2j})) = \exp \left\{ \frac{n-1}{2n} \sigma^2 \right\};
\]

(5.39) and so

\[
a_1 = e^{\bar{y}} \psi_n(1) \]

(5.40) is a minimum variance unbiased estimator of \( \alpha \). Similarly it may be shown that

\[
b_2^2 = e^{\bar{y}} \psi_n(2v_{2j}) - \psi_n \left( \frac{n-2}{n-1} v_{2j} \right) \]

(5.41) where

\[
\chi_n(t) = \psi_n(2t) - \psi_n \left( \frac{n-2}{n-1} t \right),
\]

(5.42) is a minimum variance unbiased estimator of \( \beta^2 \).
THE LOGNORMAL DISTRIBUTION

The series defining \( \psi_n(t) \) converges only slowly and a convenient asymptotic form is provided by Finney [65]:

\[
\psi_n(t) = e^{\left(1 - \frac{t(t+1)}{n} + \frac{t^3(3t^2 + 22t + 21)}{6n^2}\right) + O\left(\frac{1}{n}\right)},
\]

so that large-sample approximations for \( a_1 \) and \( b_1^2 \) may be obtained. The authors have however calculated tables of \( \psi_n(t) \) and \( \chi_n(t) \); abbreviated versions of these appear as Appendix Tables A2 and A3.

Large-sample variances of \( a_1 \) and \( b_1^2 \) may also be calculated and are found to be

\[
D^2(a_1) = \frac{\alpha^4}{n} \left(\sigma^2 + \frac{\sigma^4}{2}\right),
\]

and

\[
D^2(b_1^2) = \frac{\alpha^4}{n} \left(4\sigma^4 + 2\sigma^4(2\eta^2 + 1)\right).
\]

5.42. THE METHOD OF MOMENTS

The method of moments gives estimators \( a_2 \) and \( b_2^2 \) immediately:

\[
a_2 = t_1 = \bar{x},
\]

and

\[
b_2^2 = t_2;
\]

or, in order to obtain an unbiased estimator of \( \beta^2 \), (5.47) may be replaced by

\[
b_2^2 = \bar{x}^2 = \frac{n}{n-1} t_2.
\]

These estimators have not minimum variance, for

\[
D^2(a_2) = \frac{\alpha^4}{n} \eta^2
\]

and

\[
D^2(b_2^2) = \frac{\alpha^4}{n} (\eta^4 + 6\eta^4 + 15\eta^8 + 16\eta^8 + 2\eta^8).
\]

The large-sample efficiencies are therefore given by

\[
\text{eff.} \{a_2\} = \frac{\sigma^2 + \frac{1}{2}\sigma^4}{\eta^2},
\]

and

\[
\text{eff.} \{b_2^2\} = \frac{4\sigma^4 + 2\sigma^4(2\eta^2 + 1)}{\eta^10 + 6\eta^10 + 15\eta^8 + 16\eta^8 + 2\eta^8}.
\]

The graphs of eff. \( \{a_2\} \) and eff. \( \{b_2^2\} \) against \( \sigma^2 \) are shown in Fig. 5.3. The efficiencies decrease as \( \sigma^2 \) increases; while there is little loss of efficiency in using \( a_2 \) rather than \( a_1 \) for moderate values of \( \sigma^2 \), there is a considerable loss involved in using \( b_2^2 \) instead of \( b_1^2 \) even for small values of \( \sigma^2 \).

5.43. THE METHOD OF QUANTILES

If \( a_3 \) and \( b_3^2 \) denote estimators of \( a \) and \( \beta^2 \) determined by the method of quantiles then

\[
a_3 = \exp \left\{ m_3 + \frac{1}{2}\eta_3^2 \right\},
\]

and

\[
b_3^2 = \exp \left\{ 2m_3 + \eta_3^2 \right\} \left( \exp \left\{ \eta_3^2 \right\} - 1 \right);
\]
where $m_3$ and $s_3^2$ are estimators of the form (5.26) and (5.27). The large-sample variances are found to be

$$D^2(a_3) = \alpha^2 D^2(m_3) + D^2(s_3^2),$$

and

$$D^2(b_3^2) = \alpha^4 \left[ 4y^4 D^2(m_3) + (2y^2 + 1)^2 D^2(s_3^2) \right];$$

The large-sample variances are found to be

$$D^2(a_3) = \alpha^2 D^2(m_3) + D^2(s_3^2),$$

and

$$D^2(b_3^2) = \alpha^4 \left[ 4y^4 D^2(m_3) + (2y^2 + 1)^2 D^2(s_3^2) \right];$$

so that the large-sample efficiencies of $a_3$ and $b_3^2$ are readily derived as

$$\text{eff. } \{a_3\} = \frac{D^2(a_3)}{\text{eff. } \{a_3\}} = \frac{1 + \frac{1}{2} \sigma^2}{\text{eff. } \{m_3\} + \text{eff. } \{s_3^2\}},$$

and

$$\text{eff. } \{b_3^2\} = \frac{D^2(b_3^2)}{\text{eff. } \{b_3^2\}} = \frac{2y^4 + \sigma^2(2y^2 + 1)^2}{\text{eff. } \{m_3\} + \text{eff. } \{s_3^2\}}.$$

† The covariance of $m_3$ and $s_3^2$ is of order less than $1/n$. 

---

Fig. 5.3. Efficiency of method of moments in estimation of $\alpha$ and $\beta^2$. 

---

62
4.8 THE LOGNORMAL DISTRIBUTION

Now from the discussion of §5.2, eff. \( m_3 \) is a maximum when the quantiles used are of order 27 and 73 \%, and eff. \( \{ s_3 \} \) when the orders are 7 and 93 \%, no matter what the value of \( \sigma^2 \) may be. From the form of (5.57) and (5.58) it is clear that the efficiencies of \( a_2 \) and \( b_2^2 \) are maximized when these particular quantiles are used to estimate \( m_2 \) and \( s_2^2 \). The variation with respect to \( \sigma^2 \) of the efficiencies on this basis are shown in Fig. 5.4; the efficiency of \( a_2 \) decreases from 81 to 65 \% as \( \sigma^2 \) increases from 0 to \( \infty \), while that of \( b_2 \) always remains above 65 \%.

![Fig. 5.4. Efficiency of method of quantiles in estimation of \( \alpha \) and \( \beta^2 \).](image)

5.44. THE GRAPHICAL METHOD

Estimates of \( \alpha \) and \( \beta^2 \) may be derived from estimates of \( \mu \) and \( \sigma^2 \) obtained by the graphical method of §5.2 with the use of formulae (2.7) and (2.8).

5.5. EXPERIMENTAL RESULTS (II): THE METHODS OF §5.4 APPLIED TO THE 65 SAMPLES

Estimates of \( \alpha \) and \( \beta^2 \) were computed by the different methods for all the artificial samples and are given in Appendix Tables B3 and B4. Empirical measures of efficiency, \( \Delta (m_2) \) and \( \Delta (s_2^2) \), calculated as in §5.3, were also obtained and the results are again presented in the form of four tables.

The pattern of our conclusions is similar to that for the estimation of \( \mu \) and \( \sigma^2 \) with two important exceptions. For estimating \( \alpha \) the method of moments is quite reliable and is advisable if there is any possibility of combining the results of several samples because of the simple additive nature of the estimators. Also the graphical method seems to be quite
ESTIMATION PROBLEMS: I

**TABLE 5.5. VALUES OF \( \Delta(\alpha_i) \) FOR A GROUPING BY SAMPLE SIZE**

<table>
<thead>
<tr>
<th>Sample size ( n )</th>
<th>Finney's method</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.2127</td>
<td>0.2337</td>
<td>0.1890</td>
<td>0.1995</td>
<td>0.1935</td>
<td>0.1833</td>
</tr>
<tr>
<td>64</td>
<td>0.1171</td>
<td>0.1180</td>
<td>0.1033</td>
<td>0.1353</td>
<td>0.0999</td>
<td>0.1322</td>
</tr>
<tr>
<td>128</td>
<td>0.0759</td>
<td>0.0763</td>
<td>0.1366</td>
<td>0.0571</td>
<td>0.0582</td>
<td>0.0660</td>
</tr>
<tr>
<td>256</td>
<td>0.0676</td>
<td>0.0930</td>
<td>0.9134</td>
<td>0.0618</td>
<td>0.0624</td>
<td>0.0672</td>
</tr>
<tr>
<td>512</td>
<td>0.0501</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All samples</td>
<td>0.1382</td>
<td>0.1477</td>
<td>0.1361</td>
<td>0.1351</td>
<td>0.1287</td>
<td>0.1375</td>
</tr>
</tbody>
</table>

**TABLE 5.6. VALUES OF \( \Delta(\sigma_i) \) FOR A GROUPING BY SIZE OF \( \sigma \)**

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>Finney's method</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2-0.4</td>
<td>0.0319</td>
<td>0.0416</td>
<td>0.0499</td>
<td>0.0432</td>
<td>0.0450</td>
<td>0.0428</td>
</tr>
<tr>
<td>0.5-0.7</td>
<td>0.0128</td>
<td>0.0123</td>
<td>0.0167</td>
<td>0.0149</td>
<td>0.0140</td>
<td>0.0158</td>
</tr>
<tr>
<td>0.8-1.0</td>
<td>0.0199</td>
<td>0.0197</td>
<td>0.0196</td>
<td>0.0176</td>
<td>0.0162</td>
<td>0.0167</td>
</tr>
<tr>
<td>All samples</td>
<td>0.1382</td>
<td>0.1477</td>
<td>0.1361</td>
<td>0.1351</td>
<td>0.1287</td>
<td>0.1375</td>
</tr>
</tbody>
</table>

**TABLE 5.7. VALUES OF \( \Delta(\beta_i^2) \) FOR A GROUPING BY SAMPLE SIZE**

<table>
<thead>
<tr>
<th>Sample size ( n )</th>
<th>Finney's method</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>1.0970</td>
<td>3.8520</td>
<td>0.9755</td>
<td>6.1440</td>
<td>1.0990</td>
<td>0.9553</td>
</tr>
<tr>
<td>64</td>
<td>0.3599</td>
<td>0.6240</td>
<td>0.3550</td>
<td>0.5628</td>
<td>0.4924</td>
<td>0.4814</td>
</tr>
<tr>
<td>128</td>
<td>0.3432</td>
<td>1.1440</td>
<td>0.3413</td>
<td>1.3530</td>
<td>0.3778</td>
<td>1.1800</td>
</tr>
<tr>
<td>256</td>
<td>0.6159</td>
<td>0.2689</td>
<td>0.6179</td>
<td>0.7512</td>
<td>0.5062</td>
<td>0.4777</td>
</tr>
<tr>
<td>512</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All samples</td>
<td>0.7207</td>
<td>2.1400</td>
<td>0.7545</td>
<td>3.0210</td>
<td>0.6872</td>
<td>0.8891</td>
</tr>
</tbody>
</table>

**TABLE 5.8. VALUES OF \( \Delta(\beta_i^4) \) FOR A GROUPING BY SIZE OF \( \sigma \)**

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>Finney's method</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Graphical method (i)</th>
<th>Graphical method (ii)</th>
<th>Graphical method (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2-0.4</td>
<td>0.0276</td>
<td>0.0244</td>
<td>0.0216</td>
<td>0.0234</td>
<td>0.0232</td>
<td>0.0177</td>
</tr>
<tr>
<td>0.5-0.7</td>
<td>0.2320</td>
<td>0.2645</td>
<td>0.2919</td>
<td>0.3815</td>
<td>0.3943</td>
<td>0.3627</td>
</tr>
<tr>
<td>0.8-1.0</td>
<td>1.1900</td>
<td>3.3870</td>
<td>1.2360</td>
<td>5.5300</td>
<td>1.1250</td>
<td>1.4520</td>
</tr>
<tr>
<td>All samples</td>
<td>0.7207</td>
<td>2.1400</td>
<td>0.7545</td>
<td>3.0210</td>
<td>0.6872</td>
<td>0.8891</td>
</tr>
</tbody>
</table>

good for \( \alpha \), and for \( \beta^2 \) when the sample size is not too small or the distribution too skew. It appears that in fitting the line by eye a bias in location is often compensated by an opposite bias in the slope.

Our recommendations would then be the following. If the cost of computation is not too great, use the method of maximum likelihood;
otherwise use either the method of moments or quantiles for $\alpha$ but avoid the use of moments for $\beta^2$. If there is any possibility of the combination of samples, or the further analysis of a set of samples, the method of moments is desirable for $\alpha$. For quick estimates the graphical method is convenient but provides no check on the accuracy of the estimate.

5.6. CONFIDENCE INTERVALS

5.61. CONFIDENCE INTERVALS FOR $\mu$ AND $\sigma^2$

Exact confidence intervals may be obtained for $\mu$ and $\sigma^2$ provided the maximum-likelihood estimators $\bar{y}$ and $\hat{\sigma}^2$ are used. For $\frac{\bar{y} - \mu}{\hat{\sigma}\sqrt{n}}$ is $t$-distributed and $\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$ is $\chi^2$-distributed, each with $n-1$ degrees of freedom. If $t_{p,n-1}$ denotes the $p$ percentage point of $t$ with $n-1$ degrees of freedom, then $\left(\bar{y} - t_{p,n-1}\frac{\hat{\sigma}}{\sqrt{n}}, \bar{y} + t_{p,n-1}\frac{\hat{\sigma}}{\sqrt{n}}\right)$ is an exact $p\%$ confidence interval for $\mu$. Similarly $\left(\frac{(n-1)\hat{\sigma}^2}{\chi^2_1}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_2}\right)$ is an exact confidence interval for $\sigma^2$, where $\chi^2_1$, $\chi^2_2$ are appropriate percentage points for the $\chi^2$ distribution. If estimators other than maximum likelihood have been used, only approximate large-sample confidence intervals may be obtained. For large samples an estimator $m$ is asymptotically normal with mean $\mu$ and variance $D^2(m)$ obtained by replacing $\sigma^2$ by $\hat{s}^2$. The interval is then $\left(m - \frac{D(m)}{\sqrt{n}}, m + \frac{D(m)}{\sqrt{n}}\right)$, where $\nu$ is the appropriate $N(0,1)$ percentage point. Similar remarks apply to confidence intervals for $\sigma^2$.

It is worth noting that exact confidence intervals may be obtained for any monotonic function of $\mu$ alone, or of $\sigma^2$ alone. There are some functions of interest of this type, for instance, the median $\ell$ and the Lorenz measure of concentration (see Chapter 11) which is $2N\left(\frac{\sigma}{\sqrt{2}}, 0, 1\right) - 1$.

Another such measure is the proportion of the population less than the mean, that is

$$P(X \leq \mu) = \int_0^\mu d\Lambda(x | \mu, \sigma^2) = N\left(\frac{\mu}{\sigma}, 0, 1\right).$$

5.62. CONFIDENCE INTERVALS FOR $\alpha$ AND $\beta^2$

Theory provides no means of obtaining exact confidence intervals for $\alpha$ and $\beta^2$; all that can be said is that $a$ and $b^2$ may be treated as asymptotically normal with means $\alpha$ and $\beta^2$ and variances $D^2(a)$ and $D^2(b^2)$ respectively. Thus, if the method of moments is used, a large-sample confidence interval for $\alpha$ may be taken as $\left(a - \frac{b}{\sqrt{n}}, a + \frac{b}{\sqrt{n}}\right)$. 

63
5.7. THE GROUPING OF OBSERVATIONS

When observations are given only as grouped frequencies there are certain difficulties in the various methods which may alter the recommendations of §§5.3 and 5.5.

5.71. GROUPING FOR μ AND σ²

The method of maximum likelihood can be applied to grouped data and Gjeddebaek[90] has given an account of this application, with two tables to facilitate the calculations. Suppose that the intervals are \((x_{i-1}, x_i)\), and that \(n_i\) of the total of \(n\) observations fall within the \(i\)th interval. Then the likelihood of the sample is proportional to

\[
\prod_i \left( \Lambda(x_i \mid \mu, \sigma^2) - \Lambda(x_{i-1} \mid \mu, \sigma^2) \right)^{n_i}. 
\]  

(5.60)

The log likelihood function \(L\) is thus

\[
L = \sum_i n_i \log \left( N \left( \frac{y_i - \mu}{\sigma} \right) - N \left( \frac{y_{i-1} - \mu}{\sigma} \right) \right),
\]

(5.61)

where \(y_i = \log x_i\). The likelihood equations are

\[
\frac{\partial L}{\partial \mu} = -\frac{1}{\sigma} \sum_i n_i \frac{N' \left( \frac{y_i - \mu}{\sigma} \right) - N' \left( \frac{y_{i-1} - \mu}{\sigma} \right)}{N \left( \frac{y_i - \mu}{\sigma} \right) - N \left( \frac{y_{i-1} - \mu}{\sigma} \right)},
\]

(5.62)

\[
\frac{\partial L}{\partial \sigma^2} = - \frac{1}{2\sigma^2} \sum_i n_i \frac{N' \left( \frac{y_i - \mu}{\sigma} \right) - N' \left( \frac{y_{i-1} - \mu}{\sigma} \right)}{N \left( \frac{y_i - \mu}{\sigma} \right) - N \left( \frac{y_{i-1} - \mu}{\sigma} \right)}
\]

\[
= \frac{1}{2\sigma^2} \sum_i n_i \frac{N' \left( \frac{y_i - \mu}{\sigma} \right) - N' \left( \frac{y_{i-1} - \mu}{\sigma} \right)}{N \left( \frac{y_i - \mu}{\sigma} \right) - N \left( \frac{y_{i-1} - \mu}{\sigma} \right)},
\]

(5.63)

since \(N' \left( \frac{y_i - \mu}{\sigma} \right) = -yN' \left( \frac{y_i - \mu}{\sigma} \right)\). By introducing and tabulating the functions

\[
z_1(u, v) = \frac{N' \left( u + v \right) - N' \left( u \right)}{N \left( u + v \right) - N \left( u \right)},
\]

(5.64)

and

\[
z_2(u, v) = \frac{N' \left( u + v \right) - N' \left( v \right)}{N \left( u + v \right) - N \left( u \right)},
\]

(5.65)

Gjeddebaek simplifies the interpolation necessary for the solution of the equations. An expression for the variance matrix of the estimators is also derived. An alternative method to that of Gjeddebaek would be the use of the method of scoring as for probit analysis (see Chapter 7) giving an iterative set of equations. A further alternative method is provided by Grundy[92]; this is described in §9.6, where truncation and censorship are considered.
THE LOGNORMAL DISTRIBUTION

It should be noted that the method does not require the intervals to be equal either in the x-scale or the y-scale, but soon becomes laborious if the number of intervals is large. It is therefore mainly of use when the data are coarsely grouped.

Since the lognormal curve has high-order contact at the ends of the range, moments obtained from grouped data may be corrected in the usual way by use of Sheppard's corrections. So the method of moments may be applied to grouped data by equating the corrected moments to their theoretical values as in §5.2.

The method of quantiles may be readily applied to grouped data although it may not be possible to obtain the quantiles which provide the maximum efficiency of estimation. Two alternatives are then available: either interpolation is carried out to provide approximations to the quantiles giving maximum efficiency; or ends of intervals providing quantiles of order nearest to those giving maximum efficiency are used in formulae (5.23) and (5.24).

Little need be said about the graphical method since the data are in exactly the form required by it, and the method is easier to apply than when original observations are used. Since the reliability of the graphical procedure is unchanged by grouping, its reliability relative to that of the other methods under grouping will improve; in particular, it may be preferred to the method of quantiles when the grouping does not permit the choice of quantiles close to those of maximum efficiency; this statement has been tested empirically by the authors.

5.72. GROUPING FOR $\alpha$ AND $\beta^2$

No maximum-likelihood solution to the problem has been put forward so far. For the other methods the remarks of the preceding paragraphs hold.

5.8. SOME SPECIAL DEVICES OF ESTIMATION

In addition to the more orthodox methods of estimation, outlined in the earlier part of this chapter, in certain circumstances other procedures may be derived from certain special features of the distribution. These are illustrated by three examples.

First, it was pointed out in §5.51 that the proportion of the population below the mean, $P\{X \leq \mu\}$ is a function of $\sigma$ alone, namely $N\left(\frac{\mu}{\sigma}, 0, 1\right)$.

A possible method of estimating $\sigma$, without the labour of estimating $\mu$ at the same time, is to set the proportion of sample values below the sample mean equal to $N\left(\frac{\bar{X} - \mu}{\sigma}, 0, 1\right)$ and obtain $s$ (and hence $\sigma$) from a table of the normal integral.†

A second useful property is the simple form of the moment distribution (see §9.5). If the data are given in the form of sample moment

† Values of $N\left(\frac{\bar{X}}{\sigma^2}, 0, 1\right)$ are tabulated against $\sigma$ in Appendix Table A1.
frequencies, the parameters of the moment distribution may be estimated by a standard method, and the parameters of the basic distribution determined from these. In fact there is no reason why information may not be available in different detail on both the distribution itself and on a moment distribution. An elegant use of this approach has been made by Hinch[101] in dealing with particle-size data. This is discussed more fully in Chapter 10.

The third example involves the coefficient of variation $\gamma$ which depends only on $\sigma^2$. Suppose that a number of samples are given and that it may be assumed that they are from populations with the same $\sigma^2$ though possibly different $\mu$. Further, information is available on the mean and standard deviation only of each sample. If then the standard deviations are plotted against the means the points should lie roughly on a straight line of slope $(e^{\sigma^2} - 1)^{1/2}$ and hence $\sigma^2$ may be estimated.

5.9. Summary of Estimation Procedures for the Two-parameter Distribution

It may help the reader to summarize here the main conclusions of this chapter. This is most conveniently done by classifying the estimation problems and noting under each class the characteristics of the main methods of estimation.

5.91. Ungrouped Data: Estimation of $\mu$ and $\sigma^2$

1. Method of maximum likelihood: the estimators are sufficient and cannot therefore be bettered even in small samples; but the method is costly if the number of observations is large.

2. Method of moments: inferior in efficiency to the method of quantiles, and the efficiency falls rapidly as $\sigma^2$ increases.

3. Method of quantiles: easily applied; efficiencies of 81 and 65% respectively for $\mu$ and $\sigma^2$ are obtained, when (27, 73; 7, 93%) quantiles are used; these efficiencies are independent of $\sigma^2$.

4. Graphical method: easily applied, and simultaneously provides a test of lognormality; its efficiency is, however, not calculable, and our experiment shows it is usually less reliable than the numerical methods. The reliability is liable to decrease as $\sigma^2$ increases, but not noticeably to improve as the sample size increases above 100.

5.92. Grouped Data: Estimation of $\mu$ and $\sigma^2$

1. Method of maximum likelihood: becomes extremely cumbersome as the number of groups increases.

2. Method of moments: applied with Sheppard’s corrections; falls in efficiency rapidly as $\sigma^2$ increases.

3. Method of quantiles: the method declines in efficiency if the data are so grouped that it is necessary to choose quantiles distant from the most efficient quantiles or pairs of quantiles that are asymmetrically
placed. If this is the case it may be preferable to interpolate for the most efficient quantile pairs.

(4) **Graphical method:** since the data are in any case grouped for the application of this method, the previous remarks apply. The relative reliability of the method is therefore greater if the data are only available in grouped form.

**5.93. UNGROUPED DATA: ESTIMATION OF \( \alpha \) AND \( \beta^2 \)**

1. **Finney's method:** equivalent to maximum likelihood and therefore cannot be bettered; the \( \gamma \)-function is laborious to compute if tables, or an automatic computer, are not available.

2. **Method of moments:** not efficient for \( \beta^2 \), but good for \( \alpha \) and easy to apply; its usefulness is enhanced when there is the possibility of combining information from several samples.

3. **Method of quantiles:** easily applied by using the best quantile estimators of \( \mu \) and \( \sigma^2 \); the efficiencies vary with \( \sigma^2 \), falling to a minimum of 65% in both cases.

4. **Graphical method:** easy to derive from the graphical estimates of \( \mu \) and \( \sigma^2 \); the reported experiment shows results that are relatively better than those for \( \mu \) and \( \sigma^2 \), since it appears that there is a tendency for biases in the estimates of these parameters to counteract each other.

**5.94. GROUPED DATA: ESTIMATION OF \( \alpha \) AND \( \beta^2 \)**

1. **Method of maximum likelihood:** no tractable form has been found.

2. **Method of moments:** applied with Sheppard's corrections; remarks on efficiency for ungrouped case still apply.

3. **Method of quantiles:** uses the best available quantile estimates of \( \mu \) and \( \sigma^2 \).

4. **Graphical method:** as for ungrouped data; again, if the data are only available in grouped form, the method may be preferred to that of quantiles, and to that of moments if \( \sigma^2 \) is large.

**Note.** In the discussion of grouped data in §§ 5.7 and 5.9 the emphasis has been given to the case where the statistician receives his data in an arbitrarily grouped form. On the other hand the grouping may be carried out by the statistician himself to lighten the subsequent calculations. If then the lengths of the class intervals are chosen to be in geometric progression, the application of the method of moments to the transformed, grouped data is equivalent to applying the method to data grouped into intervals of equal length, and Sheppard's corrections may be applied to the raw moments. It is known from normal theory that this procedure is of high efficiency, provided that the grouping is not too coarse.
6.1. Introduction

When the lower bound $\tau$ of a lognormal distribution is not known from prior information, the problem of estimating the three parameters $\tau$, $\mu$ and $\sigma^2$ is more complicated than any we have yet treated. Following the lines of the last chapter we review a number of alternative procedures and, as far as possible, compare their efficiencies. The discussion for this three-parameter case is summarized in §6.4. The four-parameter distribution is treated briefly in the final sections of the chapter.

6.2. Estimation of the Parameters of the Three-Parameter Distribution

The given sample is again supposed to consist of the values $x_1 \ldots x_n$. In addition to the methods of maximum likelihood, moments, quantiles and probability paper we discuss a method, due to Cohen [42], based on the least-sample value, and another, due to Kemsley [121], which is a mixture of the methods of moments and quantiles.

6.21. The Method of Maximum Likelihood

The range of the variate now depends on $\tau$, one of the parameters to be estimated, so that the maximum-likelihood estimators cannot be assumed to possess the desirable properties of consistence, asymptotic normality and minimum variance without special investigation. Nevertheless, some writers [42, 210] have attempted to estimate $\tau$, $\mu$ and $\sigma^2$ by this method. If $\hat{\tau}$, $\hat{\mu}$ and $\hat{\sigma}^2$ denote the estimators the maximum-likelihood equations are readily obtained as

$$m_1 = \frac{1}{n} \sum \log (x - t),$$

$$s_1^2 = \frac{1}{n} \left[ \sum (\log (x - t))^2 - m_1^2 \right],$$

and

$$(s_1^2 - m_1) \Sigma \frac{1}{x - t} + \Sigma \frac{\log (x - t)}{x - t} = 0.$$

Wilson and Worcester [210] suggest solving these equations by 'trial and
error'. Cohen [42] eliminates $m_1$ and $s_1^2$ from the equations obtaining the equation
\[
\theta(t) = \frac{1}{x-t_1} \left[ -\frac{1}{n} \sum \{\log (x-t_i)\}^2 - \frac{1}{n} \sum \log (x-t_i) \right]
- \frac{1}{n^2} \left( \sum \log (x-t_i) \right)^2 + \frac{1}{n} \left( \frac{\log (x-t_i)}{x-t_1} \right) = 0,
\] (6.4)
from which to obtain $t_1$ by inverse interpolation. This is by no means an easy task even for samples of moderate size, since $\theta(t)$ is awkward to compute. The graph of $\theta(t)$ against $t$ is shown in Fig. 6.1 for one of the samples of size 64 with $\sigma = 0.7$. It will be seen that $\theta(t)$ is very sensitive to small changes of $t$ in the neighbourhood of the solution of equation (6.4).

In our view the difficulty of computation coupled with a suspicion of the underlying theory leaves little incentive to recommend the method. In the previous chapter we were able to use the variances of the maximum-likelihood estimators as a standard against which to compare the efficiencies of the other methods; here we do not feel justified in using such a procedure.

6.22. COHEN’S LEAST SAMPLE VALUE METHOD

It has been pointed out that the parameter $\tau$ determines the range of the variate; it is well known that in such cases if a sufficient estimator exists it must be a function of the least sample value. This fact underlies an alternative to the method of maximum likelihood suggested by Cohen [42]. He allows equations of the form (6.1) and (6.2) to stand, but replaces (6.3) by one based on the least sample value, say $x_\circ$. If $x_\circ$ occurs $n_0$ times in the sample it may be regarded as the sample quantile of order $n_0/n$ and the third equation equates $x_\circ$ to the population quantile of this order. Thus if $t_\circ$, $m_\circ$ and $s_\circ^2$ are the estimators, their determining equations are
\[
m_\circ = \frac{1}{n} \sum \log (x-t_\circ),
\] (6.5)
\[
s_\circ^2 = \frac{1}{n} \sum \{\log (x-t_\circ)\}^2 - m_\circ^2,
\] (6.6)
\[
x_\circ = t_\circ + e^{m_\circ + \frac{1}{2} s_\circ^2},
\] (6.7)
where $\nu$ is the $N(0,1)$ quantile of order $n_0/n$. The method may thus be regarded as a mixture of the methods of maximum likelihood and of quantiles, where the order of the quantile used is not determined in advance. Elimination of $m_\circ$ and $s_\circ^2$ from equations (6.5) to (6.7) yields the equation
\[
\phi(t_\circ) = \log (x_\circ - t_\circ) - \frac{1}{n} \sum \log (x-t_i)
- \nu \left[ \frac{1}{n} \sum \{\log (x-t_i)\}^2 - \frac{1}{n^2} \left( \sum \log (x-t_i) \right)^2 \right] = 0,
\] (6.8)
and Cohen [42] uses inverse interpolation to determine \( t_c \). The computation of \( \phi(t) \) is certainly simpler than that of \( \theta(t) \); the graph of \( \phi(t) \) against \( t \) is given in Fig. 6.1 for the same sample as for the graph of \( \theta(t) \). The authors have recently devised an automatic computing programme for the solution of this equation by the 'rule of false position'. Its application to a few artificial samples suggests that it is more reliable than the method of maximum likelihood. No expressions have yet been obtained for the variances in this case.

Fig. 6.1. The functions \( \theta(t) \) and \( \phi(t) \) for a sample of size 64.

6.23. THE METHOD OF MOMENTS

Let \( \lambda'_j \) and \( \lambda_j \) denote the population moments about the origin and about the mean respectively; then

\[
\begin{align*}
\lambda'_1 &= r + e^\mu (1 + \eta^2), \\
\lambda'_2 &= e^{2\mu} (1 + \eta^2) \eta^2, \\
\lambda'_3 &= e^{3\mu} (1 + \eta^2)^2 (\eta^4 + 3\eta^4),
\end{align*}
\]

so that

\[
\eta^2 + 3\eta = \frac{\lambda'_3}{\lambda'_2}. \tag{6.12}
\]

If \( l'_j \) and \( l_j \) denote the sample moments about the origin and mean respectively and \( l_2, m_2 \) and \( s_2^2 \) the estimators of \( r, \mu \) and \( \sigma^2 \) then the method of moments requires first the solution of the equation

\[
\eta^2 + 3\eta = \frac{l_2}{l'_2} \tag{6.13}
\]

so

\[
= k, \text{ say,}
\]

72
for \( u \); the estimators are then obtained from

\[
\begin{align*}
\hat{s}_2 &= \log (1 + u^2), \\
\hat{m}_2 &= \frac{1}{4} \left[ \log t_2 - \log [u^2(1 + u^2)] \right], \\
\text{and} \\
t_2 &= t_1 - e^{s_2}(1 + u^2). 
\end{align*}
\] (6.14, 6.15, 6.16)

Appendix Table A4 gives the values of \( u \) and \( s^2 \) corresponding to values of \( k \) equal to 0 (0, 2) to (1, 24).

This method has been applied by Wicksell [203], Gumbel [93], and Yuan [216]; equation (6.12) was presented in a slightly different form by Yuan who also published a table to assist in the solution. Since the method of moments is not an efficient method in the \( \Lambda (\mu, \sigma^2) \) case except for small values of \( \sigma^2 \), it is to be expected that the above method is also inefficient; this is corroborated by the evidence from the artificial samples. Because of this inefficiency it does not seem worthwhile to reproduce here the cumbersome formulae for the large sample variances of \( t_2, m_2 \) and \( s_2^2 \). They may be obtained by the variational method and have been calculated, at Wicksell’s suggestion, by Nydell [149].

### 6.24. The Method of Quantiles

Three sample quantiles must now be used and the obvious choice of orders is \( q, \frac{1}{2} \) and \( 1 - q \), where \( 0 < q < \frac{1}{2} \); we denote the sample quantile of order \( q \) by \( x_q \) and the quantiles of order \( q \) of the \( N(\mu, \sigma^2) \) distribution by \( \tau_q \). Then

\[
\nu_{1-q} = -\nu_q = \nu_1 
\] (6.17)
say. If \( t_3, m_3 \) and \( s_3^2 \) denote the estimators, the determining equations are

\[
\begin{align*}
x_q &= t_3 + e^{m_2} \varepsilon_3, \\
x_1 &= t_3 + e^{m_2} \\
and \\
x_{1-q} &= t_3 + e^{m_2} \varepsilon_3, 
\end{align*}
\] (6.18, 6.19, 6.20)

from which may be deduced the following system of solution:

\[
\begin{align*}
x_3 &= \frac{1}{\nu} \left( \log (x_{1-q} - x_1) - \log (x_1 - x_q) \right), \\
m_3 &= \log (x_1 - x_q) - \log (1 - e^{-m_2}) \\
\text{and} \\
t &= x_1 - e^{m_2}. 
\end{align*}
\] (6.21, 6.22, 6.23)

Again the variances of the estimators may be obtained but the formulae are cumbersome, and it is difficult to give a theoretical measure of the efficiency of this method; an empirical estimate of this is found from the artificial samples in §6.3. It is just possible that the relation (6.21) will give a negative value for \( s_3^2 \); in practice this is only likely to happen when the sample is small and especially if \( \sigma^2 \) is also small. So far we have said nothing about what value of \( q \) should be used; our conjecture is that a good general rule is to take \( q = 0.05 \).
Kemsley's method

Kemsley [121] has used an interesting estimation procedure which is a mixture of the methods of moments and quantiles. He equates the sample mean \( \bar{x} \) and the sample quantiles \( x_{q} \) and \( x_{1-q} \) to the corresponding population values; thus if \( \mu_k \) and \( s_k \) are his estimators then

\[
x_{q} = \mu_{k} + \exp \{ m_k - \nu_k \}, \tag{6.24}
\]

\[
\bar{x} = \mu_{k} + \exp \{ m_k + \frac{1}{2} s_k^2 \}, \tag{6.25}
\]

\[
x_{1-q} = \mu_{k} + \exp \{ m_k + \nu_k \}. \tag{6.26}
\]

From which the equation

\[
f(s^2_k) = \frac{\exp \{ \frac{1}{2} s^2_k \} - \exp \{ - \nu_k \}}{\exp \{ \nu_k \} - \exp \{ \frac{1}{2} s^2_k \}}
\]

may be derived. From this \( s_k^2 \) is to be determined. If \( s_k < 2\nu \), \( f(s^2_k) \) is positive, and if \( s_k > 2\nu \), \( f(s^2_k) \) is negative, and there is a range of values of \( f \) for which no real value of \( s_k^2 \) may be found. That part of the graph for which \( f(s^2_k) \) is positive is given plotted against \( s_k^2 \) in Fig. 6.2 for \( q = 0.05, 0.10 \) and \( 0.20 \). It will be seen that certain difficulties may arise: from the sample a value of \( f \) may be found which does not lead to any estimate for \( \sigma^2 \); this is certainly possible in practice as will be seen when we apply the method to the artificial samples; again it may be possible to obtain
two values of \( s^2 \) although usually it will be easy to decide which is the reasonable one. Kemysley has used the method with \( q = 0.05 \) and, though the difficult problem of obtaining expressions for the large-sample variances has not been undertaken, we conjecture that this value for \( q \) is reasonably efficient, at least for \( \sigma^2 \).

Our suggestion for the solution of equation (6.27) is first to determine \( s^2 \) approximately by reference to the graph and then to proceed to a more accurate value by inverse interpolation; \( m_k \) and \( t_k \) are then obtained in succession from the relations

\[
m_k = \log (x_{1-q} - \bar{x}) - \log \left( \exp \{ps_k\} - \exp \{\frac{1}{2}r_k^2\} \right)
\]

and

\[
t_k = x_{1-q} - \exp \{m_k + ps_k\}.
\]

6.26. THE GRAPHICAL METHOD

If \( r \) were known, an array of points lying roughly on a straight line would arise from plotting \( x - r \) against the proportion of sample values not exceeding \( x \) on logarithmic probability paper. The graphical method of estimation is therefore to try different feasible values \( t \) for \( r \) until something approximating to a straight line is arrived at. If \( t \) underestimates or overestimates \( r \) there should be a tendency towards curvature in the directions indicated in Fig. 6.3, and an adjustment should be made in the correct direction **until the inflexion change de sens. On a alors la meilleure valeur de \( t \) pour l’ajustement**, as Gibrat[88] put it. It is obvious from our description that the method is much more an art than a science, but it is often useful for a preliminary investigation (see, for example, § 11.6). It may be seen from Fig. 6.3 that if \( r \) is underestimated the slope of the array, and therefore \( \sigma^2 \), is underestimated, and conversely. The same feature applies to the numerical methods of estimation.

6.27. SPECIAL DEVICES

As for \( \Lambda(\mu, \sigma^2) \), there are one or two devices which may be used for estimation in special circumstances. For instance, the proportion in the population below the population mean is still \( N\left( \frac{\sigma}{2} \mid 0, 1 \right) \), so that an estimate of \( \sigma^2 \) may be obtained by equating the proportion below the sample mean to this value.

Again, the standard deviation \( \beta \) is related to the mean \( \alpha' \) by the relation

\[
\beta = \eta(\alpha' - \tau),
\]

where \( \eta^2 = \sigma^2 - 1 \) as before. Suppose that a number of samples are involved and that there are good reasons for believing that they come from populations with the same \( \tau \) and \( \sigma^2 \). If the sample standard deviations are plotted against the sample means, the resulting points should lie on a straight line. The slope of this line should be \((\sigma^2 - 1)^{\frac{1}{2}}\), and its intercept on the sample mean axis should be \( \tau \). Thus estimates of \( \sigma^2 \) and \( \tau \) may be obtained; the method has been used by Kleczkowski[126] (cf. § 10.6).
6.3. EXPERIMENTAL RESULTS: COMPARISON OF KEMSLEY’S METHOD AND THE METHODS OF MOMENTS AND QUANTILES

A comparison of Kemsley's method with the methods of moments and quantiles was carried out by applying them to all the artificial samples. Kemsley's method and the method of quantiles were both applied at \( q = 0.05 \) and at \( q = 0.10 \). First we record the cases in which the methods were impossible to apply: these are shown in Table 6.1 for different sample sizes and in Table 6.2 for the grouping by \( \sigma \). The conclusion from these results is that Kemsley's method should not be applied with \( q = 0.1 \); the other outcome is that failures are to be expected for small samples with small values of \( \sigma^2 \).

Values of \( \Delta(t) \), \( \Delta(m) \) and \( \Delta(x^2) \) have been calculated as in Chapter 5; for any particular method these values are based only on the samples for which the method was possible, and this should be borne in mind throughout the comparison. The tables are presented exactly as in
Chapter 5. From most points of view the quantile method using 5, 50 and 95 % points is a shade better than Kemsley's method using the mean and the 5 and 95 % points. These are decidedly better than the corresponding methods using 10 and 90 % points and than the method of moments which is clearly inefficient.

**TABLE 6.1. CASES FOR WHICH THE QUANTILE AND KEMSLEY'S METHODS WERE IMPOSSIBLE, CLASSIFIED BY SAMPLE SIZE**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Method of quantiles</th>
<th>Kemsley's method</th>
<th>Total no. of samples available</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 %</td>
<td>10 %</td>
<td>5 %</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>1</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>128</td>
<td>-</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>256</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>512</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>All samples</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

**TABLE 6.2. CASES FOR WHICH THE QUANTILE AND KEMSLEY'S METHODS WERE IMPOSSIBLE, CLASSIFIED BY SIZE OF σ**

<table>
<thead>
<tr>
<th>σ</th>
<th>Method of quantiles</th>
<th>Kemsley's method</th>
<th>Total no. of samples available</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 %</td>
<td>10 %</td>
<td>5 %</td>
</tr>
<tr>
<td>0-2-04</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0-5-07</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0-8-10</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>All samples</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

**TABLE 6.3. VALUES OF Δ(t_i) FOR A GROUPING BY SAMPLE SIZE**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Kemsley's method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 %</td>
<td>10 %</td>
<td>5 %</td>
</tr>
<tr>
<td>32</td>
<td>3-3540</td>
<td>0-3853</td>
<td>1-1850</td>
</tr>
<tr>
<td>64</td>
<td>3-7149</td>
<td>0-7650</td>
<td>0-2439</td>
</tr>
<tr>
<td>128</td>
<td>3-4573</td>
<td>0-2282</td>
<td>0-3535</td>
</tr>
<tr>
<td>256</td>
<td>3-4725</td>
<td>0-1007</td>
<td>0-4103</td>
</tr>
<tr>
<td>512</td>
<td>3-2681</td>
<td>0-0000</td>
<td>0-9955</td>
</tr>
<tr>
<td>All samples</td>
<td>1-9140</td>
<td>0-4634</td>
<td>0-6415</td>
</tr>
</tbody>
</table>

**TABLE 6.4. VALUES OF Δ(t_i) FOR A GROUPING BY SIZE OF σ**

<table>
<thead>
<tr>
<th>σ</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Kemsley's method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 %</td>
<td>10 %</td>
<td>5 %</td>
</tr>
<tr>
<td>0-2-04</td>
<td>3-2620</td>
<td>0-8098</td>
<td>0-8108</td>
</tr>
<tr>
<td>0-5-07</td>
<td>3-7774</td>
<td>0-1065</td>
<td>0-2305</td>
</tr>
<tr>
<td>0-8-10</td>
<td>3-7213</td>
<td>0-1046</td>
<td>0-7176</td>
</tr>
<tr>
<td>All samples</td>
<td>1-9140</td>
<td>0-4634</td>
<td>0-6415</td>
</tr>
</tbody>
</table>
6.4. Summary of Conclusions for the Three-Parameter Distribution

The first conclusion from the discussion of §§ 6.2 and 6.3 is that of the relative weakness of existing theory when a range parameter is to be determined from sample data. In the circumstances we recommend the use of Cohen's method (if facilities are available); this uses in (6.1)

### Table 6.5. Values of $\Delta(m_l)$ for a Grouping by Sample Size

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Kemslcy's method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>32</td>
<td>0.9335</td>
<td>0.9244</td>
<td>0.7033</td>
</tr>
<tr>
<td>64</td>
<td>0.7906</td>
<td>0.7999</td>
<td>0.7997</td>
</tr>
<tr>
<td>128</td>
<td>0.4113</td>
<td>0.2791</td>
<td>0.1314</td>
</tr>
<tr>
<td>256</td>
<td>0.4088</td>
<td>0.1347</td>
<td>0.1347</td>
</tr>
<tr>
<td>512</td>
<td>0.3949</td>
<td>0.0874</td>
<td>0.0874</td>
</tr>
<tr>
<td>All samples</td>
<td>0.7000</td>
<td>0.4765</td>
<td>0.5078</td>
</tr>
</tbody>
</table>

### Table 6.6. Values of $\Delta(m_l)$ for a Grouping by Size of $\sigma$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Kemslcy's method</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2-0.4</td>
<td>0.9031</td>
<td>0.5677</td>
<td>0.6719</td>
</tr>
<tr>
<td>0.5-0.7</td>
<td>0.6665</td>
<td>0.6212</td>
<td>0.3320</td>
</tr>
<tr>
<td>0.8-1.0</td>
<td>0.6658</td>
<td>0.4728</td>
<td>0.4964</td>
</tr>
<tr>
<td>All samples</td>
<td>0.7000</td>
<td>0.4765</td>
<td>0.5078</td>
</tr>
</tbody>
</table>

### Table 6.7. Values of $\Delta(s)$ for a Grouping by Sample Size

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Kemslcy's method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>32</td>
<td>0.1668</td>
<td>0.3223</td>
<td>0.4598</td>
</tr>
<tr>
<td>64</td>
<td>0.2861</td>
<td>0.2566</td>
<td>0.2683</td>
</tr>
<tr>
<td>128</td>
<td>0.2715</td>
<td>0.2003</td>
<td>0.2592</td>
</tr>
<tr>
<td>256</td>
<td>0.3392</td>
<td>0.1044</td>
<td>0.1293</td>
</tr>
<tr>
<td>512</td>
<td>0.3714</td>
<td>0.1044</td>
<td>0.1293</td>
</tr>
<tr>
<td>All samples</td>
<td>0.3034</td>
<td>0.2423</td>
<td>0.3060</td>
</tr>
</tbody>
</table>

### Table 6.8. Values of $\Delta(s)$ for a Grouping by Size of $\sigma$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Method of moments</th>
<th>Method of quantiles</th>
<th>Kemslcy's method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>0.2-0.4</td>
<td>0.0768</td>
<td>0.1903</td>
<td>0.2438</td>
</tr>
<tr>
<td>0.5-0.7</td>
<td>0.2934</td>
<td>0.1860</td>
<td>0.2438</td>
</tr>
<tr>
<td>0.8-1.0</td>
<td>0.3141</td>
<td>0.1215</td>
<td>0.4821</td>
</tr>
<tr>
<td>All samples</td>
<td>0.3034</td>
<td>0.2423</td>
<td>0.3060</td>
</tr>
</tbody>
</table>
and (6.2) what would be the maximum-likelihood expressions for \( \mu \) and \( \sigma^2 \) if \( \tau \) were known, and in (6.7) what would be a quantile expression for \( \tau \) if \( \mu \) and \( \sigma^2 \) were known. Combining our experience of the artificial samples with an intuition derived from the discussion of Chapter 5, however, our recommendation in most cases would be to adopt the full quantile method (which requires no iterative solutions) of equations (6.24)-(6.26), choosing whenever possible the 5 and 95% quantiles in conjunction with the median.

Perhaps it is not out of place to remark again here that estimation of the threshold parameter from the data of a single sample should be avoided whenever possible; and that Kieczkowski's device of using a graphical method to combine the evidence of several samples deserves to be considered when non-sample evidence for the value of \( \tau \) cannot be found. We draw attention to the practical danger caused by the intrusion of non-homogeneous observations in §10.10. A further disadvantage of relying on the evidence of a single sample is not only that \( \tau \) itself is difficult to determine but that the estimators of \( \mu \) and \( \sigma^2 \) are much less reliable for a three-parameter than for a two-parameter population. This is to be seen in the figures presented in Tables 6.9 and 6.10, where the \( \Delta \)-statistics for \( m \) and \( s^2 \) are compared for the quantile method alone.

**Table 6.9. Quantile Methods Compared for the Two- and Three-Parameter Distributions: by Sample Size**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( \Delta(m) )</th>
<th>( \Delta(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Lambda(\mu, \sigma^2) )</td>
<td>( \Lambda(\tau, \mu, \sigma^2) )</td>
</tr>
<tr>
<td>(27, 73%)</td>
<td>0.01558</td>
<td>0.05244</td>
</tr>
<tr>
<td>(50, 95%)</td>
<td>0.02799</td>
<td>0.06150</td>
</tr>
<tr>
<td>(75, 95%)</td>
<td>0.02791</td>
<td>0.04410</td>
</tr>
<tr>
<td>(93%)</td>
<td>0.02791</td>
<td>0.04410</td>
</tr>
<tr>
<td>(5, 95%)</td>
<td>0.02791</td>
<td>0.04410</td>
</tr>
<tr>
<td>All samples</td>
<td>0.02791</td>
<td>0.04410</td>
</tr>
</tbody>
</table>

**Table 6.10. Quantile Methods Compared for the Two- and Three-Parameter Distributions: by Size of \( \sigma \)**

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( \Delta(m) )</th>
<th>( \Delta(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Lambda(\mu, \sigma^2) )</td>
<td>( \Lambda(\tau, \mu, \sigma^2) )</td>
</tr>
<tr>
<td>(2, 4)</td>
<td>0.00397</td>
<td>0.05677</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>0.00397</td>
<td>0.05677</td>
</tr>
<tr>
<td>(3, 10)</td>
<td>0.00397</td>
<td>0.05677</td>
</tr>
<tr>
<td>All samples</td>
<td>0.00397</td>
<td>0.05677</td>
</tr>
</tbody>
</table>
6.5. Estimation of the Parameters of the Four-parameter Distribution

We have seen how much more difficult estimation becomes when we move from the two-parameter to the three-parameter case; when we proceed to the four-parameter case we can expect to be in much more serious difficulties. Every method except the method of quantiles may be dismissed immediately as mathematically intractable, and even for this method the determination is by no means easy. Four sample quantiles must be chosen and then the four equations of the form

$$\log \frac{x_i - \tau}{\theta - x_i} = \mu + \nu_i \sigma,$$

(6.31)

where $\nu_i$ is the $N(0, 1)$ quantile of order $q$, must be solved by some kind of successive approximation. Johnson[111] illustrates the method by an example on the distribution of cloudiness at Greenwich over a period of fifteen years.
CHAPTER 7
THE LOGNORMAL DISTRIBUTION AND
PROBIT ANALYSIS

Lucio. Assay the power you have.
Measure for Measure

7.1. INTRODUCTION

One of the most fertile fields of application for the lognormal distribution has been found in the statistical technique now generally known as probit analysis. This technique was finally evolved by Gaddum[75] in 1933 and Bliss[20] in 1934 after various attempts by earlier writers to establish the theory had attracted little attention. The work of Gaddum and Bliss was confined to a class of problems in biological assay which Finney in his definitive work on the subject[67] later described as "the measurement of the potency of any stimulus, physical, chemical or biological, physiological or psychological, by means of the reactions which it produces in living matter". Recently it has been recognized that the technique is applicable to a much wider range of problems: Pearson and Hartley[157] give a worked example in the study of the fall-off with distance in the blast effect of explosive charges; and the present authors[51] give an application to the study of demand for consumers' goods: this is developed in greater detail in Chapter 12.

There is already an extensive literature on this topic and we need here refer only to the recently published books of Bliss[23] and Finney[67] which provide all that could be required of a textbook of theory and method with ample illustration of their application and copious references to journal articles. Our purpose in introducing this chapter is threefold: to present briefly the essentials of probit analysis as a development of lognormal theory and to discuss the estimation problems as a continuation of the previous two chapters; to show how these estimation problems may be efficiently handled on high-speed automatic computers and how this facility enables us to investigate certain convergence problems which are difficult to resolve by analysis; and to provide the statistical foundation for the extension of probit analysis to economic problems.

7.2. QUANTAL AND QUANTITATIVE RESPONSES TO STIMULI

Because of the increasing variety of the applications of probit analysis we propose to develop the theory with as little reference as possible to any particular context. In any application there are three essential components—the stimulus, the subject and the response. For example, in biology an insect is subjected to a stimulus, say a given concentration of
LOGNORMAL DISTRIBUTION AND PROBIT ANALYSIS

a drug, and its response takes the form of death; if it survives it is said not to respond. Or sheets of cardboard are placed at given distances from a point at which a charge is detonated and a record is made of those which are perforated. Or, again, we may observe how households vary their purchases of a commodity when the price increases (a negative stimulus or repellent) or when their incomes increase (a positive stimulus). In the first two examples the characteristic response is termed quantal, that is, it is of the type 'all or nothing'; while in the last the response is quantitative, the purchases of households taking any values in a certain range. In either case it is usually found that the response is different for different subjects, even though the conditions of the experiment may be carefully controlled.

Where the response is quantal there will be for any one subject a certain critical level of intensity of the stimulus below which the subject does not respond and above which it does; this critical level we term the tolerance, following Finney. Our present interest centres on those cases in which the distribution of tolerances over the population of subjects follows the lognormal law. Such cases arise when the tolerance value may be thought of as being generated by some such process as is described in Chapter 3, and are found to be very numerous in practice when the subjects are living organisms. Thus, if in any experiment the intensity of the stimulus is represented by the variable \( t \), the proportion \( dP \) of subjects with tolerances in the interval \( (t, t + dt) \) is given by

\[
dP = d\Lambda(t | \mu, \sigma^2); \tag{7.1}
\]

and if a stimulus of intensity \( t \) is given to the whole population the proportion \( P \) responding is

\[
P = \Lambda(t | \mu, \sigma^2). \tag{7.2}
\]

In the usual form of experiment different groups of subjects are exposed to a number of preselected levels of intensity and the proportion in each group responding is determined. From these data estimates of \( \mu \) and \( \sigma^2 \) may be derived.

For quantitative responses the curve of the response \( q \) plotted against the stimulus \( t \) again follows the sigmoid form of the lognormal distribution function, asymptotically approaching a finite magnitude known as the saturation level of response. This saturation level, \( \kappa \) say, enters the equation for the response curve as a scalar factor so that the response may be measured as a proportion \( q/\kappa \) of the maximum response. The general form of the equation is

\[
q = \kappa \Lambda(t | \mu, \sigma^2), \tag{7.3}
\]

in which \( \kappa, \mu \) and \( \sigma^2 \) are parameters usually to be estimated. As it stands equation (7.3) does not allow for variation in individual subjects; it is necessary to introduce some stochastic element and this may be done in either of two ways: with an additive error term

\[
q = \kappa \Lambda(t | \mu, \sigma^2) + \epsilon, \tag{7.4}
\]
or with a multiplicative error factor

\[ q = \kappa \Lambda(t \mid \mu, \sigma^2) \epsilon, \quad (7.5) \]

where in either case \( \epsilon \) is assumed a normal variate with zero mean. Occasionally \( \kappa \) is known a priori and the estimation problem is eased.

### 7.3. Notation

In the remainder of this chapter we shall find it necessary to depart slightly from the notation we have so far adopted; this is necessary since the notation of probit analysis is now so well established that to alter it would cause confusion. We shall write

\[ P(Y) = N(Y \mid 0, 1) \]

\[ = \int_{-\infty}^{Y} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du, \quad (7.6) \]

and

\[ Z(Y) = \frac{dP(Y)}{dY} \]

\[ = \frac{1}{\sqrt{2\pi}} e^{-y^2}. \quad (7.7) \]

In equation (7.6) the \( Y \) corresponding to a given \( P \) is termed, following Gaddum, the normal equivalent deviate of \( P \); it is, indeed, the \( N(0, 1) \) quantile of order \( P \). Bliss introduced the word *probit* to denote \( Y + 5 \), claiming that the addition of the 5 eased the computations by avoiding the introduction of negative values. This device we consider artificial and of doubtful computational convenience; we shall accordingly work throughout with the theoretically more desirable normal equivalent deviates.†

If \( \alpha \) and \( \beta \) are defined by

\[ \alpha = -\frac{\mu}{\sigma} \quad (7.8) \]

and

\[ \beta = \frac{1}{\sigma}, \quad (7.9) \]

then

\[ \Lambda(t \mid \mu, \sigma^2) = P(\alpha + \beta x), \quad (7.10) \]

where \( x = \log t \); it is often more convenient to estimate \( \alpha \) and \( \beta \) rather than \( \mu \) and \( \sigma^2 \). It is now seen that the transformation to normal equivalent deviates in this case linearizes the relationship, for the normal equivalent deviate of \( P \) is

\[ P^{-1}P(\alpha + \beta x) = \alpha + \beta x. \quad (7.11) \]

### 7.4. The Estimation of Quantal Response

The estimation problem usually encountered involves the determination of estimates of \( \mu \) and \( \sigma^2 \) (or of \( \alpha \) and \( \beta \)) from information on \( g \) groups of subjects. Of the \( n_i \) subjects in the \( i \)th group, exposed to a stimulus of

† A table of \( P(Y) \) and \( Z(Y) \) is given in Appendix Table A5.
intensity $t_i$, suppose that $r_i$ respond, and $n_i - r_i$ do not respond, to the stimulus. It is assumed that the situation is purely binomial, that is to say, each subject responds with a probability $P$ independently of the behaviour of all other subjects whether in the same group or not; a statistical test of this hypothesis is desirable and is discussed later in this section.

Before proceeding to a more formal analysis we may usefully compare the problem with the quantile method of estimation described in Chapter 5. There the orders of the quantiles were fixed and the sample quantiles of these particular orders were used for estimation purposes; here, various $t_i$ are selected, and the idea behind the experiment is to discover the order of these $t_i$ when they are regarded as quantiles of the population of tolerances. These orders are, of course, estimated by the $p_i = r_i/n_i$, the proportions responding in each group; and it is worth noting that the orders are independently assessed. It is noticed at once that the data, namely, the $g$ pairs $(p_i, t_i)$, are in exactly the form required for the use of logarithmic probability paper; on the hypothesis of lognormality the points should lie roughly on a straight line and estimates of $\mu$ and $\sigma$ may be obtained by the method of §4.5. These estimates, besides providing quick, approximate results, are valuable in the procedure now to be outlined.

The method of estimation applied is that of maximum likelihood, and it is found more convenient to estimate $\alpha$ and $\beta$ rather than $\mu$ and $\sigma^2$. On the binomial assumption the likelihood of the given sample is

$$L = \prod_{i=1}^g \left( \frac{n_i}{r_i} \right) \left( \frac{P(\alpha + \beta x_i)}{Q(\alpha + \beta x_i)} \right)^{r_i} \left( \frac{Q(\alpha + \beta x_i)}{P(\alpha + \beta x_i)} \right)^{n_i - r_i},$$

(7.12)

where $Q = 1 - P$, and so the loglikelihood function $L$ as far as it involves $\alpha$ and $\beta$ is

$$L = \sum_i \left( r_i \log P + (n_i - r_i) \log Q \right).$$

(7.13)

The maximum-likelihood equations are then derived as

$$\frac{\partial L}{\partial \alpha} = \sum_i n_i x_i (p_i - P),$$

(7.14)

$$\frac{\partial L}{\partial \beta} = \sum_i n_i (p_i - P),$$

(7.15)

and the information matrix $I$ is given by

$$I = \begin{bmatrix}
-\Sigma_{xx} & -\Sigma_{wx} \\
-\Sigma_{wx} & \Sigma_{xx} - \Sigma_{xx}^2
\end{bmatrix},$$

(7.16)
where \( w = Z^2 / PQ \) is termed the weighting factor. Equations (7.14) and (7.15) are rather intractable as they stand; to solve them the method usually adopted is that of scoring\(^\dagger\) using the iterative equations

\[
\begin{bmatrix}
- E \left( \frac{\partial^2 L}{\partial \alpha^2} \right) & - E \left( \frac{\partial^2 L}{\partial \alpha \partial \beta} \right) \\
- E \left( \frac{\partial^2 L}{\partial \beta \partial \alpha} \right) & - E \left( \frac{\partial^2 L}{\partial \beta^2} \right)
\end{bmatrix}
\begin{bmatrix}
a_j - a_{j-1} \\
b_j - b_{j-1}
\end{bmatrix}
= 
\begin{bmatrix}
\frac{\partial L}{\partial \alpha} \\
\frac{\partial L}{\partial \beta}
\end{bmatrix},
\quad (7.17)
\]

which in this case reduce to

\[
I \begin{bmatrix}
a_j - a_{j-1} \\
b_j - b_{j-1}
\end{bmatrix} = 
\begin{bmatrix}
\Sigma_n w \frac{b - P}{Z} \\
\Sigma_n x w \frac{b - P}{Z}
\end{bmatrix},
\quad (7.18)
\]

where \( Z \) and \( P \) (and hence \( w \)) are calculated with \( \alpha = a_{j-1} \) and \( \beta = b_{j-1} \). To start the iterative process initial guesses \( a_0 \) and \( b_0 \) at the maximum-likelihood estimates \( a \) and \( b \) are required, and these may most conveniently be derived from the graphical analysis with the relations (7.8) and (7.9). When convergence is reached the variance matrix of the estimators \( a \) and \( b \) is given by \( I^{-1} \).

For desk computing the calculations are usually simplified by the introduction of working probits \( y \) given by

\[
y = Y + \frac{P - P}{Z} = \left( Y - \frac{P}{Z} \right) + \frac{1}{Z},
\quad (7.19)
\]

where \( Y = \alpha + \beta x \). Then equations (7.18) reduce to

\[
b_j = \frac{\Sigma_n w (x - \bar{x}) (y - \bar{y})}{\Sigma_n w (x - \bar{x})^2},
\quad (7.20)
\]

and

\[
a_j = \bar{y} - b \bar{x},
\quad (7.21)
\]

where \( \bar{x} = \Sigma_n x \Sigma_n w \) and \( \bar{y} = \Sigma_n y w \Sigma_n w \); the computations are then formally equivalent to a repeated weighted regression analysis and are facilitated by the use of tables of \( Y - P / Z, 1 / Z \) and \( w \). The values of these functions corresponding to \( Y = -4.0 \) (0.1) 4.0 are given in Appendix Table A6. The reader who is interested in the computational details is referred to the two books previously cited [20, 67].

The test of the hypothesis of pure binomiality is based on the statistic

\[
\Sigma_n \frac{(P - P)^2}{PQ} = \Sigma_n w (y - \bar{y})^2 - b^2 \Sigma_n w (x - \bar{x})^2,
\quad (7.22)
\]

which is distributed as \( \chi^2_{2n-3} \); any significant largeness of this statistic indicates some kind of heterogeneity.

\(^\dagger\) See, for example, Rao[171], where an excellent account of this method is given.
For automatic computing the authors have found it best to make use of equations (7.18); these allow easier control of scaling factors in the machine and, with a rapid and accurate method of inverting a matrix of order two, the introduction of working values is an unnecessary complication. The major difficulty to be overcome is to find an efficient way of calculating $Z(Y)$ and $P(Y)$ for any given $Y$, say in the practical range $| Y | < 4$. $Z(Y)$ is most easily obtained by using a power series for computing the exponential function. The calculation of $P(Y)$ is more troublesome; an approximate quadrature is too time-consuming; and this criticism also applies to the expansion in terms of a Gauss hypergeometric continued fraction, as suggested by Tocher [188], since it involves a number of divisions, and on most contemporary high-speed equipment it is advisable to reduce division operations to a minimum.

There is, however, a convenient approximation involving $Z(Y)$, namely,

$$P(Y) = 1 - (c_1 y + c_2 y^2 + c_3 y^3 + c_4 y^4 + c_5 y^5) Z(Y) \quad (Y \geq 0), \quad (7.23)$$

where

$$y = \frac{1}{1 + dY} \quad (7.24)$$

and

$$c_1 = 0.319381530, \quad c_4 = -1.821255978,$$

$$c_2 = -0.356563782, \quad c_5 = 1.330274429,$$

$$c_3 = 1.781477937, \quad d = 0.2316419.$$  

The automatic computer calculates $Z(Y)$ and $P(Y)$ accurately to at least six decimal places in somewhat under half a second. The iterations are carried out by inversion of the matrix $I$ at each stage and the process is automatically stopped when a preassigned order of convergence is reached; the inverse $I^{-1}$ is immediately available as the variance matrix of the estimators.

All these calculations may be carried out to a satisfactory order of convergence in under a minute for a standard problem comprising say ten observations. As an example of the application we give the results for the now classical example of the toxicity of Rotenone to Macrosiphoniella sanborni for which the relevant data are given in Table 7.1; the shape of the response curve is shown in Fig. 7.1. When plotted on logarithmic probability paper the points give the array shown in Fig. 7.2; from the fitted line $\mu$ and $\sigma$ are estimated at 1.57 and 0.55 respectively, so that we may take $a_0 = -2.80$ and $b_0 = 1.82$.

† For further details of the automatic computer actually used, the EDSAC of the Mathematical Laboratory of the University of Cambridge, and for a description of the programmes constructed see Chapter 13.

‡ The approximation used is an adaptation of that given on Sheet 45 of Approximations in Numerical Analysis, Form (15), The RAND Corporation (Hastings [100]).

§ The data are taken from Martin [145], Table 9, and are also to be found in Finney [67], Table 2.
The results of successive iterations are given in Table 7.2 together with results for some other initial values, chosen at some distance from the final values. We shall return to this point in §7.6.

The variance matrix is

\[
\begin{bmatrix}
0.1154 & -0.0676 \\
-0.0676 & 0.0428
\end{bmatrix},
\]

from which the standard errors of any derivative estimators may be obtained. The observed value of \(\chi^2\) is 1.734, and so there is no significant departure from the homogeneity hypothesis. The fitted response curve is that shown in Fig. 7.1.

### Table 7.1. Toxicity of Rotenone to 'Macrosiphoniella sanborni'

<table>
<thead>
<tr>
<th>No. of insects in group</th>
<th>Concentration (mg/l.)</th>
<th>Proportion responding</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>10.2</td>
<td>0.68</td>
</tr>
<tr>
<td>49</td>
<td>7.7</td>
<td>0.66</td>
</tr>
<tr>
<td>48</td>
<td>5.1</td>
<td>0.52</td>
</tr>
<tr>
<td>48</td>
<td>3.8</td>
<td>0.33</td>
</tr>
<tr>
<td>50</td>
<td>2.6</td>
<td>0.12</td>
</tr>
</tbody>
</table>

### Table 7.2. Iterations for Rotenone Data, Starting from Different Initial Values

<table>
<thead>
<tr>
<th>No. of iteration, (j)</th>
<th>(a_j)</th>
<th>(b_j)</th>
<th>(a_j)</th>
<th>(b_j)</th>
<th>(a_j)</th>
<th>(b_j)</th>
<th>(a_j)</th>
<th>(b_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial values</td>
<td>-2.80</td>
<td>1.82</td>
<td>2.70</td>
<td>-1.74</td>
<td>0.40</td>
<td>-0.72</td>
<td>0.60</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>-2.7892</td>
<td>1.8187</td>
<td>-5.55</td>
<td>3.62</td>
<td>-3.85</td>
<td>2.64</td>
<td>-2.29</td>
<td>1.49</td>
</tr>
<tr>
<td>2</td>
<td>-2.7892</td>
<td>1.8187</td>
<td>-4.29</td>
<td>2.34</td>
<td>-2.51</td>
<td>1.63</td>
<td>-2.74</td>
<td>1.79</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>-2.70</td>
<td>1.38</td>
<td>-2.79</td>
<td>1.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>-2.76</td>
<td>1.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>-2.79</td>
<td>1.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 7.5. A Comparison with the Logistic Curve

There has been a heated controversy\([14,15,16,17,18,72]\) in the past few years over the relative merits of the method just described and that which uses the logistic function

\[
P(Y) = \frac{1}{1 + e^{-x}}. \tag{7.25}
\]

The method of maximum likelihood applied to estimating parameters of this response curve also leads to an iterative procedure using logits. Berkson\([16]\), however, has developed a non-iterative procedure based on minimizing what he terms the logit \(\chi^2\), which leads to estimators with the same large-sample properties as maximum-likelihood estimators. His claims for small-sample advantages are made perhaps too strongly and
Fig. 7.1. Toxicity of Rotenone to Macrosiphoniella sanborni.

Fig. 7.2. Logarithmic probability graph for Rotenone data.
on too little evidence; nevertheless, the computational advantages are
great at least when desk calculators are used. Berkson's other main
criticism of the probit method is that the iterations are often stopped
before the process has converged sufficiently, a fault which gives rise to
a lack of standardization. With automatic computing machinery there
is now no disadvantage in allowing the iterations to continue until a
sufficient order of convergence is reached; so that Berkson's criticisms
may be substantially rejected by those fortunate enough to possess, or
have access to, an automatic computer.

Our own preference for the probit method rests, however, on more
positive considerations. The logistic lacks a well-recognized and manage-
able frequency distribution of tolerances which the probit curve does
possess in a natural way. Moreover, it is often necessary to average the
curve over some other characteristic of the population; it is then found
that the probit curve is much more tractable than the logistic.

<table>
<thead>
<tr>
<th>Order of quantile (%)</th>
<th>Probit quantile (mg/L)</th>
<th>Logistic quantile (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.016</td>
<td>1.016</td>
</tr>
<tr>
<td>10</td>
<td>1.294</td>
<td>1.294</td>
</tr>
<tr>
<td>15</td>
<td>1.596</td>
<td>1.596</td>
</tr>
<tr>
<td>20</td>
<td>1.919</td>
<td>1.919</td>
</tr>
<tr>
<td>25</td>
<td>2.268</td>
<td>2.268</td>
</tr>
<tr>
<td>30</td>
<td>2.635</td>
<td>2.635</td>
</tr>
<tr>
<td>35</td>
<td>2.993</td>
<td>2.993</td>
</tr>
<tr>
<td>40</td>
<td>3.333</td>
<td>3.333</td>
</tr>
<tr>
<td>45</td>
<td>3.684</td>
<td>3.684</td>
</tr>
<tr>
<td>50</td>
<td>4.052</td>
<td>4.052</td>
</tr>
<tr>
<td>55</td>
<td>4.426</td>
<td>4.426</td>
</tr>
<tr>
<td>60</td>
<td>4.805</td>
<td>4.805</td>
</tr>
<tr>
<td>65</td>
<td>5.202</td>
<td>5.202</td>
</tr>
<tr>
<td>70</td>
<td>5.613</td>
<td>5.613</td>
</tr>
<tr>
<td>75</td>
<td>6.035</td>
<td>6.035</td>
</tr>
<tr>
<td>80</td>
<td>6.467</td>
<td>6.467</td>
</tr>
<tr>
<td>85</td>
<td>6.908</td>
<td>6.908</td>
</tr>
<tr>
<td>90</td>
<td>7.368</td>
<td>7.368</td>
</tr>
<tr>
<td>95</td>
<td>7.849</td>
<td>7.849</td>
</tr>
</tbody>
</table>

There is little doubt that for an ordinary analysis the numerical
results obtained are not very different. To illustrate this we give in
Table 7.3 a comparison of the quantiles for the probit and logistic curves
both fitted by maximum likelihood to the data of the Rotenone experi-
ment (Table 7.1). The two fitted response curves are so close that it is
not practical to depict them on the same diagram.

There must now be a sufficient accumulation of data from a number of
fields to make it possible to decide which of the two response curves is
the better hypothesis. One method of comparison would be to estimate

† See, for example, equation (12.40) of Chapter 12 or §44 of Finney [67].
both by the method of maximum likelihood and discriminate by the \( \chi^2 \) measures of goodness of fit. This would be an exceedingly quick process on the automatic computer especially if, as is suggested in the following section, there is no need to obtain initial values \( a_0 \) and \( b_0 \) by graphical methods.

7.6. Convergence Problems in Quantal Response Computations

It is a source of worry to many practical workers whether the initial values they choose will lead to the maximum-likelihood solutions. The reader will have noticed that the pairs of values \( a_0 \) and \( b_0 \) in Table 7.2, while initiating iterative processes which converge to \( a \) and \( b \), are yet widely different and, in all cases except the first, at some distance from \( a \) and \( b \). It seems of interest then to pose the question: for given data what set of values of \( a_0 \) and \( b_0 \) gives rise to iterations converging to \( a \) and \( b \)? We shall term this region in the \( (a_0, b_0) \) plane the region of convergence for the particular set of data. The analytical approach to the problem is not very helpful; certainly the process will converge if \( (a_0, b_0) \) is sufficiently close to \( (a, b) \); but theory gives little clue as to what is meant by 'sufficiently close'. Some kind of empirical approach is therefore necessary.

Finney [66] has reported the results of an experiment, part of which throws some light on this problem. Twenty-one scientists having no experience in the probit method were asked to fit straight lines to two sets of data, and the first iterates from the initial values obtained from these lines were calculated. The consequences of starting with certain extreme values were also investigated. Finney's conclusions from this part of his experiment were: that except in irregular cases 'a single cycle of iteration initiated by any reasonable trial regression line will give a satisfactory approximation'; and that it seems preferable to underestimate \( b_0 \) than to overestimate it. In the remainder of this section we give an account of a fuller empirical investigation of four samples, made possible by the automatic computing programme.

The four samples were:

(i) the Rotenone-Macrosiphoniella sanborni experiment of Table 7.1,
(ii) the example given in Pearson and Hartley [157] of the blast effect of an explosive charge, and
(iii) and (iv), the insulin assays by the mouse convulsion method used by Finney in his experiment.

The data for these examples are brought together in Table 7.4. We decided to carry out a thorough study of example (i) and to confirm the results in this case by shorter investigations of the other three samples.

It was not practical in empirical work to carry out all the tests that theory would require to ensure that the process was really converging,

† We wish to thank Dr D. J. Finney for suggesting to us the problems of this section.
‡ An analysis of the situation will be found in §3.42 of Householder [108].
or diverging; some workable definitions of convergence and divergence were therefore required. Preliminary work on the four samples showed that \( a_0 \) and \( b_0 \) in the region

\[
\begin{align*}
|a_0 - a| < 2^{-7}, \\
|b_0 - b| < 2^{-7}.
\end{align*}
\]

(7.26)
gave convergence and it was thus safe to say that the process converged if

\[
\begin{align*}
|a_j - a_{j-1}| < 2^{-7}, \\
|b_j - b_{j-1}| < 2^{-7}.
\end{align*}
\]

(7.27)
for some \( j \). Again it was found satisfactory to regard the process as diverging if at some stage \( Y_j = a_j + b_j x \) was in the range \(|Y| > 4\) for some sample value \( x \) (\( P(Y) \) would then be \(<0.0000317\) or \(>0.9999683\)). The computing programme described in §7.4 was very simply modified so that the boundary of the convergence region was automatically traced out by a trial and error method. The successive iterates were printed out for each initial point so that information was given on the speeds and paths of convergence. For sample (i) a number of extra initial points were tried where the programme left some doubt about the exact position of the boundary.

Fig. 7.3 shows the region of convergence for sample (i). The bounding parallelogram is given by \( a_0 + b_0 \) max \( x = \pm 4, a_0 + b_0 \) min \( x = \pm 4 \), so that it contains all points \((a_0, b_0)\) such that \( |a_0 + b_0 x| \leq 4 \) for all \( x \). Initial values lying within this parallelogram but outside the outer elliptical boundary were found to initiate an oscillating and apparently diverging sequence of estimates \((a_j, b_j)\), eventually leading to values outside the parallelogram. A slight exception to this was provided by sample (iv) whose convergence region extends a little outside the bounding parallelogram (cf. Fig. 7.4). The inner elliptical regions break up the whole region of convergence by speed of convergence. Thus ‘4+’ between the outer and the next inner contour indicates that four or more iterations are necessary to reach the degree of convergence of (7.27), and similarly for the other contours. The results for samples (ii)-(iv) are similar; to illustrate this we give in Fig. 7.4 a comparison for the four samples: the
bounding parallelogram has been standardized to a unit square in each case by the transformation

\[
\begin{align*}
A_0 &= \frac{1}{b_0}(a_0 + b_0 \max x), \\
B_0 &= \frac{1}{b_0}(a_0 + b_0 \min x).
\end{align*}
\]

(7.28)

Fig. 7.3. Region of convergence for Rotenone data.

One surprising result is that the region of convergence is so large; many of the initial values in this region are so remote from the final values that no practical worker would dream of using them. Nevertheless, it is comforting to know that there is little need to be within a small region; though there is no doubt that, when using desk computing machines, it pays not to be too far off with the initial guesses. Support is also given to Finney's conjecture that it is safer and quicker to under-
estimate \( b \) than to overestimate it. The other interesting point is that the origin \((0, 0)\) is well within the region of convergence in each of the four cases. If this last result holds in general, or at least in a large proportion of cases, it is of great use when the calculations are carried out on an automatic machine; for it is then less trouble to take the origin as the starting point of the process (at the cost of perhaps one or two extra iterations) than to make an initial graphical estimate.

Fig. 7.4. Standardized convergence regions for the four samples of Table 7.4. The points \( \circ \) indicate the converged values.

7.7. The Estimation of Quantitative Response: the Homoscedastic Case

We deal first with the case of quantitative response for which the model is represented by equation (7.4), namely,

\[
q = k \Lambda(t; \mu, \sigma^2) + \epsilon
\]

\[
= kP(\alpha + \beta t) + \epsilon,
\]

(7.29)

where \( x = \log t, \alpha \), and \( \beta \) are related to \( \mu \) and \( \sigma \) as before (see equations (7.8) and (7.9)), and \( \epsilon \) is a \( N(0, \sigma^2) \) variate; \( \sigma^2 \) is supposed constant,
independent of $x$. Usually the information available for estimating $\kappa$, $\alpha$ and $\beta$ relates to a number $g$ of groups of subjects with all the $n_i$ subjects in the $i$th group exposed to a stimulus of intensity $t_i$; let $q_{ij}$ ($j = 1, \ldots, n_i$) denote the responses of the individual subjects of the $i$th group, $q_i$ their mean, and $q$ the mean response of the whole sample. A preliminary analysis of variance as in Table 7.5 gives an estimate $s_i^2$ of $\sigma_i^2$; $n = \Sigma n_i$ is the total sample size. From equation (7.29)

$$D^2[q \mid x] = \sigma_o^2$$

so that the variance of $q$ is constant. This hypothesis may be tested by applying the Bartlett test [11] of homogeneity to the individual group variances $s_i^2$ where,

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (q_{ij} - q_i)^2.$$  

**Table 7.5. Analysis of Variance of Quantitative Responses**

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of freedom</th>
<th>Sums of squares</th>
<th>Mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>$g - 1$</td>
<td>$\sum_{i=1}^{g} n_i(q_i - q)^2$</td>
<td>$s_o^2$</td>
</tr>
<tr>
<td>Within groups</td>
<td>$n - g$</td>
<td>$\sum_{i=1}^{g} \sum_{j=1}^{n_i} (q_{ij} - q_i)^2$</td>
<td>$s_i^2$</td>
</tr>
<tr>
<td>Total about mean</td>
<td>$n - 1$</td>
<td>$\sum_{i=1}^{g} \sum_{j=1}^{n_i} (q_{ij} - q)^2$</td>
<td></td>
</tr>
</tbody>
</table>

Graphical estimates of $\kappa$, $\mu$ and $\sigma$ may be obtained though with less reliance than for the quantal case. From a study of the response curve a guess $k_0$ is made at the value of $\kappa$, and then $f_i = q_i/k_0$ and $t_i$ are plotted on logarithmic probability paper. A tendency to systematic curvature in the array of points indicates a bad choice of $k_0$ and should be corrected. Estimates of $\mu$ and $\sigma$ (and hence of $\alpha$ and $\beta$) are then obtained as in the quantal case.

The method of maximum likelihood requires the maximization of the loglikelihood function $L$, where

$$L = -\frac{n}{2} \log \sigma_o^2 - \frac{1}{2\sigma_o^2} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \{q_{ij} - \kappa P(x + \beta t_i)\}^2,$$

so that likelihood equations are

$$o = \frac{\partial L}{\partial \kappa} = \frac{\kappa}{\sigma_o^2} \sum_{i=1}^{g} n_i P_i \frac{(q_i - P_i)}{\kappa},$$

$$o = \frac{\partial L}{\partial \beta} = \frac{\kappa^2}{\sigma_o^2} \sum_{i=1}^{g} n_i Z_i x_i \frac{(q_i - P_i)}{\kappa},$$

$$o = \frac{\partial L}{\partial \alpha} = \frac{\kappa^2}{\sigma_o^2} \sum_{i=1}^{g} n_i Z_i \frac{(q_i - P_i)}{\kappa}.$$  

(7.32)
where \( P_i = P(\alpha + \beta x_i) \) and \( Z_i = Z(\alpha + \beta x_i) \); in what follows we shall omit the suffix \( i \) for conciseness; all summations are then understood to be over groups. The information matrix \( I \) is given by

\[
I = \frac{1}{\sigma^2} \begin{bmatrix}
\sum_n P^2 & \kappa \sum_n PZx & \kappa \sum_n PZ \\
\kappa \sum_n PZx & \kappa^2 \sum_n Z^2 x \& \kappa^2 \sum_n Z^2 \\
\kappa \sum_n PZ & \kappa^2 \sum_n Z^2 x & \kappa^2 \sum_n Z^2
\end{bmatrix}
\]

so that the method of scoring leads to iterative equations which may be written in the form:

\[
\begin{bmatrix}
\sum_n P^2 & \sum_n PZx & \sum_n PZ \\
\sum_n PZx & \sum_n Z^2 x \& \sum_n Z^2 \\
\sum_n PZ & \sum_n Z^2 x & \sum_n Z^2
\end{bmatrix}
\begin{bmatrix}
k_{j-1} - k_{j-1} \\
b_{j-1} - b_{j-1} \\
\end{bmatrix}
= \begin{bmatrix}
\sum_n P \left( \frac{q}{k_{j-1}} - 1 \right) \\
\sum_n Zx \left( \frac{q}{k_{j-1}} - 1 \right) \\
\sum_n Z \left( \frac{q}{k_{j-1}} - 1 \right)
\end{bmatrix}
\]

(7.35)

The coefficients and right-hand vector of these equations are all evaluated at \( k_{j-1}, a_{j-1} \) and \( b_{j-1} \). For automatic computing the equations are best solved as they stand by the use of a quick and accurate method of inverting a third-order matrix. When a satisfactory order of convergence is reached the final inverse is easily adjusted by pre- and post-multiplying by the matrix

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & \frac{1}{k} & 0 \\
0 & 0 & \frac{1}{k}
\end{bmatrix}
\]

and by multiplying by \( \sigma^2 \) to give an estimate of \( I^{-1} \), the variance matrix of the estimators \( k, b \) and \( a \).

Some simplification similar to that for quantal estimation may be made for desk computing by the introduction of weighting factors \( w = Z^2 \), an auxiliary variable \( x' = P/Z \), and working probits \( y \) given by

\[
y = Y + \frac{q}{k} \frac{P}{Z} \\
= (Y - \frac{P}{Z}) + \frac{q}{k} \frac{1}{Z} Z
\]

(7.36)

If

\[
a'_j = a_j + b_j x - \frac{k_{j-1}}{k_j} k_{j-1} x',
\]
where $\bar{x} = \Sigma nux/\Sigma nw$ and $\bar{x}' = \Sigma nux'/\Sigma nw$, then equations (7.35) may be rewritten as

$$\begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \begin{bmatrix} k_j - k_{j-1} \\ b_j \end{bmatrix} = \begin{bmatrix} S_{xy} \\ S_{yy} \end{bmatrix},$$  

(7.37)

and

$$a_j' = \bar{y},$$

where the symbols $S$ denote weighted sums of squares and cross-products about means; thus

$$S_{xx} = \Sigma nw(x' - \bar{x}) (x - \bar{x}).$$  

(7.38)

The variance matrix of $k$ and $b$ is estimated by pre- and post-multiplying

$$S^* = S_{xx}^{-1}.$$

The variance of $a'$ is $1/\Sigma nw$ and $a'$ is uncorrelated with $k$ and $b$ so that the variance of $a$ and its covariances with $k$ and $b$ may be readily found. There is no advantage in using this method for automatic computing, since it involves more troublesome scaling problems and there is little need to reduce the inversion problem from third to second order. Tables of $Y - P/Z, 1/Z, w$ and $x'$ are provided in Appendix Table A7; the reader is referred to Finney[67] for details of the computational layout. No matter which method is employed, initial guesses $k_0, a_0$ and $b_0$ must be made and the graphical method is recommended for this purpose.

After convergence the residual mean square may be calculated by one of the alternative forms

$$\frac{1}{g-3} \Sigma n(q - kP(a + bx))^2 = \frac{k^2}{g-3} (S_{yy} - bS_{yx}).$$  

(7.39)

This statistic should be distributed as $\sigma^2_x X^2_{g-3}$, and so a test of goodness of fit is given by comparing, by a variance ratio test, this statistic with the statistic $\chi^2$ obtained from the analysis of variance (Table 7.5). If the group means only are available the analysis of variance is impossible, and the above test cannot be carried out; but the statistic (7.39) may be taken as an estimator of $\sigma^2_x$.

7.8. THE ESTIMATION OF QUANTITATIVE RESPONSE:

A SPECIAL CASE

In Chapter 12 we shall be interested in a special case of model (7.29), that for which $\beta$ is set at unity. For this model

$$q = kP(a + x) + e,$$  

(7.40)
and the estimation procedure is correspondingly simpler. The iterative
equations become
\[
\begin{bmatrix}
\Sigma nP^2 & \Sigma nPZ \\
\Sigma nPZ & \Sigma nZ^2
\end{bmatrix}
\begin{bmatrix}
k_j - k_{j-1} \\
1
\end{bmatrix}
=
\begin{bmatrix}
\Sigma nP \left( \frac{q}{k_j - k_{j-1}} - P \right) \\
\Sigma nZ \left( \frac{q}{k_j - k_{j-1}} - P \right)
\end{bmatrix},
\]
and the variance matrix of \( k \) and \( \sigma \) is estimated by pre- and post-multi-
plying
\[
\begin{bmatrix}
\Sigma nP^2 & \Sigma nPZ \\
\Sigma nPZ & \Sigma nZ^2
\end{bmatrix}^{-1}
\begin{bmatrix}
\Sigma nPZ & \Sigma nZ^2
\end{bmatrix}
\]
by
\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}.
\]
Again the equations may be simplified for desk computing by the intro-
duction of weighting factors, working probits and an auxiliary variable.
The procedure is set out by Aitchison and Brown[5].

7.9. The Estimation of Quantitative Response: The Heteroscedastic Case

The heteroscedastic case which we shall consider is that of equation
(7.5), namely,
\[
q = \kappa L \{ t | \mu, \sigma^2 \} \epsilon
= \kappa P(\alpha + \beta x) \epsilon,
\]
de \( \epsilon \) is a \( N(0, \sigma^2) \) variate, \( \sigma^2 \) being a constant. This model has been
discussed by the authors in a previous paper[7]. Here
\[
E[q | x] = \kappa P(\alpha + \beta x) \exp \{ \mu \sigma^2 \}
\]
and
\[
D^2[q | x] = \kappa^2 (P(\alpha + \beta x))^2 \exp \{ 2\sigma^2 - \sigma^2 \}
= (\exp \{ \sigma^2 \} - 1) \left[ E[q | x] \right]^2.
\]
The variance of \( q \) now depends on \( x \) and is, indeed, proportional to the
square of the mean of \( q \); this may be expressed by saying that the
coefficient of variation \( \eta(q | x) \) of \( q \) is independent of \( x \), for
\[
\eta(q | x) = \frac{D[q | x]}{E[q | x]}.
\]
The model may be analysed by rewriting it in the form
\[
\begin{align*}
\log q &= \log \kappa + \log P(\alpha + \beta x) + \epsilon \\
\text{or} \quad q' &= \kappa' + \log P(\alpha + \beta x) + \epsilon,
\end{align*}
\]
where \( q' = \log q \) and \( \kappa' = \log \kappa \). A preliminary analysis of variance of \( q' \)
between and within groups gives an estimate \( \sigma^2 \) of \( \sigma^2 \) as in the homo-
scedastic case. Graphical estimates may again be obtained by guessing.
LOGNORMAL DISTRIBUTION AND PROBIT ANALYSIS

\[ k_0 \text{ and plotting } \frac{1}{k_0} \exp \{ q'_i \}, \text{ where} \]

\[ q'_i = \frac{1}{n_i} \sum \log q_{ij}, \quad (7.46) \]

against \( t_i \) on logarithmic probability paper. This process will furnish initial estimates say \( k'_0, a_0 \) and \( b_0 \) of \( k' \), \( a \) and \( b \) with which to start the iterative solution of the maximum likelihood equations; there is no need to reproduce here the detailed algebra which proceeds exactly as for the homoscedastic case. The information matrix \( I \) is given by

\[
I = \frac{1}{\sigma^2} \begin{bmatrix}
\Sigma_n & \Sigma_n \frac{Z}{P} x & \Sigma_n \frac{Z}{P} \\
\Sigma_n \frac{Z}{P} x & \Sigma_n \frac{Z^2}{P^2} x^2 & \Sigma_n \frac{Z^2}{P} x \\
\Sigma_n \frac{Z}{P} & \Sigma_n \frac{Z^2}{P^2} x & \Sigma_n \frac{Z^2}{P^2} \\
\end{bmatrix}
\]

and the iterative scheme is

\[
\begin{bmatrix}
\Sigma_n & \Sigma_n \frac{Z}{P} x & \Sigma_n \frac{Z}{P} \\
\Sigma_n \frac{Z}{P} x & \Sigma_n \frac{Z^2}{P^2} x^2 & \Sigma_n \frac{Z^2}{P^2} x \\
\Sigma_n \frac{Z}{P} & \Sigma_n \frac{Z^2}{P^2} x & \Sigma_n \frac{Z^2}{P^2} \\
\end{bmatrix} \begin{bmatrix}
k'_{j-1} - k'_{j-1} - \log P \\
b_j - b_{j-1} \\
a_j - a_{j-1} \end{bmatrix} = \begin{bmatrix}
\Sigma_n \frac{Z}{P} (q' - k'_{j-1} - \log P) \\
\Sigma_n \frac{Z^2}{P^2} x (q' - k'_{j-1} - \log P) \\
\Sigma_n \frac{Z^2}{P^2} (q' - k'_{j-1} - \log P) \end{bmatrix},
\]

(7.47)

Again working probits, weighting factors and an auxiliary variable \( x' \) are useful.

For this model

\[
w = \frac{Z^2}{P^2},
\]

(7.48)

\[
x' = \frac{P}{Z}
\]

(7.49)

and

\[
y = Y + \frac{q' - k' - \log P}{Z/P}
\]

(7.50)

and the iterative equations become

\[
\begin{bmatrix}
S_{xx} & S_{xy} \\
S_{yy} & b \\
\end{bmatrix} \begin{bmatrix}
k'_{j-1} - k'_{j-1} - \log P \\
b_j - b_{j-1} \\
\end{bmatrix} = \begin{bmatrix}
S_{xy} \\
S_{yy} \\
\end{bmatrix}
\]

(7.51)

and

\[
a_j = \bar{y} - b_j \bar{x} - (k'_{j-1} - k'_{j-1}) x'.
\]

(7.52)

After convergence the variance matrix of \( k' \) and \( b \) is estimated by

\[
x^2 \begin{bmatrix}
S_{xx} & S_{xy} \\
S_{yy} & b \\
\end{bmatrix}^{-1},
\]

98
and the variance of $\bar{y}$, which is uncorrelated with $k$ and $b$, is $1/\Sigma nm$. The residual mean square is

$$\frac{1}{g-3} \Sigma n(q' - k' - \log P(a + bx))^2 = \frac{1}{g-3} (S_{xy} - bS_{x}y),$$

providing a test for goodness of fit as in the homoscedastic case; the reader will find details of the calculations in Aitchison and Brown[7]. Tables of $Y - P(\log P)/Z$, $x'$ and $w$ are reproduced in Appendix Table A8.

7.10. EXTENSIONS OF THE THEORY

Probit theory as described in the preceding sections may be extended in several directions; to explore these in detail would take us beyond the scope of this monograph and we content ourselves with a few brief remarks.

In quantal theory there are three main developments:

(i) **Comparisons of the effectiveness of stimuli.** Stimuli, similar in nature, often give rise to response curves with equal values of $\beta$ (or $\sigma$). The main interest in their comparative behaviour then centres on the differences in the $a$ parameter. This problem may be analysed efficiently; see, for example Finney[67].

(ii) **Several stimuli.** It may be of interest to study the reactions of subjects to a number of stimuli applied at the same time. There may exist a certain degree of interaction between the stimuli and several models have been proposed to take account of this dependence. The efficient design of the experiment also becomes a more important problem. For an analysis of models of this type the reader is referred to Finney[67], and to Hewlett and Plackett[107].

(iii) **Natural response.** A subject may show a response which is unconnected with the stimulus applied; for example in an insecticide assay some of the insects exposed may die from natural causes. It is possible with information from a control group to which no stimulus is applied to make allowances for this complication; see Finney[67].

In the case of quantitative response the first two extensions arise but methods of analysis have not been explored in detail in the literature. An effect similar to the third extension is obtained if it is possible to apply an infinite stimulus (a more meaningful case is a zero repellent) in order to estimate more efficiently the saturation level. An example of the adjustment necessary for the homoscedastic case is given by Finney[67] and for the heteroscedastic case by the authors[7].

Apart from these conceptual extensions there are many problems of statistical estimation which we must leave undiscussed. For instance, we have made no attempt to compare different estimation procedures such as those of Garwood[79], Cornfield and Mantell[43] and the many approximate methods such as Karber's method[119]. We have also ignored the problem of grouping of the stimulus variable (inherent in the example cited); one grouping problem in quan tally analysis has been satisfactorily solved by Tocher[187], but a number are yet untreated.
CHAPTER 8
COMPARISONS OF LOGNORMAL POPULATIONS

ANTIPOIUS OF SYRACUSE. Transform me then, and to your power I’ll yield.
The Comedy of Errors

8.1. USE OF T- AND F-TESTS FOR \( \mu \) AND \( \sigma^2 \)

In Chapters 10, 11 and 12 we shall emphasize that a large part of the usefulness of lognormal theory lies in the fact that with its aid a numerous class of skew distributions in a number of fields may be brought within the domain of normal test statistics. In the simple case where there are two independent samples drawn from two-parameter lognormal distributions, a \( z \) or \( F \) statistic may be computed from the transformed values to test for equality of the population variances \( \sigma^2 \); and, if no significant difference is shown, this may be followed with a \( t \)-test for the equality of the transformed population means \( \mu \). If the variances cannot be assumed equal, the testing of the transformed means may be handled by the Fisher-Behrens[13, 71] test, though it must be remembered that this is based on fiducial inference; alternatively, if the sample sizes are the same there is the result of Welch[200], who concluded that in this case no serious error will be made on proceeding as though the variances are equal; whilst if the sample sizes are unequal there is the further test proposed by Welch[200, 201] for which tables have been prepared by Aspin[9]. This is all standard theory and discussed, for example, by Kendall[123] (vol. ii, pp. 96–115). It is perhaps worth mentioning here that any statistic, such as the Lorenz index defined by equation (11.6), which is a function of \( \mu \) or \( \sigma^2 \) alone, may be handled by these methods.

8.2. COMPARISON OF ESTIMATES FOR \( \alpha \) AND \( \beta^2 \)

Occasionally the statistician may be asked to test for significant differences between two sample means, and may find on further study that the samples can be considered lognormal. A rough-and-ready large-sample theory is still provided by the fact that on the null hypothesis the statistic \( t = \bar{x}_1 - \bar{x}_2 \) is asymptotically distributed as \( N(0, \beta_1^2/n_1 + \beta_2^2/n_2) \), where the suffixes refer to the two samples. The parameters \( \beta_1^2 \) and \( \beta_2^2 \) may then be replaced by their estimates \( b_1^2 \) and \( b_2^2 \) for a large-sample test. On the other hand, the null hypothesis is that \( \alpha_1 = \alpha_2 \), and this is equivalent to the hypothesis that \( \mu_1 + \frac{1}{2}\sigma_1^2 = \mu_2 + \frac{1}{2}\sigma_2^2 \); and this equivalence shows that the means \( \alpha_1 \) and \( \alpha_2 \) may be equal even though the parent populations differ in respect both of \( \mu \) and \( \sigma^2 \). Unfortunately, there is then no test of the null hypothesis for the means \( \alpha \), since there is as yet no theory of joint confidence intervals for \( \mu \) and \( \sigma^2 \) for normal
populations: so that the statistician must confine himself to separate statements in regard to these two parameters. But the most common situation is that in which the samples can be regarded as drawn from parent populations with the same value of $\sigma^2$ but possibly different $\mu$; and this usually arises because the variate values are generated by an essentially similar process.

8.3. The Three-Parameter Distribution

The position for three-parameter distributions is less satisfactory. There is no rigorous theory for the testing of the parameter $\tau$, though for large samples recourse may be had to normal theory, using the estimated standard errors of the parameter estimates. Where three-parameter distributions have been involved it has been usual to work with the approximately normalized variate $y = \log(x - t)$, where $t$ is some estimate of $\tau$, and proceed with the two-parameter theory (see for example the application to lesion counts by Kleczkowski given in §10.4).

8.4. The Analysis of Variance

The more general comparison of samples from lognormal populations is contained in the theory of variance analysis. It is a prerequisite of this theory that there should be constant variances for all the populations involved; and if this is not the case for the original variate some transformation is sought which will ensure this property at least approximately for the transformed populations. Bartlett[12] has listed four criteria by which the success of a transformation may be judged. These are (considered after the transformation): (i) normality, (ii) independence of the variance and the mean, (iii) additivity of real effects, and (iv) efficiency of the mean of the sample as an estimator of the population mean; though, as he has pointed out, these are not independent. The first, third and fourth of these need not be further discussed, since by these criteria the success of a logarithmic transformation depends ultimately on the plausibility of such models of generation as are described in Chapter 3. Such reasoning, together with the empirical discovery that the coefficient of variation is approximately constant in the samples considered, is usually sufficient to justify the transformation; and the appropriate form of transformation is suggested by plotting the sample standard deviations against the sample means (compare §§5.9 and 6.2). Satisfactory results may however be obtained from an analysis of variance solely because a logarithmic transformation achieves stabilization (criterion (ii)). This approach has been studied by Curtiss[47] who establishes the result under fairly general conditions; and Cochran[39] advocates the transformation even when the untransformed data seem to indicate constant variances. Further discussion of the use of the transformation are to be found in Bartlett[12] and Quenouille[166, 167, 168].
CHAPTER 9
TRUNCATED AND CENSORED DISTRIBUTIONS AND THE TREATMENT OF ZERO OBSERVATIONS

First Lord. How mightily sometimes we make us comforts of our losses!
Second Lord. And how mightily some other times we drown our gain in tears!

All's Well That Ends Well

9.1. Definitions of Truncation and Censorship

Chapters 5 and 6 dealt with straightforward problems of estimation. There remain, however, a number of more complicated problems which must be disposed of before we can turn our attention from statistical theory to applications in particular fields.

A variate $X$ may be such that it appears to be lognormal except that that part of the distribution for which $X \leq \xi$ is removed, because such values of the variate either cannot occur or are not observed. The distribution of such a variate is said to be incomplete, or, more commonly, truncated, and $\xi$ is termed the point of truncation. The specification of the distribution is then:

$$P(X \leq x) = \begin{cases} 0 & (x \leq \xi) \\ \Lambda(x | \mu, \sigma^2) - \Lambda(\xi | \mu, \sigma^2) & (x > \xi) \end{cases}$$  \hspace{1cm} (9.1)

and

$$P(X \leq x) = \frac{\Lambda(x | \mu, \sigma^2)}{1 - \Lambda(\xi | \mu, \sigma^2)}$$  \hspace{1cm} (9.2)

and the $j$th moment is given by

$$E(X^j) = \int_{\xi}^{\infty} x^j d\Lambda(x | \mu, \sigma^2)$$

by Theorem 2.6,

$$= \frac{e^{\mu x + \frac{1}{2} \sigma^2 x^2}}{1 - \Lambda(\xi | \mu, \sigma^2)} \int_{\xi}^{\infty} d\Lambda(x | \mu + j\sigma^2, \sigma^2)$$

$$= e^{\mu \xi + \frac{1}{2} \sigma^2 \xi^2} \frac{1 - \Lambda(\xi | \mu + j\sigma^2, \sigma^2)}{1 - \Lambda(\xi | \mu, \sigma^2)}.$$  \hspace{1cm} (9.3)

The $j$th moment distribution may also be defined, with distribution function:

$$\int_{\xi}^{\infty} t^j d\Lambda(t | \mu, \sigma^2) = \int_{\xi}^{\infty} d\Lambda(t | \mu + j\sigma^2, \sigma^2)$$

by Theorem 2.6

$$= \Lambda(x | \mu + j\sigma^2, \sigma^2) - \Lambda(\xi | \mu + j\sigma^2, \sigma^2)$$

$$= \frac{\Lambda(x | \mu, \sigma^2) - \Lambda(\xi | \mu, \sigma^2)}{1 - \Lambda(\xi | \mu + j\sigma^2, \sigma^2)},$$  \hspace{1cm} (9.4)

which is again a truncated lognormal distribution with parameters $\mu + j\sigma^2$ and $\sigma^2$. 

102
THE LOGNORMAL DISTRIBUTION

An example of a truncated lognormal distribution is to be found in the distribution of incomes returned for income-tax purposes. It may be supposed that the incomes of all persons are lognormally distributed but the returned incomes are confined to the range above a certain minimum. Again in an expenditure inquiry there may be a minimum quantity of a commodity which it is possible to buy so that observed expenditures come from a truncated lognormal distribution.

In the taxable income example above it may be known what proportion of persons have income less than the minimum although the exact incomes are unknown. This leads to the concept of a censored distribution.

A variate \( X \) is said to have a censored lognormal distribution with point of censorship \( \xi \) if it belongs to the class of variates with

\[
P(X \leq x) = \Lambda(x | \mu, \sigma^2) \quad (x \geq \xi); \quad (9.5)
\]

in particular

\[
P(X \leq \xi) = \Lambda(\xi | \mu, \sigma^2). \quad (9.6)
\]

Moments for this distribution are undefined since the distribution is not defined precisely for every \( x \). The quantiles \( \xi_q \) of order \( q = \Lambda(\xi | \mu, \sigma^2) \) may be found and are the same as for the uncensored distribution, namely,

\[
\xi_q = e^{\mu + \tau_q \sigma}, \quad (9.7)
\]

where \( \tau_q \) is, as before, the \( N(0, 1) \) quantile of order \( q \).

The distinction between truncation and censorship thus arises from the fact that in the first the available information is confined to the range \((\xi, \infty)\), whereas in the second a limited knowledge of the variate in the range \((0, \xi)\) permits consideration of the complete range \((0, \infty)\).

The point of truncation or censorship \( \xi \) may of course be regarded as a parameter to be estimated rather than as a given point. For instance, in the example cited earlier the minimum quantity of the commodity which it is possible to buy may not be known \textit{a priori} and the point of truncation must be derived from the data. Such cases, however, must be rare and the authors feel that it should always be possible to obtain information on the point of truncation or censorship from sources other than the sample; we shall not therefore discuss this problem. Again, distributions may occur in which information is deficient for the higher values of the variate, or there may even be both lower and upper points of truncation or censorship. The theory of estimation which we now present does not cover these cases though such an extension is clearly possible.

9.2. ESTIMATION OF THE PARAMETERS OF THE TRUNCATED DISTRIBUTION

For the \( \Lambda(\mu, \sigma^2) \) distribution truncated at the known point \( \xi \) we confine our attention to the estimation of \( \mu \) and \( \sigma^2 \). If the sample is \( x_1, x_2, \ldots, x_n \) (all \( x_i > \xi \)) then the likelihood of the sample is

\[
\prod_{i=1}^{n} \frac{1}{x_i \sigma \sqrt{(2\pi)}} \exp \left\{ -\frac{1}{2\sigma^2} (\log x_i - \mu)^2 \right\} \left( 1 - \Lambda(\xi | \mu, \sigma^2) \right), \quad (9.8)
\]
and so the loglikelihood function, as far as it contains $\mu$ and $\sigma^2$, may be written

$$L = -n \log N(-\xi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum(y - \mu)^2,$$

(9.9)

where $y_i = \log x_i$, $\xi = (v - \mu)/\sigma$ and $v = \log \xi$; $N$ denotes the standardized normal distribution function. The method of maximum likelihood therefore leads to the equations

$$\frac{\partial L}{\partial \mu} = \frac{n}{\sigma} N(-\xi) - \frac{1}{\sigma^2} \sum(y - \mu),$$

(9.10)

and

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} N(-\xi) - \frac{n}{2\sigma^4} \sum(y - \mu)^2.$$

(9.11)

If $m$ and $s^2$ denote the estimators of $\mu$ and $\sigma^2$ respectively and if

$$z = \frac{v - m}{s},$$

(9.12)

these equations for determining $m$ and $s^2$ may be arranged as

$$\frac{1}{n} \sum(y - v) = \frac{N'(-z)}{N(-z)} - z = \frac{1}{g(z)},$$

say,

(9.13)

and

$$\frac{1}{n} \sum(y - v)^2 = 1 - z \left\{\frac{N'(-z)}{N(-z)} - z\right\} = \frac{g(z) - z}{g(z)}.$$

(9.14)

From equations (9.13) and (9.14)

$$\frac{1}{2} g(z) \{g(z) - z\} = \frac{n \sum(y - v)^2}{2 \sum(y - v)^2},$$

(9.15)

The right-hand side of this equation may be calculated and so $z$ may be determined by inverse interpolation as indicated by Fisher[70]. The inverse function of $\frac{1}{2} g(z) \{g(z) - z\}$ has, however, been tabulated by Hald[95]; this allows direct interpolation for $z$; the estimates $m$ and $s^2$ are then readily found from (9.12) and (9.13) by

$$s = g(z) \frac{\sum(y - v)}{n},$$

(9.16)

and

$$m = v - zs;$$

(9.17)

g(z) is also tabulated by Hald. The variance matrix of the estimators may be derived in the usual way; Hald gives an explicit expression for it and tables to simplify the evaluation.

Although in principle it would be possible to apply the method of moments by equating the expressions given by (9.3) to the first and second sample moments there seems to be no easy way of solving the resulting equations for $\mu$ and $\sigma^2$. Pearson and Lee[161] and Lee[134] have applied the methods of moments to the transformed samples and given tables to assist in the computation; Fisher[70] has shown that this
method is equivalent to that of maximum likelihood, as is evident from the expression for the moments of \( Y = \log X \), namely,

\[
E(Y - \mu) = \sigma \left\{ \frac{N'(-\xi)}{N(-\xi)} - \xi \right\}
\]

(9.18)

and

\[
E((Y - \mu)^2) = \sigma^2 \left\{ 1 - \xi \left\{ \frac{N'(-\xi)}{N(-\xi)} - \xi \right\} \right\}.
\]

(9.19)

Cohen [40, 41] has suggested an iterative procedure, based on the Newton-Raphson method, for the solution of the maximum-likelihood equations without the use of specially constructed tables; he extends his method to the case of a doubly-truncated distribution.

Finally, it is to be noted that the method of quantiles cannot be readily applied; and that the graphical method is inappropriate since, to apply it, an estimate of the area of the \( \Lambda(\mu, \sigma^2) \) curve between 0 and \( \xi \) is required and there is no information available on this.

9.3. **Estimation of the Parameters of the Censored Distribution**

We now consider a \( \Lambda(\mu, \sigma^2) \) distribution censored at \( \xi \) and suppose that the sample of size \( n \) consists of \( n_1 \) observations not greater than \( \xi \) whose exact values are unknown, and \( n_2 \) observations \( x_1, \ldots, x_{n_2} \) (all \( x_i > \xi \)). The likelihood of the sample is then

\[
\left( \frac{n}{n_1} \right) \left( \Lambda(\xi | \mu, \sigma^2) \right)^{n_1} \prod_{i=1}^{n_2} \frac{1}{x_i \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (\log x_i - \mu)^2 \right\};
\]

(9.20)

and the loglikelihood function \( L \) may be written

\[
L = n_1 \log N(\xi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum (y - \mu)^2,
\]

(9.21)

where, as before, \( y_i = \log x_i \), \( \xi = (v - \mu) / \sigma \) and \( v = \log \xi \). The maximum-likelihood equations for \( \mu \) and \( \sigma^2 \) are

\[
\frac{\partial L}{\partial \mu} = -\frac{n_1}{\sigma} N'(\xi) + \frac{1}{2\sigma^2} \sum (y - \mu),
\]

(9.22)

\[
\frac{\partial L}{\partial \sigma^2} = -\frac{n_1}{2\sigma^2} N'(\xi) - \frac{n_2}{2\sigma^2} N(\xi) - \frac{n_2}{2\sigma^4} \sum (y - \mu)^2.
\]

(9.23)

Let \( \hat{\mu} \) and \( \hat{\sigma}^2 \) denote the estimators of \( \mu \) and \( \sigma^2 \); if

\[
z = \frac{\nu - \hat{\mu}}{\hat{\sigma}}
\]

(9.24)

the likelihood equations may be written

\[
\frac{1}{n_2} \sum (y - \mu) = \frac{n_1}{n_2} N'(z) - z = \frac{1}{g(h, z)},
\]

(9.25)
say, where \( h = n_1/n \), and

\[
\frac{1}{n_2^2} \sum (y - v)^2 = 1 - z \left\{ \frac{n_1}{n} \frac{N'(z)}{N(z)} - z \right\}
\]

\[
g(h, z) = \frac{g(h, z) - z}{g(h, z)}. \quad (9.25)
\]

It follows from (9.25) and (9.26) that

\[
\frac{1}{2} g(h, z) \{ g(h, z) - z \} = n_2 \frac{\sum (y - v)^2}{2 [\sum (y - v)]^2}. \quad (9.26)
\]

The procedure follows closely that of the truncated distribution. Hald [95] has tabulated the inverse of the function \( \frac{1}{2} g(h, z) \{ g(h, z) - z \} \), so that it is possible to read off the value of \( z \) corresponding to a given value of \( h \) and a given value of the right-hand side of (9.27). A table of \( N'(z)/N(z) \) as a function of \( z \) is also given, so that

\[
g(h, z) = \frac{1}{n_1} \frac{N'(z)}{N(z)} - z \quad (9.27)
\]

may be easily calculated in order to give

\[
s = g(h, z) \frac{\sum (y - v)}{n_2} \quad (9.28)
\]

and

\[
m = v - zs. \quad (9.29)
\]

An expression for the variance matrix and auxiliary tables are also provided by Hald [95]. Stevens [180] and Gupta [94] have also made contributions to the maximum-likelihood theory of this case, and Cohen [40, 41] has applied his iterative procedure to the censored as well as to the truncated distribution.

As previously mentioned, moments cannot be determined and so the method of moments is not applicable. The method of quantiles, however, may be used provided that quantiles of order greater than \( n_1/n \) are employed; the procedure is that of § 5.2, but if \( n_1/n \) is appreciably greater than 10% it may be advantageous to use asymmetrically placed quantiles. The graphical method may also be used, the \( n_1 \) sample values in the range \((0, \xi)\) contributing towards the single point \((n_1/n, \xi)\) on the probability graph. The ease with which these last two methods may be applied makes them particularly attractive, although there is some loss of efficiency compared with the method of maximum likelihood.

### 9.4. An Application to the Artificial Samples

The reader may be interested to see the results of applying the method of maximum likelihood to truncated and censored artificial samples. Each of the samples of size 128 was truncated and censored at the point \( \xi = e^{x - \nu} \) and the estimates of \( \mu \) and \( \sigma^2 \) calculated; the procedure was repeated for \( \xi = e^{\nu} \).
Table 9.1 compares the estimates of $\mu$ so obtained with the maximum-likelihood estimates for the full samples; values of $\Delta(m)$ have been calculated as in §5.3. Table 9.2 gives the corresponding comparison for the estimates of $\sigma^2$. The results conform with expectation: comparatively little information is lost under censorship with respect to the estimation of $\mu$, even when about one-half the observations are censored; with respect to $\sigma^2$ the loss is greater but still less than that due to truncation.

### Table 9.1. Maximum-Likelihood Estimates† of $\mu$ from Full, Truncated and Censored Samples of Size 128

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Full sample estimates</th>
<th>$\xi = e^{\mu - \sigma^2}$</th>
<th>Truncated sample estimates</th>
<th>Censored sample estimates</th>
<th>$\xi = e^{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu = 0$</td>
<td>0.059</td>
<td>0.117</td>
<td>0.059</td>
<td>0.065</td>
<td>0.072</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.032</td>
<td>0.013</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.032</td>
<td>0.013</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
<tr>
<td>$\mu = 0$</td>
<td>0.047</td>
<td>0.011</td>
<td>-0.003</td>
<td>0.062</td>
<td>0.066</td>
</tr>
</tbody>
</table>

† $\mu = 0$ for all samples.

### 9.5. Distributions of Counts

A particular form of the censored distribution has been found useful as an approximation to a discrete distribution of counts[92, 166, 179, 185, 209]. This is usually done by supposing that $\phi_r$, the frequency of $r$ ($r = 0, 1, \ldots$), is given by

$$
\phi_r = \int_r^{r+1} d\Lambda(x \mid \mu, \sigma^2) = \int_{\log(r+1)}^{\log(r+1)} dN(y \mid \mu, \sigma^2).
$$

Such distributions occur in experiments on counts of insects, for example, aphides on leaves; other examples are numbers of spores on culture plates and sentence length of different authors (see §§10.6 and 10.11).

The number of zeros in the distribution appears as the part of the
normal distribution in the range \((-\infty, 0]\). The treatment suggested by Thompson [185] is to consider the variate \(Y\) where

\[
P(Y < 0) = 0,
\]

\[
P(Y = 0) = N(0 \mid \mu, \sigma^2)
\]

and

\[
P(Y \leq y) = N(y \mid \mu, \sigma^2) \quad (y > 0),
\]

TABLE 9.2. MAXIMUM-LIKELIHOOD ESTIMATES OF \(\sigma^2\) FROM FULL, TRUNCATED AND CENSORED SAMPLES OF SIZE 128

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>Full sample estimates</th>
<th>(\xi = e^{\sigma^2})</th>
<th>Full sample estimates</th>
<th>(\xi = e^{\sigma^2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.040</td>
<td>0.049</td>
<td>0.043</td>
<td>0.040</td>
</tr>
<tr>
<td>0.3</td>
<td>0.075</td>
<td>0.076</td>
<td>0.078</td>
<td>0.076</td>
</tr>
<tr>
<td>0.4</td>
<td>0.100</td>
<td>0.101</td>
<td>0.102</td>
<td>0.101</td>
</tr>
<tr>
<td>0.5</td>
<td>0.125</td>
<td>0.126</td>
<td>0.127</td>
<td>0.126</td>
</tr>
<tr>
<td>0.6</td>
<td>0.150</td>
<td>0.151</td>
<td>0.152</td>
<td>0.151</td>
</tr>
<tr>
<td>0.7</td>
<td>0.175</td>
<td>0.176</td>
<td>0.177</td>
<td>0.176</td>
</tr>
<tr>
<td>0.8</td>
<td>0.200</td>
<td>0.201</td>
<td>0.202</td>
<td>0.201</td>
</tr>
<tr>
<td>0.9</td>
<td>0.225</td>
<td>0.226</td>
<td>0.227</td>
<td>0.226</td>
</tr>
<tr>
<td>1.0</td>
<td>0.250</td>
<td>0.251</td>
<td>0.252</td>
<td>0.251</td>
</tr>
</tbody>
</table>

and to estimate the parameters \(\mu\) and \(\sigma^2\) from the transformed sample. The transformation† in this case is

\[
y = \log (r + 1),
\]

and the problem really concerns a normal distribution, censored at the origin, for which the censored portion appears as a discrete probability mass at the origin.

The estimation procedure used by Thompson is the method of moments; the first two sample moments are equated to their theoretical values and the resulting equations solved for \(\mu\) and \(\sigma^2\). Corrections to the estimators are, however, necessary because of the grouped nature of the distribution; this is complicated by the fact that the intervals are unequal on the transformed scale and that there is not high contact at the extremities of the range, so that corrections of the Sheppard type are inappropriate. Thompson has overcome this difficulty by computing corrections empirically from artificial samples; these corrections are tabulated by him.

† An alternative, \(y = \log (ar + 1)\), is suggested by Pearce [153].
A maximum-likelihood solution to the problem would also be possible using the results of Gjeddeback [90] already quoted in § 5.8 on estimation from grouped observations for the normal distribution.

9.6. GROUPED TRUNCATED AND GROUPED CENSORED DISTRIBUTIONS

If the uncensored portion of a censored distribution is available in grouped form only, the whole distribution may be considered grouped and the methods of § 5.7 applied. On the other hand, if grouping co-exists with truncation a problem of a new order arises. We shall now describe a method due to Grundy [92] which is applicable in either case, and so provides a solution to the problem of grouped, truncated data, as well as an alternative to the methods of § 5.7 for simple grouping. In particular, the method is useful for the case of discrete counts when there is some grouping of the integral values.

Grundy's contribution to the problem is essentially to provide simple expressions for adjusting the first and second 'raw' moments, computed from the grouped data, so that the adjusted moments may be used in any of the methods previously described for estimation under simple censorship or truncation. The raw moments are defined

\[ l_t^* = \sum_i f_i u_i^l \left/ \sum_i f_i \right. \]

where \( u_i = \frac{1}{2}(y_i + y_{i-1}) \) is the mid-point of the \( i \)th interval with boundaries \( y_i \) and \( y_{i-1} \) \( (y_0 = \log \xi) \) and \( f_i \) the number of sample values in the \( i \)th interval. For a truncated or censored normal distribution the first two adjusted moments (with expectations equal to the population truncated or censored moments) are given by Grundy's approximate formulae

\[ l_1^{**} = l_1^* - \frac{\sum f_i v_i^2 u_i - \mu^2 \sum f_i v_i^2}{12 \sigma^2 \sum f_i} \]

(9.37)

and

\[ l_2^{**} = l_2^* + \frac{\sigma^2 \sum f_i v_i^2 + 2 \mu \sum f_i v_i^2 u_i - 2 \sum f_i v_i^2 u_i^2}{12 \sigma^2 \sum f_i} \]

(9.38)

where \( v_i = y_i - y_{i-1} \) is the length of the \( i \)th interval. For the \( \Lambda(\mu, \sigma^2) \) distribution the intervals \( (x_{i-1}, x_i) \) are first transformed to \( (y_{i-1}, y_i) \) by \( y_i = \log x_i \), or in the case of counts by \( y_i = \log (i + 1) \). The raw moments and adjustments using initial estimates of \( \mu \) and \( \sigma^2 \) are easy to compute; and when the adjusted moments are used to obtain new values of \( \mu \) and \( \sigma^2 \) these new values may be used again in (9.37) and (9.38) to initiate a new cycle. Grundy also gives expressions for loss of information due to grouping and illustrates his method with numerical examples.

9.7. THE TREATMENT OF ZERO OBSERVATIONS

Lognormal theory cannot be applied directly to any sample which contains a zero value. Yet such samples do occur and with non-zero observations appearing to come from a lognormal population. Some
writers suggest that a positive constant should be added to all sample values before the logarithmic transformation is applied; this may be the appropriate procedure where a distribution of counts is involved as in §9.5, or where the population is of the form \( \Lambda(\tau, \mu, \sigma^2) \) with negative \( \tau \) and it happens that some sample values are zero. Others\cite{109} replace the zero values by positive constants before taking logarithms; this may be justified when it is known that the population is censored and that the reported zero values are in reality the values lying in the censored portion; in this case it would probably be better to apply the techniques of §9.3.

There are, however, situations in which neither of these devices is correct and it becomes necessary to recognize explicitly a dichotomy of the population into zero and non-zero values. For example, in a household expenditure investigation there may be certain commodities on which a definite proportion of households does not spend; the population then divides naturally into two groups, spenders and non-spenders, and the statistical model should be similarly postulated. In the next section a statistical model of this type is presented and a few of its more important properties derived; and in the following section the appropriate estimation procedures are considered. The theory is based on a more general treatment by Aitchison\cite{4}.

9.8. A Statistical Model: the \( \Delta \)-distribution

Let a population be such that there is a proportion \( \delta \) of zero values and the distribution of non-zero values is \( \Lambda(\mu, \sigma^2) \). If \( Z \) denotes the corresponding variate then

\[
\begin{align*}
P(Z < 0) &= \delta, \\
P(Z = 0) &= \delta \\
\text{and} \quad P(Z \geq z) &= \delta + (1 - \delta) \Lambda(z | \mu, \sigma^2) \quad (z > 0);
\end{align*}
\]

we then write that \( Z \) is \( \Delta(\delta, \mu, \sigma^2) \).

The distribution possesses moments of any order; the \( j \)th moment about the origin is

\[
E(Z^j) = (1 - \delta) \mu^j + \delta \eta^j, \tag{9.42}
\]

and so the mean \( \kappa \) and variance \( \rho^2 \) are given by

\[
\begin{align*}
\kappa &= (1 - \delta) \mu + \delta \eta; \\
\rho^2 &= (1 - \delta) \mu^2 + \delta \eta^2 - \delta(1 - \delta) \mu^2; \tag{9.43}
\end{align*}
\]

and

\[
\begin{align*}
E((Z - \kappa)^3) &= (1 - \delta) \alpha^3[(1 + \eta)^3 - (1 - \delta)(1 + \eta)^2 + 2(1 - \delta)^2] \tag{9.45}
\end{align*}
\]

and

\[
\begin{align*}
E((Z - \kappa)^4) &= (1 - \delta) \alpha^4[(1 + \eta)^4 - 4(1 - \delta)(1 + \eta)^3 + 6(1 - \delta)^2(1 + \eta)^2 - 3(1 - \delta)^3]. \tag{9.46}
\end{align*}
\]
THE LOGNORMAL DISTRIBUTION

As with the other lognormal distributions, a simple relation holds between the quantiles of \( \Delta(\delta, \mu, \sigma^2) \) and those of \( N(\mu, \sigma^2) \). Suppose that \( \xi_q \) is the quantile of order \( q \) of \( \Delta \). Then, if \( q < \delta \), \( \xi_q = 0 \) and if \( q > \delta \),

\[
\xi_q = \exp\{q(\mu + \nu_q \sigma)\},
\]

where

\[
q' = \frac{q - \delta}{1 - \delta}.
\]

There are analogues of the reproduction properties of the lognormal distribution. If \( Z_1 \) and \( Z_2 \) are independent and \( \Delta(\delta_1, \mu_1, \sigma_1^2) \) and \( \Delta(\delta_2, \mu_2, \sigma_2^2) \) respectively, then the product \( Z_1 Z_2 \) is

\[
\Delta(1 - (1 - \delta_1)(1 - \delta_2), \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).
\]

Clearly an extension to any finite or infinite sequence of independent \( \Delta \)-variates is possible provided that for an infinite sequence \( \prod (1 - \delta_j) \) converges. The analogue of Cramer's Theorem 1451 is also true for \( \Delta \)-variates: namely, if \( Z_1 \) and \( Z_2 \) are independent, non-negative variates and \( Z_1 Z_2 \) is a \( \Delta \)-variates the \( Z_1 \) and \( Z_2 \) are separately \( \Delta \)-variates.

Central-limit theorems also exist in this theory; a simple form is: if \( Z_j \) is a sequence of independent non-negative random variables such that

(i) \( P(Z_j = 0) = \delta_j \);
(ii) the distribution of \( Z_j \mid Z_j = 0 \) is the same for all \( j \), and
(iii) \( E[\log Z_j \mid Z_j = 0] = \mu \) and \( \text{Var}[\log Z_j \mid Z_j = 0] = \sigma^2 \) are finite, then

the product \( \prod_{j=1}^n Z_j \) is asymptotically

\[
\Delta\left\{1 - n \prod_{j=1}^n (1 - \delta_j), n\mu, n\sigma^2\right\}.
\]

This, or possibly a more general formulation, may be used in a way similar to that of Chapter 3 to explain the genesis of \( \Delta \)-type populations.

9.9. ESTIMATION FOR THE \( \Delta \)-DISTIBUTION

Suppose that a sample \( S_n \) of size \( n \) is given from \( \Delta(\delta, \mu, \sigma^2) \), that the sample contains \( n_0 \) zero values, and the remaining \( n_1 = n - n_0 \) non-zero values are \( x_i \) \( (1 \leq i \leq n_1) \). Let

\[
\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i
\]

and

\[
\bar{v}_2 = \frac{1}{n - 1} \left\{ \sum_{i=1}^{n_1} (x_i - \bar{x})^2 + n_0 \bar{x}^2 \right\}
\]

be the usual estimators of mean and variance respectively from the untransformed sample values; also write

\[
\bar{y} = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i \quad (n_1 > 0)
\]

\[
= 0 \quad (n_1 = 0)
\]

\( n_1 \)
and
\[ t^2_y = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (y_i - \bar{y})^2 \quad (n_1 > 1) \]
\[ = 0 \quad (n_1 = 0, 1). \tag{9.52} \]

where \( y_i = \log x_i \) \((1 \leq i \leq n_1)\). The likelihood of the sample is
\[ \left( \frac{n}{n_0} \right) \delta_n (1 - \delta)^{n_1} \frac{1}{(\sigma \sqrt{(2\pi)})^{n_1}} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n_1} (y_i - \mu)^2 \right), \tag{9.53} \]

which from normal theory can be expressed in the following form:
\[ \left( \frac{n}{n_0} \right) \delta_n (1 - \delta)^{n_1} \times \text{likelihood of } \bar{y} \mid \delta, \mu, \sigma^2 \]
\[ \times \text{likelihood of } t^2_y \mid \delta, \mu, \sigma^2 \]
\[ \times \text{a likelihood which is independent of } \delta, \mu, \sigma^2. \]

Hence \( n_0/n, \bar{y} \) and \( t^2_y \) are joint sufficient estimators for \( \delta, \mu \) and \( \sigma^2 \), and so any function of \( n_0/n, \bar{y} \) and \( t^2_y \) is a most efficient estimator of its expectation.

Hence if the problem is simply to estimate \( \delta, \mu \) and \( \sigma^2 \) one cannot do better than use as estimators \( n_0/n, \bar{y} \) and \( t^2_y \) respectively. The methods of moments and of quantiles are certainly less efficient; if it were desired to use a graphical method \( \delta \) could be estimated by \( n_0/n \) and the non-zero part of the sample would then lead to graphical estimates of \( \mu \) and \( \sigma^2 \) in the usual way.

Suppose, on the other hand, that the problem is to obtain most efficient estimators of \( \kappa \) and \( \rho^2 \). For this purpose functions of \( n_0/n, \bar{y} \) and \( t^2_y \) must be found whose expectations are \( \kappa \) and \( \rho^2 \). These estimators, \( k \) and \( r^2 \), are given by
\[ k = \frac{n_1}{n} e^0 \psi_n \left( \frac{1}{2} \psi_n \right) \quad (n_1 > 1) \]
\[ = \frac{x_1}{n} \quad (n_1 = 1) \]
\[ = 0 \quad (n_1 = 0), \tag{9.54} \]

and
\[ r^2 = \frac{n_1}{n} e^2 \left[ \psi_{n/2} \right] \left( 1 - \frac{n_1 - 1}{n - 1} \psi_{n/2} \right) \quad (n_1 > 1) \]
\[ = \frac{x_2^2}{n} \quad (n_1 = 1) \]
\[ = 0 \quad (n_1 = 0). \tag{9.55} \]

It would have been possible to use special definitions of \( \bar{y} \) for the case \( n_1 = 0 \) and of \( t^2_y \) for the cases \( n_1 = 0 \) or \( 1 \) and to introduce conventions about the interpretation of the function \( \psi_0(t) \) in the general formulae for \( k \) and \( r^2 \), but it is more convenient to set down explicitly the results of such an interpretation for these special cases. All that requires proof is that \( k \) and \( r^2 \) are unbiased estimators of \( \kappa \) and \( \rho^2 \) respectively, and this is readily shown. For, if \( E(X \mid Y \epsilon S) \) denotes the expectation of \( X \) conditional
on $Y$ belonging to the set $S$ for any two random variables $X$ and $Y$, then

$$E[k] = \sum_{i=0}^{n} P[n_1 = i] E[k \mid n_1 = i]$$

$$= o + P[n_1 = 1] \frac{\alpha}{n} + \sum_{i=2}^{n} P[n_1 = i] E[k \mid n_1 = i]$$

$$= P[n_1 = 1] \frac{\alpha}{n} + \sum_{i=2}^{n} P[n_1 = i] E\left(\frac{n_1}{n} \alpha \mid n_1 = i\right)$$

$$= \sum_{i=0}^{n} P[n_1 = i] E\left(\frac{n_1}{n} \alpha \mid n_1 = i\right)$$

$$= E\left(\frac{n_1}{n} \alpha \right)$$

$$= (1 - \delta) \alpha$$

and similarly

$$E(\tau^2) = \mu^2. \quad (9.56)$$

$$E(\tau^2) = \rho^2. \quad (9.57)$$

Fig. 9.1. Efficiency of estimation of $\kappa$ and $\rho^2$ in a $\Delta$-distribution for two values of $\delta$. 

113
It may also be shown that, for large \( n \) and \( \delta \) appreciably less than 1,
\[
D^2(k) = \frac{\alpha^2}{n} \left\{ \delta(1 - \delta) + \frac{1}{2}(1 - \delta)(2\sigma^2 + \sigma^4) \right\} \tag{9.58}
\]
and
\[
D^2(\sigma^2) = \frac{(1 - \delta)\alpha^4}{n} \left[ \delta(1 + \eta^2 - 2(1 - \delta))^3 + 4\sigma^2(1 + \eta^2 - (1 - \delta))^2 
+ 2\sigma^4(2(1 + \eta^2) - (1 - \delta))^2 \right]. \tag{9.59}
\]

Since
\[
D^2(\bar{x}) = \frac{(1 - \delta)\alpha^2}{n} \left\{ 1 + \eta^2 - (1 - \delta) \right\} \tag{9.60}
\]
and
\[
D^2(\nu^2_2) = \frac{(1 - \delta)\alpha^4}{n} \left[ (1 + \eta^2)^3 - 4(1 - \delta)(1 + \eta^2)^3 - (1 - \delta)(1 + \eta^2)^2 
+ 8(1 - \delta)^2(1 + \eta^2) - 4(1 - \delta)^3 \right], \tag{9.61}
\]
approximations to the efficiencies of \( \bar{x} \) and \( \nu^2_2 \) as estimators of \( \kappa \) and \( \rho^2 \) respectively may be obtained:
\[
\text{eff.} \{ \bar{x} \} = \frac{\delta + \sigma^2 + \frac{1}{2}\sigma^4}{\delta + \eta^2} \quad \tag{9.62}
\]
and
\[
\text{eff.} \nu^2_2 = \frac{\delta(1 + \eta^2 - 2(1 - \delta))^3 + 4\sigma^2(1 + \eta^2 - (1 - \delta))^2 
+ 2\sigma^4(2(1 + \eta^2) - (1 - \delta))^2}{(1 + \eta^2)^3 - 4(1 - \delta)(1 + \eta^2)^3 - (1 - \delta)(1 + \eta^2)^2 
+ 8(1 - \delta)^2(1 + \eta^2) - 4(1 - \delta)^3}. \tag{9.63}
\]

All the above theory and formulae apply to the case \( \delta = 0 \) and hence \( n_0 = 0, n_1 = n \). Graphs of the efficiencies plotted against \( \sigma^2 \) are given for \( \delta = 0 \) and \( 0.5 \) in Fig. 9.1. While \( \bar{x} \) is a very good estimator of \( \kappa \), \( \nu^2_2 \) is not a very efficient estimator of \( \rho^2 \).
CHAPTER 10
EXAMPLES OF LOGNORMAL DISTRIBUTIONS

Third Gentleman. And many other evidences proclaim her.
*The Winter’s Tale*

10.1. INTRODUCTION

The purpose of this chapter is to review briefly, and mainly by means of examples taken from the literature, occurrences of observed distributions that are adequately described by one or other of the lognormal formulae. The majority of references relate to an explicit use of the lognormal hypothesis, and we distinguish cases where distributions have been published with no mention of a mathematical model, or one other than the lognormal. The list does not pretend to be exhaustive, though it is hoped that the examples are sufficiently widely drawn to interest workers in several fields and perhaps to suggest possible extensions.

10.2. SMALL-PARTICLE STATISTICS

The first group of examples is taken from the domain of small-particle statistics, where the lognormal distribution is now well established. Distributions of small particles, such as are found as the result of natural processes in soils or rocks, or as the result of mechanical processes such as grinding, are often very skew with as much as a hundred-fold increase in size from the smallest particle to the largest. In addition, the choice of a mathematical description is restricted because investigators are often interested in a number of related particle measurements, such as their diameters, volumes and weights, which can be ordered in ascending powers of one variable; and for physical reasons it is sometimes convenient to take measurements in terms of one characteristic and translate the results into terms of another. It was early recognized, for example, by Hatch and Choate[102] and again by Krumbein[131], that the lognormal distribution offers a great advantage in this type of situation, since, if an elementary variate is lognormally distributed, so are any powers of the variate, including fractional powers (Theorem 2.1).

In 1933 Hatch[101] made the more interesting discovery that a simple relationship exists between size-frequency curves by count and by weight, this being the first reference in the literature to what the authors have termed the moment distributions of the lognormal. It is worth while, in view of the practical usefulness of Hatch’s device in the measurement of small particles, to give the gist of his argument. The problem arises when particles are sieved through metal-cloth screens of graduated mesh size and the percentage by weight of the material retained on each screen is
determined; for this is incomparably simpler than attempting to count the number of particles retained. The object of the sieving, however, may be to estimate the mean particle diameter. If \( x \) is the diameter of the particles retained, or, more accurately, if the diameters lie in the range \( (x, x + dx) \), and \( n \) is their number, the weight retained will be proportional to \( nx^3 \). If, further, the distribution obtained by regarding the relative weights as relative frequencies is \( \Lambda(\mu, \sigma^2) \), then the distribution which would be obtained by using the relative numbers of retained particles as the frequency measure is a moment distribution of this, namely, \( \Lambda_3(\mu, \sigma^2) \); and this, by Theorem 2.6, is equivalent to \( \Lambda(\mu - 3\sigma^2, \sigma^2) \).

The mean particle diameter is then easily derived as \( \exp(\mu - 0.06) \).

Hatch did not generalize the moment property from his particular result, but in the same paper published examples of empirical verification, using both weight measures and direct microscopic measures of the particle sizes. Later Krumbcin suggested the use of the number 2 as the base of the logarithmic transformation, naming this the Phi-transformation; his purpose was to make the transformed size measure conform with the standard grading of sieve-meshes.

In 1941 Kolmogoroff was prompted by the work of Rasumovsky on auriferous ores to suggest the model of genesis of the distribution which we have described in §3.6; Kolmogoroff’s approach was further discussed by Epstein in 1947. Recent applications have included those of Kottler in 1950 and of Krigel in 1951, who applied lognormal theory to gold-mine valuation problems on the Witwatersrand. For more details of applications in this field the reader is referred to the manual of Krumbcin and Pettijohn, to the book published by Herdan in 1953, and to a historical survey by Krumbcin in 1954.

10.3. Economics and Sociology

The two best documented of the fields of application of lognormal theory are the foregoing and the field of income size distributions. In regard to the latter there is the considerable empirical work of Kapteyn, Gibrat and Van der Wijk in particular; and much theoretical development was stimulated by the problems arising. There is no need to pursue this field further here, since it is dealt with in greater detail in the next chapter. Gibrat, however, studied a number of examples in economics other than those pertaining to incomes. Amongst these can be cited distributions of inheritances, bank deposits and total wealth possessed by individual persons, of industries and firms by the numbers of employed persons, of industrial profits, and of towns and communes by the numbers of inhabitants. In most of these cases Gibrat used the three-parameter distribution, but his method of estimation was graphical and it is often doubtful whether the introduction of the third parameter was justified or its interpretation reasonable. A number of similar distributions are given by Zipf, who uses a mathematical description.
of his own manufacture on which he erects some extensive sociological
theory; in fact, however, it is likely that many of these distributions can
be regarded as lognormal, or truncated lognormal, with more prosaic
foundations in normal probability theory.

Expenditures on particular commodities, or the prices paid per unit
of a commodity by individual families, are often approximately log-
normal; in the case of broadly grouped commodities, such as all food,
the simple, two-parameter hypothesis is adequate, and the logarithmic
transformation is usually necessary before proceeding to any advanced
analysis of the data of budgetary surveys. For particular commodities,
if the time period of the survey is small relative to the normal period of
purchasing, the analysis of expenditures is complicated by the existence
of a proportion of households which record no expenditure. Sometimes
the distribution can be treated as censored lognormal, but in other cases
it is necessary to treat the division between spenders and non-spenders
by the methods described in §9.7. Observed expenditure distributions
are again given by Gibrat[188], and also by Utting and Cole[190,191]; and
data on the prices paid for tea by individual families in the United
Kingdom in 1937–8 are given by Prais and Aitchison[163], who show
that the arithmetic standard deviation of price was proportional to the
mean price. The value of $\sigma^2$ for the distribution of the price paid per
unit of a commodity by families is often small, so that the distribution is
not very different in appearance from the normal; the price is the ratio
of two lognormal variates, expenditure and quantity purchased, usually
highly correlated and with approximately equal logarithmic variances;
if this common variance is written $\sigma^2$, the variance of the price variate
is $2\sigma^2(1 - \rho)$, where the coefficient of correlation $\rho$ is near to unity.

Distributions of the changes of income payments by individual states
in the U.S.A. are given by Vining[196,197], who suggests the lognormal
curve as a description. Evidence of lognormality in price statistics (the
distribution of price changes over time for a large number of com-
modities) led Davies[52] in 1946 to advocate the use of the geometric
mean in index numbers.

Other uses of the hypothesis in economic contexts are discussed in
Chapter 12.

10.4. BIOLOGY

The occurrence of lognormal distributions in biology also has been
frequently noted. As mentioned in §3.3 Cramér[46] following Wick-
sell[203] discusses the growth of an organism subject to a number of
small independent impulses acting in an ordered sequence; if the in-
fluence of each impulse is proportionate to the momentary size of the
organism, the law of proportionate effect applies, and the final size of
the organism will tend to be lognormally distributed. This hypothesis
seems to be supported in a number of instances where biological size
distributions have been observed, for example, by Hemmingsen[104];
and the references in the literature begin with Kapteyn's[115] paper of
Examples of Lognormal Distributions

More frequently in biological studies the investigator has not been interested in the size distribution of a variable per se but has nevertheless used the logarithmic transformation before proceeding to an analysis of variance (Yates[215]), or to a multiple regression (Williams[208]). Williams[207] has also discussed more extensively the use of logarithms in the interpretation of entomological problems; Sinnott[178] has used lognormal theory in a study of the relation of gene to character in quantitative inheritance; and Haldane and Kermack[97] have considered bivariate lognormal distributions in relation to allometry. The transformation \( y = \log \left( \frac{x}{(100-x)} \right) \), where \( x \) is the percentage of clover in a lawn, was suggested by Bartlett[11] in a discussion of transformations useful in applied biology: this may be regarded as a special case of the four-parameter distribution described in \( \S 2.10 \). Bernstein and Weatherall[119] give more general references in relation to statistical data drawn from biological and agricultural sources; and Van Uven[195] in 1946 makes use of the lognormal hypothesis in his book on the analysis of agricultural experiments.

10.5. Anthropometry

The distribution of bodyweights of human beings has been studied in some detail by Yuan[216] who used a normal distribution for heights and a lognormal for weights. On the other hand Camp[30] argues that even for human heights, which have long been used as an example of normality in nature, as good or possibly better fit is obtained by the lognormal, and Camp suggests the use of the latter for most anthropometric measurements. In the case of many of these, in particular of human weights, the three-parameter distribution is necessary, as recognized both by Yuan[216], and by Kemble[120, 121, 122] who has examined numerous British data.

10.6. The Abundance of Species and Discrete Counts

The law of proportionate effect can also be adduced to explain the relative abundance of species[207]. Preston[165] has published some supporting evidence, and Grundy[91] has considered the problem of sampling from a lognormal distribution of species by traps or other methods which are so arranged that individual members of the species have small and independent chances of being captured (so called Poisson sampling).

A number of biological distributions take the form of discrete counts (such as the number of fungal spores on organisms, or of aphides on the leaves of plants), and Thompson[185] and Grundy[92] have discussed such counts on the assumption that they are discrete manifestations of an underlying lognormal variate; the problem is one of grouping, with censorship or truncation also present in many cases. An interesting application to biological counts was made by Kleczkowski[126] who
counted the lesions on leaves caused by plant viruses on a large number of plants. For each plant he obtained an estimate of the arithmetic mean and standard deviation of the number of lesions per leaf; the relationship between means and standard deviations was linear. On the lognormal hypothesis the slope of the line is an estimate of $\eta = \sqrt{e^{\tau^2} - 1}$, and the intercept on the sample-mean axis is an estimate of $\tau$, the threshold parameter (cf. § 6.27). Kleczkowski used the estimate of $\tau$ so obtained to make the transformation $y = \log(x - \tau)$, where $x$ was the original count, to normalize the variate. A similar method is suggested by Davies[54] in respect of certain types of chemical test.

10.7. Household Size

The distribution of households by the numbers of resident persons is also a discrete distribution which in the opinion of the authors, who have examined numerous British data, can be approximated by the lognormal. In this case households with no resident persons are usually excluded by definition, but if the age range of persons is restricted, say to persons under 21 years, a discrete distribution is obtained with a finite proportion of zeros, comparable with those considered by Thomson. A similar distribution of French families by the number of surviving children is given by Gibrat[88].

The problem of zero values can be overcome pragmatically by assuming a value of say $-\frac{1}{2}$ or $-1$ for $\tau$, but an interesting suggestion has been made to the authors by Mr J. L. Williams. Assume a continuous two-parameter variate which is only manifested at the discrete integral points $0, 1, 2, \ldots$; if these points are regarded as the geometric means of successive class-intervals, the interval boundaries will be $0, a_1, a_2, a_3, \ldots$, and it is supposed that all the values of the variate lying between $0$ and $a_1$ will appear as $0$, between $a_1$ and $a_2$ as $1$, and so on.

It follows that $a_2 = \frac{1}{a_1}, a_3 = \frac{4}{a_1}, a_4 = \frac{9}{4a_1}$, etc., and it can be shown that, for consistency, $a_1$ is uniquely determined as equal to $2/\pi$. The proportion of zero values, in particular, is then given by

$$P[x = 0] = A\left(\frac{2}{\pi} | \mu, \sigma^2\right);$$

and the class limits so derived may be used for plotting the data on probability paper. This method is of course usable only when $\tau$ is zero; the data given by Gibrat imply a value for $\tau$ of approximately $-1.75$.

10.8. Physical and Industrial Processes

A number of examples of the distribution has been recorded in physical and industrial processes. Moroney[146] gives data on the loss angles for electrical condensers; Delaporte[56] discusses the measurable properties of iron tubes cast in mouldings; and Brownlee[28] has published data on the distributions of throughputs before failure, measured in thousands of tons, of a piece of acid plant. Day[155] states that the results of endurance
tests of many kinds (measured in terms of the effective length of life of a material or piece of equipment) are frequently lognormal, and cites the example of ball-bearing greases. The measurement of sound in decibels, referred to by the same writer, is a natural logarithmic measure and usually leads to normality. Tippett[186] also mentions examples drawn from textile research; and Moshman[147] uses the distribution to derive critical values which could be used for quality control methods, for example in grinding processes.

10.9. Astronomy

In astronomy there are references to the distributions of stars by Seidel[177] Seeliger[176], Charlier[35,37] and Wicksell[204].

10.10. Ages at First Marriage: A Difficulty with the Threshold Parameter

The studies by Wicksell[202] and by Nydell[150] of the distribution of the ages of men and women at their first marriage raise an interesting problem with regard to the threshold parameter. From the appearance of the distributions, and from a priori reasoning, the value of this parameter is greater than zero; but there is no absolute minimum age of marriage with which it can be identified, with the consequence that its value may vary in the population considered. It seems likely to the authors that the same contention can be made in the case of human body weights. A practical difficulty ensuing from this is that estimation methods based on the assumption of a single, fixed $\tau$, rely heavily on the smallest observed variate value; which is unfortunate if the sample of human beings includes a few dwarfs in the case of body weight studies, or a minority social group which practises very early marriage, in the case of Nydell’s application. In principle, of course, heterogeneity of this kind is liable to undermine all estimation procedures, but the estimation of the parameters of a three-parameter lognormal distribution is particularly sensitive to this difficulty.

10.11. Philology: Classification Problems: Time as a Logarithmic Stimulus

We conclude this chapter with a few references to some less obvious uses for the distribution. Williams[209] has used it in an analysis of the numbers of words in sample sentences written by Chesterton, Wells and Shaw, and found that these three authors differ in respect of both the parameters of location and dispersion. In a discussion of probability problems in philology Wake[198], at the suggestion of Williams, gave examples of probit analysis applied to the frequencies with which individual authors used nouns: the suggestion is that such methods may be used to determine authorship or to place literary works in order of date. There is no
need here to refer to the better-known applications of probit analysis or to record tolerance distributions which are approximately lognormal. These are discussed in Chapters 7 and 12.

An extension of the concept of abundance of biological species may be made to other fields where a classification is made on some homogeneity principle, as noted at the end of §3.6. Examples which are well described by the lognormal frequency curve are the distribution of the radical component of Chinese characters, using the number of characters containing the radical as the variate value; and the distribution of consumer goods, using the average expenditure on the good by consumers during a given period as the variate.

This section must also provide a niche for Lehfeldt[135], who in 1916 used what was effectively a quantitative probit analysis to describe the movement of demographic and economic variables over relatively long time periods (the logarithm of time being the dosage). From a series of population figures for England and Wales beginning at 5.5 millions in 1600 and ending at 32.2 millions in 1900 he predicted a population of 43.5 millions in 1960 and an ultimate maximum (to be reached about 2000) of 46.3 millions (the actual figure for June 1954 was 44.3 millions). Many workers in bio-assay have of course used time as a stimulus in connexion with the application of drugs, for example, Bliss[22] and Withell[211].
CHAPTER II
THE DISTRIBUTION OF INCOMES

Ford. I have a bag of money here troubles me.

The Merry Wives of Windsor

11.1. Conceptual Problems in Describing the Size-distribution of Incomes

Of all the size distributions of economic variables those relating to personal incomes have attracted by far the greatest attention. For the student of welfare economics the final distribution of the command over the goods and services produced by a society is of crucial importance; and the relative constancy of form of the distribution in different periods and countries has been the subject of argument since Pareto[152] first stated his general law. On a more practical level, producers of consumers' goods must study the distribution in order to estimate the probable extent of their markets, whilst administrators are confined by its form when planning the magnitude and structure of taxation. Yet this very multiplicity of ends which data on income distribution may serve has combined with certain natural difficulties to hinder agreement on a unifying core of concepts, principles and methods of analysis.

Thus from the point of view of data collection there is the problem of definition of the recipient unit, which may be the individual person, the one or more persons jointly assessed for liability to tax, or the family; and again, there is the problem of the definition of income itself, which may include current money income, wages in kind, self-produced goods, the imputed benefits obtained from the direct use of property, and goods distributed freely by the public authorities. From the point of view of analysis there is the unresolved problem of the mathematical description of the distribution, or, alternatively, of establishing some measure of the degree of inequality of income distribution without specifying a particular mathematical description.

It is beyond the scope of this chapter to treat these important problems comprehensively; our more limited objective is to discuss the lognormal distribution as a candidate for the mathematical description of given income data and its implications in relation to the measurement of the inequality of income. We include in the chapter examples of data which are adequately described by the lognormal hypothesis and cases where the fitted curve systematically departs from the observations. In short we try to assess the strength and limitations of the distribution as a tool of income analysis.

Many of the points made in the following sections have been made earlier by a number of writers and it will not always be easy to give explicit references, especially when arguments have been combined,
modified or given different emphasis. We therefore take this opportunity of acknowledging the efforts of our numerous predecessors in this field, in particular of Galton [178] and McAlister [142], Kapteyn [118], Amoroso [8], Gibrat [87, 88], D’Addario [1, 2, 3], Divisia [58], Darmois [48], Fréchet [73, 74], Marschak [144], Castellano [31], Quensel [169], Giaccardi [85, 86], Van der Vijk [205], Nicholson [148], Vining [196, 197] and Roy [173, 174, 175].

11.2. Criteria for Statistical Descriptions

Suppose then that the variable $x$ is defined as a certain measure of the income accruing to each of a number of individuals in a given population. A statistical description of the distribution of income is provided by $F(x)$, the distribution function of $x$, which will involve parameters that usually must be estimated from data. The choice of a particular mathematical form for $F(x)$ may be governed by one or more of the following four criteria:

(i) the extent to which the form of the function can be derived from realistic elementary assumptions;
(ii) the facility with which the function can be handled in analysis;
(iii) the economic significance that can be attached to its parameters;
and
(iv) the degree to which the fitted function approximates the data.

There is no reason to expect that any one mathematical form will be found superior to all others in respect of all four criteria; in particular, there may well be a conflict between the second and fourth. In many practical cases the overriding requirement may be that the function should adequately graduate the data, perhaps over a given portion of the range only. In these cases the choice may be made safely on empirical grounds and may depend on which part of the range is relevant. We recall here the finding of Quensel [169] that the lognormal curve is the better approximation in the lower range of incomes, whilst the Pareto curve is better in the higher range.

On the other hand, if the function is required primarily as a tool in a more complex analysis, departure of the fitted curve from the observations will not be important unless it is sufficient to bias the final conclusions. Here facility of manipulation may be decisive.

11.3. Models of Generation of Income Distributions

The first criterion relates to models of generation. The authors have previously [6] discussed models which lead to the Pareto distribution and compared these with models leading to the lognormal distribution. We shall not repeat the discussion here, but content ourselves with a brief discussion of the lognormal case. First there is the model whose essential features are due to Champernowne [32], although the development by this writer led to Pareto-type distributions. The model depends on the subdivision of income into discrete ranges and the specification of a matrix of transition probabilities whose typical element $(ij)$ states the
probability that an income recipient whose income at time $t$ lies in the $i$th size range will have income in the $j$th range at time $t+1$.

Given appropriate assumptions about the nature of the transition matrix, the equilibrium distribution of income to which any initial distribution may tend can be studied and conclusions reached concerning its form. Champernowne, in the paper cited, showed that if the lengths of the successive income ranges are in geometric progression, and if the transition probabilities $P_1(\cdot j)$ depend only on $t$ and on $j-i$ (that is, on the proportionate distance moved irrespective of the starting point), the equilibrium distribution may tend to that given by Pareto's law. These assumptions are similar in character to the law of proportionate effect discussed in §3.3, and if the assumption of discrete income ranges is replaced by one of continuity, the resulting model is formally identical with the breakage process of Kolmogoroff described in §3.6. For, if the probability that a person with income in the interval $(x_i, x_i + dx_i)$ at time $t$ will have income in the interval $(x_{i+1}, x_{i+1} + dx_{i+1})$ by time $t+1$ is denoted by $dG_i(x_{i+1}, x_i)$, the basic postulate of proportionate effect asserts that $dG_i(x_{i+1}, x_i)$ depends only on $t$ and on the ratio $x_{i+1}/x_i$; thus we may write:

$$dG_i(x_{i+1}, x_i) = dH_i(x_{i+1})/x_i,$$

and the transition equation becomes

$$dF_{i+1}(x_{i+1}) = \int_0^{\infty} dH_i(x_{i+1})/x_i dF_i(x_i),$$

where $F_i(x_i)$ is the distribution function of $x$ at time $t$.

Equation (11.2) is identical in form with the breakage equation (3.15) so that, from the argument given in Chapter 3, the equilibrium distribution tends to lognormality. The model is therefore seen to be included in the general case considered by Kapteyn, which is characterized by Theorem 3.1.

The weakness of considering the generation of an equilibrium income distribution as a time process involving transition probabilities is that, in general, the variance ($\sigma^2$) of the final distribution increases as the process is continued, apparently in contradiction to the material evidence. It may be, of course, that this is a genuine underlying tendency which is frustrated only by counteracting policies of governments and of the negotiating parties involved in income determination.

In support of this argument it may be said that there is consistent evidence that in a number of professions (both in the United States and in this country) the variance of the income distribution increases systematically with the age of the professions' members. So that the earnings of an individual person through life may well be described by a stochastic process of the form $x_{i+1} = \exp[f(t) + u_i]x_i$, where the function $f(t)$ is chosen to describe the path of the median income through life and $u_i$ is $N(0, \sigma_i^2)$ and independent of $t$. For doctors in general urban practices in Great Britain in 1936-8, for example, $\sigma_i^2$ was of the order of 0.01, resulting in an increase in the variance of the distribution of doctors'
earnings from 0.2 in the age-group 25-29 years to 0.5 in the age-group 65-69 years. These estimates are derived from data published in the 1946 Report of the Inter-departmental Committee on the Remuneration of General Practitioners (Cmd. 6810). In this case the stability of the complete distribution of professional earnings must depend on the assumption that a stream of new entrants is constantly entering the initial distributions with relatively small variances, to replace older members who are leaving, through death, retirement and other causes, the distributions with greater variances later in life.

On the other hand, there is the alternative formulation of Kalecki (114), described in §3.5, which constrains the value of $\sigma^2$ to remain constant; and again we would refer the reader to our own discussion, later in the same section, of a process which, conceptually, occurs without lapse of time. A similar approach has been studied in detail in two papers by Roy (173, 174), who started from the assumption that workers' earnings are related to their output of a commodity.

Roy studied the distribution of outputs by workers in a number of sample occupations in which output could be measured simply, and concluded that the evidence was on the whole favourable to the lognormal hypothesis. His tentative explanation of this result was in terms of the multiplicative central-limit theorem, being based on the proposition that the output of an individual depends on a great number of different factors which may conveniently be considered to act together in a multiplicative rather than in an additive way; though he drew attention to the possibility of the introduction of bias if the elementary factors are not independent of each other.

Passing from these alternative formulations of the law of proportionate effect, we refer briefly to an extension of the law which seems to be of some practical importance. Suppose that the law holds for income earners in each of a number of sectors of the economy and the distribution of incomes in each sector is consequently lognormal, say $\Lambda(\mu, \sigma^2)$, or $\Lambda(\log x - \frac{1}{2} \sigma^2, \sigma^2)$, where $\alpha$ is the arithmetic mean income of the sector. Then, under assumptions for which there is some empirical evidence, the overall distribution of income in all sectors is also lognormal.

The assumptions are:
(i) $\sigma^2$ is constant for all sectors,
(ii) the number of sectors is large enough for the distribution of $\alpha$ to approximate to a continuous distribution, and this distribution is lognormal, say $\Lambda(\mu_0, \sigma_0^2)$.

Then if $F(x)$ denotes the distribution function for the whole population we have immediately

$$F(x) = \int_0^\infty \Lambda(x \mid \log \alpha - \frac{1}{2} \sigma^2, \sigma^2) \ d\Lambda(\alpha \mid \mu_0, \sigma_0^2)$$

$$= \Lambda(x \mid \mu_0 - \frac{1}{2} \sigma^2, \sigma_0^2 + \sigma^2) \quad \text{from Corollary 2.2b.} \quad (11.3)$$

The evidence for these assumptions is given in the earlier paper by the authors (6); the value of $\sigma_0^2$ for industrial wage earners in Great
Britain in 1935 and 1948 was found to be approximately 0.04, or small in relation to $\sigma^2$ (about 0.5). The conclusion of this section, then, is that there exist a number of models of generation which lend plausibility to the assumption of lognormality as a simple description of income distributions.

11.4. STATISTICAL ANALYSIS OF DATA

The second criterion concerns ease of handling in statistical analysis. It will be convenient to remark briefly under three headings:

(i) The estimation of parameters: on this we need only say that there is a wide choice of methods for the lognormal distribution from which the statistician may choose rationally according to his need for speed and accuracy.

(ii) The comparison of two or more distributions, and more general applications of the analysis of variance: here the link that the lognormal distribution provides with normal theory is of great value and brings to the statistician the full facilities of existing normal test statistics. These properties of the distribution are discussed more fully in Chapter 8.

(iii) The introduction of the distribution of incomes into econometric models: it is often necessary in this field to investigate the consequences of averaging behaviouristic relationships over the distribution of incomes. Here the lognormal hypothesis seems to have considerable advantages over most other candidates. This point is taken up more fully in Chapter 12.

11.5. INTERPRETATION OF THE PARAMETERS OF THE LOGNORMAL DISTRIBUTION

Thirdly, there is the interpretation of the parameters of a lognormal distribution of incomes. The interpretation of the location parameter $\mu$ is straightforward, since (in the two-parameter case) it is the logarithm of the geometric mean income and is also the logarithm of the median income. It is to be noted that since the arithmetic mean involves both the location and dispersion parameters it is not a pure measure of the level of incomes under the lognormal hypothesis: for this the geometric mean or median is to be preferred. The dispersion parameter $\sigma^2$ is of greater interest by virtue of its relation to the concept of concentration of incomes as defined by Lorenz[141].

In the Lorenz diagram (Fig. 11.1) the proportion of income receivers having income less than $x$ is measured along the horizontal scale and the proportion of total income accruing to the same income receivers along the vertical scale. The points plotted for the various values of $x$ trace out a curve below the 45° line sloping upwards to the right from the origin. In statistical terms the curve describes the relation between the distribution function $F(x)$ and the first-moment distribution function $R(x)$, defined by

$$R(x) = \frac{\int_0^x t dF(t)}{\int_0^\infty t dF(t)}.$$  \hspace{1cm} (11.4)
The measure of income concentration which is naturally suggested by the Lorenz diagram is the ratio of the shaded area between the Lorenz curve and the 45° line to the area of the triangle under the 45° line. The measure varies from zero, when all persons have the same income (so that the 45° line may be termed the diagonal of equal distribution), to unity, when all the available income accrues to one person.

The formal definition of the measure is

\[ L = 1 - 2 \int_0^\infty F_1(x) dF(x). \]  

(11.5)

Substituting in equation (11.5) the explicit form for \( F_1(x) \) given by Theorem 2.6 we obtain, for the lognormal hypothesis,

\[
L = 1 - 2 \int_0^\infty \Lambda(x | \mu, \sigma^2) d\Lambda(x | \mu, \sigma^2) \\
= 1 - 2 \Lambda(1 | \sigma^2, 2\sigma^2) \\
= 1 - 2 N\left(-\frac{\sigma}{\sqrt{2}}, 0, 1\right) \\
= 2 N\left(-\frac{\sigma}{\sqrt{2}}, 0, 1\right) - 1;
\]  

(11.6)
which shows that the measure of concentration $L$ is monotonically related to the value of $\sigma^2$ and is independent of $\mu$.† It will also be noted, from Theorem 2.7, that there is a strong similarity between equation (11.6) and the expression for Gini's coefficient of mean difference. In fact, denoting Gini's coefficient by $G$, we have in general that

$$G = 2\alpha L,$$

where $\alpha$, as before, is the arithmetic mean income.

It follows that the parameter $\sigma^2$ may be interpreted as a measure of the concentration of incomes in a sense which is generally acceptable; and that since the value of $\sigma^2$ may be estimated from samples within calculable confidence limits so too can Lorenz's measure of concentration.

Since many empirical data have been described and analysed by means of the Lorenz diagram it is of some interest to discuss the shape of the Lorenz curve resulting from a lognormal distribution.

First, the two-parameter case. The diagonal line drawn at right angles to the diagonal of equal distribution, and defined by the equation

$$F(x) = 1 - F_1(x),$$

cuts the Lorenz curve in this case at the point \(\{F(cz), I - I\} \) corresponding to the arithmetic mean income. For

$$1 - \Lambda_1(\alpha) = 1 - N\left(\frac{\mu + \sigma^2}{2} | \mu + \sigma^2, \sigma^2\right)$$

$$= 1 - N\left(-\frac{\sigma}{2} | 0, 1\right)$$

$$= N\left(\frac{\sigma}{2} | 0, 1\right)$$

$$= \Lambda(\alpha), \text{ from equation (5.59)}; \quad (11.8)$$

or, in words, the proportion of persons with less than the mean income is the complement of the proportion of income held by these persons.‡ It also follows from the symmetry properties of the normal distribution that the Lorenz curves in this case (for all values of $\sigma^2$) are symmetrical with respect to the diagonal defined above; and that, at the points defined by (11.8), the tangents to the curves are parallel to the diagonal of equal distribution. Also no two curves of the family can intersect. These properties furnish simple tests of the two-parameter hypothesis from the appearance of the Lorenz curves.

Lorenz curves can, however, intersect in two cases of interest. First, if the data plotted on the diagram arise from a two-parameter parent distribution, but these data are available only in truncated form, as is

† Values of $L$ tabulated against $\sigma$ are given in Appendix Table A1.
‡ This proportion is tabulated against $\sigma$ in Appendix Table A1.
often the case with figures published by revenue authorities. In this case the equations determining the Lorenz curves become

\[ F(x) = \frac{N(y) - N(v)}{1 - N(v)}, \]

(11.9)

and

\[ F_1(x) = \frac{N(y - \sigma) - N(v - \sigma)}{1 - N(v - \sigma)}, \]

(11.10)

where \( y = (\log x - \mu)/\sigma \), \( v = (\log \xi - \mu)/\sigma \) and \( \xi \) is the point of truncation. Simple results do not hold for this case, but the corresponding Lorenz curves are not symmetrical. The Lorenz curves will intersect if the parameters \( \mu \) and \( \sigma^2 \) remain unchanged, but the point of truncation changes. The Lorenz curve for a two-parameter truncated distribution is given in Fig. 11.2.

Secondly, if the data arise from a three-parameter distribution in which the third parameter \( \tau \) is a (positive) threshold below which no value of income can exist. This distribution was described in §2.7, where an expression for the first-moment distribution was given (Theorem 2.10); it was shown that, because of the simple displacement of the frequency curve, the formula for Gini’s coefficient of mean difference

\[ G = \frac{2}{

Fig. 11.2. The Lorenz diagram for a truncated two-parameter distribution of incomes. The curve bounding the shaded portion represents a distribution \( \Lambda(\mu, \sigma^2) \) truncated at \( \xi = \sigma^2 \), that is, with 16% of the distribution truncated. The Lorenz curve for the full distribution is shown by the broken line.
was the same as for the two-parameter case (Theorem 2.11). It is readily seen, however, from Theorem 2.10, that Lorenz’s measure for this case $L(\tau)$ is given by

$$L(\tau) = \frac{\alpha}{\tau + a} \left[ 2N\left(\frac{\sigma}{\sqrt{2}}\right) - 1 \right]$$

$$= \frac{\alpha}{\tau + a} L(\sigma)$$

$$< L(\sigma) \quad (\tau > 0).$$

(The Lorenz measure is not meaningful for negative values of $\tau$.)

Fig. 11.3. The Lorenz diagram for a three-parameter distribution of incomes. The curve bounding the shaded portion represents a three-parameter distribution with $\tau = 0.1$, $\mu = 0$, $\sigma^2 = 0.3$. The Lorenz curve for the two-parameter distribution with $\mu = 0$ and $\sigma^2 = 0.3$ is shown by the broken line.

The tangent parallel to the diagonal of equal distribution still occurs at the arithmetic mean income $\tau + a$, but this point is always above the diagonal $F(x) = 1 - F_1(x)$; and the locus of these points of tangency for different values of $\tau$ but constant $\sigma^2$ is a vertical straight line (Fig. 11.3). Intersection is therefore not possible for constant $\sigma^2$ but will occur if either $\sigma^2$ or if both $\sigma^2$ and $\tau$ are allowed to vary. For all $\tau > 0$ the curves are not symmetrical.
11.6. THE TEST OF PRACTICE

The last criterion is the test of practice. To what extent does the lognormal hypothesis accord with observed data? Here a few general remarks are necessary. For the reasons we adduced at the beginning of this chapter the elementary forces which, if left to themselves, might result in a particular equilibrium distribution of incomes, or even in a divergent process, are rarely left to work their influence unheeded. In particular it is well known that the redistribution of incomes is often an important objective of governments when determining their taxation structure. Although these measures bear directly on the post-tax incomes, there is no doubt that they also influence the pattern of gross incomes. The measures are often sharply discontinuous and have uneven effects on the final distribution; and when the level of taxation is high and progressive, the incentive to avoid the very high rates leads to a certain arbitrariness in the definition of income received.

Again, so far as published data are concerned, it is known that these have often been smoothed, or even partly estimated, on the basis of a Pareto hypothesis, so that they cannot furnish an independent statistical test. Finally, when all the incomes in a contemporary community are considered, it cannot be overlooked that certain broad categories of income—wage incomes, property incomes, transfer incomes and so on—are generated in fundamentally different ways, so that it is doubtful whether a single model can comprehend them all. For these and many other reasons it is unlikely that actual income distributions will be as well described by any formulation which can be traced back to a simple random process as, for example, are the size distributions of small particles found in sedimentary petrology.

Perhaps the most careful and complete study of income distributions published in recent years is that of the Office of Business Economics of the United States Department of Commerce[1891, in which data from income-tax returns and field surveys have been integrated. The study covers the four years 1944, 1945, 1947 and 1950, and the tables include distributions of the recipient units and of aggregate income (the first-moment distribution) for families, consumer units, farm and non-farm families, and so on. The distributions for different years are very similar and we have chosen those shown in Figs. 11.4 and 11.5 as typical.

The distributions for 'all consumer units' all show the same systematic divergence from the lognormal curve. There are, by comparison with the numbers predicted by the latter, too many consumer units in the lowest class (less than $1000 per annum) and a less marked tendency for too many to be in the very highest class (above $20,000 per annum). Thus the higher incomes would probably be more accurately graduated by a Pareto-type curve.

The high proportion of consumer units near to the bottom of the income scale was in fact noted by the authors of the report, who were not using any mathematical hypothesis as a criterion. Reasons they put
forward for this phenomenon included: the inclusion in the lowest group of a number of part-year earnings (by persons first entering the labour market and by newly married couples being reckoned as independent units for the first time) which were "not representative of an actual command over goods and services over the full year period covered by the size distribution statistics"; the occurrence in the lowest group of a high proportion of unattached individuals as opposed to families, with the implication that these form a separate population; the inclusion of retired persons whose current incomes are regarded as a supplement only to planned drawings of accumulated savings; and the undervaluation of some incomes in the sense that the low incomes represent a greater command over goods and services than is apparent (one factor here is the valuation of farm produce consumed on the farm at farm, rather than retail, prices).†

Some insight into the type of heterogeneity present is provided by the separate distributions for non-farm families, farm operator families and unattached individuals in 1947 shown in Fig. 11.5. The unattached individuals in particular appear very heterogeneous, and the measures of location for the three classes differ considerably. As a description then of the published figures the lognormal distribution is deficient, although it would probably be less so if the true distribution of current command

† A possible method of analysis which is suggested by these considerations is to attempt to resolve the population into two or more overlapping lognormal populations. A similar problem in biology, is discussed by Harding[99] who uses normal probability paper to effect the resolution.
over goods and services could be computed. On the other hand the systematic discrepancies seem stable from year to year and the use of measures of location and concentration based on the lognormal hypothesis, or the integration of the hypothesis into econometric models, would not seriously mislead. In particular, it seems safe to conclude from Fig. 11.4 that the degree of concentration has not changed between 1944 and 1950.

The evidence studied by the authors suggests that the more homogeneous the group of income recipients is, the more likely is the lognormal curve to yield a good description of the income distribution; this is again more nearly true if the income is derived from a single source (if for example it consists entirely of earnings from employment). In this context we may quote the studies of Roy[173,174] referred to in §11.3, and a number of samples of wage and salary distributions published earlier by Gibrat[88]. An interesting series of earnings distributions in British agricultural occupation for the year 1950 has recently been made available to the authors by the Ministry of Agriculture and Fisheries;† in which the earnings of regular workers only were recorded, thus avoiding difficulties due to the presence of part-year earnings. Their logarithmic probability graphs are set out in Fig. 11.6.

These data were referred to earlier, in §3.5, where it was pointed out that the existence of a national minimum wage had led to the practice, in many agricultural occupations, of wage negotiations being conducted

† Similar data are given by Palca and Davies[151].
in terms of the 'premium', that is, the excess of the contract wage over the national minimum. During the calendar year 1950 the average value of the minimum wage was 94½d. per week, and for earnings in six out of the nine male occupations portrayed in Fig. 11.6 the three-parameter distribution, with a threshold value of about 90s., is empirically appropriate. For three of the occupations, however, the two-parameter distribution (with zero threshold) is a distinctly superior description, which suggests that here the national minimum wage had but a formal significance.

11.7. SUMMARY AND FURTHER SUGGESTIONS

If the arguments of the four preceding sections of this chapter are taken together they provide grounds for considering the lognormal family of curves as a strong candidate whenever a statistical description of...
income size distribution is required. This candidacy is strengthened when
the description must comprise incomes less than, as well as greater than,
the modal income. Of all skew, unimodal distributions the lognormal is
the easiest to manipulate in the present state of statistical theory; and
the skewness of income distributions is the characteristic which it is
least easy to disregard. Even where, as with the American data of
Figs. 11.4 and 11.5, the distribution provides a deficient description,
it may be useful in the sense of providing a sort of null hypothesis, by
which the data may be compared with the consequences of the assump-
tion of an elementary random process. Thus systematic divergence from
the fitted lognormal curve may suggest the existence of interesting
heterogeneities or peculiarities in the methods of imputing the income
values.

To some investigators the possibility of further elaborating possible
models of generation may be of more interest. To discuss possible develop-
ments would take us beyond our present boundaries, but we might
perhaps suggest accepting, provisionally, the existence of two and three-
parameter distributions for the incomes of regular workers in narrowly
defined occupations, as those depicted in Fig. 11.6, and investigating the
consequences, first of allowing individual workers to enter and leave the
occupations by a kind of birth and death process, and secondly of com-
pounding the distributions in a number of different ways, in particular
with varying values of the threshold parameter.
CHAPTER 12
APPLICATIONS OF LOGNORMAL THEORY IN THE ANALYSIS OF CONSUMERS' BEHAVIOUR

Countess. It must be an answer of most monstrous size, that must fit all demands.

All's Well That Ends Well

12.1. INTRODUCTION

In the last two chapters we have considered the lognormal distribution as an hypothesis for the size distribution of a number of economic variables. In the present chapter we suggest another class of uses for the hypothesis which may be of interest to the applied economist. Here we are concerned with the construction of mathematical models designed to describe relationships between two or more economic quantities; in particular the quantities of commodities purchased, their prices and the incomes of the purchasers. First we show briefly how the lognormal hypothesis may assist in the transition from micro-models to macro-models; and then, in more detail, we apply the theory of Chapter 7 to the general problem of market demand. We begin with a simple model which describes the distribution of purchases of an indivisible commodity among consumers with varying incomes; we then extend the theory in stages in order to cover the characteristic problems posed by family budget and time series data. In order to avoid confusion of the argument we have omitted all discussion of statistical estimation procedures; where necessary the appropriate reference is given to sections of Chapter 7. In general, it is assumed that the reader is familiar with the broad outlines of the theory of consumers' behaviour. For discussions of this the reader is referred to the works of Stone[181] and Wold[213], and for its application to the analysis of family budgets to the preceding monograph in this series by Prais and Houthakker[164]. The approach to time series adopted here in §12.13, which attempts to integrate time series and family budget data, is based on that of Stone; and the general framework of the family budget theory, in particular the approach to the problem of household composition, owes much to the work of Prais and Houthakker.

12.2. THE TRANSITION FROM MICRO-MODELS TO MACRO-MODELS

A micro-model in econometrics is a mathematical model which purports to describe some aspect of economic behaviour at the level of the individual consuming or producing unit, the family or the firm; whereas a macro-model is designed to analyse the composite behaviour of groups
of these units, the nation or the industry, within which the individual units are assumed to remain, to some degree at least, autonomous. It is usual, when undertaking an analysis of observations relating to large economic groups, to begin by assuming the relations which determine a micro-model (for it is thought that here economic intuition is of most use) and to derive the operational macro-model by the method of aggregation.

As an illustration we take a recent example cited by Prais and Houthakker[164], who derive a macro-model describing the relation between the total consumers' expenditure on a commodity and the total of consumers' expenditure on all commodities. They first assume the following micro-equation:

$$v_{ir} = \alpha_i + \beta_i \log v_{or},$$  \hspace{1cm} (12.1)

where $v_{ir}$ is the money spent by the $r$th consumer on the $i$th commodity, $v_{or}$ is the total expenditure of the same consumer on all commodities, and $\alpha_i$ and $\beta_i$ are behaviouristic parameters, assumed constant for all consumers.

The macro-equation is obtained by adding equations (12.1) over $n$ consumers each with a different value of $v_{or}$ and dividing through by $n$:

$$\bar{v}_{ir} = \frac{1}{n} \sum v_{ir} = \alpha_i + \beta_i \left( \frac{1}{n} \sum \log v_{or} \right).$$  \hspace{1cm} (12.2)

Equation (12.2) cannot be used directly in an analysis of national expenditure data since these do not include estimates of the determining variable $\left( \frac{1}{n} \sum \log v_{or} \right)$; it must therefore be rewritten in terms of

$$\bar{v}_o = \frac{1}{n} \sum v_{or},$$  \hspace{1cm} (12.3)

the total national expenditure per person on consumers' goods and services. By assuming that $v_{or}$ is lognormally distributed over consumers we obtain the required equation:

$$\bar{v}_o = \left\{ \alpha_i - \beta_i \frac{\sigma^2}{2} \right\} + \beta_i \log \bar{v}_o$$ \hspace{1cm} (12.4)

where $\sigma^2$ is the variance parameter of $v_{or}$. In this way we can construct the parameters of the macro-equation from estimates of the micro-parameters; and we may note that changes in the numerical value of $\sigma^2$ (reflecting changes in the concentration of incomes) will cause vertical displacements of the market curve (12.4).

A more elaborate use of the lognormal hypothesis has been made by Klein[127] in the second of two models constructed to analyse savings behaviour. Here aggregation occurs over a number of variables, of which income is one, and the process must take account of their joint distribution. For simplicity we show Klein's argument in respect of

\[ ^{\dagger} \text{If consumption } v_{ir} \text{ is to be non-negative, aggregation must be confined to incomes in the range } (c^{-\beta}, \infty) \text{ and (12.4) replaced by expressions derived from the truncated distribution of } \sigma. \]
income and one other variable only, the number of people in the household. The micro-equation is

\[ \frac{s_{it}}{y_{it}} = \alpha_0 + \alpha_1 \log y_{it} + \alpha_2 \log z_{it} + u_{it} \]  

(12.5)

where \( s_{it} \) = savings of the \( i \)th household in year \( t \),
\( y_{it} \) = income of the \( i \)th household in year \( t \),
\( z_{it} \) = number of persons in the \( i \)th household in year \( t \),
\( u_{it} \) = a variate representing random differences in the behaviour of individual households, and
\( \alpha_0, \alpha_1 \) and \( \alpha_2 \) are behaviour parameters.

Writing bars over the symbols to indicate their average value in the population and in particular using the notation

\[ \bar{a}_i \bar{b}_i = \frac{1}{n} \sum_i a_{it} b_{it} \]  

(12.6)

the corresponding macro-equation is

\[ \bar{s}_i = \alpha_0 \bar{y}_i + \alpha_1 \bar{y}_i \log \bar{y}_i + \alpha_2 \bar{y}_i \log \bar{z}_i + \bar{y}_i \bar{u}_i \]  

(12.7)

assuming independence between \( y \) and \( u \).

Now, using the lognormal hypothesis for \( y_{it} \), and noting that, from the definition of the simple correlation coefficient,

\[ \rho_{y_u} = \frac{\sigma_y \sigma_u}{\sigma_y \sigma_u} \]  

(12.8)

Klein derives a final equation of the form

\[ \bar{s}_i = (\alpha_0 \bar{y}_i \log \sigma_y \sigma_u) + \left( \alpha_0 + \alpha_1 \frac{\sigma_y^2}{2} \right) \bar{y}_i + \alpha_1 \bar{y}_i \log \bar{y}_i + \alpha_2 \bar{y}_i \log \bar{z}_i + \bar{y}_i \bar{u}_i, \]  

(12.9)

where \( \sigma_y \) is the standard deviation of \( y_{it} \) and \( \sigma \) the standard deviation of \( \log y_{it} \). We may comment that, though Klein does not explicitly make the assumption, the variable \( z_{it} \) is in fact approximately lognormal if measured on some scale of consumer units in which fractions are permitted; also that the aggregation would have proceeded more smoothly, avoiding the inelegant combination of \( \bar{y}_i \) and \( \log \bar{y}_i \) in (12.9), had the micro-equation been written

\[ \frac{s_{it}}{y_{it}} = \alpha_0 \bar{y}_i \bar{z}_i \bar{u}_i, \]  

(12.10)

the aggregation of which may be left to the reader.

These two examples show how the lognormal hypothesis for the distribution of some variable may be used to decide the manner in which aggregation modifies the initial micro-equation; and they may perhaps serve as an introduction to the more general problem of discovering the econometric laws which are applicable to statistical populations rather than to individual entities. In the sections which follow the aggregation process itself will be found to be of crucial importance, dominating the
form of the final statistical relationships. With this shift of emphasis it becomes possible to study the problems of demand on the basis of very weak assumptions as to the behaviour of individual consumers; for these are difficult to verify and in any case of little interest in themselves for the economist.

12.3. The Primacy of the Engel Curve in Empirical Work

The study of consumer demand divides conveniently into two parts: the study of the relation between the demand for a commodity and the consumer's income; and that of the relation between demand and the prices of consumers' goods. The former is now generally known as the study of the Engel curve, after the German economist Engel[60], who made a number of important empirical generalizations from collections of family budgets. The development of the quantitative study of the Engel curve owes its present high level to the fact that, by means of budget inquiries, economists have been able to observe a large number of consumers over a range of incomes sufficient to determine the main characteristics of the curve with some accuracy. And since, in principle at least, the observations can be taken at a single point of time, we are largely justified in applying the principle of ceteris paribus, which allows us to treat the relationship in isolation from the many other factors known to influence demand. When we consider the relation between demand and prices, however, we find a less happy situation: for the same set of prices holds at any one time for the whole community, thus ruling out the possibility of drawing inferences from an instantaneous sample of budgets; in general, price movements through time are not independent of income movements nor of each other; and the range of variation during periods in which it is safe to assume that preferences are unchanged is usually too small to investigate the characteristic reactions of consumers at all closely.

It seems then reasonable to proceed to a full specification of the demand curve by two stages: in the first selecting a form of Engel curve which seems consistent with the numerous budget data and whose characteristics conform to the demands of pure theory, and in the second considering the way in which price variations may be expected to affect the numerical values of the parameters of the Engel curves. We shall begin with what seems a conveniently simple case: the study of the demand for an indivisible good.

12.4. The Derivation of an Engel Curve for an Indivisible Good

The Engel curve, defined as a curve to be used in the analysis of budget data, is essentially a statistical concept: that is, its intent is to predict the average behaviour of a number of consumers having the same income and
faced with the same set of prices; and a necessary implication of the concept is that the preferences of the individual consumers for the commodity considered conform to some law of frequency. Consider the case of an indivisible good, say a television set, such that no individual would normally possess more than one in a given period of time. Given the price of the good (we assume for simplicity there is only one retail price) and the prices of all substitutes and complementary goods, we postulate that each consumer will reach a conclusion as to the amount of income he must have before he makes the decision to purchase. The amount of income which is just sufficient to bring him to the brink of the decision we will call his tolerance income, by analogy with the similar phenomena discussed in Chapter 7. If now a large number of consumers living in similar social environment are considered it may be supposed that their tolerance incomes, although varying, will cluster round some central value and in fact will conform approximately to some regular unimodal frequency distribution. This distribution is likely to be positively skewed since the tolerance income cannot take on negative values but is effectively unbounded in the positive range. If the tolerance income can be regarded as the product of a large number of chance influences, that is, the outcome of a generation process such as was described in Chapter 3, a simple lognormal hypothesis is an appropriate elementary assumption for its distribution. The probability that an individual consumer will be found to possess the good is then

\[ P[q = 1 \mid y] = P[x \leq y] = \Lambda(y \mid \mu, \sigma^2), \]  

where the notation \( q = 1 \) denotes possession, \( y \) is the consumer's income and \( x \) the tolerance variate with parameters \( \mu, \sigma^2 \).

This probability may be estimated by the proportion of consumers with a given income found in a cross-section study to possess the good; and the graph of the expected proportion plotted against income we may call the pseudo-Engel curve for the good. The appropriate procedure for estimating \( \mu \) and \( \sigma^2 \) in (12.11) from observations on \( y \) and \( p \), the estimate of \( P \), has been described in §7.4. This type of model has been used by Farrell [63] to describe consumer demand for motor cars in the United States. Farrell draws a distinction between the relation of income to demand or ownership for an individual family and the similar relation averaged over a number of families with the same income: the latter he terms the 'budget function', which is synonymous with our term 'Engel curve' as defined in the first sentence of this section. In a second paper Farrell [62], following a suggestion of Tobin, discusses aggregation over an assumed random (and more particularly normal or lognormal) distribution of individual consumers' preferences in more detail. The present discussion is parallel with Farrell's, though Farrell did not extend his theory to the case of the divisible good which we consider next.
12.5. The Extension of the Theory to a Divisible Good

The theory we have given for an indivisible good ignores many complications which could not be overlooked in serious empirical research. The influence of prices, the fact that the life of the good may extend over several time periods, and the range of qualities in which the good is produced for the market are some of these difficulties. But rather than attempt to deal with these we prefer now to pass to the more general problem of the divisible good. The essence of the earlier theory may easily be extended to meet this case. For suppose that expenditure on the divisible commodity may be divided into elementary units, say pennyworths. Then, as with the indivisible good, the consumer is imagined to represent his preferences in a given price situation by a number of points on the income scale at which he considers it appropriate to spend the first, second, and so on up to the \( k \)th penny on the commodity. This is represented schematically in Fig. 12.1. In the figure the positions of the 5th, 6th and 7th pennyworths are shown by heavy lines, and the remainder by broken lines; the consumer's income is supposed to be in the region of 175 units and it is only necessary for the purpose of the theory to suppose that his purchasing decisions are tolerably distinct in this region. For each consumer we then have what we may term a tolerance scheme for increments of expenditure on the commodity. The sum of the increments from 1 to \( k \) represent the consumer's saturation expenditure to which his actual expenditure would eventually expand, if his income increased without limit whilst his tolerance scheme remained unchanged.

At any finite income the consumer's actual expenditure on the commodity will be the sum \( q_r \) of the increments \( \Delta q_{ir} \) (the subscript \( i \) denotes the rank of the increment and the subscript \( r \) denotes the consumer), whole tolerance values \( x_{ir} \) are less than \( y_r \), his income:

\[
q_r = \sum_{x_{ir} \leq y_r} \Delta q_{ir}. \tag{12.12}
\]

Considering now a community of consumers with similar tolerance schemes but having incomes extending over a wide range, a community tolerance scheme can be constructed by putting together the individual schemes. The tolerance values for the community will be more densely packed than those of any one individual (and many may coincide) so that the whole will tend to a regular frequency distribution as in Fig. 12.2.

If any given income \( y \) is considered, the average expenditure per person
of those consumers having income $y$ may be obtained by adding equation (12.12) over all consumers and dividing by their number:

$$q = \frac{1}{n} \sum_{r} q_r = \frac{1}{n} \sum_{x_r \geq y} \Delta q_{ir}$$  \hspace{1cm} (12.13)

and the community saturation expenditure per person is the total area under the frequency curve divided by the number of consumers:

$$\bar{x} = \frac{1}{n} \sum_{r} \kappa_r$$  \hspace{1cm} (12.14)

By the assumption that the frequency curve of Fig. 12.2 is approximately described by the simple lognormal frequency function and has area $\bar{x}$, the equation for the Engel curve becomes

$$\bar{q} = \bar{x} \Lambda (\mu, \sigma^2),$$  \hspace{1cm} (12.15)

where $\mu, \sigma^2$ are the parameters of the tolerance scheme.

The frequency curve depicted in Fig. 12.2 would perhaps be more familiar to the economist if it were entitled 'the marginal propensity to spend on an individual good in relation to income'; for such it is (for the community) if the increments $\Delta q_{ir}$ are made indefinitely small and expressed as proportions of indefinitely small increments on the income scale. For an illustrative example computed from observed data we show in Fig. 12.3 the graphs of the marginal propensities to spend on (or the tolerance schemes for) six commodity groups which together exhaust total expenditure for industrial working-class households in 1937–8. The computations† are described more fully in an earlier paper by the present authors (Aitchison & Brown 1951).

### 12.6. The Problem of Inferior Goods

The existence of inferior goods (that is, goods on which expenditure declines as income increases) is sufficiently well established by empirical studies and cannot be ignored in any theoretical scheme. Since at zero income all expenditure must be zero, no good can be inferior below a

† The unknown parameters of (12.15) are $\bar{x}, \mu, \sigma^2$ which are to be estimated from pairs of observations of $\bar{q}, x$. Estimation procedures are discussed fully in §§7.7-7.9.
certain positive level of income however small this may be. Thus the existence of inferior goods implies that, after the $k$th positive increment of expenditure is reached in the individual tolerance scheme, further tolerance values exist corresponding to decrements of expenditure; and these, for consistency, will be not greater than $\kappa$ in number. This situation is depicted in Fig. 12.4.

In the case of goods which are generally regarded as inferior the positive increments are concentrated in a narrow range, possibly below the range of observed incomes, and their exact tolerance values will be vague or even unknown to the majority of consumers. If the positive range can be ignored completely we find† after aggregation that the form of the Engel curve becomes

$$
\bar{y} = \kappa - \kappa \Lambda(y | \mu, \sigma^2)
$$

$$
= \kappa \{1 - \Lambda(y | \mu, \sigma^2)\}
$$

$$
= \kappa \Lambda(\frac{1}{y} - \mu, \sigma^2).
$$

† Assuming $\kappa' = \kappa$. 

---

Fig. 12.3. Marginal propensities to spend on six commodity groups: British industrial working-class households, 1937-8.

Fig. 12.4. Individual tolerance scheme for expenditure on an inferior good.

Fig. 12.3. Marginal propensities to spend on six commodity groups: British industrial working-class households, 1937-8.
In general, provided a reasonable range of income separates the distribution of positive increments from that of the negative increments, a pair of Engel curves of the form (12.15) and (12.16) will probably be adequate to describe the community behaviour. If, however, an appreciable number of consumers are already decreasing their expenditure while many others have not reached their saturation levels the Engel curve will take on a more complex form. If this type of situation is met, however, we suggest that the community would be better divided into two social groups before analysis.

12.7. The Consolidation of Commodities and of Groups of Consumers

It has been noticed in empirical work carried out by the writers that Engel curves of the type (12.15) hold approximately not only for 'individual' commodities such as 'meat', 'fish', or 'dairy products', but also for composite commodities such as 'all food'.

We have also noted that the same holds true when sub-groups of consumers (families of a given composition, occupational groups and so on) are consolidated into larger groups. This process of putting together groups of commodities, groups of consumers, or data for the same consumers for several time periods, we term consolidation in distinction from aggregation as we wish to hold the latter term in reserve for the process of averaging consumption data over the distribution of incomes.

It seems worthwhile to formalize these empirical results, since consolidation of one type or another is an important feature of most empirical work, where it is usually undertaken in order to smooth out sampling irregularities by effectively increasing the size of samples. Accordingly we present the following theorem and corollary. First, it is often convenient to write equation (12.15) in standardized form:

\[ q = \kappa \Lambda(y | \mu, \sigma^2) = \kappa \Lambda(y^{1/\sigma}e^{\mu/\sigma} | 0, 1) = \kappa \Lambda(ax^\beta), \] (12.17)

where \( \alpha = e^{-\mu/\sigma} \) and \( \beta = 1/\sigma \) and the standard parameters (0, 1) are omitted.

**Theorem 12.1**

If each of a large group of individual commodities obeys the law \( q = \kappa \Lambda(ax^\beta) \) and

(i) \( \beta \) is constant in the group,

(ii) \( \kappa \) and \( \alpha \) are jointly lognormally distributed over the commodities in the group, i.e. \( F(\kappa, \alpha) = \Lambda(\kappa, \alpha | \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho^2) \)

† It may be necessary to treat the tolerance scheme for the decrements as a three-parameter distribution: \( x \) is \( \Lambda(\tau, \mu, \sigma^2) \) (cf. §2.7).

‡ What are here termed 'individual' commodities may of course be further subdivided.
then the mean expenditure $\overline{q}$ on the commodities in the whole group obeys a law of the same form

$$ \overline{q} = \kappa' \Lambda'(z'y^\theta), $$

where $\kappa'$ is the mean value of $\kappa$ in the group.

**COROLLARY 12.1a**

The result of Theorem 12.1 holds (mutatis mutandis) for the case of a single commodity consumed by each of a large number of groups of consumers.

**Proof**

$$ q = \int_0^\infty \int_0^\infty \kappa \Lambda'(z'y^\theta) dF(\kappa, \alpha) $$

$$ = \int_0^\infty \int_0^\infty \kappa \Lambda'(z'y^\theta) d\Lambda \left( \kappa \left| \mu_1 - \rho \sigma_1 \mu_2 + \rho \sigma_1 \log \alpha, (1 - \rho^2) \sigma_2^2 \right. \right) \times d\Lambda(\alpha | \mu_2, \sigma_2^2) $$

$$ = \exp \left[ \mu_1 - \rho \sigma_1 \mu_2 + \frac{1}{2} (1 - \rho^2) \sigma_2^2 \right] \int_0^\infty \Lambda'(z'y^\theta) d\Lambda(\alpha | \mu_2, \sigma_2^2) $$

$$ = \exp \left[ \mu_1 - \rho \sigma_1 \mu_2 + \frac{1}{2} (1 - \rho^2) \sigma_2^2 + \rho \sigma_1 \mu_2 + \frac{1}{2} \left( \rho \sigma_2^2 \right)^2 \right] \times \int_0^\infty \Lambda'(z'y^\theta) d\Lambda(\alpha | \mu_2 + \rho \sigma_1, \sigma_2^2) $$

from Theorem 2.6

$$ = \exp \left[ \mu_1 + \frac{1}{2} \sigma_1^2 \right] \Lambda'(\exp \left[ \frac{\mu_2 + \rho \sigma_1 \sigma_2}{\sqrt{1 + \rho^2}} \right] y^\theta (1 + \rho \alpha \sigma_2) ) \right) $$

from Corollary 2.2b

For the relevance of assumptions (i) and (ii) in the theorem and its corollary we can only appeal to the test of data. The constancy of $\beta$ does not seem to us to imply a strong constraint, since, as we shall discuss in the next section, our own investigations have revealed only slight variation in practice, and the phenomenon may be similar to the constancy of $\sigma^2$ in income distributions. It need hardly be mentioned that the theorem holds for the special cases (a) when $\kappa$ and $\alpha$ are independent and (b) when either $\kappa$ or $\alpha$ takes on fixed values.

**12.8. THE ECONOMIC INTERPRETATION OF THE PARAMETERS OF THE ENGEL CURVE**

In an investigation of the prewar British working-class budgets carried out by the writers [5] the empirical generalization was made that the parameter $\beta$ of (12.17) (or equivalently the parameter $\sigma^2$ of (12.15)) varied by very little for the different commodities and in fact was approximately equal to unity. If this generalization is valid a number of simple results hold and the manipulation of the Engel curve in contexts of varying prices and so on is made much easier. For most of the manipulations it is sufficient to assume that $\beta$ is a constant, but for the re-
mainder of this chapter we will set $\beta = 1$ and write the Engel curve in the form: \[ q = \kappa \Lambda(ay). \] (12.18)

From equation (12.18) it will be seen that the Engel curves for all commodities may be reduced to a standard form merely by varying (a) the scale of measurement of consumption of the commodity, which is determined by the factor $\kappa$, and (b) the scale of measurement of income, determined by the factor $\alpha$. We have already defined $\kappa$ as the parameter of saturation and as such it has an obvious economic meaning: we may now term the parameter $\alpha$ the parameter of cheapness since its value controls the degree to which a consumer with given income can approach his saturation consumption (cf. Aitchison and Brown [51]). It follows also that the reciprocal of the value of this parameter is equal to the median effective income: that is, the income at which, on the average, consumers are purchasing at a rate equal to one-half their saturation rate. And finally the expression for the income elasticity of demand for equation (12.18) is given by

\[ \eta = \frac{\partial \log q}{\partial \log y} = \frac{ay'\Lambda'(ay)}{\Lambda(ay)} = Z(\log \alpha + \log y) / P(\log \alpha + \log y), \] (12.19)

where $Z$ and $P$ denote the ordinate and integral of the standardized normal distribution respectively. A tabulation of $Z/P$ is given in Appendix Table A5 to assist in the calculation of $\eta$, and Fig. 12.5 shows the graph of $\eta$ plotted against values of $\frac{q}{\kappa} = \Lambda(ay)$.

Fig. 12.6 gives the results of applying equation (12.18) to the 1937–8 survey of urban working-class households. Most of the remainder of this chapter will be given to a discussion of the more important factors influencing the parameters of saturation and cheapness.

12.9. The Problem of Additivity

Those familiar with the application of probit analysis to data obtained from quantitative assays will have realized that, in biological terms, the present model of consumers’ behaviour may be described as one in which a subject (the consumer) is subjected to a stimulus (his income) and his quantitative reaction (his purchases) measured. For example Finney [67] illustrates this model with data from an experiment in which honey-bees were offered pots of honey, each containing different concentrations of a repellent, and the quantities of honey consumed from each pot were measured. The chief distinction that must now be drawn between the honey-bees and the human consumers is that, for the latter a sort of ‘feed-back’ relation must hold between their reaction and their

† The adoption of equation (12.18) in place of equation (12.17) also simplifies the statistical estimation procedure. The procedure for (12.18) is described in § 7.8.
stimulus, since the sum of their expenditure on all commodities (including 'savings' as a commodity) must equal their income.

Symbolically we have

\[ q_i = F_i(y), \]  
\[ \sum p_i q_i = y, \]  

where \( q_i \) is the quantity† of the \( i \)th commodity and \( p_i \) its price. If the function \( F_i \) in (12.20) is replaced by the simple lognormal function as in (12.18) for all \( i \), the identity (12.21) cannot strictly hold (cf. Prais[162], Worswick[214] and Champernowne[33]). The difficulty could perhaps

be avoided by the device of substituting equations of the form (12.18) for (12.20) for all commodities but one, whose purpose would be to absorb the residual expenditure. But though each consumer may use such a 'residual' commodity it is unlikely that all consumers would choose the same one; and it seems more reasonable to suppose that the individual tolerance schemes are distorted from the lognormal form sufficiently for the budget identity (12.21) to be met. With the data at present available, however, it is not worth while seeking the appropriate manner for modifying equation (12.18) and in the sections which follow

† In the previous sections of this chapter the symbol \( q \) has stood either for a physical quantity or for this quantity valued at a constant price. From now on the symbol will always denote physical quantities.
the difficulty of additivity will be ignored. In practice equation (12.21) is very nearly fulfilled over a large part of the range of observed data, the discrepancies not being statistically significant. This is ensured because the estimation procedure is applied to observations which themselves satisfy the budget identity; and Fig. 12.7, in which we depict expenditure on 6 commodity groups by means of a layer diagram where the 45° line represents the budget identity, illustrates this phenomenon.
12.10. THE INFLUENCE OF PRICES

The difficulties of establishing accurate functional relationships between quantities and prices were mentioned in §12.3. In this section we consider the influence of prices via their influences on the parameters of the Engel curve; we shall obtain our more important results without specifying the form of certain functions, a question better left to empirical investigation, though we shall indicate certain simple forms which may be taken conveniently as first approximations.

Consider then two commodities which are close substitutes, say two popular brands of cigarettes. Suppose that $K_1 = 10$ is the saturation consumption for the first brand and $K_2 = 10$ is that for the second brand (that is, preferences of consumers are about equally divided between the two brands). Then for any significant change in the relative prices in favour of the first brand we should expect $K_1$ to increase towards 20 and $K_2$ to decline towards 0 as smokers of the second brand changed over to the first. That is, we should expect a 'migration' of the tolerance schemes. The first result then is that the saturation parameter $K$ is a function of the relative prices of all commodities; the function will probably be
homogeneous of degree zero, since an equiproportionate change in all prices should result in no migration:

$$\kappa_i = K_i (p_1 \ldots p_m).$$ \hspace{1cm} (12.22)

Secondly, a change in any one price must change the effective income of the consumer, causing him to move nearer to or farther away from his saturation levels generally. We may represent this by introducing a scalar divisor of income which is again a function of all prices, that is, a price index:

$$I = I(p_1 \ldots p_m).$$ \hspace{1cm} (12.23)

This divisor will be a constant for all commodities. The Engel curve for the $i$th commodity ($i = 1, \ldots, m$) now becomes

$$q_i = \kappa_i \Lambda \left( \frac{y}{p_i} \right).$$ \hspace{1cm} (12.24)

The elasticity of $\kappa_i$ with respect to the $j$th price, written

$$\sigma_{ij} = \frac{\partial \log \kappa_i}{\partial \log p_j},$$ \hspace{1cm} (12.25)

we will call the elasticity of substitution of the commodity. That this conforms with the usual terminology of the theory of consumption may be shown by analysing the elasticity of demand with respect to a single price:

$$\eta_i = \frac{\partial \log q_i}{\partial \log p_j} = \frac{\partial I}{\partial p_j} \frac{\partial q_i}{\partial I} + \frac{\partial \kappa_i}{\partial p_j} \Lambda \left( \frac{y}{p_i} \right)$$

$$= - \left( \frac{1}{q_i} \frac{\partial I}{\partial p_j} \right) \eta_i + \sigma_{ij},$$ \hspace{1cm} (12.26)

where $\eta_i$ is the income elasticity of demand defined by equation (12.19).

Equation (12.26) is analogous to Slutsky's useful relation (cf. Wold (213)) in the theory of a single consumer; this relation partitions the price elasticity for an equilibrium budget into (a) the product of the income elasticity and the proportion of expenditure incurred on the commodity and (b) the elasticity of substitution. The same partitioning is reasonable for the present case, even though our theory has not been developed in terms of an individual consumer. Equation (12.26) thus suggests that:

$$\frac{p_j}{I} \frac{\partial I}{\partial p_j} = \frac{p_j q_i}{y},$$ \hspace{1cm} (12.27)

which is satisfied by an index of the form

$$I = \prod_j p_j^{y_j}$$ \hspace{1cm} (12.28)
where the exponential weight $\omega_j$ is the proportion of income expended on the $j$th good. The expression $\Lambda \left( x_i \frac{y_j}{I} \right)$ is then homogeneous of degree zero in income and the prices, since $\sum \omega_j = 1$ by definition. It is to be noted that $I$ is now a function of $y$ which it would be difficult to make explicit; in practice it will be necessary to work with fixed weights $\omega_j$ at least over a range of incomes, relying on the well-known property of index numbers to be relatively insensitive to changes in weights.†

The function $K_i$ in equation (12.22) is not likely to be linear, as a consideration of our cigarette example will show; we may perhaps suggest

$$K_i = \epsilon_i \prod_j p_j^y$$

(12.29)

as an approximate form which will be easy to handle. As restrictions on this function we have first its degree, namely that $\sum \sigma_{ij} = 0$, and secondly the possibility that the theory of symmetry of substitution effects, $\sigma_{ij} = \sigma_{ji}$, may be carried over from the theory of a single consumer.

### 12.11. Changes in Preferences

What corresponds to a change in preferences (not stimulated by price changes) in the theory of an individual consumer is in our system represented by an increase or decrease in the areas under the community tolerance schemes; as individual consumers change their patterns of purchasing. We may represent such a change after it has occurred by a set of multipliers $\gamma_i$ defined by reference to the base situation

$$\gamma_i = \frac{\lambda_i}{\kappa_{i0}}$$

(12.30)

Such a change must, by virtue of the budget constraint, have repercussions on the measure of effective income. In fact the effect of multiplying the saturation level is, *ceteris paribus*, to multiply expenditure on the commodity by the same factor, which is the same result as if the price of the commodity were to increase from $p_i$ to $\gamma_i p_i$. The income effect of changes in preferences may therefore be represented by corresponding changes in the price index $I$ defined by (12.28), so that the new value of the index becomes

$$I' = \prod_j (\gamma_j p_j)^{y_j}$$

$$= \Gamma I,$$

(12.31)

where

$$\Gamma = \prod_j \gamma_i.$$

(12.32)

The effect of a small change in preferences in terms of elasticities thus becomes

$$\frac{\partial \log q_i}{\partial \log \gamma_i} = -\omega_i \eta_i + 1$$

(12.33)

† There is no intention here to imply any strong preferences for price indexes of the form (12.28). For an approximative argument such as we have given a Laspeyres or Paasche type index would be just as suitable.
and
\[ \frac{\partial \log q_i}{\partial \log \gamma_j} = -\omega_j \eta_i, \]  
(12.34)

which partitions the total effects into income and specific effects, by analogy with the theory of prices; the specific effect of course vanishing in (12.34) for \( i \neq j \), so that the matrix of specific effects reduces to the unit matrix.

12.12. THE EFFECTS OF HOUSEHOLD COMPOSITION

A most important determinant of the preferences of households (we take these as our fundamental consuming units) is known† to be the household’s composition in terms of the numbers of different types of person (defined, say, by sex and age in the simple case) which it contains.

In terms of the theory of the preceding section we can therefore introduce the effects of household composition on the Engel curve equation by supposing that \( \gamma_{ir} \), the multiplier for the \( i \)th commodity and the \( r \)th type of household, be given as a function of \( n_{ir} \), the number of the \( i \)th type of person in the \( r \)th type of household:

\[ \gamma_{ir} = \gamma_i(n_{ir} \ldots n_r). \]  
(12.35)

If the functions \( \gamma_i \) are linear and homogeneous of degree one, the partial derivatives of (12.35) may be identified with what are usually known as scales of equivalent adults (or of unit consumers) for the different commodities. If the function is homogeneous but of degree less than one we have the phenomenon of ‘economies of scale’, that is, if the number of people of each type in the household is doubled, consumption of the commodity is less than doubled. If the degree is greater than one we have ‘diseconomies of scale’; so that

\[ s = 1 - d, \]  
(12.36)

where \( d \) is the degree of the function (that is, the sum of the elasticities of \( \gamma_i \) with respect to all \( n_{ir} \)), is a convenient measure of these economies or diseconomies. We refer later in this section to data which suggest that, for total food expenditure in Great Britain, \( s \) is of the order of \( +\frac{1}{4} \).

From the theory of § 12.11 it will be seen that the effect of changing the number of persons of any one type in the household (assuming prices and money income constant) is complex, since each multiplier \( \gamma_{ir} \) will change its value and the effect of each change is similar to a change in one price. Taking the elasticity of \( q_{ir} \) with respect to \( n_{ir} \) we obtain

\[ \frac{\partial \log q_{ir}}{\partial \log n_{ir}} = -\left( \sum_j \frac{\partial \log \gamma_{ir}}{\partial \log n_{ir}} \right) \eta_i + \frac{\partial \log \gamma_{ir}}{\partial \log n_{ir}}. \]  
(12.37)

The expression \( \frac{\partial \log \gamma_{ir}}{\partial \log n_{ir}} \) is the elasticity of the specific effect associated with the type of person \( t \), and from the term in brackets it may be seen that the elasticity of the income divisor \( \Gamma_r \) for the \( r \)th type of household

† Cf., for example, Prais and Houthakker[164].
with respect to \( r \), is a weighted sum of the specific elasticities, the weights \( \omega_{ir} \) again being the proportions of total expenditure disbursed on the various commodities. If all changes in preferences except those which are associated with the composition of the household are ignored, as we are entitled to do when we compare the behaviour of different types of households at the same moment of time, the number \( \gamma_r \) and the index \( \Gamma_r \) may be regarded as measures of the specific and income size of the household respectively. Estimates of these two measures for a number of types of British households in 1951–2 are given by Brown[27]; and Fig. 12.8 shows the Engel curves for four types of household in respect of total food expenditure, taken from the same source.

12.13. AGGREGATION AND THE ANALYSIS OF TIME SERIES

By the arguments of the preceding sections we have constructed an Engel curve which will obtain for a sub-population of families of given composition in a given short period \( t' \) defined by a set of prices \( p_1 \ldots p_m \):

\[
q_{ir} = \gamma_{ir} \kappa_{ir} \Lambda \left( x_i, \frac{y_{it'}}{F_{it'}} \right).
\]

(12.38)
With the aid of Theorem 12.1 and its corollary this equation can be consolidated to obtain one which will hold for the whole population (in which the measures of consumption and income are taken per person, say) over a given time period (say a year) which we will write in the form

$$ q_{it} = \kappa_{it} \Lambda \left( \alpha_i y_t \right). \quad (12.39) $$

This is the form we would require if we had a time series of family budget inquiries to analyse; for each time period $t$, independent estimates can be obtained of $\kappa_{it}$ and $\alpha_i/I_t$, and, from this series of estimates, hypotheses of the form $(12.28)$ and $(12.29)$ may be tested and the elasticities of substitution estimated. Unfortunately, the demand analyst is rather in the position of an astronomer to whom the government grants one look through a large telescope every other decade: the telescope in our case being the large-scale budget inquiries which were carried out in this country in 1937-8 and 1953-4. If we are asked to make do with data obtained, as it were, with the aid of a small magnifying glass (one observation on the whole community each year), we shall have to rely more on a priori hypotheses which the data are too weak to test, and learn to be less surprised if the predictive power of our models is unimpressive.

To find the mean consumption of the $i$th commodity by the whole nation $\bar{q}_{it}$, we aggregate over the distribution of income as in §12.2:

$$ \bar{q}_{it} = \int_0^\infty \kappa_{it} \Lambda (\alpha_i y_t / I_t) d \Lambda (y_t, \mu, \sigma^2_t) $$

$$ = \kappa_{it} \Lambda (\alpha_i y_t / I_t | 0, 1 + \sigma^2_t), \quad (12.40) $$

where the assumption is made that $y_t$ is lognormal with parameters $\mu_t, \sigma^2_t$. Since $\kappa_{it}$ and $I_t$ are functions of the prices, and since it will be necessary to place the least strain on the observed series of data, it will be preferable to estimate $\kappa_{i0}$ and $\alpha_i/I_0$ from at least one budget study made during the period and $\sigma^2_t$ from data on income distributions. The index $I_t$ may be constructed from $(12.28)$. We may then write

$$ \bar{q}_{it} (\alpha_i y_t / I_t^{-1} | 0, 1 + \sigma^2_t)^{-1} = \kappa_{it} $$

$$ = \kappa_i (p_1 \ldots p_m) $$

$$ = \epsilon_i \prod_j p_j^{\epsilon_i}, \quad \text{say}, \quad (12.41) $$

in which the prices may be regarded as regressors, and the expression on the left-hand side as a regressand; with the purpose of deriving estimates of the elasticities of substitution.

### 12.14. Concluding Remarks

The derivation of the demand curve for a typical commodity in the previous sections began with the commodity’s Engel curve, because this is more easily established by the data generated in the natural working of the economic process than the form of curve which represents the
relation between demand and prices. There is a further advantage to this order of priority. The relation between income and demand may be treated in a simple way because, in the static situation, commodity prices are given to the consumer and are unaffected by the size of his income. The effect of a price change on demand could also be treated as a simple 'stimulus-response' situation were it not for the fact that changes in prices influence the effective size of the consumer's income. The direct influence of a price change (the substitution effect) is therefore modified by the indirect influence which operates via income (the income effect). This twofold character of the price influence is illustrated by expressing the relation between the quantity demanded and a single price (all other variables assumed constant):

\[ q_i = K_i(p_j) \Lambda(z^*p_j^{-\omega_j}), \quad (12.42) \]

which shows that the appearance of the demand curve will depend on whether the income or substitution effect is dominant. A similar argument holds for the effect of changes in household composition and other factors affecting consumer preferences.

A number of the convenient properties which have been developed for the present system of relationships depend on the fact that it is fundamentally of the form

\[ \left( \frac{q_i}{a_i} \right) = f_i \left( \frac{y_i}{b_i} \right), \quad (12.43) \]

where \( a_i \) and \( b_i \) are scalar numbers; and the same properties could be derived for any system which could be similarly represented. The main point which we would wish to emphasize in conclusion, however, is that our efforts have been directed towards the discovery of relationships which characteristically do not appear until a group of consumers is studied rather than a single individual; in setting up a system of relationships of this type we have therefore preferred to make our strong assumptions statistical in nature rather than to choose from those based on economic introspection.
CHAPTER 13

COMPUTATION PROBLEMS

Duke of Milan. And here an engine fit for my proceeding.
Two Gentlemen of Verona

13.1. THE USE OF AN ELECTRONIC COMPUTER

The greater part of the calculations reported in this monograph has been carried out on an automatic digital computer. The effect of this has been partly to speed up work which would otherwise have been done on desk machines, but partly, and more importantly, to extend the range of problems which it has been found possible to treat. In the latter class must be placed the application of estimation procedures to sixty-five artificial samples, comprising some 8000 variate values drawn from specified lognormal populations; and also the work done on probit analysis, in particular in the study of convergence problems, for which some 850 iterations were performed on the Rotenone data alone. The use of automatic machines is not yet widespread amongst practising statisticians, but we predict that this is a matter which a relative short passage of time will rectify; for statisticians stand to gain as much as any other scientist from the freedom from arduous arithmetic that these machines will provide. We therefore offer the reader the following comments on the automatic programmes we found useful to construct for our purposes.

13.2. DESCRIPTION OF THE EDSAC

The machine used was the Electronic Delay Storage Automatic Calculator, built by the staff of the Mathematical Laboratory of the University of Cambridge under the direction of its Director Dr M. V. Wilkes. The EDSAC uses standard teleprinter paper tape for input and output, and a memory (at the time of the applications described) of thirty-two mercury delay lines each capable of storing sixteen long words of thirty-five binary digits, or thirty-two short words of seventeen binary digits, making a total of 1024 short words. A short or long word may represent a number (of approximately five or ten decimal digits respectively), in which case the first binary digit represents its sign, and the binary point is normally assumed to lie immediately to the right of this; numbers are therefore usually scaled before or during input to lie between $-1$ and $+1$, and appropriate steps must be taken during the calculation to prevent intermediate or final results lying outside this range. Or a short word may represent a machine order (in a single address code); these orders are normally obeyed serially starting from a given point in the memory; but special orders are available which direct the control of the machine to an order not in sequence, either unconditionally, or according to the sign of some specified quantity. The arithmetical unit
of the machine has addition, subtraction and multiplication facilities; the sum, difference or product being held in an accumulator of seventy binary digits until it is transferred to the memory. Standard sequences of orders, capable of being obeyed one or more times for a specific purpose during a calculation, such as for the taking of a square root, are known as subroutines, of which the Mathematical Laboratory has a large and continually growing library. It is, however, usually necessary to construct some new subroutines for each novel calculation. For further details on the EDSAC the reader is referred to the textbook by Wilkes, Wheeler and Gill[206], and to the supplement to this published by the Mathematical Laboratory[206]; and for discussion of automatic computers in general to the works of Bowden[25] and of Booth and Booth[24].

13.3. THE PROCESSING OF THE 65 SAMPLES

The processing of the sixty-five samples was treated as a single problem; for the reading in, or generation within the machine, of artificial samples is a lengthy process compared with the calculation of parameter estimates, and it is advisable to combine as many of the latter type of calculation as possible into one operation. The generation of artificial samples from a specified distribution is perfectly possible on a digital computer, and will undoubtedly replace the use of tables of random numbers more and more as time goes by; pseudo-random sequences of binary digits can be produced rapidly, and all that is further required is a subroutine to apply the appropriate transformation. In fact this method was not used by the authors; partly to save space in the rather limited EDSAC memory, and partly because it was felt preferable to apply the estimation procedures to published data which had already been subjected to exhaustive tests of randomness. Accordingly the necessary number of variate values were punched on tape from Wold’s Random Normal Deviates[212], which are drawn from a normal (0, 1) population and specified to three digits. These values were divided up into groups of 32, 64 and so on, corresponding to the required sample sizes, and in front of each sample was punched the allocated value of \( \sigma \) to define the given lognormal population \( \Lambda(o, \sigma^2) \). The following operations were performed on each sample. First, as each variate value \( y_i \) was read into the machine, \( y_i = \sigma u_i \) and \( x_i = e^{y_i} \) were calculated (using a standard exponential subroutine), then the first four powers of \( y \) and \( x \), suitably scaled, to accumulate towards the moments about zero; also two new subroutines arranged that the ten highest and ten lowest values of \( x \) computed up to that point were retained, and that a record of the cumulative frequency distribution of \( x \) was built up according to a preset class interval. When the end of the sample was reached, the first four moments about the mean, the coefficients of variation, skewness and kurtosis of both \( y \) and \( x \), and the value of \( u \) (from the equation \( u^3 + 3u - g_3(x) = 0 \)) were calculated. The identification of the sample, its frequency distribution, its ten highest and lowest values, and the
sample functions referred to were then printed out, and the next sample read into the machine. These printed results were sufficient for the quick application of the estimation procedures, described in Chapters 5 and 6, on desk machines and altogether about twelve hours of machine time were used in producing them, including the time taken to develop the programme. A programme was also designed for the calculation of moments and derived statistics from a grouped frequency distribution, provided that the class intervals are equal. The simplest procedure is to read into the machine simply the list of frequencies $f_i$, making sure that any zero frequencies are explicitly punched on the tape. The first four factorial moments, taking the lower bound of the first class interval as the origin and the size of class interval as unity, can then be rapidly computed by progressive summation (as described by Kendall[123], vol. 1, pp. 58–61) and the required statistics derived from these. Because of the implicit transformation of the variate to a standard scale, scaling problems are easy to handle, provided some care is taken to preserve the accuracy of the higher moments.

The estimation of the third parameter, referred to in Chapter 6, using the least-sample-value and maximum-likelihood methods, had to be treated separately, both for reasons of capacity and because the process was iterative. For the second reason it was necessary to retain the full sample of variate values in the memory during the whole calculation. The method used was to start with an initial estimate of $\tau_1$ $t_0^*$, derived from the least-sample value $x_0$. The quantity $\phi$ was then computed from equation (6.8) and the next estimate of $\tau$, $t_i^*$, from a recurrence relation based on the 'rule of false position'. The difference $|t_{i+1}^* - t_i^*|$ was then compared with a small preset quantity to determine whether convergence was reached. The main difficulty to be avoided in the application was the taking of the logarithm of too small a fraction, which would occur if any $t_i^*$ were too close to $x_0$; this was done by ensuring that $x_0 - t_i^*$ was sufficiently large, and that the process was terminated if any $t_{i+1}^*$ were greater than $t_i^*$. When the final least sample value estimate $t_0^*$ was obtained, a series of tests was performed to determine whether the maximum-likelihood estimate was closer to the true value of $\tau$ (zero) than $t_0^*$. This was done by calculating the value of the function $\theta$ (6.4) at $t_0^*$ and at $-t_0^*$ (or at $t_0^*$ and $t_i^*$ if $t_i^* < 0$ and $t_i^* > t_0^*$), and determining from these two points whether the function took on the value of zero in the range $(-t_0^*, t_0^*)$; to save printing time, a coded result was then printed which indicated which method of estimation gave results nearer to the true value of $\tau$ for the particular sample. In these two applications scaling problems presented no great difficulty but even on an automatic machine the procedure was rather lengthy for the sample sizes greater than thirty-two.

### 13.4. The Programme for Quantal Probit Analysis

The programmes developed for probit analysis were only partly designed for the purpose of this monograph; for the authors had mainly in mind
the construction of programmes which would be useful in future pra-
tical work. These programmes have been made much more economi
cal by the recent introduction into the EDSAC of the equipment known as
de the B-register; which in effect is an auxiliary arithmetical unit designed
mainly for counting the number of times a particular set of orders has
been obeyed and modifying the individual orders belonging to the set
appropriately, and therefore for allowing a computation cycle to be used
a definite number of times. In the probit applications the problem of
scanning was important, since it was desirable not to restrict the range of
data which would be handled with accuracy; and therefore arrange-
ments were made for the machine to choose its own scaling factors by
examination of the data.

The structure of the programme for quantal assay is as follows: first
the reading of the data arranged in triplets \( n_i, p_i, x_i \) with the \( n_i \) and \( x_i \)
scaled by a suitable power of ten so that \( \max n_i, x_i \) are as near to unity,
without exceeding it, as possible, together with the initial estimates
\( a_0, b_0 \) (this is now superseded as a result of the work on convergence, since
it is nearly always possible to take \( 0, 0 \) as the initial values). Then follows
the calculation of the elements of the information matrix \( I \) defined by
(7.16) and of the vector on the right-hand side of (7.18): this involves
the use of an exponential routine to compute \( Z \) and the approximation
to \( P \) given by (7.23). If the sample is large there is some danger that
overflows will occur in the accumulation of the sums \( \Sigma n_i, \Sigma n_i p_i \) and so on,
which at the same time should be kept as close to unity as possible for
reasons of accuracy. The method employed is to use an accumulator-
operation-detection order, which causes the optimum scaling factor \( 2^{-r} \) to
be applied both to the information matrix \( I \) and to the right-hand vector,
so that the factor will cancel at a later stage; the factor is also stored
in order to correct the final variance matrix. The next stage is the inver-
sion of the (2 \( \times \) 2) information matrix \( 2^{-r} I \): first the determinant is taken,
and, before any accuracy is lost by rounding, scaled by a factor \( 2^t \) such
that \( \frac{1}{2} < 2^t \mid 2^{-r} I \mid < 1 \); the reciprocal is then taken and this multiplied
with the cofactors of \( I \) giving \( 2^{-r} (I^{-1}) \). The recorded factor \( 2^{-r} \) is used
later to initiate a B-register controlled cycle which corrects the variance
matrix. The equations (7.18) are then solved for \( a_1 \) and \( b_1 \) and the test
of convergence applied; if convergence to the preassigned degree is not
reached, the process is repeated with \( a_1 \) and \( b_1 \) replacing \( a_0 \) and \( b_0 \);
otherwise the estimates are printed together with their variance matrix,
and finally, the original data are printed with the points on the fitted
curve corresponding to the observed values of \( x \), together with the test
statistic given by (7.22) which is calculated from the fitted curve.

13.5. The Programme for Quantitative Probit Analysis
The programme for the quantitative model defined by (7.29) is similar
in structure, except that since equations (7.35) are used, an inversion
subroutine for a third-order matrix is required; on the EDSAC it has
been found most economical to compute the six distinct cofactors of the information matrix to a low order of accuracy, select the largest, and from this determine the optimum scaling factor; the calculation is then repeated with the scaling factor, the reciprocal of the determinant evaluated and multiplied into the cofactors. This is not a method which obviously appears best, but it is in fact very quick and economical in the memory space it requires.

13.6. The Convergence Programme

The quantal programme was easily adapted to the study of convergence, as described in § 7.6. First a test was inserted to determine whether \(|a_i + b_1 x_i| > 4\) for any \(x_i\) during an iteration: if so the starting points of the iteration \(a_n, b_i\) were classed as ‘divergent’, as were all points \(a_{n-\tau}, b_{i-\tau}\) if any, which led through previous iterations to \(a_n, b_i\), and a suitable symbol printed. Otherwise the calculation proceeded and the resulting estimates \(a_{n+1}, b_{i+1}\) were printed. Once it was known that the convergence area was roughly elliptical, with the major axis sloping to the left from the origin (Fig. 7.3), it was easy to incorporate a decision procedure so that successive new starting points would be chosen according to the convergence or non-convergence of the previous few points. In this way the machine itself traced the shape of the convergence area, taking about an hour to do so for each sample.

13.7. The Construction of Tables for \(\psi_i(t)\) and \(\chi_i(t)\)

The other lognormal programmes designed by the authors were used for the construction of tables, especially for the tables of \(\psi_i(t)\) and \(\chi_i(t)\) given as an Appendix Tables A2, 3. These functions are ideal for automatic tabulation; their formulae are given by equations (5.40) and (5.44) respectively. The only data required are the series of constants \(2^{-3}(1, \frac{1}{2}, \ldots, \frac{1}{8})\), the initial values of \(2^{-16}, 2^{-32}\), and the two intervals by which they are to be increased.

From the initial value of \(n\), the fraction \(\frac{n - 1}{n(n + 1)}\), and from this the coefficients of the power series of (5.40) (in so far as they contain \(n\)) are computed by a recurrence relation, and held in the memory; the power series of \(t\) is then generated, combined with these coefficients and the series of constants to form the first fifteen terms of (5.40) and summed; a cyclical count is arranged to tabulate all required values of \(t\) for fixed \(n\), and then to increase \(n\) and repeat the process. The process for \(\chi_i(t)\) merely combines the calculation for two values of \(\psi_i\) according to equation (5.44); both programmes are very fast, the functional values being produced quicker than they could be copied by a typist.
REFERENCES


REFERENCES

REFERENCES

REFERENCES


REFERENCES


[134] LEE, A. (1915). Table of Gaussian 'tail' functions when the 'tail' is larger than the body. *Biometrika*, 10, 208.


REFERENCES


[156] Pearson, E. S. and Adyanthaya, N. K. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. (Second paper: Distribution of 'Student's' t.) Biometrika, 21, 259.


REFERENCES

REFERENCES


## APPENDIX A. TABLES OF FUNCTIONS

### TABLE A1. CHARACTERISTICS OF LOGNORMAL DISTRIBUTIONS

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Coefficient of variation $\eta = (e^{\sigma^2} - 1)^{1/2}$</th>
<th>Coefficient of skewness $\eta^3 + 3\eta$</th>
<th>Coefficient of kurtosis $\eta^4 + 6\eta^3 + 3(\eta^2 + 1)$</th>
<th>Ratio of mean to median $\eta^b$</th>
<th>Ratio of mean to mode $\eta^{b^2}$</th>
<th>Proportion not greater than mean $N(\eta, 0, 1)$</th>
<th>Lorentz measure $2N(\eta^2, 0, 1) - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0500</td>
<td>0.1502</td>
<td>0.0401</td>
<td>1.0013</td>
<td>1.0013</td>
<td>0.5100</td>
<td>0.0079</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1003</td>
<td>0.3018</td>
<td>0.1623</td>
<td>1.0050</td>
<td>1.0050</td>
<td>0.5199</td>
<td>0.0066</td>
</tr>
<tr>
<td>0.15</td>
<td>0.1508</td>
<td>0.4550</td>
<td>0.3719</td>
<td>1.0113</td>
<td>1.0113</td>
<td>0.5299</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2002</td>
<td>0.6143</td>
<td>0.5784</td>
<td>1.0202</td>
<td>1.0202</td>
<td>0.5399</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.25</td>
<td>0.2540</td>
<td>0.7708</td>
<td>1.0950</td>
<td>1.0317</td>
<td>1.0317</td>
<td>0.5497</td>
<td>0.0121</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3009</td>
<td>0.9415</td>
<td>1.6449</td>
<td>1.1420</td>
<td>1.1420</td>
<td>0.5596</td>
<td>0.1405</td>
</tr>
<tr>
<td>0.35</td>
<td>0.3610</td>
<td>1.1300</td>
<td>2.3534</td>
<td>1.2621</td>
<td>1.2621</td>
<td>0.5695</td>
<td>0.1679</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4166</td>
<td>1.3219</td>
<td>3.2000</td>
<td>1.3966</td>
<td>1.3966</td>
<td>0.5795</td>
<td>0.1951</td>
</tr>
<tr>
<td>0.45</td>
<td>0.4738</td>
<td>1.5277</td>
<td>4.1750</td>
<td>1.4317</td>
<td>1.4317</td>
<td>0.5890</td>
<td>0.2228</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5329</td>
<td>1.7502</td>
<td>5.3843</td>
<td>1.4762</td>
<td>1.4762</td>
<td>0.5987</td>
<td>0.2495</td>
</tr>
<tr>
<td>0.55</td>
<td>0.5913</td>
<td>1.9830</td>
<td>7.8035</td>
<td>1.5242</td>
<td>1.5242</td>
<td>0.6084</td>
<td>0.2767</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6503</td>
<td>2.2601</td>
<td>9.0732</td>
<td>1.5720</td>
<td>1.5720</td>
<td>0.6179</td>
<td>0.3027</td>
</tr>
<tr>
<td>0.65</td>
<td>0.7251</td>
<td>2.5565</td>
<td>1.3567</td>
<td>1.6212</td>
<td>1.6212</td>
<td>0.6274</td>
<td>0.3284</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7952</td>
<td>2.8883</td>
<td>1.7791</td>
<td>1.6706</td>
<td>1.6706</td>
<td>0.6371</td>
<td>0.3545</td>
</tr>
<tr>
<td>0.75</td>
<td>0.8689</td>
<td>3.2629</td>
<td>2.3540</td>
<td>1.7206</td>
<td>1.7206</td>
<td>0.6468</td>
<td>0.3794</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9418</td>
<td>3.6841</td>
<td>3.1968</td>
<td>1.7711</td>
<td>1.7711</td>
<td>0.6564</td>
<td>0.4039</td>
</tr>
<tr>
<td>0.85</td>
<td>1.0264</td>
<td>4.1768</td>
<td>4.2192</td>
<td>1.8216</td>
<td>1.8216</td>
<td>0.6660</td>
<td>0.4288</td>
</tr>
<tr>
<td>0.9</td>
<td>1.1171</td>
<td>4.7453</td>
<td>4.7141</td>
<td>1.8721</td>
<td>1.8721</td>
<td>0.6757</td>
<td>0.4522</td>
</tr>
<tr>
<td>0.95</td>
<td>1.2109</td>
<td>5.4667</td>
<td>5.7190</td>
<td>1.9229</td>
<td>1.9229</td>
<td>0.6853</td>
<td>0.4752</td>
</tr>
<tr>
<td></td>
<td>1.308</td>
<td>6.1849</td>
<td>1.1094</td>
<td>1.6487</td>
<td>4.4817</td>
<td>0.6015</td>
<td>0.5204</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>1.534</td>
<td>8.213</td>
<td>2.952</td>
<td>1.831</td>
<td>6.441</td>
<td>0.7088</td>
<td>0.6646</td>
</tr>
<tr>
<td>2</td>
<td>1.705</td>
<td>1.116</td>
<td>5.152</td>
<td>2.954</td>
<td>8.671</td>
<td>0.7367</td>
<td>0.6647</td>
</tr>
<tr>
<td>3</td>
<td>2.102</td>
<td>1.560</td>
<td>1.283</td>
<td>2.328</td>
<td>1.262</td>
<td>0.7422</td>
<td>0.6424</td>
</tr>
<tr>
<td>4</td>
<td>2.470</td>
<td>2.247</td>
<td>3.401</td>
<td>2.664</td>
<td>1.892</td>
<td>0.7560</td>
<td>0.6775</td>
</tr>
<tr>
<td>5</td>
<td>2.913</td>
<td>3.347</td>
<td>1.008</td>
<td>3.080</td>
<td>2.922</td>
<td>0.7734</td>
<td>0.7108</td>
</tr>
<tr>
<td>6</td>
<td>3.455</td>
<td>5.100</td>
<td>3.283</td>
<td>3.597</td>
<td>4.553</td>
<td>0.7881</td>
<td>0.7415</td>
</tr>
<tr>
<td>7</td>
<td>4.122</td>
<td>8.232</td>
<td>1.174</td>
<td>4.242</td>
<td>7.032</td>
<td>0.8023</td>
<td>0.7699</td>
</tr>
<tr>
<td>8</td>
<td>4.953</td>
<td>1.304</td>
<td>4.603</td>
<td>5.353</td>
<td>1.990</td>
<td>0.8159</td>
<td>0.7959</td>
</tr>
<tr>
<td>9</td>
<td>5.997</td>
<td>2.937</td>
<td>1.972</td>
<td>6.080</td>
<td>2.248</td>
<td>0.8289</td>
<td>0.8198</td>
</tr>
<tr>
<td>10</td>
<td>7.321</td>
<td>4.144</td>
<td>9.221</td>
<td>7.889</td>
<td>4.934</td>
<td>0.8413</td>
<td>0.8115</td>
</tr>
<tr>
<td>11</td>
<td>9.015</td>
<td>7.557</td>
<td>4.604</td>
<td>9.979</td>
<td>7.462</td>
<td>0.8531</td>
<td>0.8611</td>
</tr>
<tr>
<td>12</td>
<td>1.120</td>
<td>1.439</td>
<td>2.509</td>
<td>1.125</td>
<td>1.422</td>
<td>0.8643</td>
<td>0.8812</td>
</tr>
<tr>
<td>13</td>
<td>1.405</td>
<td>2.614</td>
<td>1.563</td>
<td>1.408</td>
<td>2.293</td>
<td>0.8749</td>
<td>0.8969</td>
</tr>
<tr>
<td>14</td>
<td>1.779</td>
<td>5.660</td>
<td>1.021</td>
<td>1.781</td>
<td>5.653</td>
<td>0.8849</td>
<td>0.9109</td>
</tr>
<tr>
<td>15</td>
<td>2.274</td>
<td>1.824</td>
<td>7.228</td>
<td>2.276</td>
<td>1.179</td>
<td>0.8944</td>
<td>0.9233</td>
</tr>
<tr>
<td>16</td>
<td>2.935</td>
<td>2.538</td>
<td>5.550</td>
<td>2.237</td>
<td>2.734</td>
<td>0.9032</td>
<td>0.9342</td>
</tr>
<tr>
<td>17</td>
<td>3.987</td>
<td>5.616</td>
<td>4.460</td>
<td>3.268</td>
<td>5.911</td>
<td>0.9115</td>
<td>0.9432</td>
</tr>
<tr>
<td>18</td>
<td>5.039</td>
<td>1.281</td>
<td>4.167</td>
<td>5.040</td>
<td>1.980</td>
<td>0.9192</td>
<td>0.9523</td>
</tr>
<tr>
<td>19</td>
<td>6.701</td>
<td>3.011</td>
<td>4.072</td>
<td>6.702</td>
<td>3.010</td>
<td>0.9265</td>
<td>0.9596</td>
</tr>
<tr>
<td>20</td>
<td>9.001</td>
<td>7.296</td>
<td>4.312</td>
<td>9.002</td>
<td>7.394</td>
<td>0.9332</td>
<td>0.9660</td>
</tr>
<tr>
<td>21</td>
<td>1.221</td>
<td>1.821</td>
<td>4.947</td>
<td>1.221</td>
<td>1.821</td>
<td>0.9393</td>
<td>0.9715</td>
</tr>
<tr>
<td>22</td>
<td>1.673</td>
<td>4.686</td>
<td>6.149</td>
<td>1.673</td>
<td>4.686</td>
<td>0.9452</td>
<td>0.9762</td>
</tr>
<tr>
<td>23</td>
<td>2.316</td>
<td>1.242</td>
<td>8.279</td>
<td>2.160</td>
<td>1.232</td>
<td>0.9505</td>
<td>0.9862</td>
</tr>
<tr>
<td>24</td>
<td>3.238</td>
<td>3.394</td>
<td>1.207</td>
<td>3.283</td>
<td>3.294</td>
<td>0.9554</td>
<td>0.9836</td>
</tr>
<tr>
<td>25</td>
<td>4.571</td>
<td>9.554</td>
<td>1.907</td>
<td>4.571</td>
<td>9.553</td>
<td>0.9599</td>
<td>0.9865</td>
</tr>
<tr>
<td>26</td>
<td>5.530</td>
<td>3.771</td>
<td>5.265</td>
<td>6.290</td>
<td>2.771</td>
<td>0.9641</td>
<td>0.9862</td>
</tr>
<tr>
<td>27</td>
<td>9.392</td>
<td>8.284</td>
<td>6.053</td>
<td>9.392</td>
<td>8.284</td>
<td>0.9678</td>
<td>0.9912</td>
</tr>
<tr>
<td>28</td>
<td>1.366</td>
<td>2.552</td>
<td>1.216</td>
<td>1.365</td>
<td>2.532</td>
<td>0.9715</td>
<td>0.9929</td>
</tr>
<tr>
<td>29</td>
<td>2.608</td>
<td>8.699</td>
<td>2.645</td>
<td>2.608</td>
<td>0.999</td>
<td>0.9744</td>
<td>0.9942</td>
</tr>
<tr>
<td>30</td>
<td>2.981</td>
<td>2.649</td>
<td>6.235</td>
<td>2.981</td>
<td>2.619</td>
<td>0.9772</td>
<td>0.9953</td>
</tr>
</tbody>
</table>

Note on notation: $2\times10^3 = 2.981 \times 10^3$. 

APPENDIX A
<table>
<thead>
<tr>
<th>n</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.053</td>
<td>1.054</td>
<td>1.055</td>
<td>1.056</td>
<td>1.057</td>
<td>1.058</td>
<td>1.059</td>
<td>1.060</td>
<td>1.061</td>
</tr>
<tr>
<td>0.10</td>
<td>1.091</td>
<td>1.092</td>
<td>1.093</td>
<td>1.094</td>
<td>1.095</td>
<td>1.096</td>
<td>1.097</td>
<td>1.098</td>
<td>1.099</td>
</tr>
<tr>
<td>0.15</td>
<td>1.129</td>
<td>1.130</td>
<td>1.131</td>
<td>1.132</td>
<td>1.133</td>
<td>1.134</td>
<td>1.135</td>
<td>1.136</td>
<td>1.137</td>
</tr>
<tr>
<td>0.20</td>
<td>1.167</td>
<td>1.168</td>
<td>1.169</td>
<td>1.170</td>
<td>1.171</td>
<td>1.172</td>
<td>1.173</td>
<td>1.174</td>
<td>1.175</td>
</tr>
<tr>
<td>0.25</td>
<td>1.206</td>
<td>1.207</td>
<td>1.208</td>
<td>1.209</td>
<td>1.210</td>
<td>1.211</td>
<td>1.212</td>
<td>1.213</td>
<td>1.214</td>
</tr>
<tr>
<td>0.30</td>
<td>1.245</td>
<td>1.246</td>
<td>1.247</td>
<td>1.248</td>
<td>1.249</td>
<td>1.250</td>
<td>1.251</td>
<td>1.252</td>
<td>1.253</td>
</tr>
<tr>
<td>0.35</td>
<td>1.284</td>
<td>1.285</td>
<td>1.286</td>
<td>1.287</td>
<td>1.288</td>
<td>1.289</td>
<td>1.290</td>
<td>1.291</td>
<td>1.292</td>
</tr>
<tr>
<td>0.40</td>
<td>1.324</td>
<td>1.325</td>
<td>1.326</td>
<td>1.327</td>
<td>1.328</td>
<td>1.329</td>
<td>1.330</td>
<td>1.331</td>
<td>1.332</td>
</tr>
<tr>
<td>0.45</td>
<td>1.364</td>
<td>1.365</td>
<td>1.366</td>
<td>1.367</td>
<td>1.368</td>
<td>1.369</td>
<td>1.370</td>
<td>1.371</td>
<td>1.372</td>
</tr>
<tr>
<td>0.50</td>
<td>1.404</td>
<td>1.405</td>
<td>1.406</td>
<td>1.407</td>
<td>1.408</td>
<td>1.409</td>
<td>1.410</td>
<td>1.411</td>
<td>1.412</td>
</tr>
<tr>
<td>0.55</td>
<td>1.444</td>
<td>1.445</td>
<td>1.446</td>
<td>1.447</td>
<td>1.448</td>
<td>1.449</td>
<td>1.450</td>
<td>1.451</td>
<td>1.452</td>
</tr>
<tr>
<td>0.60</td>
<td>1.484</td>
<td>1.485</td>
<td>1.486</td>
<td>1.487</td>
<td>1.488</td>
<td>1.489</td>
<td>1.490</td>
<td>1.491</td>
<td>1.492</td>
</tr>
<tr>
<td>0.65</td>
<td>1.524</td>
<td>1.525</td>
<td>1.526</td>
<td>1.527</td>
<td>1.528</td>
<td>1.529</td>
<td>1.530</td>
<td>1.531</td>
<td>1.532</td>
</tr>
<tr>
<td>0.70</td>
<td>1.564</td>
<td>1.565</td>
<td>1.566</td>
<td>1.567</td>
<td>1.568</td>
<td>1.569</td>
<td>1.570</td>
<td>1.571</td>
<td>1.572</td>
</tr>
<tr>
<td>0.75</td>
<td>1.604</td>
<td>1.605</td>
<td>1.606</td>
<td>1.607</td>
<td>1.608</td>
<td>1.609</td>
<td>1.610</td>
<td>1.611</td>
<td>1.612</td>
</tr>
<tr>
<td>0.80</td>
<td>1.644</td>
<td>1.645</td>
<td>1.646</td>
<td>1.647</td>
<td>1.648</td>
<td>1.649</td>
<td>1.650</td>
<td>1.651</td>
<td>1.652</td>
</tr>
<tr>
<td>0.85</td>
<td>1.684</td>
<td>1.685</td>
<td>1.686</td>
<td>1.687</td>
<td>1.688</td>
<td>1.689</td>
<td>1.690</td>
<td>1.691</td>
<td>1.692</td>
</tr>
<tr>
<td>0.90</td>
<td>1.724</td>
<td>1.725</td>
<td>1.726</td>
<td>1.727</td>
<td>1.728</td>
<td>1.729</td>
<td>1.730</td>
<td>1.731</td>
<td>1.732</td>
</tr>
<tr>
<td>0.95</td>
<td>1.764</td>
<td>1.765</td>
<td>1.766</td>
<td>1.767</td>
<td>1.768</td>
<td>1.769</td>
<td>1.770</td>
<td>1.771</td>
<td>1.772</td>
</tr>
<tr>
<td>1.00</td>
<td>1.804</td>
<td>1.805</td>
<td>1.806</td>
<td>1.807</td>
<td>1.808</td>
<td>1.809</td>
<td>1.810</td>
<td>1.811</td>
<td>1.812</td>
</tr>
<tr>
<td>1.05</td>
<td>1.844</td>
<td>1.845</td>
<td>1.846</td>
<td>1.847</td>
<td>1.848</td>
<td>1.849</td>
<td>1.850</td>
<td>1.851</td>
<td>1.852</td>
</tr>
<tr>
<td>1.10</td>
<td>1.885</td>
<td>1.886</td>
<td>1.887</td>
<td>1.888</td>
<td>1.889</td>
<td>1.890</td>
<td>1.891</td>
<td>1.892</td>
<td>1.893</td>
</tr>
<tr>
<td>1.15</td>
<td>1.925</td>
<td>1.926</td>
<td>1.927</td>
<td>1.928</td>
<td>1.929</td>
<td>1.930</td>
<td>1.931</td>
<td>1.932</td>
<td>1.933</td>
</tr>
<tr>
<td>1.20</td>
<td>1.966</td>
<td>1.967</td>
<td>1.968</td>
<td>1.969</td>
<td>1.970</td>
<td>1.971</td>
<td>1.972</td>
<td>1.973</td>
<td>1.974</td>
</tr>
<tr>
<td>1.25</td>
<td>2.007</td>
<td>2.008</td>
<td>2.009</td>
<td>2.010</td>
<td>2.011</td>
<td>2.012</td>
<td>2.013</td>
<td>2.014</td>
<td>2.015</td>
</tr>
<tr>
<td>1.30</td>
<td>2.048</td>
<td>2.049</td>
<td>2.050</td>
<td>2.051</td>
<td>2.052</td>
<td>2.053</td>
<td>2.054</td>
<td>2.055</td>
<td>2.056</td>
</tr>
<tr>
<td>1.35</td>
<td>2.089</td>
<td>2.090</td>
<td>2.091</td>
<td>2.092</td>
<td>2.093</td>
<td>2.094</td>
<td>2.095</td>
<td>2.096</td>
<td>2.097</td>
</tr>
<tr>
<td>1.40</td>
<td>2.130</td>
<td>2.131</td>
<td>2.132</td>
<td>2.133</td>
<td>2.134</td>
<td>2.135</td>
<td>2.136</td>
<td>2.137</td>
<td>2.138</td>
</tr>
<tr>
<td>1.45</td>
<td>2.171</td>
<td>2.172</td>
<td>2.173</td>
<td>2.174</td>
<td>2.175</td>
<td>2.176</td>
<td>2.177</td>
<td>2.178</td>
<td>2.179</td>
</tr>
<tr>
<td>1.50</td>
<td>2.212</td>
<td>2.213</td>
<td>2.214</td>
<td>2.215</td>
<td>2.216</td>
<td>2.217</td>
<td>2.218</td>
<td>2.219</td>
<td>2.220</td>
</tr>
<tr>
<td>1.55</td>
<td>2.253</td>
<td>2.254</td>
<td>2.255</td>
<td>2.256</td>
<td>2.257</td>
<td>2.258</td>
<td>2.259</td>
<td>2.260</td>
<td>2.261</td>
</tr>
<tr>
<td>1.60</td>
<td>2.294</td>
<td>2.295</td>
<td>2.296</td>
<td>2.297</td>
<td>2.298</td>
<td>2.299</td>
<td>2.300</td>
<td>2.301</td>
<td>2.302</td>
</tr>
<tr>
<td>1.65</td>
<td>2.335</td>
<td>2.336</td>
<td>2.337</td>
<td>2.338</td>
<td>2.339</td>
<td>2.340</td>
<td>2.341</td>
<td>2.342</td>
<td>2.343</td>
</tr>
<tr>
<td>1.70</td>
<td>2.376</td>
<td>2.377</td>
<td>2.378</td>
<td>2.379</td>
<td>2.380</td>
<td>2.381</td>
<td>2.382</td>
<td>2.383</td>
<td>2.384</td>
</tr>
<tr>
<td>1.75</td>
<td>2.417</td>
<td>2.418</td>
<td>2.419</td>
<td>2.420</td>
<td>2.421</td>
<td>2.422</td>
<td>2.423</td>
<td>2.424</td>
<td>2.425</td>
</tr>
<tr>
<td>1.80</td>
<td>2.458</td>
<td>2.459</td>
<td>2.460</td>
<td>2.461</td>
<td>2.462</td>
<td>2.463</td>
<td>2.464</td>
<td>2.465</td>
<td>2.466</td>
</tr>
<tr>
<td>1.85</td>
<td>2.500</td>
<td>2.501</td>
<td>2.502</td>
<td>2.503</td>
<td>2.504</td>
<td>2.505</td>
<td>2.506</td>
<td>2.507</td>
<td>2.508</td>
</tr>
<tr>
<td>1.90</td>
<td>2.541</td>
<td>2.542</td>
<td>2.543</td>
<td>2.544</td>
<td>2.545</td>
<td>2.546</td>
<td>2.547</td>
<td>2.548</td>
<td>2.549</td>
</tr>
<tr>
<td>1.95</td>
<td>2.583</td>
<td>2.584</td>
<td>2.585</td>
<td>2.586</td>
<td>2.587</td>
<td>2.588</td>
<td>2.589</td>
<td>2.590</td>
<td>2.591</td>
</tr>
<tr>
<td>2.00</td>
<td>2.624</td>
<td>2.625</td>
<td>2.626</td>
<td>2.627</td>
<td>2.628</td>
<td>2.629</td>
<td>2.630</td>
<td>2.631</td>
<td>2.632</td>
</tr>
</tbody>
</table>
### APPENDIX A

#### TABLE A2 (continued)

<table>
<thead>
<tr>
<th>R</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>1.0507</td>
<td>1.0508</td>
<td>1.0508</td>
<td>1.0508</td>
<td>1.0508</td>
<td>1.0509</td>
<td>1.0509</td>
<td>1.0509</td>
<td>1.0510</td>
</tr>
<tr>
<td>1.00</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
<td>1.0504</td>
</tr>
<tr>
<td>1.05</td>
<td>1.0528</td>
<td>1.0532</td>
<td>1.0536</td>
<td>1.0540</td>
<td>1.0544</td>
<td>1.0548</td>
<td>1.0552</td>
<td>1.0556</td>
<td>1.0560</td>
</tr>
<tr>
<td>1.10</td>
<td>1.0556</td>
<td>1.0560</td>
<td>1.0564</td>
<td>1.0567</td>
<td>1.0571</td>
<td>1.0575</td>
<td>1.0579</td>
<td>1.0583</td>
<td>1.0587</td>
</tr>
<tr>
<td>1.15</td>
<td>1.0586</td>
<td>1.0590</td>
<td>1.0594</td>
<td>1.0597</td>
<td>1.0601</td>
<td>1.0605</td>
<td>1.0609</td>
<td>1.0613</td>
<td>1.0617</td>
</tr>
<tr>
<td>1.20</td>
<td>1.0617</td>
<td>1.0621</td>
<td>1.0625</td>
<td>1.0629</td>
<td>1.0633</td>
<td>1.0637</td>
<td>1.0641</td>
<td>1.0645</td>
<td>1.0649</td>
</tr>
<tr>
<td>1.25</td>
<td>1.0649</td>
<td>1.0653</td>
<td>1.0657</td>
<td>1.0661</td>
<td>1.0665</td>
<td>1.0669</td>
<td>1.0673</td>
<td>1.0677</td>
<td>1.0681</td>
</tr>
<tr>
<td>1.30</td>
<td>1.0681</td>
<td>1.0685</td>
<td>1.0689</td>
<td>1.0693</td>
<td>1.0697</td>
<td>1.0701</td>
<td>1.0705</td>
<td>1.0709</td>
<td>1.0713</td>
</tr>
<tr>
<td>1.35</td>
<td>1.0713</td>
<td>1.0717</td>
<td>1.0721</td>
<td>1.0725</td>
<td>1.0729</td>
<td>1.0733</td>
<td>1.0737</td>
<td>1.0741</td>
<td>1.0745</td>
</tr>
<tr>
<td>1.40</td>
<td>1.0745</td>
<td>1.0749</td>
<td>1.0753</td>
<td>1.0757</td>
<td>1.0761</td>
<td>1.0765</td>
<td>1.0769</td>
<td>1.0773</td>
<td>1.0777</td>
</tr>
<tr>
<td>1.45</td>
<td>1.0777</td>
<td>1.0781</td>
<td>1.0785</td>
<td>1.0789</td>
<td>1.0793</td>
<td>1.0797</td>
<td>1.0801</td>
<td>1.0805</td>
<td>1.0809</td>
</tr>
<tr>
<td>1.50</td>
<td>1.0809</td>
<td>1.0813</td>
<td>1.0817</td>
<td>1.0821</td>
<td>1.0825</td>
<td>1.0829</td>
<td>1.0833</td>
<td>1.0837</td>
<td>1.0841</td>
</tr>
<tr>
<td>1.55</td>
<td>1.0841</td>
<td>1.0845</td>
<td>1.0849</td>
<td>1.0853</td>
<td>1.0857</td>
<td>1.0861</td>
<td>1.0865</td>
<td>1.0869</td>
<td>1.0873</td>
</tr>
<tr>
<td>1.60</td>
<td>1.0873</td>
<td>1.0877</td>
<td>1.0881</td>
<td>1.0885</td>
<td>1.0889</td>
<td>1.0893</td>
<td>1.0897</td>
<td>1.0901</td>
<td>1.0905</td>
</tr>
<tr>
<td>1.65</td>
<td>1.0905</td>
<td>1.0909</td>
<td>1.0913</td>
<td>1.0917</td>
<td>1.0921</td>
<td>1.0925</td>
<td>1.0929</td>
<td>1.0933</td>
<td>1.0937</td>
</tr>
<tr>
<td>1.70</td>
<td>1.0937</td>
<td>1.0941</td>
<td>1.0945</td>
<td>1.0950</td>
<td>1.0954</td>
<td>1.0958</td>
<td>1.0962</td>
<td>1.0966</td>
<td>1.0970</td>
</tr>
<tr>
<td>1.75</td>
<td>1.0970</td>
<td>1.0974</td>
<td>1.0978</td>
<td>1.0982</td>
<td>1.0986</td>
<td>1.0990</td>
<td>1.0994</td>
<td>1.0998</td>
<td>1.1002</td>
</tr>
<tr>
<td>1.80</td>
<td>1.1002</td>
<td>1.1006</td>
<td>1.1010</td>
<td>1.1014</td>
<td>1.1018</td>
<td>1.1022</td>
<td>1.1026</td>
<td>1.1030</td>
<td>1.1034</td>
</tr>
</tbody>
</table>

**Notes:**

- The table continues with similar entries for values of R ranging from 0.95 to 1.80.
- Each row represents the values for R at different intervals, typically increasing by 0.05.
- The values are likely to be related to a specific formula or calculation, though the exact nature is not specified in the image provided.
<table>
<thead>
<tr>
<th>x</th>
<th>190</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>750</th>
<th>1000</th>
<th>5000</th>
<th>50000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.0510</td>
<td>1.0510</td>
<td>1.0511</td>
<td>1.0511</td>
<td>1.0512</td>
<td>1.0512</td>
<td>1.0512</td>
<td>1.0512</td>
<td>1.0512</td>
</tr>
<tr>
<td>0.10</td>
<td>1.0495</td>
<td>1.0495</td>
<td>1.0496</td>
<td>1.0496</td>
<td>1.0497</td>
<td>1.0497</td>
<td>1.0498</td>
<td>1.0498</td>
<td>1.0498</td>
</tr>
<tr>
<td>0.15</td>
<td>1.0480</td>
<td>1.0480</td>
<td>1.0481</td>
<td>1.0482</td>
<td>1.0483</td>
<td>1.0483</td>
<td>1.0484</td>
<td>1.0484</td>
<td>1.0485</td>
</tr>
<tr>
<td>0.20</td>
<td>1.0466</td>
<td>1.0466</td>
<td>1.0467</td>
<td>1.0468</td>
<td>1.0469</td>
<td>1.0469</td>
<td>1.0470</td>
<td>1.0470</td>
<td>1.0471</td>
</tr>
<tr>
<td>0.25</td>
<td>1.0451</td>
<td>1.0451</td>
<td>1.0452</td>
<td>1.0453</td>
<td>1.0454</td>
<td>1.0454</td>
<td>1.0455</td>
<td>1.0455</td>
<td>1.0456</td>
</tr>
<tr>
<td>0.30</td>
<td>1.0437</td>
<td>1.0437</td>
<td>1.0438</td>
<td>1.0439</td>
<td>1.0440</td>
<td>1.0440</td>
<td>1.0441</td>
<td>1.0441</td>
<td>1.0442</td>
</tr>
<tr>
<td>0.35</td>
<td>1.0424</td>
<td>1.0424</td>
<td>1.0425</td>
<td>1.0426</td>
<td>1.0427</td>
<td>1.0427</td>
<td>1.0428</td>
<td>1.0428</td>
<td>1.0429</td>
</tr>
<tr>
<td>0.40</td>
<td>1.0411</td>
<td>1.0411</td>
<td>1.0412</td>
<td>1.0413</td>
<td>1.0414</td>
<td>1.0414</td>
<td>1.0415</td>
<td>1.0415</td>
<td>1.0416</td>
</tr>
<tr>
<td>0.45</td>
<td>1.0399</td>
<td>1.0399</td>
<td>1.0400</td>
<td>1.0401</td>
<td>1.0402</td>
<td>1.0402</td>
<td>1.0403</td>
<td>1.0403</td>
<td>1.0404</td>
</tr>
<tr>
<td>0.50</td>
<td>1.0387</td>
<td>1.0387</td>
<td>1.0388</td>
<td>1.0389</td>
<td>1.0390</td>
<td>1.0390</td>
<td>1.0391</td>
<td>1.0391</td>
<td>1.0392</td>
</tr>
<tr>
<td>0.55</td>
<td>1.0375</td>
<td>1.0375</td>
<td>1.0376</td>
<td>1.0377</td>
<td>1.0378</td>
<td>1.0378</td>
<td>1.0379</td>
<td>1.0380</td>
<td>1.0381</td>
</tr>
<tr>
<td>0.60</td>
<td>1.0364</td>
<td>1.0364</td>
<td>1.0365</td>
<td>1.0366</td>
<td>1.0367</td>
<td>1.0367</td>
<td>1.0368</td>
<td>1.0369</td>
<td>1.0370</td>
</tr>
<tr>
<td>0.65</td>
<td>1.0353</td>
<td>1.0353</td>
<td>1.0354</td>
<td>1.0355</td>
<td>1.0356</td>
<td>1.0356</td>
<td>1.0357</td>
<td>1.0358</td>
<td>1.0359</td>
</tr>
<tr>
<td>0.70</td>
<td>1.0342</td>
<td>1.0342</td>
<td>1.0343</td>
<td>1.0344</td>
<td>1.0345</td>
<td>1.0345</td>
<td>1.0346</td>
<td>1.0347</td>
<td>1.0348</td>
</tr>
<tr>
<td>0.75</td>
<td>1.0331</td>
<td>1.0331</td>
<td>1.0332</td>
<td>1.0333</td>
<td>1.0334</td>
<td>1.0334</td>
<td>1.0335</td>
<td>1.0336</td>
<td>1.0337</td>
</tr>
<tr>
<td>0.80</td>
<td>1.0320</td>
<td>1.0320</td>
<td>1.0321</td>
<td>1.0322</td>
<td>1.0323</td>
<td>1.0323</td>
<td>1.0324</td>
<td>1.0325</td>
<td>1.0326</td>
</tr>
<tr>
<td>0.85</td>
<td>1.0309</td>
<td>1.0309</td>
<td>1.0310</td>
<td>1.0311</td>
<td>1.0312</td>
<td>1.0312</td>
<td>1.0313</td>
<td>1.0314</td>
<td>1.0315</td>
</tr>
<tr>
<td>0.90</td>
<td>1.0298</td>
<td>1.0298</td>
<td>1.0299</td>
<td>1.0300</td>
<td>1.0301</td>
<td>1.0301</td>
<td>1.0302</td>
<td>1.0303</td>
<td>1.0304</td>
</tr>
<tr>
<td>0.95</td>
<td>1.0287</td>
<td>1.0287</td>
<td>1.0288</td>
<td>1.0289</td>
<td>1.0290</td>
<td>1.0290</td>
<td>1.0291</td>
<td>1.0292</td>
<td>1.0293</td>
</tr>
<tr>
<td>1.00</td>
<td>1.0276</td>
<td>1.0276</td>
<td>1.0277</td>
<td>1.0278</td>
<td>1.0279</td>
<td>1.0279</td>
<td>1.0280</td>
<td>1.0281</td>
<td>1.0282</td>
</tr>
</tbody>
</table>

TABLE A2 (continued)
### TABLE A3. THE FUNCTION $x_n(t)$

<table>
<thead>
<tr>
<th>$t$</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.10</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>0.15</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
</tr>
<tr>
<td>0.20</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>0.25</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td>0.30</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td>0.35</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>0.40</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>0.45</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
<td>0.062</td>
</tr>
<tr>
<td>0.50</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
<td>0.071</td>
</tr>
</tbody>
</table>

### TABLE A3 (continued)

<table>
<thead>
<tr>
<th>$n$</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
</tr>
<tr>
<td>0.10</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>0.15</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
</tr>
<tr>
<td>0.20</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
</tr>
<tr>
<td>0.25</td>
<td>0.117</td>
<td>0.117</td>
<td>0.117</td>
<td>0.117</td>
<td>0.117</td>
<td>0.117</td>
<td>0.117</td>
<td>0.117</td>
</tr>
<tr>
<td>0.30</td>
<td>0.139</td>
<td>0.139</td>
<td>0.139</td>
<td>0.139</td>
<td>0.139</td>
<td>0.139</td>
<td>0.139</td>
<td>0.139</td>
</tr>
<tr>
<td>0.35</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
<td>0.162</td>
</tr>
<tr>
<td>0.40</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
</tr>
<tr>
<td>0.45</td>
<td>0.211</td>
<td>0.211</td>
<td>0.211</td>
<td>0.211</td>
<td>0.211</td>
<td>0.211</td>
<td>0.211</td>
<td>0.211</td>
</tr>
<tr>
<td>0.50</td>
<td>0.238</td>
<td>0.238</td>
<td>0.238</td>
<td>0.238</td>
<td>0.238</td>
<td>0.238</td>
<td>0.238</td>
<td>0.238</td>
</tr>
</tbody>
</table>
## APPENDIX A

### TABLE A3 (continued)

<table>
<thead>
<tr>
<th>( \times )</th>
<th>190</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>750</th>
<th>1000</th>
<th>5000</th>
<th>( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.0538</td>
<td>0.0538</td>
<td>0.0538</td>
<td>0.0539</td>
<td>0.0539</td>
<td>0.0539</td>
<td>0.0539</td>
<td>0.0539</td>
<td>0.0539</td>
</tr>
<tr>
<td>0.10</td>
<td>0.1159</td>
<td>0.1159</td>
<td>0.1160</td>
<td>0.1161</td>
<td>0.1161</td>
<td>0.1162</td>
<td>0.1162</td>
<td>0.1162</td>
<td>0.1162</td>
</tr>
<tr>
<td>0.15</td>
<td>0.1872</td>
<td>0.1873</td>
<td>0.1875</td>
<td>0.1876</td>
<td>0.1877</td>
<td>0.1878</td>
<td>0.1879</td>
<td>0.1880</td>
<td>0.1880</td>
</tr>
<tr>
<td>0.20</td>
<td>0.2689</td>
<td>0.2690</td>
<td>0.2694</td>
<td>0.2697</td>
<td>0.2697</td>
<td>0.2700</td>
<td>0.2701</td>
<td>0.2704</td>
<td>0.2704</td>
</tr>
<tr>
<td>0.25</td>
<td>0.3620</td>
<td>0.3622</td>
<td>0.3630</td>
<td>0.3634</td>
<td>0.3635</td>
<td>0.3640</td>
<td>0.3642</td>
<td>0.3646</td>
<td>0.3647</td>
</tr>
<tr>
<td>0.30</td>
<td>0.4680</td>
<td>0.4682</td>
<td>0.4696</td>
<td>0.4702</td>
<td>0.4705</td>
<td>0.4712</td>
<td>0.4714</td>
<td>0.4721</td>
<td>0.4723</td>
</tr>
<tr>
<td>0.35</td>
<td>0.5683</td>
<td>0.5687</td>
<td>0.5696</td>
<td>0.5716</td>
<td>0.5720</td>
<td>0.5731</td>
<td>0.5735</td>
<td>0.5744</td>
<td>0.5747</td>
</tr>
<tr>
<td>0.40</td>
<td>0.6796</td>
<td>0.6799</td>
<td>0.6829</td>
<td>0.6852</td>
<td>0.6861</td>
<td>0.6880</td>
<td>0.6888</td>
<td>0.6908</td>
<td>0.6913</td>
</tr>
<tr>
<td>0.45</td>
<td>0.7952</td>
<td>1.0331</td>
<td>1.0385</td>
<td>1.0612</td>
<td>1.0625</td>
<td>1.0651</td>
<td>1.0666</td>
<td>1.0689</td>
<td>1.0696</td>
</tr>
<tr>
<td>0.50</td>
<td>1.2478</td>
<td>1.2490</td>
<td>1.2561</td>
<td>1.2598</td>
<td>1.2616</td>
<td>1.2649</td>
<td>1.2664</td>
<td>1.2700</td>
<td>1.2709</td>
</tr>
<tr>
<td>0.55</td>
<td>1.4078</td>
<td>1.4092</td>
<td>1.4176</td>
<td>1.4234</td>
<td>1.4258</td>
<td>1.4302</td>
<td>1.4321</td>
<td>1.4368</td>
<td>1.4398</td>
</tr>
<tr>
<td>0.60</td>
<td>1.7147</td>
<td>1.7166</td>
<td>1.7287</td>
<td>1.7349</td>
<td>1.7390</td>
<td>1.7436</td>
<td>1.7461</td>
<td>1.7522</td>
<td>1.7538</td>
</tr>
<tr>
<td>0.65</td>
<td>2.0215</td>
<td>1.9939</td>
<td>2.0904</td>
<td>2.0173</td>
<td>2.0214</td>
<td>2.0265</td>
<td>2.0317</td>
<td>2.0355</td>
<td>2.0414</td>
</tr>
<tr>
<td>0.70</td>
<td>2.3098</td>
<td>2.3046</td>
<td>2.3442</td>
<td>2.3341</td>
<td>2.3395</td>
<td>2.3483</td>
<td>2.3523</td>
<td>2.3592</td>
<td>2.3647</td>
</tr>
<tr>
<td>0.75</td>
<td>2.6484</td>
<td>2.6522</td>
<td>2.6767</td>
<td>2.6892</td>
<td>2.6959</td>
<td>2.7069</td>
<td>2.7120</td>
<td>2.7244</td>
<td>2.7275</td>
</tr>
<tr>
<td>0.80</td>
<td>3.0200</td>
<td>3.0408</td>
<td>3.0711</td>
<td>3.0806</td>
<td>3.0951</td>
<td>3.1066</td>
<td>3.1159</td>
<td>3.1204</td>
<td>3.1343</td>
</tr>
<tr>
<td>0.90</td>
<td>3.9515</td>
<td>3.9386</td>
<td>4.0044</td>
<td>4.0278</td>
<td>4.0410</td>
<td>4.0612</td>
<td>4.0709</td>
<td>4.0913</td>
<td>4.1002</td>
</tr>
<tr>
<td>0.95</td>
<td>4.4094</td>
<td>4.4090</td>
<td>4.5539</td>
<td>4.5824</td>
<td>4.5986</td>
<td>4.6231</td>
<td>4.6349</td>
<td>4.6638</td>
<td>4.6708</td>
</tr>
</tbody>
</table>
### APPENDIX A

**Table A4. The Solution of the Equation** $u^3 + 3u = k$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$u$</th>
<th>$u^3$</th>
<th>$k$</th>
<th>$u$</th>
<th>$u^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2</td>
<td>0.6666</td>
<td>0.4444</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1236</td>
<td>0.0174</td>
<td>0.6</td>
<td>0.1974</td>
<td>0.3822</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2662</td>
<td>0.6598</td>
<td>1.0</td>
<td>0.3222</td>
<td>0.9888</td>
</tr>
<tr>
<td>1.4</td>
<td>0.4386</td>
<td>0.7590</td>
<td>1.6</td>
<td>0.4033</td>
<td>0.2178</td>
</tr>
<tr>
<td>1.8</td>
<td>0.5456</td>
<td>0.2607</td>
<td>2.0</td>
<td>0.5961</td>
<td>0.3040</td>
</tr>
<tr>
<td>2.2</td>
<td>0.6442</td>
<td>0.3471</td>
<td>2.4</td>
<td>0.6903</td>
<td>0.3927</td>
</tr>
<tr>
<td>2.6</td>
<td>0.7246</td>
<td>0.4315</td>
<td>2.8</td>
<td>0.7770</td>
<td>0.4723</td>
</tr>
<tr>
<td>3.0</td>
<td>0.8177</td>
<td>0.5120</td>
<td>3.2</td>
<td>0.8569</td>
<td>0.5506</td>
</tr>
<tr>
<td>3.4</td>
<td>0.8946</td>
<td>0.5880</td>
<td>3.6</td>
<td>0.9310</td>
<td>0.6242</td>
</tr>
<tr>
<td>3.8</td>
<td>0.9661</td>
<td>0.6593</td>
<td>4.0</td>
<td>1.0000</td>
<td>0.6931</td>
</tr>
<tr>
<td>4.2</td>
<td>1.0328</td>
<td>0.7272</td>
<td>4.4</td>
<td>1.0645</td>
<td>0.7576</td>
</tr>
<tr>
<td>4.6</td>
<td>1.0953</td>
<td>0.7883</td>
<td>4.8</td>
<td>1.1252</td>
<td>0.8180</td>
</tr>
<tr>
<td>5.0</td>
<td>1.1542</td>
<td>0.8468</td>
<td>5.2</td>
<td>1.1824</td>
<td>0.8746</td>
</tr>
<tr>
<td>5.4</td>
<td>1.2058</td>
<td>0.9016</td>
<td>5.6</td>
<td>1.2355</td>
<td>0.9270</td>
</tr>
<tr>
<td>5.8</td>
<td>1.2625</td>
<td>0.9532</td>
<td>6.0</td>
<td>1.2879</td>
<td>0.9738</td>
</tr>
<tr>
<td>6.2</td>
<td>1.3127</td>
<td>1.0018</td>
<td>6.4</td>
<td>1.3369</td>
<td>1.0251</td>
</tr>
<tr>
<td>6.6</td>
<td>1.3695</td>
<td>1.0477</td>
<td>6.8</td>
<td>1.3937</td>
<td>1.0707</td>
</tr>
</tbody>
</table>
### Table A5. The Normal Integral $P$, the Ordinate $Z$ and the Elasticity $Z/P$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$P(Y)$</th>
<th>$Z(Y)$</th>
<th>$Z(P)$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-4.0$</td>
<td>0.00003</td>
<td>4.225</td>
<td>0.00003</td>
<td>0.99997</td>
</tr>
<tr>
<td>$-3.9$</td>
<td>0.00005</td>
<td>4.131</td>
<td>0.00002</td>
<td>0.99995</td>
</tr>
<tr>
<td>$-3.8$</td>
<td>0.00007</td>
<td>4.036</td>
<td>0.00002</td>
<td>0.99993</td>
</tr>
<tr>
<td>$-3.7$</td>
<td>0.00011</td>
<td>3.940</td>
<td>0.00002</td>
<td>0.99991</td>
</tr>
<tr>
<td>$-3.6$</td>
<td>0.00016</td>
<td>3.846</td>
<td>0.00001</td>
<td>0.99989</td>
</tr>
<tr>
<td>$-3.5$</td>
<td>0.00023</td>
<td>3.751</td>
<td>0.00001</td>
<td>0.99987</td>
</tr>
<tr>
<td>$-3.4$</td>
<td>0.00034</td>
<td>3.653</td>
<td>0.00001</td>
<td>0.99986</td>
</tr>
<tr>
<td>$-3.3$</td>
<td>0.00046</td>
<td>3.564</td>
<td>0.00001</td>
<td>0.99985</td>
</tr>
<tr>
<td>$-3.2$</td>
<td>0.00069</td>
<td>3.472</td>
<td>0.00001</td>
<td>0.99984</td>
</tr>
<tr>
<td>$-3.1$</td>
<td>0.00097</td>
<td>3.376</td>
<td>0.00001</td>
<td>0.99983</td>
</tr>
<tr>
<td>$-3.0$</td>
<td>0.00135</td>
<td>3.283</td>
<td>0.00001</td>
<td>0.99982</td>
</tr>
<tr>
<td>$-2.9$</td>
<td>0.00187</td>
<td>3.191</td>
<td>0.00001</td>
<td>0.99980</td>
</tr>
<tr>
<td>$-2.8$</td>
<td>0.00256</td>
<td>3.066</td>
<td>0.00001</td>
<td>0.99978</td>
</tr>
<tr>
<td>$-2.7$</td>
<td>0.00347</td>
<td>2.990</td>
<td>0.00001</td>
<td>0.99976</td>
</tr>
<tr>
<td>$-2.6$</td>
<td>0.00466</td>
<td>2.914</td>
<td>0.00001</td>
<td>0.99974</td>
</tr>
<tr>
<td>$-2.5$</td>
<td>0.00611</td>
<td>2.832</td>
<td>0.00001</td>
<td>0.99972</td>
</tr>
<tr>
<td>$-2.4$</td>
<td>0.00810</td>
<td>2.731</td>
<td>0.00001</td>
<td>0.99969</td>
</tr>
<tr>
<td>$-2.3$</td>
<td>0.01072</td>
<td>2.644</td>
<td>0.00001</td>
<td>0.99967</td>
</tr>
<tr>
<td>$-2.2$</td>
<td>0.01390</td>
<td>2.552</td>
<td>0.00001</td>
<td>0.99964</td>
</tr>
<tr>
<td>$-2.1$</td>
<td>0.01786</td>
<td>2.462</td>
<td>0.00001</td>
<td>0.99961</td>
</tr>
<tr>
<td>$-2.0$</td>
<td>0.02275</td>
<td>2.373</td>
<td>0.00001</td>
<td>0.99958</td>
</tr>
<tr>
<td>$-1.9$</td>
<td>0.02872</td>
<td>2.285</td>
<td>0.00001</td>
<td>0.99955</td>
</tr>
<tr>
<td>$-1.8$</td>
<td>0.03593</td>
<td>2.197</td>
<td>0.00001</td>
<td>0.99952</td>
</tr>
<tr>
<td>$-1.7$</td>
<td>0.04426</td>
<td>2.110</td>
<td>0.00001</td>
<td>0.99948</td>
</tr>
<tr>
<td>$-1.6$</td>
<td>0.05460</td>
<td>2.024</td>
<td>0.00002</td>
<td>0.99944</td>
</tr>
<tr>
<td>$-1.5$</td>
<td>0.06601</td>
<td>1.939</td>
<td>0.00002</td>
<td>0.99940</td>
</tr>
<tr>
<td>$-1.4$</td>
<td>0.07876</td>
<td>1.854</td>
<td>0.00002</td>
<td>0.99935</td>
</tr>
<tr>
<td>$-1.3$</td>
<td>0.09260</td>
<td>1.770</td>
<td>0.00002</td>
<td>0.99931</td>
</tr>
<tr>
<td>$-1.2$</td>
<td>0.10757</td>
<td>1.687</td>
<td>0.00002</td>
<td>0.99926</td>
</tr>
<tr>
<td>$-1.1$</td>
<td>0.12367</td>
<td>1.606</td>
<td>0.00002</td>
<td>0.99921</td>
</tr>
<tr>
<td>$-1.0$</td>
<td>0.14086</td>
<td>1.525</td>
<td>0.00002</td>
<td>0.99916</td>
</tr>
<tr>
<td>$-0.9$</td>
<td>0.16016</td>
<td>1.446</td>
<td>0.00002</td>
<td>0.99911</td>
</tr>
<tr>
<td>$-0.8$</td>
<td>0.18116</td>
<td>1.370</td>
<td>0.00002</td>
<td>0.99906</td>
</tr>
<tr>
<td>$-0.7$</td>
<td>0.20116</td>
<td>1.299</td>
<td>0.00002</td>
<td>0.99900</td>
</tr>
<tr>
<td>$-0.6$</td>
<td>0.23745</td>
<td>1.235</td>
<td>0.00002</td>
<td>0.99894</td>
</tr>
<tr>
<td>$-0.5$</td>
<td>0.29854</td>
<td>1.141</td>
<td>0.00002</td>
<td>0.99888</td>
</tr>
<tr>
<td>$-0.4$</td>
<td>0.33458</td>
<td>1.009</td>
<td>0.00002</td>
<td>0.99882</td>
</tr>
<tr>
<td>$-0.3$</td>
<td>0.38249</td>
<td>0.968</td>
<td>0.00002</td>
<td>0.99876</td>
</tr>
<tr>
<td>$-0.2$</td>
<td>0.42074</td>
<td>0.929</td>
<td>0.00002</td>
<td>0.99870</td>
</tr>
<tr>
<td>$-0.1$</td>
<td>0.46107</td>
<td>0.893</td>
<td>0.00002</td>
<td>0.99864</td>
</tr>
<tr>
<td>$0.0$</td>
<td>0.50000</td>
<td>0.791</td>
<td>0.00002</td>
<td>0.99858</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Z(Y)$</th>
<th>$Z(P)$</th>
<th>$P(Y)$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.798</td>
<td>0.50000</td>
<td>0.3984</td>
<td>$0.798$</td>
</tr>
</tbody>
</table>

$Z(Y)$ is the ordinate corresponding to $Y$. $Z(P)$ is the elasticity of $Z$ with respect to $P$. $P(Y)$ is the integral of $P$ from $Y$ to $0$.
### Table A6. Factors for a Quantal Probit Analysis

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$Y - P_Z$</th>
<th>$1 Z$</th>
<th>$W = Z^P PQ$</th>
<th>$Y - P Z$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-4.0$</td>
<td>-4.2367</td>
<td>74.72</td>
<td>0.09056</td>
<td>-74.68</td>
<td>4.0</td>
</tr>
<tr>
<td>$-3.9$</td>
<td>-4.1421</td>
<td>70.34</td>
<td>0.0882</td>
<td>-70.30</td>
<td>3.9</td>
</tr>
<tr>
<td>$-3.8$</td>
<td>-4.0478</td>
<td>65.25</td>
<td>0.08618</td>
<td>-65.21</td>
<td>3.8</td>
</tr>
<tr>
<td>$-3.7$</td>
<td>-3.9536</td>
<td>60.54</td>
<td>0.08467</td>
<td>-60.50</td>
<td>3.7</td>
</tr>
<tr>
<td>$-3.6$</td>
<td>-3.8600</td>
<td>55.00</td>
<td>0.08235</td>
<td>-55.00</td>
<td>3.6</td>
</tr>
<tr>
<td>$-3.5$</td>
<td>-3.7665</td>
<td>49.60</td>
<td>0.08027</td>
<td>-49.62</td>
<td>3.5</td>
</tr>
<tr>
<td>$-3.4$</td>
<td>-3.6734</td>
<td>44.31</td>
<td>0.07832</td>
<td>-44.30</td>
<td>3.4</td>
</tr>
<tr>
<td>$-3.3$</td>
<td>-3.5806</td>
<td>39.15</td>
<td>0.07641</td>
<td>-39.15</td>
<td>3.3</td>
</tr>
<tr>
<td>$-3.2$</td>
<td>-3.4882</td>
<td>34.04</td>
<td>0.07463</td>
<td>-34.04</td>
<td>3.2</td>
</tr>
<tr>
<td>$-3.1$</td>
<td>-3.3962</td>
<td>29.04</td>
<td>0.07297</td>
<td>-29.04</td>
<td>3.1</td>
</tr>
<tr>
<td>$-3.0$</td>
<td>-3.3046</td>
<td>24.14</td>
<td>0.07139</td>
<td>-24.14</td>
<td>3.0</td>
</tr>
<tr>
<td>$-2.9$</td>
<td>-3.2134</td>
<td>19.34</td>
<td>0.07083</td>
<td>-19.34</td>
<td>2.9</td>
</tr>
<tr>
<td>$-2.8$</td>
<td>-3.1220</td>
<td>14.64</td>
<td>0.07030</td>
<td>-14.64</td>
<td>2.8</td>
</tr>
<tr>
<td>$-2.7$</td>
<td>-3.0307</td>
<td>10.03</td>
<td>0.06980</td>
<td>-10.03</td>
<td>2.7</td>
</tr>
<tr>
<td>$-2.6$</td>
<td>-2.9432</td>
<td>5.53</td>
<td>0.06932</td>
<td>-5.53</td>
<td>2.6</td>
</tr>
<tr>
<td>$-2.5$</td>
<td>-2.8543</td>
<td>1.15</td>
<td>0.06887</td>
<td>-1.15</td>
<td>2.5</td>
</tr>
<tr>
<td>$-2.4$</td>
<td>-2.7655</td>
<td>0.85</td>
<td>0.06844</td>
<td>-0.85</td>
<td>2.4</td>
</tr>
<tr>
<td>$-2.3$</td>
<td>-2.6768</td>
<td>0.50</td>
<td>0.06804</td>
<td>-0.50</td>
<td>2.3</td>
</tr>
<tr>
<td>$-2.2$</td>
<td>-2.5879</td>
<td>0.18</td>
<td>0.06766</td>
<td>-0.18</td>
<td>2.2</td>
</tr>
<tr>
<td>$-2.1$</td>
<td>-2.4989</td>
<td>0.00</td>
<td>0.06730</td>
<td>-0.00</td>
<td>2.1</td>
</tr>
<tr>
<td>$-2.0$</td>
<td>-2.4099</td>
<td>0.00</td>
<td>0.06705</td>
<td>-0.00</td>
<td>2.0</td>
</tr>
<tr>
<td>$-1.9$</td>
<td>-2.3215</td>
<td>0.00</td>
<td>0.06674</td>
<td>-0.00</td>
<td>1.9</td>
</tr>
<tr>
<td>$-1.8$</td>
<td>-2.2328</td>
<td>0.00</td>
<td>0.06644</td>
<td>-0.00</td>
<td>1.8</td>
</tr>
<tr>
<td>$-1.7$</td>
<td>-2.1452</td>
<td>0.00</td>
<td>0.06615</td>
<td>-0.00</td>
<td>1.7</td>
</tr>
<tr>
<td>$-1.6$</td>
<td>-2.0579</td>
<td>0.00</td>
<td>0.06587</td>
<td>-0.00</td>
<td>1.6</td>
</tr>
<tr>
<td>$-1.5$</td>
<td>-1.9705</td>
<td>0.00</td>
<td>0.06560</td>
<td>-0.00</td>
<td>1.5</td>
</tr>
<tr>
<td>$-1.4$</td>
<td>-1.8839</td>
<td>0.00</td>
<td>0.06535</td>
<td>-0.00</td>
<td>1.4</td>
</tr>
<tr>
<td>$-1.3$</td>
<td>-1.7972</td>
<td>0.00</td>
<td>0.06511</td>
<td>-0.00</td>
<td>1.3</td>
</tr>
<tr>
<td>$-1.2$</td>
<td>-1.7104</td>
<td>0.00</td>
<td>0.06488</td>
<td>-0.00</td>
<td>1.2</td>
</tr>
<tr>
<td>$-1.1$</td>
<td>-1.6233</td>
<td>0.00</td>
<td>0.06464</td>
<td>-0.00</td>
<td>1.1</td>
</tr>
<tr>
<td>$-1.0$</td>
<td>-1.5367</td>
<td>0.00</td>
<td>0.06444</td>
<td>-0.00</td>
<td>1.0</td>
</tr>
<tr>
<td>$-0.9$</td>
<td>-1.4501</td>
<td>0.00</td>
<td>0.06415</td>
<td>-0.00</td>
<td>0.9</td>
</tr>
<tr>
<td>$-0.8$</td>
<td>-1.3635</td>
<td>0.00</td>
<td>0.06388</td>
<td>-0.00</td>
<td>0.8</td>
</tr>
<tr>
<td>$-0.7$</td>
<td>-1.2769</td>
<td>0.00</td>
<td>0.06362</td>
<td>-0.00</td>
<td>0.7</td>
</tr>
<tr>
<td>$-0.6$</td>
<td>-1.1902</td>
<td>0.00</td>
<td>0.06337</td>
<td>-0.00</td>
<td>0.6</td>
</tr>
<tr>
<td>$-0.5$</td>
<td>-1.0035</td>
<td>0.00</td>
<td>0.06313</td>
<td>-0.00</td>
<td>0.5</td>
</tr>
<tr>
<td>$-0.4$</td>
<td>-0.8168</td>
<td>0.00</td>
<td>0.06289</td>
<td>-0.00</td>
<td>0.4</td>
</tr>
<tr>
<td>$-0.3$</td>
<td>-0.6290</td>
<td>0.00</td>
<td>0.06266</td>
<td>-0.00</td>
<td>0.3</td>
</tr>
<tr>
<td>$-0.2$</td>
<td>-0.4413</td>
<td>0.00</td>
<td>0.06244</td>
<td>-0.00</td>
<td>0.2</td>
</tr>
<tr>
<td>$-0.1$</td>
<td>-0.2535</td>
<td>0.00</td>
<td>0.06222</td>
<td>-0.00</td>
<td>0.1</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0655</td>
<td>0.00</td>
<td>0.06202</td>
<td>-0.00</td>
<td>0.0</td>
</tr>
</tbody>
</table>
### TABLE A7. FACTORS FOR A HOMOSCEDASTIC QUANTITATIVE PROBIT ANALYSIS

<table>
<thead>
<tr>
<th>( Y )</th>
<th>( Y - Z )</th>
<th>( x = P / Z )</th>
<th>( 1 / Z )</th>
<th>( w = Z )</th>
<th>( Y - P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.0</td>
<td>-4.2957</td>
<td>0.9267</td>
<td>7472</td>
<td>0.00000</td>
<td>7471.9</td>
</tr>
<tr>
<td>-3.9</td>
<td>-4.1424</td>
<td>0.9241</td>
<td>5034</td>
<td>0.00000</td>
<td>5035.6</td>
</tr>
<tr>
<td>-3.8</td>
<td>-4.0478</td>
<td>0.9278</td>
<td>7425</td>
<td>0.00000</td>
<td>7425.0</td>
</tr>
<tr>
<td>-3.7</td>
<td>-3.9353</td>
<td>0.9238</td>
<td>2354</td>
<td>0.00000</td>
<td>2353.9</td>
</tr>
<tr>
<td>-3.6</td>
<td>-3.8660</td>
<td>0.9200</td>
<td>1354</td>
<td>0.00000</td>
<td>1354.0</td>
</tr>
<tr>
<td>-3.5</td>
<td>-3.7665</td>
<td>0.9166</td>
<td>1146</td>
<td>0.00000</td>
<td>1145.6</td>
</tr>
<tr>
<td>-3.4</td>
<td>-3.6734</td>
<td>0.9134</td>
<td>8115</td>
<td>0.00000</td>
<td>8115.7</td>
</tr>
<tr>
<td>-3.3</td>
<td>-3.5806</td>
<td>0.9086</td>
<td>5708</td>
<td>0.00000</td>
<td>5708.2</td>
</tr>
<tr>
<td>-3.2</td>
<td>-3.4882</td>
<td>0.9022</td>
<td>4194</td>
<td>0.00000</td>
<td>4193.6</td>
</tr>
<tr>
<td>-3.1</td>
<td>-3.3962</td>
<td>0.8952</td>
<td>3061</td>
<td>0.00000</td>
<td>3058.1</td>
</tr>
<tr>
<td>-3.0</td>
<td>-3.3046</td>
<td>0.8876</td>
<td>2256</td>
<td>0.00000</td>
<td>2253.3</td>
</tr>
<tr>
<td>-2.9</td>
<td>-3.2034</td>
<td>0.8790</td>
<td>16830</td>
<td>0.00000</td>
<td>16787.2</td>
</tr>
<tr>
<td>-2.8</td>
<td>-2.9226</td>
<td>0.8528</td>
<td>12534</td>
<td>0.00000</td>
<td>12502.0</td>
</tr>
<tr>
<td>-2.7</td>
<td>-2.7577</td>
<td>0.8337</td>
<td>9590</td>
<td>0.00000</td>
<td>9586.8</td>
</tr>
<tr>
<td>-2.6</td>
<td>-2.5453</td>
<td>0.8062</td>
<td>7562</td>
<td>0.00000</td>
<td>7372.6</td>
</tr>
<tr>
<td>-2.5</td>
<td>-2.4513</td>
<td>0.7513</td>
<td>5795</td>
<td>0.00011</td>
<td>5656.1</td>
</tr>
<tr>
<td>-2.4</td>
<td>-2.3760</td>
<td>0.7061</td>
<td>44634</td>
<td>0.00050</td>
<td>44280.1</td>
</tr>
<tr>
<td>-2.3</td>
<td>-2.2786</td>
<td>0.6516</td>
<td>33302</td>
<td>0.00080</td>
<td>32492.2</td>
</tr>
<tr>
<td>-2.2</td>
<td>-2.0919</td>
<td>0.4391</td>
<td>28189</td>
<td>0.00126</td>
<td>27797.4</td>
</tr>
<tr>
<td>-2.1</td>
<td>-2.0402</td>
<td>0.4062</td>
<td>22738</td>
<td>0.00193</td>
<td>22330.9</td>
</tr>
<tr>
<td>-2.0</td>
<td>-2.0214</td>
<td>0.4214</td>
<td>18522</td>
<td>0.00292</td>
<td>18100.2</td>
</tr>
<tr>
<td>-1.9</td>
<td>-1.9370</td>
<td>0.4376</td>
<td>15240</td>
<td>0.00431</td>
<td>14802.6</td>
</tr>
<tr>
<td>-1.8</td>
<td>-1.8751</td>
<td>0.4551</td>
<td>12066</td>
<td>0.00823</td>
<td>11611.4</td>
</tr>
<tr>
<td>-1.7</td>
<td>-1.7739</td>
<td>0.4759</td>
<td>10633</td>
<td>0.00885</td>
<td>10365.4</td>
</tr>
<tr>
<td>-1.6</td>
<td>-1.6940</td>
<td>0.4940</td>
<td>9015</td>
<td>0.01230</td>
<td>8832.4</td>
</tr>
<tr>
<td>-1.5</td>
<td>-1.6158</td>
<td>0.5158</td>
<td>7721</td>
<td>0.01677</td>
<td>7570.1</td>
</tr>
<tr>
<td>-1.4</td>
<td>-1.5390</td>
<td>0.5394</td>
<td>65888</td>
<td>0.02422</td>
<td>64194.4</td>
</tr>
<tr>
<td>-1.3</td>
<td>-1.4668</td>
<td>0.5649</td>
<td>58151</td>
<td>0.02937</td>
<td>56705.7</td>
</tr>
<tr>
<td>-1.2</td>
<td>-1.3986</td>
<td>0.5946</td>
<td>51407</td>
<td>0.03771</td>
<td>50167.1</td>
</tr>
<tr>
<td>-1.1</td>
<td>-1.3227</td>
<td>0.6227</td>
<td>45593</td>
<td>0.04266</td>
<td>44675.5</td>
</tr>
<tr>
<td>-1.0</td>
<td>-1.1557</td>
<td>0.6557</td>
<td>41387</td>
<td>0.05385</td>
<td>40922.1</td>
</tr>
<tr>
<td>-0.9</td>
<td>-1.1167</td>
<td>0.6837</td>
<td>37712</td>
<td>0.07080</td>
<td>37185.5</td>
</tr>
<tr>
<td>-0.8</td>
<td>-1.0531</td>
<td>0.7133</td>
<td>34519</td>
<td>0.09392</td>
<td>34080.2</td>
</tr>
<tr>
<td>-0.7</td>
<td>-0.9749</td>
<td>0.7479</td>
<td>32025</td>
<td>0.09750</td>
<td>31627.5</td>
</tr>
<tr>
<td>-0.6</td>
<td>-0.9125</td>
<td>0.7830</td>
<td>30010</td>
<td>0.11104</td>
<td>29680.0</td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.8624</td>
<td>0.8176</td>
<td>28404</td>
<td>0.12395</td>
<td>28060.5</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.8167</td>
<td>0.8537</td>
<td>27154</td>
<td>0.13586</td>
<td>26776.0</td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.7718</td>
<td>0.8918</td>
<td>26220</td>
<td>0.14518</td>
<td>25802.2</td>
</tr>
<tr>
<td>-0.2</td>
<td>-0.7259</td>
<td>0.9259</td>
<td>25573</td>
<td>0.14921</td>
<td>25481.6</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.6893</td>
<td>0.9593</td>
<td>25162</td>
<td>0.15757</td>
<td>25099.0</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.2533</td>
<td>1.12533</td>
<td>25666</td>
<td>0.15915</td>
<td>25533.0</td>
</tr>
</tbody>
</table>
### Table A8. Factors for a Heteroscedastic Quantitative Probit Analysis

<table>
<thead>
<tr>
<th>Y</th>
<th>$Y - Z \log P$</th>
<th>$x^* = \frac{P}{Z}$</th>
<th>$w = \frac{Z^*}{\mu}$</th>
<th>$w^* = \frac{Z^*}{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.0</td>
<td>-1.5483</td>
<td>0.2367</td>
<td>17.8256</td>
<td>17.8256</td>
</tr>
<tr>
<td>-4.9</td>
<td>-1.4929</td>
<td>0.2218</td>
<td>17.0796</td>
<td>17.0796</td>
</tr>
<tr>
<td>-4.8</td>
<td>-1.4374</td>
<td>0.2078</td>
<td>16.3237</td>
<td>16.3237</td>
</tr>
<tr>
<td>-4.7</td>
<td>-1.3817</td>
<td>0.1938</td>
<td>15.5672</td>
<td>15.5672</td>
</tr>
<tr>
<td>-4.6</td>
<td>-1.3259</td>
<td>0.1799</td>
<td>14.8103</td>
<td>14.8103</td>
</tr>
<tr>
<td>-4.5</td>
<td>-1.2699</td>
<td>0.1660</td>
<td>14.0529</td>
<td>14.0529</td>
</tr>
<tr>
<td>-4.4</td>
<td>-1.2137</td>
<td>0.1521</td>
<td>13.2951</td>
<td>13.2951</td>
</tr>
<tr>
<td>-4.3</td>
<td>-1.1574</td>
<td>0.1382</td>
<td>12.5372</td>
<td>12.5372</td>
</tr>
<tr>
<td>-4.2</td>
<td>-1.1009</td>
<td>0.1243</td>
<td>11.7793</td>
<td>11.7793</td>
</tr>
<tr>
<td>-4.1</td>
<td>-1.0442</td>
<td>0.1104</td>
<td>11.0214</td>
<td>11.0214</td>
</tr>
<tr>
<td>-4.0</td>
<td>-0.9874</td>
<td>0.0965</td>
<td>10.2632</td>
<td>10.2632</td>
</tr>
<tr>
<td>-3.9</td>
<td>-0.9303</td>
<td>0.0826</td>
<td>9.5050</td>
<td>9.5050</td>
</tr>
<tr>
<td>-3.8</td>
<td>-0.8730</td>
<td>0.0687</td>
<td>8.7469</td>
<td>8.7469</td>
</tr>
<tr>
<td>-3.7</td>
<td>-0.8155</td>
<td>0.0548</td>
<td>7.9887</td>
<td>7.9887</td>
</tr>
<tr>
<td>-3.6</td>
<td>-0.7577</td>
<td>0.0409</td>
<td>7.2307</td>
<td>7.2307</td>
</tr>
<tr>
<td>-3.5</td>
<td>-0.6997</td>
<td>0.0270</td>
<td>6.4727</td>
<td>6.4727</td>
</tr>
<tr>
<td>-3.4</td>
<td>-0.6415</td>
<td>0.0131</td>
<td>5.7146</td>
<td>5.7146</td>
</tr>
<tr>
<td>-3.3</td>
<td>-0.5830</td>
<td>0.0026</td>
<td>4.9566</td>
<td>4.9566</td>
</tr>
<tr>
<td>-3.2</td>
<td>-0.5243</td>
<td>0.0013</td>
<td>4.1987</td>
<td>4.1987</td>
</tr>
<tr>
<td>-3.1</td>
<td>-0.4652</td>
<td>0.0001</td>
<td>3.4406</td>
<td>3.4406</td>
</tr>
<tr>
<td>-3.0</td>
<td>-0.4059</td>
<td>0.0001</td>
<td>2.6826</td>
<td>2.6826</td>
</tr>
<tr>
<td>-2.9</td>
<td>-0.3466</td>
<td>0.0001</td>
<td>1.9245</td>
<td>1.9245</td>
</tr>
<tr>
<td>-2.8</td>
<td>-0.2863</td>
<td>0.0001</td>
<td>1.1664</td>
<td>1.1664</td>
</tr>
<tr>
<td>-2.7</td>
<td>-0.2259</td>
<td>0.0001</td>
<td>0.4082</td>
<td>0.4082</td>
</tr>
<tr>
<td>-2.6</td>
<td>-0.1653</td>
<td>0.0001</td>
<td>0.1501</td>
<td>0.1501</td>
</tr>
<tr>
<td>-2.5</td>
<td>-0.1042</td>
<td>0.0001</td>
<td>0.0920</td>
<td>0.0920</td>
</tr>
<tr>
<td>-2.4</td>
<td>-0.0421</td>
<td>0.0001</td>
<td>0.0339</td>
<td>0.0339</td>
</tr>
<tr>
<td>-2.3</td>
<td>-0.0813</td>
<td>0.0001</td>
<td>0.0153</td>
<td>0.0153</td>
</tr>
<tr>
<td>-2.2</td>
<td>-0.1200</td>
<td>0.0001</td>
<td>0.0067</td>
<td>0.0067</td>
</tr>
<tr>
<td>-2.1</td>
<td>-0.1587</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-2.0</td>
<td>-0.2074</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.9</td>
<td>-0.2561</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.8</td>
<td>-0.3048</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.7</td>
<td>-0.3535</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.6</td>
<td>-0.4022</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.5</td>
<td>-0.4509</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.4</td>
<td>-0.5095</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.3</td>
<td>-0.5681</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.2</td>
<td>-0.6267</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.1</td>
<td>-0.6853</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-1.0</td>
<td>-0.7439</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.9</td>
<td>-0.8025</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.8</td>
<td>-0.8611</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.7</td>
<td>-0.9197</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.6</td>
<td>-0.9783</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.5</td>
<td>-1.0369</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.4</td>
<td>-1.1021</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.3</td>
<td>-1.1673</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.2</td>
<td>-1.2325</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.1</td>
<td>-1.2977</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.0</td>
<td>-1.3629</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
APPENDIX B. THE RESULTS OF APPLYING THE DIFFERENT METHODS OF ESTIMATION TO THE 65 ARTIFICIAL SAMPLES

### Table B1. Estimates of $\mu$ for $\Lambda(\mu, \sigma^2)$

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\sigma$</th>
<th>Maximum likelihood</th>
<th>Moments</th>
<th>Quantiles 27% or 73%</th>
<th>Graphical (i)</th>
<th>Graphical (ii)</th>
<th>Graphical (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.2</td>
<td>0.0423</td>
<td>0.0417</td>
<td>-0.00561</td>
<td>-0.00547</td>
<td>-0.00547</td>
<td>-0.00396</td>
</tr>
<tr>
<td>32</td>
<td>0.2</td>
<td>0.0387</td>
<td>0.0323</td>
<td>0.00660</td>
<td>0.00431</td>
<td>0.00451</td>
<td>0.00512</td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>0.0353</td>
<td>0.0336</td>
<td>-0.00345</td>
<td>-0.00016</td>
<td>-0.00173</td>
<td>-0.01087</td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>0.0023</td>
<td>0.0054</td>
<td>-0.00105</td>
<td>0.00127</td>
<td>0.00288</td>
<td>0.00168</td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>0.0063</td>
<td>0.0732</td>
<td>0.00580</td>
<td>0.00770</td>
<td>0.00755</td>
<td>0.00683</td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>0.0134</td>
<td>0.0140</td>
<td>0.00140</td>
<td>0.00076</td>
<td>0.00371</td>
<td>0.00101</td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.1513</td>
<td>0.157</td>
<td>0.1425</td>
<td>0.1422</td>
<td>0.1448</td>
<td>0.1653</td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.0223</td>
<td>0.0232</td>
<td>0.00560</td>
<td>0.00379</td>
<td>0.00123</td>
<td>0.0128</td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>0.0176</td>
<td>0.0134</td>
<td>0.00330</td>
<td>-0.00028</td>
<td>0.00151</td>
<td>0.00306</td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>0.1958</td>
<td>0.205</td>
<td>0.01560</td>
<td>0.01937</td>
<td>0.02754</td>
<td>0.02617</td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>0.2514</td>
<td>0.2682</td>
<td>0.0242</td>
<td>0.0214</td>
<td>0.0312</td>
<td>0.03105</td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>0.00077</td>
<td>0.00280</td>
<td>0.0122</td>
<td>0.0177</td>
<td>0.0226</td>
<td>0.0161</td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>0.1930</td>
<td>0.1066</td>
<td>0.0139</td>
<td>0.0236</td>
<td>0.01522</td>
<td>0.01508</td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>0.0865</td>
<td>0.0223</td>
<td>0.0126</td>
<td>0.02422</td>
<td>0.01856</td>
<td>0.01266</td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>0.1650</td>
<td>0.2660</td>
<td>0.0228</td>
<td>0.02442</td>
<td>0.02707</td>
<td>0.00529</td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>0.0266</td>
<td>0.0207</td>
<td>0.0270</td>
<td>0.03974</td>
<td>0.03272</td>
<td>0.03444</td>
</tr>
<tr>
<td>32</td>
<td>1.0</td>
<td>0.2713</td>
<td>0.3097</td>
<td>0.2700</td>
<td>0.3258</td>
<td>0.3616</td>
<td>0.3016</td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>0.0277</td>
<td>0.0279</td>
<td>0.0300</td>
<td>0.0361</td>
<td>0.0186</td>
<td>0.02115</td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>0.0023</td>
<td>0.0027</td>
<td>0.01190</td>
<td>0.00444</td>
<td>0.00028</td>
<td>0.00006</td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>0.0923</td>
<td>0.0039</td>
<td>-0.0896</td>
<td>-0.0572</td>
<td>-0.0729</td>
<td>-0.0681</td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>0.0105</td>
<td>0.0106</td>
<td>-0.0223</td>
<td>-0.02235</td>
<td>-0.00588</td>
<td>-0.00116</td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>0.0595</td>
<td>0.052</td>
<td>-0.0060</td>
<td>-0.00581</td>
<td>-0.01055</td>
<td>-0.00943</td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>0.00111</td>
<td>0.0015</td>
<td>-0.0010</td>
<td>0.00089</td>
<td>0.00068</td>
<td>0.000145</td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>0.00831</td>
<td>0.0096</td>
<td>0.0025</td>
<td>0.00173</td>
<td>0.00260</td>
<td>0.00173</td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>0.00831</td>
<td>0.0096</td>
<td>0.0025</td>
<td>0.00173</td>
<td>0.00260</td>
<td>0.00173</td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.0088</td>
<td>0.0168</td>
<td>0.0021</td>
<td>0.00161</td>
<td>0.00327</td>
<td>0.00050</td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.00268</td>
<td>0.0059</td>
<td>0.0210</td>
<td>0.0161</td>
<td>0.02274</td>
<td>0.00498</td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>0.1925</td>
<td>0.1871</td>
<td>-0.1390</td>
<td>-0.1621</td>
<td>-0.1859</td>
<td>-0.1621</td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>0.00220</td>
<td>0.0034</td>
<td>0.0013</td>
<td>0.00089</td>
<td>0.003276</td>
<td>0.00008</td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>0.00356</td>
<td>0.00558</td>
<td>0.0924</td>
<td>0.00294</td>
<td>0.1331</td>
<td>0.1049</td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>0.00311</td>
<td>0.0037</td>
<td>0.0820</td>
<td>0.00293</td>
<td>0.0915</td>
<td>0.0553</td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>0.01313</td>
<td>0.1908</td>
<td>0.0945</td>
<td>0.1177</td>
<td>0.1597</td>
<td>0.2069</td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>0.00210</td>
<td>0.0732</td>
<td>-0.0225</td>
<td>-0.0221</td>
<td>-0.0101</td>
<td>-0.1297</td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>0.1583</td>
<td>0.1803</td>
<td>0.1300</td>
<td>0.1723</td>
<td>0.1723</td>
<td>0.1151</td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>0.00347</td>
<td>0.0018</td>
<td>0.0020</td>
<td>0.00181</td>
<td>0.00008</td>
<td>0.00172</td>
</tr>
</tbody>
</table>
### APPENDIX B

#### TABLE B1 (continued)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\sigma$</th>
<th>Maximum likelihood</th>
<th>Moments</th>
<th>Quantiles 27%</th>
<th>73%</th>
<th>Graphical (i)</th>
<th>Graphical (ii)</th>
<th>Graphical (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.0010</td>
<td>0.0006</td>
<td>-0.00100</td>
<td>-0.00020</td>
<td>-0.00059</td>
<td>-0.00059</td>
<td>0.00040</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.0024</td>
<td>0.0022</td>
<td>-0.00210</td>
<td>-0.00243</td>
<td>-0.00276</td>
<td>-0.00276</td>
<td>0.00342</td>
</tr>
<tr>
<td>128</td>
<td>0.3</td>
<td>0.0018</td>
<td>0.0010</td>
<td>-0.00130</td>
<td>-0.00122</td>
<td>-0.00122</td>
<td>-0.00122</td>
<td>0.00000</td>
</tr>
<tr>
<td>128</td>
<td>0.3</td>
<td>0.0057</td>
<td>0.0054</td>
<td>0.00575</td>
<td>0.00524</td>
<td>0.00545</td>
<td>0.00545</td>
<td>0.00377</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>0.0049</td>
<td>0.0049</td>
<td>-0.00520</td>
<td>-0.00610</td>
<td>-0.00270</td>
<td>-0.00270</td>
<td>-0.00210</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>0.0091</td>
<td>0.0075</td>
<td>0.00820</td>
<td>0.003770</td>
<td>0.00159</td>
<td>0.00159</td>
<td>0.00325</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>0.0043</td>
<td>0.0055</td>
<td>0.00200</td>
<td>0.00482</td>
<td>0.00270</td>
<td>0.00270</td>
<td>0.00200</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.0006</td>
<td>0.00134</td>
<td>0.00075</td>
<td>0.00222</td>
<td>-0.0164</td>
<td>-0.0164</td>
<td>-0.02265</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.0061</td>
<td>0.0090</td>
<td>-0.00690</td>
<td>-0.00222</td>
<td>0.00020</td>
<td>0.00020</td>
<td>0.00257</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.0037</td>
<td>0.0093</td>
<td>-0.00660</td>
<td>-0.00488</td>
<td>0.00459</td>
<td>0.00459</td>
<td>0.01068</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>0.0053</td>
<td>0.0127</td>
<td>0.00770</td>
<td>0.00969</td>
<td>0.00953</td>
<td>0.00953</td>
<td>0.02313</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>0.0079</td>
<td>0.0065</td>
<td>-0.00690</td>
<td>0.00661</td>
<td>-0.00931</td>
<td>-0.00931</td>
<td>-0.10049</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>0.0076</td>
<td>0.0086</td>
<td>0.00480</td>
<td>0.00497</td>
<td>0.00387</td>
<td>0.00387</td>
<td>0.1102</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>0.0038</td>
<td>0.0067</td>
<td>0.00080</td>
<td>0.00589</td>
<td>0.00493</td>
<td>0.00493</td>
<td>0.0545</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>0.0125</td>
<td>0.0121</td>
<td>0.00900</td>
<td>0.01339</td>
<td>0.1194</td>
<td>0.1194</td>
<td>0.2319</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>0.0053</td>
<td>0.0085</td>
<td>0.00315</td>
<td>0.00812</td>
<td>0.0076</td>
<td>0.0076</td>
<td>0.1200</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>0.0010</td>
<td>0.0079</td>
<td>-0.00500</td>
<td>0.00468</td>
<td>-0.1278</td>
<td>-0.1278</td>
<td>-0.1604</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>0.0010</td>
<td>0.0082</td>
<td>0.2530</td>
<td>0.1167</td>
<td>0.1167</td>
<td>0.1167</td>
<td>0.6662</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>0.0026</td>
<td>0.0015</td>
<td>0.00340</td>
<td>0.00461</td>
<td>0.00292</td>
<td>0.00292</td>
<td>0.0227</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>0.0020</td>
<td>0.0020</td>
<td>-0.00220</td>
<td>-0.00284</td>
<td>-0.00315</td>
<td>-0.00315</td>
<td>-0.0297</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>0.0090</td>
<td>0.0046</td>
<td>-0.00459</td>
<td>-0.00574</td>
<td>-0.00190</td>
<td>-0.00190</td>
<td>-0.0747</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>0.0032</td>
<td>0.0016</td>
<td>-0.00290</td>
<td>-0.00020</td>
<td>-0.00165</td>
<td>-0.00165</td>
<td>-0.1456</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>0.0042</td>
<td>0.00437</td>
<td>0.00930</td>
<td>0.00965</td>
<td>0.00662</td>
<td>0.00662</td>
<td>0.0408</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>0.0011</td>
<td>0.00372</td>
<td>0.00080</td>
<td>0.00666</td>
<td>0.1423</td>
<td>0.1423</td>
<td>0.6140</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>0.0026</td>
<td>0.0066</td>
<td>-0.00250</td>
<td>0.00680</td>
<td>-0.0239</td>
<td>-0.0239</td>
<td>-0.0257</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>0.0053</td>
<td>0.1136</td>
<td>0.00900</td>
<td>0.2280</td>
<td>0.0402</td>
<td>0.0402</td>
<td>0.3362</td>
</tr>
<tr>
<td>512</td>
<td>0.5</td>
<td>0.0031</td>
<td>0.0371</td>
<td>0.03000</td>
<td>0.0431</td>
<td>0.0361</td>
<td>0.0361</td>
<td>0.0602</td>
</tr>
<tr>
<td>512</td>
<td>0.5</td>
<td>0.0050</td>
<td>0.0310</td>
<td>0.0420</td>
<td>0.0458</td>
<td>0.0408</td>
<td>0.0408</td>
<td>0.0408</td>
</tr>
<tr>
<td>512</td>
<td>0.9</td>
<td>0.0080</td>
<td>0.1156</td>
<td>0.0630</td>
<td>0.1077</td>
<td>0.1190</td>
<td>0.1190</td>
<td>0.0953</td>
</tr>
</tbody>
</table>
### Table B2. Estimates $s^2$ of $\sigma^2$ for $\Lambda(\mu, \sigma^2)$

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\sigma$</th>
<th>$s^2$</th>
<th>Maximum likelihood</th>
<th>Moments</th>
<th>Quantiles 7% 93%</th>
<th>Graphical (i)</th>
<th>Graphical (ii)</th>
<th>Graphical (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.2</td>
<td>0.0236</td>
<td>0.0239</td>
<td>0.0318</td>
<td>0.0324</td>
<td>0.0398</td>
<td>0.0455</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.2</td>
<td>0.0240</td>
<td>0.0282</td>
<td>0.0509</td>
<td>0.0609</td>
<td>0.0403</td>
<td>0.0629</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>0.0212</td>
<td>0.1380</td>
<td>0.1125</td>
<td>0.1399</td>
<td>0.0911</td>
<td>0.2024</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>0.0295</td>
<td>0.0822</td>
<td>0.1019</td>
<td>0.1121</td>
<td>0.1047</td>
<td>0.1087</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>0.1192</td>
<td>0.1652</td>
<td>0.1329</td>
<td>0.1436</td>
<td>0.1305</td>
<td>0.1307</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>0.1356</td>
<td>0.1904</td>
<td>0.1555</td>
<td>0.1634</td>
<td>0.1356</td>
<td>0.1423</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.2183</td>
<td>0.2191</td>
<td>0.1639</td>
<td>0.2373</td>
<td>0.1868</td>
<td>0.3341</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.2512</td>
<td>0.2303</td>
<td>0.2565</td>
<td>0.2829</td>
<td>0.2791</td>
<td>0.2748</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>0.3266</td>
<td>0.3098</td>
<td>0.2751</td>
<td>0.3557</td>
<td>0.3148</td>
<td>0.3006</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>0.4390</td>
<td>0.4285</td>
<td>0.3865</td>
<td>0.4448</td>
<td>0.3352</td>
<td>0.3798</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>0.5064</td>
<td>0.2653</td>
<td>0.3831</td>
<td>0.3331</td>
<td>0.3127</td>
<td>0.2950</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>0.5681</td>
<td>0.5287</td>
<td>0.5669</td>
<td>0.4735</td>
<td>0.2533</td>
<td>0.2597</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>0.6517</td>
<td>0.4823</td>
<td>0.6176</td>
<td>0.6239</td>
<td>0.6168</td>
<td>0.7152</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>0.6576</td>
<td>0.4572</td>
<td>0.5472</td>
<td>0.5497</td>
<td>0.5472</td>
<td>0.6314</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>0.7860</td>
<td>0.7507</td>
<td>0.5957</td>
<td>0.6318</td>
<td>0.5273</td>
<td>0.5545</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>0.8578</td>
<td>0.7811</td>
<td>0.6828</td>
<td>0.7818</td>
<td>0.7708</td>
<td>1.1681</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.0</td>
<td>0.7236</td>
<td>0.4730</td>
<td>0.6066</td>
<td>0.5336</td>
<td>0.4506</td>
<td>0.5371</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.0</td>
<td>0.6038</td>
<td>0.4908</td>
<td>0.5399</td>
<td>0.5085</td>
<td>0.4156</td>
<td>0.4032</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>0.0291</td>
<td>0.0383</td>
<td>0.0350</td>
<td>0.0332</td>
<td>0.0435</td>
<td>0.0446</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>0.0421</td>
<td>0.0400</td>
<td>0.0373</td>
<td>0.0311</td>
<td>0.0800</td>
<td>0.0406</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>0.0879</td>
<td>0.0880</td>
<td>0.0865</td>
<td>0.0816</td>
<td>0.1089</td>
<td>0.0947</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>0.1606</td>
<td>0.1523</td>
<td>0.1651</td>
<td>0.1649</td>
<td>0.1814</td>
<td>0.0734</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>0.1456</td>
<td>0.1491</td>
<td>0.1564</td>
<td>0.1333</td>
<td>0.1416</td>
<td>0.1829</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>0.1152</td>
<td>0.1164</td>
<td>0.1379</td>
<td>0.1174</td>
<td>0.1043</td>
<td>0.1257</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>0.1726</td>
<td>0.1570</td>
<td>0.2497</td>
<td>0.1510</td>
<td>0.1459</td>
<td>0.1861</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>0.1879</td>
<td>0.1613</td>
<td>0.2092</td>
<td>0.1924</td>
<td>0.2303</td>
<td>0.2066</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.3103</td>
<td>0.3007</td>
<td>0.2772</td>
<td>0.3781</td>
<td>0.3709</td>
<td>0.3478</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.4542</td>
<td>0.3351</td>
<td>0.4665</td>
<td>0.5468</td>
<td>0.4778</td>
<td>0.5052</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.5010</td>
<td>0.5231</td>
<td>0.4688</td>
<td>0.6334</td>
<td>0.5622</td>
<td>0.6542</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>0.4782</td>
<td>0.4784</td>
<td>0.4982</td>
<td>0.4571</td>
<td>0.5604</td>
<td>0.5814</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>0.7197</td>
<td>0.8757</td>
<td>0.8000</td>
<td>0.7965</td>
<td>0.5431</td>
<td>0.8189</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>0.7979</td>
<td>0.8044</td>
<td>0.6834</td>
<td>0.9177</td>
<td>0.8590</td>
<td>0.9900</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>0.6132</td>
<td>0.6459</td>
<td>0.8091</td>
<td>0.8299</td>
<td>0.6016</td>
<td>0.9116</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>0.7200</td>
<td>0.6164</td>
<td>0.7194</td>
<td>0.6098</td>
<td>0.8345</td>
<td>0.6978</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>0.9305</td>
<td>0.7713</td>
<td>0.9720</td>
<td>0.8747</td>
<td>0.7554</td>
<td>0.8998</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>0.8724</td>
<td>0.8660</td>
<td>1.0339</td>
<td>0.8511</td>
<td>0.8598</td>
<td>0.9104</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>$\sigma$</td>
<td>$\sigma^2$</td>
<td>Maximum likelihood</td>
<td>Moments</td>
<td>Quantiles 7%</td>
<td>93%</td>
<td>Graphical (i)</td>
<td>Graphical (ii)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>---------</td>
<td>-------------------</td>
<td>---------</td>
<td>-------------</td>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.04</td>
<td>0.0491</td>
<td>0.0414</td>
<td>0.0410</td>
<td></td>
<td>0.0422</td>
<td>0.0418</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.04</td>
<td>0.0373</td>
<td>0.0394</td>
<td>0.0416</td>
<td></td>
<td>0.0363</td>
<td>0.0403</td>
</tr>
<tr>
<td>128</td>
<td>0.3</td>
<td>0.09</td>
<td>0.0735</td>
<td>0.0738</td>
<td>0.0692</td>
<td></td>
<td>0.1020</td>
<td>0.0840</td>
</tr>
<tr>
<td>128</td>
<td>0.3</td>
<td>0.09</td>
<td>0.0771</td>
<td>0.0611</td>
<td>0.0930</td>
<td></td>
<td>0.0785</td>
<td>0.0766</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>0.16</td>
<td>0.1646</td>
<td>0.1829</td>
<td>0.1559</td>
<td></td>
<td>0.1951</td>
<td>0.1830</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>0.25</td>
<td>0.3170</td>
<td>0.3214</td>
<td>0.3012</td>
<td></td>
<td>0.3470</td>
<td>0.3629</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>0.25</td>
<td>0.2761</td>
<td>0.3060</td>
<td>0.2884</td>
<td></td>
<td>0.2692</td>
<td>0.2550</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.36</td>
<td>0.2611</td>
<td>0.2677</td>
<td>0.3102</td>
<td></td>
<td>0.2774</td>
<td>0.2632</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.36</td>
<td>0.3172</td>
<td>0.3259</td>
<td>0.2968</td>
<td></td>
<td>0.3219</td>
<td>0.3293</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>0.49</td>
<td>0.5271</td>
<td>0.4427</td>
<td>0.5101</td>
<td></td>
<td>0.5174</td>
<td>0.5537</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>0.49</td>
<td>0.5197</td>
<td>0.4553</td>
<td>0.4723</td>
<td></td>
<td>0.4834</td>
<td>0.4937</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>0.64</td>
<td>0.5787</td>
<td>0.4991</td>
<td>0.6307</td>
<td></td>
<td>0.5951</td>
<td>0.6212</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>0.64</td>
<td>0.5778</td>
<td>0.5775</td>
<td>0.6136</td>
<td></td>
<td>0.6237</td>
<td>0.6438</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>0.81</td>
<td>0.7474</td>
<td>0.5795</td>
<td>0.7099</td>
<td></td>
<td>0.7699</td>
<td>0.7761</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>0.81</td>
<td>0.8540</td>
<td>0.5271</td>
<td>0.6174</td>
<td></td>
<td>0.6601</td>
<td>0.6407</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>1.00</td>
<td>0.8789</td>
<td>0.7245</td>
<td>0.6853</td>
<td></td>
<td>0.7887</td>
<td>1.0150</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>1.00</td>
<td>0.9008</td>
<td>0.9729</td>
<td>0.9069</td>
<td></td>
<td>0.9369</td>
<td>0.8917</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>0.16</td>
<td>0.1547</td>
<td>0.1798</td>
<td>0.1609</td>
<td></td>
<td>0.1667</td>
<td>0.1533</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>0.16</td>
<td>0.1609</td>
<td>0.1612</td>
<td>0.1534</td>
<td></td>
<td>0.1590</td>
<td>0.1655</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>0.36</td>
<td>0.3543</td>
<td>0.5077</td>
<td>0.3263</td>
<td></td>
<td>0.3323</td>
<td>0.3502</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>0.36</td>
<td>0.3333</td>
<td>0.2995</td>
<td>0.3553</td>
<td></td>
<td>0.4153</td>
<td>0.3860</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>0.64</td>
<td>0.5105</td>
<td>0.5290</td>
<td>0.5842</td>
<td></td>
<td>0.5418</td>
<td>0.5420</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>0.64</td>
<td>0.7184</td>
<td>0.8177</td>
<td>0.7014</td>
<td></td>
<td>0.8754</td>
<td>0.5245</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>1.00</td>
<td>0.8672</td>
<td>0.6683</td>
<td>0.9521</td>
<td></td>
<td>0.9404</td>
<td>0.9507</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>1.00</td>
<td>0.9727</td>
<td>0.7441</td>
<td>0.9664</td>
<td></td>
<td>0.9761</td>
<td>0.9761</td>
</tr>
<tr>
<td>512</td>
<td>0.5</td>
<td>0.25</td>
<td>0.2938</td>
<td>0.2157</td>
<td>0.2331</td>
<td></td>
<td>0.2290</td>
<td>0.1893</td>
</tr>
<tr>
<td>512</td>
<td>0.7</td>
<td>0.41</td>
<td>0.4894</td>
<td>0.4739</td>
<td>0.4324</td>
<td></td>
<td>0.4284</td>
<td>0.4695</td>
</tr>
<tr>
<td>512</td>
<td>0.9</td>
<td>0.61</td>
<td>0.8755</td>
<td>0.7816</td>
<td>0.9226</td>
<td></td>
<td>0.9474</td>
<td>0.7320</td>
</tr>
<tr>
<td>Sample size</td>
<td>$\sigma$</td>
<td>$\alpha$</td>
<td>Finney Moments</td>
<td>Quantiles</td>
<td>Graphical (i)</td>
<td>Graphical (ii)</td>
<td>Graphical (iii)</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>----------------</td>
<td>-----------</td>
<td>---------------</td>
<td>---------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.2</td>
<td>1.0020</td>
<td>0.9708</td>
<td>0.9707</td>
<td>0.9617</td>
<td>0.9665</td>
<td>0.9658</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.2</td>
<td>1.0020</td>
<td>1.0764</td>
<td>1.0740</td>
<td>1.0892</td>
<td>1.0790</td>
<td>1.0782</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>1.0460</td>
<td>1.0210</td>
<td>1.0257</td>
<td>1.0220</td>
<td>1.0561</td>
<td>1.0287</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>1.0460</td>
<td>1.0486</td>
<td>1.0476</td>
<td>1.0413</td>
<td>1.0712</td>
<td>1.0344</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>1.0834</td>
<td>1.1363</td>
<td>1.1341</td>
<td>1.1339</td>
<td>1.1804</td>
<td>1.1511</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>1.0834</td>
<td>1.0558</td>
<td>1.0578</td>
<td>1.0658</td>
<td>1.1049</td>
<td>1.1051</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>1.1333</td>
<td>1.2978</td>
<td>1.2967</td>
<td>1.2516</td>
<td>1.2900</td>
<td>1.2946</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>1.1333</td>
<td>1.2045</td>
<td>1.2016</td>
<td>1.2132</td>
<td>1.2471</td>
<td>1.2507</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.2129</td>
<td>1.2085</td>
<td>1.1800</td>
<td>1.1913</td>
<td>1.1883</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.5200</td>
<td>1.5261</td>
<td>1.4287</td>
<td>1.4367</td>
<td>1.5837</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>1.2718</td>
<td>1.5427</td>
<td>1.5417</td>
<td>1.5132</td>
<td>1.5265</td>
<td>1.5837</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>1.2718</td>
<td>1.2800</td>
<td>1.2877</td>
<td>1.4505</td>
<td>1.3749</td>
<td>1.3404</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>1.3778</td>
<td>1.1384</td>
<td>1.1184</td>
<td>1.1930</td>
<td>1.1391</td>
<td>1.2097</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>1.3778</td>
<td>1.1684</td>
<td>1.1976</td>
<td>1.1630</td>
<td>1.2446</td>
<td>1.2180</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>1.4995</td>
<td>1.9338</td>
<td>2.0112</td>
<td>1.8755</td>
<td>1.9734</td>
<td>1.8021</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>1.4995</td>
<td>1.9724</td>
<td>1.9327</td>
<td>1.0714</td>
<td>1.1825</td>
<td>1.1126</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.00</td>
<td>1.6487</td>
<td>1.9338</td>
<td>1.8550</td>
<td>1.6658</td>
<td>1.9479</td>
<td>1.9260</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.00</td>
<td>1.6487</td>
<td>1.7961</td>
<td>1.7421</td>
<td>1.7995</td>
<td>1.7811</td>
<td>1.7574</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>1.0020</td>
<td>1.0384</td>
<td>1.0482</td>
<td>1.0592</td>
<td>1.0536</td>
<td>1.0412</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>1.0020</td>
<td>1.0231</td>
<td>1.0299</td>
<td>1.0047</td>
<td>1.0925</td>
<td>1.0979</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>1.0460</td>
<td>0.9993</td>
<td>0.9937</td>
<td>0.9148</td>
<td>0.9907</td>
<td>0.9808</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>1.0460</td>
<td>1.0360</td>
<td>1.0300</td>
<td>1.0403</td>
<td>1.0097</td>
<td>0.9527</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>1.0834</td>
<td>0.9969</td>
<td>0.9968</td>
<td>0.9972</td>
<td>0.9866</td>
<td>0.9779</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>1.0834</td>
<td>1.0476</td>
<td>1.0481</td>
<td>1.0544</td>
<td>1.0550</td>
<td>0.9618</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>1.1333</td>
<td>1.1048</td>
<td>1.1004</td>
<td>1.1538</td>
<td>1.0927</td>
<td>1.1065</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>1.1333</td>
<td>1.0724</td>
<td>1.0694</td>
<td>1.1643</td>
<td>1.0958</td>
<td>1.1281</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.1777</td>
<td>1.1820</td>
<td>1.1695</td>
<td>1.2349</td>
<td>1.2156</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.3093</td>
<td>1.2793</td>
<td>1.2865</td>
<td>1.2742</td>
<td>1.2666</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>1.2776</td>
<td>1.2352</td>
<td>1.2285</td>
<td>1.2474</td>
<td>1.2716</td>
<td>1.2358</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>1.2776</td>
<td>1.2776</td>
<td>1.2814</td>
<td>1.3817</td>
<td>1.3622</td>
<td>1.3889</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>1.3771</td>
<td>1.4231</td>
<td>1.4163</td>
<td>1.5261</td>
<td>1.4661</td>
<td>1.4924</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>1.3771</td>
<td>1.2998</td>
<td>1.2997</td>
<td>1.4017</td>
<td>1.3842</td>
<td>1.3806</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>1.4995</td>
<td>1.7081</td>
<td>1.6766</td>
<td>1.6472</td>
<td>1.6074</td>
<td>1.5800</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>1.4995</td>
<td>1.4688</td>
<td>1.4044</td>
<td>1.3068</td>
<td>1.4060</td>
<td>1.5026</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>1.6487</td>
<td>1.4855</td>
<td>1.4701</td>
<td>1.3815</td>
<td>1.6597</td>
<td>1.7332</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>1.6487</td>
<td>1.4855</td>
<td>1.4402</td>
<td>1.7272</td>
<td>1.6830</td>
<td>1.7044</td>
<td></td>
</tr>
</tbody>
</table>
## APPENDIX B

### TABLE B3 (continued)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>Moments</th>
<th>Quantiles</th>
<th>Graphical (i)</th>
<th>Graphical (ii)</th>
<th>Graphical (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.2</td>
<td>1.0202</td>
<td>1.0214</td>
<td>1.0215</td>
<td>1.0234</td>
<td>1.0260</td>
<td>1.0266</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>1.0202</td>
<td>1.0200</td>
<td>1.0186</td>
<td>1.0209</td>
<td>1.0220</td>
<td>1.0226</td>
</tr>
<tr>
<td>128</td>
<td>0.3</td>
<td>1.0460</td>
<td>1.0769</td>
<td>1.0713</td>
<td>1.0694</td>
<td>1.0511</td>
<td>1.0518</td>
</tr>
<tr>
<td>128</td>
<td>0.3</td>
<td>1.0460</td>
<td>1.0689</td>
<td>1.0676</td>
<td>1.0660</td>
<td>1.0589</td>
<td>1.0645</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>1.0833</td>
<td>1.1180</td>
<td>1.1204</td>
<td>1.1252</td>
<td>1.1120</td>
<td>1.1213</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>1.0833</td>
<td>1.1180</td>
<td>1.1204</td>
<td>1.1252</td>
<td>1.1120</td>
<td>1.1213</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>1.1331</td>
<td>1.2132</td>
<td>1.2190</td>
<td>1.2161</td>
<td>1.2121</td>
<td>1.2136</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>1.1331</td>
<td>1.1472</td>
<td>1.1589</td>
<td>1.1565</td>
<td>1.1517</td>
<td>1.1513</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.1990</td>
<td>1.1888</td>
<td>1.1843</td>
<td>1.1735</td>
<td>1.1809</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.2191</td>
<td>1.2242</td>
<td>1.2333</td>
<td>1.2344</td>
<td>1.2351</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>1.2776</td>
<td>1.4708</td>
<td>1.4717</td>
<td>1.4726</td>
<td>1.4700</td>
<td>1.4764</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>1.2776</td>
<td>1.6571</td>
<td>1.6759</td>
<td>1.6818</td>
<td>1.6805</td>
<td>1.6805</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>1.3771</td>
<td>1.4139</td>
<td>1.4024</td>
<td>1.4381</td>
<td>1.4324</td>
<td>1.4321</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>1.3771</td>
<td>1.4107</td>
<td>1.4254</td>
<td>1.4274</td>
<td>1.4290</td>
<td>1.4300</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>1.5993</td>
<td>1.8243</td>
<td>1.5970</td>
<td>1.7254</td>
<td>1.6634</td>
<td>1.6218</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>1.5993</td>
<td>1.5228</td>
<td>1.4111</td>
<td>1.5855</td>
<td>1.4777</td>
<td>1.4747</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>1.6487</td>
<td>1.5794</td>
<td>1.6689</td>
<td>1.7034</td>
<td>1.5576</td>
<td>1.4618</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>1.6487</td>
<td>1.8167</td>
<td>1.8453</td>
<td>2.0843</td>
<td>1.8113</td>
<td>1.7613</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>1.0833</td>
<td>1.1129</td>
<td>1.1157</td>
<td>1.1213</td>
<td>1.1382</td>
<td>1.1127</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>1.0833</td>
<td>1.0616</td>
<td>1.0620</td>
<td>1.0625</td>
<td>1.0526</td>
<td>1.0595</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.1830</td>
<td>1.2083</td>
<td>1.2154</td>
<td>1.2327</td>
<td>1.2160</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>1.1972</td>
<td>1.2371</td>
<td>1.2434</td>
<td>1.2393</td>
<td>1.2499</td>
<td>1.2273</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>1.3771</td>
<td>1.2397</td>
<td>1.2446</td>
<td>1.2517</td>
<td>1.2392</td>
<td>1.2235</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>1.3771</td>
<td>1.4153</td>
<td>1.4357</td>
<td>1.4315</td>
<td>1.4164</td>
<td>1.4066</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>1.6487</td>
<td>1.5309</td>
<td>1.5019</td>
<td>1.5700</td>
<td>1.6179</td>
<td>1.5550</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>1.6487</td>
<td>1.6345</td>
<td>1.6213</td>
<td>1.7009</td>
<td>1.7279</td>
<td>1.8708</td>
</tr>
<tr>
<td>512</td>
<td>0.5</td>
<td>1.1331</td>
<td>1.1597</td>
<td>1.1539</td>
<td>1.1578</td>
<td>1.1591</td>
<td>1.1674</td>
</tr>
<tr>
<td>512</td>
<td>0.7</td>
<td>1.2776</td>
<td>1.4054</td>
<td>1.4086</td>
<td>1.4211</td>
<td>1.3934</td>
<td>1.3867</td>
</tr>
<tr>
<td>512</td>
<td>0.9</td>
<td>1.4993</td>
<td>1.6747</td>
<td>1.6693</td>
<td>1.6508</td>
<td>1.5450</td>
<td>1.6406</td>
</tr>
</tbody>
</table>
### TABLE B4: ESTIMATES $b^2$ OF $\beta^2$ FOR $\Lambda(\mu, \sigma^2)$

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\sigma$</th>
<th>$\beta^2$</th>
<th>Finney</th>
<th>Moments</th>
<th>Quantiles</th>
<th>Graphical (i)</th>
<th>Graphical (ii)</th>
<th>Graphical (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.2</td>
<td>0.0425</td>
<td>0.0251</td>
<td>0.0235</td>
<td>0.0299</td>
<td>0.0356</td>
<td>0.0379</td>
<td>0.0393</td>
</tr>
<tr>
<td>32</td>
<td>0.2</td>
<td>0.0425</td>
<td>0.0086</td>
<td>0.0313</td>
<td>0.0619</td>
<td>0.0793</td>
<td>0.0878</td>
<td>0.0966</td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>0.1030</td>
<td>0.1261</td>
<td>0.1607</td>
<td>0.2444</td>
<td>0.1677</td>
<td>0.1809</td>
<td>0.2210</td>
</tr>
<tr>
<td>32</td>
<td>0.3</td>
<td>0.1030</td>
<td>0.1664</td>
<td>0.0970</td>
<td>0.1163</td>
<td>0.1361</td>
<td>0.1293</td>
<td>0.1303</td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>0.2036</td>
<td>0.1664</td>
<td>0.1473</td>
<td>0.1800</td>
<td>0.2079</td>
<td>0.2047</td>
<td>0.1881</td>
</tr>
<tr>
<td>32</td>
<td>0.4</td>
<td>0.2036</td>
<td>0.1643</td>
<td>0.1741</td>
<td>0.2131</td>
<td>0.1526</td>
<td>0.1530</td>
<td>0.1728</td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.3647</td>
<td>0.4131</td>
<td>0.4258</td>
<td>0.2726</td>
<td>0.4312</td>
<td>0.3449</td>
<td>0.6705</td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.3647</td>
<td>0.4164</td>
<td>0.3674</td>
<td>0.4304</td>
<td>0.4760</td>
<td>0.3609</td>
<td>0.4231</td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>0.6211</td>
<td>0.1959</td>
<td>0.5328</td>
<td>0.4454</td>
<td>0.6063</td>
<td>0.5924</td>
<td>0.5120</td>
</tr>
<tr>
<td>32</td>
<td>0.6</td>
<td>0.6211</td>
<td>0.2776</td>
<td>1.2661</td>
<td>0.9721</td>
<td>1.3400</td>
<td>0.9739</td>
<td>1.1599</td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>1.0321</td>
<td>1.1335</td>
<td>0.8238</td>
<td>1.0718</td>
<td>0.9994</td>
<td>0.9202</td>
<td>0.8729</td>
</tr>
<tr>
<td>32</td>
<td>0.7</td>
<td>1.0321</td>
<td>1.0190</td>
<td>0.6251</td>
<td>1.4638</td>
<td>1.2628</td>
<td>1.1637</td>
<td>0.5320</td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>1.7002</td>
<td>1.1851</td>
<td>0.8204</td>
<td>1.4319</td>
<td>1.4177</td>
<td>1.1804</td>
<td>1.5749</td>
</tr>
<tr>
<td>32</td>
<td>0.8</td>
<td>1.7002</td>
<td>0.8321</td>
<td>1.1250</td>
<td>1.1984</td>
<td>1.5197</td>
<td>0.9438</td>
<td>2.1983</td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>2.8052</td>
<td>6.0773</td>
<td>18.6230</td>
<td>4.5144</td>
<td>28.6642</td>
<td>3.5015</td>
<td>4.9882</td>
</tr>
<tr>
<td>32</td>
<td>0.9</td>
<td>2.8052</td>
<td>6.0660</td>
<td>1.5079</td>
<td>0.9555</td>
<td>1.7711</td>
<td>1.4433</td>
<td>2.0665</td>
</tr>
<tr>
<td>32</td>
<td>1.0</td>
<td>4.6708</td>
<td>3.7416</td>
<td>2.1430</td>
<td>2.8635</td>
<td>2.3814</td>
<td>2.0111</td>
<td>2.4231</td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>0.0425</td>
<td>0.0445</td>
<td>0.0436</td>
<td>0.0400</td>
<td>0.0273</td>
<td>0.0282</td>
<td>0.0693</td>
</tr>
<tr>
<td>64</td>
<td>0.2</td>
<td>0.0425</td>
<td>0.0466</td>
<td>0.0434</td>
<td>0.0489</td>
<td>0.0441</td>
<td>0.0297</td>
<td>0.0438</td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>0.1030</td>
<td>0.0917</td>
<td>0.0923</td>
<td>0.0962</td>
<td>0.0938</td>
<td>0.1037</td>
<td>0.0957</td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>0.1030</td>
<td>0.0658</td>
<td>0.0596</td>
<td>0.0613</td>
<td>0.0564</td>
<td>0.1086</td>
<td>0.0617</td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>0.2036</td>
<td>0.1547</td>
<td>0.1523</td>
<td>0.1800</td>
<td>0.1500</td>
<td>0.1909</td>
<td>0.1905</td>
</tr>
<tr>
<td>64</td>
<td>0.4</td>
<td>0.2036</td>
<td>0.1553</td>
<td>0.1376</td>
<td>0.1944</td>
<td>0.1379</td>
<td>0.1923</td>
<td>0.1476</td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>0.3647</td>
<td>0.2417</td>
<td>0.2091</td>
<td>0.3659</td>
<td>0.1962</td>
<td>0.1953</td>
<td>0.2192</td>
</tr>
<tr>
<td>64</td>
<td>0.5</td>
<td>0.3647</td>
<td>0.2388</td>
<td>0.2095</td>
<td>0.3154</td>
<td>0.2543</td>
<td>0.2306</td>
<td>0.2660</td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.6211</td>
<td>0.0960</td>
<td>0.2403</td>
<td>0.3535</td>
<td>0.6771</td>
<td>0.6784</td>
<td>0.9499</td>
</tr>
<tr>
<td>64</td>
<td>0.6</td>
<td>0.6211</td>
<td>0.0861</td>
<td>0.0638</td>
<td>0.9087</td>
<td>0.9956</td>
<td>0.8238</td>
<td>0.8933</td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>1.0321</td>
<td>1.6621</td>
<td>1.6040</td>
<td>2.0104</td>
<td>2.3739</td>
<td>1.9038</td>
<td>2.3964</td>
</tr>
<tr>
<td>64</td>
<td>0.7</td>
<td>1.0321</td>
<td>0.8974</td>
<td>0.9407</td>
<td>0.9538</td>
<td>0.1440</td>
<td>0.2783</td>
<td>0.9447</td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>1.7002</td>
<td>2.0624</td>
<td>1.6341</td>
<td>2.8816</td>
<td>2.5407</td>
<td>1.6425</td>
<td>2.9093</td>
</tr>
<tr>
<td>64</td>
<td>0.8</td>
<td>1.7002</td>
<td>2.7579</td>
<td>3.0209</td>
<td>3.0209</td>
<td>4.0132</td>
<td>3.0165</td>
<td>4.1813</td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>2.8052</td>
<td>3.4972</td>
<td>2.5736</td>
<td>3.8382</td>
<td>3.6703</td>
<td>2.9056</td>
<td>1.8692</td>
</tr>
<tr>
<td>64</td>
<td>0.9</td>
<td>2.8052</td>
<td>2.1884</td>
<td>1.8396</td>
<td>2.0909</td>
<td>3.8699</td>
<td>2.3832</td>
<td>2.6884</td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>4.6708</td>
<td>4.9360</td>
<td>3.6690</td>
<td>5.6351</td>
<td>4.7324</td>
<td>3.3904</td>
<td>4.5167</td>
</tr>
<tr>
<td>64</td>
<td>1.0</td>
<td>4.6708</td>
<td>2.9342</td>
<td>1.9944</td>
<td>5.5787</td>
<td>2.9192</td>
<td>3.1602</td>
<td>2.1554</td>
</tr>
</tbody>
</table>
### APPENDIX B

**TABLE B4 (continued)**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\sigma$</th>
<th>$\beta^1$</th>
<th>Finney Moments</th>
<th>Quantiles</th>
<th>Graphical (i)</th>
<th>Graphical (ii)</th>
<th>Graphical (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.0425</td>
<td>0.0430</td>
<td>0.0444</td>
<td>0.0427</td>
<td>0.0451</td>
<td>0.0440</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.0426</td>
<td>0.0432</td>
<td>0.0425</td>
<td>0.0425</td>
<td>0.0477</td>
<td>0.0405</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.0430</td>
<td>0.0480</td>
<td>0.0467</td>
<td>0.0487</td>
<td>0.0477</td>
<td>0.0477</td>
</tr>
<tr>
<td>128</td>
<td>0.2</td>
<td>0.0432</td>
<td>0.0436</td>
<td>0.0412</td>
<td>0.0453</td>
<td>0.0481</td>
<td>0.0675</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>0.0296</td>
<td>0.0213</td>
<td>0.0283</td>
<td>0.0116</td>
<td>0.0390</td>
<td>0.0239</td>
</tr>
<tr>
<td>128</td>
<td>0.4</td>
<td>0.0296</td>
<td>0.0224</td>
<td>0.0277</td>
<td>0.0249</td>
<td>0.0243</td>
<td>0.0249</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>0.0247</td>
<td>0.0568</td>
<td>0.0494</td>
<td>0.0469</td>
<td>0.0247</td>
<td>0.0247</td>
</tr>
<tr>
<td>128</td>
<td>0.5</td>
<td>0.0247</td>
<td>0.0482</td>
<td>0.0415</td>
<td>0.0434</td>
<td>0.0426</td>
<td>0.0590</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.0211</td>
<td>0.0412</td>
<td>0.0412</td>
<td>0.0430</td>
<td>0.0393</td>
<td>0.0393</td>
</tr>
<tr>
<td>128</td>
<td>0.6</td>
<td>0.0211</td>
<td>0.0335</td>
<td>0.0320</td>
<td>0.0394</td>
<td>0.0376</td>
<td>0.0376</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>1.0821</td>
<td>1.4039</td>
<td>1.4276</td>
<td>1.4375</td>
<td>1.4957</td>
<td>1.5531</td>
</tr>
<tr>
<td>128</td>
<td>0.7</td>
<td>1.0821</td>
<td>0.7020</td>
<td>0.8063</td>
<td>0.6831</td>
<td>0.8763</td>
<td>0.7972</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>1.7002</td>
<td>1.5467</td>
<td>1.2877</td>
<td>1.8178</td>
<td>1.7577</td>
<td>1.7319</td>
</tr>
<tr>
<td>128</td>
<td>0.8</td>
<td>1.7002</td>
<td>1.3366</td>
<td>1.6066</td>
<td>1.5926</td>
<td>1.8930</td>
<td>1.8688</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>2.8052</td>
<td>2.8725</td>
<td>1.9930</td>
<td>4.2454</td>
<td>2.8663</td>
<td>2.9216</td>
</tr>
<tr>
<td>128</td>
<td>0.9</td>
<td>2.8052</td>
<td>3.0502</td>
<td>2.4084</td>
<td>2.5366</td>
<td>2.3290</td>
<td>2.1293</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>4.8708</td>
<td>3.4171</td>
<td>2.7317</td>
<td>4.7169</td>
<td>2.9127</td>
<td>2.9323</td>
</tr>
<tr>
<td>128</td>
<td>1.0</td>
<td>4.8708</td>
<td>4.6840</td>
<td>5.6479</td>
<td>7.4529</td>
<td>4.9115</td>
<td>4.8775</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>0.2036</td>
<td>0.2075</td>
<td>0.2120</td>
<td>0.2350</td>
<td>0.2890</td>
<td>0.2831</td>
</tr>
<tr>
<td>256</td>
<td>0.4</td>
<td>0.2036</td>
<td>0.1970</td>
<td>0.1981</td>
<td>0.1969</td>
<td>0.1993</td>
<td>0.2172</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>0.2211</td>
<td>0.2840</td>
<td>0.6695</td>
<td>0.4887</td>
<td>1.0704</td>
<td>0.5731</td>
</tr>
<tr>
<td>256</td>
<td>0.6</td>
<td>0.2211</td>
<td>0.6435</td>
<td>0.7553</td>
<td>0.5971</td>
<td>0.7792</td>
<td>0.7012</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>1.7002</td>
<td>1.0016</td>
<td>1.0073</td>
<td>1.3350</td>
<td>1.9120</td>
<td>1.6770</td>
</tr>
<tr>
<td>256</td>
<td>0.8</td>
<td>1.7002</td>
<td>2.0818</td>
<td>2.7190</td>
<td>2.8831</td>
<td>2.8900</td>
<td>2.9288</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>4.7608</td>
<td>3.3640</td>
<td>2.2427</td>
<td>3.9218</td>
<td>1.9023</td>
<td>3.8383</td>
</tr>
<tr>
<td>256</td>
<td>1.0</td>
<td>4.7608</td>
<td>3.6768</td>
<td>2.9330</td>
<td>3.8780</td>
<td>2.4118</td>
<td>4.8000</td>
</tr>
<tr>
<td>512</td>
<td>0.5</td>
<td>0.3637</td>
<td>0.3586</td>
<td>0.3223</td>
<td>0.3519</td>
<td>0.3342</td>
<td>0.3555</td>
</tr>
<tr>
<td>512</td>
<td>0.7</td>
<td>1.0321</td>
<td>1.0796</td>
<td>1.0420</td>
<td>1.0556</td>
<td>0.1911</td>
<td>0.8816</td>
</tr>
<tr>
<td>512</td>
<td>0.9</td>
<td>2.8052</td>
<td>3.8799</td>
<td>3.7699</td>
<td>4.3256</td>
<td>1.4902</td>
<td>3.0180</td>
</tr>
<tr>
<td>512</td>
<td>1.0</td>
<td>4.7608</td>
<td>3.8799</td>
<td>3.7699</td>
<td>4.3256</td>
<td>1.4902</td>
<td>3.0180</td>
</tr>
</tbody>
</table>

---

*Note: The table continues with similar entries.*
### Table B5. Estimates \( \lambda, m \) and \( \sigma^2 \) of \( \tau, \mu \) and \( \sigma^2 \) by the Method of Quantiles for \( \Lambda(\tau, \mu, \sigma^2) \)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( \sigma )</th>
<th>( \lambda )</th>
<th>( m )</th>
<th>( \sigma^2 )</th>
<th>Sample size</th>
<th>( \sigma )</th>
<th>( \lambda )</th>
<th>( m )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.02</td>
<td>-0.0019</td>
<td>0.0671</td>
<td>0.0062</td>
<td>128</td>
<td>0.02</td>
<td>-0.0019</td>
<td>0.0671</td>
<td>0.0062</td>
</tr>
<tr>
<td>32</td>
<td>0.03</td>
<td>-0.0272</td>
<td>0.0804</td>
<td>0.0058</td>
<td>128</td>
<td>0.03</td>
<td>-0.0272</td>
<td>0.0804</td>
<td>0.0058</td>
</tr>
<tr>
<td>32</td>
<td>0.04</td>
<td>-0.0435</td>
<td>0.0946</td>
<td>0.0052</td>
<td>128</td>
<td>0.04</td>
<td>-0.0435</td>
<td>0.0946</td>
<td>0.0052</td>
</tr>
<tr>
<td>32</td>
<td>0.05</td>
<td>-0.1475</td>
<td>0.1061</td>
<td>0.0236</td>
<td>128</td>
<td>0.05</td>
<td>-0.1475</td>
<td>0.1061</td>
<td>0.0236</td>
</tr>
<tr>
<td>32</td>
<td>0.06</td>
<td>-0.0259</td>
<td>0.0777</td>
<td>0.0051</td>
<td>128</td>
<td>0.06</td>
<td>-0.0259</td>
<td>0.0777</td>
<td>0.0051</td>
</tr>
<tr>
<td>32</td>
<td>0.07</td>
<td>-0.0363</td>
<td>0.1213</td>
<td>0.0360</td>
<td>128</td>
<td>0.07</td>
<td>-0.0363</td>
<td>0.1213</td>
<td>0.0360</td>
</tr>
<tr>
<td>32</td>
<td>0.08</td>
<td>-0.1001</td>
<td>0.2617</td>
<td>0.0691</td>
<td>128</td>
<td>0.08</td>
<td>-0.1001</td>
<td>0.2617</td>
<td>0.0691</td>
</tr>
<tr>
<td>32</td>
<td>0.09</td>
<td>-0.1637</td>
<td>0.2766</td>
<td>0.0570</td>
<td>128</td>
<td>0.09</td>
<td>-0.1637</td>
<td>0.2766</td>
<td>0.0570</td>
</tr>
<tr>
<td>32</td>
<td>1.00</td>
<td>-0.2661</td>
<td>0.2469</td>
<td>0.0569</td>
<td>128</td>
<td>1.00</td>
<td>-0.2661</td>
<td>0.2469</td>
<td>0.0569</td>
</tr>
<tr>
<td>64</td>
<td>0.02</td>
<td>-0.0096</td>
<td>0.0476</td>
<td>0.0169</td>
<td>256</td>
<td>0.02</td>
<td>-0.0096</td>
<td>0.0476</td>
<td>0.0169</td>
</tr>
<tr>
<td>64</td>
<td>0.03</td>
<td>0.0407</td>
<td>0.0167</td>
<td>0.0104</td>
<td>256</td>
<td>0.03</td>
<td>0.0407</td>
<td>0.0167</td>
<td>0.0104</td>
</tr>
<tr>
<td>64</td>
<td>0.04</td>
<td>-0.0443</td>
<td>0.0392</td>
<td>0.0055</td>
<td>256</td>
<td>0.04</td>
<td>-0.0443</td>
<td>0.0392</td>
<td>0.0055</td>
</tr>
<tr>
<td>64</td>
<td>0.05</td>
<td>-0.0095</td>
<td>0.0850</td>
<td>0.0132</td>
<td>256</td>
<td>0.05</td>
<td>-0.0095</td>
<td>0.0850</td>
<td>0.0132</td>
</tr>
<tr>
<td>64</td>
<td>0.06</td>
<td>0.0090</td>
<td>0.0465</td>
<td>0.0125</td>
<td>256</td>
<td>0.06</td>
<td>0.0090</td>
<td>0.0465</td>
<td>0.0125</td>
</tr>
<tr>
<td>64</td>
<td>0.07</td>
<td>-0.1485</td>
<td>0.0273</td>
<td>0.0238</td>
<td>256</td>
<td>0.07</td>
<td>-0.1485</td>
<td>0.0273</td>
<td>0.0238</td>
</tr>
<tr>
<td>64</td>
<td>0.08</td>
<td>-0.0489</td>
<td>0.0762</td>
<td>0.0306</td>
<td>256</td>
<td>0.08</td>
<td>-0.0489</td>
<td>0.0762</td>
<td>0.0306</td>
</tr>
<tr>
<td>64</td>
<td>0.09</td>
<td>-0.2634</td>
<td>0.0502</td>
<td>0.0489</td>
<td>256</td>
<td>0.09</td>
<td>-0.2634</td>
<td>0.0502</td>
<td>0.0489</td>
</tr>
<tr>
<td>64</td>
<td>1.00</td>
<td>-0.4256</td>
<td>0.0307</td>
<td>0.1155</td>
<td>256</td>
<td>1.00</td>
<td>-0.4256</td>
<td>0.0307</td>
<td>0.1155</td>
</tr>
<tr>
<td>64</td>
<td>1.01</td>
<td>-0.0099</td>
<td>0.0403</td>
<td>0.0403</td>
<td>1029</td>
<td>1.01</td>
<td>-0.0099</td>
<td>0.0403</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.02</td>
<td>-0.0094</td>
<td>0.0502</td>
<td>0.0403</td>
<td>1029</td>
<td>1.02</td>
<td>-0.0094</td>
<td>0.0502</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.03</td>
<td>-0.0091</td>
<td>0.0615</td>
<td>0.0403</td>
<td>1029</td>
<td>1.03</td>
<td>-0.0091</td>
<td>0.0615</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.04</td>
<td>-0.0094</td>
<td>0.0728</td>
<td>0.0403</td>
<td>1029</td>
<td>1.04</td>
<td>-0.0094</td>
<td>0.0728</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.05</td>
<td>-0.0091</td>
<td>0.0841</td>
<td>0.0403</td>
<td>1029</td>
<td>1.05</td>
<td>-0.0091</td>
<td>0.0841</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.06</td>
<td>-0.0094</td>
<td>0.0954</td>
<td>0.0403</td>
<td>1029</td>
<td>1.06</td>
<td>-0.0094</td>
<td>0.0954</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.07</td>
<td>-0.0091</td>
<td>0.1067</td>
<td>0.0403</td>
<td>1029</td>
<td>1.07</td>
<td>-0.0091</td>
<td>0.1067</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.08</td>
<td>-0.0094</td>
<td>0.1180</td>
<td>0.0403</td>
<td>1029</td>
<td>1.08</td>
<td>-0.0094</td>
<td>0.1180</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.09</td>
<td>-0.0091</td>
<td>0.1293</td>
<td>0.0403</td>
<td>1029</td>
<td>1.09</td>
<td>-0.0091</td>
<td>0.1293</td>
<td>0.0403</td>
</tr>
<tr>
<td>64</td>
<td>1.10</td>
<td>-0.0094</td>
<td>0.1406</td>
<td>0.0403</td>
<td>1029</td>
<td>1.10</td>
<td>-0.0094</td>
<td>0.1406</td>
<td>0.0403</td>
</tr>
</tbody>
</table>
INDEX

Abundance of species, 103
Additivity of Engel curves, 131-4
Age at first marriage, 105
Aggregation, 122-3, 128-9, 138-9
Agricultural earnings, 2-6, 119
Analogue devices, 23-4, 28
Analysis of variance, 79, 82, 86
Anthropometry, 103
Arithmetic mean, 2, 85, 111, 113, 115
of $A(\mu, \sigma^2)$, 8, 44-50
of $\Lambda(\gamma, \mu, \sigma^2)$, 15
of sample, 38
of $\Delta(\delta, \mu, \sigma^2)$, 95-9
Arithmetic variance, 44-50, 85
of $A(\mu, \sigma^2)$, 8
of sample, 38
of $\Delta(\delta, \mu, \sigma^2)$, 95-9
Artificial samples, 5, 28-36, 37, 43, 61, 91-3, 141-3
Astronomy, 105
Automatic computing, 6, 24, 28, 29, 30, 66, 71, 74-5, 80-1, 82-4
Bartlett, M. S., 79, 86
Berkson, J., 72, 74
Biology, 102-3
Bliss, C. I., 4, 66, 106
Breakage, theory of, 5, 26, 100-1
Censorship, 4-6, 90-3, 103
defined, 88
Central limit theorems, 13-14, 22-4, 27-9, 66, 110
Champenowone, D. G., 108-9
Cheapness, parameter of, 131
Classification, theory of, 27, 106
Coefficient of mean difference, 12-13, 15, 113
Coefficient of variation, 8, 15, 142
Cohen, A. C., 4, 55-7, 66, 75, 80-1, 98-9
Comparisons of parameter estimates, 6, 85-6, 111
Concentration of incomes, 111
Confidence intervals, 90, 85
Consolidation of commodities and of groups of consumers, 129-30
Convergence, 66, 75-8, 80-1, 83, 145
Convolution property, 11
Counts, discrete, 4, 92-5, 103
Cramér, H., 22
Definition, of $A(\mu, \sigma^2)$, 7
of $\Lambda(\gamma, \mu, \sigma^2)$, 14
of $A(\delta, \gamma, \mu, \sigma^2)$, 17
of $A(\delta, \gamma, \mu, \sigma^2)$, 95
of truncated distribution, 97
of censored distribution, 88
Econometrics scale in consumption, 137
Econometrics, size distribution in, 101-2
EDSAC, 29-30, 71 n.
described, 141-2
Efficiency, 40-1, 43, 46-8, 98-9
Elasticity, income, 131
Elasticity, price, 135
Elasticity, substitution, 135, 139-40
Electronic computing, see Automatic computing
Engel, E., 124
Engel curves, defined, 124, 129-39
Estimators, criteria for, 37
Finney, D. J., 45-6, 49-54, 66-7, 75, 75 n., 77, 79 n., 81, 84
Fisher, R. A., 85, 86
Four-parameter $A$-distribution, 4, 17, 26, 65, 109
Gaddum, J. H., 4, 66
Galton, F., 2-3, 22, 31
Geary, R. C., 30, 34
Generation, Models of, 20, 108-11
Genesis of $A$-distributions, 20, 24, 96
Geometric mean, 2-3, 11, 14
Gibrat, R., 3, 101, 102, 104, 108
Gini, C., 13, 113
Gjeddebaek, N. F., 51
Graphical methods, 5, 37, 42-4, 48-50, 91, 97
Grouping of observations, 5, 51-4, 94-5
Grundy, P. M., 51, 94, 103
Haldane, J. B. S., 22 n., 103
Hatch, T., 5, 53, 100-1
Hazen, A., 31
Heterogeneity in income distribution, 117, 120
Heteroscedasticity, 82-3
Homogeneity, 72, 79
Homoscedasticity, 78-9, 82
Household composition, 137-8
Household size, 104, 123
Income distribution, 24-5, 88, 107-20, 123, 129, 139
Indivisible goods, 124-5
Industrial processes, 104-5
Inferior goods, 127-9
Information matrix, 69-71, 80-1, 83, 144-5
Iterative procedures, 6, 57, 70-2, 83-9, 90, 143-5
Kalecki, M., 24-5, 110
Kapteyn, J. C., 4, 59, 101, 109
Kemaley, W. F. F., 55, 59, 61-3, 103
Klein, L. R., 122-3
Kolmogoroff, A. N., 5, 26, 101
Krumbein, W. C., 5
Kurtosis, 8-9, 34-5
INDEX

Least sample value method, 55-7, 143
Life-cycle in income, 109-10
Logarithmic probability paper, 30-3, 42, 69
Logistic curve, 72, 74-5
Logits, 72
Lorenz, M. O., 111
measure, 50, 85; defined 112
diagrams, 111-15
McAlister, D., 2-3, 22
Macromodels, 121-3
Marginal propensity to spend, 127-8
Matrix inversion methods, 144-5
Maximum likelihood, 5, 35, 39, 43-6, 69, 55-7, 69, 79, 89-93, 97, 143
Median, 9, 15
Median effective income, defined, 131
Micromodels, 121-3
Mode, 9, 15
Moment distributions, 1, 12, 15, 100-1, 111
Moments, 8, 15, 38, 87, 95, 142
calculation by summation, 143
method of, in estimation, 5, 38-9, 43-4, 46, 49-51, 57, 89, 91, 94, 97
Multiplicative errors, 1-2, 68, 110
Multivariate A distribution, 11-12, 123
Natural response, 84
Negatively skew distributions, 16, 26
Normal equivalent deviate, 68
Normal integral, approximation to, 71
Notation of monograph, 6
of probit analysis, 68
Pearson, E. S., 34, 66, 89
Pearson, K., 3, 21
Philology, 105
Phi-transformation, 101
Physical processes, 104-5
Preferences of consumers, 136-7
Prais, S. J., 102, 121-2
Prices, 103, 134-6, 138-9, 140
Probit analysis, 4, 6, 66-84, 105-6, 125
automatic computing, 143-5
Probit, 4, defined, 68, see Probit analysis
Proportionate effect, law of, 3, 22, 109
Quantal response to a stimulus, defined, 67
Quantiles, 9, 15, 17, 32, 40-2, 74, 88
method of, in estimation, 5, 43-4, 46-50, 56, 61-5, 91, 97
Quantitative response to a stimulus, defined, 67
Region of convergence, 75-8
Reproductive properties, 9-11
Response, 66-7
Rote none, 71-8, 141
Roy, A. D., 108, 110
Rule of false position, 57, 143
Saturation, level of response to a stimulus, 67
Saturation expenditure
defined, 126
parameter, defined, 131
Savings, 123-3
Scaling, in computation, 142-4
Scoring, method of, 70, 70n., 80
Series representations, 17-19
Sheppard’s corrections, 59-4
Skewness, 8-9, 34-5, 142
Small particle statistics, 5, 100-1
Sociology, 101-2
Special devices of estimation, 59-3, 60
Stabilization of variance, 86
Stimulus, 3, 66-9, 84, 131
Subject (probit analysis), 66-7
Substitution, 134-6
Tests of lognormality, 30
Three-parameter A-distribution, 14, 25, 29, 55-64, 114-15, 119
Threshold, 3, 26, 104-5, 143
Time as a stimulus, 106
Time series, 121, 124, 138-9
Tolerance
defined, 67
income, 125
scheme, 126-7
Transformation
criteria for, 86
logarithmic, 3, 21, 86, 101
to normal equivalent deviates, 68
Translation, 21
Truncation, 6-5
defined, 87, 88-94, 103, 113
Upper bound, 16-17
Van Uven, 3, 103
Variance analysis, see Analysis of variance
Weber, H.-Fechner, C. T., law, 3
Wicksell, S. D., 3, 4, 17, 22, 50, 105
Williams, J. L., 104
Wold, H., 29-3, 142
Yuan, P. T., 4, 58, 103
Zero values, 5-6, 94-9
ψ-function, defined, 45, 97, 145
χ-function, defined, 45, 145
STATISTICAL PREDICTION
ANALYSIS

J. AITCHISON
Titular Professor in Statistics
University of Glasgow

I.R. DUNSMORE
Lecturer in Statistics
University of Sheffield

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON NEW YORK NEW ROCHELLE
MELBOURNE SYDNEY
Contents

Preface ix

1 Introduction page 1
   1. The nature of statistical prediction analysis 1
   2. Some examples 1
   3. Examples of choice of future experiment 4
   4. Examples of detection of future experiment 6

2 Predictive distributions 17
   1. Notation and terminology 17
   2. Two sources of information and their combination 18
   3. The predictive density function 19
   4. Predictive distributions for the standard situations 26
   5. Prediction for regression models 32
   6. An application to medical prognosis 35

3 Decisive prediction 45
   1. Point prediction 45
   2. An application to machine tool replacement 50
   3. Set prediction 54
   4. Further analysis of the machine tool problem 57
   5. Computational aids for more general utility functions 60
   6. An alternative formulation 64

4 Informative prediction 68
   1. Introduction 68
   2. Bayesian informative prediction 69
   3. Region of previous experience 72
   4. An application to metabolite excretion rates 73
   5. Bayesian coverage 77
   6. Statistical tolerance regions: role of coverage distributions 79

5 Mean coverage tolerance prediction 88
   1. Introduction 88
   2. Interpretation of similar mean coverage tolerance predictors 91
   3. Additional criteria 92
   4. Relationship to Bayesian coverage 93
Contents

5. Mean coverage tolerance predictors for the binomial case 94
6. Mean coverage tolerance predictors for the Poisson case 97
7. Mean coverage tolerance predictors for the gamma case 99
8. Mean coverage tolerance predictors for the normal case 101
9. Mean coverage tolerance predictors for the multinormal case 104
10. Mean coverage tolerance predictors for the two-parameter exponential case 105

6 Guaranteed coverage tolerance prediction 110
1. Introduction 110
2. Interpretation of similar guaranteed coverage tolerance predictors 113
3. Guaranteed coverage tolerance predictors for the binomial and Poisson cases 114
4. Guaranteed coverage tolerance predictors for the gamma case 117
5. Guaranteed coverage tolerance predictors for the normal case 118
6. Guaranteed coverage tolerance predictors for the multinormal case 123
7. Guaranteed coverage tolerance predictors for the two-parameter exponential case 124

7 Other approaches to prediction 131
1. Introduction 131
2. Linear utility structure 131
3. Frequentist decision theory 132
4. Frequentist linear utility theory 133
5. Prior probabilities on the sample space 135
6. Empirical Bayes prediction 137
7. Distribution-free prediction 138

8 Sampling inspection 146
1. Introduction 146
2. Fixed-size single-sample destructive testing 146
3. Role of mean coverage and guaranteed coverage tolerance predictors in sampling plans 152
4. Optimum choice of fixed sample size 154
5. Sequential predictive sampling inspection 156

9 Regulation and optimisation 162
1. Introduction 162
2. Point regulation 163
3. Set regulation 170
4. Regulation problems with a finite index set 172
5. Optimisation 173

10 Calibration 181
1. The nature of a calibration problem 181
2. The calibrative distribution 185
Contents

3. Calibrative distributions for the normal case 189
4. Two applications of calibrative distributions 192
5. The search for tractable prior distributions 197
6. Calibration under a utility structure 198
7. Some comparisons of methods 199
8. An application to antibiotic assay 205

11 Diagnosis 212
1. The nature of a diagnostic problem 212
2. The diagnostic distribution 215
3. An illustrative example 221
4. Monitoring for atypicality 224
5. Estimative and predictive diagnosis 227
6. An application to differential diagnosis of Conn’s syndrome 231
7. Diagnosis under a utility structure 234

12 Treatment allocation 238
1. The nature of a treatment allocation problem 238
2. The prognostic distributions and utility structure 240
3. Two applications 241
4. Treatment allocation to meet a required specification 246

Appendix I 249
Appendix II 258
Bibliography 259
Author index 267
Subject index 269
Example and problem index 273
Preface

Prediction by its derivation (L. *prædicere*, to say before) means literally the stating beforehand of what will happen at some future time. It is an occupational hazard of many professions: meteorologist, doctor, economist, market researcher, engineering designer, politician and pollster. It is indeed a precarious game because any specific prediction can eventually be compared with the actuality. Many prophets of doom predicting that the world will end at 12.30 on 7 May are left in quieter mood by 12.31. Prediction is a problem simply because of the presence of uncertainty. Seldom, if ever, is it a case of logical deduction; almost inevitably it is a matter of induction or inference. Probabilistic and statistical tools are therefore necessary components of any scientific approach to the formalisation of prediction problems.

In this book we shall be concerned with prediction not only in this narrow sense of making a reasoned statement about what is likely to happen in some future situation but with a much wider class of problems. Any inferential problem whose solution depends on our envisaging some future occurrence will be termed a problem of statistical prediction analysis. The presentation in chapter 1 of a selection of motivating examples illustrates the nature and diversity of statistical prediction analysis, and serves as an introduction to the ingredients of the problem.

A science historian, writing on the development of the concepts and practice of prediction, would probably start by pointing out how primitive man was compelled to attempt prediction, for example the forecasting of the date on which the local river would flood. He might trace how traditions of prediction by magic and sorcery gave way to a realisation that past experience and observation often prove a reliable guide to future events, quoting as an example some well-known folk-lore rhyme about the weather, such as 'A red sky at night is a shepherd's delight'. He would move forward in time to the scientific revolution of the sixteenth and seventeenth centuries highlighting the great advances in descriptive astronomy culminating in what is possibly the greatest predictive achievement of all — the nautical almanack. He might then record the origins of the realisation that there is a relationship between the reliability of a prediction and the inherent variability of the data or the difficulty of accurate measurement, and he could indicate how the science of
statistics in its sophisticated present-day form had emerged from such origins. Eventually he might feel the need for a chapter to explain the attitudes of present-day statisticians to the problem of prediction.

His enquiry into such attitudes would soon reveal an interesting though puzzling situation. If his first step were the natural one of taking stock of well-established theory and application of prediction by surveying statistical text-books he would be singularly disappointed. A search of their indexes would reveal only a small minority which listed the term 'prediction' or its near equivalents. Thorough reading of the texts would bring only a little further enlightenment. He would find a few which quoted a 'prediction interval' or a 'confidence interval for a future observation' towards the end of their discussion on regression analysis, where the regression data are to be used to indicate what is likely to happen when the basic experiment is performed at some specified value of the controllable variable. Moreover in very few texts would he find any clear statement of the principles on which such a prediction interval is based and of the method by which the interval is evolved from these principles. He might fare a little better with the small number of texts which present the 'tolerance interval' approach to prediction, but again he would be left with doubts about the basis, interpretation and usefulness of such intervals.

If he broadened his survey to include research literature and specialist books he would certainly find a fuller account of the tolerance interval approach but it is unlikely that all his doubts would be removed. He would also find a well-developed and still developing theory of prediction for stationary processes. On discovering this he might be forgiven for expressing surprise that a more fully developed theory apparently existed for this more complicated situation than for essentially simpler situations. Apart from these expositions he would find only a mixture of ad hoc techniques of forecasting by trend curves, exponential weighting, etc. He might then begin to wonder why it is that statisticians have devoted so much time, energy and skill to the fields of estimation, hypothesis-testing and experimental design to the comparative neglect of prediction analysis which is surely at the heart of many statistical applications.

Much of statistical analysis is concerned with making inferences about unknown distribution parameters. For example, given the outcomes from an experiment which is known to be $N(\mu, \sigma^2)$ is it reasonable to suppose that $\mu > 0$; what is a confidence interval for $\mu$? Now the purpose of such inference statements is surely to convey to some second party information about what is likely to happen if the experiment is performed again, or perhaps repeated a number of times. It is surprising therefore that greater thought has not been given to the more direct practical type of inference, where statements are required for what is likely to occur when future experiments are performed. Indeed it is common practice for a statistician first to obtain from the
experimental outcome an estimate of the indexing parameter of some class of distributions describing an experiment, and subsequently to use the estimate as if it were the true value to allow prediction. It is paradoxical that while the folly of this approach is pointed out in simple situations there is all too ready acceptance of it in more complicated situations.

This book is an attempt to present certain aspects of statistical prediction theory within a unified framework and notation. As we have already indicated statistical prediction analysis will be here considered in a wide sense, to include any form of statistical analysis where consideration of what may happen (or indeed, to be slightly esoteric at this early stage, what may already have happened) at performances of some future experiment or experiments is essential to the formulation of the problem. The development is considered from both a frequentist and a Bayesian viewpoint for it is our belief that there are situations which are essentially frequentist and other situations which are essentially Bayesian, and the particular type of analysis appropriate to the situation in hand should be used.

It is necessary to provide a comprehensive, yet clear, notation for all the possible distributions involved. Further the notation must not become over-elaborate, for example, when distinguishing between prior and posterior distributions. We believe we have achieved the necessary balance. The notation is introduced as required, mainly in chapters 1, 2 and 3, but a complete list is provided in appendix I.

We are grateful to Dr A.F. Lever of the Medical Research Council Blood Pressure Unit, Western Infirmary, Glasgow, for the data on Conn’s syndrome and to Dr M. Damkjaer Nielsen of Glostrup Hospital, Copenhagen, for the data on Cushing’s syndrome, used for illustrative purposes. The typing of the various drafts and the final version of this book was undertaken by Mrs. I.U. Adey, Miss E.M. Nisbet and Mrs M.S. Robertson. Their care, patience and good humour in the face of a continually changing manuscript played a major role in its eventual completion, and we wish to record our sincere thanks to them.

Glasgow, Sheffield

March 1975

J.A., I.R.D.
1

Introduction

1.1 The nature of statistical prediction analysis

An essential feature of statistical prediction analysis is that it involves two experiments \( e \) and \( f \). From the information which we gain from a performance of \( e \), the informative experiment, we wish to make some reasoned statement concerning the performance of \( f \), the future experiment. In order that \( e \) should provide information on \( f \) there must be some link between these two experiments. Throughout this book we shall deal with problems where this link is through the indexing parameter of the two experiments \( e \) and \( f \), and so we make the following assumption.

**Assumption 1** The class of probability models which form the possible descriptions of \( e \) and the class of possible models for \( f \) have the same index set \( \Theta \), and the true models have the same (though unknown) index \( \theta^* \).

A further general feature of all the problems we shall consider is contained in the following independence assumption.

**Assumption 2** For given index \( \theta \) the experiments \( e \) and \( f \) are independent.

By adopting this second assumption we deliberately exclude a range of prediction problems in which \( f \) is a continuation of some stochastic process of which \( e \) records a realisation to date. Techniques such as forecasting by exponential weighting, linear least squares prediction and time series analysis are thus outside the scope of this book.

To give some idea of the wide applicability of statistical prediction analysis as defined above and to motivate the development of appropriate theory we devote the remainder of this chapter to the presentation of specific prediction problems. All these problems are later analysed and extended in the sections indicated in the text.

1.2 Some examples

In its most direct form statistical prediction analysis may simply be the
provision of some probabilistic statement about the likely outcome of the
derformance of $f$.

<table>
<thead>
<tr>
<th>Table 1.1 Survival times (weeks) of 20 carcinoma patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
</tr>
<tr>
<td>51</td>
</tr>
</tbody>
</table>

**Example 1.1**

**Medical Prognosis.** The data of table 1.1 are the survival times (weeks) of 20 patients presenting with a certain type of carcinoma and receiving treatment of preoperative radiotherapy followed by radical surgery. On the basis of this information what can appropriately be said about the future of a new patient with this type of carcinoma and assigned to this form of treatment? Clearly any rational statement would regard 100 weeks survival as much more plausible than 500 weeks survival, but how should such views be summarised and quantified? What is a reasonable assessment of the probability that the patient will survive 100 weeks?

In this example the informative experiment $e$ consists of recording the survival times of the 20 patients already treated. The future experiment $f$ consists of treating the new patient similarly and recording his survival time. If no change in the treatment has been made since the conducting of $e$, then $e$ and $f$ consist respectively of 20 replicates and a single replicate of the same basic trial (record the survival time of a treated patient) and are independent. Assumptions 1 and 2 are therefore satisfied.

Attempts to quantify medical prognoses are of vital importance when similar information on an alternative treatment, for example radical surgery followed by postoperative radiotherapy, is available and a choice has to be made between treatments for a particular patient.

A detailed analysis and developments of this example are given in §2.6. There are many less direct forms of statistical analysis than that of providing probabilistic statements. For example the problem may be one of choosing between alternative courses of action. If the consequences of taking a course of action depend on the outcome of $f$, then we still technically describe the problem as one of statistical prediction analysis.

**Example 1.2**

**Machine tool replacement.** Table 1.2 shows the recorded lifetimes of 24 machine tools of a certain type. In a factory using one of these machine tools the question of the best inspection and replacement policy is under discussion. If the tool wears out while unattended there is a loss of $\xi = 1.8$ per minute until such time as it is inspected (and immediately replaced). To have an inspector in attendance at the machine tool costs $\eta = 2.4$ per minute. If the
Introduction

Table 1.2 Lifetimes (minutes) of 24 machine tools

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>62</td>
<td>111</td>
<td>47</td>
<td>57</td>
<td>14</td>
<td>290</td>
<td>118</td>
<td>19</td>
<td>4</td>
<td>17</td>
<td>46</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>239</td>
<td>9</td>
<td>140</td>
<td>89</td>
<td>94</td>
<td>217</td>
<td>35</td>
<td>103</td>
<td>28</td>
<td>37</td>
<td>111</td>
<td></td>
</tr>
</tbody>
</table>

tool is replaced before it wears out an overhead cost $\xi = 54$ is incurred and also a loss at rate $\xi$ per minute of its unused lifetime has to be debited. What is the optimum policy on the basis of all the information?

Here again the experiments $e$ and $f$ are easily identified. If we regard a basic trial as consisting of the recording of the lifetime of a single machine tool then $e$ and $f$ consist of 24 replicates and a single replicate of the basic trial, and satisfy assumptions 1 and 2. Again we clearly wish to infer something about the performance of $f$ (the current machine tool) from the information contained in $e$. But a statement about the relative plausibilities of the possible outcomes of $f$ is not sufficient. We must decide on one of many possible courses of action, namely the time periods during which we wish an inspector to be present to investigate whether the tool is still functioning or has already broken down. Suppose for simplicity that the only courses of action open to us are to select a time, $a$ say, at which to send in an inspector with instructions to replace the tool immediately. The consequences of taking action $a$ depend on the outcome of $f$, the actual lifetime $y$ of the current machine tool. For if $a > y$ there is lost production time $a - y$ with a corresponding loss of $\xi(a - y)$, whereas if $y > a$, the overhead scrapping loss $\xi$ is incurred together with a debit of $\xi(y - a)$ for unused productive capacity. Thus a prediction associated with the performance of the future experiment $f$ is necessary for any rational analysis of the problem but the prediction is a means to the end of selecting an appropriate course of action. We have thus here a less direct form of statistical prediction analysis than in our previous example.

The analysis of this problem is developed in §§3.2, 3.4.

Example 1.3

A quality control problem. Items are produced independently in large batches by a firm. The items may be either effective or defective and it is recognised by both manufacturer and customer that batches vary considerably in the number of effectives they contain. The terms of a suggested contract between manufacturer and customer require the manufacturer to test destructively 5 of the components of each batch. The remainder of the batch is to be supplied in packets of 25 with an accompanying statement about the maximum number of defectives each packet contains. The contract further requires that for at least 90 per cent of such batches the statement will be true for at least 80 per cent of packets. What statement strategy will fulfil the terms of this contract?

Here the future experiment $f$ envisaged is the observation of the number of defectives in a packet of 25 components from a batch. The information that
we have available consists of the observation of the number \( x \) of defectives in a sample of 5 components.

How we use the information in \( e \) to meet the requirements on \( f \) and how we interpret the 90 per cent and 80 per cent in the statements are discussed later in §6.3.

In the examples so far discussed we have attempted to illustrate some of the basic structure of statistical prediction analysis. The main feature emerging is the need to make some prediction of the outcome \( y \) of a future experiment \( f \) based on the outcome \( x \) of an informative experiment \( e \). The relevance of the information obtained from \( e \) to the future experiment \( f \) is contained in what can conveniently be called the predictive density function. This concept is dealt with in detail in chapter 2 and is central to much of the subsequent analysis. For example the predictive density function provides the quantification of medical prognosis that we seek for example 1.1. In chapter 3 the introduction of utility functions to quantify the measures of gains or losses involved in making a prediction enables decisive prediction problems such as example 1.2 to be analysed in detail. Informative prediction problems in which no such measures are available, as in example 1.3, are dealt with in chapters 4, 5 and 6. A substantial part of these chapters presents the theory of tolerance regions from a fresh viewpoint. Some interesting relationships between decisive and informative prediction are developed in chapter 7, which also reviews other approaches such as empirical Bayes and distribution-free prediction. The remainder of the book is then devoted to specific areas of application and particularly to even more indirect forms of statistical prediction analysis. Some of these forms are now illustrated by examples.

1.3 Examples of choice of future experiment

There are many problems in which there is a whole class \( F \) of possible future experiments and the problem is to determine which future experiment \( f \) satisfies certain desirable properties. In choosing this experiment we have to envisage its performance and for this reason such problems fall within the scope of statistical prediction analysis. We consider here two representative examples.

Example 1.4

A problem of laminate design. In the manufacture of a laminate several sheets of material are superimposed. The sheets are liable to contain flaws and the total number \( y \) of flaws in the finished product can be measured by an X-ray device. The durability of the product increases as the number \( t \) of component sheets increases, but there is an upper limit \( y_0 \) to the total number of flaws which can be allowed before the product is rejected.
Introduction

Table 1.3 Numbers of flaws in nine laminate specimens

<table>
<thead>
<tr>
<th>Number of sheets</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of flaws</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

In a pilot experiment 9 specimens of the product were made with 5, 6, 7, ..., 13 sheets of material superimposed with resulting flaws as shown in table 1.3. The management wishes to market as durable a product as possible but has decided that at most \( v_0 = 7 \) flaws can be allowed in any product. If the profit per component sheet for accepted products is ten times the cost per component sheet for rejected products how many sheets should be superimposed?

In order to resolve this problem we have to consider the number of flaws which may result if we use \( t \) sheets. Thus we are forced to envisage a whole class of possible experiments

\[
F = \{ f_t : t = 1, 2, \ldots \},
\]

where \( f_t \) denotes the experiment of counting the total number of flaws in a laminate of \( t \) sheets. The problem is then to choose which future experiment \( f_t \) gives as high durability as possible and yet attempts to meet the flaw limitation. The information available comes from an informative experiment \( e \) yielding the data of table 1.3 and which could formally be written in the form

\[
e = \{ f_5, f_6, \ldots, f_{13} \},
\]
a set of independent performances of \( f_5, f_6, \ldots, f_{13} \). The direct prediction problem for \( f_t \) enters into our attempts to balance our desire to increase \( t \) (and hence the durability) and our concern that the number of flaws may increase beyond the acceptable limit. This regulation example is analysed in §9.3.

Example 1.5

Maximising yield of an industrial process. The yields (kg) shown in table 1.4 were obtained in an experiment in which an industrial process was run successively at 5 different temperatures and 3 different pressures, each combination of temperature and pressure being used twice. What combination of temperature and pressure should be used in order to maximise the yield in a future run of the process?

Here the problem is to determine at what combination \( t = (t_1, t_2) \) of temperature \( t_1 \) and pressure \( t_2 \) to run the process. Denote by \( f_t \) or \( f_{t_1, t_2} \) the future experiment which records the yield from an operation of the industrial process at temperature \( t_1 \) and pressure \( t_2 \). We are thus forced to consider the class of future experiments
Introduction

Table 1.4 Yields (kg) from 30 process runs at different temperature-pressure combinations

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Pressure (atmospheres)</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>65, 68</td>
<td>70, 72</td>
<td>73, 74</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>72, 70</td>
<td>75, 75</td>
<td>77, 76</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>73, 75</td>
<td>81, 83</td>
<td>79, 78</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>76, 75</td>
<td>81, 79</td>
<td>75, 77</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>76, 76</td>
<td>78, 80</td>
<td>76, 73</td>
<td></td>
</tr>
</tbody>
</table>

\[ F = \{ f_t : t \in T \}, \]

where \( T \) denotes the set of possible temperature–pressure combinations. The informative experiment consists of 30 independent experiments (process runs) each of \( f_t \) type. To choose an experiment from the set \( F \) we must again envisage the outcomes of such future experiments and so are involved in statistical prediction analysis. This *optimisation* problem is analysed in §9.5.

1.4 Examples of detection of future experiment

A common statistical problem is to detect which one of a class \( F \) of ‘future’ experiments has already been performed from the information from \( e \) and the known outcome \( y \) of the performed future experiment. While in such circumstances it may seem strange to use the term statistical prediction analysis we shall see that we are led inevitably to the same concepts of prediction as we have already encountered. Indeed we have to envisage prediction for each of the possible future experiments in \( F \). Two examples illustrate the nature of the problem here.

Example 1.6

*Antibiotic assay.* When a droplet of specified volume of an antibiotic is placed on an infected medium on a Petri dish and kept under controlled conditions the antibiotic clears a circular area of the medium. Moreover the diameter of the cleared area depends on the concentration of the antibiotic although this relationship is not a deterministic one. The idea underlying a biological assay of the unknown concentration of antibiotic in a blood specimen from a patient is to place droplets of standard antibiotic at different known concentrations and droplets from the specimen on the same batch of infected medium (fig. 1.1). The problem is then to infer from the relative sizes of the diameters associated with droplets of known concentration and of the diameters associated with droplets of the unknown concentration as much as possible about the unknown concentration. Such a direct comparison between the patient’s specimen and the standard is usually necessary because the relationship between diameter and concentration usually varies from batch to batch of the medium.
Fig. 1.1 Typical clearance circles in an antibiotic assay. For circles from standard droplets the known concentrations (mcg/ml) are shown. The three circles labelled ? are from droplets from a single specimen of unknown concentration.

Table 1.5 shows the results of an experiment to investigate the feasibility and reliability of this type of assay for a particular antibiotic. From the same batch of infected medium 20 Petri dishes were prepared and on each one droplet at each of seven concentrations (mcg/ml) was placed and the resulting clearance diameters recorded. Typical questions that we have to be in a position to answer are the following.

(1) Suppose that a droplet from a particular specimen has been placed on medium from this batch and has cleared a circle of diameter 19 mm. What can we infer about the concentration of antibiotic in the specimen?

(2) If three droplets from the same blood specimen have given clearance diameters of 18.0, 19.5 and 19.5 mm, what can be inferred about the concentration of the antibiotic in the specimen?
Table 1.5 Clearance diameters (mm) on 20 Petri dishes for different concentrations (mcg/ml) of antibiotic

<table>
<thead>
<tr>
<th>Dish no.</th>
<th>Concentration (mcg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>21.5</td>
</tr>
<tr>
<td>2</td>
<td>24.5</td>
</tr>
<tr>
<td>3</td>
<td>21.0</td>
</tr>
<tr>
<td>4</td>
<td>20.5</td>
</tr>
<tr>
<td>5</td>
<td>21.5</td>
</tr>
<tr>
<td>6</td>
<td>22.0</td>
</tr>
<tr>
<td>7</td>
<td>21.0</td>
</tr>
<tr>
<td>8</td>
<td>20.5</td>
</tr>
<tr>
<td>9</td>
<td>21.5</td>
</tr>
<tr>
<td>10</td>
<td>22.5</td>
</tr>
<tr>
<td>11</td>
<td>22.5</td>
</tr>
<tr>
<td>12</td>
<td>22.5</td>
</tr>
<tr>
<td>13</td>
<td>21.5</td>
</tr>
<tr>
<td>14</td>
<td>22.0</td>
</tr>
<tr>
<td>15</td>
<td>23.5</td>
</tr>
<tr>
<td>16</td>
<td>21.5</td>
</tr>
<tr>
<td>17</td>
<td>21.0</td>
</tr>
<tr>
<td>18</td>
<td>22.5</td>
</tr>
<tr>
<td>19</td>
<td>22.0</td>
</tr>
<tr>
<td>20</td>
<td>21.0</td>
</tr>
</tbody>
</table>

An entry * indicates that no measurable clearance was achieved.

(3) For medical reasons we may wish to be reasonably sure that the concentration quoted for a particular specimen is within 10 per cent of the true value. How many droplets from the specimen should be used to achieve this reliability?

Let \( f \) denote an experiment which records the clearance diameter of a droplet of concentration \( t \) of the antibiotic, and consider the class

\[ F = \{ f : t \in T \} \]

of experiments indexed by the set \( T \) of all possible concentrations. Then the informative experiment \( e \) which yields the data of table 1.5 is clearly of regression type and may be expressed briefly as

\[ e = \{ f_1, \ldots, f_{140} \}, \]

a set of 140 independent experiments, where \( f_1, \ldots, f_{140} \) are the concentrations associated with the 140 droplets. The data from this informative experiment thus consist of 140 pairs \( (t_1, x_1), \ldots, (t_{140}, x_{140}) \) of concentrations and corresponding clearance diameters, and can thus be set out in a typical scatter diagram (fig 1.2). For convenience a logarithmic concentration scale has been used.

In considering the first question posed we have the outcome \( y = 19 \) of some experiment from the class \( F \), say \( f_u \) where \( u \) is the unknown concentration.
Fig. 1.2 Variability of clearance diameter with concentration of antibiotic. The line shown joins the mean clearance diameters at the different concentrations.
<table>
<thead>
<tr>
<th>Patient no.</th>
<th>Age (Years)</th>
<th>Na (meq/l)</th>
<th>K (meq/l)</th>
<th>CO₂ (meq/l)</th>
<th>Renin (meq/l)</th>
<th>Aldosterone (meq/l)</th>
<th>Systolic (mm Hg)</th>
<th>Diastolic (mm Hg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>40</td>
<td>140.6</td>
<td>2.3</td>
<td>30.3</td>
<td>4.6</td>
<td>121.0</td>
<td>192</td>
<td>107</td>
</tr>
<tr>
<td>A2</td>
<td>37</td>
<td>143.0</td>
<td>3.1</td>
<td>27.1</td>
<td>4.5</td>
<td>15.0</td>
<td>230</td>
<td>150</td>
</tr>
<tr>
<td>A3</td>
<td>34</td>
<td>140.0</td>
<td>3.0</td>
<td>27.0</td>
<td>0.7</td>
<td>19.5</td>
<td>200</td>
<td>130</td>
</tr>
<tr>
<td>A4</td>
<td>48</td>
<td>146.0</td>
<td>2.8</td>
<td>33.0</td>
<td>3.3</td>
<td>30.0</td>
<td>213</td>
<td>125</td>
</tr>
<tr>
<td>A5</td>
<td>41</td>
<td>138.7</td>
<td>3.6</td>
<td>24.1</td>
<td>4.9</td>
<td>20.1</td>
<td>163</td>
<td>106</td>
</tr>
<tr>
<td>A6</td>
<td>22</td>
<td>143.7</td>
<td>3.1</td>
<td>28.0</td>
<td>4.2</td>
<td>33.0</td>
<td>190</td>
<td>120</td>
</tr>
<tr>
<td>A7</td>
<td>27</td>
<td>137.3</td>
<td>2.5</td>
<td>29.6</td>
<td>5.4</td>
<td>52.1</td>
<td>220</td>
<td>140</td>
</tr>
<tr>
<td>A8</td>
<td>18</td>
<td>141.0</td>
<td>2.5</td>
<td>30.0</td>
<td>2.5</td>
<td>50.2</td>
<td>210</td>
<td>135</td>
</tr>
<tr>
<td>A9</td>
<td>53</td>
<td>143.8</td>
<td>2.4</td>
<td>32.2</td>
<td>1.5</td>
<td>68.9</td>
<td>160</td>
<td>105</td>
</tr>
<tr>
<td>A10</td>
<td>54</td>
<td>144.6</td>
<td>2.9</td>
<td>29.5</td>
<td>3.0</td>
<td>144.7</td>
<td>213</td>
<td>135</td>
</tr>
<tr>
<td>A11</td>
<td>50</td>
<td>139.5</td>
<td>2.3</td>
<td>26.0</td>
<td>2.6</td>
<td>31.2</td>
<td>205</td>
<td>125</td>
</tr>
<tr>
<td>A12</td>
<td>44</td>
<td>144.0</td>
<td>2.2</td>
<td>33.7</td>
<td>3.9</td>
<td>65.1</td>
<td>263</td>
<td>133</td>
</tr>
<tr>
<td>A13</td>
<td>44</td>
<td>145.0</td>
<td>2.7</td>
<td>33.0</td>
<td>4.1</td>
<td>38.0</td>
<td>203</td>
<td>115</td>
</tr>
<tr>
<td>A14</td>
<td>66</td>
<td>140.2</td>
<td>3.1</td>
<td>29.1</td>
<td>4.7</td>
<td>43.1</td>
<td>195</td>
<td>115</td>
</tr>
<tr>
<td>A15</td>
<td>39</td>
<td>144.7</td>
<td>2.9</td>
<td>27.4</td>
<td>0.9</td>
<td>65.1</td>
<td>180</td>
<td>120</td>
</tr>
<tr>
<td>A16</td>
<td>46</td>
<td>139.0</td>
<td>3.1</td>
<td>31.4</td>
<td>2.8</td>
<td>192.7</td>
<td>228</td>
<td>133</td>
</tr>
<tr>
<td>A17</td>
<td>48</td>
<td>144.8</td>
<td>1.9</td>
<td>33.5</td>
<td>3.8</td>
<td>103.5</td>
<td>205</td>
<td>132</td>
</tr>
<tr>
<td>A18</td>
<td>38</td>
<td>145.7</td>
<td>3.7</td>
<td>27.4</td>
<td>2.8</td>
<td>42.6</td>
<td>203</td>
<td>117</td>
</tr>
<tr>
<td>A19</td>
<td>60</td>
<td>144.0</td>
<td>2.2</td>
<td>33.0</td>
<td>3.2</td>
<td>92.0</td>
<td>220</td>
<td>120</td>
</tr>
<tr>
<td>A20</td>
<td>44</td>
<td>143.5</td>
<td>2.7</td>
<td>27.5</td>
<td>3.6</td>
<td>74.5</td>
<td>210</td>
<td>114</td>
</tr>
</tbody>
</table>

(continued)
Table 1.6 (continued)

<table>
<thead>
<tr>
<th>Patient no.</th>
<th>Age (Years)</th>
<th>Na (meq/l)</th>
<th>K (meq/l)</th>
<th>CO(_2) (meq/l)</th>
<th>Renin (meq/l)</th>
<th>Aldosterone (meq/l)</th>
<th>Blood Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>46</td>
<td>140.3</td>
<td>4.3</td>
<td>23.4</td>
<td>6.4</td>
<td>27.0</td>
<td>270 160</td>
</tr>
<tr>
<td>B2</td>
<td>35</td>
<td>141.0</td>
<td>3.2</td>
<td>25.0</td>
<td>8.8</td>
<td>26.3</td>
<td>210 130</td>
</tr>
<tr>
<td>B3</td>
<td>50</td>
<td>141.2</td>
<td>3.6</td>
<td>25.8</td>
<td>4.1</td>
<td>20.9</td>
<td>181 113</td>
</tr>
<tr>
<td>B4</td>
<td>41</td>
<td>142.0</td>
<td>3.0</td>
<td>22.0</td>
<td>4.7</td>
<td>20.4</td>
<td>260 160</td>
</tr>
<tr>
<td>B5</td>
<td>57</td>
<td>143.5</td>
<td>4.2</td>
<td>27.8</td>
<td>4.3</td>
<td>23.7</td>
<td>185 125</td>
</tr>
<tr>
<td>B6</td>
<td>57</td>
<td>139.7</td>
<td>3.4</td>
<td>28.0</td>
<td>5.2</td>
<td>46.0</td>
<td>240 130</td>
</tr>
<tr>
<td>B7</td>
<td>48</td>
<td>141.1</td>
<td>3.6</td>
<td>25.0</td>
<td>2.5</td>
<td>37.3</td>
<td>197 120</td>
</tr>
<tr>
<td>B8</td>
<td>60</td>
<td>141.0</td>
<td>3.8</td>
<td>26.0</td>
<td>6.5</td>
<td>23.4</td>
<td>211 118</td>
</tr>
<tr>
<td>B9</td>
<td>52</td>
<td>140.5</td>
<td>3.3</td>
<td>27.0</td>
<td>4.2</td>
<td>24.0</td>
<td>168 104</td>
</tr>
<tr>
<td>B10</td>
<td>49</td>
<td>140.0</td>
<td>3.6</td>
<td>26.0</td>
<td>6.3</td>
<td>39.8</td>
<td>220 120</td>
</tr>
<tr>
<td>B11</td>
<td>49</td>
<td>140.0</td>
<td>4.4</td>
<td>25.6</td>
<td>5.1</td>
<td>47.0</td>
<td>190 125</td>
</tr>
<tr>
<td>Undiagnosed</td>
<td>50</td>
<td>143.3</td>
<td>3.2</td>
<td>27.0</td>
<td>8.5</td>
<td>51.0</td>
<td>210 130</td>
</tr>
</tbody>
</table>

A: adenoma
B: bilateral hyperplasia
The problem here differs from that of previous examples in that the 'future' experiment has already been performed; we know its outcome but we do not know which experiment in the class $F$ has been performed. We have thus a kind of inverse prediction problem, but we shall find in its solution that the same concepts of statistical prediction analysis arise as in more direct problems.

The second and third questions are very similar in nature. To deal with the second question we have the outcome $(y_1, y_2, y_3)$ of the 'future' experiment $\{f_0, f_a, f_u\}$ consisting of three replicates of $f_a$, and again we face the problem of detecting what $f_a$ has been performed. To analyse the third question we have to envisage the consequences of the 'future' experiment consisting of $K$ replicates of some $f_a$ and investigate how these consequences vary with $K$.

All these questions, which fall within the general class of calibration problems, are fully analysed in §10.8.

Example 1.7

Preoperative medical diagnosis. Recently at operation it has been discovered that a rare syndrome of hypertension (Conn's syndrome) can be due to either (1) a benign tumour (adenoma) in the adrenal cortex or (2) a more diffuse condition (bilateral hyperplasia) of the adrenal glands. The assessment of treatment, which may range from total adrenalectomy, through removal of an adenomatous adrenal gland if locatable, to drug therapy, is now recognised to depend on the diagnostic assessment and on a number of factors external to the diagnostic assessment. It is therefore highly desirable to be in a position to obtain preoperatively a reasonable diagnostic assessment of the relative plausibilities of the two types. Since radiology is not yet a reliable diagnostic tool in this differential diagnostic problem, the possibility of diagnosing preoperatively from eight diagnostic tests has been considered. Table 1.6 shows the results of these eight preoperative tests on 31 patients who have in earlier years been operated on and their type, 1 or 2, definitely established. The eight tests have now been carried out on a new patient known to have Conn's syndrome but of as yet undiagnosed form. The results of these eight tests are shown in the final row of table 1.6.

The reader may recognise this problem under some different statistical heading such as discriminant analysis. We shall show that we gain both in simplicity and in practical application by considering it within the unifying framework of statistical prediction analysis. Indeed its form is similar to that of example 1.6, the only difference being that in the calibration example the index set $T$ for the class $F$ is continuous whereas in the present context $T$ is discrete. Let $f_t$ denote the experiment of recording the eight test results of a patient in category $t$; and let $F = \{f_t: t = 1, 2\}$,
so that \( T = \{1, 2\} \) is the index set of the class \( F \). If we regard our ‘future’ experiment \( f \) as the recording of the eight test results for the new patient then while we know the outcome of the future experiment we do not know which experiment of \( F \) has been performed. As in the calibration problem we shall find that prediction plays a central role in the satisfactory analysis of this diagnostic problem. The informative experiment \( e \) which yields the data of table 1.6 consists essentially of 20 replicates (with different adenoma patients) of \( f_1 \) and 11 replicates (with different bilateral hyperplasia patients) of \( f_2 \).

For the further analysis of this diagnostic problem see §11.6.

Problems

For each of the following prediction problems identify the informative experiment \( e \) and the future experiment \( f \). These problems will be set for fuller analysis in later chapters, and there is no need to attempt to resolve them or to answer the questions posed.

1.1 The biparietal diameter of the foetal head can be measured by ultrasonic cephalometry while the baby is still in the mother’s womb. This measurement gives a good indication of the size of the baby. To allow for the possibility of inducing labour prematurely if the baby is not developing, it would be useful to have some indication of the range of likely values of the diameter during the 7th month of pregnancy for normal babies. Provide such a range of normality if the measurements shown below are those taken during the 7th month for 50 normal babies who were born at full term.

<table>
<thead>
<tr>
<th>Biparietal diameter (cm)</th>
<th>8.91</th>
<th>9.53</th>
<th>9.23</th>
<th>8.04</th>
<th>8.42</th>
<th>8.85</th>
<th>7.51</th>
<th>8.59</th>
<th>8.06</th>
<th>8.93</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.98</td>
<td>9.25</td>
<td>8.80</td>
<td>8.61</td>
<td>8.62</td>
<td>8.84</td>
<td>8.24</td>
<td>8.16</td>
<td>9.20</td>
<td>8.10</td>
</tr>
<tr>
<td></td>
<td>9.02</td>
<td>8.15</td>
<td>8.06</td>
<td>8.71</td>
<td>8.91</td>
<td>9.08</td>
<td>8.25</td>
<td>8.71</td>
<td>8.88</td>
<td>7.96</td>
</tr>
<tr>
<td></td>
<td>9.46</td>
<td>8.93</td>
<td>8.54</td>
<td>9.78</td>
<td>8.39</td>
<td>8.65</td>
<td>7.94</td>
<td>8.48</td>
<td>8.76</td>
<td>9.20</td>
</tr>
<tr>
<td></td>
<td>8.78</td>
<td>8.86</td>
<td>9.00</td>
<td>9.21</td>
<td>8.44</td>
<td>9.20</td>
<td>8.58</td>
<td>8.98</td>
<td>8.76</td>
<td>9.06</td>
</tr>
</tbody>
</table>

1.2 Close study of ten patients discharged from hospital after treatment for a chronic disorder has been undertaken and the ten observed times to first relapse found to have average 100 days.

The clinic is now trying to formulate a reasonable policy for recall time of patients which, for administrative purposes, must be the same for all patients. On the assumption that times to relapse have an exponential distribution what recall time for future patients would you suggest to ensure that 95 per cent of patients are recalled for retreatment before relapse?

A further attempt to rationalise policy suggests that the disutility for any patient who relapses before recall is 3 units for each day until recall day,
whereas for patients who have not relapsed by recall day the disutility can be regarded as 1 unit for each day by which recall precedes the day on which relapse would have taken place. How does this affect the previous policy? What is the expected disutility per patient with this policy?

1.3 An income tax department has been asked to investigate the consequences of changing from a tax based on personal income to one based on household income, and to set a minimum taxable household income so that at least 25 per cent of households are exempt from the tax. The incomes (in £s) of 50 randomly selected households are as shown below; and previous studies suggest that logarithms of household incomes are normally distributed. The department likes to be 90 per cent sure of any statement it issues. What minimum taxable household income should be quoted?

1.4 In a study on the setting strength of a woodwork adhesive the following procedure is carried out. Adhesive is applied to two strips of wood. After an interval \( t_1 \) the two strips are clamped together at right angles for time \( t_2 \). Then the force \( x \) (in kgs) required to prize the strips apart is measured. In the study the two factors \( f \) and \( t_2 \) are varied as shown, 48 experiments being performed following the prescribed procedure for 3 repetitions of each of the 16 possible pairings.

What values of \( f \), \( t_2 \) should the manufacturer recommend if the user wishes to maximise the strength of his joints?
1.5 The technique of radiocarbon dating can be used to date archaeological specimens. The amount of radiocarbon in the specimen is determined by measuring the rate of disintegration of Carbon 14 in the specimen. From this measurement a radiocarbon age can be evaluated. A calibration curve is obtained by assessing the radiocarbon ages of samples of the bristlecone pine tree. The true ages of these samples can be accurately determined from the tree ring markings, and the bristlecone pine tree is particularly suitable for the purpose since the species is capable of survival for over 4500 years. The data are shown below for the relevant period of history.

A newly found archaeological specimen has a radiocarbon age of 4010 years. What is its true age if prior archaeological knowledge suggests a value in the region 3700 to 4500 years old?

| Radiocarbon age (yrs) | 3300 | 3500 | 3690 | 3510 | 3800 | 3800 |
| Bristlecone pine age (yrs) | 3500 | 3600 | 3720 | 3750 | 3910 | 4030 |
| Radiocarbon age (yrs) | 4090 | 3910 | 4110 | 4000 | 4200 | 4210 |
| Bristlecone pine age (yrs) | 4210 | 4220 | 4370 | 4440 | 4600 | 4620 |

1.6 When running under control a complex chemical process yields an output for which two quantitative characteristics are positively correlated. There are two points a and b at which the process may run out of control, but it is difficult to determine which without costly dismantling. When the process runs out of control the correlation between the characteristics is thought to be less strong or even negative. In the most recent 26 production runs after each of which dismantling of a and b was carried out to determine any loss of control results were obtained as in the table below.

The problem is to decide on the basis of the characteristics observed for the output from a new run whether to investigate at points a and b before the next run. The table below shows the losses involved in the various (action, fault) combinations.

<table>
<thead>
<tr>
<th>Loss of control</th>
<th>Observed characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>(36.5, 66.4), (29.8, 61.7), (33.0, 63.8), (39.0, 68.2), (35.0, 66.7)</td>
</tr>
<tr>
<td></td>
<td>(35.8, 67.4), (40.7, 72.2), (35.5, 68.4), (34.8, 65.2), (37.0, 67.0)</td>
</tr>
<tr>
<td></td>
<td>(40.7, 70.6), (33.4, 65.3), (34.5, 68.0), (31.3, 63.4), (38.9, 71.2)</td>
</tr>
<tr>
<td>At a only</td>
<td>(38.6, 66.5), (37.4, 67.3), (38.8, 65.4)</td>
</tr>
<tr>
<td>At b only</td>
<td>(33.5, 67.4), (34.2, 70.5), (32.6, 69.4), (34.1, 68.9), (35.5, 69.2)</td>
</tr>
<tr>
<td>At both a and b</td>
<td>(32.8, 68.2), (32.2, 66.8), (33.1, 66.5)</td>
</tr>
</tbody>
</table>
### Introduction

<table>
<thead>
<tr>
<th>Action taken</th>
<th>Loss of control</th>
<th>At a only</th>
<th>At b only</th>
<th>At both a and b</th>
</tr>
</thead>
<tbody>
<tr>
<td>No dismantling</td>
<td>0 5 6 9</td>
<td>2 0 4 3</td>
<td>4 3 0 both a and b</td>
<td></td>
</tr>
<tr>
<td>Dismantle a</td>
<td>3 4 0 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dismantle b</td>
<td>4 3 3 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dismantle both a and b</td>
<td>4 3 3 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2

Predictive distributions

2.1 Notation and terminology

Throughout this book we make use of the increasingly familiar device of labelling functions by their arguments. Suppose that we specify a density function $p(\theta)$ on a space $\Theta$ and that, for each $\theta$ in $\Theta$, we specify a density function $p(y|\theta)$ on a space $Y$. The vertical bar preceding $\theta$ denotes conditioning on this known value of the index. The absolute or unconditional density function $p(y)$ on $Y$ is then given by

$$p(y) = \int_{\Theta} p(y|\theta)p(\theta)d\theta. \quad (2.1)$$

Note that $p(\theta)$ and $p(y)$ are completely different functions. We shall never have to attach numerical values to $\theta$ and $y$ in the theoretical development; these symbols will therefore always be present to ensure the distinction between the two functions. No confusion arises from this device and the advantages of it are overwhelming. By its use we can express concepts with greater clarity and can avoid overdecorating the structure of the analysis.

We use the term density function not only in its usual context of describing a spread of the unit of probability over a continuous space but also to describe the breaking up of the unit of probability into discrete pieces on a discrete space. For example we speak not only of the exponential density function

$$p(y|\theta) = \exp(-\theta y) \quad (y > 0)$$
on the positive real line, but also of the Poisson density function

$$p(y|\theta) = \theta^y \exp(-\theta)/y! \quad (y = 0, 1, 2, ... )$$
on the set of non-negative integers. Note also the device of stating the effective sample space, the domain of non-zero probability density, within brackets in the specification of the density function. Moreover, in theoretical developments where spaces may be either continuous or discrete, we always use the integral sign $\int$, as in (2.1), to indicate the mathematical process of accumulation, whether the process is integration or summation.
Predictive distributions

For standard distributions a mnemonic notation, which will allow complicated results to be simply stated and readily assimilated, will be introduced as it is required. For reference purposes, however, appendix I provides a complete list of this notation.

2.2 Two sources of information and their combination

Suppose that there is some future experiment \( f \) in which we are interested, and that for its probabilistic description we have a sample space \( Y \) and a class of possible density functions

\[
\{p(y|\theta): \theta \in \Theta\}
\]
on \( Y \). The indexing set \( \Theta \) is assumed to be known but the true index \( \theta^* \) is not known. This uncertainty about the true index is the source of the statistical nature of the prediction problem. If we knew \( \theta^* \) we would know precisely the density function \( p(y|\theta^*) \) describing our uncertainty about the outcome of the future experiment \( f \). And we could do no better than this.

Our uncertainty about the true index may be alleviated by information from two sources.

(i) Prior information about \( \theta \). At this stage of statistical prediction analysis we adopt a Bayesian approach. We assume that while we do not know the true index \( \theta^* \) we can place plausibilities or probabilities on the various possible \( \theta \) in \( \Theta \). More precisely we assume that we have a known density function \( p(\theta) \) on \( \Theta \). Later in chapters 4, 5 and 6 we shall examine the considerable alteration and complications to the statistical analysis when such information is not available.

(ii) Informative experiment. We suppose that it is possible to perform some informative experiment \( e \) with sample space \( X \). The way in which the outcome of such an experiment conveys information about the true index is through assumption 1 of §1.1, that the index set of the class of possible density functions describing \( e \) is also \( \Theta \) and that the true index operating is \( \theta^* \), the same as for \( f \). The class of density functions for \( e \) can then be denoted by

\[
\{p(x|\theta): \theta \in \Theta\}
\]
on \( X \).

The information we obtain by observing \( x \) at the performance of the informative experiment \( e \) clearly influences the prior plausibilities \( p(\theta) \) we have attached to the possible indices. The updating of the plausibilities in the light of the observation \( x \) is effected by the mechanism of Bayes's theorem, which in our notation leads to the posterior plausibility function

\[
p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)},
\]

(2.2)
Predictive distributions

where

\[ p(x) = \int_\Theta p(\theta) p(x|\theta) d\theta. \]  \hspace{1cm} (2.3)

We introduce in table 2.1 some notation for the standard probability distributions which may apply to the informative experiment, together with the conjugate prior distributions for the parameters. We emphasise again that a complete list of the notation used is contained in appendix 1, and therein can be found the meanings of such symbols as \( D(g, h), S^d, \Gamma_d, \) etc., which are used in table 2.1. The choice of family of prior plausibility functions for any particular situation is inevitably a nice judgement between mathematical tractability and practical considerations, and we adopt the now familiar Bayesian device of using the rich families of conjugate prior distributions. By choosing such conjugate prior distributions we ensure that the prior and the posterior distributions are both of the same family. Also shown in the final column of the table are the particular members of these families which correspond to vague prior information. The simple rules of updating from prior to posterior plausibility function are shown in table 2.3 for the standard distributions; the use of this table is explained in §2.4.

2.3 The predictive density function

So far we have used the information available to make a plausibility assessment about the unknown index. But this is not our final objective. We are interested in the outcome \( y \) of the future experiment \( f \) and so we want to use our assessment about the plausibility of indices to induce the plausibilities attaching to the unknown outcome. We are uncertain about which density function \( p(y|\theta) \) applies to \( f \) but we have assessed the plausibility \( p(\theta|x) \) of \( \theta \). The essence of the Bayesian approach to prediction problems is the concept of the probability distribution of a future observation \( y \) given the outcome \( x \) of the informative experiment. Clearly the plausibility of \( y \) given \( p(\theta) \) and \( x \) is expressed by

\[ p(y|x) = \int_\Theta p(y|\theta) p(\theta|x) d\theta, \]  \hspace{1cm} (2.4)

and we term this function on \( Y \) the predictive density function \( p(y|x) \) for \( y \) given \( p(\theta) \) and \( x \). The concept has been growing in popularity recently and is at the heart of much of the subsequent analysis of this book. Because of this importance we provide a formal definition.

**Definition 2.1**

*Predictive density function.* For a future experiment \( f \) with class of density functions \( \{p(y|\theta): \theta \in \Theta\} \) on \( Y \),
<table>
<thead>
<tr>
<th>Notation and name</th>
<th>Random variable and domain</th>
<th>Probability density function</th>
<th>Parameter restrictions</th>
<th>Vague prior information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Binomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bi(n, (\theta))</td>
<td>(x = 0, 1, \ldots, n)</td>
<td>(\binom{n}{x} \theta^x (1-\theta)^{n-x})</td>
<td>(0 \leq \theta \leq 1)</td>
<td></td>
</tr>
<tr>
<td>2 Poisson</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Po((\theta))</td>
<td>(x = 0, 1, 2, \ldots)</td>
<td>(\theta^x \exp(-\theta) / x!)</td>
<td>(\theta &gt; 0)</td>
<td></td>
</tr>
<tr>
<td>3 Gamma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ga((\theta))</td>
<td>(x &gt; 0)</td>
<td>(\theta^x \Gamma(k) \exp(-\theta x) / \Gamma(k))</td>
<td>(\theta &gt; 0, k &gt; 0)</td>
<td></td>
</tr>
<tr>
<td>Ex((\theta)) = Ga(1, (\theta))</td>
<td>(x &gt; 0)</td>
<td>(\theta \exp(-\theta x))</td>
<td>(\theta &gt; 0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.1 (continued)

<table>
<thead>
<tr>
<th>Notation and name</th>
<th>Random variable and domain</th>
<th>Probability density function</th>
<th>Parameter restrictions [Vague prior information]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Multinomial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mu((n, \Theta))</td>
<td>(x = (x_1, \ldots, x_d))</td>
<td>(\binom{n}{x_1} \ldots \binom{n}{x_d} (1 - \sum \theta_i)^{n - \sum x_i})</td>
<td>(0 &lt; \theta_i &lt; 1) (\sum \theta_i &lt; 1)</td>
</tr>
<tr>
<td>Multinomial</td>
<td>(x_i = 0, 1, \ldots, n_i)</td>
<td>(\sum x_i &lt; n)</td>
<td></td>
</tr>
<tr>
<td>Di((g, h))</td>
<td>(\Theta = (\theta_1, \ldots, \theta_d))</td>
<td>(\frac{1}{D(g, h)} \prod \theta_i^{-1} \exp \left(-\frac{1}{2} \Theta (x - \mu)^T \Sigma^{-1} \Theta \right))</td>
<td>(g_1 &gt; 0, h &gt; 0) (g \to 0, h \to 0)</td>
</tr>
<tr>
<td>Dirichlet</td>
<td>(0 &lt; \theta_i &lt; 1,)</td>
<td>(\sum \theta_i &lt; 1)</td>
<td></td>
</tr>
<tr>
<td>5 Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No((\mu, \tau))</td>
<td>(x \in \mathbb{R}^d)</td>
<td>(\frac{1}{(2\pi)^{d/2}} \exp \left(-\frac{1}{2} \tau (x - \mu)^T \tau (x - \mu)\right))</td>
<td>(\mu \in \mathbb{R}^d, \tau &gt; 0)</td>
</tr>
<tr>
<td>Normal</td>
<td>(\mu, \tau)</td>
<td>(\mu \in \mathbb{R}^d, \tau &gt; 0)</td>
<td></td>
</tr>
<tr>
<td>NoCh((b, c, g, h))</td>
<td>(\mu \in \mathbb{R}^d, \tau &gt; 0)</td>
<td>(p(\mu</td>
<td>\tau) = \text{No}(b, c^T)) (p(\tau) = \text{Ch}(g, h)): (\frac{1}{</td>
</tr>
<tr>
<td>Normal chi-squared</td>
<td>(\mu \in \mathbb{R}^d, \tau &gt; 0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NoWd((b, c, g, h))</td>
<td>(\mu, \tau)</td>
<td>(p(\mu</td>
<td>\tau) = \text{No}(b, c^T)) (p(\tau) = \text{Wd}(g, h)): (\frac{1}{</td>
</tr>
<tr>
<td>Normal-Wishart</td>
<td>(\mu \in \mathbb{R}^d, \tau \in \mathbb{R}^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Multinormal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NoG((b, c, I, \gamma))</td>
<td>(x = (x_1, \ldots, x_G))</td>
<td>(\frac{1}{(2\pi)^{d/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right))</td>
<td>(\mu \in \mathbb{R}^d) (\Sigma \in \mathbb{S}^d)</td>
</tr>
<tr>
<td>Normal</td>
<td>(x \in \mathbb{R}^d)</td>
<td>(\frac{1}{(2\pi)^{d/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right))</td>
<td>(\mu \in \mathbb{R}^d) (\Sigma \in \mathbb{S}^d)</td>
</tr>
<tr>
<td>NoWdG((b, c, I, \gamma))</td>
<td>(\mu, \tau)</td>
<td>(p(\mu</td>
<td>\tau) = \text{NoG}(b, c\tau)) (p(\tau) = \text{WdG}(g, h)): (\frac{1}{</td>
</tr>
<tr>
<td>Normal-Wishart</td>
<td>(\mu \in \mathbb{R}^d, \tau \in \mathbb{R}^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notation and name</td>
<td>Random variable and domain</td>
<td>Probability density function</td>
<td>Parameter restrictions</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>7 Exponential (two-parameter)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El(μ, τ)</td>
<td>x</td>
<td>( r \exp{-r(μ - x)} )</td>
<td>( μ \in \mathbb{R}_1, r &gt; 0 )</td>
</tr>
<tr>
<td>Exponential left-sided</td>
<td>( x &lt; μ )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ErGa(b, c, g, h)</td>
<td>((μ, τ))</td>
<td>( p(μτ) \text{ is } Er(b, cτ) ) ( p(τ) \text{ is } Ga(g, h) )</td>
<td>( b \in \mathbb{R}_1, c &gt; 0 ), ( g &gt; 0, h &gt; 0 ) ( [b \to -\infty, c \to 0, g \to 0, h \to 0] )</td>
</tr>
<tr>
<td>Exponential (right-sided)-gamma</td>
<td>( μ \in \mathbb{R}_1, τ &gt; 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Er(μ, τ)</td>
<td>x</td>
<td>( r \exp{-r(μ - x)} )</td>
<td>( μ \in \mathbb{R}_1, τ &gt; 0 )</td>
</tr>
<tr>
<td>Exponential right-sided</td>
<td>( x &gt; μ )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElGa(b, c, g, h)</td>
<td>((μ, τ))</td>
<td>( p(μτ) \text{ is } El(b, cτ) ) ( p(τ) \text{ is } Ga(g, h) )</td>
<td>( b \in \mathbb{R}_1, c &gt; 0 ), ( g &gt; 0, h &gt; 0 ) ( [b \to -\infty, c \to 0, g \to 0, h \to 0] )</td>
</tr>
<tr>
<td>Exponential (left-sided)-gamma</td>
<td>( μ \in \mathbb{R}_1, τ &gt; 0 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
an informative experiment \( e \) with class, of density functions

\[ \{ p(x|\theta) : \theta \in \Theta \} \text{ on } X, \]

and a prior density function \( p(\theta) \) on \( \Theta \), the **predictive density function** \( p(y|x) \) for \( f \) is defined by

\[
p(y|x) = \frac{\int_{\theta} p(y|\theta)p(\theta|x) \, d\theta}{\int_{\theta} p(x|\theta)p(\theta) \, d\theta}
\]  

(2.5)

Fig. 2.1 Basic steps leading to the predictive density function.

Fig. 2.1 shows the basic steps leading to the predictive density function. The mechanism for arriving at node B, where the lines from \( p(\theta) \) and \( p(x|\theta) \) converge, is an operation of Bayes's theorem by (2.2) and (2.3). The subsequent process for obtaining the predictive density function at node A is an accumulation (2.4): the \( p(y|\theta) \) and \( p(\theta|x) \) on the lines leading to A are multiplied and accumulated over the common symbol \( \theta \), that is, either integrated or summed over \( \Theta \). The notation to be used for the more commonly encountered predictive distributions is shown in table 2.2. These distributions are also included in the complete list of notation in appendix I.

---

**Predictive distributions**

\[ p(y|x) = \frac{\int_{\theta} p(y|\theta)p(\theta|x) \, d\theta}{\int_{\theta} p(x|\theta)p(\theta) \, d\theta} \]  

(2.5)
**Table 2.2 Predictive density functions**

<table>
<thead>
<tr>
<th>Notation and name</th>
<th>Random variable and domain</th>
<th>Probability density function</th>
<th>Parameter restrictions</th>
</tr>
</thead>
</table>
| 1. BeBi($n, g, h$)  
Beta-binomial | $y$  
y = 0, 1, ..., $n$ | $\frac{B(g+y, h+n-y)}{B(g, h)}$ | $n$ positive integer,  
g > 0, $h > 0$. |
| 2. NeBi($n, l$)  
Negative-binomial | $y$  
y = 0, 1, 2, ... | $\binom{y+n-1}{n-1} l^y (1-l)^n$ | $n$ positive integer,  
0 < $l$ < 1. |
| 3. InBe($k, g, h$)  
Inverse-beta | $y$  
y > 0 | $y^{g-1} e^{-Wy}$ | $k > 0, g > 0, h > 0$. |
| 4. DiMu($n, g, h$)  
Dirichlet-multinomial | $y_i$ = ($y_{i1}, y_{i2}, ..., y_{id}$)  
y = 0, 1, ..., $n$ | $\frac{D(g+y, h+n-\sum_i y_i)}{D(g, h)}$ | $g = (g_1, ..., g_d)$,  
gi > 0,  
h > 0. |
| 5. St($k, b, c$)  
Student | $m$  
m $\in \mathbb{R}$ | $\frac{1}{B\left(\frac{1}{2}, \frac{1}{2}\right) \left(1 + \frac{(c+b)^2}{2(k+\frac{1}{2})}\right)^{k+1/2}}$ | $b \in \mathbb{R}$,  
c > 0, $k > 0$. |
| Si($k, g, h$)  
Siegel | $v$  
v > 0 | $\frac{v^{g-1}}{B\left(\frac{1}{2}, \frac{1}{2}\right) \left(1 + \frac{h^2}{2(k+\frac{1}{2})}\right)^{k+1/2}}$ | $k > 0$,  
g > 0, $h > 0$. |
| StSi($k; b; c; g; h$)  
Student-Siegel | $(m, v)$  
m $\in \mathbb{R}$, v > 0 | $\frac{v^{g-1}}{D\left(\frac{1}{2}, \frac{1}{2}\right) \left(1 + \frac{(c+b)^2}{2(k+\frac{1}{2})}\right)^{k+1/2}}$ | $b \in \mathbb{R}$,  
c > 0,  
k > 0, g > 0, $h > 0$. |
<table>
<thead>
<tr>
<th>Notation and name</th>
<th>Random variable and domain</th>
<th>Probability density function</th>
<th>Parameter restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. $S_{td}(k, b, c)$ Student</td>
<td>$m \in \mathbb{R}^d$</td>
<td>$\frac{1}{D(| \frac{1}{2} (k-d+1) |_{k c}^{1/2} (1 + (m-b)'(k c)^{-1}(m-b))^{(k+1)v/2})}$</td>
<td>$b \in \mathbb{R}^d$, $c \in \mathbb{S}^d$, $k &gt; d-1$.</td>
</tr>
<tr>
<td>$S_{td}(k, g, h)$ Siegel</td>
<td>$\nu \in \mathbb{S}^d$</td>
<td>$\frac{1}{V(k-d-1)v/2} \frac{1}{B_d(\frac{1}{2} k, \frac{1}{2} g)</td>
<td>h^{d/2} (1 + h^v (k+g)v/2)}$</td>
</tr>
<tr>
<td>$S_{td}(k; b; c; g; h)$ Student-Siegel</td>
<td>$m \in \mathbb{R}^d$, $\nu \in \mathbb{S}^d$</td>
<td>$\frac{1}{\Gamma_d(\frac{1}{2} (k+g+1)) \Gamma_d(\frac{1}{2} g) \pi^{d/2} (k c)^{1/2}</td>
<td>h^{d/2} (1 + (k c)^{-1} (m-b)'(m-b)^2 + h^v (k+g+1) v/2)}$</td>
</tr>
</tbody>
</table>
26 Predictive distributions

Example 2.1

Prediction for the gamma distribution. As an illustration we take the reader through the formal stages of the derivation of the predictive density function for the gamma family. Here \( e \) is described by

\[
p(x|\theta) = \frac{\theta^k x^{k-1} \exp(-\theta x)}{\Gamma(k)} \quad (x > 0).
\]

The conjugate prior distribution for \( \theta \) is also of gamma type, namely

\[
p(\theta) = \frac{h^g \theta^{g-1} \exp(-h \theta)}{\Gamma(g)} \quad (\theta > 0).
\]

For the case of a single \( \text{Ga}(k, \theta) \) observation \( x \) a straightforward application of Bayes's theorem yields

\[
p(\theta|x) = \text{Ga}(G, H) \text{ where } G = g + k, H = h + x.
\]

The extension to the case where \( e \) consists of \( n \) replicates of a \( \text{Ga}(k, \theta) \) experiment with observation \( x = (x_1, x_2, \ldots, x_n) \) is straightforward since \( \sum x_i \) is sufficient for \( \theta \) with probability distribution \( \text{Ga}(nk, \theta) \). Hence

\[
p(\theta|x) = \text{Ga}(G, H) \text{ where } G = g + nk, H = h + \sum x_i.
\]

To cover this possibility we take

\[
p(y|\theta) = \text{Ga}(K, \theta)
\]

where \( K \) is not necessarily equal to \( k \). It follows that the predictive density function is given by

\[
p(y|x) = \int_0^\infty \frac{\theta^K y^{K-1} \exp(-\theta y) H^G \theta^{G-1} \exp(-H \theta)}{\Gamma(K) \Gamma(G)} \, d\theta
\]

\[
= \frac{H^G \theta^{G-1}}{B(K, G)(H + y)^{G+K}} \quad (y > 0),
\]

so that

\[
p(y|x) = \ln\text{Be}(K, G, H).
\]

2.4 Predictive distributions for the standard situations

It is convenient to have available for easy reference a means of deriving such predictive distributions for standard situations. Table 2.3 is set out in a form
corresponding to the derivation scheme of fig. 2.1, with the density functions in the following array:

\[
\begin{array}{ccc}
  p(x|\theta) & p(\theta) & p(\theta|x) \\
  p(y|\theta) & p(\theta|x) \\
  p(y|x) & & \\
\end{array}
\]

For instance, the five steps (2.6) to (2.10) of example 2.1 can be read immediately from the five entries of the array

\[
\begin{array}{ccc}
  Ga(k, \theta) & Ga(g, h) & \\
  Ga(K, \theta) & Ga(G, H) \\
  InBe(K, G, H) & & \\
\end{array}
\]

in section 3 of the table.

Although cases 1–3 of table 2.3 apparently deal with a single observation \(x\) from the informative experiment they are easily adapted to information consisting of a set \(x = (x_1, \ldots, x_n)\) of observations from \(n\) replicates. We have already seen an instance of this adaptation for the gamma case in example 2.1, and for the binomial, Poisson and gamma distributions the form of cases 1, 2 and 3 in table 2.3 can be directly used. For each of these distributions \(\theta\) is one-dimensional and all the information about \(\theta\) provided by the informative experiment is contained in the value of the sufficient statistic \(\Sigma x_i\) which itself has a density function of the given form. For example, if each \(x_i\) is Po(\(\theta\)) then the sufficient statistic \(\Sigma x_i\) is Po(\(n\theta\)). Thus from the table the posterior plausibility function \(p(\theta|x)\) is simply Ga(\(g + \Sigma x_i, h + n\)) and we can proceed to the construction of the predictive density function.

For the normal and multinormal distributions (cases 5 and 6 of table 2.3) the results are presented in a general way to allow for wide applicability. To motivate this form of presentation we consider the following situation.

Suppose that the informative experiment consists of \(n\) replicates of a No(\(\mu, \tau\)) experiment with outcomes \(x_1, \ldots, x_n\). Here \(\theta = (\mu, \tau)\) is two-dimensional and the data can be reduced to a sufficient statistic \((m, v)\), where

\[
m = \bar{x} = \sum x_i/n, \quad v = \sum (x_i - \bar{x})^2.
\]

We can thus envisage the informative experiment as producing simply \((m, v)\); moreover

\[
p(m, v|\theta) = p(m|\mu, \tau)p(v|\tau),
\]

so that, for given \((\mu, \tau)\), \(m\) and \(v\) are independent with distributions of the following forms:
<table>
<thead>
<tr>
<th>Table 2.3 Construction of predictive distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x \mid \theta) )</td>
</tr>
<tr>
<td>( p(x \mid \theta) )</td>
</tr>
</tbody>
</table>

1. **Binomial**
   - \( \text{Bi}(n, \theta) \)
   - \( \text{Bi}(N, \theta) \)
   - \( \text{BeBi}(N, G, H) \)
   - \( \text{Be}(g, h) \)
   - \( \text{Be}(G, H) \)
   - where \( G = g + x, \quad H = h + n - x. \)

2. **Poisson**
   - \( \text{Po}(k \theta) \)
   - \( \text{Po}(K \theta) \)
   - \( \text{NeBi} \left(G, \frac{K}{K + H}\right) \)
   - \( \text{Ga}(g, h) \)
   - \( \text{Ga}(G, H) \)
   - where \( G = g + x, \quad H = h + k. \)

3. **Gamma**
   - \( \text{Ga}(k, \theta) \)
   - \( \text{Ga}(K, \theta) \)
   - \( \text{InBe}(K, G, H) \)
   - \( \text{Ga}(g, h) \)
   - \( \text{Ga}(G, H) \)
   - where \( G = g + k, \quad H = h + x. \)

4. **Multinomial**
   - \( \text{Mu}(n, \theta) \)
   - \( \text{Mu}(N, \theta) \)
   - \( \text{DiMu}(N, G, H) \)
   - \( \text{Di}(g, h) \)
   - \( \text{Di}(G, H) \)
   - where \( G = g + x, \quad H = h + n - 1. \)
\[
\begin{align*}
5 \text{ Normal} & \\
\{ & \text{No}(\mu, \kappa) \\
& \text{Ch}(\nu, \tau) \\
& \{ & \text{No}(\mu, \kappa) \\
& \text{Ch}(\lambda, \tau) \\
\text{St} \left\{ G, B, \left( \frac{1}{K} + \frac{1}{C} \right) \frac{H}{G} \right\} & \\
\text{Si} (G, \lambda, H) & \\
\text{StSi} \left\{ G; B, \left( \frac{1}{K} + \frac{1}{C} \right) \frac{H}{G}; \lambda, H \right\} & \\
\end{align*}
\]

6 Multinormal
\[
\begin{align*}
\{ & \text{No}_d(\mu, k \tau) \\
& \text{Wid}(\nu, \tau) \\
& \{ & \text{No}_d(\mu, k \tau) \\
& \text{Wid}(\lambda, \tau) \\
\text{Std} \left\{ G, B, \left( \frac{1}{K} + \frac{1}{C} \right) \frac{H}{G} \right\} & \\
\text{Si}(G, \lambda, H) & \\
\text{StSi}(G, B, \left( \frac{1}{K} + \frac{1}{C} \right) \frac{H}{G}; \lambda, H) & \\
\end{align*}
\]

NoCh(b, c, g, h)

NoCh(B, C, G, H) where

\[
\begin{align*}
B &= C^{-1}(cb + km), \\
C &= c + k, \\
G &= g + v + \Delta(c), \\
H &= h + v + \frac{ck}{c + k} (m - b)^{t}, \\
\Delta(c) &= \begin{cases} \\
0 & (c = 0), \\
1 & (c > 0). \\
\end{cases}
\end{align*}
\]

NoWid(b, c, g, h)

NoWid(B, C, G, H) where

\[
\begin{align*}
B &= C^{-1}(cb + km), \\
C &= c + k, \\
G &= g + v + \Delta(c), \\
H &= h + v + \frac{kc}{k + c} (m - b)(m - b)^{t}. \\
\end{align*}
\]
Table 2.3 (continued)

<table>
<thead>
<tr>
<th>$p(x \mid \theta)$</th>
<th>$p(\theta)$</th>
<th>$p(y \mid x)$</th>
</tr>
</thead>
</table>

7 Exponential (two-parameter)

\[ \begin{align*}
&\text{EiGa}(b, c, g, h) \\
&\text{EiGa}(B, C, G, H) \quad \text{where } B = \min(m, b), \quad C = c + k, \quad G = g + v + \Delta(c), \quad H = h + v + \omega_m(b, c, k)(m - b).
\end{align*} \]

\[
\begin{align*}
\ln Be(\lambda, G, H) &= \frac{CK}{B(1, G)(C + K)H} \left\{ 1 + H^{-1} \omega_M(B, C, K)(M - B) \right\}^{-\lambda - 1} \\
\ln Be(\lambda, G, H) &= \frac{CKV^{-1}}{D(1, G, \lambda)(C + K)H^{\lambda + 1}} \left\{ 1 + H^{-1} \omega_M(B, C, K)(M - B) + H^{-1} V \right\}^{-\lambda - 1}
\end{align*}
\]
Predictive distributions

\[ p(m|\mu, \tau) = \text{No}(\mu, k\tau), \]
\[ p(v|\tau) = \text{Ch}(v, \tau), \]

where \( k = n, \nu = n - 1 \). If the future experiment is to consist of \( N \) replicates of the \( \text{No}({\mu}, \tau) \) experiment we may be interested in such summary statistics as

\[ M = \bar{y} = \frac{1}{N} \sum y_i, V = \sum (y_i - \bar{y})^2, \]

which again are independent for given \((\mu, \tau)\) and have distributions of the forms:

\[ p(M|\mu, \tau) = \text{No}(\mu, K\tau), \]
\[ p(V|\tau) = \text{Ch}(\lambda, \tau), \]

with \( K = N, \lambda = N - 1 \). It is thus convenient to set out the predictive analysis in the following array:

\[
\begin{pmatrix}
    p(x|\theta) & p(m|\mu, \tau) & p(\theta) \\
    p(v|\tau) & p(v|\theta) \\
    p(y|\theta) & p(M|\mu, \tau) & p(\theta|x) \\
    p(\theta|x) \\
    p(y|x) & p(M|x) & p(V|x) \\
    p(V|x) \\
    p(M, V|x)
\end{pmatrix}
\]

With such generality the results apply to even more complicated situations, as in the regression problem in §2.6. The result applies equally to simpler situations than that of the motivating example. For instance, if the informative experiment and the future experiment are each a single replicate of a \( \text{No}({\mu}, \tau) \) experiment then we simply apply case 5 of table 2.3 with \( k = 1, \nu = 0, v = 0, K = 1 \), to obtain \( p(M|x) \) as the predictive density function of the outcome \( y (= M) \) of the future experiment.

For the multinormal case \( m \) and \( M \) are, of course, vectors and \( v \) and \( V \) are matrices but the concepts are completely analogous to those of the univariate normal case.

The idea underlying the presentation of the two-parameter exponential distribution (case 7 of table 2.3) is similar. For example, if the informative experiment \( e \) consists of \( n \) replicates of a \( \text{Er}({\mu}, \tau) \) experiment and we suppose that only the first \( r \) order statistics \( x_{(1)}, ..., x_{(r)} \) \((2 \leq r \leq n)\) are available then

\[ m = x_{(r)}, \]
32 Predictive distributions

\[ v = x_{(1)} + \ldots + x_{(r)} + (n - r)x_{(r)} - nx_{(1)} \]

are sufficient for \((\mu, \tau)\), and are independent with

\[
\begin{align*}
p(m | \mu, \tau) &= \text{Er}(\mu, m\tau), \\
p(v | \tau) &= \text{Ga}(\nu, \tau),
\end{align*}
\]

where \(\nu = r - 1\). Table 2.3 then provides the appropriate apparatus for the construction of the predictive density function for a \(\text{Er}(\mu, K\tau)\) future experiment.

2.5 Prediction for regression models

A wide range of prediction problems occurs in regression situations; we have already seen a problem of this type in example 1.5. In such problems the future experiment \(f\) records the response (value of the dependent variable) made by some experimental unit to a known stimulus \(t\) (value of the explanatory or regressor variable), and so is better denoted by \(f_t\) than by \(f\). We are thus led to consider a class

\[
F = \{f_t : t \in T\}
\]

of experiments, where each \(f_t\) has the same sample space \(Y\) and \(T\) is the set of possible stimuli. The dependence on \(t\) of the density function of \(f_t\) is also made explicit by the notation \(p(y | t, \theta)\). Note that in specifying the density function for \(f_t\) we have used a common \(\theta\)-index for every \(t\) rather than a possibly different index, say \(\theta_t\), for each \(t\). We can achieve this by setting \(\theta = \{\theta_t : t \in T\}\), for example, and we then gain some simplification in later developments, particularly in chapters 10 and 11.

The informative experiment \(e\) consists of a set of \(n\) independent experiments of given types \(f_{t_1}, \ldots, f_{t_n}\) and provides a set of observations \((t_1, x_1), \ldots, (t_n, x_n)\), which can be portrayed as the usual scatter diagram in the \((t, x)\) plane. To simplify notation we denote this complete set of observations by \(z\), and write \(t = (t_1, \ldots, t_n), x = (x_1, \ldots, x_n)\) when we wish to isolate the \(t\) and \(x\) components of \(z\).

At the centre of our analysis again will be the predictive density function for \(f_t\), now denoted by \(p(y | t, z)\). Corresponding to (2.5) we have

\[
p(y | t, z) = \int_\Theta p(y | t, \theta)p(\theta | z) d\theta,
\]

where

\[
p(\theta | z) = \frac{p(\theta) \prod_{i=1}^{n} p(x_i | t_i, \theta)}{\int_\Theta p(\theta) \prod_{i=1}^{n} p(x_i | t_i, \theta) d\theta}.
\]
In accordance with our definition above of $\theta$ as a common index for all $t$ we write $p(\theta | z)$ rather than $p(\theta | t, x)$ in (2.12). Despite the apparent complexity of (2.11) and (2.12) the derivation of the predictive regression density function can be easily obtained for standard situations by reference to the tables already presented in this chapter. We illustrate the technique for the binomial, Poisson, gamma and normal regression models.

(i) Binomial. Here $f_t$ is Bi($t, \theta$) and $x$, the total number of successes in $f_t$, $f_{t_1}, \ldots, f_{t_n}$, is sufficient for $\theta$ and is a Bi($\Sigma f_t$, $\theta$) random variable. Hence given $\Sigma f_t$ we can use case 1 of table 2.3 with $n = \Sigma f_t$ and $N = t$ to derive $p(y | t, z)$.

(ii) Poisson. Here $f_t$ is Po($t\theta$) and, for given $\Sigma f_t$, $x = \Sigma x_t$ is sufficient for $\theta$ with a Po($\Sigma f_t \theta$) distribution.

(iii) Gamma. Similar remarks apply if $f_t$ is Ga($t, \theta$). Here, for given $\Sigma f_t$, $x = \Sigma x_t$ is sufficient for $\theta$ with distribution Ga($\Sigma f_t$, $\theta$).

(iv) Normal. The case of normal regression can also be analysed by means of table 2.3. Here $f_t$ is described by a No($\alpha + \beta t$, $\tau$) density function. Let

$$\bar{t} = \frac{\sum f_t}{n}, \bar{x} = \frac{\sum x_t}{n},$$

$$S(t, t) = \sum (t_i - \bar{t})^2, S(t, x) = \sum (t_i - \bar{t})(x_i - \bar{x}),$$

$$S(x, x) = \sum (x_i - \bar{x})^2.$$ 

Then the estimated regression coefficient is

$$\hat{\beta} = S(t, x)/S(t, t)$$

and the residual sum of squares is

$$v = \sum (x_i - \bar{x} - \hat{\beta}(t_i - \bar{t}))^2 = S(x, x) - \frac{(S(t, x))^2}{S(t, t)}.$$

Then, for given $S(t, t)$,

$$\bar{x} + \hat{\beta}(t - \bar{t})$$

are jointly sufficient for

$$\mu = \alpha + \beta t$$

and $\tau$.

Also $\bar{x} + \hat{\beta}(t - \bar{t})$ and $v$ are independently distributed as

$$\text{No} \left[ \mu, \tau \left/ \left( \frac{1}{n} + \frac{(t - \bar{t})^2}{S(t, t)} \right) \right. \right]$$

and $\text{Ch}(n - 2, \tau)$. 
This is thus a special case of the normal array in case 5 of table 2.3 if we take
\[
\frac{1}{k} = \frac{1}{n + \frac{(t - \bar{t})^2}{S(t, t)}} \quad \text{and} \quad \nu = n - 2.
\]

**Example 2.2**

*A bird-nesting problem.* Four pairs of a rare species of bird nested for the first time in Scotland last season. The observed number of eggs in the four nests were 2, 3, 3, 4 and from these nests 1, 2, 3, 3 nestlings survived the season. At the start of the current season a new pair has a nest with 3 eggs. What are the chances that at least 2 nestlings will survive the season?

This is a case of binomial regression with \( \theta \) representing the probability that an egg from a nest will give rise to a surviving nestling, with

\[
t_i = 2 \quad 3 \quad 3 \quad 4
\]
\[
x_i = 1 \quad 2 \quad 3 \quad 3
\]

and \( t = 3 \). Since the rare species was nesting under completely new conditions last season we use the vague prior distribution of table 2.1 with

\[
p(\theta) = \text{Be}(g, h) \quad (g \to 0, \ h \to 0).
\]

Hence, from (i) above, the predictive density function for \( y \), the number of surviving nestlings in the new nest with \( t = 3 \) eggs, is given by table 2.3 as

\[
p(y | t, z) = \text{BeBi}(t, \Sigma x_i, \Sigma t_i - \Sigma x_i)
\]
\[
= \text{BeBi}(3, 9, 3).
\]

We show the numerical values of this distribution in table 2.4, and so assess the chance that at least 2 nestlings will survive the season to be 0.371 + 0.453 = 0.824.

<table>
<thead>
<tr>
<th>( y )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(y</td>
<td>t, z) )</td>
<td>0.027</td>
<td>0.148</td>
<td>0.371</td>
</tr>
<tr>
<td>( p(y</td>
<td>t, \theta(z)) )</td>
<td>0.016</td>
<td>0.141</td>
<td>0.422</td>
</tr>
</tbody>
</table>

A technique of prediction commonly practiced is to regard the problem as primarily one of estimating \( \theta \), in this case by

\[
\hat{\theta}(z) = \frac{\sum x_i}{\sum t_i} = 0.75
\]

say; and then to use the density function \( p(y | t, \theta) \) with \( \theta \) equal to \( \hat{\theta}(z) \). This density function is shown in the last row of table 2.4 and is clearly quite different from \( p(y | t, z) \). The difference arises from the fact that the last row
uses $\hat{\theta}(z)$ as if it were the true index, whereas the construction of the predictive density function in (2.11) adopts the more reasonable approach of averaging the density functions $p(y \mid t, \theta)$, the weighting factors being

$$p(\theta \mid z) = \text{Be}(9, 3)$$

with mean value 0.75.

### 2.6 An application to medical prognosis

We conclude this chapter with a practical application which highlights the idea of predictive distributions. A recurring prediction problem in medical practice is that of prognosis. There the clinician in deciding which of the alternative available treatments is best for a particular patient must attempt to make prognoses of the future possible paths of the patient's illness on the different treatments. In any quantitative approach to such problems a first step is to provide for each treatment an appropriate predictive distribution for the future experiment which records the patient's progress. In the simple prognostic problem of example 1.1 a suitable measure of a patient's progress or effectiveness of treatment is taken to be survival time.

**Table 2.5 Aces and survival times of 20 carcinoma patients**

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Survival time (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>25</td>
</tr>
<tr>
<td>54</td>
<td>45</td>
</tr>
<tr>
<td>37</td>
<td>238</td>
</tr>
<tr>
<td>47</td>
<td>194</td>
</tr>
<tr>
<td>51</td>
<td>16</td>
</tr>
<tr>
<td>48</td>
<td>23</td>
</tr>
<tr>
<td>42</td>
<td>30</td>
</tr>
<tr>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>45</td>
<td>22</td>
</tr>
<tr>
<td>33</td>
<td>123</td>
</tr>
<tr>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td>34</td>
<td>41</td>
</tr>
<tr>
<td>66</td>
<td>45</td>
</tr>
<tr>
<td>44</td>
<td>19</td>
</tr>
<tr>
<td>64</td>
<td>23</td>
</tr>
<tr>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td>56</td>
<td>40</td>
</tr>
<tr>
<td>43</td>
<td>35</td>
</tr>
<tr>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>40</td>
<td>91</td>
</tr>
</tbody>
</table>

We assume that the 20 survival times of table 2.5 form a random sample from a lognormal distribution (Aitchison and Brown, 1957). By considering log (survival time) we can convert the data to a random sample $x_1, \ldots, x_n$ ($n = 20$) from a $\text{No}(\mu, \tau)$ distribution. We then reduce the sample to the sufficient statistics

$$m = \sum x_i/n, \quad v = \sum (x_i - m)^2,$$

which are independently distributed as

$\text{No}(\mu, \tau), \text{Ch}(n - 1, \tau)$.

If we adopt a vague prior distribution for $(\mu, \tau)$ as in case 5 of table 2.1 with

$$p(\mu, \tau) = \text{NoCh}(b, c, g, h) \quad (c \rightarrow 0, g \rightarrow 0, h \rightarrow 0)$$

We assume that the 20 survival times of table 2.5 form a random sample from a lognormal distribution (Aitchison and Brown, 1957). By considering log (survival time) we can convert the data to a random sample $x_1, \ldots, x_n$ ($n = 20$) from a $\text{No}(\mu, \tau)$ distribution. We then reduce the sample to the sufficient statistics

$$m = \sum x_i/n, \quad v = \sum (x_i - m)^2,$$

which are independently distributed as

$\text{No}(\mu, \tau), \text{Ch}(n - 1, \tau)$.

If we adopt a vague prior distribution for $(\mu, \tau)$ as in case 5 of table 2.1 with

$$p(\mu, \tau) = \text{NoCh}(b, c, g, h) \quad (c \rightarrow 0, g \rightarrow 0, h \rightarrow 0)$$

We assume that the 20 survival times of table 2.5 form a random sample from a lognormal distribution (Aitchison and Brown, 1957). By considering log (survival time) we can convert the data to a random sample $x_1, \ldots, x_n$ ($n = 20$) from a $\text{No}(\mu, \tau)$ distribution. We then reduce the sample to the sufficient statistics

$$m = \sum x_i/n, \quad v = \sum (x_i - m)^2,$$

which are independently distributed as

$\text{No}(\mu, \tau), \text{Ch}(n - 1, \tau)$.

If we adopt a vague prior distribution for $(\mu, \tau)$ as in case 5 of table 2.1 with

$$p(\mu, \tau) = \text{NoCh}(b, c, g, h) \quad (c \rightarrow 0, g \rightarrow 0, h \rightarrow 0)$$
we have case 5 of table 2.3 with

\[ k = n = 20, \ m = 3.83, \ \nu = n - 1 = 19, \ \nu = 23.22, \ K = 1. \]

We are led through the posterior distribution

\[ p(\mu, \tau | \mathbf{x}) = \text{NoCh}(m, n, \nu, v) \]

to the predictive density function
Predictive distributions

\[ p(y|x) = St(19, 3.83, 1.28) \]

for \( y \), the logarithm of survival time of a new patient on treatment of pre-operative radiotherapy followed by radical surgery.

The probability that a new patient will survive 150 weeks is then easily assessed by (A26) of appendix I as

\[ \int_{5.01}^{\infty} p(y|x)dy = 0.5 f_{0.346}(9.5, 0.5) \]

\[ = 0.16. \]

Example 2.3

Age-specific medical prognosis. Table 2.5 shows the ages of the 20 carcinoma patients just considered. We now pose the question: to what extent is the reliability of prognosis improved by knowledge of this factor?

Fig. 2.2 shows a scatter diagram of age \( t \) and logarithm \( (x) \) of survival time and we recognise the possibility of explaining part of the variability of survival time in terms of a linear regression of \( x \) on \( t \). Following the standard normal regression calculations of §2.5 we compute the basic quantities

\[ n = 20, \bar{t} = 46.6, \bar{x} = 3.83, \]

\[ S(t, t) = 1460.8, S(t, x) = -104.25, S(x, x) = 23.22, \]

from which we obtain

\[ \nu = 18, \hat{\beta} = -0.0714, \nu = 15.78. \]

A standard t-test shows that the regression coefficient is significantly different from zero at the 1 per cent level.

The predictive density function for \( y \), the logarithm of the survival time of a new patient aged \( t \) and placed on the specified treatment, is

\[ p(y|t, z) = St(18, 7.16 - 0.0714t, 0.9205 + 0.000600(t - 46.6)^2). \]

The fact that the third parameter of this age-adjusted distribution is, for all \( t \) in the range of the informative experiment, less than the corresponding value 1.28 of the unadjusted predictive distribution is an indication of the greater precision of the age-adjusted analysis.

The age-specific nature of the survival rate can be presented in terms of probability curves for different survival times. Fig. 2.3 shows these curves for 100, 150 and 200 weeks and illustrates clearly how survival probability decreases with increasing age.
Fig. 2.3 Age-specific probability curves for different survival times.
Predictive distributions

History

The concept of the predictive distribution, at least in particular simple situations, has a long tradition dating back at least to Laplace. His 'rule of succession' was an attempt to answer the question: if in \( n \) binomial trials success has occurred \( n \) times what is the probability of success in the \((n + 1)\)th trial? This question and its generalisation 'If we obtain \( x \) successes in \( n \) binomial trials what is the probability of \( y \) successes in the next \( N \) trials?' have often been the subject of discussion by philosophers and statisticians; for some comments on the history of this particular problem see the contribution by E.S. Pearson to the discussion of Aitchison (1964) and Thatcher (1964).

Fisher (1935) provided a fiducial type derivation for a predictive distribution. The widening of the concept and in particular its more extended use as an effective tool of data analysis are however fairly recent. Jeffreys (1961, p. 143) obtains the predictive distributions for the mean and sample variance of a second sample from a normal distribution based on the observations from a first sample and a vague prior distribution on the parameter. Geisser (1964) applies this to a diagnostic problem and Guttman and Tiao (1964) use predictive distributions for the normal and two-parameter exponential families. Zellner and Chetty (1965) derive the predictive distributions for the multivariate regression model. All these papers use only vague prior distributions. Aitchison and Sculthorpe (1965) analyse cases 1, 2, 3 and 5 of table 2.3 with general prior distributions. Raiffa and Schlaifer (1961, chapters 9–13) contains most of the relevant distribution theory of table 2.3 for the standard families. The idea of predictive distributions is not discussed by them but the derivation of

\[
p(\gamma) = \int_\theta p(\gamma|\theta) p(\theta) d\theta
\]

is given.


Problems

2.1 For the class of Pareto density functions

\[
p(x|\theta) = Pa(k, \theta) = \frac{\theta k^\theta}{x^{\theta+1}} \quad (x > k)
\]

find a suitable conjugate class of prior density functions on the parameter set \( \Theta = \{\theta : \theta > 0\} \).
40 Predictive distributions

Suppose that \( x = (x_1, \ldots, x_n) \) is a set of \( n \) independent observations from the above distribution. Obtain the predictive density function \( p(y|x) \) based on this information and on a typical member of the conjugate class of prior density functions. Show that the probability that a future outcome exceeds \( ke^d \), where \( d > 0 \), can be expressed in the form

\[
\left[ \frac{H}{H+d} \right]^G.
\]

2.2 An informative experiment consists of \( n \) replicates yielding random variables \( x_1, \ldots, x_n \), each distributed uniformly over the interval \((-\theta, \theta)\), where \( \theta > 0 \). The prior density function \( p(\theta) \) is \( P(\theta) \). Show that the predictive distribution for a future observation \( y \) which is also, for given \( \theta \), uniformly distributed in \((-\theta, \theta)\) has density function

\[
\frac{H}{2G(H+1)} \begin{cases} |y| \leq G, \\ \frac{HG^H}{2(H+1)|y|^{H+1}} & (|y| > G), \end{cases}
\]

where \( G = \max \{g, |x_1|, \ldots, |x_n|\}, \quad H = h + n. \)

2.3 Show that if

\[
p(x|\theta) = \frac{1}{\pi \{1 + (x-\theta)^2\}} \quad (x \in \mathbb{R}^1),
\]

\[
p(y|\theta) = \frac{1}{\pi \{1 + (y-\theta)^2\}} \quad (y \in \mathbb{R}^1),
\]

and if the prior distribution is diffusely uniform over \( \Theta = \mathbb{R}^1 \), then

\[
p(y|x) = \frac{1}{2\pi \{1 + \frac{1}{4}(y-x)^2\}} \quad (y \in \mathbb{R}^1).
\]

2.4 Show that, for an informative experiment with

\[
p(x|\theta) = \frac{1}{2} \exp(-|x-\theta|) \quad (x \in \mathbb{R}^1),
\]

for a future experiment with

\[
p(y|\theta) = \frac{1}{2} \exp(-|y-\theta|) \quad (y \in \mathbb{R}^1),
\]

and for a prior which is uniformly diffuse over the real line parameter space, the predictive density function is given by
Predictive distributions

\[ p(y|x) = \frac{1}{2} (1 + |y - x|) \exp(-|y - x|) \quad (y \in \mathbb{R}). \]

2.5 A completely new extrusion process for the manufacture of artificial fibre is under investigation. It is assumed that the distribution of flaws along the length of the fibre follows a Poisson process. The numbers of flaws in five fibres of lengths 10, 15, 25, 30 and 40 metres were found to be 3, 2, 7, 6, 10 respectively. What is the probability that a fibre of length 60 metres will contain at most 14 flaws?

2.6 A machine uses \( N \) components and fails as soon as one or other of its components fails. The lifetimes of components manufactured during a long production run are known to be independently distributed as \( \text{Ex}(\theta) \), and \( \theta \) is assumed to vary from run to run according to a \( \text{Ga}(g, h) \) distribution. From a recent production run life-testing of \( n \) components gave lifetimes \( x_1, \ldots, x_n \). On the basis of all this information show that an appropriate distribution for the lifetime \( y \) of a machine constructed of components from this run has density function

\[ \frac{NGH^G}{(H+Ny)^{G+1}} \quad (y > 0), \]

where

\[ G = g + n, \quad H = h + \sum_{i=1}^{n} x_i. \]

2.7 Suppose that independent identically distributed \( \text{Ex}(\theta) \) random variables \( x_1, \ldots, x_n \) constitute the informative experiment and that the prior distribution for \( \theta \) is \( \text{Ga}(g, h) \). Show that the predictive distribution for a future experiment described by independent identically distributed \( \text{Ex}(\theta) \) random variables \( y_1, \ldots, y_N \) has density function

\[ p(y|x) = \frac{\Gamma(G+N)}{\Gamma(G)} \frac{H^G}{(H+y_1+\ldots+y_N)^{G+N}} \quad (y_1 > 0, \ldots, y_N > 0), \]

where

\[ G = g + n, \quad H = h + \sum_{i=1}^{n} x_i. \]

Deduce the distribution of \( z = \min(y_1, \ldots, y_N) \), and confirm that it is the same as the machine lifetime distribution of problem 2.6.

2.8 Consider a multinomial experiment with \( N + 1 \) categories, of which \( N \) are 'success' categories with probabilities \( \alpha_1, \ldots, \alpha_N \), the remaining failure category
42 Predictive distributions

Having probability $\alpha_{N+1} = 1 - \alpha_1 - \ldots - \alpha_N$. Suppose that we record the numbers $y_1, \ldots, y_N$ of the different types of success that occur before the $k$th failure is recorded. Then $y_1, \ldots, y_N$ is said to have a negative multinomial distribution, written $\text{NeMu}(k; \alpha_1, \ldots, \alpha_N, \alpha_{N+1})$.

For a Poisson informative experiment with $x$ distributed as $\text{Po}(\theta)$, for a $\text{Ga}(g, h)$ prior distribution for $\theta$, and for a future experiment yielding counts $y_1, \ldots, y_N$ which, for given $\theta$, are independent $\text{Po}(t_1 \theta), \ldots, \text{Po}(t_N \theta)$, show that the predictive joint distribution of $y_1, \ldots, y_N$ is

$$\text{NeMu} \left( g + x; \frac{t_1}{h + t + \sum t_i}, \ldots, \frac{t_N}{h + t + \sum t_i}, \frac{h + t}{h + t + \sum t_i} \right).$$

Deduce the predictive distribution for the total count $y_1 + \ldots + y_N$.

2.9 Each run of a process produces a large batch of ball bearings whose diameters (mm) are $\text{No}(\mu, \tau)$ distributed. The process is not sufficiently under control to achieve the target values $\mu = 8, \tau = 100$ on each run, and study of a large number of previous batches suggests that the variability of $(\mu, \tau)$ from batch to batch follows a $\text{NoCh}(8, 2, 1000, 10)$ distribution.

From a particular batch a sample of 10 ball bearings is chosen at random and their diameters (mm) are found to be

8.07, 8.15, 7.97, 7.85, 8.02, 8.17, 8.11, 8.09.

A further 15 bearings are selected at random from the batch. What are the probabilities

(i) that the mean diameter of the 15 bearings is less than 8.10;
(ii) that the standard deviation of the 15 diameters is less than 0.1?

2.10 The lifetimes (hours) of certain components are known to be independently distributed as $\text{Ex}(\tau)$. Ten of these components were simultaneously put into continuous use at an unrecorded time $\mu$ after their joint purchase and so far five of them have failed at times 2364, 2532, 2575, 2900, 3412 hours after purchase. An eleventh component from the same purchase is about to be put into continuous use. On the basis that your prior information about $\mu$ and $\tau$ is extremely imprecise (except, of course, that you know that $\mu > 0$) how do you assess the chances that this eleventh component will still be functioning after 2000 hours?

2.11 Suppose that the machine of problem 2.6 continues functioning until the $K$th failure among its $N$ components. Also suppose that the life testing of the $n$ components from the recent production run is continued only until
the $k$th component fails, and that these successive lifetimes are $x_1, ..., x_k$.
Show that, given all this information, the probability that a machine survives
at least time $t$ is assessed as
\[
N \binom{N-1}{K-1} \sum_{i=0}^{K-1} (-1)^i \binom{K-1}{i} \frac{1}{(N-K+i+1)} \times \left( \frac{h+\nu}{h+\nu + t(N-K+i+1)} \right)^{x+k},
\]
where $\nu = x_1 + ... + x_k + (n-k)x_h$.

2.12 To study the effectiveness of a fertiliser additive on the productivity of
a certain variety of tomato plant, eight plants were grown in compost into
which different strengths of the additive had been mixed. The resulting yields
were as follows:

<table>
<thead>
<tr>
<th>Strength of additive</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield (kg)</td>
<td>1.28</td>
<td>1.84</td>
<td>1.37</td>
<td>1.50</td>
<td>2.20</td>
<td>1.65</td>
<td>3.22</td>
<td>3.00</td>
</tr>
</tbody>
</table>

On the basis of vague prior information concerning all the parameters involved
how would you assess the probabilities
(i) that a single tomato plant grown in compost with additive at strength 4.5
will yield at least 2.5 kg;
(ii) that a single tomato plant grown in compost with additive at strength 8
will yield between 2.5 and 3.5 kg;
(iii) that eight tomato plants all grown in compost with additive at strength
5.5 will yield altogether more than 20 kg?

2.13 Let $x_1, ..., x_n$, associated with an informative experiment, be independent
identically distributed random variables, each uniformly distributed over the
unit interval $(0, \theta + 1)$, and let $y$, associated with a future experiment, be
also uniformly distributed over $(\theta, \theta + 1)$. Suppose that the index set $\Theta$
is the interval $(b, c)$ and that the prior distribution is the uniform distribution over
$(b, c)$. Write
\[
m = \min \{x_1, ..., x_n\}, \quad M = \max \{x_1, ..., x_n\}.
\]
Show that, if the intervals $(b, c)$ and $(M-1, m)$ intersect and if
\[
B = \max \{b, M-1\}, \quad C = \min \{c, m\},
\]
then
\[
(C-B)p(y|x) = \begin{cases} 
\frac{y-B}{B} & (B \leq y \leq C), \\
\frac{C-B}{B} & (C \leq y \leq B + 1), \\
\frac{C+1-y}{B+1} & (B + 1 \leq y \leq C + 1).
\end{cases}
\]
What conclusions would you reach if the intervals \((b, c)\) and \((M - 1, m)\) were disjoint?

### 2.14

The two-parameter Pareto distribution, written \(\text{Pa}(\mu, \tau)\), has density function

\[
\tau \mu^\tau x^{\tau - 1} \quad (x \geq \mu).
\]

Suppose that \(x_1, \ldots, x_n\) are independent \(\text{Pa}(\mu, \tau)\) random variables. Show that

\[
m = \min(x_1, \ldots, x_n)
\]

and

\[
v = \log \frac{x_1 \ldots x_n}{m^n}
\]

are sufficient statistics for \((\mu, \tau)\) and are independently distributed as \(\text{Pa}(\mu, n\tau)\) and \(\text{Ga}(n - 1, \tau)\).

For prior distribution of Pareto-Gamma type, written \(\text{PaGa}(b, c, g, h)\) and with density function

\[
p(\mu, \tau) = \frac{c r^{c - 1}}{b^c} \frac{h^{g + 1} \tau^g e^{-h\tau}}{\Gamma(g)} \quad (0 < \mu < b, \tau > 0),
\]

show that the posterior distribution is \(\text{PaGa}(B, C, G, H)\), where

\[
B = \min(b, m),
\]

\[
C = c + n,
\]

\[
G = g + n - 1 + \Delta(c),
\]

\[
H = h + v + \omega_m(b, c, n) \log(m/b).
\]

Find the predictive distribution for a future observation \(y\) which, for given \((\mu, \tau)\), is distributed as \(\text{Pa}(\mu, \tau)\).
3

Decisive prediction

3.1 Point prediction

If we are asked to predict the outcome of a performance of a future experiment \( f \) our answer will clearly depend on how we view the consequences of being wrong. More specifically we may attempt to assess the relative consequences of being 'close' to the realised outcome and of being 'badly' wrong. If we can quantify these visualised consequences then we can present the problem as one of statistical decision theory. Since in constructing the predictive density function \( p(y|x) \) we have already carried out the information-extraction part of the problem we have a particularly simple confrontation in this decision problem. The components are as follows.

(i) Parameter set. The unknown outcome of the future experiment \( f \) plays the role of an unknown state of nature, so that \( Y \), the sample space of \( f \), is the parameter set of the statistical decision problem. Our assessment of the plausibility of a particular \( y \) at the time of making a decision is \( p(y|x) \), the predictive density at \( y \).

(ii) Action set. The set \( A \) of possible actions is simply a reproduction of \( Y \), since any element of \( Y \) is a possible prediction \( a \).

(iii) Utility function. Associated with each prediction or action \( a \) and each realisable outcome \( y \) there is a utility or value \( U(a, y) \). We thus suppose defined a function \( U \) on the product domain \( A \times Y \).

Standard statistical decision theory then directs us to choose as optimum \( a^* \) a prediction which maximises the expected utility

\[
U(a) = \int_Y U(a, y)p(y|x)dy.
\]  

(3.1)

Thus

\[
U(a^*) = \max_A U(a).
\]  

(3.2)

The optimum prediction depends on the particular \( x \) observed, and as we allow \( x \) to run through \( X \) we generate a function with domain \( X \) and range space \( A (= Y) \). Such a function, instructing as to what simple prediction or action \( a \) to adopt when faced with any outcome \( x \) of the informative experiment, may be termed a simple predictor.
Definition 3.1

Simple predictor. A simple predictor $\delta$ is a function

$$\delta : X \rightarrow A = Y.$$  \hfill (3.3)

What we have determined by the construction above is an optimum simple predictor $\delta^\ast$ defined by the optimum property

$$\max U\{\delta^\ast(x)\} = \max_A U(a).$$  \hfill (3.4)

Note that all we are doing here is to construct the Bayesian decision procedure relative to $p(y \mid x)$.

We now examine some simple results which depend only on the form of the utility function and not on the specific form of predictive density function. We confine attention to the one-dimensional versions of these results where $Y$ is the real line or a subset of the real line.

All-or-nothing point prediction. How should a predictor be constructed if it is desperately important to predict the true outcome? For such a problem the natural formulation is to consider the limiting case $(e \rightarrow 0)$ of the following utility specification:

$$U(a, y) = \begin{cases} 1 & (y - e < a < y + e), \\ 0 & \text{otherwise.} \end{cases}$$  \hfill (3.5)

Then, by (3.1),

$$U(a) = \int_{a-e}^{a+e} p(y \mid x) dy = 2e p(a \mid x) + o(e) \quad (e \rightarrow 0),$$

given simple regularity conditions on $p$. To maximise $U(a)$ we must maximise $p(a \mid x)$, and so we must adopt the intuitively reasonable procedure of predicting the most plausible outcome on the basis of the information we have. Here $\delta^\ast(x)$ is determined by

$$p\{\delta^\ast(x) \mid x\} = \max_A p(a \mid x).$$  \hfill (3.6)

In words: for an all-or-nothing utility structure the optimum simple prediction is the mode of the predictive distribution.

Linear loss point prediction. If the loss is zero when we predict correctly but is otherwise proportional to the distance of the prediction from the actual outcome then
Decisive prediction

\[ U(a, y) = \begin{cases} 
-\xi(a - y) & (y < a), \\
-\eta(y - a) & (y \geq a), 
\end{cases} \tag{3.7} \]

where \( \xi > 0, \eta > 0 \). Then

\[ U(a) = -\xi \int_{-\infty}^{a} (a - y)p(y|x)dy - \eta \int_{a}^{\infty} (y - a)p(y|x)dy \]

and the first and second derivatives of \( U(a) \) with respect to \( a \) are

\[ U'(a) = -\xi \int_{-\infty}^{a} p(y|x)dy + \eta \int_{a}^{\infty} p(y|x)dy, \]

\[ U''(a) = - (\xi + \eta)p(a|x) < 0. \]

Hence \( U(a) \) has a maximum where \( U'(a) = 0 \), so that the maximising \( a^* \) is given by

\[ \int_{-\infty}^{a^*} p(y|x)dy = \frac{\eta}{\xi + \eta}. \tag{3.8} \]

Thus for a linear loss utility structure the optimum simple prediction is the \( \eta/(\xi + \eta) \)-quantile of the predictive distribution. Only the relative value \( \eta/\xi \) is of importance and so, for example, we could assume that \( \xi = 1 \) without loss of generality. However it is convenient to retain both \( \xi \) and \( \eta \) in our formulation.

Note that if \( Y \) is one of the standard discrete spaces consisting of a set of integers we have to choose \( a^* \) such that

\[ \sum_{y=0}^{a^*-1} p(y|x) < \frac{\eta}{\xi + \eta} \quad \text{and} \quad \sum_{y=a^*}^{\infty} p(y|x) > \frac{\eta}{\xi + \eta}. \tag{3.9} \]

We shall see later (§7.2) that this form of decisive prediction is related to a particular form of informative prediction.

**Quadratic loss point prediction.** If the loss is zero for a correct prediction and is proportional to the square of the error for a wrong prediction then we have

\[ U(a, y) = -(a - y)^2 \quad (a \in A, y \in Y) \tag{3.10} \]

and

\[ U(a) = - \int_{Y} (a - y)^2 p(y|x)dy \]

\[ = - V(y|x) - (a - E(y|x))^2, \]

where \( E \) and \( V \) denote the operations of evaluating the mean and variance of the indicated distribution. Hence \( U(a) \) takes its largest value, \( -V(y|x) \), when \( a = E(y|x) \) and so we set
### Table 3.1 Characteristics of predictive distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mode</th>
<th>q-quantile ( q = \frac{\eta}{\xi + \eta} )</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(y</td>
<td>x) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BeBi ( (N, G, H) )</td>
<td>( \begin{cases} 0 &amp; \text{if } G &lt; 1 \ \frac{(G-1)(N+1)}{G+H-2} &amp; \text{if } G \geq 1 \end{cases} )</td>
<td>largest integer ( a^* ) such that ( P_{\text{by}}(N + G + H - 1, a^* + G - 1, N, a^* - 1) &lt; q ) where ( P_{\text{by}} ) is defined in (A21) of Appendix I</td>
<td>( \frac{NG}{G+H} )</td>
</tr>
<tr>
<td>NeBi ( \left( \frac{K}{K+H} \right) )</td>
<td>( \begin{cases} 0 &amp; \text{if } G &lt; 1 \ \frac{(G-1)K}{H} &amp; \text{if } G \geq 1 \end{cases} )</td>
<td>largest integer ( a^* ) such that ( I_{K+H}(a^<em>+G)^{(a^</em>+G)}(a^*) &gt; 1 - q ) where ( I ) is defined in (A19) of Appendix I</td>
<td>( \frac{GK}{H} )</td>
</tr>
<tr>
<td>InBe ( (K, G, H) )</td>
<td>( H(K-1) )</td>
<td>solution ( a^* ) of ( I_{a^<em>+H}(a^</em>, G) = q ) where ( I ) is defined in (A19) of Appendix I</td>
<td>( \frac{KH}{G-1} )</td>
</tr>
<tr>
<td>St ( (G, B, \left( \frac{1}{K + \frac{1}{C}} \right)^{\frac{1}{2}} )</td>
<td>( B )</td>
<td>( B + \left( \frac{1}{K + \frac{1}{C}} \right)^{\frac{1}{2}} t(G; q) ) where ( t ) is defined in (A17) of Appendix I</td>
<td>( B )</td>
</tr>
</tbody>
</table>
Table 3.1 (continued)

| $p(y|x)$ | Mode | $q$-quantile ($q = \frac{n}{\xi + \eta}$) | Mean |
|----------|------|------------------------------------------|------|
| \[ \frac{CKG}{(C + K)H} \left( 1 + \frac{\omega_y(B, C, K)}{H} (y - B) \right)^{(G + 1)} \] | \[ \frac{B + H}{C} \frac{H}{C} \left( \frac{C + K}{K} \right)^{-1/G} \] if $q < \frac{K}{C + K}$ | \[ B + \frac{(C - K)H}{CK(G - 1)} \] |
| | \[ \frac{B - H}{K} \frac{H}{C} \left( 1 - q \right)^{-1/G} \] if $q > \frac{K}{C + K}$ | |
\[
\delta^*(x) = E(y|x), \quad (3.11)
\]

In words: for a quadratic loss structure the optimum prediction is the mean of the predictive distribution. For a discrete predictive distribution we would select the \(a\)-value which is closest to \(E(y|x)\).

To help in the applications of these types of prediction we provide in Table 3.1 the mode, \(q\)-quantile and mean for each of the one-dimensional predictive distributions of Table 2.2 for which a simple formulation is possible.

### 3.2 An application to machine tool replacement

In the preceding section we have selected some simple utility functions and obtained some general results about optimum prediction. In any real application, however, an appropriate utility function should be constructed to meet the particular circumstances of the problem. We shall now illustrate various aspects of such decisive prediction by determining an optimum inspection and replacement policy for the machine tool problem of example 1.2. In the present section we recognise three different types of policy as point prediction decisions and determine which particular predictions are optimum for the three types. In order to see the structure of the analysis we retain the symbols \(\xi, \eta, \xi^*\) of example 1.2 rather than use their specific values. Later in §3.4 we shall see that the three types of policy are special cases of a more general policy which can be represented as an interval prediction decision. There we shall use the specific values of \(\xi, \eta, \xi^*\) and the data to pinpoint the overall optimum policy.

**Policy 1.** Send in the inspector at time \(a\) to replace the machine tool immediately.

This is tantamount to predicting that the machine will wear out at time \(a\), and we have to decide which is the best \(a^*\) to select.

First we resolve the question of the predictive distribution. We make the reasonable assumption that the individual lifetimes are independently and exponentially distributed, say as \(\text{Ex}(\theta)\), so that the sum \(x(=1939)\) of the \(n(=24)\) lifetimes is a sufficient statistic for \(\theta\). We may thus take as the informative experiment density function

\[
p(x|\theta) = \text{Ga}(n, \theta).
\]

Suppose that we adopt a conjugate prior distribution on \(\Theta\) with density function

\[
p(\theta) = \text{Ga}(\gamma, \beta).
\]

The future experiment of interest consists in the recording of the lifetime \(y\) of a further machine tool and so we have
Decisive prediction

\[ p(y|\theta) = \text{Ex}(\theta) = \text{Ga}(1, \theta). \]

Application of case 3 of table 2.3 yields the predictive density function

\[ p(y|x) = \text{InBe}(1, G, H), \quad (3.12) \]

where \( G = g + n, H = h + x. \)

The utility function \( U(a, y) \) is readily obtained. If \( y \leq a \) the machine tool is worn out before the inspector arrives and time \( a - y \) of production is lost, so that the utility is then \(-\xi(a - y)\). If \( y > a \) then the machine tool is still functioning, and therefore there is a scrapping loss of \( y \). We have also lost a time \( y - a \) of production and so the total utility is \(-\xi - \xi(y - a)\). Hence

\[ U(a, y) = \begin{cases} -\xi(a - y) & (y \leq a), \\ -\xi - \xi(y - a) & (y > a). \end{cases} \quad (3.13) \]

We then have immediately, by (3.1), that

\[ U(a) = -\xi \int_0^a (a - y)p(y|x)dy - \int_a^\infty (\xi + \xi(y - a))p(y|x)dy, \]

so that

\[ U'(a) = \xi p(a|x) - \xi \int_0^a p(y|x)dy + \xi \int_a^\infty p(y|x)dy = \xi \left[ \frac{\xi G}{\xi H} \left( \frac{H}{H + a} \right)^{G + 1} + 2 \left( \frac{H}{H + a} \right)^G - 1 \right]. \]

To see whether

\[ U'(a) = 0 \quad (a > 0) \]

has a solution we set

\[ w = \frac{H}{H + a} \quad (3.14) \]

and consider whether the equation

\[ f(w) = \frac{\xi G}{\xi H} w^{G+1} + 2w^G - 1 = 0 \quad (0 < w < 1) \quad (3.15) \]

has a solution. Since \( f(0) = -1, f(1) > 0 \) and \( f'(w) > 0 \quad (0 < w < 1) \) we see that there is in fact a unique solution, \( w^* \) say. Thus

\[ a^* = \frac{1 - w^*}{w^*} H \quad (3.16) \]
Fig. 3.1 Graphs of utility functions (3.21) for two predictions $a, a'$ with $a < a'$.

![Graphs of utility functions](image)

(i) $\xi < \eta$

(ii) $\xi \geq \eta$

provides the optimum prediction of the machine tool lifetime, or equivalently, the optimum time at which to replace.

After some simple integration and tedious algebra it can be shown that

$$U(a^*) = \frac{\xi H}{G(G-1)w^*} \{2w^*G - G^2w^* + G^2 - 1\}. \tag{3.17}$$

We now consider an alternative policy.

**Policy 2.** Send in the inspector at time $a$ to replace the machine tool immediately if it has already worn out, and otherwise to attend the tool until it wears out.

For this policy the predictive distribution is as before but the utility function is now

$$U(a, y) = \begin{cases} -\xi(a - y) & (y < a), \\ -\eta(y - a) & (y \geq a). \end{cases} \tag{3.18}$$

(A loss of production is involved if the inspector arrives too late, a cost of labour if he arrives too early.) This is the linear loss utility structure (3.7) and so the optimum $a^*$ is the $\eta/((\xi + \eta)$-quantile of the predictive distribution, satisfying

$$\int_0^{a^*} p(y|x) \, dy = \frac{\eta}{\xi + \eta}.$$ 

Direct integration and solving the resulting equation for $a^*$ gives
Again simple integration and some algebra yield the result

\[ U(a^*) = -\frac{\xi G a^*}{G - 1}. \]  

Let us consider also a third policy.

**Policy 3.** Make the inspector attend the tool from its start until it wears out or for time \( a \), whichever is the shorter. If the machine has not worn out by time \( a \) replace it then.

Here

\[ U(a, y) = \begin{cases} -\eta y & (y \leq a), \\ -\xi - \eta a - \xi(y - a) & (y > a). \end{cases} \]  

We can immediately distinguish between two cases (i) \( \xi < \eta \) and (ii) \( \xi \geq \eta \). For each of these cases fig. 3.1 shows the graphs of \( U(a, y) \) against \( y \) for two predictions \( a \) and \( a' \) with \( a < a' \). In case (ii) \( U(a, y) \leq U(a', y) \) for every \( y \). For this case it follows that \( a^* \to \infty \), as we might expect since \( \xi \geq \eta \) means that the loss of production is greater than or equal to the cost of labour per unit time. We therefore restrict attention to the case

\[ \xi < \eta \]  

for which we shall obtain a non-trivial prediction.

The expected utility can here be expressed as

\[ U(a) = -\eta E(y|x) - \int_a^\infty \{ \xi + \eta(a - y) + \xi(y - a) \} p(y|x)dy. \]

Hence

\[ U'(a) = \xi p(a|x) - (\eta - \xi) \int_a^\infty p(y|x)dy \]

\[ = 0 \text{ at } a^* = \frac{\xi G}{\eta - \xi} - H \text{ if } \frac{\xi G}{\eta - \xi} > H, \]

\[ < 0 \text{ for all } a > 0 \text{ otherwise.} \]

The maximising \( a^* \) is thus given by

\[ a^* = \begin{cases} \frac{\xi G}{\eta - \xi} - H & \text{if } \frac{\xi G}{\eta - \xi} > H, \\ 0 & \text{if } \frac{\xi G}{\eta - \xi} \leq H. \end{cases} \]
Decisive prediction

(If the scrapping loss is small enough it pays to replace the machine tool im-
mmediately.) Again a simple expression for the maximum expected utility can
be obtained:

\[
U(a^*) = \begin{cases} 
-\frac{1}{G-1} \left( \eta H - \xi \left( \frac{H}{H + a^*} \right)^G \right) & (a^* > 0), \\
-\xi - \frac{\xi H}{G-1} & (a^* = 0).
\end{cases}
\]

(3.24)

Set prediction

Often there may be no great pressure to pinpoint the actual outcome of the
future experiment but rather a need to ensure that an interval, region or set is
provided which contains the realised outcome. For the purposes of design, for
example, it may not be so important that we know that a component functions
at a specified value of some characteristic as to be fairly sure that its operation
lies within some acceptable range. For such predictive problems we naturally
take as our action set \( A \) not the set \( Y \) of possible outcomes but the class \( \mathcal{J} \) of
(measurable) subsets of \( Y \). Presumably if we specify a set prediction \( a \) we are
happy if the actual outcome \( y \) falls in \( a \), unhappy if \( y \) falls outside \( a \) and, if
degrees of happiness are allowable, most unhappy when \( y \) falls well outside \( a \).
If we can quantify this then we again have a utility structure with utility
\( U(a, y) \) defined for every \( a \in A \) and \( y \in Y \).

We shall examine in some detail the case where \( Y \) is the whole or part of
the real line and where \( a \) is restricted by the practical requirements of the
problem to be an interval \((a_1, a_2)\). As for simple prediction we choose an
interval \( a \) which maximises the expected utility

\[
U(a) = \int_Y U(a, y) \rho(y | x) dy.
\]

The tensions in the construction of the utility function here will be on the
one hand a desire to keep the interval short and so provide a useful prediction
and on the other hand a fear of the losses involved if the interval fails to
capture the outcome.

The definition of a simple predictor extends in an obvious way to this more
general situation.

**Definition 3.2**

Set predictor. A set predictor \( \delta \) is a function

\[
\delta : X \rightarrow A = \mathcal{J}.
\]

(3.25)

The notion of optimality is again as defined in (3.2). It is possible that the
optimum set prediction may reduce to a point prediction. As in §3.1 we can
Decisive prediction

investigate the possibility of some general results, which depend on the special form of the utility function but not on the form of the predictive distribution.

All-or-nothing set prediction. Here we consider a utility structure which awards a unit if and only if the interval contains the outcome of \( f \) and where there is a cost proportional to the length of the interval. More precisely,

\[
U(a_1, a_2, y) = \begin{cases} 
1 - \gamma (a_2 - a_1) & (a_1 \leq y \leq a_2), \\
- \gamma (a_2 - a_1) & \text{otherwise}.
\end{cases}
\] (3.26)

The expected utility is then

\[
U(a_1, a_2) = \int_{a_1}^{a_2} p(y \mid x) dy - \gamma (a_2 - a_1)
= \int_{a_1}^{a_2} \{p(y \mid x) - \gamma\} dy.
\]

If therefore we place in \((a_1^*, a_2^*)\) all values of \( y \) such that \( p(y \mid x) > \gamma \) we will maximise \( U(a_1, a_2) \). Hence

\[
(a_1^*, a_2^*) = \{y : p(y \mid x) > \gamma\},
\] (3.27)
as illustrated in fig. 3.2.

If \( \gamma \) is of such a magnitude that no values of \( y \) satisfy \( p(y \mid x) > \gamma \) then a limiting argument similar to that for the all-or-nothing point prediction of §3.1 shows that the best interval prediction degenerates into a point prediction, the mode of the predictive density function \( p(y \mid x) \).

---

Fig. 3.2 Optimum set prediction for utility function (3.26).
Decisive prediction

Linear utility interval prediction. Suppose that we assert that an interval \((a_1, a_2)\) will contain the outcome \(y\) of the future experiment \(f\). Suppose further that we incur penalties if in fact \(y\) lies below \(a_1\) or above \(a_2\), and these penalties are proportional to the amounts by which \(y\) escapes the interval. If the cost attaching to the interval used is proportional to the length of the interval, say \(\gamma(a_2 - a_1)\), then we are led to investigate the piecewise-linear utility function

\[
U(a_1, a_2, y) = \begin{cases} 
-\xi(a_1 - y) - \gamma(a_2 - a_1) & (y < a_1), \\
-\gamma(a_2 - a_1) & (a_1 \leq y < a_2), \\
-\eta(y - a_2) - \gamma(a_2 - a_1) & (y > a_2),
\end{cases}
\]

where \(\xi, \eta, \gamma\) are all positive constants. Since only relative values of \(\xi, \eta\) and \(\gamma\) are of importance we take \(\gamma = 1\) without loss of generality. By varying \(\xi\) and \(\eta\) we allow for the possibility of differential losses associated with the interval overshooting or undershooting the true \(y\).

If \(1/\xi + 1/\eta < 1\) we find that maximisation of the expected utility leads to selecting an interval \((a_1^*, a_2^*)\) given by the solutions of

\[
\frac{1}{\xi} \int_{a_1^*}^{a_2^*} p(y|x)\,dy = 1/\eta.
\]

If however \(1/\xi + 1/\eta > 1\) the optimal interval prediction degenerates into a point prediction \(a^*\) given by the solution of

\[
\frac{1}{\xi} \int_{-\infty}^{a_2^*} p(y|x)\,dy = \frac{\eta}{\xi + \eta}.
\]

Fig. 3.3 shows the graph of the utility function (3.28) for (i) \(1/\xi + 1/\eta < 1\), (ii) \(1/\xi + 1/\eta > 1\).
Decisive prediction is inadmissible since there exists a point prediction, namely $a_0 = (\xi a_1 + \eta a_2) / (\xi + \eta)$, with uniformly larger expected utility.

3.4 Further analysis of the machine tool problem

As an example of set or interval prediction we can consider a more general policy than we have so far investigated.

**Policy 4.** Send in the inspector at time $a_1$ and have him replace the machine tool as soon as it wears out or at time $a_2$ whichever is earlier.

One interpretation of such a policy is that we are regarding $(a_1, a_2)$ as a prediction interval since it is clearly our hope that the lifetime $\nu$ will fall in the interval $(a_1, a_2)$. Note that the three policies previously considered are all special cases of this policy. For policy 1, $a_2 = a_1$; for policy 2, $a_2 = \infty$; for policy 3, $a_1 = 0$. It is thus of interest to ask under what conditions one of those three policies will turn out to be optimum within this more general policy framework.

For this case,

$$U(a_1, a_2, \nu) = \begin{cases} -\xi (a_1 - \nu) & (\nu < a_1), \\ -\eta (\nu - a_1) & (a_1 \leq \nu \leq a_2), \\ -\xi (a_2 - a_1) - \xi (\nu - a_2) & (\nu > a_2). \end{cases}$$

(3.31)

For example if $\nu > a_2$, a scrapping loss $\xi$ is involved, there is a cost of inspection for the whole interval, and we also have to debit for the unused lifetime $\nu - a_2$ of the machine tool. Again we assume that $\xi < \eta$. The expected utility can be expressed as the sum of a function of $a_1$ only and a function of $a_2$ only:

$$U(a_1, a_2) = -\xi \int_0^{a_1} (a_1 - \nu) p(\nu | x) d\nu - \eta \int_{a_1}^{a_2} (\nu - a_1) p(\nu | x) d\nu$$

$$- \int_{a_2}^{\infty} [(\nu - a_2) + \xi (\nu - a_2)] p(\nu | x) d\nu$$

$$= F_1(a_1) + F_2(a_2),$$

where

$$F_1(a_1) = \eta a_1 - (\xi + \eta) \int_0^{a_1} (a_1 - \nu) p(\nu | x) d\nu,$$

$$F_2(a_2) = -\eta E(\nu | x) - \int_{a_2}^{\infty} [(\nu - a_2) + \xi (\nu - a_2)] p(\nu | x) d\nu.$$
We have that
\[ \frac{\partial U}{\partial a_1} = \eta - (\xi + \eta) \int_0^a p(y|x)dy \]
\[ = \eta - (\xi + \eta) \left[ 1 - \left( \frac{H}{H + a_1} \right)^G \right] \]
and
\[ \frac{\partial U}{\partial a_2} = \xi p(a_2|x) - (\eta - \xi) \int_{a_1}^a p(y|x)dy \]
\[ = \frac{(\eta - \xi)H^G}{(H + a_2)^{G+1}} \left( \frac{\xi G}{\eta - \xi} - H - a_2 \right). \]

Note first that the equations
\[ \frac{\partial U}{\partial a_1} = \frac{\partial U}{\partial a_2} = 0 \quad (a_1, a_2 > 0) \]
have roots
\[ a_1^* = H \left( \left( \frac{\xi + \eta}{\xi} \right)^{1/G} - 1 \right), \quad (3.32) \]
\[ a_2^* = \frac{\xi G}{\eta - \xi} - H, \quad (3.33) \]
provided
\[ \frac{\xi G}{\eta - \xi} > H, \quad (3.34) \]
see policy 3. From now on we shall suppose that this condition holds.

Then it can easily be shown that at \((a_1^*, a_2^*)\)
\[ \frac{\partial^2 U}{\partial a_1^2} < 0, \quad \frac{\partial^2 U}{\partial a_2^2} < 0, \quad \frac{\partial^2 U}{\partial a_1 \partial a_2} = 0. \]

Hence, if \(a_1^* < a_2^*\), \(U\) attains its maximum value within the region \(0 < a_1 < a_2\)
at the local maximum \((a_1^*, a_2^*)\); otherwise \(U\) has its maximum on the boundary \(a_2 = a_1\).

Hence \((a_1^*, a_2^*)\) is the optimum interval if and only if
\[ \frac{\xi G}{\eta - \xi} > H \left( \frac{\xi + \eta}{\xi} \right)^{1/G} (> H). \quad (3.35) \]

In this case the maximum utility is
\[ U(a_1^*, a_2^*) = - \frac{1}{G-1} \left\{ \xi G a_1^* - \xi \left( \frac{H}{H + a_2^*} \right)^G \right\}. \quad (3.36) \]
For the case
\[
\frac{\xi G}{(\eta - \xi)} \leq H \left( \frac{\xi + \eta}{\xi} \right)^{VG}
\]
(3.37)

there is no local maximum in the region \(0 < a_1 < a_2\) but we know that the absolute maximum lies on the boundary \(a_2 = a_1\). In this case we obtain a point estimate which is the solution of
\[
\frac{\xi G}{\xi H} \left( \frac{H}{H + a} \right)^{G+1} + 2 \left( \frac{H}{H + a} \right)^G - 1 = 0,
\]
that is the same solution as for policy 1; see (3.15).

We are now in a position to examine the consequences of using the specific values \(\xi = 1.8, \eta = 2.4, \xi = 54\). If we assume that information is vague concerning \(\theta\) before the evidence of the lifetimes of the 24 machine tools then by case 3 of table 2.1 we take \(\gamma = h = 0\), so that \(G = n = 24, H = x = 1939\). We note that \(\xi < \eta\), that the first condition in (3.23) is satisfied and that condition (3.35) is also satisfied. Straightforward calculations then lead to optimum policies and maximised expected utilities as shown in table 3.2. Since policies 1, 2 and 3 are special cases of policy 4 the optimum form of policy 4 must be overall best, but table 3.2 shows that the advantage over the optimum form of policy 2 is slight.

<table>
<thead>
<tr>
<th>Policy</th>
<th>Optimum (a^<em>) or ((a_1^</em>, a_2^*))</th>
<th>Maximised (U(a^<em>)) or (U(a_1^</em>, a_2^*))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.6</td>
<td>-131.5</td>
</tr>
<tr>
<td>2</td>
<td>69.7</td>
<td>-130.9</td>
</tr>
<tr>
<td>3</td>
<td>221.0</td>
<td>-202.2</td>
</tr>
<tr>
<td>4</td>
<td>69.7, 221.0</td>
<td>-130.7</td>
</tr>
</tbody>
</table>

For the specific relative values of \(\xi, \eta\) and \(\xi\) the choice of optimum policy is delicately balanced as we can easily demonstrate by considering changes in the rate \(\eta\) of labour cost for fixed values of \(\xi\) and \(\xi\). Fig. 3.4 shows the graphs of \(a_1^*\) and \(a_2^*\) and fig. 3.5 the graphs of the maximised expected utilities for each of the four policies, plotted against \(\eta\) in the neighbourhood of \(\eta = 2.4\). As \(\eta\) increases from 2.4 the advantage of policy 4 over policy 2 increases, while the advantage over policy 1 decreases. For \(\eta > 2.445\) policy 1 is overall best. As \(\eta\) decreases from 2.4 the advantage of policy 4 over policy 2 decreases as the appropriate inspection period for policy 4 expands due to the reduction of labour costs. For \(\eta < 1.8\) condition (3.22) is not satisfied and policy 2 is the optimum form. Only when \(\eta = 0\) is policy 3 worth considering.
2.3  2.4  2.5

\[ a_1^* = a_2^* \]

Fig. 3.4 Graphs of \( a_1^* \) and \( a_2^* \) against \( \eta \) for the machine tool problem.

3.5 Computational aids for more general utility functions

Earlier, in §§3.1 and 3.3, we obtained for some simple utility structures some general results applicable whatever the predictive distribution. The example we have just studied shows that for more complicated utility functions we are forced to take account of the particular form of \( p(y|x) \) to obtain a useful result. If the utility function can be expressed as a polynomial in \( y \), or is piecewise polynomial in \( y \), then some help can be given in the evaluation of \( U(a) \). For any polynomial or piecewise polynomial in \( y \) can be expressed as a linear combination of simpler utilities of one of the following forms:

\[
U_j(a, y) = \begin{cases} 
  y(y-1) \ldots (y-j+1) & (y < a), \\
  0 & (y \geq a);
\end{cases} 
\]  

(3.38)
Fig. 3.5 Graphs of maximised expected utilities against $\eta$ for the four policies.
Decisive prediction

\[ U_j(a, y) = \begin{cases} 
  y & (y < a), \\
  0 & (y \geq a);
\end{cases} \quad (3.39) \]

\[ U_j(a, y) = \begin{cases} 
  (y - r)^j & (y < a), \\
  0 & (y \geq a).
\end{cases} \quad (3.40) \]

The first is a form best suited to the investigation of discrete \( Y \), the second and third are the better forms for the investigation of continuous \( Y \).

Example 3.1

**Linear utility point prediction.** For

\[ U(a, y) = \begin{cases} 
  -\xi (a - y) & (y < a), \\
  -\eta (y - a) & (y \geq a),
\end{cases} \quad (3.41) \]

we have

\[ U(a, y) = \xi U_1(a, y) - \xi a U_0(a, y) + \eta a \{1 - U_0(a, y)\} \\
  - \eta \{U_1(\infty, y) - U_1(a, y)\} \quad (3.42) \]

for forms (3.38) and (3.39); and

\[ U(a, y) = \xi U_1(a, y) - \xi (a - r) U_0(a, y) + \eta (a - r) \{1 - U_0(a, y)\} \\
  - \eta \{U_1(\infty, y) - U_1(a, y)\} \quad (3.43) \]

for form (3.40).

Example 3.2

**Interval prediction.** When a finite prediction interval \((a_1, a_2)\) is required and when losses are quadratic in the distance of \( y \) from the interval, and also increase quadratically with distance inside the interval (the deeper \( y \) is contained in the interval the more we may have used too large a prediction interval) we may have a utility function

\[ U(a_1, a_2, y) = \begin{cases} 
  -(y - a_1)^2 & (y < a_1), \\
  -(y - a_1)(a_2 - y) & (a_1 \leq y \leq a_2), \\
  -(y - a_2)^2 & (y \geq a_2).
\end{cases} \quad (3.44) \]

This is again easily expressed in terms of the \( U_j \) functions. For example, in terms of the form (3.39),

\[ U(a_1, a_2, y) = -2U_2(a_1, y) + (3a_1 + a_2)U_1(a_1, y) \\
  - a_1(a_1 + a_2)U_0(a_1, y) + 2U_2(a_2, y) \\
  - (a_1 + 3a_2)U_1(a_2, y) + a_2(a_1 + a_2)U_0(a_2, y) \\
  + 2a_2U_1(\infty, y) - a_2^2 - U_2(\infty, y). \quad (3.45) \]
Table 3.3 Evaluation of $U_1(\omega)$ for standard predictive distributions

<table>
<thead>
<tr>
<th>Predictive distribution</th>
<th>$U_1(\omega, \gamma)$</th>
<th>$U_1(\omega)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeBi $(N, G, H)$</td>
<td>$N! B(G+j, H) / (N-j)! B(G, H)$</td>
<td>$P_{NH}(N+G+H-1, G+a-1, N-j, a-1-j)$</td>
</tr>
<tr>
<td>NeBi $\left(G, \frac{K}{K+H}\right)$</td>
<td>$(G+j-1)! K^j / (G-1)! H(K+H)^{(G+j,a)}$</td>
<td>$(N-I) B(G, H) B(G-1)! H^j (G+j/H) G^j (G-1)!$</td>
</tr>
<tr>
<td>InBe $(K, G, H)$</td>
<td>$H! B(K+j, G)/B(K, G) I_{a/H+a} (K+j, G+j-1)$</td>
<td>$(G-K)! H^j (G+j/IK) G^j (G-1)!$</td>
</tr>
<tr>
<td>St $\left(G, B, \left(\frac{1}{K+C}\right)^{-1/G}\right)$</td>
<td>$r = B$</td>
<td>$(a-B)^j (a-B)^{H} + (a-B)^3$</td>
</tr>
</tbody>
</table>

where $\gamma = \frac{(a-B)^j (a-B)^{H} + (a-B)^3}{G}$
Integrating \( U(a, y) \) or \( U(a_1, a_2, y) \) with respect to \( p(y \mid x) \) over \( Y \) to obtain the expected utilities \( U(a) \) and \( U(a_1, a_2) \) is a linear operation, and so we get precisely the forms (3.42), (3.43) and (3.45) above with the \( y \) omitted for these expected utilities in terms of

\[
U_f(a) = \int_Y U_f(a, y)p(y \mid x)dy. \tag{3.46}
\]

It is therefore useful to have a catalogue of \( U_f(a) \) corresponding to the simple \( U_f(a, y) \) utility functions and the standard predictive distributions. These are provided in table 3.3.

### 3.6 An alternative formulation

The expected utility \( U(a) \), which is the criterion for optimum prediction, can be expressed in an alternative form. For, under regularity conditions allowing a change of order of integration, we have the following development:

\[
U(a) = \int_Y U(a, y)p(y \mid x)dy
= \int_Y U(a, y) \left( \int_\Theta p(y \mid \theta)p(\theta \mid x)d\theta \right) dy \quad \text{by (2.4)}
= \int_\Theta \left( \int_Y U(a, y)p(y \mid \theta)dy \right)p(\theta \mid x)d\theta
= \int_\Theta V(a, \theta)p(\theta \mid x)d\theta, \tag{3.47}
\]

where

\[
V(a, \theta) = \int_Y U(a, y)p(y \mid \theta)dy. \tag{3.48}
\]

This formulation shows that we have a second way of viewing the problem because \( U(a) \) has been expressed as the expectation of the induced utility \( V(a, \theta) \) with respect to the posterior plausibility function \( p(\theta \mid x) \). From any basic utility function \( U(a, y) \), showing the assessment of utilities in terms of \( a \) and \( y \), we can derive by (3.48) an equivalent induced utility \( V(a, \theta) \), an assessment of utilities in terms of \( a \) and the index \( \theta \).

Not every function on \( A \times \Theta \) necessarily arises from a function defined on \( A \times Y \) by the relationship (3.48). For example

\[
V(a_1, a_2, \theta) = \begin{cases} 
1 - \gamma(a_2 - a_1) & \text{if } \int_{a_1}^{a_2} p(y \mid \theta)dy \geq c, \\
\gamma(a_2 - a_1) & \text{otherwise,}
\end{cases} \tag{3.49}
\]

is such a \( V \) function. This prompts us to ask the question: Are there any
situations of prediction where a $V$ formulation of utility applies and where
there is no corresponding $U$ specification? To answer this in the affirmative
we need only envisage the future experiment $f$ being repeated a large number
of times and an interval $(a_1, a_2)$ being regarded as of any use at all if it con-
tains at least a proportion $c$ of the outcomes from these replicates. We could
then score 1 for success, 0 for failure. If the cost of an interval is proportional
to its length we would then have the $V$ function as specified in (3.49).

We shall see in §7.3 that the $U$ and $V$ specifications also provide a means of
distinguishing between two fundamental types of informative prediction
approaches.

History

The foundations of decision theory were laid down by Wald (1950) and the
subject has expanded rapidly since then along both classical and Bayesian lines;
see, for example, De Groot (1970), Ferguson (1967), Raiffa and Schlaifer (1961).
The computational aids for more general utility functions in §3.5 and the
alternative formulation of §3.6 are both discussed in Aitchison and Sculthorpe
(1965).

Problems

3.1 For the situations of problems 2.1–2.4 find the all-or-nothing simple pre-
dictors, the linear loss simple predictors and the quadratic loss simple predictors
as described in §3.1.

3.2 Complete the analysis of problem 1.2.

3.3 The research and development department associated with the extrusion
process of problem 2.5 has been attempting to formulate the problem of pre-
dicting the number of flaws in 60 metre lengths of fibre as a decision problem.
Two suggestions concerning the possible losses involved have been put forward:
(i) a prediction is useless unless it is correct,
(ii) the loss involved in a wrong prediction is equal to $10d^2$, where $d$ is the
difference between the predicted and the actual number of flaws.

What predictions would you advocate on the basis of these suggestions?
What is the expected loss per prediction in case (ii)?

3.4 A delicatessen store owner has to sign a long-term contract whereby a
fixed amount of a perishable new delicacy will be delivered to him daily. Any
quantity sold on the day of delivery brings a profit of 3p per g whereas any
quantity unsold involves a loss of 4p per g. After discussion with the store
Decisive prediction

owner you feel that it is safe to assume that daily demand will be normally distributed but you remain vague about the values of the mean and variance parameters. During the 10 days before the contract has to be signed the daily amounts (g) demanded are

\[ 815, 920, 880, 830, 1125, 845, 990, 1200, 844, 1015. \]

Determine the optimum fixed daily order for the owner and the expected profit with this order.

3.5 In order to obtain favourable terms a theatre ticket agency has, at the end of the first week of a new show, to give a firm commitment to take a fixed number of theatre seats for each daily performance. The agency reckons that the daily demand of its clientele is Poisson-distributed but it is very vague about the mean parameter. The numbers for the first six performances are 8, 6, 3, 7, 2, 5. For each ticket sold the agency makes a profit of 50p, for each ticket unsold, a loss of £1. What fixed number of seats per day should the agency order, and what is its expected profit with this number?

3.6 Analyse the problem of interval prediction with the piecewise-linear utility function of the form (3.28) for the situations in problems 2.1 and 2.3.

3.7 Consider again the machine tool replacement problem (example 1.2, §§3.2, 3.4) and derive the optimal actions for the following policies (a) and (b).

(a) Send in the inspector at time \( a \) for a given interval of time of length \( T \). Replace the machine tool as soon as it wears out or at the end of the interval whichever is earlier. If the machine tool has worn out before the inspector arrives the inspector replaces it and can be reallocated to another job at zero cost. Otherwise he is assigned for the whole period.

(b) As for policy (a) except that the inspector is contracted from outside and cannot be reallocated to another job.

Re-examine the relevant policies if a machine tool which is replaced while it is still functioning can be sold and some of its cost recouped, i.e. \( \xi < 0 \).

3.8 In a certain country income tax policy is to tax only incomes above \( k \), the tax being a fixed amount \( q \). There is now a proposal to increase this tax from \( q \) to \( r \) for taxable incomes above \( l \), where \( l > k \), but to decrease the amount from \( q \) to \( s \) for taxable incomes below \( l \). It is assumed that currently taxable incomes follow a \( \text{P} \& (k, \theta) \) distribution, where little is known about the indexing parameter \( \theta \). A random sample of \( n \) currently taxable incomes shows recorded incomes \( x_1, \ldots, x_n \).

Show that the proposal will provide a higher tax yield than current policy
Decisive prediction

if \( I < k(\bar{x}_g/k)^k \), where \( h = n \left[ (q-s)/(q-s) \right]^{1/n} - 1 \) and \( \bar{x}_g \) is the geometric mean of \( x_1, \ldots, x_n \).

3.9 A production run of an industrial chemical process produces a large number of drums of sterilising liquid, which vary in acidity according to a \( \text{No}(\mu, \tau) \) distribution, where \( \tau \) is known. The mean acidity \( \mu \) of a run is not known, but \( \mu \) is known to vary from run to run according to a \( \text{No}(b, c) \) distribution, where \( b \) and \( c \) are known. If an alkalising agent of strength \( a \) is injected into a drum of acidity \( y \) the resulting liquid has acidity \( y - a \) if \( a \leq y \) or alkalinity \( a - y \) if \( a > y \). The liquid is fully effective only when it is neutral; the loss of effectiveness from acid or alkaline liquid is proportional to the acidity or alkalinity, the factor of proportionality for acidity being thrice that for alkalinity. Administrative considerations dictate that each drum of a given run should receive the same strength of alkalising agent, and that only \( n \) drums from a run can be tested for acidity prior to treatment of the run.

Determine the optimum strength of alkalising agent for a run whose sampled drums have shown acidities \( x_1, \ldots, x_n \).

3.10 Suppose that in a decisive prediction problem

\[
A = Y = R^n, \\
U(a, y) = -(a - y)'M(a - y),
\]

where \( M \) is a positive definite matrix. Show that the optimum prediction is

\[
a^* = E(y|x)
\]

and that the maximised expected utility is

\[
-\text{trace } MV(y|x),
\]

where \( E(y|x) \) and \( V(y|x) \) are the mean and covariance matrix of the predictive distribution.
4

Informative prediction

4.1 Introduction

As already pointed out in chapter 2 the essence of prediction analysis from the Bayesian point of view is the construction of the predictive density function $p(y|x)$. When the appropriateness of predictions can be assessed in terms of a utility function the decisive prediction approach of chapter 3 resolves the prediction problem by maximising expected utility, expectation being taken with respect to the predictive density function. When there is difficulty in specifying the utility function and yet we want to convey some summary form of information about the plausible outcomes of the future experiment $f$, some principle other than the maximisation of expected utility must be introduced. To draw a clear distinction we suggest the term informative prediction to describe the use of any principle of prediction which does not require the specification of a utility function. As in previous chapters we use $X$ and $Y$ to denote the sample spaces of the informative experiment $e$ and the future experiment $f$, and $Y$ to denote the event space of $f$ or the class of measurable subsets of $Y$.

Definition 4.1

Informative predictor. An informative predictor is a function

$$\delta : X \rightarrow Y$$

which satisfies some probabilistic relation based on the probability measures associated with $p(x|\theta), p(y|\theta)$ and, in the case of a Bayesian informative predictor, $p(\theta)$.

An informative predictor thus instructs the experimenter what region of the future experiment sample space he must use if his informative experiment yields $x$: he must use the informative prediction $\delta(x)$.

The object of an informative prediction region is clearly to narrow the range of possibility of a future observation from $Y$, so that receivers of this information may more readily plan or take appropriate action. Thus a manufacturer of components may wish to give to his customers some indication of the likely range of values within which a component characteristic (such
as electrical resistance, lifetime) may lie. Again, a medical research worker who has devised a method of determining the urinary excretion rate of some steroid metabolite will wish to convey to other workers the "normal range": he wants to provide a prediction interval within which other workers can be reasonably sure that most excretion rates of healthy persons lie. There is in such predictions, as in all statistical problems, a conflict between usefulness and validity. If we want to be absolutely sure that our prediction region will capture the actual outcome of a future experiment then we should quote $Y$ as our prediction region. If we want to convey an extremely useful prediction then we should attempt to say that some specific $a \in Y$ will happen at the future performance, and probably place a forlorn hope on this point prediction $a$ actually occurring. Clearly some compromise between these two extremes is necessary. In chapter 3 we saw how this can be achieved, when the effects of the conflict may be expressed by a suitable utility function, through the sophisticated yet simple principles of statistical decision theory. In the present and the following two chapters we study more primitive, and in some respects more complicated, analyses of such situations. These analyses have arisen to meet the needs of situations where it is difficult to set down a specific utility function — for example, where a manufacturer is trying to meet the needs of many customers.

In the present chapter we first discuss various aspects of such informative prediction from a Bayesian viewpoint using the predictive distribution. We then set the basis of the non-Bayesian or frequentist approach to informative prediction by introducing the general concept of the coverage distribution. Study of this distribution leads to two main types of prediction which have come to be known under the general heading of statistical tolerance regions. The two types — mean coverage and guaranteed coverage tolerance predictions — are separately developed in chapters 5 and 6.

4.2 Bayesian informative prediction

From a Bayesian viewpoint the natural way to assess the effectiveness of an informative prediction $a$ is in terms of the probability that a future outcome lies in $a$. This probability, assessed in terms of the predictive distribution, we term the Bayesian cover provided by $a$.

**Definition 4.2**

**Bayesian cover.** The informative prediction $a \subset Y$ is said to have Bayesian cover $\kappa$ if and only if

$$P(a | x) \equiv \int_a p(y | x)dy = \kappa.$$  

(4.2)
Example 4.1

Bayesian cover of replacement policy. The Bayesian cover provided by the optimum prediction \((a_1^*, a_2^*)\) associated with policy 4 of §3.4 is, when condition (3.35) holds,

\[
\int_{a_1^*}^{a_2^*} p(y|x) dy = \left( \frac{H}{H + a_1^*} \right)^G - \left( \frac{H}{H + a_2^*} \right)^G
\]

\[
= \frac{\xi}{\xi + \eta} - \left( \frac{(\eta - \xi)H}{\xi G} \right)^G
\]

\[
= 0.35
\]

for \(\xi = 1.8, \eta = 2.4, \xi = 54\). The practical interpretation of Bayesian cover here is the assessed probability that a machine tool will actually wear out during the attendance period. We are prepared to accept a low cover or probability 0.35 that the attendant will be present when the tool fails because of relatively high attendance costs.

There may be many predictions providing Bayesian cover equal to a specified value \(K\). A natural way of removing this ambiguity is to build up a prediction region by placing into it the most plausible values of \(y\). This leads to the following formal definition.

Definition 4.3

Most plausible Bayesian prediction. A prediction \(a\) is a most plausible Bayesian prediction of cover \(K\) if \(a\) has the form

\[
a = \{ y : p(y|x) > \gamma \}
\]

(4.4)

where \(\gamma\) is determined by

\[
P(a|x) = K.
\]

(4.5)

We have already met an example of such Bayesian prediction in (3.27) for the all-or-nothing decisive interval prediction of (3.26). This correspondence gives us an interpretation for \(\gamma\) in (4.4), as the cost per unit interval in the equivalent all-or-nothing decisive prediction. Fig. 4.1 also gives a diagrammatic view of the determination of \(\gamma\) in the construction of a most plausible Bayesian prediction \(a\) of cover \(K\); we must adjust the horizontal line at height \(\gamma\) until the shaded area above the corresponding interval \(a\) and under the prediction density curve is \(K\).

Example 4.2

Most plausible Bayesian predictor for tool lifetime. Fig. 4.2 shows the graph of the predictive density function (3.12) associated with the machine tool.
Fig. 4.1 A most plausible Bayesian prediction of cover $\kappa$.

Fig. 4.2 Predictive density function (3.12) for the machine tool problem.

replacement problem of §§3.2 and 3.4. From the shape of the graph we see immediately that any interval $a$ satisfying (4.4) must be of the form $(0, a)$ and $a$ is determined by (4.5).
Informative prediction

\[ \int_0^a \frac{GH^G}{(H + y)^{\nu + 1}} \, dy = \kappa, \]

so that

\[ a = H \{(1 - \kappa)^{-\nu G} - 1\}. \quad (4.6) \]

Moreover

\[ \gamma = G(1 - \kappa)^{1 - (1/G)/H}. \quad (4.7) \]

To obtain 95 per cent cover with this type of interval we would have to be prepared to attend the tool until 257.8 minutes after the start of its operation. The cost per minute associated with the equivalent all-or-nothing interpretation is then \( \gamma = 0.0005 \).

4.3 Region of previous experience

An interesting application of most plausible Bayesian prediction occurs in finding statistical expression for such terms as 'within previous experience' and 'outside previous experience'. Suppose that \( x = (x_1, \ldots, x_n) \) is a set of the outcomes of \( n \) replicates of some basic experiment, and we are assessing whether an outcome \( y \) from a newly performed experiment \( f \) can reasonably be regarded as within the experience of the \( n \) previous results. We first construct the predictive density function \( p(y|x) \) which provides us with a measure of the plausibility of \( y \) based on previous experience \( x \). We can then compare this with the plausibility, as assessed by this predictive distribution, of each of the actually observed \( x_i \), namely \( p(x_i|x) \). If and only if \( p(y|x) \) is greater than the minimum of these \( p(x_i|x) \) do we say that \( y \) is within previous experience. Thus our region of previous experience is

\[ a = \{y : p(y|x) \geq \min_i p(x_i|x)\}. \quad (4.8) \]

Such a region of previous experience is thus a most plausible Bayesian prediction with \( \gamma \) specified by

\[ \gamma = \min_i p(x_i|x). \quad (4.9) \]

Note that the concept of the region of previous experience differs from that of range of previous experience defined as

\( \min x_i, \max x_i \)

for the one-dimensional case. For instance in example 4.2 the region of previous experience is \((0, 290)\) whereas the range is \((4, 290)\). Indeed the short lifetimes omitted from the range are the most plausible under the predictive distribution.
Cover of the region of previous experience. It is sometimes of interest to compute the cover associated with a prediction region. The interpretation of cover here is simply the probability that a new observation will fall within previous experience. For example we consider the multivariate normal situations in which we are led to predictive distributions of generalised Student form.

Suppose that

\[ p(y | x) \text{ is } \text{Std}(k, b, c). \]

Then if we write

\[ \lambda = \max_i (x_i - b)'c^{-1}(x_i - b) \]  

(4.10)

the region of previous experience simplifies to

\[ a = \{y : (y - b)'c^{-1}(y - b) \leq \lambda\} \]  

(4.11)

and so is the interior of an ellipsoid of concentration of the Student distribution.

The cover is given by

\[ \int_a p(y | x) dy. \]

Although this integral appears complicated it can be reduced to a standard function by the following series of elementary transformations and steps.

(i) \( y - b = \sqrt{k}Wz \) where \( W'cW = I_d \) is the unit matrix of order \( d \), the dimension of the multivariate distribution.

(ii) \( z \to \) generalised spherical polar coordinates \( (r, \psi_1, ..., \psi_{d-1}) \).

(iii) Since the integrand factorises to \( f(r)g(\psi) \) and the region of integration is determined only by a restriction on \( r \), we can easily integrate out \( \psi_1, ..., \psi_{d-1} \), obtaining

\[ \int_0^2 \int_0^{\sqrt{\lambda/k}} \frac{r^{d-1}}{(1 + r^2)^{(k+1)/2}} dr. \]

(iv) \( u = r^2/(1 + r^2) \) or \( r = \sqrt{u/(1 - u)} \) then gives

\[ \int_a p(y | x) dy = \int_0^{\lambda/(\lambda+k)} \frac{\lambda^{d/2} - 1 \cdot (1 - \mu)^{(k-d+1)/2} - 1}{B\{\frac{1}{2}, \frac{1}{2}(k - d + 1)\}} d\mu \]

an incomplete beta function, as defined in (A19) of appendix I.

4.4 An application to metabolite excretion rates

Table 4.1 shows the urinary excretion rates (mg/24h) of two steroid
Table 4.1 Urinary excretion rates (mg/24h) of cortisol and cortisone for 27 cases of Cushing's syndrome with adrenal hyperplasia

<table>
<thead>
<tr>
<th>Patient no.</th>
<th>Cortisol</th>
<th>Cortisone</th>
<th>Value of quadratic form</th>
<th>Patient no.</th>
<th>Cortisol</th>
<th>Cortisone</th>
<th>Value of quadratic form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.41</td>
<td>0.38</td>
<td>1.66</td>
<td>15</td>
<td>0.48</td>
<td>0.36</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.18</td>
<td>2.97</td>
<td>16</td>
<td>0.80</td>
<td>0.39</td>
<td>2.42</td>
</tr>
<tr>
<td>3</td>
<td>0.26</td>
<td>0.15</td>
<td>0.79</td>
<td>17</td>
<td>0.40</td>
<td>0.24</td>
<td>1.12</td>
</tr>
<tr>
<td>4</td>
<td>0.34</td>
<td>0.33</td>
<td>1.51</td>
<td>18</td>
<td>0.22</td>
<td>0.10</td>
<td>3.46</td>
</tr>
<tr>
<td>5</td>
<td>1.12</td>
<td>0.60</td>
<td>4.71</td>
<td>19</td>
<td>0.24</td>
<td>0.24</td>
<td>1.44</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
<td>0.14</td>
<td>2.36</td>
<td>20</td>
<td>0.56</td>
<td>0.42</td>
<td>1.48</td>
</tr>
<tr>
<td>7</td>
<td>0.20</td>
<td>0.16</td>
<td>0.96</td>
<td>21</td>
<td>0.40</td>
<td>0.16</td>
<td>2.50</td>
</tr>
<tr>
<td>8</td>
<td>0.26</td>
<td>0.18</td>
<td>0.27</td>
<td>22</td>
<td>0.88</td>
<td>0.48</td>
<td>2.94</td>
</tr>
<tr>
<td>9</td>
<td>0.56</td>
<td>0.32</td>
<td>0.79</td>
<td>23</td>
<td>0.44</td>
<td>0.26</td>
<td>0.24</td>
</tr>
<tr>
<td>10</td>
<td>0.26</td>
<td>0.20</td>
<td>0.30</td>
<td>24</td>
<td>0.24</td>
<td>0.16</td>
<td>0.53</td>
</tr>
<tr>
<td>11</td>
<td>0.16</td>
<td>0.13</td>
<td>1.90</td>
<td>25</td>
<td>0.27</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>12</td>
<td>0.56</td>
<td>0.33</td>
<td>0.79</td>
<td>26</td>
<td>0.18</td>
<td>0.18</td>
<td>1.92</td>
</tr>
<tr>
<td>13</td>
<td>0.33</td>
<td>0.08</td>
<td>11.49</td>
<td>27</td>
<td>0.60</td>
<td>0.35</td>
<td>1.03</td>
</tr>
<tr>
<td>14</td>
<td>0.26</td>
<td>0.22</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Informative prediction

metabolites, cortisol and cortisone, determined by paperchromatography for 27 patients diagnosed as having Cushing's syndrome with adrenal hyperplasia. There are other forms of Cushing's syndrome and other steroid metabolites which are relevant to the differential diagnosis of the syndrome (§ 11.1) but we restrict attention here to the construction of a region of previous experience for cortisol and cortisone, since this problem is sufficient to illustrate the technique.

The first task is to construct the predictive distribution. To cope with the skewness of the data we recognise the approximate lognormality and work throughout with logarithms (to the base 10) of the excretion rates. The transformed data then arise from \( n = 27 \) replicates of a multinormal \( \mathcal{N}_d(\mu, \Sigma) \) experiment with dimension \( d = 2 \). These data can then be summarised sufficiently in the two-dimensional vector mean \( \mu \) and the \( 2 \times 2 \) matrix \( \Sigma \) of corrected sums of squares and cross-products. We are thus concerned with case 6 of table 2.3 with \( \mu \) and \( \Sigma \) independently distributed as \( \mathcal{N}_2(\mu, n\Sigma) \) and \( \mathcal{W}_2(n - 1, \tau) \). We are interested in the problem of predicting the two excretion rates for a new hyperplasia patient, and so in a \( \mathcal{N}_2(\mu, \Sigma) \) future experiment \( f \). Following through the construction process of table 2.3 using the vague prior for \( (\mu, \tau) \) we arrive at the predictive distribution

\[
S_{t_2} \left\{ a - 1, m, \left(1 + \frac{1}{n}\right) \frac{\Sigma}{n - 1} \right\}.
\]

(4.13)

For the given data

\[
n - 1 = 26, \quad m = \begin{bmatrix} -0.4650 \\ -0.6403 \end{bmatrix}, \quad \frac{n(n - 1)}{n + 1} \Sigma^{-1} = \begin{bmatrix} 47.8802 & -42.1236 \\ -42.1236 & 58.1050 \end{bmatrix},
\]

The values of the quadratic forms in (4.10) are then easily computed and are also shown in table 4.1. The maximum of these is \( \lambda = 11.49 \) and so the region of previous hyperplasia experience is bounded by the ellipse

\[
47.8802(y_1 + 0.4650)^2 + 58.1050(y_2 + 0.6403)^2 - 84.2472(y_1 + 0.4650)(y_2 + 0.6403) = 11.49.
\]

Fig. 4.3 shows the positions of the 27 cases in the cortisol–cortisone plane and the bounding ellipse. The cover provided by this region of previous experience is, by (4.12),

\[
\lambda_{0.99} (1, 12.5) = 0.99,
\]

so that we would expect only about one in a hundred of hyperplasia patients...
Informative prediction

Fig. 4.3 Region of previous experience of 27 patients having Cushing's syndrome with adrenal hyperplasia and the relative positions of 10 other patients.

to have cortisol—cortisone values outside the ellipse. Table 4.2 shows the excretion rates for ten new patients, their quadratic form values and the confirmed diagnostic form of Cushing's syndrome for the eight patients found to be suffering from it. The point representations of the ten patients are shown in fig. 4.3. All six of the new hyperplasia cases A–F fall within the region of previous hyperplasia experience, and two non-hyperplasia cases G and H, fall outside the region. The fact that the other two non-hyperplasia cases fall within the hyperplasia region is a reminder that the construction of such a region of previous experience does not provide a statistical means of diagnosing between hyperplasia and non-hyperplasia. Any such differential diagnostic system must be built not only on hyperplasia experience but also on information about the excretion rates in non-hyperplasia cases. Regions of previous experience for the other forms of Cushing's syndrome do in fact intersect the hyperplasia region of fig. 4.3. What the construction of a region of hyperplasia
Table 4.2 Urinary excretion rates (mg/24h) of cortisol and cortisone for ten new patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Cortisol</th>
<th>Cortisone</th>
<th>Value of quadratic form</th>
<th>Confirmed differential diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.32</td>
<td>0.18</td>
<td>0.41</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>B</td>
<td>1.12</td>
<td>0.32</td>
<td>7.59</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>C</td>
<td>1.12</td>
<td>0.48</td>
<td>4.74</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>D</td>
<td>0.48</td>
<td>0.31</td>
<td>0.41</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>E</td>
<td>0.96</td>
<td>0.32</td>
<td>5.33</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>F</td>
<td>1.04</td>
<td>0.40</td>
<td>4.69</td>
<td>Hyperplasia</td>
</tr>
<tr>
<td>G</td>
<td>11.20</td>
<td>0.75</td>
<td>59.47</td>
<td>Ectopic carcinoma</td>
</tr>
<tr>
<td>H</td>
<td>0.04</td>
<td>0.03</td>
<td>17.57</td>
<td>Normal</td>
</tr>
<tr>
<td>I</td>
<td>0.07</td>
<td>0.10</td>
<td>9.40</td>
<td>Normal</td>
</tr>
<tr>
<td>J</td>
<td>1.60</td>
<td>0.40</td>
<td>11.19</td>
<td>Adrenal carcinoma</td>
</tr>
</tbody>
</table>

Experience does provide a means of monitoring new cases diagnosed as hyperplasia to ensure that they conform reasonably with previous experience of that particular form. See §11.4 for further discussion of monitoring for atypicality in diagnostic problems.

4.5 Bayesian coverage

If we apply a predictor \( \delta \) to a number of informative experiments then the cover provided by the predictions supplied by \( \delta \) will vary from informative experiment to informative experiment. For the purpose of studying this variability it is worth making a formal definition.

**Definition 4.4**

Bayesian coverage of a predictor. The *Bayesian coverage* of a predictor \( \delta : X \to \mathbb{Y} \) is the statistic

\[
\int \delta(x) p(y|x) \, dy
\]

(4.14)

defined on the sample space \( X \) of the informative experiment. The *mean Bayesian coverage* of a predictor \( \delta \) is then

\[
\int_X p(x) dx \int \delta(x) p(y|x) \, dy,
\]

(4.15)

and can be interpreted as the long-run average Bayesian cover provided by the predictor if it is applied to a sequence of outcomes of the informative experiment.

**Example 4.3**

*Machine tool replacement.* If we use policy 3 in §3.2 we are expecting the machine to fail during the time-interval \((0, a^*)\) and so in effect are using an
78 Informative prediction

interval predictor

\[
\delta(x) = \begin{cases} 
\text{interval prediction } 0, \frac{\xi G}{\eta - \xi} - (h + x) & \text{if } x < \frac{\xi G}{\eta - \xi} - h, \\
\text{point prediction } \{0\} & \text{otherwise.}
\end{cases}
\]

Thus the Bayesian coverage of \( \delta \) is

\[
1 - \left(1 - \frac{(h + x)(\eta - \xi)}{\xi G}ight) \quad \text{if } x < \frac{\xi G}{\eta - \xi} - h,
\]

\[
0 \quad \text{otherwise.}
\]

(4.16)

![Bayesian coverage graph](image)

Fig. 4.4 Relationship of Bayesian coverage (4.16) to \( x \) for the machine tool problem.

The relationship of Bayesian coverage to \( x \) is shown in fig. 4.4. Here

\[
p(x) = \ln Be(n, g, h)
\]

and the mean coverage is
Informative prediction

\[ \int_0^{\xi G} \left( 1 - \left( \frac{\eta - \xi}{\xi G} \right)^G (h + x)^G \right) \rho(x) dx \]

\[ = I_{1 - \rho(n, g)} - \frac{\rho^n (1 - \rho)^n}{n B(n, g)}, \quad (4.17) \]

where

\[ \rho = \frac{h(\eta - \xi)}{\xi (g + n)}. \quad (4.18) \]

4.6 Statistical tolerance regions: role of coverage distributions

Much research effort has been devoted to methods of constructing prediction regions under the restriction that no prior distribution on \( \theta \) is available. The resulting tolerance regions are made to satisfy certain probabilistic statements in much the same way that confidence intervals for an unknown parameter value are constructed. Unfortunately the more complicated predictive situation requires a more sophisticated probabilistic statement and this extra complexity is often a deterrent to potential users. We shall therefore attempt in this section to give a relatively fresh account of the principles in terms of the unifying concept of the coverage distribution. Also an adequate notation is required to make clear the meaning of the probabilistic statements that form the basis of statistical tolerance regions. We continue to label the informative experiment by \( e \) and the future experiment by \( f \) but now place additional identifiers on the informative and future experiment density functions \( \rho_e(x|\theta) \) and \( \rho_f(y|\theta) \) and the corresponding probability measures \( P_e(\cdot|\theta) \) and \( P_f(\cdot|\theta) \) over \( X \) and \( Y \). We shall also require the product density function

\[ \rho_{ef}(x, y|\theta) = \rho_e(x|\theta) \rho_f(y|\theta) \quad (4.19) \]

associated with the combined experiment \( (e, f) \) and the corresponding probability measure \( P_{ef}(\cdot|\theta) \) over \( X \times Y \). To emphasise that we are not now dealing with predictions or prediction regions based on predictive distributions we shall use the established term tolerance region or tolerance prediction in what follows.

Cover of a tolerance region. A basic concept in all work on tolerance regions is that of the cover provided by such a region \( a \subset Y \). We define this to be \( P_f(a|\theta) \), where \( P_f(\cdot|\theta) \) denotes the probability measure associated with the future experiment \( f \) for which a prediction region is required. The relative frequency interpretation of cover is the following: if we use \( a \) as a tolerance region for a long run of independent repetitions of the future experiment then \( a \) will successfully cover or contain a proportion \( P_f(a|\theta) \) of the outcomes from these repetitions. Note that the cover depends not only on the region \( a \) but also on the index \( \theta \). We take account of this in our formal definition.
Informative prediction

Definition 4.5

Cover of a tolerance region. The cover of a tolerance region $a$ with respect to $\theta$ (or, for the sake of brevity, at $\theta$) is

$$P_f(a|\theta). \quad (4.20)$$

Example 4.4

Exponential case. Suppose that

$$P_e(-\theta x, y > 0), \quad P_f(y|x) = \theta \exp(-\theta y)(y > 0). \quad (4.21)$$

Consider the tolerance interval $\delta(x) = (q_1 x, q_2 x)$, where $0 < q_1 < q_2$. Then the cover provided by this tolerance interval is

$$P_f \{q_1 x, q_2 x|\theta\} = \int_{q_1 x}^{q_2 x} \theta \exp(-\theta y)dy \quad (4.22)$$

Coverage and coverage distribution of a tolerance predictor. In the preceding example a repetition of the informative experiment would almost certainly result in a different outcome $x$, and consequently a different tolerance interval $\delta(x)$ with different cover $P_f \{\delta(x)|\theta\}$. This inherent variability in $x$ and the variability it induces in $P_f \{\delta(x)|\theta\}$ play a central role in the definition and construction of statistical tolerance regions. As a crucial first step towards the construction of such regions we therefore introduce the following definition.

Definition 4.6

Coverage and coverage distribution. For a predictor $\delta$ the coverage at $\theta$ is the random variable

$$P_f \{\delta(x)|\theta\} \quad (4.23)$$

and its distribution is termed the coverage distribution.

A predictor has a coverage distribution corresponding to each index $\theta$. Since we do not know the true value of $\theta$ and are currently assuming that no prior distribution on $\Theta$ is available all we can hope to provide is a predictor $\delta$ whose coverage distributions all satisfy some desirable property. Since any coverage distribution is a distribution over the interval $[0, 1]$ we would ideally want all the coverage distributions to be concentrated on $I$ (fig. 4.5a). Since this is clearly unattainable what we seek is some requirement on the coverage distributions which forces them to approximate to this ideal. These requirements are of two main types.

(1) Ensure that, for every $\theta \in \Theta$, the mean of the coverage distribution is
reasonably high, say at least \( c \) (fig. 4.5b). This criterion produces what is commonly termed a \textit{mean coverage tolerance region}.

(2) Ensure that, for every \( \theta \in \Theta \), the bulk of the coverage distribution is above some specified cover value, \( c \) say. To guarantee that at least a proportion \( g \) of the coverage distribution lies above \( c \) for every possible \( \theta \) is
Informative prediction

essentially to place a restriction on a quantile rather than the mean. We re-
quire (fig. 4.5c) that the \((1 - q)\)-quantile lies at or to the right of \(c\) for every
\(\theta\). Such a requirement leads to what may be termed a quantile-guaranteed or,
more briefly, a guaranteed coverage tolerance region.

Example 4.5

Two simple coverage distributions. To illustrate the nature of coverage distri-
butions we study two extremely simple cases.

(i) Exponential distribution. Suppose that

\[
p_e(x|\theta) = \theta \exp(-\theta x)(x > 0), \quad p_f(y|\theta) = \theta \exp(-\theta y)(y > 0),
\]

and that we consider the predictor \(\delta\) defined by

\[
\delta(x) = (qx, \infty),
\]

where \(q > 0\). The coverage is then

\[
P_f(\delta(x)|\theta) = \int_{\theta x}^{\infty} \theta \exp(-\theta y)dy = \exp(-\theta qx),
\]

and to find the coverage distribution we have to find the distribution of
\(z = \exp(-\theta qx)\), where \(x\) is exponentially distributed. The density function of
the coverage is therefore

\[
(1/q)z^{1/(1-q)} \quad (0 < z < 1).
\]

Fig. 4.6 shows the graph of this density function. Since

\[
E(z) = 1/(q + 1)
\]
we can attain the mean coverage requirement of (1) by setting
\[ q = (1 - c)/c. \]  
(4.28)

Also since
\[ \int_c^1 \frac{1}{q} z^{1-q} \, dz = 1 - c^{1/q} \]
we obtain the quantile property of (2) by setting
\[ q = \frac{\log c}{\log (1 - g)}. \]  
(4.29)

For this predictor the coverage distribution does not depend on \( \theta \). This will not always be the case. For instance if we consider the predictor \( \delta \) defined by
\[ \delta(x) = (q + x, \infty), \]  
(4.30)
the density function of the coverage is given by
\[ p(z) = \begin{cases} \exp (\theta q) & (0 < z < \exp (-\theta q)), \\ 0 & (\exp (-\theta q) < z < 1), \end{cases} \]  
(4.31)
which depends crucially on \( \theta \).

We have thus within this simple example all the ingredients which are present in the more complex situations dealt with later in chapters 5 and 6.

(ii) Binomial trial. A second exceedingly simple illustrative example is specified by the informative and future experiment density functions:
\[ p_e(0|0) = \theta, \quad p_e(1|0) = 1 - \theta, \]
\[ p_f(0|\theta) = \theta, \quad p_f(1|\theta) = 1 - \theta; \]  
(4.32)
that is, both experiments are simple binomial trials. Let us consider the predictor \( \delta \) specified by
\[ \delta(0) = \{0\}, \quad \delta(1) = \{1\}, \]
and expressing the view that what has happened in the informative experiment will happen in the future experiment. The coverage statistic then takes two possible values
\[ p_f(\delta(0)|\theta) = p_f(0|\theta) = \theta, \]
\[ p_f(\delta(1)|\theta) = p_f(1|\theta) = 1 - \theta, \]  
(4.33)
and since these values are taken with (informative experiment) probabilities \( \theta \) and \( 1 - \theta \), we see that the coverage \( z \) has density function \( p(z) \) specified by
\[ p(\theta) = \theta, \quad p(1 - \theta) = 1 - \theta. \]  
(4.34)
Informative prediction

History

The concept of Bayesian cover is used by Aitchison and Sculthorpe (1965). Regions of previous experience for the multinormal case are simply ellipsoids of concentration, and so have a long ancestry. For their application to a medical situation see, for example, Ferriss et al (1970). The assessment of the probability (4.12) of obtaining a new case outside previous experience through the use of the predictive distribution is reported by Aitchison and Kay (1975).

Problems

4.1 Given that

\[ p(\theta) \text{ is } \text{Ga}(g, h), \]
\[ p(x|\theta) \text{ is } \text{Ga}(n, \theta), \]
\[ p(y|\theta) \text{ is } \text{Ex}(\theta), \]

find \( p(y|x) \), \( E(y|x) \) and \( V(y|x) \).

What are the Bayesian covers of the following prediction intervals?

(i) The optimum prediction interval of the form \( (0, a) \) with respect to the utility function

\[ U(a, y) = \begin{cases} -\xi(a - y) & (y < a), \\ -\eta(y - a) & (y \geq a). \end{cases} \]

(ii) The prediction interval \( \{k(h + x), (h + x)/k\} \), where \( k < 1 \). Show that, if \( n > 2 \), the Bayesian cover lies between \( (1 - k)/(1 + k)^{n-1} \) and \( 1/(1 + k)^n \).

(iii) The prediction interval \( \{0, E(y|x)\} \). Show that as \( n \to \infty \) the Bayesian cover tends to \( 1 - e^{-1} \). Can you suggest an alternative explanation of this simple result?

4.2 Complete problem 1.1 interpreting 'range of normality' as

(i) the most plausible Bayesian prediction interval of cover 0.95,
(ii) the region of previous experience,

in each case assuming that the biparietal diameters are normally distributed and making use of an appropriate vague prior distribution.

4.3 For the predictive distributions associated with problems 2.1-2.4 obtain most plausible Bayesian prediction intervals of cover \( k \).

4.4 For problem 2.9 find most plausible Bayesian prediction intervals of cover 0.99 for
Informative prediction

(i) the mean of the 15 further bearings,
(ii) the standard deviation of the 15 further bearings.

Obtain the interval of diameter experience associated with the sample of 10 ball bearings. What is the probability that the diameter of another ball bearing from the batch will be outside previous experience?

4.5 For problem 2.10 obtain a most plausible Bayesian prediction of cover 0.90 for the lifetime of the eleventh component.

4.6 For problem 2.12 obtain a most plausible Bayesian prediction of cover 0.95 for the total yield of the eight tomato plants all grown in compost with additive at strength 5.5.

4.7 The times (hours) to first breakdown of 20 machines were measured in an informative experiment and were as follows.

<table>
<thead>
<tr>
<th>Time (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4  10  62  119  74  24  13  29  19  18  57  23  47  409  19  208  13  209  46  188</td>
</tr>
</tbody>
</table>

We assume that these are independent observations on Ex(θ) random variables where little prior information on θ is available. Five similar machines are to be used simultaneously (and independently) in a laboratory and we wish to give prediction intervals for

(i) the time to the first failure, \( y_{11} \),
(ii) the time to the last failure, \( y_{15} \),
(iii) the total running time of the machines, \( \sum_i y_i \).

Derive the predictive density functions in each case and hence obtain Bayesian predictive intervals with cover 0.95.

4.8 Find the two-dimensional regions of previous experience for
(a) the adenoma patients,
(b) the bilateral hyperplasia patients,
associated with the K and CO2 measurements only of table 1.6.

What are the probabilities
(i) that a new adenoma patient will fall within previous adenoma experience;
(ii) that a new bilateral hyperplasia patient will fall within previous bilateral hyperplasia experience?

How many bilateral hyperplasia patients fall within previous adenoma experience, and how many adenoma patients fall within previous bilateral hyperplasia experience?

4.9 The urinary excretion rates (mg/24h) of three steroid metabolites A, B and C are shown in the following table for 12 normal healthy adults. Construct
Informative prediction

A region of previous experience for this set of data. Assess the probability that another normal healthy adult will have metabolite results outside this region of previous experience.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>3.0</td>
<td>4.2</td>
<td>1.4</td>
</tr>
<tr>
<td>5.7</td>
<td>8.0</td>
<td>3.2</td>
</tr>
<tr>
<td>1.6</td>
<td>7.7</td>
<td>3.4</td>
</tr>
<tr>
<td>4.0</td>
<td>4.7</td>
<td>2.4</td>
</tr>
<tr>
<td>4.3</td>
<td>5.5</td>
<td>4.5</td>
</tr>
<tr>
<td>3.3</td>
<td>3.5</td>
<td>6.8</td>
</tr>
<tr>
<td>6.1</td>
<td>2.8</td>
<td>4.8</td>
</tr>
<tr>
<td>3.2</td>
<td>2.8</td>
<td>3.4</td>
</tr>
<tr>
<td>3.5</td>
<td>5.7</td>
<td>4.0</td>
</tr>
<tr>
<td>2.2</td>
<td>8.7</td>
<td>5.3</td>
</tr>
<tr>
<td>4.9</td>
<td>9.4</td>
<td>3.7</td>
</tr>
<tr>
<td>2.6</td>
<td>5.5</td>
<td>4.1</td>
</tr>
</tbody>
</table>

For the following three new cases first attempt to assign an index of abnormality intuitively, and then compute such an index.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>6.0</td>
<td>5.1</td>
<td>7.0</td>
</tr>
<tr>
<td>4.0</td>
<td>8.2</td>
<td>3.8</td>
</tr>
<tr>
<td>5.5</td>
<td>4.7</td>
<td>6.5</td>
</tr>
</tbody>
</table>

4.10 Let \( \delta(x) \) denote a most plausible Bayesian prediction of cover \( \kappa \) based on data \( x \) from an informative experiment. What is the coverage distribution of the predictor \( \delta \) and what is the Bayesian coverage?

4.11 Find the density function of the coverage of the predictor \( \delta \) defined by

\[
\delta(x) = (qx, \infty),
\]

where \( q > 0 \), for the prediction problem:

\[
p_x(x|\theta) = Ga(k, \theta) \quad p(y|\theta) = Ex(\theta).
\]

Is it possible to obtain mean coverage and guaranteed coverage tolerance predictors for this case?

4.12 An attempt has been made to design 'self-destructive' components in such a way that each has a minimum lifetime \( \mu \) and that, subsequent to that minimum period, each component follows a failure pattern which is a Poisson process of high intensity \( \tau \). For any production run control has not yet reached the stage where \( \mu \) and \( \tau \) can be predetermined, but study of a number of trial production runs suggests that the joint variability in \( \mu \) and \( \tau \) follows an
Informative prediction

EIGa(b, c, g, h) distribution. What is now proposed is that from the large number of components of each production run a random sample of $n$ components should be observed and the lifetimes $x_1, \ldots, x_n$ determined. Then for each production run it is proposed to quote an interval within which 99 per cent of lifetimes for that run will lie. It is understood that for any one production run this may be misleading but it is hoped that over a number of batches this policy of quoting intervals will be effective.

Advise the management on how to construct such intervals. How would your answer be affected if one and only one component (say the $n$th) was still functioning at time $t$ when an interval had to be quoted?

For an entirely new trial production process about whose characteristics $\mu$ and $\tau$ little is known the lifetimes (days) of a random sample of 10 components from the large production run are given below. What proportion of the other components do you assess have lifetimes within the same range?


4.13 In a certain factory each production run produces a large number of items. For each of $k$ such production runs some items have been tested and found to have the following characteristics (measurements of quality).

<table>
<thead>
<tr>
<th>Production run</th>
<th>Measurements of quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$w_{11}, w_{12}, \ldots, w_{1n}$</td>
</tr>
<tr>
<td>2</td>
<td>$w_{21}, w_{22}, \ldots, w_{1n}$</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>$k$</td>
<td>$w_{k1}, w_{k2}, \ldots, w_{kn}$</td>
</tr>
</tbody>
</table>

The possibility that the distribution of the characteristic differs with production run has now been realised. It is therefore proposed to test $n$ items from each production run and to make some statement about the distribution of characteristics associated with that production run. For a production run with test components giving characteristic values $x_1, x_2, \ldots, x_n$ what can be said about the distribution of the characteristic in the large number of untested items from that run?
5

Mean coverage tolerance prediction

5.1 Introduction

The motivation underlying the search for mean coverage tolerance predictors is an attempt to ensure that the average cover provided by the predictor does not fall too low. Since the mean of the coverage distribution of a predictor depends in general on $\theta$ this aim can be secured only by constructing the predictor in such a way that none of these possible means fall below some pre-assigned cover, say $c$. We can express this idea formally in terms of the following definition.

**Definition 5.1**

Mean coverage tolerance predictor. A predictor $\hat{\delta}$ is a tolerance predictor of mean coverage $c$ if

$$\inf_{\theta} E[P_T(\delta(x)|\theta)] = c,$$

where

$$E[P_T(\delta(x)|\theta)] = \int_{X} P_T(\delta(x)|\theta) p_e(x|\theta) \, dx.$$  (5.2)

For the case of discrete distributions it is necessary to relax the definition, the equality sign being replaced by $\geq$.

Also the concept of similarity may apply.

**Definition 5.2**

Similar mean coverage tolerance predictor. A predictor $\hat{\delta}$ is a tolerance predictor of similar mean coverage $c$ if

$$E[P_T(\delta(x)|\theta)] = c \text{ for every } \theta \in \Theta.$$  (5.3)

We have extended the usual definition of mean coverage tolerance predictors, the similar form of definition 5.2, to the more general form of definition 5.1. While for some standard situations such as the normal and gamma cases similar mean coverage tolerance predictors can be found there are other situations, for example the binomial case, where no such similar mean coverage tolerance predictors may exist. It is for this reason that we
Mean coverage tolerance prediction

introduce the less demanding requirement (5.1) of definition 5.1. The distinction between the two definitions is analogous to the concepts of size and exact or similar size of critical regions in the classical theory of hypothesis testing.

Example 5.1

Two simple mean coverage tolerance predictors

(i) Exponential distribution. We have already seen in example 4.5(i) a case of a similar mean coverage tolerance predictor. Indeed there are circumstances in which more than one such predictor may exist.

\[
\frac{c}{1-c}
\]

Fig. 5.1 Locus of \((q_1, q_2)\) for which \(\frac{1}{q_1 + 1} - \frac{1}{q_2 + 1} = c\) for given \(c\).
Consider the predictor
\[ \delta(x) = (q_1 x, q_2 x) \]
discussed in example 4.4. From our previous work we easily calculate the integral in (5.2) as
\[
\int_0^\infty \left\{ \exp(-\theta q_1 x) - \exp(-\theta q_2 x) \right\} \theta \exp(-\theta x) dx = \frac{1}{q_1 + 1} - \frac{1}{q_2 + 1}.
\]
Since this is independent of \( \theta \), to satisfy (5.3) we need only select values of \( q_1 \) and \( q_2 \) such that \( q_2 > q_1 \) and
\[
\frac{1}{q_1 + 1} - \frac{1}{q_2 + 1} = c. \tag{5.4}
\]
It is easily seen from fig. 5.1 that this is possible for any value \( c \) in \( 0 < c < 1 \) and indeed that there is an embarrassingly large number of such \((q_1, q_2)\) combinations.

Table 5.1 The eight possible predictors \( \delta_i \) for the simple binomial example

<table>
<thead>
<tr>
<th>( i )</th>
<th>( \delta_i(0) )</th>
<th>( \delta_i(1) )</th>
<th>Mean coverage at ( \theta )</th>
<th>Infimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{0}</td>
<td>{0, 1}</td>
<td>( \theta^2 + (1 - \theta) )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>2</td>
<td>{0, 1}</td>
<td>{1}</td>
<td>( \theta + (1 - \theta)^2 )</td>
<td>( \frac{3}{4} )</td>
</tr>
<tr>
<td>3</td>
<td>{0}</td>
<td>{1}</td>
<td>( \theta^2 + (1 - \theta)^2 )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>4</td>
<td>{0}</td>
<td>{0}</td>
<td>( \theta )</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>{1}</td>
<td>{1}</td>
<td>( 1 - \theta )</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>{1}</td>
<td>{0}</td>
<td>( 2(1 - \theta) )</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>{0, 1}</td>
<td>{0}</td>
<td>( \theta (2 - \theta) )</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>{1}</td>
<td>{0, 1}</td>
<td>( 1 - \theta^2 )</td>
<td>0</td>
</tr>
</tbody>
</table>

(ii) Binomial trials. We now develop example 4.5(ii). This is such a simple case that we can enumerate all the possible predictors (table 5.1). We can then calculate the mean coverage for each of these eight predictors and these are also shown in table 5.1. We see immediately that the last five predictors are useless as mean coverage tolerance predictors since the infimum (5.1) associated with them is zero. The first three predictors have some possible use. We see that for \( c < \frac{1}{2} \) we can obtain a tolerance predictor, for example \( \delta_1 \), of mean coverage \( c \), but that for \( c > \frac{1}{2} \) no tolerance predictor of mean coverage \( c \) exists. Notice that we have omitted a decision rule of the form
\[
\delta_9(0) = \{0, 1\}, \delta_9(1) = \{0, 1\}.
\]
Although this would give mean coverage of 1 for all \( \theta \), it has no practical relevance and should not be considered as a tolerance predictor.

For this binomial trials situation no similar mean coverage tolerance predictor exists.
Mean coverage tolerance prediction

5.2 Interpretation of similar mean coverage tolerance predictors

Suppose that we wish to apply a predictor \( \delta \) repeatedly on a number \( r \) of occasions on which nature may present us with different unknown indices according to an unknown density function \( p(\theta) \). Then in the \( r \) applications of the predictor we may obtain \( \theta_1, \ldots, \theta_r \), and information \( x_1, \ldots, x_r \), in the \( r \) associated informative experiments. We would then use regions \( \delta(x_1), \ldots, \delta(x_r) \), and the associated actual covers obtained would be \( P_r[\delta(x_1) | \theta_1], \ldots, P_r[\delta(x_r) | \theta_r] \). One criterion sometimes suggested is that the long-run average cover should attain some specified level \( c \), usually not too far below 1: for large \( r \)

\[
P_r[\delta(x_1) | \theta_1] + \ldots + P_r[\delta(x_r) | \theta_r] = c.
\]

The counterpart of long-run average in a statistical model is the appropriate expectation, in this case with respect to the joint distribution of \( x \) and \( \theta \), namely \( p(\theta)p_e(x | \theta) \). Thus we would seek a predictor \( \delta \) satisfying

\[
E[P_r[\delta(\cdot) | \theta]] = c \quad \text{(5.6)}
\]

or

\[
\int \int p(\theta)p_e(x | \theta) p(\theta) dx d\theta = c. \quad \text{(5.7)}
\]

We can express this double accumulation (integral or summation) as a repeated accumulation under very wide conditions:

\[
\int_\Theta p(\theta) \left( \int_x P_r[\delta(x) | \theta] p_e(x | \theta) dx \right) d\theta = c. \quad \text{(5.8)}
\]

If we can obtain a predictor which sets the inner accumulation equal to \( c \) for every \( \theta \), then the overall double accumulation will also be equal to \( c \) and condition (5.3) will be satisfied, whatever the nature of \( p(\theta) \). This is the basis of definition 5.2 of a tolerance predictor of similar mean coverage \( c \). Such a predictor has the property that if applied to a long run of similar informative experiments it provides long-run average cover \( c \) over the set of associated future experiments to which it is applied. Note that what is being said is not the following: if there is a single informative experiment with outcome \( x \) then the average cover provided by \( \delta(x) \) over a long sequence of future experiments is \( c \). The cover is in fact \( P_r[\delta(x) | \theta] \) for each of these future experiments.

Some such principle as that outlined above concerning what happens in a long run of repetitions of \((e, f)\) has to be invoked to make use of mean coverage tolerance predictors.

There are other possible interpretations of the relation (5.3). For we may develop the integral in the following way:

290
Mean coverage tolerance prediction

\[
\int_X P_f(\delta(x) | \theta) P_e(x | \theta) \, dx = \int_X \left\{ \int_{\delta(x)} P_f(y | \theta) \, dy \right\} P_e(x | \theta) \, dx
\]

\[
= P_{ef}(x, y : y \in \delta(x) | \theta). 
\] (5.9)

Suppose that we can devise \( \delta \) in such a way that this is equal to \( c \) for every \( \theta \in \Theta \). Then in a long run of informative experiments each followed by a single prediction using the predictor \( \delta \) the proportion of successful predictions (where the actual \( y \) observed falls in the corresponding prediction region \( \delta(x) \) in the long run is \( c \). There is an obvious extension to the form of (5.1).

5.3 Additional criteria

We have seen in example 5.1(i) that a large number of tolerance predictors of mean coverage \( c \) may exist. The nature of the practical situation may, however, dictate just which of this class of possible tolerance predictors is a sensible one to use. For example the underlying density functions may describe the lifetime of a component from some production process. The manufacturer may wish to quote some minimum lifetime which he hopes will convey to the purchasers the quality of his product. To do this he will quote only prediction intervals of the form \((q_1 x, \infty)\). In our analysis above for the exponential distribution this would involve setting \( q_2 = \infty \), and then there is a unique \( q_1 = (1 - c) / c \) leading to a given mean coverage \( c \).

If the practical dictates of the situation do not lead to a unique predictor then some new principle has to be introduced to arrive at a satisfactory result. This may take the form of a principle of symmetry, the region having as its centre the most likely future observation or the mean future observation. We shall see how this principle is used in normal theory in §5.8. Some other principles depend on the notion that a large prediction region is less good than a smaller one. For instance, in example 5.1(i), the length of the interval is \((q_2 - q_1) x\) so that the tolerance predictor of mean coverage \( c \) of the minimum length will minimise \( q_2 - q_1 \) subject to the condition (5.1). Since in the \((q_1, q_2)\) plane of fig. 5.1 lines of the form

\[
q_2 - q_1 = \text{constant}
\]

each have unit gradient and since the gradient of the curve (5.4) everywhere exceeds 1 for \( q_2 \geq q_1 \geq 0 \) it is clear that the minimum occurs with \( q_1 = 0 \), \( q_2 = c / (1 - c) \).

One important principle of exclusion not yet mentioned is that of invariance.

In example 5.1(i) the lifetime of a component may be the variable under question. Suppose that the predictor \( \delta \) specifies an interval \([\delta_1(x), \delta_2(x)]\).

Since the prediction arrived at must be the same whether time is measured in
Mean coverage tolerance prediction

Mean coverage tolerance prediction

**Mean coverage tolerance prediction**

minutes, microseconds or years we clearly require that

\[ \delta_1(\lambda x) = \lambda \delta_1(x), \delta_2(\lambda x) = \lambda \delta_2(x) \quad (\lambda > 0); \quad (5.10) \]

the prediction procedure should be invariant under the group of transformations \( x \rightarrow \lambda x \) (the group of changes of scale). Notice that the class of density functions already possesses the necessary property: if \( x \) is \( \text{Ex}(\theta) \) then \( \lambda x \) is \( \text{Ex}(\theta/\lambda) \).

What limitation to the form of \( \delta \) do the relations (5.10) imply? The answer to this question is easily seen since (5.10) are simply the defining relations of proportional functions \( \delta_1(x) = q_1 x, \delta_2(x) = q_2 x \). We thus see that the restriction to invariant predictors in this case requires that we take the proportional function already considered in our analysis above. Thus it still leaves a wide class of possible tolerance predictors satisfying (5.4).

5.4 Relationship to Bayesian coverage

Mean coverage tolerance predictors have a strong Bayesian property which makes them particularly attractive. The following result provides this by pinpointing the relationship of the mean coverage of such a predictor to the concept of Bayesian coverage.

A tolerance predictor \( \delta \) of similar mean coverage \( c \) has mean Bayesian coverage \( c \) whatever the prior density function \( p(\theta) \) on \( \Theta \).

\[ \text{(5.11)} \]

The proof is as follows. Since \( \delta \) has similar mean-coverage \( c \) we have, by (5.3),

\[ \int_X P_T[\delta(x) \mid \theta] p_e(x \mid \theta) \, dx = c \quad \text{for every} \quad \theta \in \Theta, \]

so that

\[ \int_X \int_{\Theta} P_T(y \mid \theta) p_e(x \mid \theta) \, dx \, dy = c \quad \text{for every} \quad \theta \in \Theta. \]

Hence, for every prior density function \( p(\theta) \) on \( \Theta \),

\[ \int_\Theta \int_X \int_{\Theta} P_T(y \mid \theta) p_e(x \mid \theta) p(\theta) \, dx \, dy \, d\theta = c. \quad (5.12) \]

Since, by (2.2),

\[ p_e(x \mid \theta) p(\theta) = p(x \mid \theta) \]

we may rewrite (5.12) as

292
Mean coverage tolerance prediction

\[
\int_{X} p(x) \int_{\delta(x)} \left( \int_{\Theta} p(y|\theta) p(\theta|x) d\theta \right) dx dy = c
\]

and so, by (2.4),

\[
\int_{X} p(x) \int_{\delta(x)} p(y|x) dy dx = c.
\]  

(5.13)

The left side of (5.13) is the mean Bayesian coverage as defined in (4.15), and the result is therefore established.

The condition of similarity can be dropped from the statement (5.11) with a minor modification:

\[A\text{ tolerance predictor } \delta \text{ of mean coverage } c \text{ has mean Bayesian coverage at least } c \text{ whatever the prior density function } p(\theta) \text{ on } \Theta.\]  

(5.14)

5.5 Mean coverage tolerance predictors for the binomial case

Suppose that the informative experiment \( e \) is Bi(n, \( \theta \)) and the future experiment \( f \) is Bi(N, \( \theta \)). We consider first the construction of a tolerance predictor \( \delta \) of mean coverage \( c \) and of the one-sided form

\[ \delta(x) = \{0, 1, ..., e(x)\}. \]  

(5.15)

Here the coverage (4.23) is

\[ e(x) \sum_{N - y}^{N} \binom{N}{y} \theta^{y}(1 - \theta)^{N-y} \]  

(5.16)

and has a distribution which is not independent of \( \theta \). This causes some difficulty in the construction of suitable \( e(x) \). We must choose the region of summation (fig. 5.2) of the mean coverage, which is the double sum

\[ \sum_{x=0}^{n} \sum_{y=0}^{n} \binom{N}{y} \theta^{y}(1 - \theta)^{N-y} \binom{n}{x} \theta^{x}(1 - \theta)^{n-x}, \]  

(5.17)

in such a way that (5.17) is at least \( c \) for every \( \theta \) in the interval \([0, 1]\). Now the typical term of this double sum can be rewritten as

\[ \binom{n + N}{x + y} \theta^{x+y}(1 - \theta)^{n+x+y} \binom{n}{x} \binom{N}{y} / \binom{n + N}{x + y} \]  

(5.18)

and the determination of \( e(x) \) is made easier by a change in the method of summation. Consider the transformation...
Mean coverage tolerance prediction

Fig. 5.2 Region of summation for the double sum associated with mean coverage (5.17).

\[ U = x + y, \quad V = x. \] (5.19)

The double sum can then be expressed as the repeated sum

\[
\sum_{u=0}^{n+N} \binom{n+N}{u} \theta^u (1-\theta)^{n+N-u} \sum_{v: v + \epsilon(v) > u} \binom{n}{v} \binom{N}{u-v} \binom{n+N}{u}. \] (5.20)
Mean coverage tolerance prediction

Our object will then be achieved if we choose $\epsilon(v)$ such that

$$\sum_{\{\nu: v + \epsilon(v) \geq u\}} \binom{n}{\nu} \binom{N}{N-\nu} \geq c, \quad (5.21)$$

for every non-negative integer $u$. The inequality (5.21) will be satisfied in as economic a way as possible (that is with as small tolerance intervals as possible) if we take

$$\epsilon(x) = \min \{z: \Pr_{n+N, z, n, x} < 1 - c\} - x - 1, \quad (5.22)$$

where $\Pr_{n+N, z, n, x}$ is the hypergeometric function defined in (A21) of appendix I and tabulated by Lieberman and Owen (1961).

Notice that we cannot obtain a tolerance predictor $\delta$ of similar mean coverage $c$.

By an argument analogous to those above we can obtain a tolerance predictor of mean coverage $c$ of the other one-sided form

$$\{\epsilon(x), \epsilon(x) + 1, \ldots, N\} \quad (5.23)$$

by setting

$$\epsilon(x) = \max \{z: \Pr_{n+N, z, n, x-1} \geq c\} - x + 1. \quad (5.24)$$

Tolerance prediction problems associated with the binomial family can usually be expressed in terms of one or other of the one-sided forms (5.15) and (5.23).

Example 5.2

A quality control problem. Items are produced independently in batches of 30 by a firm. Items may be either effective or defective and it is recognised by both manufacturer and customer that batches vary considerably in the number of effectives they contain. The terms of a suggested contract between manufacturer and customer require the manufacturer to test destructively 5 of the 30 components from each batch, and to supply the remaining 25 components in a packet with an accompanying statement about the maximum number of defectives in the packet. The contract further requires that the statements must be correct for at least 80 per cent of batches. What statement strategy will fulfil the terms of the contract?

For a particular batch let $\theta$ denote the probability that an item will be defective, so that we can regard $e$ and $f$ as $\text{Bi}(5, \theta)$ and $\text{Bi}(25, \theta)$ experiments. If for an outcome $x$ of $e$ we state that the number $y$ of defective items in the batch will be in the interval $\delta(x)$ then the contract requires that
Mean coverage tolerance prediction

\[ P_{ef}(x, y) : y \in \delta(x) \mid \theta \geq 0.8 \quad (5.25) \]

whatever the value of \( \theta \). This is precisely the form (5.9) so that what we require is a mean coverage tolerance predictor of form (5.15) with \( c = 0.8, n = 5, N = 25 \). Consider the construction of the tolerance prediction corresponding to \( x = 1 \). The hypergeometric function within the braces of (5.22) is then

\[ P_{hy}(30, 5, 1). \]

From the Lieberman and Owen (1961) tables we have

\[ P_{hy}(30, 15, 5, 1) = 0.1648, \quad P_{hy}(30, 14, 5, 1) = 0.2095, \]

so that, by (5.22),

\[ \epsilon(1) = 15 - 1 - 1 = 13. \]

The complete mean coverage tolerance predictor is set out in table 5.2.

| Table 5.2 Mean coverage predictor for the quality control problem
| \( \delta(x) = \{0, 1, \ldots, \epsilon(x)\} \) |
|---|---|---|---|---|---|
| \( x \) | 0   | 1   | 2   | 3   | 4   |
| \( \epsilon(x) \) | 7   | 13  | 17  | 21  | 24  |
| 5       | 25  |

5.6 Mean coverage tolerance predictors for the Poisson case

Consider the construction of a tolerance predictor \( \delta \) of the form

\[ \delta(x) = \{0, 1, \ldots, \epsilon(x)\} \quad (5.26) \]

for a future \( Po(K \theta) \) experiment \( f \), based on the information from a \( Po(k \theta) \) informative experiment \( e \) and with mean coverage \( c \). Similarly to the binomial case we have to ensure that the inequality

\[ \sum_{x=0}^{\infty} \sum_{y=0}^{\epsilon(x)} \frac{\exp(-K \theta)}{y!} \frac{\exp(-k \theta)}{x!} \frac{(K \theta)^y}{(k \theta)^x} \geq c \quad (5.27) \]

holds for every \( \theta > 0 \). Again we consider an alternative way of carrying out the double summation. The double sum may be written as

\[ \sum \sum \left( \begin{array}{c} x+y \\ x \end{array} \right) \left( \begin{array}{c} k \\ k+K \end{array} \right)^x \left( \begin{array}{c} K \\ k+K \end{array} \right)^y \frac{\exp(-(k+K) \theta)}{(k+K)\theta}^{x+y} \frac{(x+y)!}{(x+y)!} \quad (5.28) \]

where the region of summation is again as shown in fig. 5.2 with \( n \to \infty, N \to \infty \).

Again the change of variables

\[ u = x + y, \quad v = x \]
Mean coverage tolerance prediction

gives

\[ \sum_{u=0}^{\infty} \frac{\exp\left[-(k+K)\theta\right]\{(k+K)\theta\}^u}{u!} \]

\[ x \sum_{v:s+v+\epsilon(v) > u} \binom{u}{v} \left( \frac{k}{k+K} \right)^v \left( \frac{K}{k+K} \right)^{u-v} \]

(5.29)

The inequalities analogous to (5.21) then lead us to set

\[ \epsilon(x) = \min\{z: I_{k/(k+K)}(x+1, z) \geq c\} - 1, \]

(5.30)

where \( I \) is the incomplete beta function defined in (A19) of Appendix I and tabulated by Pearson (1934).

For a predictor of the form

\[ \delta(x) = \{\epsilon(x), \epsilon(x) + 1, \ldots\} \]

(5.31)

and with mean coverage \( c \) we set

\[ \epsilon(x) = \max\{z: I_{k/(k+K)}(x, z) \leq 1 - c\}. \]

(5.32)

Example 5.3

Seed germination. A rare plant bears many seeds of low germination rate. A seedsman divides each of his boxes of 12 such plants at random into two sets of 6, sows the seeds from one set for his own crops next season, records the number of germinating seeds from the set, and packets the seeds from the other set with a statement about the minimum number of seeds from the packet likely to germinate. If he hopes that 95 per cent of his packets should contain correct statements what minimum number of germinating seeds should he attribute to a packet associated with 6 plants yielding 7, 3, 5, 5, 4, 6 germinating seeds?

Let us assume that the number of germinating seeds from a plant is distributed as \( \text{Po}(\theta) \), where \( \theta \) may vary from box to box. The informative experiment \( e \) is then effectively \( \text{Po}(6\theta) \) with observation \( x = 7 + 3 + \ldots + 6 = 30 \); the future experiment \( f \), which records the number \( y \) of germinating seeds, is also \( \text{Po}(6\theta) \). To achieve 95 per cent statements we require to quote minima \( \epsilon(x) \) satisfying

\[ P_{ef}\{(x, y): y \geq \epsilon(x)\theta\} \geq 0.95 \]

(5.33)

for every \( \theta > 0 \). This is of form (5.9) and we are thus led to seek a mean coverage tolerance predictor of type (5.31) with \( k = K = 6, c = 0.95 \). The incomplete beta function within the braces of (5.32) is, for \( x = 30 \),

\[ I_{0.95}(30, z). \]
Since
\[ I_{0.5}(30, 18) = 0.0395, \quad I_{0.5}(30, 19) = 0.0557, \]
we have
\[ e(30) = 18. \]

5.7 Mean coverage tolerance predictors for the gamma case

In example 5.1(i) we saw that it is possible to construct a tolerance predictor of similar mean coverage \( c \) for the simple exponential case. This particular case can be included as a special case of the following more general formulation.

Suppose that
\[ p_x(x \mid \theta) = \text{Ga}(k, \theta), \quad p_f(y \mid \theta) = \text{Ga}(K, \theta). \]

For the reasons of invariance suggested in §5.3 it is natural to consider a predictor of the form
\[ \delta(x) = (q_1 x, q_2 x). \quad (5.34) \]

We can obtain from this finite-interval predictor one-sided predictors by considering the cases \( q_1 = 0 \) and \( q_2 = \infty \) as special forms.

The coverage of \( \delta \) is
\[ \int_{q_1 x}^{q_2 x} \frac{\theta^K y^{K-1} \exp(-\theta y)}{\Gamma(K)} dy \quad (5.35) \]

which has a distribution independent of \( \theta \), so that in mean coverage calculations we can conveniently set \( \theta = 1 \). The mean coverage is then given by (5.9) as
\[ P_{ef}\{ (x, y) : q_1 x \leq y \leq q_2 x \}, \]
where \( x, y \) are independent \( \text{Ga}(k, 1) \) and \( \text{Ga}(K, 1) \) random variables. This mean coverage can thus be expressed as
\[ P_{ef}\left\{ (x, y) : \frac{1}{1 + q_2} \leq \frac{x}{x + y} \leq \frac{1}{1 + q_1} \right\}. \quad (5.36) \]

Since \( x/(x + y) \) has a \( \text{Be}(k, K) \) distribution, we see that we obtain mean coverage \( c \) if we choose \( c_1, c_2 \) such that \( 0 \leq c_2 \leq c_1 \leq 1 \) and \( c_1 - c_2 = c \) and then set, in the quantile notation of appendix I,
\[ \frac{1}{1 + q_1} = \text{Be}(k, K; c_1), \quad (5.37) \]
\[ \frac{1}{1 + q_2} = \text{Be}(k, K; c_1); \quad \text{(5.38)} \]

see fig. 5.3. Thus we obtain the following general result for this case. The tolerance predictor

\[ \delta (x) = \begin{cases} \frac{1 - \text{Be}(k, K; c_1)}{\text{Be}(k, K; c_1)} x, & \text{Be}(k, K; c_2) \end{cases} \quad \text{(5.39)} \]

has similar mean coverage \( c \).

The following semi-infinite interval predictors are derived as special cases. The tolerance predictors

\[ \delta (x) = \begin{cases} 0, & 1 - \text{Be}(k, K; 1 - c) \end{cases} \quad \text{Be}(k, K; 1 - c) \quad \text{x} \quad \text{(5.40)} \]

and

\[ \delta (x) = \begin{cases} 1 - \text{Be}(k, K; c) \end{cases} \quad \text{Be}(k, K; c) \quad \text{x, } \infty \quad \text{(5.41)} \]

each have similar mean coverage \( c \).
Mean coverage tolerance prediction

5.8 Mean coverage tolerance predictors for the normal case

We suppose throughout this section that

\[ p_f(Y | \theta) = \text{No}(\mu, K\tau) \]

and that the informative experiment \( e \) provides \((m, v)\), where \( m \) and \( v \) are independent with \( \text{No}(\mu, K\tau) \) and \( \text{Ch}(v, \tau) \) distributions.

For example we may take

\[ m = \bar{x} = \frac{1}{n} \sum x_i, \quad v = S(x, \bar{x}), \]

\[ k = n, \quad v = n - 1, \]

for the simple normal prediction problem where \( e \) consists of \( n \) replicates of a \( \text{No}(\mu, \tau) \) experiment. We find it more convenient to use \( v \) than the conventional standard deviation statistic \( s \) given by \( s = (\bar{v}/\nu)^{1/2} \).

One-sided tolerance predictors. We consider here a predictor \( \delta \) defined by

\[ \delta (m, v) = (-\infty, m + qv^{1/2}), \tag{5.42} \]

where \( q \) is a constant to be determined. The choice of the form \( m + qv^{1/2} \) is based on an invariance property, that the predictor should be invariant under the group of transformations \( m \rightarrow m + g, \quad v \rightarrow hv (h > 0) \). The coverage of \( \delta \) is

\[ \Phi \{ (K\tau)^{1/2} (m + qv^{1/2} - \mu) \} \tag{5.43} \]

and has a distribution which is clearly independent of \( \theta = (\mu, \tau) \). In computing the mean coverage we can thus simplify matters by setting \( (\mu, \tau) = (0, 1) \). If therefore we use the abbreviated notation,

\[ p(m) = \text{No}(0, k), \quad p(v) = \text{Ch}(v, 1), \quad p(\nu) = \text{No}(0, K), \tag{5.44} \]

we can express the mean coverage as

\[ \int_{-\infty}^{\infty} \int_{0}^{\infty} \Phi \{ K^{1/2} (m + qv^{1/2}) \} p(m) p(\nu) dm d\nu \]

\[ = P_{ef} \{ (m, v, y) : y \leq m + qv^{1/2} \} \]

\[ = P_{ef} \left\{ (m, v, y) : (y - m) \left/ \left( \frac{v}{\nu} \left( \frac{1}{k} + \frac{1}{K} \right) \right) \right. \right\}^{1/2} \leq q \left/ \left( \frac{v}{\nu} \left( \frac{1}{k} + \frac{1}{K} \right) \right) \right. \right\}^{1/2} \]

Now

\[ (y - m) \left/ \left( \frac{v}{\nu} \left( \frac{1}{k} + \frac{1}{K} \right) \right) \right. \right\}^{1/2} \]
Mean coverage tolerance prediction

is distributed as \( t(\nu) \) and so to achieve mean coverage \( c \) (for every \( \theta \)) we may set

\[
q \sqrt{\left(1 + \frac{1}{k + \frac{1}{K}}\right)}^{1/2} = t(\nu; c).
\]

Thus we have the following result.

The tolerance predictor

\[
\delta(m, \nu) = \left[-\infty, m + \left(1 + \frac{1}{k + \frac{1}{K}}\right)^{1/2} t(\nu; c)\nu^{1/2}\right]
\]

(5.46)

has similar mean coverage \( c \).

In exactly the same manner it can be shown that

\[
\delta(m, \nu) = \left[m - \left(1 + \frac{1}{k + \frac{1}{K}}\right)^{1/2} t(\nu; c)\nu^{1/2}, \infty\right]
\]

(5.47)

has similar mean coverage \( c \).

Symmetric two-sided tolerance predictors. The prediction intervals obtained in (5.46) and (5.47) are semi-infinite, and this may not suit the particular practical problem. Where it is desirable to quote a finite interval a popular tolerance predictor is one that is symmetric about \( m \), namely of the form

\[
\delta(m, \nu) = (m - \nu^{1/2} m + \nu^{1/2})
\]

(5.48)

The coverage,

\[
\Phi\{K^{1/2}(m + \nu^{1/2} - \mu)\} - \Phi\{K^{1/2}(m - \nu^{1/2} - \mu)\}
\]

(5.49)

has again a distribution which is independent of \((\mu, \tau)\) and the same device of considering the special case \((\mu, \tau) = (0, 1)\) may be used in evaluating the mean coverage:

\[
\int_{-\infty}^{\infty} \int_{0}^{\infty} \Phi\{K^{1/2}(m + \nu^{1/2})\} - \Phi\{K^{1/2}(m - \nu^{1/2})\} p(m)p(\nu)dmd\nu
\]

(5.50)

Again the fact that \((y - m)/\left(1 + 1/k + 1/K\right)^{1/2}\) is distributed as \( t(\nu) \) gives an immediate determination of

\[
q = \left(1 + \frac{1}{k + \frac{1}{K}}\right)^{1/2} t(\nu; \frac{1}{2}(c + 1))
\]
Mean coverage tolerance prediction

yielding the following result.

The tolerance predictor

$$\left[ m - \frac{1}{\nu} \left( \frac{1}{\kappa} + \frac{1}{K} \right) \right]^{1/2} t \{ \nu; \frac{1}{2} (1 + c) \nu^{1/2}, m + \left[ \frac{1}{\nu \kappa} \left( \frac{1}{\kappa} + \frac{1}{K} \right) \right]^{1/2} t \{ \nu; \frac{1}{2} (1 + c) \nu^{1/2} \}$$

(5.51)

has similar mean coverage c.

Non-symmetric two-sided tolerance predictors. It is clear from the construction of (5.51) that it is easy to construct a tolerance predictor of mean coverage c of the form

$$\delta(m, \nu) = (m + q_1 \nu^{1/2}, m + q_2 \nu^{1/2})$$

(5.52)

where q_1 is not necessarily equal to -q_2. This is achieved by setting

$$q_1 = \left( \nu \kappa \left( \frac{1}{\kappa} + \frac{1}{K} \right) \right)^{1/2} t(\nu; c_1)$$

(5.53)

$$q_2 = \left( \nu \kappa \left( \frac{1}{\kappa} + \frac{1}{K} \right) \right)^{1/2} t(\nu; c_2)$$

(5.54)

where 0 < c_1 < c_2 < 1 and c_2 - c_1 = c. The symmetric case is given by c_2 = \frac{1}{2} (1 + c), c_1 = \frac{1}{2} (1 - c), but there are many other (c_1, c_2) combinations.

The length of the interval provided by \delta depends on the magnitude of q_2 - q_1 and it can be shown that this is a minimum for fixed c when q_1 = q_2, the symmetric case.

Example 5.4

Crop prediction. A potato farmer wishes to gauge the weight of crop he is likely to produce from the potatoes he has planted in separate fields so that he may provide relevant information to prospective buyers. A few days before complete harvesting he selects at random 24 square metre units in a field and measures the weight of crop produced in each unit. He then wishes to provide a 95 per cent mean coverage tolerance interval for the yield per square metre of the crop. (This can easily be converted to an interval for the total yield in the field.)

We assume that the weights in each square metre are independent No(\mu, \tau) where (\mu, \tau) may vary from field to field. The informative experiment thus effectively provides independent m and \nu where m is No(\mu, 24\tau) and \nu is Ch(23, \tau). Suppose the values obtained for the 24 measurements are as in table 5.3. Then m = 7.97 and \nu = 57.87. Hence substitution in (5.51) gives a tolerance interval of similar mean coverage 0.95 of the form (4.62, 11.32).

In the same way we can obtain one sided intervals (-\infty, 10.74) or (5.19, \infty) of similar mean coverage 0.95 by substitution in (5.46) or (5.47).
5.9 Mean coverage tolerance predictors for the multinormal case

It is relatively easy to obtain a similar mean coverage tolerance predictor for the multinormal case analogous to (5.48) for the normal case. Suppose that

\[ p_T(y|\theta) = N_{d}(\mu, K\tau) \]  

(5.55)

and that the informative experiment provides \((m, v)\), where \(m\) and \(v\) are independently distributed as \(N_{d}(\mu, k\tau)\) and \(W_{d}(\nu, \tau)\). We consider a predictor \(\delta\) giving predictive regions which are ellipsoidal in shape:

\[ \delta(m, v) = \{ y : (y - m)'v^{-1}(y - m) \leq q \}, \]  

(5.56)

the counterpart of (5.48). The problem is to determine \(q\) such that \(\delta\) has mean coverage \(c\). By (5.9) the mean coverage can be expressed as

\[ P_r\{ (m, v, y) : (y - m)'v^{-1}(y - m) \leq q \}. \]  

(5.57)

It is again easily verified that the value of this is independent of \(\theta = (\mu, \tau)\), and so similar mean coverage tolerance predictors can be constructed. For simplicity of evaluation we can thus set \((\mu, \tau) = (0, I_d)\), for which case \(y - m\) is

\[ N_{d}\left( 0, \left( \frac{1}{k} + \frac{1}{K} \right)^{-1} I_d \right) \]

independently of \(v\), which is \(W_{d}(\nu, I_d)\).

Two routes to the determination of \(q\) are available. The first shows that the quadratic form

\[ \nu(y - m)'v^{-1}(y - m)\left( \frac{1}{k} + \frac{1}{K} \right) \]

has a \(T^2\) distribution (Anderson, 1958), which through its relation to the \(F\) distribution, yields

\[ q = \left( \frac{1}{k} + \frac{1}{K} \right) \frac{d}{\nu - d + 1} F(d, \nu - d + 1; c). \]  

(5.58)

The other exploits the property (4.12) by observing that if \(W'W = \nu\) then

\[ z = \nu^{1/2}W^{-1}(y - m)\left( \frac{1}{k} + \frac{1}{K} \right)^{1/2}\]  

is \(St_d(\nu, 0, I_d)\).
Mean coverage tolerance prediction

Hence (5.57) can be expressed as

\[ P \left\{ z: z'z \leq \nu q/\left(\frac{1}{k} + \frac{1}{K}\right) \right\} , \]

which by (4.12) can be evaluated as

\[ I_{r(r+1)} \left\{ \left( q - d - \frac{1}{r} \right) \right\} , \]

where

\[ r = q/\left(\frac{1}{k} + \frac{1}{K}\right) . \]

Hence

\[ q = \left( \frac{1}{k} + \frac{1}{K} \right) \frac{\text{Be} \{ \frac{1}{r}, \frac{1}{r+1}; c \}}{1 - \text{Be} \{ \frac{1}{r}, \frac{1}{r+1}; c \}} \]  

(5.59)

which is equivalent to (5.58).

5.10 Mean coverage tolerance predictors for the two-parameter exponential case

Suppose that for the future experiment

\[ p(r|\theta) = \text{Er}(\mu, Kr) \]  

(5.60)

and that the information from \( e \) can be condensed to \((m, v)\), where \( m \) and \( v \)

are independent with density functions

\[ p(m|\mu, \tau) = \text{Er}(\mu, k\tau), \quad p(v|\tau) = \text{Ga}(v, \tau). \]  

(5.61)

One-sided tolerance predictor. We consider a predictor of the form

\[ \delta(m, v) = (m + qv, \infty) \]  

(5.62)

where \( q \geq 0 \). Since \( m \geq \mu \) and \( v \geq 0 \) the restriction to the case \( q > 0 \) ensures that \( m + qv \geq \mu \) and avoids a complication of the coverage distribution which otherwise arises. The coverage of the predictor (5.62) is

\[ \exp \{-Kr(m + qv - \mu)\} \]  

(5.63)

which has a distribution independent of \( \theta = (\mu, \tau) \). Thus in computing mean coverage we can consider the convenient case \((\mu, \tau) = (0, 1)\). The mean coverage is

\[ \int_{0}^{\infty} \int_{0}^{\infty} \exp(-Km - qKv)p(m)p(v)dm\,dv \]

\[ = \int_{0}^{\infty} \exp(-Km)p(m)dm \int_{0}^{\infty} \exp(-qKv)p(v)dv \]  

(5.64)
Mean coverage tolerance prediction

where

\[ p(m) = \text{Er}(0, k) = \text{Ex}(k), \]
\[ p(v) = \text{Ga}(\nu, 1). \]

Integration of (5.64) now gives mean coverage as

\[
\frac{k}{k + K} \frac{1}{(1 + qK)^v} \tag{5.65}
\]

with the following conclusion.

The tolerance predictor

\[
\delta(m, v) = \left[ m + \frac{v}{K} \left( \frac{k}{c(k + K)} \right)^{1/v} - 1 \right], \tag{5.66}
\]

has similar mean coverage c. Note that a condition for the existence of such a predictor with \( q > 0 \) is that

\[
\frac{k}{c(k + K)} > 1,
\]

that is

\[
k > cK/(1 - c). \tag{5.67}
\]

The critical quantity \( k/(k + K) \) involved in this condition is, of course, simply the mean coverage of the tolerance predictor \( \delta(m, v) = (m, \infty) \).

The relaxing of the condition that \( q > 0 \) alters expression (5.63) for the coverage of (5.62) to

\[
\begin{cases} 
\exp\{-Kr(m + qv - \mu)\} & \text{if } m + qv \geq \mu, \\
1 & \text{if } m + qv < \mu, 
\end{cases} \tag{5.68}
\]

and the ensuing mean coverage computation is more involved. This complication can be avoided by ensuring that the informative experiment is large enough for (5.67) to hold.

Two-sided tolerance predictor. As an example of a tolerance predictor providing a finite interval we consider

\[
\delta(m, v) = (m, m + qv) \tag{5.69}
\]

with \( q > 0 \). The technique of evaluating mean coverage is now so familiar that we omit details here and quote the conclusion.
The tolerance predictor
\[ \delta(m, v) = \left[ m, m + \frac{v}{K} \left( 1 + \frac{c(k + K)}{k} \right) \right] \]  
(5.70)

has similar mean coverage \( c \). Note again the condition \( k > cK/(1 - c) \) for the existence of such a predictor.

**History**

The concept of mean coverage tolerance predictors was introduced by Wilks (1941) who constructed such predictors for the normal case. For the multivariate normal case, see Chew (1966) and Guttman (1970). Most writers appear to have concentrated on the normal case. For some of the other cases, see Alchison and Sculthorpe (1965). See also Bain and Weeks (1965), Paulson (1943) and Proschan (1953).

**Problems**

5.1 Suppose that the informative experiment \( e \) is \( \text{Ge}(\theta) \), with density function
\[ p_e(x|\theta) = \theta^x (1 - \theta) \quad (x = 0, 1, 2, \ldots), \]
where \( 0 < \theta < 1 \), and that the future experiment \( f \) is \( \text{Ge}(\theta) \). Derive tolerance predictors of mean coverage \( c \) of the form

- (i) \( \delta(x) = \{0, 1, \ldots, e(x)\} \),
- (ii) \( \delta(x) = \{e(x), e(x) + 1, \ldots\} \).

5.2 Suppose that the future experiment \( f \) is described by a Pareto distribution \( \text{Pa}(k, \theta) \), with density function
\[ p_f(y|\theta) = \frac{\theta \theta^x}{x^{k+1}} \quad (x > k), \]
where \( \theta > 0 \) and \( k \) is a known constant, and that the informative experiment provides \( x_1, x_2, \ldots, x_n \) from \( n \) replicates of \( \text{Pa}(k, \theta) \). Show that
\[ v = \log \left( \frac{x_1 x_2 \ldots x_n}{k^n} \right) \]
is sufficient for \( \theta \) with \( p_v(v|\theta) = \text{Ga}(n, \theta) \). Derive mean coverage tolerance predictors of the forms

- (i) \( \delta(v) = (k + qv, \infty) \),
- (ii) \( \delta(v) = (k, k + qv) \),
- (iii) \( \delta(v) = (k + q_1 v, k + q_2 v) \).
5.3 After the completion of the first three out of each set of eight binomial trials, each with outcome either response or non-response, an attempt is to be made to estimate the minimum number of responses that will be obtained in the complete set. This estimation attempt is to be made for each of a long series of such sets, and it is hoped that about 80 per cent of the attempts will be successful. Provide a suitable procedure to attain this target.

5.4 Standard bales of cloth of length 50 metres have 'blemishes' distributed along their length according to a Poisson process. For each bale of cloth the number of blemishes is counted along 10 metres chosen at random and then a label is attached to the bale stating the estimated maximum number of blemishes for that bale. The manufacturer wants the information on 95 per cent of these labels to be correct.

What maximum number of blemishes should be quoted for standard bales whose sampled 10 metres have shown (i) 1 blemish, (ii) 7 blemishes?

5.5 Light bulbs are produced independently in batches of 50. The lifetimes of bulbs in a batch may be assumed to be described by an \( \text{Ex}(\theta) \) distribution, but \( \theta \) may vary from batch to batch. Before marketing a batch the manufacturer wishes to make some statement about the minimum lifetime of a bulb from the batch, and to be correct in his statements 95 per cent of the time. He therefore selects 5 bulbs at random from the batch and measures their lifetimes simultaneously. What statement should he make if the lifetimes (in hours) are 111, 86, 110, 50, 16?

Suppose that the manufacturer decides that he cannot afford to spend more than 72 hours on testing the bulbs from any batch. He therefore counts the number out of the 5 which have failed in that time. Can you help him now to make the relevant statement about the minimum lifetime of a bulb in the remainder of the batch?

5.6 Consider again the 'self-destructive' components described in problem 4.12. These have been designed so that each lifetime is described by an \( \text{Er}(\mu, \tau) \) random variable. Suppose, however, that we are unable to specify how \( \mu, \tau \) vary over production runs. Each run produces 5 components. It is so important to the customer to have an accurate description of the lifetime that he persuades the manufacturer to test 4 of the 5 components and then provide him with a statement about the minimum lifetime of the remaining component. The manufacturer realises that there is not much evidence available and says that he can only be correct in his statements 75 per cent of the time. What minimum lifetime should he report if the 4 tested components had lifetimes 306.5, 301.1, 304.2, 320.3 mins?
5.7 After completing a standard linear regression analysis in which you have
analysed the relationship of a response to a stimulus you are asked to express
a view as to what response will occur if a stimulus of strength \( t \) is applied to a
new experimental unit.

Can you state an interval within which you can claim 'with 95 per cent
confidence' that the new response will lie?

Apply the theory you have developed to the following data to make a
statement concerning the use of the stimulus at strength 1.5.

<table>
<thead>
<tr>
<th>Stimulus strength</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

5.8 In the analysis of the metabolic excretion rates in §4.4 a region of
previous experience was constructed. This had the property that the proba-
bility of a new hyperplasia patient having metabolic rates within the region
is assessed (in terms of the predictive distribution) to be 0.99. For this set
of data construct an elliptical tolerance prediction of mean coverage 0.99,
and compare this with the region of previous experience.

5.9 Show that the region of previous experience based on independently
distributed \( \text{N}_d(\mu, \Sigma) \) vectors \( x_1, \ldots, x_n \) is a similar mean coverage tolerance
predictor with mean coverage \( c \), where \( c \) is given by

\[
F(d, n - d, c) = \frac{n(n-d)}{d(n+1)} \max_i (x_i - \bar{x})'v^{-1}(x_i - \bar{x}),
\]

where

\[
\bar{x} = \frac{1}{n} (x_1 + \ldots + x_n), \quad v = \sum_i (x_i - \bar{x})(x_i - \bar{x})'.
\]
Guaranteed coverage tolerance prediction

6.1 Introduction

We saw in §4.6 that the mathematical achievement of a guaranteed coverage tolerance predictor is to ensure that the bulk of its coverage distribution lies above a specified value, conveniently termed the guaranteed coverage, say c.

To make this idea concrete we had to specify what we meant by "bulk" and we chose a definition involving quantiles of the coverage distributions. This approach can be conveniently expressed in terms of the concept of the guarantee of a predictor. The first step is to define the guarantee function of a predictor δ in providing cover c; this is a function of θ, the indexing parameter (cf. the concept of power function). We define $g(\theta | \delta, c)$ to be the probability that δ produces a region with cover of at least c, that is

$$g(\theta | \delta, c) = P_{\theta} \{ x : P_{\theta} \{ \delta(x) | \theta \} \geq c | \theta \}. \tag{6.1}$$

This quantity, for given θ, is simply the amount of the coverage distribution to the right of c. We can then define the guarantee of δ in providing coverage c.

Definition 6.1

Guarantee provided by a predictor. The guarantee of coverage c provided by a predictor δ is

$$g(\delta, c) = \inf_\theta g(\theta | \delta, c). \tag{6.2}$$

The formal definition of a guaranteed coverage tolerance predictor is then immediate.

Definition 6.2

Guaranteed coverage tolerance predictor. A $(c, g)$ guaranteed coverage tolerance predictor δ provides coverage c with guarantee g if $g(\delta, c) = g$, that is if

$$\inf_\theta P_{\theta} \{ x : P_{\theta} \{ \delta(x) | \theta \} \geq c | \theta \} = g. \tag{6.3}$$

For brevity, we shall refer to such a predictor as a $(c, g)$ tolerance predictor. For the case of discrete distributions the equality sign in definition 6.2 is relaxed to $\geq$. 

110
Guaranteed coverage tolerance prediction

It can happen that the guarantee function is constant, that is, does not depend on $\theta$. For such predictors the familiar terminology of similarity is used, thus giving the following definition.

Definition 6.3

**Similar guaranteed coverage tolerance predictor.** A tolerance predictor $\delta$ has similar guaranteed coverage $c$ with guarantee $g$ if

$$g(\theta | \delta, c) = g \text{ for every } \theta \in \Theta. \tag{6.4}$$

Note that as in the case of mean coverage tolerance predictors we have again extended the usual definitions and that our similar guaranteed coverage tolerance predictors are the usual frequentist guaranteed coverage predictors.

Example 6.1

(i) Exponential distribution. We have already seen in example 4.5(i) that a $(c, g)$ tolerance predictor exists. Here we rework the example in terms of the guarantee function. We have, for $\delta(x) = (q x, \infty)$,

$$g(\theta | \delta, c) = P_x \{x : \exp(-q \theta x) \geq c | \theta\}$$

$$= P_x \left\{x : x \leq -\frac{1}{q \theta} \log c | \theta\right\}$$

$$= 1 - e^{\frac{-c}{q \theta}}, \tag{6.5}$$

which is independent of $\theta$, so that

$$g(\delta, c) = 1 - e^{\frac{-c}{q \theta}}, \tag{6.6}$$

and we obtain a similar $(c, g)$ tolerance predictor by setting

$$1 - e^{\frac{-c}{q \theta}} = g,$$

that is,

$$q = \frac{\log c}{\log (1 - g)} \tag{6.7}$$

as before.

That the problem is not always trivial can be seen if we consider a predictor $\delta(x) = (q_1 x, q_2 x)$ which provides finite intervals. Then

$$g(\theta | \delta, c) = P_x \{x : \exp(-q_1 \theta x) - \exp(-q_2 \theta x) \geq c | \theta\}. \tag{6.8}$$

It is easy to see that this is independent of $\theta$ and so $g(\delta, c)$ is the value of $g(\theta | \delta, c)$ at any convenient value of $\theta$, say $\theta = 1$. But the evaluation of

$$g(\delta, c) = \int_{\left\{x : \exp(-q_1 x) - \exp(-q_2 x) \geq c\right\}} \exp(-x) dx \tag{6.9}$$
and the solution of \( g(\delta, c) = g \) can be achieved only by numerical methods. Indeed there is a whole infinity of possible pairs \((q_1, q_2)\), lying in the \((q_1, q_2)\) plane along a continuous curve (fig. 6.1) joining \( \{0, \log(1 - c)/\log g\} \) to \( \{\log c/\log(1 - g), \infty\} \), the latter corresponding to the \((c, g)\) tolerance predictor of the form \((q_1, x, \infty)\) which we have just been investigating.

(ii) Binomial trials. The guarantee function for each of the eight predictors of example 5.1(ii) can readily be constructed. For example, for \( c > \frac{1}{2} \) which would normally be a practical requirement,

\[
g(\theta | \delta_3, c) = \begin{cases} 
1 - \theta & (\theta < 1 - c), \\
0 & (1 - c < \theta < c), \\
\theta & (\theta \geq c),
\end{cases}
\]  

(6.10)
Guaranteed coverage tolerance prediction

and so

\[ g(\delta_3, c) = 0. \]  \hspace{1cm} (6.11)

The guarantee is also zero for predictors \( \delta_4, \ldots, \delta_8 \) of table 5.1. However

\[ g(\delta_1, c) = g(\delta_2, c) = 1 - c, \]

and hence \((c, g)\) tolerance predictors exist for this case only if \(1 - c \geq g\).

6.2 Interpretation of similar guaranteed coverage tolerance predictors

Users of guaranteed coverage predictors are often unclear about their suitability and it is therefore important to give specific interpretations of their properties. One such interpretation is to relate the concepts of guaranteed coverage \( c \) and guarantee \( g \) to the performance of the predictor \( \delta \) in a series of repeated applications, one following each of a series of informative experiments. Suppose that we regard as successful (although in practice we may have very limited means of assessing success) any particular prediction that provides cover at least \( c \). One measure of the effectiveness of a predictor could then be taken as the long-run proportion of successes. We thus have a sequence of situations \((\theta_1, x_1), \ldots, (\theta_r, x_r)\) and the corresponding predictions \( \delta(x_1), \ldots, \delta(x_r) \). If we define

\[ t_6(\theta, x) = \begin{cases} 1 & \text{if } P_{\theta}(\delta(x)|\theta) \geq c, \\ 0 & \text{otherwise,} \end{cases} \]  \hspace{1cm} (6.12)

then we have a counting random variable for success. We wish the long-run proportion of success to be approximately (or at least) \( g \), where \( g \) will not be too far from 1, so that

\[ t_6(\theta_1, x_1) + \ldots + t_6(\theta_r, x_r) = g. \]  \hspace{1cm} (6.13)

Now the model counterpart of long-run average is expectation and so we wish to choose \( \delta \) such that

\[ \mathbb{E}[t_6(\cdot, \cdot)] = g. \]

This means that we require

\[ \int_{\Theta} \int_X t_6(\theta, x) p_n(x|\theta)p(\theta)d\theta dx = g, \]  \hspace{1cm} (6.14)

where \( p(\theta) \) is the (possibly unknown) density with which nature selects \( \theta \).

Taking a frequentist approach to this we first write it in the form:
114 Guaranteed coverage tolerance prediction

\[ \int_{\Theta} P(\theta) \left( \int_{X} t_{\theta}(x, x) \cdot p_{\theta}(x|\theta) \, dx \right) \, d\theta = g, \quad (6.15) \]

and note that this relation will be satisfied for all \( p(\theta) \) if we can arrange that the inner integral is equal to \( g \), that is,

\[ \int_{X} t_{\theta}(x, x) \cdot p_{\theta}(x|\theta) \, dx = g \text{ for every } \theta \in \Theta. \quad (6.16) \]

We can express this more as a probabilistic statement in terms of \( P_{e} \) and \( P_{t} \) probability measures, for

\[ \int_{X} t_{\theta}(x, x) \cdot p_{\theta}(x|\theta) \, dx = \int_{\{x : P_{t}(\delta(x)|\theta) > c\}} 1 \cdot p_{\theta}(x|\theta) \, dx \]

\[ = P_{e} \left[ \{x : P_{t}(\delta(x)|\theta) \geq c|\theta\} \right]. \quad (6.17) \]

Thus it is reasonable to use a \((c, g)\) predictor if a sequence of predictions is envisaged, one for each informative experiment, and if we want to guarantee that a proportion \( g \) of these regions will be reliable in the sense that they provide cover at least \( c \).

6.3 Guaranteed coverage tolerance predictors for the binomial and Poisson cases

The easiest approach to guaranteed coverage predictors for the standard discrete distributions — the binomial and Poisson distributions — is through the theory of confidence intervals for the distribution parameter — the binomial success probability \( \theta \) or the Poisson mean \( \theta \). We describe the construction of predictors of the form

\[ \delta(x) = \{0, 1, \ldots, e(x)\}. \quad (6.18) \]

For any given \( \theta \in \Theta \) we can determine \( d_{\theta}(\theta) \), the smallest integer \( d \) satisfying the inequality

\[ \sum_{y=0}^{d} \sum_{y=0}^{d} \cdot p_{t}(y|\theta) \geq c. \quad (6.19) \]

Note that for both the binomial and Poisson distributions \( d_{\theta}(\theta) \) is an increasing function of \( \theta \). If we knew \( \theta \) then \( d_{\theta}(\theta) \) could be regarded as a \((c, 1)\) tolerance predictor. It is the fact that we do not know \( \theta \) that makes us lower our sights below \( 1 \) to \( g \) for the guarantee. But we can find a confidence interval for \( \theta \) by standard methods. Let \( u_{\theta}(x) \) be an upper confidence limit for \( \theta \) at confidence level \( g \). Then

\[ P_{e} \left\{ x : u_{\theta}(x) \geq \theta |\theta \right\} \geq g \quad (\theta \in \Theta) \quad (6.20) \]
**Guaranteed coverage tolerance prediction**

so that

\[ P_{\theta} \{ x : d_{\theta} \{ u_{\theta}(x) \} \geq d_{\theta}(\theta) \} \geq g \quad (\theta \in \Theta), \]  

(6.21)

and if we set

\[ \epsilon(x) = d_{\theta} \{ u_{\theta}(x) \} \]  

(6.22)

the interior inequality is equivalent to

\[ P_{\theta} \{ \delta(x) \} \geq c. \]  

(6.23)

Hence \( \delta \) satisfies definition 6.2 and so is a \((c, g)\) tolerance predictor.

**Binomial case.** The problem remaining for special cases is the determination of the functions \( u_{\theta} \) and \( d_{\theta} \). The basis of their determination for the binomial family lies in the relationship between the binomial distribution function and the incomplete beta function:

\[ d(n) = \frac{1}{\Gamma(d+1,n-d)}. \]  

(6.24)

As in §5.5 we consider the case of a Bi\((n, \theta)\) informative experiment and a Bi\((N, \theta)\) future experiment. The upper confidence limit \( u_{\theta}(x) \) for \( \theta \) at confidence level \( g \) is the solution of

\[ I_{\theta}(x+1,n-x) = g \]  

(6.25)

for \( \theta \); alternatively Pearson and Hartley (1966) provide \( u_{\theta}(x) \) directly in their table 41 for \( g = 0.975, 0.995 \). Moreover \( d_{\theta}(\theta) \) is the minimum integer \( d \) satisfying

\[ I_{\theta}(d+1,N-d) \leq 1 - c. \]  

(6.26)

**Example 6.2**

A quality control problem. Recall example 1.3. For a particular batch let \( \theta \) denote the probability that an item will be defective, so that we can regard \( e \) as Bi\((5, \theta)\) and a typical \( f \) (determining the number of defectives in a packet) as Bi\((25, \theta)\). If for an outcome \( x \) of \( e \) we state on each packet that the number \( Y \) of defective items will be in the interval \( \delta(x) \) of form (6.18) then the contract requires that

\[ P_{\theta} \{ x : P_{\theta} \{ \delta(x) \} \geq 0.8 \} \geq 0.9 \]  

(6.27)

whatever the value of \( \theta \). Comparing this with definition 6.2 we clearly require a \((c, g)\) tolerance predictor with \( c = 0.8 \) and \( g = 0.9 \).

Consider the construction of this type of tolerance prediction corresponding to \( x = 1 \). First from (6.25) we have to solve

\[ I_{\theta}(d+1,N-d) \leq 1 - c. \]  

(6.26)
Guaranteed coverage tolerance prediction

\[ I_0(2, 4) = 0.9 \]  

(6.28)

for \( \theta \) to obtain \( u_g(1) \). From the Pearson (1934) tables we obtain

\[ u_g(1) = 0.584. \]  

(6.29)

Then from (6.26) we obtain \( \epsilon(1) \) as the minimum integer \( d \) satisfying

\[ I_{0.584}(d + 1, 25 - d) \leq 0.2. \]  

(6.30)

From the Pearson (1934) tables we have

\[ I_{0.584}(16 + 1, 25 - 16) \approx 0.22, \]
\[ I_{0.584}(17 + 1, 25 - 17) \approx 0.12, \]

so that

\[ \epsilon(1) = 17. \]  

(6.31)

The complete guaranteed coverage tolerance predictor is set out in table 6.1. Note that the criterion of a \((c, g)\) tolerance predictor is so much more demanding than a \(c\)-mean coverage tolerance predictor that much wider prediction intervals (cf. table 5.2) result.

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon(x) )</td>
<td>11</td>
<td>17</td>
<td>21</td>
<td>24</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 6.1 Guaranteed coverage predictor for the quality control problem, with \((c, g) = (0.80, 0.90)\)

Poisson case. The determination of the functions \( u_\theta \) and \( d_\epsilon \) for a \( \text{Po}(k\theta) \) informative experiment and a \( \text{Po}(K\theta) \) future experiment is similar to that for the binomial case, with the incomplete gamma function \( I \) replacing the incomplete beta function \( J \). The basic relationship between the Poisson distribution function and the incomplete gamma function takes the form:

\[ \sum_{x=0}^{d} \frac{\exp(-\theta)\theta^x}{x!} = 1 - I_{0}(d + 1). \]  

(6.32)

The upper confidence limit \( u_\theta(x) \) for \( \theta \) at confidence level \( g \) is then the solution of

\[ J_{K\theta}(x + 1) = g \]  

(6.33)

for \( \theta \); for the particular cases \( g = 0.95, 0.975, 0.99, 0.999 \), Pearson and Hartley (1966) provide \( ku_\theta(x) \) directly in their table 40. Similarly \( d_\epsilon(\theta) \) is the minimum integer \( d \) satisfying

\[ J_{K\theta}(d + 1) \leq 1 - c. \]  

(6.34)
6.4 Guaranteed coverage tolerance predictors for the gamma case

Here as in §5.7 we are interested in finding a tolerance predictor \( \delta \), based on the information \( x \) from a \( \text{Ga}(k, \theta) \) experiment for the outcome of a \( \text{Ga}(K, \theta) \) experiment and of the form

\[
\delta(x) = (0, qx). \tag{6.35}
\]

The coverage is then

\[
\int_0^x \frac{\theta^k}{\Gamma(K)} y^{K-1} \exp(-\theta y) \, dy \tag{6.36}
\]

and we have seen in §5.7 that the coverage distribution is independent of \( \theta \).

In evaluating the guarantee function we can therefore conveniently set \( \theta = 1 \). We then have

\[
g(0 | \delta, c) = P_e \left\{ x : \int_0^x \frac{\theta^k}{\Gamma(K)} y^{K-1} \exp(-\theta y) \, dy \geq c \right\}
= P_e \{ x : qx \geq \text{Ga}(K, 1; c) \}
= P_e \left\{ x : x \geq \frac{1}{q} \text{Ga}(K, 1; c) \right\}. \tag{6.37}
\]

We shall thus obtain the value \( g \) for this if we set \( (1/q) \text{Ga}(K, 1; c) \) equal to the \( (1-g) \)-quantile of the \( \text{Ga}(k, 1) \) distribution. Thus we have

\[
\frac{1}{q} \text{Ga}(K, 1; c) = \text{Ga}(k, 1; 1-g)
\]

and so

\[
q = \frac{\text{Ga}(K, 1; c)}{\text{Ga}(k, 1; 1-g)}. \tag{6.38}
\]

We therefore have the following result.

The predictor

\[
\delta(x) = \begin{cases} 0, & \frac{\text{Ga}(K, 1; c)}{\text{Ga}(k, 1; 1-g)} x \\ \end{cases} \tag{6.39}
\]

is a similar \((c, g)\) tolerance predictor.

The corresponding \((c, g)\) tolerance predictor \( \delta \) of the form

\[
\delta(x) = (qx, \infty) \tag{6.40}
\]

can easily be shown to have

\[
q = \frac{\text{Ga}(K, 1; 1-c)}{\text{Ga}(k, 1; g)}. \tag{6.41}
\]
Many \((c, g)\) finite interval predictors can be obtained. For
\[
\delta(x) = [q_1 x, q_2 x]
\]
we have a \((c, g)\) similar guaranteed coverage tolerance predictor if we choose
\[
q_1 = \frac{\text{Ga}(K, 1; c_1)}{\text{Ga}(K, 1; g_2)}, \quad q_2 = \frac{\text{Ga}(K, 1; c_2)}{\text{Ga}(K, 1; g_1)}
\]
where \(c_2 - c_1 = e. g_2 - g_1 = g\). The choice of specific \(c_1, c_2, g_1, g_2\) will depend on what other feature, such as shortness of interval, we require of the predictor.

### 6.5 Guaranteed coverage tolerance predictors for the normal case

We suppose that the future experiment and the informative experiment are as described in the corresponding section §5.8 on mean coverage tolerance predictors.

**One-sided tolerance predictors.** Consider a predictor \(\delta\), based on the information \((m, v)\) as specified in §5.8, and of the form
\[
\delta(m, v) = (-\infty, m + qv^{1/2}).
\]
The coverage is then \(\Phi\{(K^2/2)(m + qv^{1/2}) - \mu\}\) and we have seen in our considerations of mean coverage tolerance predictors that the coverage distribution is independent of \(\theta = (\mu, \tau)\), so that we can consider the case \(\theta = (0, 1)\) in evaluating the guarantee function associated with \(\delta\). We then have
\[
g(\theta|\delta, c) = P_{\Theta}(\theta, \delta): \Phi\{K^{1/2}(m + qv^{1/2})\} \geq c
\]
\[
= P_{\Theta}\{m, v: K^{1/2}(m + qv^{1/2}) \geq \Phi^{-1}(c)\} = P_{\Theta}\{m, v: -mk^{1/2} + \left(\frac{k}{K}\right)^{1/2} \Phi^{-1}(c) \left(\frac{v}{\nu}\right)^{1/2} \leq (k\nu)^{1/2} q\}.
\]
Since \(m\) and \(v\) can be taken to be independently distributed as \(\text{No}(0, k)\) and \(\text{Ch}(v, 1)\) respectively the statistic on the left side of the inequality in (6.45) has the non-central \(t\)-distribution \(t\{v, (k/K)^{1/2} \Phi^{-1}(c)\}\). (See Resnikoff and Lieberman, 1957, and Resnikoff, 1962.) In order to obtain a similar \((c, g)\) predictor, that is with
\[
g(\theta|\delta, c) = g\text{ for every } \theta \in \Theta,
\]
we simply set
\[
q(kv)^{1/2} = t\left(v, \left(\frac{k}{K}\right)^{1/2} \Phi^{-1}(c); g\right).
\]
We thus have the following results.

The predictor
\[
\delta(m, v) = \left[ -\infty, m + \left( \frac{v}{k\nu} \right)^{1/2} t \left( \nu, \frac{k}{K} \right) \Phi^{-1}(c); \nu \right] \tag{6.46}
\]
is a similar \((c, g)\) tolerance predictor.

Similarly there can be derived the following result for the other kind of one-sided region.

The predictor
\[
\delta(m, v) = \left[ m - \left( \frac{v}{k\nu} \right)^{1/2} t \left( \nu, \frac{k}{K} \right) \Phi^{-1}(c); \nu \right], \tag{6.47}
\]
is a similar \((c, g)\) tolerance predictor.

**Symmetric two-sided tolerance predictors.** For a predictor
\[
\delta(m, v) = (m - q\nu^{1/2}, m + q\nu^{1/2}) \tag{6.48}
\]
the coverage is
\[
\Phi\{(K\nu^{1/2}(m + q\nu^{1/2} - \mu)) - \Phi\{(K\nu^{1/2}(m - q\nu^{1/2} - \mu)\}. \tag{6.49}
\]
Here again the guarantee function is constant for all \(\theta = (\mu, \tau)\) and so the problem is essentially that of evaluating
\[
P_e[(m, v): \Phi\{K^{1/2}(m + q\nu^{1/2})\} - \Phi\{K^{1/2}(m - q\nu^{1/2})\} \geq c], \tag{6.49}
\]
again with \(m\) and \(v\) independently distributed as \(\text{No}(0, k)\) and \(\text{Ch}(v, 1)\). The problem is thus the computational one of evaluating the double integral of \(p(m)p(v)\) over the region of integration \(R\) shown in fig. 6.2. The shape is determined by the following factors concerning the coverage
\[
\Phi\{K^{1/2}(m + q\nu^{1/2})\} - \Phi\{K^{1/2}(m + q\nu^{1/2})\}; \tag{6.49}
\]
(i) The coverage is a strictly increasing function of \(v\).
(ii) The coverage is a symmetric function in \(m\).
(iii) For fixed \(m, \Phi\{K^{1/2}(m + q\nu^{1/2})\} - \Phi\{K^{1/2}(m - q\nu^{1/2})\} = c\) has a unique solution in \(v\).

The solution of the computational problem here has had a long and interesting history (Wald and Wolfowitz, 1946; Wallis, 1951; Lieberman, 1957; Weissberg and Beatty, 1960; Ellison, 1964). The most recent and very satisfactory method for the computation of \(q\) for given \(k\) and \(v\) is given by Howe (1969) and uses only tables of normal and chi-squared quantiles. Fig. 6.3 shows the flow chart for the computation of \(q\) and for determining the order of the approximation in terms of the closeness of the guarantee attained to the guarantee aimed at. For example, if \(\nu < \nu_g\) and we use the \(q\) shown, then
Fig. 6.2 Region of integration $R$ in (6.49) for which
$\Phi \{K^{1/2}(m + qv^{1/2})\} - \Phi \{K^{1/2}(m - qv^{1/2})\} > c$.

Howe (1969) also shows numerically that the order of approximation is often very much better than that ascribed by this general result, especially for cases where $v$ and $k^2$ are of the same order of magnitude. The route along the flow chart depends on the relative magnitudes of the effective sample size $k$ associated with the estimate of the mean $\mu$ and the number $v$ of degrees of freedom associated with the estimate of variance.

**Example 6.3**

*Contract for design components.* The distributions of the characteristics of components from a simultaneous production run on each of four different machines are known from past experience to be well approximated by normal
Guaranteed coverage tolerance prediction

Write \( \xi_g = N(0, 1; \frac{1}{2}(1 + g)) \)
\( \eta_g = \chi^2(\nu, 1 - g) \)

Compute
\( \nu_g = \left(1 + \frac{1}{\xi_g^2}\right)\left(\frac{\nu}{K}\right) \)

Is \( \nu < \nu_g \)?

Yes

\( q = \left(\frac{1}{\nu} + \frac{1}{K}K\frac{\xi_g^2}{\eta_g}\left(1 + \frac{K^2(1 - \nu - 2 - \eta_g)}{2(k + K)^2}\right)\right)^{\frac{1}{3}} \)

Compute
\( s = 1 + \frac{K\xi_g^2}{\nu} + \frac{K^2(3 - \xi_g^2)}{6k^2} \)

No

\( q = \left(\frac{1}{\nu} + \frac{\xi_g^2}{2k}\right)^{\frac{1}{2}} \)

Order

\( O\left(\max\left(\frac{K^2\nu^3}{K^2}, \frac{k^3}{k}\right)\right) \)

Order

\( O\left(\max\left(\frac{K^2\nu^3}{K^2}, \frac{k^3}{k^2\nu}\right)\right) \)

Fig. 6.3 Flow chart for computation of two-sided normal tolerance limits.
Guaranteed coverage tolerance prediction

Table 6.2 Characteristics of the 52 sampled components

<table>
<thead>
<tr>
<th>Machine</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.3, 51.2, 48.4, 46.1, 51.6, 51.4, 51.7, 48.1, 51.6, 46.6, 49.2, 46.5, 48.0</td>
</tr>
<tr>
<td>2</td>
<td>56.2, 57.3, 54.5, 53.6, 51.7, 55.5, 54.1, 53.0, 52.1, 51.3, 54.6, 55.1, 53.2</td>
</tr>
<tr>
<td>3</td>
<td>59.0, 59.7, 58.2, 56.7, 59.6, 60.2, 61.5, 62.5, 61.3, 58.5, 60.6, 61.3, 57.7</td>
</tr>
<tr>
<td>4</td>
<td>49.1, 49.6, 44.5, 45.9, 46.0, 48.2, 49.6, 52.5, 48.8, 45.6, 47.2, 45.6, 47.7</td>
</tr>
</tbody>
</table>

...distributions with the same variance. The distribution means, however, vary from machine to machine and, even for the same machine, from run to run; moreover the common variance may differ from run to run. A designer contracts to accept the complete output from the first machine subject to the following conditions. The producer will determine for each production run the characteristics of 13 components from each machine. On the basis of this information he will quote to the designer an interval which purports to contain 95 per cent of the component characteristics supplied to the designer from that run. It is also agreed that at least 95 per cent of such quotations should be correct. Table 6.2 provides the characteristics associated with the 52 sampled components from one run. What quotation should the producer provide?

Let \( x_{ir} (r = 1, \ldots, 13) \) denote the observations on the \( i \)th machine \((i = 1, \ldots, 4)\). Then the underlying model asserts that \( x_{ir} \) is \( \text{No}(\mu, \tau) \) and that all 52 observations are independent. Since the contract is concerned only with a future experiment which is \( \text{No}(\mu_1, \tau) \), we can condense the information in the \( x_{ir} \) in the usual analysis-of-variance approach to

\[
m = x_1 = \sum_{r=1}^{13} x_{ir}/13, \quad (6.51)
\]

\[
v = \sum_{i=1}^{4} \sum_{r=1}^{13} (x_{ir} - x_{i.})^2. \quad (6.52)
\]

In this condensation we are therefore envisaging an informative experiment of type 5 in table 2.3, with \( m \) distributed as \( \text{No}(\mu_1, 13\tau) \) and \( v \) as \( \text{Ch}(48, \tau) \). The requirements of the contract are then met by a (0.95, 0.95) tolerance predictor of the form

\[
(m - qv^{1/2}, m + qv^{1/2}), \quad (6.53)
\]

where \( q \) is determined by the flow chart of fig. 6.3 with \( c = 0.95, g = 0.95, k = 13, v = 48, K = 1 \).
Guaranteed coverage tolerance prediction

Since

\[ \nu_v = \left( 1 + \frac{1}{(1.96)^2} \right) (13)^2 = 213 \]  \hspace{1cm} (6.54)

we have \( \nu < \nu_v \) and so the computations follow the left-hand branch, leading to

\[ q = 0.359. \]  \hspace{1cm} (6.55)

From the data of table 6.2,

\[ m = 49.36, \nu = 188.96 \]  \hspace{1cm} (6.56)

so that the appropriate quotation to the designer is the interval

\[ (44.4, 54.3). \]  \hspace{1cm} (6.57)

The order of the approximation computed from the flow diagram is here

\[ \max \left( \frac{(48)^1}{(13)^2}, \frac{(48)^1}{(13)^2} \right) = 0.15, \]  \hspace{1cm} (6.58)

which appears remarkably poor. Howe shows, however, that for this particular \( \nu, k \) and \( K \) the actual guarantee is 0.9526 compared with the specified target of 0.95.

6.6 Guaranteed coverage tolerance predictors for the multinormal case

It is clear that the already non-trivial computational problem of constructing guaranteed coverage tolerance predictors for the univariate case of §6.5 becomes even more complicated for the multinormal case. The simplest form of predictor to consider is again the elliptical type of form (5.56). A successful computational technique, due to Guttman (1970), is

(i) to obtain, for given \( q \), approximations to the mean and variance of the coverage distribution,

(ii) to 'fit' a beta distribution, say Be \( \{ r(q), s(q) \} \), with the same mean and variance,

(iii) to 'adjust' \( q \) until the fitted beta distribution has its \( g \)-quantile equal to \( c \), namely

\[ I_c \{ r(q), s(q) \} = g. \]  \hspace{1cm} (6.59)

For details and tables of \( q \) for \( d = 2, 3, 4 \) and \( c, g = 0.75, 0.90, 0.95, 0.99 \), see Guttman (1970).
Guaranteed coverage tolerance prediction

6.7 Guaranteed coverage tolerance predictors for the two-parameter exponential case

We now consider the construction of $(c, g)$-tolerance predictors when the basic situation is of two-parameter exponential type as described in §5.10.

To obtain an interval predictor of the form

$$(m + qv, \infty)$$

(6.60)

which gives coverage at least $c$ with guarantee $g$ we require that

$$P_e\left(m, v\right): \int_{m+qv}^{\infty} p(y)dy \geq c = g$$

(6.61)

or

$$P_e\left(m, v\right): m + qv \approx \frac{1}{K} \log c = g.$$  

(6.62)

(Here as in §5.10 the parameters $(\mu, \tau)$ can be set equal to $(0, 1)$.) The awkward computational aspect here is that the form of the left hand side depends crucially on the relative magnitudes of $c$ and $g$. This dependence can be clearly seen in table 6.3 which gives the guarantee $g$ corresponding to given $c$ and $q$.

Note how the form of this measure depends on the value of $q$. The results are easily established when the double integral is expressed as a repeated integral.

<table>
<thead>
<tr>
<th>Range of $q$</th>
<th>Guarantee</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q &lt; 0$</td>
<td>$1 - \frac{e^{h/K}}{(1 - kq)^y}$</td>
</tr>
<tr>
<td>$0 &lt; q &lt; \frac{1}{k}$</td>
<td>$\frac{1}{K} \log c \int_{0}^{(kq/K)\log c (\nu + 1)} p(m) dm$</td>
</tr>
<tr>
<td>$q = \frac{1}{k}$</td>
<td>$J_{(1/(Kq)) \log c (\nu + 1)} - \frac{e^{h/K}}{Kq} \int_{0}^{1} \exp (kqv)p(u) du$</td>
</tr>
<tr>
<td>$q &gt; \frac{1}{k}$</td>
<td>$J_{(1/(Kq)) \log c (\nu + 1)} - \frac{e^{h/K}}{Kq} \int_{0}^{1} \exp (kqv)p(u) du$</td>
</tr>
</tbody>
</table>

The computational problem is, of course, to determine $q$ for given $c, g, k, \nu$ and $K$, but table 6.3 shows that the form of $q$ will depend on the relative magnitudes of $c, g, k, \nu$ and $K$. The table can be converted into a flow diagram (fig. 6.4) to determine which form of table 6.3 should be used. From the first form the guarantee of the interval predictor $(m, \infty)$ is $1 - e^{h/K}$. Hence, if
Guaranteed coverage tolerance prediction

\[ 1 - e^{h/K} = g, \] we take \( q = 0. \) If \( 1 - e^{h/K} < g \) we have not achieved the desired guarantee and so will have to widen the interval \((m, \infty)\) by taking a negative \( q. \)

Solving

\[ 1 - \frac{e^{h/K}}{(1 - kq)^p} = g \quad (6.63) \]

for \( q \) then provides the stated formula for \( q. \) If \( 1 - e^{h/K} > g \) then we shall have to shorten the interval \((m, \infty)\), thus taking a positive \( q. \) Note that \((m, \infty)\) has acted as a criterion interval at this first node of the flow diagram. At the second node a similar type of argument applies, the criterion interval \((m + (1/k)v, \infty)\) of guarantee \(J_{-(h/K)1og e(v + 1)}\) providing the basis for the routing of the computations.

At first sight some of the computations seem rather laborious but in what follows we shall show that the worst is most unlikely to arise in practice, and we shall give a simple desk computer method for the others.

\[ \text{Compute } 1 - e^{h/K} \]

\[ \begin{align*}
&> g \\
&= g \\
&< g \\
\end{align*} \]

\begin{align*}
&\text{Compute } J_{-(h/K)1og e (v + 1)} \\
&\text{Take } q = 0 \\
&\text{Take } q = -1 \left( \frac{e^{h/K}}{1 - g} \right)^{1/v} - 1 \\
&\text{Take } q = \frac{1}{k} \\
&\text{Solve (6.67) for } q \\
\end{align*} \]

Fig. 6.4 Flow diagram for computation of a \((c, g)\) guaranteed coverage tolerance interval for the two-parameter exponential distribution.
First we shall arrive at the second node (and so the awkward-looking computations) if and only if

\[
\frac{k}{K} > \frac{\log(1-g)}{\log c}
\]  

(6.64)

Table 6.4 shows the critical value of \(k/K\) for practical values of \((c, g)\) in the following sense: if the value of \(k/K\) is no greater than the number shown then, regardless of the value of \(v\), we have \(q < 0\) and use the simple form. (We shall assume throughout that interest is in reasonably high values of \(c\) and \(g\), say \(c > 0.9, g > 0.9\); the selection made is intended to illustrate the concepts.) From this table the experimenter will be aware when he chooses \(k\) relative to \(K\) at the outset of his experiment just which branch of the computations he will face.

We shall now show that in a practical situation it is most unlikely that we would travel down the \(> g\) branch leading from the second node (the most difficult computationally). Since \(J_{(h/K) \log c}(v + 1)\) decreases as \(v\) increases we can verify the practical assertion by determining for ranges of \(k/K\) in excess of the critical values of table 6.4, minimum \(v\) for which

\[
J_{(h/K) \log c}(v + 1) < g.
\]

(6.65)

Table 6.5 gives these minima for \(g = 0.9\). For values of \(g\) higher than 0.9 the critical value of \(v\) is less than the corresponding value for \(g = 0.9\). To place this in a life-testing context we would expect an experimenter wishing to obtain reasonably satisfactory insight into the properties of his items to see a life-testing experiment through to a value of \(v\) at least that shown in the table. For example, at the \((0.9, 0.9)\) level it seems most unlikely that in testing 50 items he would be happy with his experiment if he had not seen through to failure at least \(r = v + 1 = 3\) items for the case \(K = 1\). We have thus taken the view that it is best to concentrate our energies in the investigation of the computation arising from the \(< g\) branch leading from the second node. The computations here are within the scope of speedy computation on a desk computer. (Along the \(> g\) branch it seems that numerical integration is necessary, which complicates the computation, although it is well within the scope of a moderate automatic computer.)
Guaranteed coverage tolerance prediction

Table 6.5 Minimum \( v \) for which guarantee of \((m + (1/k)v, -)\)
is less than \( g = 0.9 \)

<table>
<thead>
<tr>
<th>( k/K )</th>
<th>( \nu )</th>
<th>( k/K )</th>
<th>( \nu )</th>
<th>( k/K )</th>
<th>( \nu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1</td>
<td>30</td>
<td>1</td>
<td>300</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>100</td>
<td>2</td>
<td>400</td>
<td>2</td>
</tr>
<tr>
<td>100</td>
<td>7</td>
<td>200</td>
<td>6</td>
<td>500</td>
<td>2</td>
</tr>
<tr>
<td>200</td>
<td>15</td>
<td>300</td>
<td>10</td>
<td>600</td>
<td>3</td>
</tr>
</tbody>
</table>

To complete this investigation we have therefore to consider, for values of \( v \) exceeding those of table 6.5, or more precisely for configurations of \((c, g, k, K, \nu)\) satisfying

\[ J_{-\frac{k}{K}} \log e (\nu + 1) < g < 1 - e^{\frac{k}{K}}, \]  

(6.66)

the solution of the equation

\[ G(q) = g, \]  

(6.67)

where

\[ G(q) = \left( \begin{array}{c} 0 < q \leq \frac{1}{k} \end{array} \right) \]

\[ = - \frac{1}{K} \int_0^\nu \log e \left( \frac{-\log c - Km}{Kq} \right) p(v) dv \int_0^{Kq} \left( \frac{-\log c - Km}{Kq} \right)^{\nu - 1} \]

\[ \times \exp \left( \frac{-\log c - Km}{Kq} \right) k \exp (- km) dm \]

\[ = - \frac{ukc^{\frac{k}{K}}}{(1-kq)^{\nu+1}} \int_{-\frac{k}{K}} (1-kq)/(Kq) \log e (\nu + 1). \]

(6.68)

(6.69)

Since \( G'(q) < 0 \) \((0 < q \leq 1/n)\) the function \( G(q) \) steadily decreases from \( 1 - e^{k/K} \) through \( g \) to \( J_{-\frac{k}{K}} \log e (\nu + 1). \)

We have expressed \( G(q) \) and \( G'(q) \) very simply in terms of incomplete gamma ratio functions \( J_j \), and so, with a suitable table of these functions (Pearson, 1922), or with a suitable subroutine, for example, using

\[ J_e (\nu) = 1 - \exp (-x) \sum_{k=0}^{\nu} \frac{x^{k-1}}{(\nu-k)!!}, \]  

(6.70)
Guaranteed coverage tolerance prediction

for their computation, \( q \) can very quickly be determined by the Newton—Raphson iterative procedure for the solution of (6.67), namely

\[
q_r = q_{r-1} + \frac{g - G(q_{r-1})}{G'(q_{r-1})}
\]

For the range of \( c \) and \( g \) already described, for the range of \( k/k \) in table 6.4, and for the complete range of \( \nu \) exceeding the values quoted in table 6.5, we have explored numerically this computational procedure. As a consequence of this investigation we recommend as initial value for the iterative procedure:

\[
q_0 = \frac{1}{k} \left( \frac{k}{1-g} \right)^\nu
\]

This clearly lies in the correct interval \( 0 < q < 1/k \) and is in fact the solution that we would obtain by setting both \( J \) values equal to 1 in (6.68).

History

The concept of a guaranteed coverage tolerance predictor was first considered by Wald and Wolfowitz (1946) who also obtained a solution to the computational problem for the finite interval normal case. As already mentioned in §6.5 the development of this particular problem was continued through the work of Wallis (1951), who showed its extension from the case \( \nu = k - 1 \) to the regression situation; Lieberman (1957) and Weissberg and Beatty (1960), who provided tables for \( q \); Ellison (1964) and Howe (1969), who allow the direct computation through the flow chart of fig. 6.3. As already mentioned Guttman (1970) provides a treatment of the multinormal case, and also of the \( \text{Ex}(0) \) case. For other distributions the development is less clear. For tables and applications for the binomial, Poisson and gamma cases see Aitchison (1963) and Aitchison and Sculthorpe (1964). As far as we can determine the treatment of the two-parameter exponential distribution given in §6.7 is new. See also Aitchison (1964) for a discussion of the relationship of guaranteed coverage predictors to a Bayesian form of \((c,g)\) prediction, not discussed in this book.

Problems

6.1 For the case where the informative experiment \( e \) and the future experiment \( f \) are both described by \( \text{Ge}(\theta) \) random variables, derive guaranteed coverage tolerance predictors of the form

\[
(i) \quad \delta(x) = \{0, 1, \ldots, e(x)\},
\]

\[
(ii) \quad \delta(x) = \{e(x), e(x) + 1, \ldots\}.
\]
6.2 (i) For the binomial case where \( e, f \) are described by \( \text{Bi}(n, \theta), \text{Bi}(N, \theta) \) random variables, derive a guaranteed coverage tolerance predictor of the form
\[
\delta(x) = \{e(x), e(x) + 1, \ldots, N\}.
\]
(ii) For the Poisson case where \( e, f \) are described by \( \text{Po}(k\theta), \text{Po}(K\theta) \) random variables, derive a guaranteed coverage tolerance predictor of the form
\[
\delta(x) = \{e(x), e(x) + 1, \ldots\}.
\]

6.3 Derive \((c, g)\) guaranteed coverage tolerance predictors of the form \((k + qy, \infty)\) for the Pareto distribution described in problem 5.2.

6.4 An ornithologist wishes to predict the numbers of eggs laid by colonies of birds in different locations. Suppose that the numbers laid by different females in a year are described by independent \( \text{Po}(\theta) \) random variables, where \( \theta \) may vary with location. The ornithologist obtains information by selecting a random sample of 20 nests in a colony. Suppose these yield a total of 43 eggs. Provide a statement concerning the number of eggs laid by a bird in the colony which will be correct for at least 90 per cent of the birds in the colony and which will be true for 95 per cent of colonies.

6.5 Complete the analysis of problem 1.3.

6.6 Close study of ten patients discharged from hospital after treatment for a chronic disorder has been undertaken. The times to relapse of the first six to relapse out of the ten patients were 45, 52, 67, 84, 108, 135 days, the remaining four patients showing no sign of relapse at the 135th day since discharge. The clinic is now trying to formulate a reasonable policy for the recall time of patients which, for administrative purposes, must be the same for all patients. It wants to be reasonably certain (say 95 per cent) that 95 per cent of patients are recalled for treatment before relapse. On the assumption that times to relapse have a two-parameter exponential distribution what policy do you recommend?

6.7 Reconsider problem 5.5. Suppose now that the light bulbs are produced in larger batches than 50. Before marketing the manufacturer wishes to make a statement to the effect that at least 95 per cent of the bulbs in a batch will have lifetimes greater than a given value, and he wishes to be correct in his statements for 95 per cent of all batches. Obtain the relevant guaranteed coverage tolerance predictor for the data provided.

Find the limiting tolerance predictor of mean coverage 0.95 as the size of the batch increases from 50.
6.8 A shoe manufacturer assumes that the distribution of foot length and foot breadth in his women customers is bivariate normal. The relevant information from a sample study of 200 women customers is shown below. For shoe size 6 designed to meet the requirements of his women customers of foot length 26 cm the manufacturer wants to cater in breadth sizes for 95 per cent of these customers. For what range of foot breadth should he design to be 90 per cent certain of meeting this requirement?

<table>
<thead>
<tr>
<th>Foot length (cm)</th>
<th>Foot breadth (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean</td>
<td>24.3</td>
</tr>
<tr>
<td>Sample standard deviation</td>
<td>3.4</td>
</tr>
<tr>
<td>Sample correlation coefficient between two measurements</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Guaranteed coverage tolerance prediction
7
Other approaches to prediction

7.1 Introduction
In this chapter we discuss some further aspects concerning decisive and informative prediction. We recall that the decisive prediction approach can be used when knowledge of the predictive distribution \( p(y|x) \) is available together with a utility function \( U(a, y) \). When no utility specification is available one has to use an informative prediction approach and obtain either a most plausible Bayesian predictor or a frequentist tolerance predictor.

7.2 Linear utility structure
There is an interesting tie up between decisive prediction and Bayesian informative prediction for the case where we can assume a linear utility structure. For example suppose that we take the linear loss structure, (37), that is

\[
U(a, y) = \begin{cases} 
-\xi(a - y) & (y < a), \\
-\eta(y - a) & (y \geq a), 
\end{cases}
\]

for the prediction of a one-sided interval \((-\infty, a)\). From equation (3.8) we have that the optimal \( a^* \) is given by the solution of

\[
\int_{-\infty}^{a^*} p(y|x)dy = \frac{\eta}{\xi + \eta}.
\]

This interval obviously corresponds to an informative Bayesian prediction interval (not necessarily the most plausible) of coverage \( \kappa = \eta/(\xi + \eta) \). The provision of this alternative view of Bayesian cover intervals may make them more attractive to some users in that they feel that it is easier to assess the relative cost factor \( \eta/\xi \) than the more nebulous cover \( \kappa \). For example a statistician who uses a 95 per cent cover interval is behaving in approximately the same way as a statistician who regards the proportional loss caused by outcomes above the limit to be 19 times more serious than that caused by outcomes inside the interval. Table 7.1 illustrates the relationship between \( \eta/\xi \) and \( \kappa \) for several cases.

We may investigate a similar relationship for a linear loss structure of the
Other approaches to prediction

Table 7.1 Bayesian cover corresponding to relative cost factor \( \eta/\xi \) for utility function (7.1)

<table>
<thead>
<tr>
<th>( \eta/\xi )</th>
<th>( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.33</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>9</td>
<td>0.90</td>
</tr>
<tr>
<td>19</td>
<td>0.95</td>
</tr>
<tr>
<td>99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

form (3.28), that is,

\[
U(a_1, a_2, y) = \begin{cases} 
-\xi(a_1 - y) - (a_2 - a_1) & (y < a_1), \\
-\eta(y - a_2) - (a_2 - a_1) & (a_1 < y < a_2), \\
-\eta(y - a_2) - (a_2 - a_1) & (y > a_2),
\end{cases} \quad (7.3)
\]

for the prediction of a two-sided interval \((a_1, a_2)\). From equations (3.29) we have that the optimal \((a_1^*, a_2^*)\) are given by the solutions of

\[
\int_{a_1^*}^{a_1} p(y|x)dy = 1/\xi, \quad \int_{a_2^*}^{a_2} p(y|x)dy = 1/\eta, \quad (7.4)
\]

provided \(1/\xi + 1/\eta < 1\). This interval has Bayesian coverage \(\kappa = 1 - 1/\xi - 1/\eta\).

Thus, for example, if the differential losses associated with being outside the interval are 40 times greater than the cost per unit length of interval used, then the interval corresponds to a 95 per cent cover interval. Table 7.2 gives the Bayesian cover associated with the optimal intervals obtained for several \((\xi, \eta)\) values.

Table 7.2 Bayesian cover \(\kappa\) corresponding to \((\xi, \eta)\) for utility function (7.3)

<table>
<thead>
<tr>
<th>( \xi )</th>
<th>( \eta )</th>
<th>( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.50</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.90</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>0.95</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.99</td>
</tr>
<tr>
<td>( \xi )</td>
<td>( \eta )</td>
<td>( \kappa )</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>0.43</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>0.70</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>0.925</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>0.94</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.985</td>
</tr>
</tbody>
</table>

7.3 Frequentist decision theory

If we are unwilling to assign a prior plausibility function \(p(\theta)\) on \(\theta\), the predictive density function \(p(y|x)\) cannot be determined, and so the Bayesian decisive prediction approach of chapter 3 is no longer available. The corresponding frequentist utility approach lies along a path beset with difficulties. Although the utility \(U(a, y)\) is an appealing one to the practical man it has some considerable conceptual difficulties for the frequentist. If a prediction
is required for only one performance of \( \phi \) it is natural, in accord with frequentist decision theory, to introduce a predictor \( \delta \) and take the expectation of \( U(\delta(x), y) \) with respect to the informative density \( p(x | \theta) \). The resulting expectation depends on \( (\theta, y) \), the unknown state of nature; the presence of both \( \theta \) and \( y \) makes the usual awkward feature of frequentist theory, namely the difficulty of finding pivotal statistics, even more embarrassing. The dependence of \( y \) on \( \theta \) through \( p(y | \theta) \) is waiting to be used but there is no obvious frequentist way to introduce it.

If a series of replicates of \( \phi \) is to be conducted and the prediction region to be used in each replicate, the frequentist could then obtain the induced utility \( V(\delta(x), \theta) \) from

\[
V(\delta(x), \theta) = \int_{\mathcal{Y}} U(\delta(x), y) p(y | \theta) dy.
\]

He would then proceed in the usual way by basing his considerations on

\[
G(\delta, \theta) = \int_{\mathcal{X}} V(\delta(x), \theta) p(x | \theta) dx
\]

and searching to find an optimal \( \delta \) for all \( \theta \) — if one exists. The method implicitly requires that utilities are additive over replicates and in fact \( V(\delta(x), \theta) \) corresponds to the average utility per replicate, thus providing a measure of the effectiveness of the prediction region. However this approach can lead to an inconsistency in interpretation (see Aitchison and Sculthorpe, 1965) and there is little doubt that the frequentist is on the safer logical ground — though farther removed from practical considerations — when he confines himself to utility functions of the form \( V(\delta, \theta) \) without recourse to the more practical \( U(\delta, y) \).

### 7.4 Frequentist linear utility theory

As an example where it is in fact possible to maximise \( G(\delta, \theta) \) for all \( \theta \) we consider the often amenable normal case with a linear utility function. We take \( p(y | \theta) \) to be a \( N(\mu, K \tau) \) density and suppose that the informative experiment \( e \) provides us with a sufficient statistic \( (m, v) \) for \( (\mu, \tau) \) — so that \( x = (m, v) \) — and that \( m \) and \( v \) are independent with \( N(\mu, k \tau) \) and \( \chi^2(\nu, \tau) \) distributions. In such circumstances we saw in chapter 5 that a tolerance interval of mean cover \( c \) of the form \((-\infty, m + q v^{1/2})\) has

\[
q = \left( \frac{1}{\nu} \left( \frac{1}{k} + \frac{1}{K} \right) \right)^{1/2} t(\nu; c).
\]

We now find the frequentist decisive predictor of the form \( \delta(x) = (-\infty, a) \) when we have available a utility function
Other approaches to prediction

\[ U(δ(x), y) = \begin{cases} -ξ(a - y) & (y < a), \\ -η(y - a) & (y ≥ a). \end{cases} \quad (7.8) \]

Just as in mean cover analysis it seems sensible to consider limits of the form \( a = δ(m, v) = m + qv^{1/2} \). We attempt to find a \( q \), if any, which maximises \( G(q, θ) \) for all \( θ \); note that we write \( G(q, θ) \) for \( G(δ, θ) \) since \( q \) completely specifies \( δ \). It follows that

\[ G(q, θ) = \int_{-∞}^{∞} \int_{0}^{∞} V(m + qv^{1/2}, θ) p(m, v|θ) dm dv, \quad (7.9) \]

where

\[ V(m + qv^{1/2}, θ) = \eta(m + qv^{1/2} - μ) - (ξ + η) \]
\[ \times \int_{-∞}^{m + qv^{1/2}} (m + qv^{1/2} - y)p(y|θ)dy. \quad (7.10) \]

It is fairly easy to establish that a maximising value of \( q \) necessarily occurs where the derivative of \( G(q, θ) \) with respect to \( q \) is zero. The derivative equation is obtained as

\[ \int_{-∞}^{∞} \int_{0}^{∞} \{ -ξ + η\} P_r(m + qv^{1/2}|θ) v^{1/2} p(m, v|θ) dm dv = 0, \quad (7.11) \]

where \( P_r(m + qv^{1/2}|θ) \) is an abbreviation for \( P_r((-∞, m + qv^{1/2})|θ) \). Now

\[ P_r(m + qv^{1/2}|θ) = Φ((Kv)^{1/2}(m - μ + qv^{1/2})). \]

If we introduce the change of variables

\[ M = (kv)^{1/2}(m - μ), \quad V = τv, \]

we reduce the left-hand side of (7.11), after the cancellation of a factor to

\[ \int_{-∞}^{∞} \int_{0}^{∞} \{ -ξ + η\} Φ\left( (Kv)^{1/2}(M + qV^{1/2}) \right) p(M)p(V) dMdV, \quad (7.12) \]

where \( p(M) \) and \( p(V) \) are \( N(0, 1) \) and \( \chi^2(\nu + 1) \) densities respectively. Note that we use \( p(V) \) here to denote not the density function of \( V = τv \) which is \( \chi^2(\nu) \) but the naturally arising factor in \( V \) in the integrand which turns out to be a \( \chi^2(\nu + 1) \) density function. In considering (7.12) we can therefore treat \( M \) and \( V \) as independent with the specified densities and if we introduce another \( N(0, 1) \) variable \( W \), say, independent of \( (M, V) \), we see that the derivative equation (7.11) yields
Other approaches to prediction

\[ \Pr \left( W < K^{1/2} \left( \frac{M}{k^{1/2}} + q \nu^{1/2} \right) \right) = \frac{\eta}{\xi + \eta} . \]

It follows immediately, since

\[ \frac{(W/K^{1/2} - M/k^{1/2})(\nu + 1)^{1/2}}{\left( \frac{1}{k} + \frac{1}{K} \right)^{1/2} \nu^{1/2}} \]

is distributed as \( t(\nu + 1) \), that

\[ q = \left( \left( \frac{1}{k} + \frac{1}{K} \right) (\nu + 1) \right)^{1/2} t(\nu + 1; \eta/(\xi + \eta)) . \]  \hspace{1cm} (7.13)

Thus the upper normal linear utility frequentist limit is \( m + q\nu^{1/2} \), where \( q \) is given by (7.13).

There is here no relation between \( c \) and \( \eta/\xi \) (independent of \( \nu \)) which leads to equality of expressions (7.7) and (7.13); the complete equivalence of the informative prediction and linear utility intervals, discussed in §7.2 for the Bayesian approach, is thus absent. When, however, the statistic \( \nu \) is based on a large sample so that \( \nu \) is appreciable, we see from (7.7) and (7.13) that the two intervals are for all practical purposes the same if \( c = \eta/(\xi + \eta) \). Since the direct interpretation of \( c \) is not without difficulty (see Aitchison and Sculthorpe, 1965, p. 477) the new interpretation arising from this correspondence provides a useful alternative view of mean cover intervals. An interval with mean cover \( c \) is for appreciable \( \nu \) the same as a linear utility interval with \( \eta/\xi \) factor equal to \( c/(1 - c) \). Table 7.1 can again be used to illustrate the relationship if we substitute expected cover \( c \) for Bayesian cover \( k \). For example, a statistician who uses a 95 per cent expected cover interval is behaving in approximately the same way as a statistician who regards the proportional loss caused by outcomes above the limit to be 19 times more serious than that caused by outcomes inside the interval.

While we have developed the theory for intervals of type \( (-\infty, a) \) it is clear that a similar development is possible for intervals of type \( (a, \infty) \), and that the comments on correspondence between mean cover and linear utility intervals remain unchanged.

7.5 Prior probabilities on the sample space

In this section we briefly investigate an alternative approach to Bayesian prediction problems which in some cases may be more palatable to the practical man.

The possible probabilistic descriptions of the experiment form a class of density functions \( \{ p(\cdot|\theta) : \theta \in \Theta \} \) on the sample space \( Y \) and indexed by
Other approaches to prediction

parameters in the parameter space $\Theta$. We have adopted the common (Bayesian) practice of developing the analysis by the use of a prior density function $p(\theta)$ on $\Theta$. It does not seem to be generally realised, although it is implicit in the concepts of conditional probability and was discussed by Raiffa and Schlaifer (1961, chapter 1), that there is automatically induced a prior density $p(y)$ on $Y$ through the relationship

$$p(y) = \int_\Theta p(y|\theta) p(\theta) d\theta. \quad (7.14)$$

If it is impossible to perform any informative experiment then the only source of information is the prior plausibility function $p(\theta)$. We would then use $p(y)$ as the predictive density function.

The notion of the induced prior $p(y)$ suggests an alternative Bayesian approach with a different starting point. A practical man may find it easier to specify a prior $p(y)$ than to specify a prior $p(\theta)$. He may do this in an approximate way by constructing a histogram for the interpretation of his thoughts on the likely values of $y$. To a Bayesian the density function of importance is $p(y|x)$. The following question now arises: given $p(y)$, can one determine $p(y|x)$?

The solution of such a problem is not generally easy. One possible approach is to try to obtain $p(\theta)$ from the inversion of (7.14). This raises the topic of identifiability of mixtures of distributions and reference may be made for example to Teicher (1960, 1961, 1963). Difficulties arise over the uniqueness of the solution. For a given $p(y|\theta)$ and $p(y)$ there will not in general be a unique $p(\theta)$ which induces $p(y)$. This is true for example where the class of density functions $p(\cdot|\theta) : \theta \in \Theta$ is the class of normal distributions or of gamma distributions.

If one can satisfactorily obtain $p(\theta)$, however, the way is clear through the use of Bayes's theorem and

$$p(y|x) = \int_\Theta p(y|\theta) p(\theta|x) d\theta$$

to the determination of $p(y|x)$.

The general implications of starting with a prior $p(y)$ instead of a prior $p(\theta)$ can in some cases be illustrated neatly. Consider for example the case in which $p(y|\theta)$ is Po($\theta$), and so $E(y|\theta) = V(y|\theta) = \theta$. We then have that

$$E(y) = E\{E(y|\theta)\} = E(\theta),$$

and

$$V(y) = E\{V(y|\theta)\} + V\{E(y|\theta)\} = E(\theta) + V(\theta).$$

From these relationships comparisons may be made concerning the corresponding choices of $p(y)$ or $p(\theta)$. The choice of a $p(y)$ which is concentrated
Other approaches to prediction

about its mean corresponds to a choice of \( p(\theta) \) similarly concentrated about its mean. At the other extreme the case where little is known about \( p(y) \) with \( V(y) \) large corresponds to the idea of prior ignorance on \( \theta \) with \( V(\theta) \) large.

7.6 Empirical Bayes prediction

In many situations we entertain the idea of the parameter \( \theta \) being a random variable in the sense that in repetitions of the experiment different values of \( \theta \) may be the true parameter value. In some situations, however, we may feel that we are not in a position to specify the distribution of \( \theta \) — the randomness alone is not sufficient. Nor are we willing to assign a uniform or diffuse prior to \( \theta \) to represent ignorance, which is in fact a very specialised assumption.

Although von Mises (1942) had discussed a similar topic the first real approach to such problems was by Robbins (1955), and it was he who coined the phrase 'empirical Bayes approach'. The assumption that \( \theta \) is a random variable imposes a plausibility function \( p(\theta) \) on \( \Theta \), but the randomness does not indicate the actual form of the distribution. This marginal distribution is therefore assumed unknown.

Robbins laid the foundations of the approach for obtaining the Bayes estimate of the parameter \( \theta \) in the 'future' experiment and found in several instances that a direct estimate of \( p(\theta) \) is not necessary. Instead attention is focussed on the marginal density function

\[
p(y) = \int_\Theta p(y|\theta)p(\theta) d\theta, \quad (7.15)
\]

and it is sufficient simply to find an estimate of this from \( x_1, x_2, \ldots, x_n \).

We can adapt the empirical Bayes approach to our problem of prediction. A sequence of experiments yields results \( x_1, x_2, \ldots, x_n \). We do not know the values \( \theta_1, \theta_2, \ldots, \theta_n \) of the parameters in the experiments but each is assumed to be an observation of a random variable \( \theta \) with unknown probability density function \( p(\theta) \). We require an informative predictor \( a \) for the future observation \( y \) in the next experiment, in which the unknown parameter value is \( \theta_{n+1} \). Using a quadratic utility function \(- (a - y)^2\) we find that the expected utility is given by

\[
-\int_y (a - y)^2 p(y) dy. \quad (7.16)
\]

This is maximised when \( a \) is the mean value of \( p(y) \). If \( p(\theta) \) is unknown we are unable to evaluate \( p(y) \) and so cannot obtain the Bayes decision rule. In the empirical Bayes approach we use the results \( x_1, x_2, \ldots, x_n \) to obtain an estimate \( \hat{p}(y|x) \) of \( p(y) \). For a discrete \( p(y) \) we estimate \( p(y) \) by \( f_n(y)/n \), where \( f_n(y) \) denotes the number of observations in the informative experiment.
having the value \( y \). It follows immediately that the expected value of \( p(y|x) \) is simply the sample mean \( \bar{x} \) of the informative experiment. For a continuous density \( p(y) \) we would obtain our estimate \( p(y|x) \) in the form of a histogram, and approximate the expected value of \( p(y) \) by the sample mean.

As in the Robbins estimation case mentioned above we see that it is unnecessary to obtain an approximation for \( p(\theta) \). We focus our attention instead on the marginal density \( p(y) \) and obtain an empirical estimate \( p(y|x) \).

The difficulty presented by the mixture problem in the similar approach of §7.5 is thus avoided.

One severe assumption that is sometimes made is that the form of the prior distribution is known, but the particular member of the family is unknown. Consider for example the empirical Bayes approach to the following situation. Suppose that we have an independent set of observations \( x_1, x_2, ..., x_n \) where \( p(x|\theta) = \text{No}(\theta, 1) \) and for the future experiment \( p(y|\theta) = \text{No}(\theta, 1) \). We assume that \( p(\theta) = \text{No}(b, c) \) where \( b, c \) are unknown. We have that

\[
p(y) = \text{No} \left( b, \left(1 + \frac{1}{c}\right)^{-1} \right).
\]

We then obtain estimates of \( b, c \) from the results \( x_1, x_2, ..., x_n \) of the informative experiment. For example we may take here

\[
\hat{b} = \bar{x}, \hat{c} = \left(\frac{n - 1}{v} - 1\right)^{-1},
\]

where

\[
\bar{x} = \frac{1}{n} \sum x_i, \quad v = S(x, x),
\]

provided that \( v > n - 1 \). Hence our estimated predictive distribution is

\[
p(y|x) = \text{No}(\bar{x}, (n - 1) v^{-1}).
\]

Further references to the empirical Bayes approach may be found in Maritz (1970).

7.7 Distribution-free prediction

We conclude this chapter by taking an even more extreme starting point. Not only are we unable to assign a prior plausibility function — we cannot even identify the parameters or the form of distribution describing the informative and future experiments. The problem of whether or not to use a parametric form of density function to describe the future and informative experiment is one which requires very careful consideration and to which statisticians have probably paid too little attention. The art of approximation in the
whole of applied mathematics is a subtle one, and the only justification for
an approximation must be that it is of sufficient practical validity, or alter-
natively it is robust against departures from approximating assumptions under
which it was derived. In statistical work the need to make an assumption about
the functional form of the density may occasionally be obviated by the use
of what have come to be termed non-parametric or distribution-free techniques.
In prediction analysis these lead to the use of order statistics in frequentist
inference prediction. The presentation of such results within our general
framework is the main purpose of this section. Non-parametric methods are
at present outside the reach of Bayesian techniques since there exists no
satisfactory way of specifying measures on spaces of functions. There is, how-
ever, some Bayesian justification for the use, when there is a state of 'prior
ignorance', of the empirical distribution as the predictive distribution \( p(y|x) \),
and for such cases it is of course then relatively easy to obtain the corre-
spanding predictions.

Preliminary distribution theory. We first set out some definitions and properties
of Dirichlet and beta distributions which will be required.

A Dirichlet density \( D_i(g, h) \) is defined on the simplex

\[
\left\{ (u_1, ..., u_d) : u_i > 0 (i = 1, ..., d), \sum_{i=1}^{d} u_i = 1 \right\}
\]

in \( d \)-dimensional Euclidean space by

\[
\frac{1}{D(g, h)} \prod_{i=1}^{d} u_i^{g_i-1} u_2^{g_2-1} \cdots u_d^{g_d-1} (1 - u_1 - \cdots - u_d)^{h-1}. \quad (7.17)
\]

The special case where \( d = 1 \) is the more familiar beta type of distribution
over the interval \((0, 1)\).

Note that \( \mathbb{E}(u_i) = g_i/(g_1 + \cdots + g_d + h) \) \((i = 1, ..., d)\). \quad (7.18)

The first important property of Dirichlet distributions is that certain con-
densations preserve the Dirichlet form in the following way. Let \( t_1, t_2, ..., t_j \)
denote the sums of \( j \) non-overlapping subsets of \( u_1, ..., u_d \); these sets need
not necessarily exhaust the \( u_s \). Then let \( k = (k_1, k_2, ..., k_j) \) denote the sums
of the corresponding subsets of \( g_1, ..., g_d \), and \( l = g_1 + \cdots + g_d + h - k_1 -
\cdots - k_j \). Then \( (t_1, ..., t_j) \) has a \( \text{Di}(k, l) \) density. In particular \( u_1 + \cdots + u_d \) has
a \( \text{Di}(g_1 + \cdots + g_d + h) \) density.

Also of importance in distribution-free prediction analysis is a related
ordered Dirichlet distribution \( \text{Di}^*(g, h) \). This can be derived from \((u_1, ..., u_d)\)
by the transformation
Other approaches to prediction

\[ t_1 = u_1, \]
\[ t_2 = u_1 + u_2, \]
\[ ... \]
\[ t_d = u_1 + u_2 + ... + u_d, \]

and gives as density, over the ordered simplex

\[ \{(t_1, ..., t_d): 0 \leq t_1 \leq t_2 \leq ... \leq t_d \leq 1\} \]

in d-dimensional Euclidean space,

\[
\frac{1}{D(g, h)} t_1^{g_1-1} (t_2 - t_1)^{g_2-1} ... (t_d - t_{d-1})^{g_d-1} (1 - t_d)^{h-1}. \tag{7.19} \]

The important property for this distribution, and the one which corresponds to the property quoted for the Di distribution, is the following. If \((s_1, ..., s_j)\) is any subset of \((t_1, ..., t_d)\) and \(k = (k_1, ..., k_j)\) gives the corresponding \(g_i\) values, then \((s_1, ..., s_j)\) has a \(D^{*}(k, l)\) density, where

\[ l = g_1 + ... + g_d + h - k_1 - ... - k_j. \]

Now we quote the basic distribution property of order statistics on which distribution-free prediction depends. Suppose that \(x_1, ..., x_n\) are the outcomes of \(n\) replicates of an experiment described by some continuous distribution function \(F\), and that the ordered outcomes are \(X(1), ..., X(n)\). Let \(t_i = F(x(i))\). Then the distribution of \((t_1, ..., t_n)\) is \(D^{*}(1', 1)\) and it is the non-dependence of this distribution on \(F\) which permits the derivation of distribution-free techniques. It then follows immediately that \((u_1, u_2, ..., u_n)\) defined by

\[ u_i = t_i - t_{i-1} \quad (i = 1, 2, ..., n) \]

has a \(D(1', 1)\) density.

**Distribution-free tolerance intervals.** Suppose that the informative experiment is \(n\) replicates of the future experiment and that associated with the future experiment is a distribution function \(F\). We shall consider what are the consequences of using as either a mean coverage or a guaranteed coverage tolerance interval the interval \((x(r), x(s))\), where \(x(r)\) and \(x(s)\) are the \(r\)th and \(s\)th order statistics of set \(x = (x_1, ..., x_n)\) of outcomes of the \(n\)-replicate experiment. The direction of investigation will be to determine the \(c\) values corresponding to the use of this interval (i) for the mean coverage approach and (ii) for the guaranteed coverage approach for a specified \(g\) value. Having found these \(c\) values we can then attempt to find \(r\) and \(s\) which provide satisfactory \(c\) values.
Mean coverage. In the mean coverage analysis we wish to obtain the expectation of

$$P_r((x(r), x(\omega)) = F(x(\omega)) - F(x(r)),$$

expectation being with respect to the $n$-replicate experiment (or its condensation in terms of $x(r), x(\omega)$). Now, in fact, in the notation introduced,

$$F(x(\omega)) - F(x(r)) = t_s - t_r = u_{r+1} + \ldots + u_s,$$

and so has a $\text{Di}(s - r, n + 1 - (s - r))$ distribution. Hence, from (7.18) its expectation is $(s - r)/(n + 1)$. Thus to obtain a $c$ mean coverage tolerance interval we have to attempt to choose $r$ and $s$ so that

$$c = \frac{s - r}{n + 1} \quad (7.20)$$

It is immediately clear, as we might have expected from the very limited class of predictors which we are considering, that by no means all values of $c$ are possible, and it is therefore sensible to select $(r, s)$ if possible in such a way that the chosen $c$ value is exceeded. If a finite prediction interval is required we cannot obtain a mean cover value greater than $(n - 1)/(n + 1)$, and $r = 1$ and $s = n$ provide the maximum value of $c$. Similar considerations apply to infinite intervals where we take $r = 0$ and $x(\omega) = -\infty$ or $s = n + 1$ and $x(n+1) = +\infty$. For finite intervals it is a common practice to take 'symmetric' intervals, i.e., with $s = n - r + 1$. Thus $(x(r), x(n - r + 1))$ supplies a tolerance interval of mean coverage $(n - 2r + 1)/(n + 1)$.

A more general result may be stated. If we term the interval between consecutive order statistics (including conventionally $x(0) = -\infty$ and $x(n+1) = +\infty$) a block, then we may say that the region consisting of $b$ such blocks provides a tolerance region of mean coverage $b/(n + 1)$.

Guaranteed coverage. The $c$ values corresponding to the use of $(x(r), x(\omega))$ as a $(c,g)$ guaranteed coverage tolerance interval is given by

$$P_e\{x : F(x(\omega)) - F(x(r)) \geq c\} = g. \quad (7.21)$$

Now since $F(x(\omega)) - F(x(r))$ has a $\text{Di}(s - r, n + 1 - (s - r))$ density this is given by

$$\int_c^{\infty} \frac{u^{s-r-1}(1-u)^{n+1-(s-r)-1}}{B(s-r,n+1-(s-r))} \, du = 1 - I_c\{s-r,n+1-(s-r)\}.$$
For a given $g$ and given $c$ the problem is therefore to select $r$ and $s$ such that

$$1 - I_c(s - r, n + 1 - (s - r)) = g. \quad (7.22)$$

Again it is not necessarily possible to choose $r$ and $s$ so that $c$ is exactly obtained and it is then necessary to choose so that $c$ is just exceeded. This, of course, may prove to be impossible. It is in fact useful before conducting an informative experiment to know whether a desired $c$ is attainable. From (7.22) it is clear that $c$ will be attainable with a finite interval provided

$$1 - I_c(n - 1, 2) > g.$$

For given $g$ and $c$ it is easy to obtain the minimum $n$ to meet the requirements from tables of incomplete beta function, or from special tables (Owen, 1962) constructed for the direct determination of $n$.

We now consider two examples illustrating the distribution-free approach for tolerance intervals. In the first example we investigate what, if anything, is lost by following the above approach when an assumption of normality in the underlying distribution is reasonable. In the second we consider a less standard situation.

**Example 7.1**

*Crop Prediction.* Recall example 5.4 on crop prediction. It is known that a normality assumption is reasonable. Consider however the distribution-free approach. The maximum value of $c$ we can obtain with a two-sided interval is 0.92 from (7.20), and this is provided by the interval $(4.1, 11.4)$. This is simply the range of the observations. The corresponding symmetric mean coverage interval obtained from (5.51) is $(5.01, 10.93)$. Notice the widening in the interval when the normality assumption is not made.

We turn now to the guaranteed coverage tolerance interval. Again consider the interval $(4.1, 11.4)$. This provides cover 0.847 with guarantee 0.90. The corresponding interval obtained by Howe’s technique (see § 6.5) is $(6.0, 9.9)$.

**Example 7.2**

*Component production.* In the manufacture of a certain type of component it is possible for an airlock to get into the component. Such a component tends to suffer a reduction in its lifetime. Unfortunately it is not possible to detect the affected components. The lifetimes (in days) obtained for 185 components are summarised in the histogram in fig. 7.1, the ordered sample containing values $(29.0, 30.7, 31.0, ..., 92.1, 92.6, 92.7, 93.2, 94.1)$. The distribution of lifetimes appears to be bimodal. A parametric approach would be to consider a mixture of two normal distributions, say, corresponding to the two types of component. This would involve problems in estimating the mixing parameter for use in the prediction of future lifetimes and lead to
non-standard situations which we have not covered. The distribution free approach here provides a simple alternative. For example the interval \((31.0, 92.7)\) is a frequentist tolerance interval of mean coverage \(c = 0.968\). Similarly this interval provides cover 0.986 with guarantee 0.95.

**History**

The frequentist decision theory approach and the similarity between linear utility and informative prediction intervals are discussed in Aitchison and Sculthorpe (1965) and Aitchison (1966).

The empirical Bayes approach was introduced by Robbins (1955) and heralded as a breakthrough by Neyman (1962). Maritz (1970) provides a good introduction to, and survey of, the topic and supplies a comprehensive bibliography.

The distribution-free approach to tolerance predictors is given by Wilks (1941) in the first important paper on tolerance regions. Guttman (1970, chapter 2) derives the relevant theory and provides several references.
Other approaches to prediction

Problems

7.1 For the case of the gamma distribution with
\[ p(x | \theta) = \text{Ga}(k, \theta), \]
\[ p(y | \theta) = \text{Ga}(K, \theta), \]
show that a frequentist linear utility interval of the form \( \delta(x) = (0, qx) \) based on the utility function
\[ U(\delta(x), y) = \begin{cases} -q(x - y) & (y < qx), \\ -\eta(y - qx) & (y \geq qx), \end{cases} \]
is given by
\[ q = \frac{1 - \text{Be}[k + 1, K; \xi/(\xi + \eta)]}{\text{Be}[k + 1, K; \xi/(\xi + \eta)]}. \]
Compare this interval with the tolerance predictor (5.40) of similar mean coverage \( c \).

7.2 For the gamma case with
\[ p(x | \theta) = \text{Ga}(k, \theta), \quad p(y | \theta) = \text{Ga}(K, \theta), \]
find the frequentist linear utility interval of the form \( \delta(x) = (q_1x, q_2x) \) and based on the utility function
\[ U(\delta(x), y) = \begin{cases} \xi(y - q_1x) & (y \leq q_1x), \\ q_1x - y & \{q_1x < y \leq \frac{1}{2}(q_1x + q_2x)\}, \\ y - q_2x & \{\frac{1}{2}(q_1x + q_2x) < y \leq q_2x\}, \\ \lambda(q_2x - y) & (y > q_2x). \end{cases} \]

7.3 Suppose that
\[ p(y | \theta) = \text{Ex}(0), \quad p(y) = \ln\text{Be}(1, 8, h). \]
What can be deduced about \( p(\theta) \)? If further
\[ p(x | \theta) = \text{Ga}(k, \theta) \]
what can be deduced about \( p(y | x) \)?

7.4 Derive an empirical Bayes predictive distribution for each of the following cases.
Other approaches to prediction

(i) $p(x_i|\theta) = \text{Po} (\theta)$ \hspace{1cm} (i = 1, 2, ..., n),
$p(y|\theta) = \text{Po} (\theta),
p(\theta) = \text{Ga} (g, h)$ \hspace{1cm} with $g, h$ unknown.

(ii) $p(x_i|\theta) = \text{Ex} (\theta)$ \hspace{1cm} (i = 1, 2, ..., n),
$p(y|\theta) = \text{Ex} (\theta),
p(\theta) = \text{Ga} (g, h)$ \hspace{1cm} with $g, h$ unknown.

7.5 In an effort to cut costs a builder is considering whether he can install smaller domestic storage tanks than the ones he uses at present in his houses. He wishes to decide the minimum size of tank that he can reasonably install in his standard 3-bedroomed houses. The water supply to the storage tank is through a valve which opens whenever water is drawn off and closes when the level has been restored to the 'Full' mark. The supply rate is constant when in use and is such that for the domestic demand pattern he is considering the tanks at present in use never became empty and each day began at the 'Full' mark. For a random selection of his standard houses he monitors the water usage by measuring the maximum drop in level from the 'Full' mark on 365 days. The results set out as an ordered sample consist of values 6.1, 6.5, 6.5, ..., 47.5, 50.1, 54.3, 55.9, 57.1 litres. Can you assist the builder to reduce his storage tank size?
8

Sampling inspection

8.1 Introduction

The myriad of possible statistical sampling inspection procedures forces us to consider in detail only a very small selection in a book of this size. We would need a separate book to do justice to the huge variety of plans. In this chapter therefore we show how some standard plans come within the framework of decisive prediction, and how the framework can readily cope with less standard problems. The application of prediction theory to this area will provide some additional justification and motivation for some of these plans. We hope that those selected will be sufficient to indicate the direction of analysis to any reader with a specific problem.

We consider both fixed size sample and sequential sampling schemes. Wetherill (1966) and Wetherill and Campling (1966) also provide a decision theory approach to sampling inspection but do not consider predictive distributions.

8.2 Fixed-size single-sample destructive testing

We consider first a fixed-size single-sample plan for deciding whether to accept or reject a batch. For a process which produces an item at each of a number of independent operations we may imagine as our basic future experiment the determination of the quality \( y \) of a single item. This quality \( y \) may be a simple counting variable taking the value 1 for an effective and 0 for a defective item, or may be more sophisticated, for example the lifetime of a component or the degree of purity of a chemical preparation. We suppose that the probabilistic mechanism which describes the production of the variable \( y \) is a density function \( p(y \mid \theta) \) on \( Y \) where, as in previous work, \( \theta \) is an indexing parameter with density function \( p(\theta) \). We suppose further that altogether \( N \) items have been produced by independent operations of the system and that the informative experiment consists of the destructive testing of \( n \) components, whose qualities \( x_1, \ldots, x_n \) are determined by the testing. The informative experiment \( e \) is then described by the density function

\[
p(x \mid \theta) = p(x_1 \mid \theta) \cdots p(x_n \mid \theta).
\]
The utility structure is such that the realised utility depends on the as yet unknown qualities $v_1, ..., v_{N-n}$ of the untested items. It also depends on which of the available possible actions is adopted. For simplicity we suppose that the action space $A$ consists of just two possible actions:

\[ a_1: \text{ market the untested items, } \]
\[ a_2: \text{ scrap the untested items. } \] (8.1)

We can thus express the utility in the form $U(a, y_1, ..., y_{N-n})$ or $U(a, y)$. Note that we have not made allowance in this basic utility for the cost of testing the items. This is not necessary at the terminal stage of determining the appropriate action after the testing is complete. It is, however, very relevant to the preposterior analysis for deciding how many items should be tested, and will appear in §8.4.

The inferential aspect of the problem produces the predictive density function

\[ p(y_1 | x) = f \prod_{i=1}^{N-n} p(y_i | \theta) \, p(\theta | x) \, d \theta. \] (8.2)

Note that the distribution of $y_1, ..., y_{N-n}$ for given $x$ is not necessarily a product distribution. In general the function $p(y_1 | x)$ will not be the product of component functions $p(y_1 | x), ..., p(y_{N-n} | x)$.

The optimum action is then easily decided. According to standard statistical decision theory we should take that action which maximises

\[ U(a) = \int_Y U(a, y) \cdot p(y | x) \, dy, \] (8.3)

where the bold $dy$ and $Y$ indicate that integration is over the $(N-n)$-dimensional space, the product space of $N-n$ identical spaces $Y_1, ..., Y_{N-n}$, each of type $Y$.

A great simplification occurs if the utilities are additive over items; in other words, if we can envisage a component utility attaching to each item, this utility depending only on the quality of the item, and if the utility of the batch of $N-n$ items is the sum of these component utilities. Mathematically, if $U(a, y)$ denotes the utility of a component of quality $y$ then

\[ U(a, y) = U(a, y_1) + ... + U(a, y_{N-n}). \] (8.4)

Such a utility function would of course not apply if the utility of the batch depended on some overall property of the batch, for example, if it were of value only if it contained at most $d$ defective items.

The simplification with utility function (8.4) arises because we can rewrite (8.3) as
The contribution from the first term of the sum can (provided the change of order of operations is allowable) be written as

\[
\int_{Y_1} U(a, y_1) dy_1 \int_\Theta p(y_1 | \theta) p(\theta | x) d\theta
\]

\[
\times \int_{Y_2, Y_3, \ldots, Y_{N-n}} p(y_2 | \theta) \ldots p(y_{N-n} | \theta) dy_2 \ldots dy_{N-n}
\]

\[
= \int_{Y_1} U(a, y_1) p(y_1 | x) dy_1.
\]

Hence

\[
U(a) = (N - n) \int_Y U(a, y) p(y | x) dy. \tag{8.6}
\]

Example 8.1

Batch acceptance. A manufacturer must decide whether to market or scrap a batch of \(N\) items which he has produced. By destructive testing he determines the nature \(x_1, x_2, \ldots, x_n\) of a random sample of \(n\) items, each being classified as either effective \((x_i = 1)\) or defective \((x_i = 0)\). If \(x = \sum x_i\) records the number of effective items a suitable description is a binomial trials model with \(p(x | \theta) = B(n, \theta)\). The acceptability of the remaining \(N - n\) items in the batch for a customer depends on the characteristics \(y_1, y_2, \ldots, y_{N-n}\) of the items, where

\[
p(y_i = 1 | \theta) = \theta \quad \text{and} \quad p(y_i = 0 | \theta) = 1 - \theta \quad (i = 1, 2, \ldots, N - n).
\]

If we take \(p(\theta) = B(g, h)\) it follows from table 2.3 and (8.2) that the predictive distribution for \(y = (y_1, y_2, \ldots, y_{N-n})\) is given by

\[
p(y | x) = \frac{B(G + \sum y_i, H + N - n - \sum y_i)}{B(G, H)}
\]

\[
(y: y_i = 0, 1; \; i = 1, 2, \ldots, N - n), \tag{8.7}
\]

where \(G = g + x, H = h + n - x\).

Suppose that the customer considers the batch effective only if there are at least \(q\) effective items remaining, and that the manufacturer provides a 'money-back' guarantee if the batch does not meet this specification. Then
for the two actions (8.1) the manufacturer assesses his gains as follows:

\[
U(a_1, y) = \left\{ \begin{array}{ll}
K_1 & \text{if } \sum_{i=1}^{N-n} y_i \geq q, \\
-K_2 & \text{if } \sum_{i=1}^{N-n} y_i < q, 
\end{array} \right.
\]

\[
U(a_2, y) = -L,
\]

where \(K_1, K_2, L\) are non-negative constants. We then have

\[
U(a_1) = (K_1 + K_2) \sum_{i=0}^{N-n-q} \binom{N-n}{i} \frac{B(G + N - n - i, H + i)}{B(G, H)} - K_2,
\]

and

\[
U(a_2) = -L.
\]

We select the action giving \(\max \{U(a_1), U(a_2)\}\). Notice that we need investigate only the case \(L < K_2\); for otherwise the optimal action would be to market every batch irrespective of the \(x\)-value, thus neglecting any information obtained from sampling.

Consider now a numerical solution. Suppose that \(N = 100, n = 10\) and \(q = 80\). Also suppose that the manufacture and testing of the batch costs £40 and the selling price of the batch is £60. Thus \(K_1 = 20\) and \(K_2 = 40\). If the batch is rejected the manufacturer can sell it for scrap for £1; hence \(L = 39\). The manufacturer has been operating the process for some time and assesses his knowledge of \(\theta\) by \(p(\theta) = Be(9, 1)\). This gives a prior mean success rate of 0.90 and, for example, \(P(\theta > 0.7) = 0.96\). This indicates that the manufacturer has fairly strong prior ideas about \(\theta\).

The values of \(U(a_1)\) for the possible \(x\)-values are shown in table 8.1. Our optimal action is to market the batch if we get 5 or more effective items in the sample; otherwise scrap the batch. Notice how critical the scrapping loss is in this situation.

The manufacturer here has a good deal of prior information on \(\theta\) and this affects his decision rule to a considerable degree. He is able to allow up to 5 defective items in a sample of 10 before he scraps the batch because he is so confident that \(\theta\) will be large; and the probability of obtaining as low an \(x\)-value as 5 consequently will be very small. If the process is new and the manufacturer thinks any value of \(\theta\) equally plausible, so that \(p(\theta) = Be(1, 1)\), the corresponding values of \(U(a_1)\) are also shown in table 8.1, and the optimal action is to market the batch only if 7 or more effective items are found.

The expected utilities shown in table 8.1 may seem somewhat strange and indicate that the manufacturer has little time left before going out of business!
Before he performs his experiment his ideas about the value of \( x \) he is likely to obtain are summarised by the prior predictive distribution

\[
p(x) = \int_{\Theta} p(x|\theta) p(\theta) d\theta.
\]  

(8.10)

Here this is given by \( \text{BeBi}(10, g, h) \). If the manufacturer follows his optimal choice he will expect to gain

\[
\sum_{x=0}^{10} \max \{ U(a_1), U(a_2) \} p(x).
\]  

(8.11)

For the two prior functions considered, namely \( \text{Be}(9, 1) \) and \( \text{Be}(1, 1) \) the prior predictive densities \( p(x) \) are shown in table 8.1, and the expected utilities (8.11) are 0.20 and -32.18 respectively. The foolhardiness of the manufacturer's venture is amply demonstrated in the latter case.

The above serves as an introduction to preposterior analysis. We return to this concept in §8.4.

Suppose that instead of selling the remainder of the batch as a whole, the items are sold individually with a double-your-money back guarantee for an item which is defective. The utility to the manufacturer may then be different and a sensible assessment would be of the form (8.4) with components

\[
U(a_1, y_i) = \begin{cases} k_1 & \text{if } y_i = 1, \\ -k_2 & \text{if } y_i = 0, \end{cases}
\]  

(8.12)

\[
U(a_2, y_i) = -l,
\]
Sampling inspection

\((i = 1, 2, \ldots, N - n)\), where \(k_1, k_2, l\) are non-negative constants. Thus if he sells an effective item he will make a profit of \(k_1\) per item; if he markets a defective item he has to return twice the selling price with a corresponding loss of \(k_2\) per item. Again there is a scrapping loss, assessed as \(l\) per scrapped item. Thus each individual item makes its own independent contribution to the overall utility of the action. Now

\[
p(y_i|x) = \begin{cases} \frac{G}{G + H} & \text{if } y_i = 1, \\ \frac{H}{G + H} & \text{if } y_i = 0, \end{cases}
\]

and so

\[
U(a_1) = \frac{(N-n)(k_1G-k_2H)}{G + H} = (N-n) \left( \frac{(k_1 + k_2)G}{G + H} - k_2 \right).
\]

\[
U(a_2) = -(N-n)l.
\]

Here it is sensible to assume \(l < k_2\).

The determination of the optimal action follows directly. Consider again a numerical solution with \(N = 100, n = 10\). Suppose that each item costs £0.40 to produce and is sold at £0.60; hence \(k_1 = 0.2, k_2 = 1\). A rejected item may be sold for scrap for £0.05, giving \(l = 0.35\). (These figures are not necessarily compatible with the values for \(K_1, K_2, L\) assumed earlier.) Again we take \(p(\theta) = \text{Be}(9, 1)\). Table 8.2 provides the values of \(U(a_1)\) for the possible outcomes \(x\) of the informative experiment. The optimal terminal decision rule is therefore the following.

If \(x = 0, 1\), scrap the batch;
otherwise, market the batch.

Table 8.2 Values of \(U(a_1)\), the expected utility (8.14), for batch acceptance with \(k_1 = 0.2, k_2 = 1.0\), corresponding to the two cases \(\text{Be}(9, 1)\) and \(\text{Be}(1, 1)\) for \(p(\theta)\)

<table>
<thead>
<tr>
<th>(x)</th>
<th>(U(a_1)) corresponding to (p(\theta) = \text{Be}(9, 1))</th>
<th>(U(a_1)) corresponding to (p(\theta) = \text{Be}(1, 1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-41.4</td>
<td>-81</td>
</tr>
<tr>
<td>1</td>
<td>-36.0</td>
<td>-72</td>
</tr>
<tr>
<td>2</td>
<td>-30.6</td>
<td>-63</td>
</tr>
<tr>
<td>3</td>
<td>-25.2</td>
<td>-54</td>
</tr>
<tr>
<td>4</td>
<td>-19.8</td>
<td>-45</td>
</tr>
<tr>
<td>5</td>
<td>-14.4</td>
<td>-36</td>
</tr>
<tr>
<td>6</td>
<td>-9.0</td>
<td>-27</td>
</tr>
<tr>
<td>7</td>
<td>-3.6</td>
<td>-18</td>
</tr>
<tr>
<td>8</td>
<td>1.8</td>
<td>-9</td>
</tr>
<tr>
<td>9</td>
<td>7.2</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>12.6</td>
<td>9</td>
</tr>
</tbody>
</table>
This may seem a somewhat unlikely optimal rule. As before, the prior distribution is playing a dominant role—we are so sure $\theta$ will be large (and consequently the batch will be acceptable) that only the most extreme values of $x$ will cause us to scrap the batch. Again as comparison we list the values of $U(a_1)$ in Table 8.2 for the uniform prior $p(\theta) = \text{Be}(1, 1)$. The terminal decision rule for this prior would be the following.

If $x = 0, 1, ..., 5$, scrap the batch;
otherwise, market the batch.

### 8.3 Role of mean coverage and guaranteed coverage tolerance predictors in sampling plans

The decision-theoretic approach of §8.2 to sampling inspection throws some light on commonly suggested sampling plans in quality control. Such sampling plans usually make the demand that for a batch to be marketable some statistical tolerance limit constructed from a random sample from the batch should meet some quality requirement such as the exceeding of some specified quality.

**Mean coverage predictors.** Suppose that there is a critical quality level $q$ for each item. For any marketed item with quality level $q$ or more there is a profit of $k_1$, whereas for any marketed item with quality level below $q$ there is a loss of $k_2$. The loss involved in scrapping a batch is $l$ per item. Thus we can specify the utility structure (8.4) with components

$\begin{align*}
U(a_1, y_i) &= \begin{cases} k_1 (y_i \geq q), \\ -k_2 (y_i < q), \end{cases} \\
U(a_2, y_i) &= -l \\
&\quad (i = 1, 2, ..., N-n).
\end{align*}$

This is simply a generalisation of utility function (8.12). It follows that

$\begin{align*}
U(a_1) &= (N-n) \left( (k_1 + k_2) \int_q^\infty p(y|x) dy - k_2 \right) \\
U(a_2) &= -(N-n)l.
\end{align*}$

We shall take a decision to market if and only if $U(a_1) > U(a_2)$, that is, if and only if

$\int_q^\infty p(y|x) dy > \frac{k_2 - l}{k_1 + k_2}.$
Notice we again need to assume \( l < k_2 \) for a realistic situation. The inequality (8.17) can be expressed more familiarly by noting that it is equivalent to the lower Bayesian informative predictor of cover \((k_2 - l)/(k_1 + k_2)\) computed from the sample exceeding the critical quality \( q \); in other words, if we used an informative prediction interval \((q, \infty)\) for the quality of a future item the Bayesian cover provided by it would be at least \((k_2 - l)/(k_1 + k_2)\). The importance of this result is that it gives a tangible meaning of the cover associated with the sampling plan in terms of more directly assessable profits and losses; cf. §7.2.

**Guaranteed coverage predictors.** Suppose that the batch is large and that it is effective if and only if the proportion of items in the batch of quality \( q \) or more is at least \( c \). Under such circumstances it is then necessary to employ the alternative utility formulation of §3.6. If the profit from an effective batch is \( K_1 \), the loss from a defective batch \( K_2 \), and the loss from a scrapped batch \( L \) then we can set

\[
V(a_1, \theta) = \begin{cases} 
K_1 & \text{if } P_f\{ (q, \infty) | \theta \} > c, \\
-K_2 & \text{otherwise.}
\end{cases} \tag{8.18}
\]

This is a generalisation of the situation covered by utility function (8.8). Then

\[
U(a_1) = \int_\theta V(a_1, \theta) p(\theta | x) d\theta 
= (K_1 + K_2) \int_{\{ \theta : P_f( (q, \infty) | \theta ) > c \}} p(\theta | x) d\theta - K_2, \tag{8.19}
\]

and

\[
U(a_2) = \int_\theta V(a_2, \theta) p(\theta | x) d\theta 
= -L.
\]

Hence we should market if and only if

\[
P[ \theta : P_f\{ (q, \infty) | \theta \} > c | x ] > \frac{K_2 - L}{K_1 + K_2}. \tag{8.20}
\]

We can express this marketing rule very simply in terms of what may be called a Bayesian guaranteed coverage tolerance limit. In chapters 4 and 6 we did not define such an interval. However the extension to such a concept is obviously straightforward. On the basis of the sample information \( x \) construct a lower \((c, g)\) Bayesian guaranteed coverage limit with \( g = (K_2 - L)/(K_1 + K_2)\).
Sampling inspection

If the tolerance limit falls below the critical quality level \( q \) the batch should be marketed; otherwise it should be scrapped.

8.4 Optimum choice of fixed sample size

In our study of sampling plans so far we have considered only the terminal stage of the analysis, that is the problem of determining the appropriate action given the sample design and after the testing is complete. The choice of experiment or sampling design is obviously important and we now investigate this preposterior analysis.

We will again confine our attention to the fixed sample size destructive testing design of §8.2. The problem is to select the size \( n \) of the sample which we should take in the informative experiment. An extension of the notation is required. Suppose we rewrite \( U(a, y) \) as \( U(n, x, a, y) \), that is the utility of drawing a sample of size \( n \), observing outcomes \( x = (x_1, x_2, ..., x_n) \) and choosing action \( a \) when \( y = (y_1, y_2, ..., y_{N-n}) \) is the outcome of the future experiment. The terminal analysis leads us to select the action which maximises

\[
U(n, x, a, y) = \int_y U(n, x, a, y) p(y|\theta) d\theta.
\]

(Strictly we should indicate the dependence of \( p(y|x) \) on \( n \) by writing \( p(y|x, n) \). However we shall retain our usual notation for the predictive density function.) Thus the expected utility of sampling \( n \) items and observing \( x \) is given by

\[
U(n, x) = \max_{a \in A} U(n, x, a).
\]

In a fixed sample size plan we must select \( n \) before experimentation. Although we do not know at that stage which \( x \) will obtain, we do have a distribution over the possible values given by

\[
p(x) = \int_\theta p(x_1|\theta) ... p(x_n|\theta) p(\theta) d\theta.
\]

We may therefore evaluate the expected utility of performing an informative experiment with sample size \( n \), namely

\[
U(n) = \int_x U(n, x) p(x) dx.
\]

This preposterior analysis supplies the optimal size of sample—the value of \( n \) which maximises \( U(n) \).

Although no theoretical problems present themselves in the preposterior analysis, the actual determinations of the optimal sample size in a practical
situation may be tedious. As an illustration we investigate the preposterior analysis for example 8.1.

Example 8.1 (continued)

Consider the case of selecting a sample of size $n$ from the batch of $N$ items, and observing $x = (x_1, x_2, ..., x_n)$. Suppose that we consider the case where the remaining items are sold individually so that we may take as our utility function

$$U(n, x, a_j, y) = \sum_{i=1}^{N-n} U(n, x, a_j, y_i) - \gamma n,$$  \hspace{1cm} (8.25)

where $U(n, x, a_j, y_i)$ is given in (8.12) ($i = 1, 2, ..., N - n$) and where $x = \Sigma x_j$. The utility function now includes a factor $\gamma n$, with $\gamma > 0$, which gives a measure of the cost of sampling $n$ items. Thus $\gamma$ may be termed the cost per item sampled. Inevitably this involves some compromise between the gains and losses accrued through the information obtained by destroying a defective item on the one hand and a good item on the other. From (8.21) we have the generalisation of (8.14):

$$U(n, x, a_1) = (N - n) \left( \frac{(k_1 + k_2)(g + x)}{g + h + n} - k_2 \right) - \gamma n,$$

$$U(n, x, a_2) = -l(N - n) - \gamma n,$$  \hspace{1cm} (8.26)

so that the terminal analysis yields

$$U(n, x) = \begin{cases} U(n, x, a_1) & \text{if } n \geq x \geq Q, \\ U(n, x, a_2) & \text{otherwise}, \end{cases}$$  \hspace{1cm} (8.27)

where

$$Q = \frac{(g + h + n)(k_2 - l)}{k_1 + k_2} - g.$$  \hspace{1cm} (8.28)

Notice that if $Q \leq 0$ we will always take action $a_1$, and if $Q > n$ the action $a_2$.

The prior predictive distribution $p(x)$ given by (8.23) is BeBi $(n, g, h)$.

Substitution in (8.24) yields

$$U(n) = \begin{cases} \frac{(N - n)(k_1 g - k_2 h)}{g + h} - \gamma n & \text{if } Q \leq 0, \\ U^*(n) - \gamma n & \text{if } 0 < Q \leq n, \\ -l(N - n) - \gamma n & \text{if } Q > n, \end{cases}$$  \hspace{1cm} (8.29)
where

\[ U^*(n) = -(N-n) \left( \frac{(k_1 + k_2)g}{g + h + n} - k_2 + l \right) \sum_{x=0}^{Q_1} p(x) \]

\[ + \frac{(N-n)(k_1 + k_2)}{g + h + n} \sum_{x=0}^{Q_1} xp(x) + \frac{(N-n)(k_1 g - k_2 h)}{g + h} \]

(8.30)

Table 8.3 Values of \( U(n) \), the expected utility (8.29) of sample size \( n \) corresponding to the two cases \( Be(9, 1) \) and \( Be(1, 1) \) for \( p(\theta) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( U(n) ) corresponding to ( p(\theta) = Be(9, 1) )</th>
<th>( Be(1, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.00</td>
<td>-35.00</td>
</tr>
<tr>
<td>1</td>
<td>7.92 - ( \gamma )</td>
<td>-27.23 - ( \gamma )</td>
</tr>
<tr>
<td>2</td>
<td>7.84 - 2( \gamma )</td>
<td>-26.13 - 2( \gamma )</td>
</tr>
<tr>
<td>3</td>
<td>7.76 - 3( \gamma )</td>
<td>-24.74 - 3( \gamma )</td>
</tr>
<tr>
<td>4</td>
<td>7.68 - 4( \gamma )</td>
<td>-24.00 - 4( \gamma )</td>
</tr>
<tr>
<td>5</td>
<td>7.60 - 5( \gamma )</td>
<td>-23.41 - 5( \gamma )</td>
</tr>
<tr>
<td>6</td>
<td>7.52 - 6( \gamma )</td>
<td>-22.83 - 6( \gamma )</td>
</tr>
<tr>
<td>7</td>
<td>7.44 - 7( \gamma )</td>
<td>-22.48 - 7( \gamma )</td>
</tr>
<tr>
<td>8</td>
<td>7.36 - 8( \gamma )</td>
<td>-21.98 - 8( \gamma )</td>
</tr>
<tr>
<td>9</td>
<td>7.28 - 9( \gamma )</td>
<td>-21.72 - 9( \gamma )</td>
</tr>
<tr>
<td>10</td>
<td>7.20 - 10( \gamma )</td>
<td>-21.27 - 10( \gamma )</td>
</tr>
<tr>
<td>11</td>
<td>7.12 - 11( \gamma )</td>
<td>-21.02 - 11( \gamma )</td>
</tr>
<tr>
<td>12</td>
<td>7.04 - 12( \gamma )</td>
<td>-20.65 - 12( \gamma )</td>
</tr>
<tr>
<td>13</td>
<td>6.96 - 13( \gamma )</td>
<td>-20.38 - 13( \gamma )</td>
</tr>
<tr>
<td>14</td>
<td>6.88 - 14( \gamma )</td>
<td>-20.07 - 14( \gamma )</td>
</tr>
</tbody>
</table>

If we take the same numerical example as in §8.2, with \( N = 100, k_1 = 0.2, k_2 = 1, l = 0.35, g = 9, h = 1 \), we find that, provided \( \gamma > 0 \), it is not worth while to sample the batch. The optimal value of \( n \) is 0. We have that for \( n = 0, 1, 2, ..., 6, Q < 0 \) anyway, and so in such cases action \( a_1 \) is optimal. The values of \( U(n) \) are shown in table 8.3 for \( n = 0, 1, ..., 14 \), so that the overall expected gain to the manufacturer is £8.00.

For comparison we again look at the case where \( p(\theta) = Be(1, 1) \). Table 8.3 gives the corresponding values of \( U(n) \). Suppose the manufacturer assesses \( \gamma = 0.40 \). We have that the optimal sample size is \( n = 8 \) and the batch is marketed only if 5 or more effective items are observed. Notice however that the manufacturer will always expect to lose, and so an investigation of his costs, his guarantee, his prior information and his plant is in order!

### 8.5 Sequential predictive sampling inspection

One possible alternative to the fixed-size, single-sample destructive testing plan for batch acceptance or rejection is a sequential scheme. Here we test one item at a time, and after each trial we assess the situation and select one of
three possible actions, namely,

\[ a_1 : \text{market the untested items}, \]
\[ a_2 : \text{scrap the untested items}, \]
\[ a_3 : \text{test another item}. \]

The utility structure will depend on the number \((N - n)\) of items left and we explicitly show this by writing the utility in the form \(U(a, y_{N-n})\), where \(y_{N-n} = (y_1, y_2, \ldots, y_{N-n})\). (Note that we have reverted to our shortened notation of \(\S 8.2, 8.3\).) As in \(\S 8.4\) there is a cost \(\gamma\) attached to testing an item.

Let \(F_n(x_n)\) denote the maximum expected gain from pursuing an optimal policy after \(n\) items have been tested with results \(x_n = (x_1, x_2, \ldots, x_n)\). Then the principle of optimality leads to the following relationships.

\[
F_n(x_n) = \max \left\{ \int \int \cdots \int_{Y_1, X_2, \ldots, Y_{N-n}} U(a_1, y_{N-n}) p(y_{N-n} | x_n) dy_{N-n}, \right. \\
\left. \int \int \cdots \int_{Y_1, X_2, \ldots, Y_{N-n}} U(a_2, y_{N-n}) p(y_{N-n} | x_n) dy_{N-n}, \right. \\
\left. \int_{Y} F_{n+1} \{(x_n, y)\} p(y | x_n) dy - \gamma \right\} \\
(n = 0, 1, 2, \ldots, N-1),
\]

\[(8.31)\]

Again considerable simplifications occur if the utilities are additive over items as in \(\S 8.4\) of \(\S 8.2\). For then

\[
F_n(x_n) = \max \left\{ (N-n) \int_{Y} U(a_1, y) p(y | x_n) dy, \right. \\
\left. (N-n) \int_{Y} U(a_2, y) p(y | x_n) dy, \int_{Y} F_{n+1} \{(x_n, y)\} p(y | x_n) dy - \gamma \right\}.
\]

\[(8.32)\]

**Example 8.1 (continued)**

Suppose that we consider testing the batch of size \(N\) with a sequential sampling scheme and that utility specification \((8.12)\) is appropriate. Then, writing \(x = \sum_{i} x_i\), we have from \(8.32)\)
Sampling inspection

\[ F_n(x) = \max \left( (N - n) \left( \frac{(k_1 + k_2)G}{G + H} - k_2 \right), \right. \\
\left. -(N - n) l, \frac{H}{G + H} F_{n+1}(x) + \frac{G}{G + H} F_{n+1}(x + 1) - \gamma \right) \]

\( (n = 0, 1, ..., N - 1; x = 0, 1, ..., n), \)

\( F_N(x) = 0, \quad (8.33) \)

with \( G = g + x, H = h + n - x. \) The optimal strategy must then be obtained by a standard dynamic programming technique.

To provide a numerical solution we take the case where \( N = 20, \) the other constants being as in § 8.2, 8.4, namely \( k_1 = 0.2, k_2 = 1.0, l = 0.35, \gamma = 0.40. \) Consider first the case where \( p(\theta) = \text{Be}(1, 1). \) From (8.33) we have

\[ F_n(x) = \max \left[ (20 - n) \left( \frac{1.2(1 + x)}{n + 2} - 1 \right), -0.35(20 - n), \right. \\
\left. \frac{n + 1 - x}{n + 2} F_{n+1}(x) + \frac{x + 1}{n + 2} F_{n+1}(x + 1) - 0.40 \right] \]

\( (n = 0, 1, ..., 19; x = 0, 1, ..., n), \)

\( F_{20}(x) = 0. \quad (8.34) \)

Fig. 8.1 shows the optimal actions for each possible \((n, x)\) position. The procedure is to start at the origin and continue sampling until a boundary is reached. The path followed moves at each step either horizontally along one square for a defective item or diagonally upwards across one square for an effective item.

For the case where \( p(\theta) = \text{Be}(9, 1) \) it turns out that the optimal action is to accept the batch without sampling; compare the similar situation in § 8.4 with \( N = 100. \) For the more uncertain prior situation just considered it is worth while obtaining some information by testing before coming to a decision.

History

There is a vast literature on sampling inspection models, a not negligible proportion of which is concerned with Bayesian models; for example Guthrie and Johns (1959), Lindley and Barnett (1965) and Wetherill and Campling (1966). In the decision theoretic models the tendency is to work with utility.
specifications of the form \( V(a, \theta) \), see §3.6, and not to consider explicitly the concept of predictive distributions. Draper and Guttman (1968a, 1968b) use predictive distributions in their Bayesian model, albeit only for vague prior distributions.

Wetherill (1966) gives a good review of sequential decision models.
Sampling inspection

Problems

8.1 Follow through the terminal and preposterior analysis of Example 8.1 if \( p(x|\theta) \) is \( \text{No}(\mu, \tau) \) instead of binomial and if utility specification (8.15) holds.

8.2 A manufacturer must decide whether to market or destroy a batch of \( N \) machine tools which he has produced. The suitability of the machine tools in a batch for a customer depends on their lifetimes. The manufacturer gets some information from the destructive testing of \( n \) machine tools whose lifetimes turn out to be \( x_1, x_2, \ldots, x_n \), these being independent observations on an \( \text{Ex}(\theta) \) random variable. The lifetimes of the remaining machine tools remain unknown however. It is recognised that \( \theta \) may vary from batch to batch, a suitable description of this variation being \( p(\theta) = \text{Ga}(g, h) \). A machine tool is considered of beneficial use to the customer only if its lifetime is greater than \( q \); the contract specifies therefore the replacement of non-useful components. The manufacturer's assessment of this contract yields the utility structure (8.15). Should he market the batch?

Suppose we now consider the choice of sample size \( n \). Determine the optima sample sizes for each of the following specifications of the constants:

\[
\begin{align*}
N &= 100, \quad g = 1, \quad h = 10, \quad q = 1, \\
k_1 &= 1, \quad k_2 = 4, \quad l = 0, \\
\text{cost per item sampled } \gamma &= 1.0, 0.5, 0.1, 0.05, 0.01.
\end{align*}
\]

8.3 Consider again the batch acceptance situation of example 8.1. Suppose that the sample taken is of size \( n_1 \), and that the manufacturer can either accept or reject the batch or can take a further sample of size \( n_2 \) before deciding. With an additive utility structure

\[
\begin{align*}
U(\text{accept}, y_i) &= \begin{cases} 
k_1 & \text{if } y_i = 1, \\
-k_2 & \text{if } y_i = 0,
\end{cases} \\
U(\text{reject}, y_i) &= -l,
\end{align*}
\]

for the remaining items together with a cost \( \gamma \) per sampled item, which action would you take in this double-sampling scheme if the first sample yields observations \( x_1, x_2, \ldots, x_n \) ?

8.4 Reconsider example 8.1 as a sequential prediction sampling inspection problem with the following changes.

\[
N = 10, \quad k_1 = 0.2, \quad k_2 = 1.0, \quad l = 0.35,
\]
Sampling inspection

(i) $\gamma = 0.40, \ g = 1, \ h = 1,$
(ii) $\gamma = 0.40, \ g = 9, \ h = 1,$
(iii) $\gamma = 0.30, \ g = 1, \ h = 1,$
(iv) $\gamma = 0.30, \ g = 9, \ h = 1.$

Compare your solutions with the optimum fixed size sample tests.
9
Regulation and optimisation

9.1 Introduction

In the next three chapters we focus attention on some predictive problems which occur in essentially regression-type situations. The distinctive features of regulation, optimisation, calibration and diagnosis have already been indicated in examples 1.4-1.7 in chapter 1 and are formally recognised in the classification of prediction problems in appendix II. In these regression-type situations we envisage a typical experiment $f_t$, performed at the particular value $t$ of an 'independent' variable and resulting in the observation of a $y$-value. There is thus a class $F$ of 'future' experiments:

$$F = \{ f_t : t \in T \},$$

where each $f_t$ has the same outcome space $Y$ and $T$ acts as index set for the class $F$. The informative experiment $e$ consists of a single performance of each of the $n$ independent experiments $f_1, f_2, \ldots, f_n$, and we shall denote the set of outcomes by $x = (x_1, \ldots, x_n)$, and write $t = (t_1, \ldots, t_n)$, $z = \{(t_1, x_1), \ldots, (t_n, x_n)\}$ as in §2.5.

In a problem of control type we aim to obtain a specific $y$-value, $y = y_0$ say, and require to find a corresponding $t$-value which is optimum in some defined sense. In particular, in problems of regulation we know the value $y_0$ for which we are aiming and wish to regulate or control the $t$-value in an effort to attain this $y_0$. In optimisation problems we wish to choose the $t$-value in order to maximise (minimise) the $y$-value subject to constraints and are usually unaware of the optimal value. In problems of calibration and diagnosis we are given the $y$-value, and require some point or interval estimate for the corresponding $t$-value. The main difference between calibration and diagnosis on the one hand and regulation and optimisation on the other is that for the latter cases the choice of $f_t$ is ours, whereas for calibration and diagnosis we have no control over which member of $F$ is performed.

In this chapter we deal with regulation and optimisation problems. Since we are going to carry out a particular experiment $f_t$ in the future (on the assumption that there is a unique optimal $t$-value), we are concerned only with point prediction. There is no advantage in providing an interval or set estimate, and indeed such a procedure may well lead to confusion.
The derivation of the appropriate predictive density function \( p(y | t, z) \) performs the information-extraction process in the problem. As illustrated in §2.5 we can use Table 2.3 to obtain these densities.

9.2 Point regulation

We again present the problem as one of statistical decision theory with the following components.

1. **Parameter space.** \( Y \) plays the role of the set of unknown parameters, the unknown outcome of the future experiment \( f_t \) playing the role of an unknown state of nature. The predictive density \( p(y | t, z) \) provides an assessment of the plausibility of a particular \( y \).

2. **Action set.** The action set \( A \) is simply the set \( T \), since we can choose to perform the future experiment at any \( t \in T \).

3. **Utility function.** A sensible utility specification is one which for given \( t \) compares the \( y \)-value obtained with the optimal \( y_0 \). We thus obtain a utility structure \( U(t, y, y_0) \) by defining a function on the product domain \( A \times Y \times Y_0 \).

We require to maximise the expected utility

\[
U(t, y_0) = \int_Y U(t, y, y_0) p(y | t, z) dy.
\] (9.1)

Since we have complete control over the choice of future experiment \( f_t \) we simply maximise \( U(t, y_0) \) with respect to \( t \) to obtain the best member \( f* \) of \( F \) to perform. Thus

\[
U(f*, y_0) = \max_{F} U(t, y_0). \quad (9.2)
\]

We now mention briefly a few simple utility functions \( U(t, y, y_0) \) suitable for some regulation problems. Since the aim in regulation is to perform an experiment \( f_t \) which will give us a value of \( y \) very close to a specified \( y_0 \), we shall naturally attach penalties to values of \( y \) which differ from \( y_0 \).

**All-or-nothing utility.** If it is imperative that the \( y \)-value obtained should be very near \( y_0 \), the natural formulation for the utility is as the limiting case \((\epsilon \to 0)\) of utility function:

\[
U(t, y, y_0) = \begin{cases} 
1 & (y_0 - \epsilon \leq y \leq y_0 + \epsilon), \\
0 & \text{otherwise}.
\end{cases} \quad (9.3)
\]

Then

\[
U(t, y_0) = \int_{y_0-\epsilon}^{y_0+\epsilon} p(y | t, z) dy \\
= 2\epsilon p(y_0 | t, z) + o(\epsilon) \quad (\epsilon \to 0) \quad (9.4)
\]
provided certain simple regularity conditions on \( p \) are satisfied. We therefore
need to maximise \( p(y_0|t, z) \) with respect to \( t \). This is intuitively reasonable
and is analogous to maximum likelihood estimation. We are selecting the \( t \)-
value which gives most support for \( y_0 \) in the sense that it produces the largest
predictive probability of obtaining \( y_0 \).

**Linear utility.** Often a more realistic utility function is the piecewise linear
type:

\[
U(t, y, y_0) = \begin{cases} 
-\xi(y_0 - y) & (y < y_0), \\
-\eta(y - y_0) & (y > y_0), 
\end{cases}
\]

where \( \xi, \eta > 0 \). We see that \( U(t, y_0, y_0) \) is the maximum value. Then

\[
U(t, y_0) = -\xi \int_{-\infty}^{y_0} (y_0 - y) p(y|t, z) \, dy
- \eta \int_{y_0}^{\infty} (y - y_0) p(y|t, z) \, dy
= -\xi \int_{-\infty}^{y_0} (y_0 - y) p(y|t, z) \, dy
- \eta \{E(y|t, z) - y_0\}. \tag{9.6}
\]

Our task is to select the value \( t^* \) of \( t \) which maximises \( U(t, y_0) \). Unlike the
direct case of linear-loss point prediction (§3.1) no simple interpretation, such
as a quantile of the predictive density function, is available in the case of
linear utility regulation.

**Quadratic loss.** Here

\[
U(t, y, y_0) = -(y - y_0)^2. \tag{9.7}
\]

Then

\[
U(t, y_0) = -V(y|t, z) - (\{E(y|t, z) - y_0\})^2, \tag{9.8}
\]

and we must again choose the \( t^* \) value of \( t \) which maximises \( U(t, y_0) \).

**Cost function.** A function \( K(t) \) which takes into account the cost of performing
the future experiment at value \( t \), but which we assume to be independent of
\( y \), may easily be incorporated into the utility specification.

The method of subsequently maximising \( U(t, y_0) \) must depend largely on
its form and on the computing facilities available. In table 9.1 we list the
optimum \( t \)-values in cases where simple formulations are available. In other
cases, for example where the predictive density function is Student, iterative
or search techniques are required, and we provide such an application in
| $p(y|t, z)$ | Utility function | Linear-loss | Quadratic-loss |
|------------|-----------------|-------------|---------------|
| BeBi($t, G, H$) | $\frac{y_0}{G} (G + H - 1)$ | $\frac{(G + H + 1)(2y_0 - 1)}{2(G + 1)} + 1$ |
| NeBi $\left(\frac{G}{t + H}\right)$ | $\frac{H}{G} y_0$ | Solution $t^*$ of $I_{t^*}^*(t^*, H)(y_0, G + 1) = \frac{G - 2}{H} y_0 - \frac{1}{t^*}$ |
| InBe($t, G, H$) | $\left[\frac{G}{H} y_0 + 1\right]$ (integer) | $\frac{H(2y_0 - 1)}{2(G + 1)}$ |
example 9.2. First, however, we consider an illustrative example for one of the cases shown in table 9.1.

Example 9.1

**Particle emission.** The number of radioactive particles emitted in a unit time period by a sample of chemical compound depends on the amount \( t \) of radioactive material contained in the sample. Past experiments with prepared amounts \( t_1, \ldots, t_n \) of the radioactive material gave radioactive counts \( x_1, \ldots, x_n \). For future purposes we want a sample which will emit \( y_0 \) radioactive particles in the unit time period. What amount of radioactive material should be used in preparation of the sample?

Assuming the usual Poisson model for radioactive counts we may take \( f_t \), the experiment which records the number of radioactive particles from a sample with amount \( t \) of the radioactive material, as \( \text{Po}(t) \). From sufficiency considerations we can clearly consider the informative experiment as recording \( x = x_1 + \ldots + x_n \) and then

\[
p(x|t, \theta) = \text{Po}(\Sigma t \theta)
\]

and

\[
p(y|t, \theta) = \text{Po}(t \theta).
\]

Assuming a conjugate prior \( p(\theta) = G \alpha, H \beta \) we immediately obtain from table 2.3 that

\[
p(y|t, z) = \text{NeBi} \left( G, \frac{t}{H + t} \right),
\]

where \( G = g + x, H = h + \Sigma t_i \).

If it is vital to obtain \( y_0 \), use of the all-or-nothing utility leads us to maximise

\[
t^* \frac{t^*}{(H + t)^{y_0 + \xi}}
\]

with respect to \( t \). Hence

\[
t^* = \frac{H}{G} y_0.
\]

This is sensible when one recalls that \( E(y|t, z) = (G/H) t \). We are therefore choosing the \( t \)-value which makes the expected value of \( y \) equal to \( y_0 \).

Suppose that a linear utility function of form (9.5) is suitable. Then we find that the expected utility is maximised when \( t^* \) is the solution of

\[
I_{u(t+H)}(y_0, G + 1) = \frac{\xi}{\xi + \eta}.
\]
3. In the derivation use is made of relationship (A23) in appendix I between the negative binomial and beta distributions. As \( \eta/\xi \) increases we see that \( t^* \) decreases. This is because as \( \eta/\xi \) increases the penalty for obtaining a \( y \)-value greater than \( y_0 \) becomes relatively larger. We therefore aim for smaller \( y \)-values. Since on the average \( y \) increases as \( t \) increases (\( E(y|t, \theta) = t\theta \)), our choice of \( t^* \) decreases. It is possible to display the dependence of \( t^* \) on \( \eta/\xi \) by plotting the graph (\( \eta/\xi, t^* \)). To illustrate this point fig. 9.1 shows the graph for the case where \( y_0 = 2, G = 25, H = 14 \). When \( \eta/\xi \) is large (> 10 say) small variations in its value are of little importance. For smaller values of \( \eta/\xi \), however, any variation is critical.

For a quadratic utility function,

\[
U(t, y_0) = -V(y|t, z) - \{E(y|t, z) - y_0\}^2 = \frac{-Gt(H + t)}{H^2} - \left( \frac{Gt}{H} - y_0 \right)^2.
\]

(9.12)

Thus, provided \( y_0 \neq 0 \), we find that
We now consider an example where the maximisation problems are more involved and iterative or search techniques are required.

Example 9.2
Suppressor drug dose level. In a clinical trial of a new drug aimed at suppressing the level of a certain body hormone to a satisfactory level, 23 randomly chosen patients were each given a different dose of the drug and their subsequent hormone level recorded. We require to recommend a dose \( t \) of a drug which will suppress the level of the body hormone in a patient to a level \( y_0 \).

Suppose that we may take \( f_t \), the experiment which records hormone level after application of dose \( t \) of the drug, to be described by a \( \text{NO}(\alpha + \beta t, \tau) \) density. The informative experiment \( e \) then consists of the independent experiments \( f_1, f_2, \ldots, f_{23} \) with recorded hormone levels \( x_1, \ldots, x_{23} \). The parameters \( \alpha \) and \( \beta \) occur only in the combination \( \mu = \alpha + \beta t \) in \( f_t \), and hence we may summarise the information from \( e \) by using

\[
m = \bar{x} + \hat{\beta}(t - \bar{t}), \quad v = \sum (x_i - \bar{x} - \hat{\beta}(t_i - \bar{t}))^2,
\]

where \( \hat{\beta} \) is the least squares estimate of \( \beta \). The statistic \( (m, v) \) is sufficient for \((\mu, \tau)\) given \( S(t, t) \), and has a distribution of the form

\[
p(m, v | \mu, \tau) = p(m | \mu, \tau) p(v | \tau),
\]

\[
p(m | \mu, \tau) = \text{NO}(\mu, \tau), \quad p(v | \tau) = \text{Ch}(v, \tau),
\]

where

\[
\frac{1}{k} = \frac{1}{n} + \frac{(t - \bar{t})^2}{S(t, t)}, \quad v = n - 2.
\]

We can therefore use table 2.3 in chapter 2 to update a prior \( \text{NOCh}(\beta, c, g, h) \) density for \((\mu, \tau)\) and so obtain the predictive density function

\[
p(y | t, z) = \text{St} \left[ G, B, \left( 1 + \frac{1}{C} \right) \frac{H}{G} \right]
\]

where \( B, C, G, H \) are as defined in case 5 of table 2.3.

Suppose that any excess of hormone is assessed to be \( \eta \) times as harmful as a similar deficit. Then we may use

\[
U(t, y, y_0) = \begin{cases} 
- (y_0 - y) & (y < y_0), \\
- \eta (y - y_0) & (y \geq y_0).
\end{cases}
\]

The expected utility is given by
Regulation and optimisation

\[ U(t, y_0) = \left( \left( 1 + \frac{1}{C} \right)^2 \right) \left\{ \eta w - \left( 1 + \eta \right) \frac{G + w^2}{G - 1} \psi_G(w) \right\} \]

\[ = \left( 1 + \frac{1}{C} \right)^2 \left( 1 + \frac{1}{C} \right)^{-1/2} \]

where \( w = (y_0 - B) \left( \left( 1 + \frac{1}{C} \right)^2 \right)^{-1/2} \)

\( \psi_G, \Psi_G \) are the probability density function and distribution function respectively of a \( t(G) \) random variable. In general no simple expression can be found for the \( t \)-value which maximises \( U(t, y_0) \). Even the assumption of prior ignorance with \( \rho(y|t, z) = S[\nu, m, \{1 + (1/k)\}^\nu] \) does not simplify matters.

A similar problem exists if we consider the quadratic loss function (9.7). Then

\[ U(t, y_0) = - \frac{(C + 1)H}{C(G - 2)} - (B - y_0)^2. \]

If we set the derivative with respect to \( t \) equal to zero then we have to find the solution of a polynomial equation of degree 5 in \( t \). For the case of prior ignorance, however, a solution is readily obtained, and is given by

\[ \hat{t} = \hat{t} + \frac{(y_0 - \bar{x})}{\hat{\beta} + \frac{\nu}{\hat{\beta}(\nu - 2)} S(t, t)}. \]

The simplest classical solution would be to perform the future experiment at the value

\[ t = \frac{y_0 - \bar{x}}{\hat{\beta}} = \hat{t} + \frac{y_0 - \bar{x}}{\hat{\beta}}, \]

the result of solving the regression equation for \( t \) and substituting \( y = y_0 \). The use of such an estimate is, however, questionable since the variance of a predicted y-value using this \( t \) increases as \( t \) increases; in other words, the variance depends on the value \( y_0 \).

We now consider a specific numerical example. Suppose that the recommended hormone level \( y_0 \) is 65 and that any excess of hormone is twice as harmful as a similar deficit. We have available the results shown in table 9.2 from the informative experiment in which 23 patients received varying doses. We have

\[ y_0 = 65, \eta = 2, n = 23, \]

\[ \bar{x} = 50.87, \hat{t} = 7.60, \]

\[ \hat{\beta} = 12.14, S(t, t) = 26.11, \nu = 667.98. \]
170 Regulation and optimisation

For simplicity we assume prior ignorance, so that with utility function (9.16) the expected utility \( U(t, y_0) \) reduces to

\[
U(t, y_0) = 1.10 \left( t^2 - 15.21t + 85.07 \right)^{1/2} \\
\times \{2w - (3.15 + 0.15w^2) \psi_{21}(w) - 3w\psi_{21}(w)\} \tag{9.21}
\]

where

\[
w = \frac{96.42 - 11.00t}{(t^2 - 15.21t + 85.07)^{1/2}}.
\]

A simple search technique reveals that \( U(t, y_0) \) takes its maximum value when \( t = 8.54 \).

This may be compared with (i) the classical estimate (9.20) which gives \( t = 8.77 \) and (ii) the estimate (9.19) associated with the quadratic utility structure which gives \( t = 8.76 \).

9.3 Set regulation

Sometimes in regulation problems the objective may be not to obtain a particular value \( y = y_0 \) but to ensure that the \( y \)-value lies in some subset \( Y_0 \) of \( Y \). Although we seek to regulate the outcome within a set we are still concerned with finding a single element \( t \) from \( T \). Simple utility functions corresponding to those of the previous section are straightforwardly derived, but their use, even for the standard distributions, leads to formulations which are not as simple as those previously encountered. It seems that trial-and-error computations are necessary. We leave to the reader the simple task of formulating the optimisation problems and discovering the difficulties. We follow here the easier path of an illustrative example.

Example 9.3 Laminate design. Recall example 1.4 on laminate design discussed in chapter 1. If, for any particular sheet, we assume that the number of flaws is Poisson-distributed with mean \( \theta \), it follows immediately that the conditional distribution of the total number \( y \) of flaws given that \( t \) independently produced sheets have been superimposed is Poisson with mean \( \theta t \), that is, \( p(y | t, \theta) \) is \( \text{Po}(\theta t) \). A suitable utility function for this problem may take the following form:

\[
U(t, y, y_0) = \begin{cases} 
K_1(t) & (y \leq y_0), \\
-K_2(t) & (y > y_0), 
\end{cases}
\tag{9.22}
\]

where \( K_1(t), K_2(t) \) are cost functions. If \( y \leq y_0 \) we shall be able to sell the product and thus make some profit \( K_1(t) \), which we assume to depend only on the number \( t \) of sheets. If, however, \( y > y_0 \) the product will be rejected...
Table 9.2 Dosages of drug and body hormone levels for 23 patients

<table>
<thead>
<tr>
<th>Dose</th>
<th>Hormone level</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.4</td>
<td>9.4</td>
</tr>
<tr>
<td>7.8</td>
<td>7.1</td>
</tr>
<tr>
<td>6.8</td>
<td>65</td>
</tr>
<tr>
<td>8.5</td>
<td>35</td>
</tr>
<tr>
<td>5.9</td>
<td>62</td>
</tr>
<tr>
<td>6.4</td>
<td>26</td>
</tr>
<tr>
<td>8.8</td>
<td>43</td>
</tr>
<tr>
<td>8.0</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dose</th>
<th>Hormone level</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>8.3</td>
</tr>
<tr>
<td>6.1</td>
<td>8.3</td>
</tr>
<tr>
<td>8.2</td>
<td>8.3</td>
</tr>
<tr>
<td>6.1</td>
<td>8.3</td>
</tr>
<tr>
<td>7.0</td>
<td>8.3</td>
</tr>
<tr>
<td>7.5</td>
<td>8.3</td>
</tr>
<tr>
<td>9.7</td>
<td>8.3</td>
</tr>
<tr>
<td>9.3</td>
<td>8.3</td>
</tr>
</tbody>
</table>

and there will be a loss $K_{2}(t)$ associated with this, which we have again assumed to be a function of $t$. One could introduce some factor comparing $y$ with the critical value $y_0$, but this increases the complexity of the maximisation of the expected utility. If the main interest is the acceptance or rejection of the finished product and not in the amount by which the number of faults falls short of, or exceeds, the critical value, then the above utility function seems the most satisfactory. For simplicity we take $K_1(t) = k_1 t$, $K_2(t) = k_2 t$, where $k_1$, $k_2$ are positive constants.

From table 2.3 we have that

$$p(y|t,z) = \text{NeBi} \left( G, \frac{t}{t+H} \right),$$

where $G = g + x$, $H = h + \Sigma t$, as in (9.9). If we use (A23) of appendix I the expected utility may be written in the form

$$U(t, y_0) = k_1 t - (k_1 + k_2) t_1 U(t, y_0 + 1),$$

and we select the $t$-value $t^*$ which gives the largest value of this expected utility. For the numerical example given in example 1.4 with $k_1 = 10k_2$, we have, for prior ignorance on $\theta$,

$$U(t, y_0) = k_2 t \{10 - 11U(t, s)\} (8.49).$$

Evaluating $U(t, y_0)/k_2$ for $t = 0, 1, 2, \ldots$ we obtain the solution $t^* = 9$, and it follows that the optimal policy is to produce an article by superimposing 9 sheets.

Similarly we may obtain optima for different values of the ratio $k_1/k_2$, and table 9.3 gives a selection of results.

Table 9.3 Optimal number of sheets in laminate design problem

<table>
<thead>
<tr>
<th>$k_1/k_2$</th>
<th>Optimal number of sheets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
9.4 Regulation problems with a finite index set

Certain types of classification problems may be considered under this statistical model. Suppose that there are \( k \) categories, indexed by the \( t \)-values \( t = 1, 2, \ldots, k \), to which an item may belong. In the future experiment \( f \), we may wish to obtain a certain \( y \)-value (\( y = y_0 \), say). For which category do we carry out the experiment? This is simply a regulation problem in which the index set \( T \) is not continuous, but finite. Again we consider only a simple example and leave the reader to derive solutions for other situations.

Suppose that we take \( p(y|\theta) = N_0(\mu_0, \tau) \), where \( \theta = (\mu_1, \mu_2, \ldots, \mu_k, \tau) \), and assume that \( p(\theta) = N_0(\mu_0, \tau) \), that is, \( p(\mu_0) = N_0(\mu_0, \tau) \), \( p(\tau) = Ch(g, h) \). Let \( n_i \) observations in the informative experiment be in the \( t \)th category and suppose that \( m_t \) is the mean of the corresponding \( x \)-values \( (t = 1, 2, \ldots, k) \). Then the predictive density \( p(y|z) \) is

\[
\text{St}\{G, B_t, (1 + C^{tt})H/G\} \quad (9.24)
\]

where

\[
\begin{align*}
B &= (B_1, B_2, \ldots, B_k)' = C^{-1}(Dm + cb), \\
C &= c + D, \\
G &= g + n, \\
H &= h + xx' + b'B - B'C'B, \\
m &= (m_1, m_2, \ldots, m_k)', \\
D &= \text{diag}(n_1, n_2, \ldots, n_k), \\
n &= \sum_{t=1}^k n_t, \\
C^{-1} &= \{\psi_{ij}\} (i, j = 1, 2, \ldots, k). \\
\end{align*}
\]

If we use utility function (9.5) then

\[
U(t, y_0) = \left(1 + C^{tt}\right)^{1/2} \left[ \eta w_t - \frac{(\xi + \eta) G + w_t^2}{G - 1} \right] \psi_G(w_t), \\
- \left(\xi + \eta\right) w_t \psi_G(w_t), \quad (9.26)
\]

where

\[
w_t = (y_0 - B_t) \left(1 + C^{tt}\right)^{1/2} \quad (t = 1, 2, \ldots, k).
\]

We can carry out the maximisation simply by evaluating \( U(t, y_0) \) for the \( k \) different values of \( t \).
Regulation and optimisation

Table 9.4 Lengths of 42 articles produced by 6 machines

<table>
<thead>
<tr>
<th>Machine</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>x (mm)</td>
<td>10.6</td>
<td>10.6</td>
<td>10.9</td>
<td>11.0</td>
<td>10.9</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>11.2</td>
<td>11.0</td>
<td>11.2</td>
<td>10.9</td>
<td>11.1</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>10.9</td>
<td>10.8</td>
<td>11.1</td>
<td>11.2</td>
<td>10.6</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>11.3</td>
<td>10.8</td>
<td>11.2</td>
<td>10.9</td>
<td>10.9</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>11.2</td>
<td>11.2</td>
<td>11.3</td>
<td>10.6</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>10.4</td>
<td>11.3</td>
<td>11.3</td>
<td>10.7</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>10.6</td>
<td>11.2</td>
<td>11.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.3</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.3</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Table 9.5 Values of expected utility $U(t, y_j)$ (9.26) for the 6 machines

<table>
<thead>
<tr>
<th>t</th>
<th>$n_t$</th>
<th>$U(t, y_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>-0.19</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>-0.25</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-0.24</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>-0.20</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>-0.25</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>-0.41</td>
</tr>
</tbody>
</table>

Example 9.4

Component length control. An item can be produced on any one of six machines I, II, ..., VI. The data in table 9.4 represent the lengths (mm) of a sample of items produced by the machines. If we require to produce an item of length 11.0 mm, which machine do we use?

If we assume prior ignorance on $\theta$ we have that

$$B_t = m_t, \quad C^{tt} = 1/n_t, \quad G = n, \quad H = \sum_{i=1}^{n} x_i^2 - \frac{k}{1}, \quad \sum_{t=1}^{n} m_t^2.$$

Taking $\xi = \eta = 1$ we find that the values of $U(t, y_0)$ given by (9.26) are as shown in table 9.5. The maximum value of $U(t, y_0)$ thus occurs when $t = 1$. We therefore perform the future experiment on Machine I.

9.5 Optimisation

In a number of problems of a regression nature we are interested in determining the value of the independent variable $t$ at which we should perform a future experiment in order to obtain an optimal value of the dependent variable $y$. In some cases (for instance, certain curvilinear regression problems) the optimal value may simply be the maximum (minimum) value of $y$ we can hope to obtain. In other cases the cost of performing the experiment $f_t$ may
Regulation and optimisation

vary with \( t \) and we shall then want to balance this cost against the benefit from a large (small) \( y \)-value. For optimisation problems, in contrast to regulation problems, we do not know the specific \( y_0 \) we are trying to obtain. Our objective is to optimise the outcome \( y \) of \( f_t \) subject to certain restrictions.

In the classification of prediction problems (appendix II) the components of the optimisation problem are as follows. The future experiment is one of the class \( F = \{ f_t : t \in T \} \); the predictive density function is of the form \( p(y \mid t, z) \); the action set \( A \) is \( T \); and the domain of the utility function is \( A \times Y \). The only difference from regulation problems is therefore the change in the specification of the utility function, brought about by our lack of knowledge of \( y_0 \).

A suitable specification for the cases mentioned above takes the form

\[
U(t, y) = y - K(t),
\]

where \( K(t) \) is a function which takes account of the cost of performing the future experiment \( f_t \), but which is assumed to be independent of \( y \). The expected utility is then

\[
U(t) = \int_Y U(t, y) p(y \mid t, z) dy = E(y \mid t, z) - K(t).
\]

Obviously the solution to the optimisation problem depends critically on \( K(t) \) and the form of \( T \). For the standard distributions of cases 1, 2 and 3 in table 2.3, the expected values \( E(y \mid t, z) \) are linear in \( t \). Hence use of a linear cost function \( K(t) \) would result in the optimum \( t \) value being one of the extreme values of \( T \).

If we are concerned with maximising a profit we may extend the above specification to the form

\[
U(t, y) = f(y) - K(t),
\]

where \( f(y) \) represents the return or gain from an experiment with outcome \( y \).

In these optimisation problems we are trying to determine the conditions (the value of the independent variable \( t \), which may be vector-valued) under which we should perform the future experiment in order to obtain the optimum result. The classical approach involves response surfaces and their polynomial representation. Typically, to find a maximum response, the response surface is first estimated, for example by the method of least squares, and then the optimum point found by differentiation. Techniques such as the method of steepest ascent seek an approach to a stationary point, possibly by some sequential method of experimentation. We are here concerned more with the terminal rather than the preposterior analysis, however, and so we are not concerned with the design of the informative experiment \( e \).
Consider first the case where there is an independent variable $t$ which is real-valued. If we can assume that we are dealing with normal variability from the response surface then we can take

$$p(y|t, \theta) = N(\phi(t, \beta), \tau),$$

where $\theta = (\beta, \tau)$ and $\phi(t, \beta)$ is the response surface. Thus, for polynomial regression, we have that

$$\phi(t, \beta) = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_r t^r,$$

where $\beta = (\beta_0, \beta_1, ..., \beta_r)$. The informative experiment consists of performances of $f_1, f_2, ..., f_n$ yielding observations $x_1, x_2, ..., x_n$.

By a simple extension of the arguments in §2.5 we can again make use of case 5 in table 2.3. For, if

$$T = \begin{pmatrix} 1 & t_1 & t_1^2 & ... & t_1^r \\ 1 & t_2 & t_2^2 & ... & t_2^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & ... & t_n^r \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and if $\hat{\beta} = (T' T)^{-1} T' x$, the least-squares estimate of $\beta$, then $m = T_0 \hat{\beta}$ and $v = x' x - x' T \hat{\beta}$ are jointly sufficient for $\mu = \phi(t, \beta)$ and $\tau$. Furthermore $m$ and $v$ are independently distributed with

$$p(m|\mu, \tau) = N(\mu, \{T_0(T'T)^{-1} T_0\}'^{-1} \tau),$$

$$p(v|\tau) = Ch(n-r-1, \tau).$$

The conjugate prior distribution for $(\mu, \tau)$ is taken to be $N(0, c, g, h)$. Hence we have that

$$p(y|t, z) = St\left(G, B, \left(1 + \frac{1}{C}\right) H \right)$$

where $B, C, G, H$ are as given in case 5 of table 2.3 with

$$k = (T_0(T'T)^{-1} T_0)^{-1}, K = 1, \nu = n-r-1.$$

With utility function (9.27) we then have that

$$U(t) = B - K(t).$$

If we assume prior ignorance on $(\mu, \tau)$ that is $c \to 0, g \to 0, h \to 0$, and let $K(t)$ be constant over the region of interest we see that $B = m = T_0(T'T)^{-1} T' x$. 

374
Table 9.6 Responses with varying amounts of drug for 34 patients

<table>
<thead>
<tr>
<th>Drug</th>
<th>Response</th>
<th>Drug</th>
<th>Response</th>
<th>Drug</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1.8</td>
<td>0.6</td>
<td>3.3</td>
<td>1.5</td>
<td>5.0</td>
</tr>
<tr>
<td>0.3</td>
<td>1.1</td>
<td>0.7</td>
<td>3.5</td>
<td>1.6</td>
<td>6.6</td>
</tr>
<tr>
<td>0.4</td>
<td>3.0</td>
<td>0.9</td>
<td>6.4</td>
<td>1.7</td>
<td>5.7</td>
</tr>
<tr>
<td>0.4</td>
<td>1.3</td>
<td>1.0</td>
<td>5.8</td>
<td>1.8</td>
<td>5.5</td>
</tr>
<tr>
<td>0.4</td>
<td>3.2</td>
<td>1.2</td>
<td>6.0</td>
<td>1.9</td>
<td>5.9</td>
</tr>
<tr>
<td>0.5</td>
<td>4.9</td>
<td>1.3</td>
<td>5.0</td>
<td>2.0</td>
<td>5.2</td>
</tr>
<tr>
<td>0.5</td>
<td>4.4</td>
<td>1.4</td>
<td>6.1</td>
<td>2.1</td>
<td>4.0</td>
</tr>
<tr>
<td>0.5</td>
<td>2.7</td>
<td>1.4</td>
<td>5.5</td>
<td>2.2</td>
<td>4.7</td>
</tr>
<tr>
<td>0.5</td>
<td>2.1</td>
<td>1.4</td>
<td>5.0</td>
<td>2.3</td>
<td>2.7</td>
</tr>
<tr>
<td>0.6</td>
<td>3.1</td>
<td>1.4</td>
<td>6.6</td>
<td>2.4</td>
<td>3.5</td>
</tr>
<tr>
<td>0.6</td>
<td>3.4</td>
<td>1.4</td>
<td>5.9</td>
<td>2.6</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Then the optimum \( t^* \) at which we should perform the future experiment corresponds exactly to the maximum of the fitted least-squares regression curve and would be the optimum obtained by classical methods.

Example 9.5

Drug response. In a series of medical experiments 34 patients suffering from a certain complaint were treated with varying amounts of a drug and their responses recorded. It is known that underdoses and overdoses generally lead to smaller values of the response. The results of the series of trials of the experiment are shown in table 9.6, where \( x \) denotes the response measured when an amount \( t \) ml of the drug was used. One factor of interest would be to find the amount of the drug needed to maximise the response for a future patient.

Suppose that we assume a quadratic response surface, so that \( r = 2 \). For prior ignorance on \( \theta \) we find that

\[
\hat{\theta} = \begin{bmatrix} -0.40 \\ 8.65 \\ -2.98 \end{bmatrix}
\]

Hence

\[
U(t) = -0.40 + 8.65t - 2.98t^2 - K(t).
\]

If \( K(t) \) is a constant over the region of interest then

\[
t^* = 1.45.
\]

The cost of an individual experiment may consist of two parts — the cost of the amount \( t \) of the drug and the charge for performing the treatment. We could then take \( K(t) = k_1 + k_2t \) where \( k_1, k_2 \) are positive constants. In this case we should perform the future experiment with an amount \( t \) of drug given by
8.65 - k_2 \over 5.96}

provided that \( k_2 < 8.65 \).

We may extend the ideas to the case where the independent variable \( t \) is vector-valued. For example, for two independent variables \( t_1 \) and \( t_2 \) a polynomial representation of the response surface is

\[
\phi(t_1, t_2, \beta) = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_{11} t_1^2 + \beta_{12} t_1 t_2 + \beta_{22} t_2^2 + \ldots
\]

The predictive density function is as before except that

\[
T = \begin{bmatrix}
1 & t_{11} & t_{11}^2 & t_{11} t_{21} & t_{21}^2 & \\
1 & t_{12} & t_{12}^2 & t_{12} t_{22} & t_{22}^2 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \\
1 & t_{1n} & t_{2n} & t_{1n} t_{2n} & t_{2n}^2 &
\end{bmatrix}
\]

and \( T_0 = [1 \ t_1 \ t_2 \ t_1^2 \ t_1 t_2 \ t_2^2 \ldots] \).

Computations are considerably eased if \( \theta \) is well-designed and the \( t_1, t_2 \) values scaled.

**Example 9.6**

*Maximising the yield of an industrial process.* Recall example 1.5. In a balanced informative experiment the yields (kg), recorded in table 1.4, were obtained when an industrial process was run successively at five different temperatures and three different pressures, each combination of temperature and pressure being used twice. For simplification of the calculations we scale the values of temperature to \(-2, -1, 0, 1, 2\) and of pressure to \(-1, 0, 1\), so that \( \Sigma t_{1i} = \Sigma t_{2i} = 0 \). Also we take the response surface to be of second degree and given by

\[
\phi(t_1, t_2, \beta) = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_{11} (t_1^2 - 2) + \beta_{12} t_1 t_2 \\
+ \beta_{22} (t_2^2 - 3/2).
\]

This slight alteration in the response curve, and the corresponding change in \( T \) ensure that for the balanced design given \( T' T \) is diagonal, and so we have greatly simplified the problem of inverting a \( 6 \times 6 \) matrix.

If we again assume prior ignorance on \( \theta \) we have that
\[ \hat{\beta} = (T'T)^{-1} T'x = \begin{bmatrix} 75.27 \\ 1.53 \\ 1.60 \\ -1.00 \\ -1.10 \\ -3.20 \end{bmatrix} \]

so that

\[ U(t_1, t_2) = 75.27 + 1.53t_1 + 1.60t_2 - 1.00(t_1^2 - 2) \\
- 1.10 t_1t_2 - 3.20(t_2^2 - 3/2) - K(t). \]

Hence for a constant cost function the optimal operational values are

\[ t_1^* = 0.69, \quad t_2^* = 0.13, \]

that is at 76.9°C and 1.28 atmospheres.

**History**

The Bayesian decision theory models derived in the chapter are given in Dunsmore (1969). Zellner and Chetty (1965) consider a similar regulation problem with a quadratic loss function for the multiple regression model. Lindley (1968) also considers this setting with the additional option of choosing which independent variables should be used in the attempt to control the \( y \)-value. His loss function contains a factor which depends on the values of the independent variables used.

The classical approach to optimisation is through the field of response surfaces; see, for example, Davies (1960).

Similar problems arise in the theory of stochastic control. There is considerable literature in this field and much Bayesian decision theory work; see, for example, Aoki (1967) and Sawagari, Sunahara and Nakamizo (1967).

**Problems**

9.1 A faculty of a state university is faced with the problem of the number of applicants to whom it should offer places. Study of the acceptance rate \( \theta \) in previous years has shown a variation well described by a \( \text{Be}(G, H) \) distribution. The optimum number of students in first year is reckoned to be \( y_0 \). The disutilities of missing this target are estimated at \( \xi \) for each student place unfilled and \( \eta \) for each student in excess of the target. How many students should be offered places?
Regulation and optimisation

9.2 Complete the analysis of problem 1.4.

9.3 The rate of flow of air through doorways in a certain type of hospital area has been found to be approximately linearly related to door area for fixed temperature difference between the rooms separated by the door. For the temperature drop proposed between two rooms the data pairs (area, airflow) available are \((t_1, x_1), \ldots, (t_n, x_n)\). From considerations of comfort it has been agreed that the desirable airflow rate is \(Y_0\), and that any rate of airflow in excess of \(Y_0\) is three times as uncomfortable as the rate of airflow which falls a corresponding amount short of \(Y_0\). What door area should be adopted? With this adoption for what proportion of time is the airflow likely to exceed \(Y_0\)?

9.4 The crushing strength of mortar varies at different times after mixing. For mortar prepared from two different types of cement \(A, B\) the crushing strengths are observed at six different times after mixing, four replications of each observation being made with the following results (in kg/cm\(^2\)).

<table>
<thead>
<tr>
<th>Days, (t)</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>350</td>
<td>420</td>
<td>480</td>
<td>560</td>
<td>610</td>
</tr>
<tr>
<td></td>
<td>280</td>
<td>390</td>
<td>460</td>
<td>520</td>
<td>550</td>
<td>580</td>
</tr>
<tr>
<td>(A)</td>
<td>300</td>
<td>330</td>
<td>450</td>
<td>530</td>
<td>540</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>310</td>
<td>350</td>
<td>480</td>
<td>550</td>
<td>550</td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>310</td>
<td>450</td>
<td>510</td>
<td>600</td>
<td>670</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>340</td>
<td>460</td>
<td>530</td>
<td>610</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>350</td>
<td>490</td>
<td>550</td>
<td>590</td>
<td>670</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>310</td>
<td>430</td>
<td>560</td>
<td>560</td>
<td>710</td>
</tr>
</tbody>
</table>

For each of the mortar mixes determine after how many days you could continue with construction if a crushing strength of 400 kg/cm\(^2\) is required. If you use utility function (9.5) what happens if \(\xi \gg \eta\)?

Which cement would you use if you require a crushing strength of (at least) 400 kg/cm\(^2\) after 6 days?

9.5 For the preparation of a new household insecticide a manufacturer wishes to determine the quantity \(t\) of the highly potent active ingredient which he should include. He carries out a series of trials with various levels \(t_1, t_2, \ldots, t_n\) of this ingredient and records the numbers of insects killed as shown below.

What minimum level of the active ingredient should he use in his insecticide if he requires a 75 per cent success rate?
Suppose further that a utility structure of the form (9.5) is suitable with $\xi = 2$, $\eta = 1$ and with a cost component $k \log t$. What level would you now suggest?
Calibration

10.1 The nature of a calibration problem

Calibration is commonly regarded as the process whereby the scale of a measuring instrument is determined or adjusted on the basis of an informative or ‘calibration’ experiment. For example, if we wish to calibrate an unscaled thermometer we might note the position \( x_1 \) on the liquid scale when the thermometer is immersed in boiling water at atmospheric pressure, that is, corresponding to temperature \( t_1 (= 100 \, ^\circ C) \); and the position \( x_2 \) when the immersion is in ice, say corresponding to temperature \( t_2 (= 0 \, ^\circ C) \). We might then divide the scale between \( x_1 \) and \( x_2 \) into 100 equal divisions so that, when the thermometer is immersed into some other substance, we are able to deduce very simply from the \( x \)-scale the corresponding temperature of the substance. In this example the use of the calibration experiment yielding trial records \((t_1, x_1), (t_2, x_2)\) is straightforward since there is, or at least we are assuming that there is, a one-to-one correspondence between the \( x \)-scale and the temperature or \( t \)-scale. But the same type of problem arises commonly in a less simple form, for usually, as in the following examples, there is no unique \( x \) corresponding to a given \( t \).

Example 10.1

Measuring water content of soil specimens. Two methods are available for obtaining the water content in soil specimens. The first method, performed in the laboratory, is very accurate but is expensive and tedious to operate. The second method, which can be performed on site, is much quicker and cheaper, but is less accurate. It is intended that for future samples the second method be used and from the value obtained some estimate be made of the reading which the accurate first method would have given. Information on the relative values given by the two methods is obtained from a calibration experiment in which the water contents of 16 naturally occurring soil specimens were measured with results as shown in table 10.1.

The data \( z \) from the calibration experiment consist of the 16 paired measurements

\[
z = \{(t_i, x_i): i = 1, ..., 16\},
\]
Table 10.1 Water contents (percentages by weight) of 16 soil specimens determined by two methods

<table>
<thead>
<tr>
<th>Serial no. of specimen</th>
<th>Laboratory method</th>
<th>On-site method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35.3</td>
<td>23.7</td>
</tr>
<tr>
<td>2</td>
<td>27.6</td>
<td>20.2</td>
</tr>
<tr>
<td>3</td>
<td>36.2</td>
<td>24.5</td>
</tr>
<tr>
<td>4</td>
<td>21.6</td>
<td>15.8</td>
</tr>
<tr>
<td>5</td>
<td>39.8</td>
<td>29.2</td>
</tr>
<tr>
<td>6</td>
<td>24.1</td>
<td>17.8</td>
</tr>
<tr>
<td>7</td>
<td>16.1</td>
<td>10.1</td>
</tr>
<tr>
<td>8</td>
<td>27.5</td>
<td>19.0</td>
</tr>
<tr>
<td>9</td>
<td>33.1</td>
<td>24.3</td>
</tr>
<tr>
<td>10</td>
<td>12.8</td>
<td>10.6</td>
</tr>
<tr>
<td>11</td>
<td>23.1</td>
<td>15.2</td>
</tr>
<tr>
<td>12</td>
<td>19.6</td>
<td>11.4</td>
</tr>
<tr>
<td>13</td>
<td>26.1</td>
<td>19.7</td>
</tr>
<tr>
<td>14</td>
<td>19.3</td>
<td>12.7</td>
</tr>
<tr>
<td>15</td>
<td>18.8</td>
<td>12.6</td>
</tr>
<tr>
<td>16</td>
<td>39.8</td>
<td>31.8</td>
</tr>
</tbody>
</table>

Fig. 10.1 Scatter diagram of laboratory and on-site measurements of 16 soil specimens.
where $t_i$ and $x_i$ denote measurements by the laboratory and on-site methods respectively. The associated scatter diagram of fig. 10.1 shows that we can no longer make the simplifying assumption that there is a unique $x$-value corresponding to a given $t$-value. For example $t_5 = t_{16} = 39.8$, but $x_5 \neq x_{16}$. To construct an appropriate calibration model we must therefore regard the on-site determination of water content corresponding to a true water content $t$ determined by the laboratory method, as a random experiment $f_t$, say, with outcome space $X$. The calibration experiment then consists of recording

$$ t = (t_1, \ldots, t_{16}), $$

the vector of true values and the vector

$$ x = (x_1, \ldots, x_{16}), $$

where $x_1, \ldots, x_{16}$ are the outcomes of independent performances of $f_{t_1}, \ldots, f_{t_{16}}$. Moreover, for a new soil specimen with on-site measurement $y$ the corresponding true value $u$ is unknown, and so we are forced to consider the class $F$ of experiments:

$$ F = \{f_t : t \in T\}, $$

where $T$ is the set of possible true water contents.

We have thus a regression-type framework similar to that used for the regulation and optimisation problems of chapter 9. The difference is that whereas in the previous problems we had to select some suitable future experiment $f_t$, here the 'future' experiment $f_u$ has already been performed and our objective is to attempt to identify the index $u$ from which the known outcome $y$ has arisen. A calibration problem is thus a kind of inverse prediction, a problem of retrospection, but we shall see shortly that in its resolution the predictive distribution plays a central role.

**Example 10.2**

**Calibration of an autoanalyser.** It is proposed to install a new autoanalyser in a hospital for the routine determination of the concentration of a certain enzyme in blood plasma samples. The enzyme concentration can be determined accurately by a long and costly laboratory method whereas the autoanalyser method is quick and cheap. It is known from the considerable past experience that the samples presented for analysis have enzyme concentrations (meq/l) which are normally distributed with mean 4.6 and standard deviation 0.8.

To evaluate the effectiveness of the autoanalyser 9 plasma samples, selected to cover the range of enzyme concentrations, have each been divided into four aliquots, one aliquot being assigned to the laboratory method and the other three to separate analyser determinations. The results are shown
in table 10.2. In the future it is hoped that only one such aliquot need be analysed by the autoanalyser to provide a reliable estimate of enzyme concentration. Is this a reasonable hope?

For an aliquot from a new plasma sample the autoanalyser gives a reading of 3.8 meq/l. What can be said about enzyme concentration?

Fig. 10.2 shows the \((t, x)\) scatter diagram for the data from this calibration experiment. As appears reasonable from this diagram, we make the assumption that all autoanalyser determinations are statistically independent even though they may be associated with aliquots from the same plasma sample. Thus the data from this informative experiment constitute

\[ z = \{(t_i, x_i): i = 1, \ldots, 27\}, \]

with \((t_1, x_1) = (3.0, 2.3), (t_2, x_2) = (3.0, 2.4), \ldots, (t_{27}, x_{27}) = (6.2, 5.2)\). These are trial records of independent performances of \(f_1, \ldots, f_{27}\), where \(f_i\) denotes the experiment which records the autoanalyser determination associated with an enzyme concentration \(t\) determined by the laboratory method.

The important difference between example 10.1 and the present example lies in the manner in which the \(t\)-values \(t_1, \ldots, t_n\) in the calibration experiment have arisen. In the former, the \(t\)-values were naturally occurring and if it is to be assumed that future soil specimens arise in the same way as in the past then the calibration experiment provides some indication of what this pattern is. In the latter, the blood samples used in the informative experiment have been deliberately selected to give a reasonable cover of the set \(T\), and improve the 'design' of the calibration experiment. Thus while their choice presumably reflects some view as to the future pattern of \(t\)-values the calibration experiment itself provides no new information about that pattern.

To distinguish clearly between these two types of calibration experiment we shall use the terms *natural* and *designed*.
10.2 The calibrative distribution

The calibration problem can now be stated in general terms. Let us denote by the term subject any candidate for calibration such as the soil specimens and blood samples of the examples in §10.1. Associated with each subject is a unique element of a specified set $T$ and we shall refer to this element as the index of the subject. In our two examples the indices are the accurate laboratory determinations of water content and enzyme concentration. The aim of calibration is to identify the indices of subjects as clearly or as closely as possible. For each subject it is possible to perform a trial yielding an outcome or measurement in a specified record set $X$. Such a trial is an experiment belonging to a class

$$F = \{f_t : t \in T\}$$

of experiments. The trial associated with a subject whose index is $t$ is $f_t$. It is assumed that for a number $n$ of subjects it has been possible to record both the index and the corresponding measurement. The $n$ trial records
(t_1, x_1), ..., (t_n, x_n) thus constitute a calibration experiment with data

$$z = \{(t_1, x_1), ..., (t_n, x_n)\},$$

which we may also conveniently write in the form

$$z = (t, x),$$

where

$$t = (t_1, ..., t_n), \quad x = (x_1, ..., x_n).$$

A new subject of unknown index \( u \) is presented and when the associated trial is performed the measurement obtained is \( y \). In the light of our knowledge of this measurement \( y \) and of the data \( z \) of the calibration experiment what can we say about the unknown index \( u \)? Clearly this will depend on any assumptions we make concerning the probabilistic mechanism by which trial records are generated, and these we now examine. Our objective is to try to arrive at a probabilistic description \( p(u|y, z) \), expressing the plausibility of the various \( u \) for the given values of \( y \) and \( z \). This distribution we shall refer to as the calibrative distribution.

**Calibrative distribution for a natural calibration experiment.** We now set out carefully the assumptions under which the calibrative distribution is derived. Only by exposing the assumptions can we be in a position to examine whether for a particular application the calibrative distribution is appropriate. In each of our applications we shall be concerned with parametric models and so this constitutes our first calibration assumption.

**C1**

The class of probability models describing the generation of a trial record is parametric.

Denote the parameter set by \( \Omega \), so that the density function corresponding to parameter \( \omega \in \Omega \) is \( p(t, x|\omega) \). The next two assumptions concern the nature of the parameter \( \omega \). We envisage \( \omega \) as having two components \( \psi \in \Psi \) and \( \theta \in \Theta \) and the assumptions C2 and C3 give meanings to these two aspects of the parameter.

**C2**

$$p(t|\psi, \theta) = p(t|\psi).$$

Here we are simply stating that \( \psi \) is the parameter which is concerned with the natural arrival pattern of the indices or \( t \)-values. Given \( \psi \), knowledge of \( \theta \) in no way affects the arrival pattern. We shall thus term \( \psi \) the *arrival* parameter.

**C3**

$$p(x|t, \psi, \theta) = p(x|t, \theta).$$
This assumption concerns the description of the trial \( f_t \) and asserts that its description involves only the \( \theta \) component of the parameter. The parameter \( \theta \) is therefore important in describing the structure of the regression experiments in the class \( F \), and so we shall term it the *structural parameter*.

Assumptions C1–3 then combine to give us the description of the variability in trial records:

\[
p(t, x|\omega) = p(t, x|\psi, \theta) = p(t|\psi)p(x|t, \theta). \tag{10.4}
\]

Another simplifying assumption is that subjects or trial records are statistically independent. This is certainly an assumption that may require very careful scrutiny for each application. In §10.1 we have already examined its relevance for example 10.2.

**C4**

For any set \( z \) of trial records \( z_1, z_2, \ldots \)

\[
p(z|\omega) = \prod p(z_i|\omega).
\]

We now make an assumption about our state of knowledge about the arrival parameter \( \psi \) and the structural parameter \( \theta \) prior to the calibration experiment. This states that our initial sources of information are independent.

**C5**

\[
p(\psi, \theta) = p(\psi)p(\theta).
\]

Finally, since we are dealing with the case of a natural calibration experiment we assume that further subjects presenting will follow the same pattern of arrival as the original trial subjects.

**C6**

\[
p(u|\psi, \theta) = p(u|\psi),
\]

which is of the same form as the \( p(t|\psi) \) of C2.

In following through the consequences of these assumptions we must first recognise that we have three unknown indices or parameters, \( u, \psi \) and \( \theta \), and that the set of data is \( y, z \). The technique for obtaining \( p(u|y, z) \) is then simply to use the assumptions and Bayes's theorem to obtain first \( p(u, \psi, \theta|y, z) \) and then to integrate out \( \psi, \theta \) to derive the marginal \( p(u|y, z) \) as the calibrative distribution. First we obtain the forms of the prior \( p(u, \psi, \theta) \) and of the likelihood \( p(y, z|u, \psi, \theta) \) as consequences of the assumptions. We have

\[
p(u, \psi, \theta) = p(u|\psi, \theta)p(\psi)p(\theta)
\]

\[
= p(u|\psi)p(\psi)p(\theta) \tag{10.5}
\]

by C5 and C6.
Calibration

Also since \((u, y)\) is a trial record like \((t_1, x_1), \ldots, (t_n, x_n)\) we have

\[
p(u, y; z|\psi, \theta) = p(u, y|\psi, \theta) \prod_{i=1}^{n} p(z_i|\psi, \theta) \text{ by C4}
\]

\[
= p(u|\psi)p(y|u, \theta) \prod_{i=1}^{n} p(t_i|\psi) \prod_{i=1}^{n} p(x_i|t_i, \theta)
\]

\[
= p(u|\psi)p(y|u, \theta)p(t|\psi)p(x|t, \theta)
\]

by C1–3 \quad (10.6)

in an obvious shortened notation. Hence, from C6,

\[
p(y, z|u, \psi, \theta) = \frac{p(u, y; z|\psi, \theta)}{p(u|\psi)}
\]

\[
= p(y|u, \theta)p(t|\psi)p(x|t, \theta)
\]

by (10.6). Now applying Bayes’s theorem and using (10.5) and (10.7) we have

\[
p(u, \psi, \theta|y, z) \propto p(u, \psi, \theta)p(y, z|u, \psi, \theta)
\]

\[
= p(u|\psi)p(\psi|\theta)p(t|\psi)p(x|t, \theta)p(y|u, \theta)
\]

\[
= p(u|\psi)p(\psi|\theta)p(\theta|z)p(y|u, \theta),
\]

(10.8)

where

\[
p(\psi|t) \propto p(\psi)p(t|\psi),
\]

(10.9)

\[
p(\theta|z) \propto p(\theta)p(x|t, \theta)
\]

(10.10)

are the post-calibration experiment probability assessments for \(\psi\) and \(\theta\). Now integrating out \(\psi\) and \(\theta\) we have the calibrative distribution:

\[
p(u, y, z) \propto p(u|t)p(y|u, z),
\]

(10.11)

where

\[
p(u|t) = \int_{\psi} p(u|\psi)p(\psi|t) d\psi,
\]

(10.12)

\[
p(y|u, z) = \int_{\theta} p(y|u, \theta)p(\theta|z) d\theta.
\]

(10.13)

We can now see the relevance of the predictive distribution to the calibration problem, for (10.13) is simply the predictive distribution associated with the ‘future’ experiment \(f_u\) and based on data \(z\). Indeed (10.11) takes the form of Bayes’s theorem, as it could be applied after the data \(z\) or \((t, x)\) of the calibration experiment are known. The prior plausibility assessment is then \(p(u|t)\) based on the pattern we have seen in \(t_1, \ldots, t_n\), and the ‘likelihood’ of Bayes’s theorem takes the form of the predictive distribution.
Calibration

189

Calibrative distribution associated with a designed calibration experiment.

When the \( t \)-values \( t_1, \ldots, t_n \) for the calibration experiment are selected by the experimenter then the calibration experiment provides no direct information concerning the plausibilities of various \( u \). We no longer require to describe how trial records \( (t_1, x_1), \ldots, (t_n, x_n) \) are generated but only the nature of the given regression experiments \( f_{t_1}, \ldots, f_{t_n} \). For this we have the independence assumption:

\[ C'1 \]
\[ p(x|t, \theta) = \prod p(x_i|t_i, \theta). \]

The second assumption is concerned with a typical new subject with trial record \( (u, y) \):

\[ C'2 \]
\[ p(u, y|\theta) = p(u)p(y|u, \theta). \]

This incorporates the assumption that the calibration experiment provides no information about \( u \), since it implies that \( p(u|\theta) = p(u) \). If, however, we select the \( t \)-values in the informative experiment we clearly have some view as to what indices are likely to turn up, and the onus is on us to describe this view directly in \( p(u) \). The calibrative distribution can then be arrived at by the following sequence of steps:

\[ p(u, \theta|y, t, x) \propto p(u, \theta)p(y|t, x, u, \theta) \]
\[ \quad \propto p(u)p(\theta)p(y|u, \theta)p(x|t, \theta) \]
\[ \quad \propto p(u)p(y|u, \theta)p(\theta|x) \quad (10.14) \]

and integrating out \( \theta \),

\[ p(u|y, z) \propto p(u)p(y|u, z). \quad (10.15) \]

The relation (10.15) takes exactly the same form as (10.11) with, of course, \( p(u|\theta) \), the adjusted assessment for \( u \), replaced by the direct assessment \( p(u) \).

10.3 Calibrative distributions for the normal case

We confine ourselves throughout the remainder of this chapter to the case of linear regression with independently normally distributed errors. It is convenient to collect here the relevant distribution results preliminary to their application in the next section. The reader will find examples of other standard situations in the problems at the end of the chapter; see also Dunsmore (1968).

For the calibration experiment, whether natural or designed, we have the normal regression model for \( f_t \), which sets

\[ p(x|t, \theta) = N(\alpha + \beta t, \tau), \]
Calibration

where $\Theta = (\alpha, \beta, \tau)$; and for the future experiment $f_u$

$$p(y | u, \Theta) = \text{No}(\alpha + \beta u, \tau).$$

Since our interest is in the combination $\mu = \alpha + \beta u$ and $\tau$ we can, as in our discussion of example 9.2 in §9.2, summarise the data in the trial records $(t_1, x_1), \ldots, (t_n, x_n)$ constituting $e$ by

$$m = \bar{x} + \beta (u - \bar{t}), \quad v = \sum (x_i - \bar{x} - \beta (t_i - \bar{t}))^2.$$

Then $(m, v)$ satisfies the conditions of case 5 of table 2.3 with

$$\frac{1}{k} = \frac{1}{n} + \frac{(u - \bar{t})^2}{S(t, 1)}, \quad v = n - 2, \quad K = 1.$$

With a NoCh$(b, c, g, I)$ prior density function on $(\mu, \tau)$ we arrive at a Student form for the predictive distribution as defined in (10.13):

$$p(y | u, z) = \text{St}(G, B, \left(1 + \frac{1}{C} \frac{H}{G}\right);$$

the particular form of this when we adopt the vague prior distribution for $(\mu, \tau)$ is:

$$p(y | u, z) = \text{St}(\nu, m, \left(1 + \frac{1}{K} \frac{v}{\nu}\right). \quad (10.16)$$

For a natural calibration experiment the assumption of normality for $p(t | \psi)$, say No$(\lambda, \rho)$, together with a NoCh prior on $(\lambda, \rho)$ and the data $t_1, \ldots, t_n$ from $e$, leads again through a simple application of case 5 of table 2.3 to a Student form for $p(u | t)$ as defined in (10.12). The particular form of this when we adopt the vague prior distribution for $(\lambda, \rho)$ is:

$$p(u | t) = \text{St}(n - 1, \bar{t}, \left(1 + \frac{1}{n} \frac{S(t, 1)}{n - 1}\right). \quad (10.17)$$

Substituting for $p(u | t)$ and $p(y | u, z)$ in (10.11) we see that $p(u | y, z)$ is proportional to the product of two Student-type density functions, and in general this does not reduce to any standard distribution. This, however, does not detract from the applicability of the method since it is an elementary computational exercise, trivial even on a small computer, to obtain the appropriate conversion factor to yield the appropriate density function $p(u | y, z)$. This technique is illustrated in the first of the applications of the next section.

For a designed calibration experiment the source of information on which to base the choice of $p(u)$ in (10.15) is outwith the calibration experiment. Its choice must in some sense reflect the opinion which governed the selection of the $t$-values $t_1, \ldots, t_n$. Again the product $p(u) p(y | u, z)$ is unlikely to yield a standard density function but our previous remarks regarding simple computation apply equally here.
Much research effort has gone into attempts to find some suitable $p(u)$ which will yield a $p(y | u, z)$ of standard form. While we feel that this search for tractability is not really necessary in view of the simplicity of the computational problem we report on it briefly in § 10.5. Since much of the argument will there centre on $p(y | u, z)$ as a function of $u$ we set this out here in order to prepare the way for § 10.5. After some algebraic rearrangement we can express $p(y | u, z)$, in so far as it contains $u$, as being proportional to

$$\left( n - 3 + \frac{n(n - 3)}{(n + 1)S(t, t)}(u - \bar{t})^2 \right)^{(n-1)/2} \left( n - 2 + \frac{(u - u_l)^2}{Q(z, y)} \right)^{(n-1)/2},$$

(10.18)

where

$$u_l = \bar{t} + \frac{S(t, x)}{S(x, x)}(y - \bar{x}),$$

(10.19)

and

$$Q(z, y) = \frac{vS(t, t)}{(n - 2)S(x, x)} \left( 1 + \frac{1}{n} + \frac{(y - \bar{x})^2}{S(x, x)} \right).$$

(10.20)

All the results so far have been concerned with a single performance of the future experiment $f_u$. There are circumstances where it may be necessary to perform several replicates of $f_u$ to obtain a precise enough calibrative distribution. We thus adjust our analysis to cover the case where $K$ replicates of $f_u$ have been performed with outcomes $y_1, \ldots, y_K$. By sufficiency arguments for $\mu = \alpha + \beta u$ and $\tau$ we can clearly restrict attention to a condensed future experiment which records $(M, V)$, where

$$M = \bar{y}, \quad V = \Sigma(y_i - \bar{y})^2.$$

(10.21)

The distributions of $M$ and $V$, for given $u$ and $\theta = (\alpha, \beta, \tau)$, are independent and take the following forms:

$$p(M | u, \theta) = N_0(\mu, K\tau), \quad p(V | u, \theta) = C(h, K - 1, \tau).$$

(10.22)

We have thus again case 5 of table 2.3 and can immediately arrive at the predictive distribution for $(M, V)$. The particular form corresponding to the vague prior on $(\mu, \tau)$ is the following:

$$p(M, V | u, m, v) = StSi \left\{ n - 2; m, \left( \frac{1}{k} + \frac{1}{K} \right) \nu; K - 1, \nu \right\},$$

(10.23)

where

$$\frac{1}{k} = \frac{1}{n} + \frac{(u - \bar{t})^2}{S(t, t)}, \quad \nu = n - 2,$$

(10.24)
to which we apply whatever prior distribution for \( u \) is appropriate. We shall see an application of this result in the analysis of example 10.2 in §10.4.

Again we record the form of \( p(M, V|u, m, v) \) as a function of \( u \), so that we can readily study the search for a tractable \( p(u) \) in §10.5. In this case the function is proportional to

\[
\left\{ \frac{n + K - 4 + \frac{nK(n + K - 4)}{(n + K)S(t, t)} (u - \bar{t})^2 \right\}^{(n + K - 3)/2} \left\{ \frac{n + K - 3 + \frac{(u - u_0)^2}{Q(z, y)}}{(n + K - 2)/2} \right\}^{(n + K - 3)/2},
\]

where in this case

\[
\alpha = \frac{S(t, x)}{S(x, x) + V} (M - \bar{x}),
\]

and

\[
Q(z, y) = \frac{(v + V)S(t, t)}{(n + K - 3) [S(x, x) + V]} \left\{ \frac{1}{K} + \frac{1}{n} + \frac{(M - \bar{x})^2}{S(x, x) + V} \right\}. \tag{10.27}
\]

**10.4 Two applications of calibrative distributions**

We now apply the results of the preceding two sections to the two motivating examples of §10.1.

**Example 10.1 (continued)**

*Measuring water content of soil specimens.* Straightforward regression calculations yield the following values in the notation of §10.3:

\[
\begin{align*}
n &= 16, \quad \hat{\alpha} = -1.60, \quad \hat{\beta} = 0.770, \\
\bar{t} &= 26.3, \quad S(t, t) = 1066.5, \quad V = 29.6.
\end{align*}
\]

We recall that this is a case of a natural calibration experiment so that we require the two factors of (10.11). These are given by substitution of the above values in (10.16) and (10.17):

\[
\begin{align*}
p(y|u, z) &= \text{St}(14, -1.60 + 0.770u, 2.24 + 0.00198 (u - 26.3)^2), \\
p(u|t) &= \text{St}(15, 26.3, 75.5). \tag{10.28}
\end{align*}
\]

The numerical method of constructing the calibrative density function for a given \( y \) is first to compute the product (10.11) for \( u = b, b + h, b + 2h, \ldots, b + (N - 1)h \), where \( b + Nh = c \), and where \( (b, c) \) can easily be chosen to provide a sufficiently wide range for \( u \), and \( h \) chosen sufficiently small to make
Calibration

\[ \sum_{u=b}^{b+(N-1)h} p(u|t)p(y|u, z) \]

or some more sophisticated numerical integration formula, such as Simpson's rule, a good approximation to

\[ \int_T p(u|t)p(y|u, z)du. \]

The calibrative density function can then be simply graphed for the specification

\[ p(u|y, z) = \frac{p(u|t)p(y|u, z)}{h \sum p(u|t)p(y|u, z)} (u = b, b + h, \ldots, c). \] (10.30)

Fig. 10.3 Change from \( p(u|t) \) to \( p(u|y, z) \) for an observed on-site determination of water content \( y = 18 \).
Fig. 10.3 shows the change from $p(u|t)$ to $p(u|y, z)$ for an observed on-site determination of water content $y = 18$. Fig. 10.4 shows the calibrative density functions corresponding to $y = 10, 18, 26$. Note that there is very little difference in the shape of these three curves, and that the central 95 per cent of each calibrative distribution extends over a water content range of about 8 per cent. If no improvement of the on-site method is possible then the only opportunity to reduce the uncertainty of the calibration would be through the possibility of replicating the on-site determination. For example, if two determinations by the on-site method give $y_1 = 17, y_2 = 19$, then we require to apply the analysis of (10.21) to (10.24) with $K = 2$. We have $M = 18, V = 2$ and the predictive distribution of (10.28) is replaced by the following particular form of (10.23):

$$p(M, V|u, m, v) = StSi(14; -1.60 + 0.770u, 1.188 + 0.00198(u - 26.3)^2; 1, 29.56). \quad (10.31)$$

We can again apply the technique of (10.30) to obtain the calibrative distribution. Fig. 10.5 shows the extent to which the calibration is improved when two replicates are used instead of just one.
Fig. 10.5 Comparison of calibrative density functions for one and two replicates.

Example 10.2 (continued)

Calibration of an autoanalyser. As we have already pointed out in §10.1 this is an example of a designed calibration experiment and so the calibrative density function is evaluated by use of (10.15). In fact we are given a firm basis for the specification of \( p(u) \) in the information that the samples presented for analysis are likely to have enzyme concentrations which are normally distributed.
with mean 4.6 and standard deviation 0.8. Hence

\[ p(u) = \text{No}(4.6, 1/0.64). \quad (10.32) \]

The calibration experiment provides the following regression values,

\[
\begin{align*}
n &= 27, & \alpha &= -0.0624, & \beta &= 0.814, \\
\bar{x} &= 4.6, & S(t, t) &= 28.80, & v &= 0.223,
\end{align*}
\]

from which we construct the predictive density function from (10.16):

\[
p(y|u, z) = \text{St}\{25, -0.0624 + 0.814u, \\
0.00926 + 0.000310 (u - 4.6)^2 \}. \quad (10.33)
\]

Fig. 10.6 Calibrative density functions for the autoanalyser problem.

We can then apply a computational technique similar to (10.30) to obtain the calibrative density function corresponding to \( y = 3.8 \). Fig. 10.6 shows this density function. The calibrative density function corresponding to other
Calibration

...y-values in the range likely to occur is very similar in shape. If this degree of reliability available from one aliquot is not satisfactory then we can investigate along the lines of (10.21)–(10.24) the consequences of using more aliquots. Thus we could readily see how many aliquots are required to provide a specified degree of reliability. To indicate the extent of such an increased reliability we show also in fig. 10.6 the calibration density functions corresponding to the following cases:

\[
\begin{align*}
K &= 3, \quad M = 3.8, \quad V = 0.02; \\
K &= 10, \quad M = 3.8, \quad V = 0.09.
\end{align*}
\]

10.5 The search for tractable prior distributions

Suppose that in constructing the calibrative density function of (10.11) or of (10.15) we adopt an entirely mathematical approach and specify \( p(u|t) \) or \( p(u) \) in such a way that \( p(u|y, z) \) turns out to be a standard function. While such an approach may be considered a rearrangement of priorities, a case of the end justifying the means, some practical justification for such a choice can be supplied in certain situations. For the normal case a choice of a tractable prior is influenced by the expression (10.18) for \( p(y|u, z) \) as a function of \( u \). The denominator of (10.18) depends on \( y \), the outcome of the future experiment. Since our choice of \( p(u|t) \) or \( p(u) \) is made prior to the future experiment it seems unreasonable to allow \( p(u|t) \) or \( p(u) \) to depend on \( y \). It is therefore impossible to eliminate the denominator from (10.18) and we turn our attention to the numerator. If we take \( p(u|t) \) or \( p(u) \) to be

\[
St\left\{n - 3, \frac{1}{n} \left[ \frac{5(t, t)}{n - 3} \right] \right\}, \quad (10.34)
\]

then either (10.11) or (10.15) gives the very simple result that

\[
p(u|y, z) = St\{n - 2, u_t, Q(z, y)\}, \quad (10.35)
\]

where \( u_t \) is as given by (10.19) and \( Q(z, y) \) by (10.20).

It now remains to examine whether there is any non-mathematical justification for the choice of (10.34). When dealing with a natural calibration experiment in §10.3 we saw that \( p(u|t) \) there takes the form (10.17). Comparison of (10.17) with (10.34) shows that the tractable prior distribution will be a reasonable approximation to (10.17) when \( n \) is reasonably large. For the water content calibration experiment fig. 10.3 provided the calibrative density function for \( y = 18 \). The tractable form (10.35) which here takes the particular form

\[
St(14, 25.5, 3.61)
\]
is in fact indistinguishable in fig. 10.3 from this calibrative density function. For a designed calibration experiment any justification is less direct. The mean and variance of the suggested prior distribution (10.34) are \( \bar{t} \) and \( (n + 1)S(t, t) / (n(n - 5)) \). If the designer of the calibration experiment has chosen the \( t \)-values \( t_1, \ldots, t_n \) so that their mean and variance reflect the pattern of \( t \)-values he expects in the future then for reasonably large \( n \) the distribution (10.34) may give a satisfactory approximation. It should be emphasised, however, that there is no guarantee that such a design for the calibration experiment is optimum in the sense that it produces the most precise calibration distributions. For the enzyme calibration problem of §10.4 the one-aliquot calibrative density function for \( y = 3.8 \) already shown in fig. 10.6 is again indistinguishable from the tractable calibrative density function (10.35) which here takes the particular form

\[
\text{St}(25, 4.74, 0.0138).
\]

For the \( K \)-replicate future experiment discussed in §10.3 it is clear from (10.25) that tractability arises from the specification of \( p(u | t) \) or \( p(u) \) as

\[
\text{St} \left( n + K - 4, \bar{t}, \left( \frac{1}{n} + \frac{1}{K} \right) \frac{S(t, t)}{n + K - 4} \right).
\]

Then, from either (10.11) or (10.15),

\[
p(u | y, z) = \text{St}(n + K - 3, u_t, Q(z, y)).
\]

Unfortunately the choice of (10.36) as prior density function appears to be fairly arbitrary from a practical viewpoint. The prior mean and variance are given by \( \bar{t} \) and \( (n + K)S(t, t) / (nK(n + K - 6)) \). Whilst in certain circumstances the mean of \( t \) may be satisfactory, the variance is harder to justify since it depends on \( K \), the number of future replicates. Indeed we have the rather paradoxical situation that the more future observations there are (that is, the larger \( K \) is), the smaller is the prior variance. We see therefore that in this search for tractability there are certain difficulties involved in the case where \( K > 1 \). We reiterate our comment in §10.3 that tractability in these calibration problems can lead to an unnecessary departure from reality.

10.6 Calibration under a utility structure

A statistical decision theory model for calibration has the following components.

Parameter space. The set of unknown parameters is the index set \( T \) for the class \( F \) of possible experiments, the unknown index \( u \) playing the role of the unknown state of nature.
Calibration

Action set. The action set $A$ is the set $T$ for point calibration or a class of subsets of $T$ for interval calibration.

Utility function. It seems sensible to define a utility function $U(a, u)$ which attaches a utility to each combination of possible calibration point or set $a$ and each possible index $u$. The basic problem is to maximise the expected utility $U(y, a)$ of having obtained results $z$ from the informative experiment and $y$ from the 'future' experiment, and of having selected a region $a \in T$.

We have that

$$U(y, a) = \int_T U(a, u) p(u|y, z) du.$$  \hspace{1cm} (10.38)

where $p(u|y, z)$ is the calibrative density function. Since we have full control over the selection of $a$ we choose the index or interval which maximises $U(y, a)$, perhaps subject to some constraint. We note here a special feature of the analysis, namely the dependence of the expected utility on $y$. We are indeed using $y$ to obtain information about $\theta$ additional to that already obtained from $e$. For some simple forms of $U(a, u)$, completely analogous to the utility functions considered in §3.1 and 3.3, we simply list in table 10.3 the optimum calibrations in terms of $p(u|y, z)$.

10.7 Some comparisons of methods

The calibrative distribution $p(u|y, z)$ is the cornerstone of the Bayesian model for the inverse prediction problem. It supplies the complete summary of our views on the value of $u$. From it we can provide either a point estimate (for example, the mean or mode of $p(u|y, z)$) or a set prediction (for example, the most plausible Bayesian interval of cover $\kappa$). Alternatively the Bayesian decision theory approach of §10.6 can provide such predictions.

The classical approach to the calibration problem has been subjected to much investigation (see references at the end of the chapter). We consider first the single-replicate future experiment situation for the normal case. The classical point prediction $\hat{u}$ is obtained simply by inverting the least squares regression line

$$y = \hat{a} + \hat{\beta} u$$

to give

$$\hat{u} = \bar{t} + \frac{(y - \bar{y})}{\hat{\beta}}$$

$$= \bar{t} + \frac{S(t, t)}{S(t, x)} (y - \bar{y}).$$  \hspace{1cm} (10.39)
Table 10.3 Optimal calibrations associated with some utility functions

| Utility function | Optimal prediction in terms of \( p(y | x, z) \) |
|------------------|-------------------------------------------------|
| \( U(a, u) \)    |                                                 |
| \( 1 \)          | \( a - \epsilon < u < a + \epsilon \) mode    |
| \(-\xi(a - u)\)  | \( u < a \) \frac{\eta}{\xi + \eta} \text{ th quantile} |
| \(-\eta(u - a)\) | \( u > a \) \frac{\eta}{\xi + \eta} \text{ th quantile} |
| \(- (a - u)^2 \) | mean                                            |

Set prediction

\[
\begin{cases}
1 - \gamma(a_2 - a_1) & a_1 < u < a_2 \\
- \gamma(a_3 - a_1) & \text{otherwise}
\end{cases}
\]
\[
\begin{aligned}
&{-\xi(a_2 - u) - (a_2 - a_1)} & u < a_1 \\
&{-\eta(u - a_2) - (a_3 - a_1)} & a_1 < u < a_3 \\
&{-\eta(u - a_3) - (a_3 - a_1)} & u > a_3
\end{aligned}
\]

If one considers \( u \) as an unknown parameter then \( \hat{u} \) is simply the maximum likelihood estimate.

The inverse estimator (10.19) given by
\[
u_f = \bar{y} + S(t, x) \left( y - \bar{x} \right)
\] (10.40)

has also been proposed. Formally this can be obtained by fitting a line \( t = \gamma + \delta x \) to the data by means of least squares. This estimate has the advantage over \( \hat{u} \) of having a finite mean square error, at least for \( n \geq 4 \).

Recall that \( \hat{u}_f \) is the mean value of the calibrative distribution (10.37). From a Bayesian viewpoint the inverse estimator can be thought of as a shift of \( \hat{u} \) towards the prior mean which gives the smallest adjustments to \( \hat{u} \) when the data are most informative. We note that in all situations \(|\hat{u}_f - \bar{t}| < |\hat{u} - \bar{t}|\), with equality only in the case of an exact relationship. Thus the more informative the data the more we move from the prior mean towards the estimate \( \hat{u} \).

Several authors have decried the use of \( u_f \) (for example, Berkson (1969), Williams (1969a), Halperin (1970), Martinelle (1970)), stating categorically...
that \( \hat{u} \) is to be preferred to \( u_l \) despite the fact that \( \hat{u} \) has an undefined mean and infinite mean square error.

Difficulties can also be encountered in the classical confidence interval approach although in most practical situations the analysis is straightforward. The classical confidence region is obtained from the fact that

\[
\frac{y - \bar{x} - \hat{\beta}(u - i)}{\left(1 + \frac{1}{n} + \frac{(u - i)^2}{S(t, t)}\right)^{1/2}} \left(\frac{v}{n-2}\right)^{1/2} \text{ is } \text{St}(n - 2, 0, 1),
\]

so that a 100 \((1 - \alpha)\) per cent region is given by

\[
\left\{ u: \frac{(n-2)\hat{\beta}^2(u - u)^2}{v \left[1 + \frac{1}{n} + \frac{(u - i)^2}{S(t, t)}\right]} < F(1, n - 2; 1 - \alpha) \right\}.
\]

This yields a confidence region of one of three possible forms which depend on the value of

\[
R = \frac{(n-2)\hat{\beta}^2 S(t, t)}{v}.
\]

We have the regions

\[
\{ u: u_1 < u < u_2 \} \quad \text{if } R > F(1, n - 2; 1 - \alpha),
\]

\[
\left\{ u: u \leq u_1 \text{ or } u \geq u_2 \right\} \quad \text{if } R \frac{1 + \frac{1}{n} + \frac{(i - i)^2}{S(t, t)}}{1 + \frac{1}{n} + \frac{(u - i)^2}{S(t, t)}} x F(1, n - 2; 1 - \alpha) < R < F(1, n - 2; 1 - \alpha),
\]

\[-\infty, \infty \] \quad \text{if } R \frac{1 + \frac{1}{n} + \frac{(i - i)^2}{S(t, t)}}{1 + \frac{1}{n} + \frac{(u - i)^2}{S(t, t)}} x F(1, n - 2; 1 - \alpha),
\]

where \( u_1, u_2 (u_1 < u_2) \) are given by

\[
\bar{t} + \frac{\hat{\beta}(y - \bar{x})}{w} \pm \frac{1}{w} \left(\frac{v}{n-2}\right)^{1/2} F(1, n - 2; 1 - \alpha)
\]

\[
\times \left(\left[1 + \frac{1}{n} + \frac{(y - \bar{x})^2}{S(t, t)}\right]^{1/2}
\right)
\]

(10.43)
Calibration

with

\[ w = \beta^2 - \frac{vF(1, n-2; 1-\alpha)}{(n-2)S(t, t)}. \]

There is thus a chance that the 100(1 - \alpha) per cent confidence region will be given by (\(-\infty, \infty\)) and so be rendered useless. This unlimited region occurs when the hypothesis \(\beta = 0\) cannot be rejected on the information available. Of course the probability of this may be very small, thus negating the importance of this contingency. In much the same way the fact that \(\tilde{u}\) has undefined mean and infinite mean square error may not be disastrous — the distribution of \(\tilde{u}\) may still be quite well behaved.

Hoadley (1970) suggests a third reason why \(\tilde{u}\) may be considered unsatisfactory. The data from \(e\) contain some information about the precision of \(\tilde{u}\). Thus if \(\tilde{u}\) is known to be unreliable, less weight or importance should be attached to it. This suggests that the Bayesian approach is more realistic.

A similar situation exists when we consider the case of the \(K\)-replicate future experiment discussed in §10.3. Here three point estimates have been proposed, namely

(i) the classical estimate

\[ \tilde{u} = \bar{t} + \frac{M - \bar{x}}{\beta}; \tag{10.44} \]

(ii) the inverse estimate suggested by Krutchkoff (1967)

\[ \tilde{u} = \bar{t} + \frac{S(t, x)}{S(x, x)} (M - \bar{x}) \]

\[ = \bar{t} + \frac{S(t, x)}{\beta^2 S(t, t)} + v (M - \bar{x}); \tag{10.45} \]

(iii) Halperin's (1970) family of modified inverse estimates

\[ u(r) = \bar{t} + \frac{rS(t, x)}{r_2 S(t, t)} (M - \bar{x}). \tag{10.46} \]

These compare with the mean \(u_1\) of the calibrative distribution (10.37) where

\[ u_1 = \bar{t} + \frac{S(t, x)}{S(x, x)} + v (M - \bar{x}). \tag{10.47} \]

The balance of opinion favours (10.44) from the three classical estimates (10.44), (10.45) and (10.46). The point estimate \(u_1\) has one appealing property. The variation in \(y_1, y_2, \ldots, y_K\), as measured by \(V\), is incorporated in the estimate in such a way that the more variation there is in the future observations the nearer is the predictive mean to \(\bar{t}\). From a Bayesian viewpoint with prior mean \(\bar{t}\) this is an intuitively pleasing property for one should be less
willing to change one's prior view if the additional data are very variable. None of the other estimates take this variation into account in obtaining a point estimate. Notice that

$$|u_f - \bar{r}| \leq |\tilde{u} - \bar{r}| \leq |u - \bar{r}|,$$

so that $u_f$ places less weight on $\tilde{u}$ than $\tilde{u}$ does and is less removed from the sample mean $\bar{r}$ than $\bar{u}$. If $r > 1$ (and one of the important cases considered by Halperin is where $r = K$) then we have further that

$$|u_f - \bar{r}| \leq |\tilde{u} - \bar{r}| \leq |u(r) - \bar{r}| \leq |\tilde{u} - \bar{r}|.$$

As in the single replicate case difficulties can be encountered in obtaining a confidence interval. The extension from (10.42) is straightforward. Suffice it to say that for suitable cases, that is if

$$\frac{(n + K - 3)}{v + V} \frac{\hat{\beta}^2 S(t, t)}{F(1, n + K - 3; 1 - \alpha)},$$

the 100(1 - $\alpha$) per cent confidence region is given by

$$\bar{r} + \frac{\hat{\beta}(M - \bar{x})}{W} \pm \frac{1}{W} \left( \frac{v + V}{n + K - 3} \right) F(1, n + K - 3; 1 - \alpha)$$

$$\times \left( \frac{1}{n + K} \right) W + \frac{(M - \bar{x})^2}{S(t, t)} \right) 1/2$$

(10.48)

where

$$W = \hat{\beta}^2 \frac{(v + V) F(1, n + K - 3; 1 - \alpha)}{(n + K - 3) S(t, t)}.$$

This interval is not positioned symmetrically about the classical estimate $\bar{u}$. The variation in the values $y_1, y_2, ... , y_K$ is now taken into account, and it can be seen that, for a given $K$, the more the variation the wider is the corresponding confidence interval.

Halperin's estimate $u(r)$ does not allow an exact interval estimate of $u$. He feels that in most practical situations there would be little to choose between $u(r)$ and $\tilde{u}$, and so prefers $\tilde{u}$ since an exact interval estimate is usually obtainable.

As a simple illustration to compare the methods we consider again example 10.1. The results of the informative experiment are shown in fig. 10.1. We will consider several different situations for the 'future' experiment in each of which $M = 18$, but where $K$ and the variation in $y_1, y_2, ... , y_K$ alter. These specifications are listed in table 10.4.

**Point prediction.** The classical estimate $\hat{u}$ (10.44), the inverse estimate $\tilde{u}$ (10.45) and Halperin's estimate $u(r)$ (10.46) are invariant for all the
Calibration

Table 10.4. Several future experiments in example 10.1 with \( M = 18 \).

<table>
<thead>
<tr>
<th>Specification</th>
<th>1</th>
<th>2(i)</th>
<th>2(ii)</th>
<th>2(iii)</th>
<th>3(i)</th>
<th>3(ii)</th>
<th>3(iii)</th>
<th>3(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>( Y )</td>
<td>18</td>
<td>17.5</td>
<td>17.0</td>
<td>16.0</td>
<td>17.8</td>
<td>17.6</td>
<td>17.0</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td>19.5</td>
<td>19.0</td>
<td>20.0</td>
<td>18.0</td>
<td>18.2</td>
<td>18.6</td>
<td>19.0</td>
<td>19.0</td>
</tr>
<tr>
<td>( V )</td>
<td>0</td>
<td>0.50</td>
<td>2.0</td>
<td>8.0</td>
<td>0.08</td>
<td>0.56</td>
<td>2.00</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Table 10.5. Bayesian and classical 95 per cent set prediction

<table>
<thead>
<tr>
<th>Specification</th>
<th>Bayesian (10.49)</th>
<th>Classical (10.48)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(21.40, 29.56)</td>
<td>(21.22, 29.63)</td>
</tr>
<tr>
<td>2(i)</td>
<td>(22.61, 28.35)</td>
<td>(22.47, 28.39)</td>
</tr>
<tr>
<td>2(ii)</td>
<td>(22.54, 28.42)</td>
<td>(22.39, 28.46)</td>
</tr>
<tr>
<td>2(iii)</td>
<td>(22.30, 28.68)</td>
<td>(22.11, 28.74)</td>
</tr>
<tr>
<td>3(i)</td>
<td>(23.17, 27.78)</td>
<td>(23.05, 27.80)</td>
</tr>
<tr>
<td>3(ii)</td>
<td>(23.16, 27.80)</td>
<td>(23.04, 27.82)</td>
</tr>
<tr>
<td>3(iii)</td>
<td>(23.11, 27.86)</td>
<td>(22.98, 27.88)</td>
</tr>
<tr>
<td>3(iv)</td>
<td>(22.97, 28.00)</td>
<td>(22.82, 28.03)</td>
</tr>
</tbody>
</table>

specifications. We find that

\[ \hat{u} = 25.44, \quad \tilde{u} = 25.48 \]

and

\[ r \]

<table>
<thead>
<tr>
<th>( u(r) )</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.48</td>
<td>25.48</td>
<td>25.46</td>
<td>25.45</td>
<td>25.44</td>
<td>25.44</td>
<td>25.44</td>
</tr>
</tbody>
</table>

The only estimate which varies with \( K \) and \( V \) is \( u_1 \) (10.47), and we obtain

<table>
<thead>
<tr>
<th>Specification</th>
<th>1</th>
<th>2(i)</th>
<th>2(ii)</th>
<th>2(iii)</th>
<th>3(i)</th>
<th>3(ii)</th>
<th>3(iii)</th>
<th>3(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>25.48</td>
<td>25.48</td>
<td>25.48</td>
<td>25.49</td>
<td>25.48</td>
<td>25.48</td>
<td>25.48</td>
<td>25.49</td>
</tr>
</tbody>
</table>

We see that for given \( K \) the more variation there is in \( Y_1, Y_2, \ldots, Y_K \), the less we move away from the prior mean \( t = 26.30 \).

Set Prediction. No difficulty arises in the classical confidence interval approach in this situation. In table 10.5 we give the 95 per cent intervals for each of the specifications. We also give the 95 per cent (shortest) Bayesian intervals, obtained from (10.37) and given by

\[ u_1 \pm \sqrt{Q(\tau, y)} t (n + K - 3; 1 - \frac{1}{2} \alpha). \]  

(10.49)

In each case we see that the Bayesian interval is the shorter.
Table 10.6 Bayesian set predictions from utility functions

<table>
<thead>
<tr>
<th>Specification</th>
<th>1</th>
<th>2(ii)</th>
<th>3(iii)</th>
<th>Bayesian cover</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>(19.83, 31.13)</td>
<td>(21.41, 29.56)</td>
<td>(22.19, 28.77)</td>
<td>0.99</td>
<td>200</td>
</tr>
<tr>
<td>0.05</td>
<td>(21.48, 29.48)</td>
<td>(22.59, 28.38)</td>
<td>(23.13, 27.83)</td>
<td>0.95</td>
<td>40</td>
</tr>
<tr>
<td>0.10</td>
<td>(22.30, 28.68)</td>
<td>(23.18, 27.79)</td>
<td>(23.61, 27.35)</td>
<td>0.88</td>
<td>17</td>
</tr>
<tr>
<td>0.20</td>
<td>(23.30, 27.66)</td>
<td>(23.90, 27.06)</td>
<td>(24.19, 26.77)</td>
<td>0.73</td>
<td>7</td>
</tr>
<tr>
<td>0.30</td>
<td>(24.12, 26.83)</td>
<td>(24.49, 26.47)</td>
<td>(24.68, 26.29)</td>
<td>0.51</td>
<td>4</td>
</tr>
</tbody>
</table>

If the Bayesian decision theory approach is used the specification of $\gamma$ or $\xi$, $\eta$ in the utility functions in table 10.3 is of importance. Table 10.6 shows the intervals obtained for three of the specifications from table 10.4 for a selection of suitable values of $\gamma$. In column 5 the (approximate) predictive probabilities of the selected regions are shown. (These can easily be found by use of (10.37) and table 9 in Pearson and Hartley, 1966.) It so happens that, for a given $\gamma$, the predictive probabilities are almost identical for each specification. This is because there is not very much difference between the $t$-distributions with 14, 15 and 16 degrees of freedom.

Suppose we use the last utility of table 10.3. For simplicity we assume $\xi = \eta$. The approximate values of $\xi$ which correspond to the intervals obtained in table 10.6 are shown in column 6 of that table.

Basically the imposition of the extra structure afforded by the utility functions leads to different ways of summarising the predictive density function $p(u|\gamma, \xi)$.

10.8 An application to antibiotic assay

The analysis of example 1.6 (calibration by biological assay) may now be completed with the already familiar tools developed earlier in this chapter. We need draw attention only to some particular aspects.

First from the scatter diagram of fig. 1.2 we see that the regression of clearance diameter on the logarithm of concentration is not linear over the whole range of concentrations explored. Some alternatives are open to us. For example, we may attempt to adopt a suitable non-linear form for the regression, or try to obtain some transformation of the clearance diameters which will linearise the regression, or restrict our operations to the existing linear stretch of fig. 1.2. It is often convenient in assay problems to adopt the third alternative since there is usually the opportunity of investigating successive dilutions of a specimen and so obtaining a particular dilution lying within the linear part of the concentration range. For our particular problem we confine attention to the range $t = 1$ to $t = 3$ for $\log_2$ (concentration). The informative experiment then consists of 60 trial records (the central three columns of table 1.5), and the usual regression calculations in the notation of §10.3 give:
The calibration experiment is here designed and so we have to choose a sensible \( p(u) \) from other considerations. If preliminary dilutions have convinced us that the current specimen has a concentration within the range \( 1 < u < 3 \) then we should confine \( p(u) \) to this range; the uniform distribution over this range may then be a reasonable basis for calibration. If, however, there is the possibility that the concentration of the specimen lies outside the range \( 1 < u < 3 \) then \( p(u) \) must reflect this possibility. There is indeed a certain attraction in this second alternative since it provides a useful form of monitoring whether further dilution is required. We therefore adopt this alternative and assume that \( p(u) \) is the improper uniform prior over the real line. We can thus set \( p(u) = 1 \) in (10.15). If we construct the calibrative density function \( p(u | y, z) \) on this basis and find that it does not assign high plausibility to the calibration range \( 1 < u < 3 \) then we require further dilution or possibly have over-diluted, and the calibrative density function will indicate which of these is the appropriate conclusion. In other words, we can construct a rule for reassaying the specimen at a different dilution of the following form: if

\[
\int_1^3 p(u | y, z) \, du > c
\]

use the calibrative density function \( p(u | y, z) \) for the assay; otherwise reassay at a different dilution.

We now proceed to answer the three questions posed in example 1.6 on the basis of the vague prior on \( (\alpha, \beta, \tau) \) or equivalently on \( (\mu, \tau) \) where

\[ \mu = \alpha + \beta u. \]

(1) In the notation and terminology of §10.3 we have a single replicate of \( f \) giving an observation \( y = 19 \). The predictive density function required for (10.15) is then of the form (10.16) and using the regression calculations already made we obtain

\[
p(y | u, z) = St \{ 58; 15.1 + 1.94u, 0.847 + 0.0208(u - 2.00)^2 \}.
\]

Application of the technique used in (10.30) with \( y = 19 \) gives the calibrative density function shown in fig. 10.7.

(2) For this calibration we have 3 clearance diameters so that in the notation and terminology of (10.21)–(10.24) we have

\[
K = 3, \quad M = 19.0, \quad V = 1.50,
\]

and

\[
p(M, V | u, m, v) = StSi \{ 58; 15.1 + 1.94u, 0.292 + 0.0208(u - 2)^2; 2, 48.3 \}.
\]
Again it is easy to apply (10.30) for the given \((M, V)\) to obtain the more precise calibrative density function also shown in fig. 10.7.

(3) To resolve this problem we have to determine how large \(K\) must be to ensure that the resulting calibrative density is highly concentrated round its mode. If we interpret 'reasonably' to be 90 per cent then we would require 90 per cent of the plausibility assigned by the calibrative density function to be within 10 per cent of its modal concentration, or in terms of the \(\log_2 (\text{concentration})\) scale within \(\log_2 0.9 = -0.15\) below and within \(\log_2 1.1 = 0.14\) above the modal \(u\). To investigate this we can easily apply the graphing technique to representative data based on different \(K\), say with

\[
M = 19, \quad V = (K - 1) \frac{v}{n - 2}
\]

until we obtain the desired accuracy, in the present case until we obtain by numerical integration.
Calibration

\[ \int_{1.85}^{2.14} p(u | y, z) du > 0.90. \]

By this technique we find that it is necessary to take \( K \) as big as 60 to obtain such accuracy and this is unlikely to be a practical proposition.

History

Calibration is a topic which many textbooks containing details of regression methods tend to avoid. Two which escape this criticism are Bowker and Lieberman (1959) and Brownlee (1960).


Eisenhart (1939) had earlier discussed both the classical and inverse estimates for bivariate situations and concluded that it was clear which should be chosen in any circumstance, the decision resting simply on the nature of \( t_1, t_2, \ldots, t_n \).


Problems

10.1 Complete the analysis of Problem 1.5.

10.2 In a new production process items arrive at an inspection point according to a Poisson process with parameter \( \psi \). The inspector either passes the item or rejects the item, in which case he channels it back for reprocessing. Suppose the individual items each have probability \( \theta \) of being accepted. Little is known initially about \( \theta \) and \( \psi \). A count is made both of acceptable items and scrapped items during \( n \) trial eight-hour shifts. Suppose that in future only the acceptable items produced during a shift are counted. Can you estimate the number of scrapped items?

10.3 In a medical experiment testing the reactions of patients the number \( x \) of current responses given in a fixed period of time is related to the amount \( t \) of a stimulus in the patient's bloodstream. Suppose that \( p(x | t, \theta) \) is \( \text{Po}(t \theta) \). It is difficult to evaluate \( t \) and in future the amount of stimulus is to be estimated from the number of correct responses which the patient gives. To obtain information the experiment is simulated by injecting a series of patients with the dosages shown below and recording the numbers of correct responses.
For a new patient who records 5 correct answers what can be said about the level of stimulus in his bloodstream if initially it was thought equally likely to be any value in the range 0 to 1 ml?

\[
\begin{array}{cccccccc}
 r (\text{ml}) & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\
 x & 2 & 3 & 2 & 4 & 4 & 6 & 6 & 8 & 7
\end{array}
\]

10.4 An inexpensive quick chromatographic method for determining the excretion rate (mg/24h) of a certain steroid metabolite in urine has been developed. It is hoped that this method may in future replace the long and costly, though accurate, bioassay technique currently used. The considerable past experience of bioassays has shown that the excretion rates analysed are approximately normally distributed with mean 2 mg/24h and standard deviation 0.5 mg/24h. To explore the possibilities of the new method, aliquots from a number of urine samples are available. The experimenter has made bioassay determinations on one aliquot from each urine sample and selected a subset which he felt gave adequate coverage of the range of excretion rates. Three other aliquots from each urine sample of this subset were then assigned to the chromatographic method and the results are shown in the table.

Explore the possibilities of using the chromatographic method in future.

<table>
<thead>
<tr>
<th>Serial no. of urine sample</th>
<th>Excretion rate (mg/24h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bioassay method</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>1.20</td>
</tr>
<tr>
<td>4</td>
<td>1.40</td>
</tr>
<tr>
<td>5</td>
<td>1.60</td>
</tr>
<tr>
<td>6</td>
<td>1.80</td>
</tr>
<tr>
<td>7</td>
<td>2.00</td>
</tr>
<tr>
<td>8</td>
<td>2.20</td>
</tr>
<tr>
<td>9</td>
<td>2.40</td>
</tr>
<tr>
<td>10</td>
<td>2.60</td>
</tr>
<tr>
<td>11</td>
<td>2.80</td>
</tr>
<tr>
<td>12</td>
<td>3.30</td>
</tr>
<tr>
<td>13</td>
<td>3.50</td>
</tr>
</tbody>
</table>

10.5 The gain in weight of a rat depends on the amount of a certain vitamin in its diet. A random sample of 25 batches of feed are known to contain concentrations \( t_1, t_2, \ldots, t_{25} \) of the vitamin. In the informative experiment 25 rats were used, each rat being assigned to a different batch, and the weight gains \( x_1, x_2, \ldots, x_{25} \) shown below are noted after each has received the same amount of feed. The concentration of the vitamin in a further batch is unknown. The weight gains of \( K \) rats fed from this batch with the same amount
of feed as before are shown below for several specifications, each of which has $M = 8.4$. Compare the different calibrations of §10.7 for the specifications given.

<table>
<thead>
<tr>
<th>Specification</th>
<th>1</th>
<th>2(i)</th>
<th>2(ii)</th>
<th>2(iii)</th>
<th>3(i)</th>
<th>3(ii)</th>
<th>3(iii)</th>
<th>3(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$y = (y_1, \ldots, y_K)$</td>
<td>8.4</td>
<td>8.2</td>
<td>7.9</td>
<td>7.4</td>
<td>8.3</td>
<td>8.0</td>
<td>7.7</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>8.6</td>
<td>8.9</td>
<td>8.4</td>
<td>8.4</td>
<td>8.2</td>
<td>8.0</td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10.6 In the bioassay of an antibiotic by the clearance circle technique the following table gives the clearance diameters associated with standard antibiotic preparations of known dilution, together with the six clearance diameters from each of two antibiotic specimens of unknown 'dilutions'. What can usefully be inferred about these unknown dilutions?

<table>
<thead>
<tr>
<th>Standard antibiotic preparations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilution</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$\frac{1}{8}$</td>
</tr>
<tr>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>$\frac{1}{32}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Antibiotic specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specimen no</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

10.7 The calibration problem associated with (10.21) (10.24) envisages an informative experiment with $m$ and $v$ distributed independently as $\text{No}(\mu, K\tau)$ and $\text{Ch}(v, \tau)$ and a 'future' experiment $f_r$ with $M$ and $V$ distributed independently as $\text{No}(\mu, K\tau)$ and $\text{Ch}(K - 1, \tau)$. Show that the calibrative density function is
Calibration

exactly the same as that obtained from an informative experiment $e'$ with $m$ and $v + V$ distributed independently as $\text{No}(\mu, k\tau)$ and $\text{Ch}(v + K - 1, \tau)$ and a future experiment $f'_u$ yielding only $M$ which is $\text{No}(\mu, K\tau)$. Can you provide an intuitive argument for this result?
11

Diagnosis

11.1 The nature of a diagnostic problem

Although for this chapter we have used a title which has often a medical connotation the problem arises in many other fields — for example, in the diagnosis of a fault in a complex industrial process, in categorising an archaeological or anthropological specimen. From an expository point of view, however, the nature of a diagnostic problem is most easily described, and the corresponding theory is best developed, within the context of a specific situation. For this purpose we have selected a medical problem concerning the differential diagnosis of three forms or types of a particular syndrome on the basis of two diagnostic tests or observable features. We have deliberately selected this three-type two-feature problem because it allows the maximum exploitation of diagrammatic means of expressing concepts and analyses. All the concepts and analyses carry over straightforwardly into higher dimensional problems. Indeed the introductory illustrative problem which we now present is a subproblem extracted from a larger real one.

Example 11.1

Differential diagnosis of Cushing's syndrome. Cushing's syndrome is a rare hypersensitive disorder associated with the over-secretion of cortisol by the adrenal cortex. For illustrative purposes we confine ourselves here to three 'types' of the syndrome, those types in which the cause of this over-secretion is actually within the adrenal gland itself. The types are

\[ a: \text{adenoma,} \]
\[ b: \text{bilateral hyperplasia,} \]
\[ c: \text{carcinoma,} \]

and we investigate the possibilities of distinguishing the types on the basis of two observable 'features', the determination by paperchromatography of the urinary excretion rates (mg/24h) of two steroid metabolites, tetrahydrocortisone and pregnanetriol. Table 11.1 gives these rates for 21 patients with Cushing's syndrome, who in the past have all been operated on and the particular type \( a, b \) or \( c \) histopathologically determined. For each of these past cases the verified type of the syndrome and the feature vectors can be represented by a
Table 11.1 Urinary excretion rates (mg/24h) of two steroid metabolites for 21 patients with Cushing's syndrome

<table>
<thead>
<tr>
<th>Case no.</th>
<th>Tetrahydrocortisone</th>
<th>Pregnanetriol</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>3.1</td>
<td>11.70</td>
</tr>
<tr>
<td>a2</td>
<td>3.0</td>
<td>1.30</td>
</tr>
<tr>
<td>a3</td>
<td>1.9</td>
<td>0.10</td>
</tr>
<tr>
<td>a4</td>
<td>3.8</td>
<td>0.04</td>
</tr>
<tr>
<td>a5</td>
<td>4.1</td>
<td>1.10</td>
</tr>
<tr>
<td>a6</td>
<td>1.9</td>
<td>0.40</td>
</tr>
<tr>
<td>b1</td>
<td>8.3</td>
<td>1.00</td>
</tr>
<tr>
<td>b2</td>
<td>3.8</td>
<td>0.20</td>
</tr>
<tr>
<td>b3</td>
<td>3.9</td>
<td>0.60</td>
</tr>
<tr>
<td>b4</td>
<td>7.8</td>
<td>1.20</td>
</tr>
<tr>
<td>b5</td>
<td>9.1</td>
<td>0.60</td>
</tr>
<tr>
<td>b6</td>
<td>15.4</td>
<td>3.60</td>
</tr>
<tr>
<td>c1</td>
<td>10.2</td>
<td>6.40</td>
</tr>
<tr>
<td>c2</td>
<td>9.2</td>
<td>7.90</td>
</tr>
<tr>
<td>c3</td>
<td>9.6</td>
<td>3.10</td>
</tr>
<tr>
<td>c4</td>
<td>53.8</td>
<td>2.30</td>
</tr>
<tr>
<td>c5</td>
<td>15.8</td>
<td>7.60</td>
</tr>
</tbody>
</table>

point in a two-dimensional diagram, as in fig. 11.1. In this diagram logarithmic scales have been used to accommodate the large proportional differences that occur in these excretion rates. This diagram can be thought of as representing past experience or past case records; the results indeed constitute our informative experiment $e$. In fig. 11.1 there is clearly some degree of separation of the three types and so some hope that the excretion rates will be of some diagnostic value for future cases.

If we denote by $f_t$ the 'experiment' which records the two urinary excretion rates of a patient who has type $t$, then when we know that we are dealing with a Cushing patient we are performing an experiment from the class

$$F = \{f_t : t \in T\}, \text{ where } T = \{a, b, c\},$$

of possible experiments. We thus can write

$$e = \{f_{t_1}, \ldots, f_{t_{21}}\},$$

a set of 21 independent component experiments, where $t_1, \ldots, t_{21}$ are the known types of the 21 patients.

Suppose that a new patient, who is already known to have Cushing's syndrome but of as yet unknown type, has urinary excretion rates of 9.0 mg/24h of tetrahydrocortisone and of 1.50 mg/24h of pregnanetriol. On the basis of our past experience — the informative experiment $e$ — and on the basis of the
Fig. 11.1 Scatter diagram of urinary excretion rates of tetrahydrocortisone and pregnanetriol.

- adenoma
- bilateral hyperplasia
- carcinoma
Diagnosis

outcome of these diagnostic tests on the new patient what plausibilities ought we to attach to the three types for this new patient? This is the nature of a diagnostic problem. If we denote the unknown type of the new patient by \( u \) then his results on the diagnostic tests constitute an outcome of a performance of \( f_u \). We thus have a problem analogous to the calibration problem of chapter 10. We have already performed 'future' experiment \( f_u \) from the class \( F \) but we do not know whether \( u \) is \( a, b \) or \( c \). The only difference between calibration and diagnosis lies in the fact that the set \( T \) is usually a continuum in a calibration problem whereas \( T \) is a finite set in a diagnostic problem.

We have described diagnosis as the assessment of the plausibilities of the various possible types rather than the definite allocation of the patients to one of the types. In other words we are regarding diagnosis more as a problem of inference than one of decision. We shall return to this point in § 11.7, where we discuss the problem of diagnosis under a utility structure, and again in chapter 12, where diagnosis is seen as a kind of mental resting place in the search for a suitable treatment.

11.2 The diagnostic distribution

Just as in §10.2 we recognised the calibrative distribution as a sensible means of describing the plausibility of the possible indices of the new subject, so we shall be able to construct a diagnostic distribution which, for a new case, assigns plausibilities to each of the finite set of types. As in our treatment of calibration we again set down explicitly a set of assumptions as a basis for the development of statistical diagnosis. For many diagnostic situations only eight simple assumptions seem to be required, and acceptance of these has far-reaching consequences. We do not imply that all diagnostic problems conform to this pattern; our purpose is rather to emphasise the underlying assumptions so that it is easier to examine the relevance of the predictive method to a new situation and to pinpoint where any adjustment may have to be made.

For convenience of reference we first present compactly the eight assumptions, then discuss their interpretation and relevance, and finally follow through their consequences.

Assumptions

D1
Each case belongs to one and only one of a finite set \( T = \{1, \ldots, r\} \) of possible types.

D2
For each case a finite set \( G = \{1, \ldots, d\} \) of possible features may be observed, any observed feature vector falling within a given sample space \( X \).
The class of probabilistic models considered as possible descriptions of the generation of case records (that is, types and associated feature vectors) is indexed by a finite-dimensional vector \((\Psi, \theta)\) with \(\Psi \in \Psi, \theta \in \Theta\).

In the following three assumptions \((t, x)\) with \(t \in T, x \in X\) is a typical case record, and \(z = \{z_1, \ldots, z_n\}\) any given set of \(n\) case records.

\[p(t|\Psi, \theta) = p(t|\Psi) \quad (t \in T) \text{ for every } \Psi \in \Psi, \theta \in \Theta.\]
\[p(x|t, \Psi, \theta) = p(x|t, \theta) \quad (x \in X) \text{ for every } t \in T, \Psi \in \Psi, \theta \in \Theta.\]
\[p(z|\Psi, \theta) = \prod_{p=1}^{n} p(z_p|\Psi, \theta) \text{ for every set } z = \{z_1, \ldots, z_n\}\]

of case records.

\[p(\Psi, \theta) = p(\Psi)p(\theta) \quad (\Psi \in \Psi, \theta \in \Theta).\]

For a case of unknown type \(u\)
\[p(u|\Psi) = \psi_u (u = 1, \ldots, d),\]
where \(\psi_u\) is the \(u\)th component of \(\Psi\), and
\[\Psi = \{\psi: \psi > 0, \Sigma_T \psi = 1\}\]
is the \(d\)-dimensional simplex.

There are two aspects of assumption D1 — the exhaustiveness and exclusiveness of the categories of \(T\). The first aspect asserts that a case arriving for diagnosis by a system based on \(T\) does fall into one of the types of \(T\). In example 11.1 if we adopt \(T = \{a, b, c\}\) we have then no direct means of categorising say a patient who displays all the symptoms of Cushing's syndrome but who turns out to have a fourth type \(d\) of the disease. If we are to use this assumption wisely therefore we will choose as sets of types for consideration only those which have this property of exhaustiveness. Even for a set \(T\) chosen for this property we would still be wise to allow for the possibility of misdirected cases. We shall see in §11.4 that we can to some extent monitor for this possibility of wrong referral. The second aspect of exclusiveness asserts that a case cannot belong to more than one type, or, in medical terminology, have a dual or multiple pathology. If in example 11.1 it were possible for a patient to have both an adenoma and bilateral hyperplasia then to meet the assumption we would have to consider four categories (a) adenoma only, (b) hyperplasia only,
Diagnosis

(c) carcinoma, (d = ab) adenoma and hyperplasia.

Assumption D2 simply asserts that for each case there are certain features (for example, (1) urinary excretion rate of tetrahydrocortisone, (2) diastolic blood pressure, (3) sex, and so on) which may be observed. The sample space \( X \) in a medical diagnostic situation may be large.

Assumption D3 expresses the usual statistical hope that we can describe the variability we observe in the case records in terms of some family of distributions indexed by a finite parameter. The partition of the parameter into \( \Psi \) and \( \Theta \) is given meaning by the following two assumptions.

Assumption D4 asserts that, as far as the probabilistic mechanism which determines the type of a case is concerned, if we know \( \Psi \) then we need not know \( \Theta \). In other words the parameter \( \Psi \) characterises the arrival pattern of types and so we shall term it the arrival parameter.

For diagnosis to be feasible we must hope that the way in which feature vectors arise depends to some extent on the type of the case. Assumption D5 states that this conditional distribution of feature vector given the type does not depend on the arrival parameter. In other words \( \Theta \) is that aspect of the parameter which characterises the dependence of feature vector on type. For this reason we refer to \( \Theta \) as the structural parameter.

If in example 11.1 we were to make the assumption that the three distributions of feature vectors associated with the three types are bivariate normal with mean vectors \( \mu_1, \mu_2, \mu_3 \) and covariance matrices \( \Sigma_1, \Sigma_2, \Sigma_3 \) then the structural parameter \( \Theta \) could be taken as \((\mu_1, \Sigma_1, \mu_2, \Sigma_2, \mu_3, \Sigma_3)\). In this case the distribution of feature vectors for the first type is \( N_{2d}(\mu_1, \Sigma_1^{-1}) \) and so \( p(x|t = a, \Theta) \) could be reduced to \( p(x|t = a, \mu_1, \Sigma_1) \), but the form of the theoretical development is kept clearer if we retain the complete \( \Theta \) in the parametrisation of each of the feature vector distributions.

Assumption D6 states that, for given arrival and structural parameters, case records are statistically independent. While this is a sufficiently realistic assumption in a great number of diagnostic problems it would require careful reconsideration if any of the types were contagious or had a substantial genetic character. For example, if in two case records \((t_1, x_1), (t_2, x_2)\) the vectors \(x_1\) and \(x_2\) provide evidence of a family relationship then in such a situation there might be an implied dependence between \(t_1\) and \(t_2\).

Assumption D7 states that prior to investigation of past case records any information that we have concerning the arrival pattern is independent of any information we have concerning the feature structure of the types. We shall see that the prior independence postulated in this assumption persists in a posterior form when the data on which updating is based consists of case records, which by definition are associated with cases of known types. In its posterior form it appears to be a tacit assumption of most statistical diagnostic methods. We emphasise that at this point of the formulation we are making no
other particular assumptions about the form of \( p(\Psi) \) and \( p(\Theta) \).

The nature of the arrival parameter is clarified by assumption D8. For a given arrival pattern vector \( \Psi \) the \( r \) components \( \psi_u (u = 1, \ldots, r) \) are the probabilities we immediately associate with the \( r \) possible types without any additional evidence. The vector \( \Psi \) is thus a probability vector, with elements summing to unity, which justifies the restriction to the \( r \)-dimensional simplex \( \Psi \).

In the analysis of calibration problems we had to distinguish between natural and designed calibration experiments, and we require to make the same distinction here. To achieve this we can consider assumption D4 in greater detail. As it stands this assumption does not specify the way in which the probabilistic mechanism determining the type of a past case depends on \( \Psi \). If the past case records constituting the informative experiment have arisen naturally with the same arrival pattern as is anticipated for new cases then we would make the further assumption that

\[
p(t|\Psi) = \psi_t
\]

for a past case record \((t, x)\). In some circumstances such a 'natural' informative experiment may be difficult to achieve. For example, if the incidence of one of the types is small then in order to obtain sufficient information on the feature structure for that type it may be necessary to seek the referral of such types from sources other than those immediately available. For such selected past case records we have a 'designed' informative experiment, and we can recognise this feature in assumption D4 by making the further assumption that

\[
p(t|\Psi) = p(t).
\]

We can thus retain all our assumptions while recognising that D4 simply asserts a possible dependence on \( \Psi \). For a natural informative experiment (11.1) holds and the informative experiment will clearly alter our prior assertion \( p(\Psi) \). For a designed informative experiment (11.2) holds and our prior distribution \( p(\Psi) \) will be unaltered by such an experiment.

We now suppose that we have the case records \( z = \{z_1, \ldots, z_n\} \), where \( z_i = (t_i, x_i) \), of \( n \) diagnosed cases, and the feature vector \( y \) of an undiagnosed case of unknown type \( u \). Thus the complete parameter is \((\Psi, \Theta, u)\), the data are \((y, z)\) and the general problem to resolve is how a prior distribution \( p(\Psi, \Theta, u) \) on the parameter set is transformed to a posterior distribution \( p(\Psi, \Theta, u|y, z) \) on the basis of the data \((y, z)\) and the set \( D \) of assumptions D1–D8. The change is of course brought about by an application of Bayes's theorem with likelihood \( p(y, z|\Psi, \Theta, u) \), and so to obtain insight into the process we must first discover what the implications of \( D \) are for the prior distribution and the likelihood. The structure of the prior is readily obtained:
Diagnosis

\[ p(\Psi, \theta, u) = p(\Psi, \theta) p(u | \Psi, \theta) \]

\[ = p(\Psi, \theta) p(u | \Psi) \text{ by D4} \]

\[ = p(\Psi) p(\theta) p(u | \Psi) \text{ by D7.} \quad (11.3) \]

To obtain the likelihood we note that

\[ p(y, z | \Psi, \theta, u) = p(u, y, z | \Psi, \theta) / p(u | \Psi, \theta) \]

\[ = \frac{p(u, y | \Psi, \theta) \prod_{i=1}^{n} p(x_i | \Psi, \theta)}{p(u | \Psi, \theta)} \text{ by D6} \]

\[ = p(y | u, \Psi, \theta) \prod_{i=1}^{n} p(t_i | \Psi, \theta) p(x_i | t_i, \Psi, \theta) \]

\[ = p(y | u, \theta) \prod_{i=1}^{n} p(t_i | \Psi) \prod_{i=1}^{n} p(x_i | t_i, \theta) \text{ by D4 and D5} \]

\[ = p(y | u, \theta) p(t | \Psi) p(x | t, \theta) \quad (11.4) \]

in an obvious shortened notation, where \( t = (t_1, \ldots, t_n), x = (x_1, \ldots, x_n). \)

Then, by Bayes's Theorem and using the symbol \( \propto \) to indicate that the factor of proportionality does not depend on \( \Psi, \theta \) or \( u \), we have

\[ p(\Psi, \theta, u | y, z) \propto p(\Psi) p(\theta) p(u | \Psi) p(y | u, \theta) p(t | t, \theta) \]

\[ \propto p(\Psi | t) p(u | \Psi) p(\theta | z) p(y | u, \theta), \quad (11.5) \]

where

\[ p(\Psi | t) = p(\Psi) p(t | \Psi) / \int_{\Psi} p(\Psi) p(t | \Psi) d\Psi, \quad (11.6) \]

\[ p(\theta | z) = p(\theta | t, x) \]

\[ = p(\theta) p(x | t, \theta) / \int_{\theta} p(\theta) p(x | t, \theta) d\theta. \quad (11.7) \]

If we then write

\[ p(u | t) = \int_{\Psi} p(u | \Psi) p(\Psi | t) d\Psi, \quad (11.8) \]

\[ p(y | u, z) = \int_{\theta} p(y | u, \theta) p(\theta | z) d\theta, \quad (11.9) \]

we can express the final general result in the following form:

\[ p(\Psi, \theta, u | y, z, D) = \frac{p(\Psi | t) p(u | \Psi) p(\theta | z) p(y | u, \theta)}{\sum_{u \in T} p(u | t) p(y | u, z)} \quad (11.10) \]
Diagnosis

where the inclusion of the conditioning \( D \) is to emphasise the dependence of the form on the assumptions. The full generality of this result would be required for such problems as updating the diagnostic method on the basis of new cases of unconfirmed type (Aitchison, Habbema and Kay, 1975); for our immediate diagnostic objective we are interested in the marginal distribution for \( u \):

\[
p(u | y, z) = \frac{p(u | t) p(y | u, z)}{\sum_{u \in T} p(u | t) p(y | u, z)},
\]

(11.11)

the density function of the diagnostic distribution.

For a natural informative experiment an interpretation of (11.11) is that it is simply the conversion of an assessment \( p(u | t) \) after the types \( t \) of the past cases are known, but prior to any information concerning the \( n + 1 \) feature vectors \( y, z \), to a posterior assessment \( p(u | y, t, x) \) by way of Bayes's theorem and with \( p(y | u, t, x) \) or \( p(y | u, z) \) playing the role of the likelihood function. Special interest then centres on \( p(u | t) \) and \( p(y | u, z) \).

First we note that, from (11.8) and D8,

\[
p(u | t) = \int_\Psi p(\Psi | t) d\Psi = E(\Psi | t).
\]

(11.12)

Hence in so far as inference concerning the category of the new case is concerned uncertainty about \( \Psi \) is involved only in the form \( E(\Psi | t) \). We do not have to be able to provide a complete picture of the uncertainty in \( p(\Psi | t) \) but only the mean vector \( E(\Psi | t) \) of this distribution.

The distribution \( p(y | u, z) \) defined by (11.9) is the now familiar predictive distribution. For a new case in known category \( u \) it provides, on the basis of prior information \( p(\theta) \), the past records \( z \) and the assumptions \( D \), an assessment of the probabilities of the possible feature vectors \( y \) we may observe on the case.

For a designed informative experiment the consequences of (11.2), (11.6) and (11.8) are that

\[
p(u | t) = p(u),
\]

(11.13)

and we re-emphasise that the specification of \( p(u) \) must then be based on sources other than the informative experiment.

We shall apply the predictive diagnostic method expressed by (11.11) to the case where the distributions of feature vectors for each given type are multinormal, say \( N_d(\mu, \Sigma) \) for type \( t \). Suppose that of the \( n \) past case records \( n_t \) are of type \( t (t = 1, \ldots, r) \) so that \( n_1 + \ldots + n_r = n \). For the later application we record here the appropriate predictive distributions for vague
NoWi_d priors, and must distinguish between two situations. The first is where we make no assumption about the equality of the covariance matrices of the r feature distributions. If \( \bar{x}_t \) and \( S_t \) denote the vector mean and the covariance matrix for the \( n_t \) feature vectors of type \( t \) then \( m_t = \bar{x}_t \) and \( v_t = (n_t - 1)S_t \) are independently distributed as \( \text{NoW}_d(\mu_t, n_t \tau_t) \) and \( \text{Wi}_d(n_t - 1, \tau_t) \). We can then apply case 6 of table 2.3 to obtain

\[
p(y | u, z) = \text{St}_d \left( \frac{n_u - 1}{n_u} \right, m_u, \left( 1 + \frac{1}{n_u} \right) \frac{v_u}{n_u - 1} \right).
\]

(11.14)

The second situation is where we make the assumption that the covariance matrices of the r feature distributions are equal, say \( \tau_t = \tau \in \mathbb{T} \). If \( S \) is the pooled covariance matrix of r different sets of observed feature vectors,

\[
S = \sum_{t \in \mathbb{T}} (n_t - 1) S_t / (n - r),
\]

(11.15)

then \( v = (n - r)S \) is distributed as \( \text{Wi}_d(n - r, \tau) \) independently of \( m_t = \bar{x}_t \) \((t \in \mathbb{T})\). Then again by case 6 of table 2.3 we have

\[
p(y | u, z) = \text{St}_d \left( n - r, m_u, \left( 1 + \frac{1}{n_u} \right) \frac{v}{n - r} \right).
\]

(11.16)

We also record the easily verifiable fact that if for a natural informative experiment we adopt for \( p(\Psi) \) the vague Dirichlet distribution \( \text{Di}(g, h) \) with \( g \to 0, h \to 0 \) then

\[
p(u | t) = \frac{n_u}{n},
\]

(11.17)

the proportion of type \( u \) cases in the past case records.

11.3 An illustrative example

We now illustrate the use of the diagnostic distribution by applying the results of §11.2 to example 11.1. We make the assumption that the logarithms of the excretion rates are bivariate normal for each type and record the required summary of the data of table 11.1, first transformed by taking natural logarithms:

\[
n_1 = 6, \quad n_2 = 10, \quad n_3 = 5; \quad r = 3, \quad d = 2;
\]

\[
\bar{x}_1 = \begin{bmatrix} 1.0433 \\ -0.6034 \end{bmatrix}, \quad \bar{x}_2 = \begin{bmatrix} 2.0073 \\ -0.2060 \end{bmatrix}, \quad \bar{x}_3 = \begin{bmatrix} 2.7097 \\ 1.5998 \end{bmatrix};
\]
Fig. 11.2 Reference triangle for plausibility assessments about Cushing’s syndrome.

\[
S_1 = \begin{bmatrix}
0.11069 & 0.12389 \\
0.12389 & 4.08910
\end{bmatrix},
\quad S_2 = \begin{bmatrix}
0.21187 & 0.32413 \\
0.32413 & 0.72030
\end{bmatrix},
\quad S_3 = \begin{bmatrix}
0.55522 & -0.24224 \\
-0.24224 & 0.28850
\end{bmatrix},
\quad S = \begin{bmatrix}
0.26060 & 0.14265 \\
0.14265 & 1.56012
\end{bmatrix}.
\]

There is some evidence in \(S_1, S_2, S_3\) that it would be unreasonable to adopt the assumption that \(\tau_1 = \tau_2 = \tau_3\), so that we shall first use form (11.14) in our diagnostic assessment. The data in table 11.1 did not in fact arise from a natural informative experiment and so we shall quote results for the case where \(p(u) = 1/3\) for \(u = 1, 2, 3\). The diagnostic distribution for other \(p(u)\) is easily derived by a simple weighting of the quoted diagnostic distribution.

To present the results of the application of (11.11) we use a simple diagrammatic representation. In fig. 11.2 the equilateral triangle \(abc\) with unit altitude is such that for any point \(P\) within it the sum of the perpendiculars
Diagnosis

Fig. 11.3 Diagnostic assessments on the basis of unequal covariance matrices for the 21 past cases of Cushing’s syndrome.

- adenoma
- bilateral hyperplasia
- carcinoma

$P_a, P_b, P_c$ is unity. Thus any statement which places plausibilities $P_a, P_b, P_c$ on the three types $a, b, c$ can be represented by a unique point $P(P_a, P_b, P_c)$ in the triangle $abc$; and conversely each point in the triangle represents a unique plausibility statement about $a, b, c$. Roughly speaking with this representation the nearer a point is to a vertex the more plausible the type associated with that vertex is being regarded. The distribution (11.11) associated with any feature vector $y$ thus gives a point within triangle $abc$. Application of this diagnostic method to each of the feature vectors of table 11.1 produces 21 such points (fig. 11.3) and some measure of the potential of the method is given by the extent to which these points are close to the correct vertex.

For comparison purposes we show in fig. 11.4 the corresponding points which arise if the assumption of equal covariance matrices is adopted. As might have been anticipated in this particular example recognition of the
Fig. 11.4 Diagnostic assessments on the basis of equal covariance matrices for the 21 past cases of Cushing's syndrome.
- adenoma
- bilateral hyperplasia
- carcinoma

evidence of unequal covariance matrices leads on the whole to a firmer diagnostic view.

Fig. 11.5 shows the diagnostic distributions based on (11.14) for six new cases listed in table 11.2. We shall return to a discussion of these cases in §11.4.

11.4 Monitoring for atypicality

In discussing the aspect of assumption D1 concerning the exhaustiveness of the set $T$ of possible types we indicated that some form of monitoring is desirable to ensure that the decision to stream a new case into the speciality characterised by $T$ has not been unreasonable. We have already met the basic
Fig. 11.5 Diagnostic assessments on the basis of unequal covariance matrices of 6 new cases of Cushing's syndrome.

Table 11.2 Urinary excretion rates (mg/24h) of two steroid metabolites for 6 undiagnosed patients with Cushing's syndrome

<table>
<thead>
<tr>
<th>Case no.</th>
<th>Tetrahydrocortisone</th>
<th>Pregnantriol</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1</td>
<td>5.1</td>
<td>0.4</td>
</tr>
<tr>
<td>u2</td>
<td>12.9</td>
<td>5.0</td>
</tr>
<tr>
<td>u3</td>
<td>13.0</td>
<td>0.8</td>
</tr>
<tr>
<td>u4</td>
<td>2.6</td>
<td>0.1</td>
</tr>
<tr>
<td>u5</td>
<td>30.0</td>
<td>0.1</td>
</tr>
<tr>
<td>u6</td>
<td>20.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

idea of monitoring in the discussion of the notion of 'past experience' in §4.3. A method of achieving a form of monitoring takes the following form. On the basis of past experience z of a particular type u the probability distribution associated with the feature vector y of any new patient is simply the predictive distribution \( p(y|u, z) \). Suppose that we are concerned about how typical of disease category u a patient with observed feature vector \( y_0 \) is? We can regard any case with feature vector y for which \( p(y|u, z) \geq p(y_0|u, z) \) as more typical.
of \( u \) than patient \( y_0 \), or equivalently \( y_0 \) as less typical than \( y \). We can thus construct for patient \( y_0 \) a sensible index \( J_u(y_0) \) of atypicality relative to disease category \( u \) as the probability (on the basis of past experience) that another patient is more typical than him. Thus

\[
J_u(y_0) = \int \left\{ y : p(y|u, z) > p(y_0|u, z) \right\} p(y|u, z) \, dy. \tag{11.18}
\]

Thus \( J_u \) is measured on the scale \((0, 1)\) with 0 indicating the absolutely typical and 1 complete atypicality. If we find

\[
\min_{u \in T} J_u(y_0) \tag{11.19}
\]

near 1 then we would be right to suspect that the patient may have been channelled into the wrong set \( T \) of types, and take some appropriate action.

In a multivariate normal setting the predictive distribution takes a Student form such as (11.14) or (11.16), say

\[
p(y|u, z) = \text{Std}(k, b, c).
\]

Then following a mathematical development similar to that of §4.3 we arrive at a simple expression for the atypicality index for type \( u \) of a patient with feature vector \( y \):

\[
J_u(y) = I_{q(y) < I_{q(y_0)} + k_1 \{ \frac{1}{2}d, \frac{1}{2}(k - d + 1) \}}. \tag{11.20}
\]

where

\[
q(y) = (y - b)'c^{-1}(y - b). \tag{11.21}
\]

For the 21 past cases of Cushing's syndrome the most atypical for each of the three types are as follows:

(a) Case \( a_4 \) with atypicality \( J_{a} = 0.56 \),
(b) Case \( b_3 \) with atypicality \( J_{b} = 0.75 \),
(c) Case \( c_4 \) with atypicality \( J_{c} = 0.52 \).

Table 11.3 shows the atypicality indices \( J_{a}, J_{b}, \) and \( J_{c} \) for the six new cases of Cushing's syndrome presented in table 11.2, together with the confirmed histopathological type after operation, where appropriate. For cases \( u_1 \) to \( u_4 \) these histopathological types are all in agreement with the predictive diagnostic assessment of fig. 11.5. For case \( u_5 \) all three atypicality indices are large and we would be hesitant to regard the diagnostic distribution as expressing an appropriate view for this case. Rather we would draw attention to the possibility that the case has been wrongly referred to the set \( T \) of types. In fact the data for \( u_5 \) arose from an irregularity in collection of the urine specimen, and is
Diagnosis

Table 11.3 Atypicality indices for the six new cases of Cushing's syndrome and the actual type confirmed by histopathology

<table>
<thead>
<tr>
<th>Case no.</th>
<th>I_a</th>
<th>I_b</th>
<th>I_c</th>
<th>Confirmed type</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1</td>
<td>0.60</td>
<td>0.25</td>
<td>0.975</td>
<td>b</td>
</tr>
<tr>
<td>u2</td>
<td>0.95</td>
<td>0.84</td>
<td>0.02</td>
<td>c</td>
</tr>
<tr>
<td>u3</td>
<td>0.95</td>
<td>0.80</td>
<td>0.91</td>
<td>b</td>
</tr>
<tr>
<td>u4</td>
<td>0.21</td>
<td>0.86</td>
<td>0.993</td>
<td>a</td>
</tr>
<tr>
<td>u5</td>
<td>0.991</td>
<td>0.9999</td>
<td>0.984</td>
<td>Irregular urine collection†</td>
</tr>
<tr>
<td>u6</td>
<td>0.981</td>
<td>0.978</td>
<td>0.89</td>
<td>d†</td>
</tr>
</tbody>
</table>

† See text.

presented here to show that atypicality indices can serve as a useful safeguard, particularly in higher-dimensional problems where the simplicity of fig. 11.1 is not available. Case u6 is actually of a type d of Cushing's syndrome not included in the set T. It is presented here as a sharp reminder that while monitoring for atypicality is a necessary discipline of any sensible diagnostic system it is not sufficient to guarantee that a new case outside T will not be differentially diagnosed within T in an apparently satisfactory way. For a full diagnostic analysis of Cushing's syndrome it is necessary to include type d in T and to extend the dimensionality of the feature vector from 2 to 15.

For any new case, in addition to reporting on atypicality we can also state whether or not it is outside previous experience of any particular type, in the sense of §4.3. In terms of table 11.3, for example, a new case is outside previous experience of type b if its atypicality is greater than the case of greatest atypicality in the basic set. Thus while case u3 is being diagnosed correctly as of type b it is outside previous experience of type b. This is not surprising since our previous experience of type b is small, being confined to ten case records. Indeed we can indicate how likely we are to see a new case of type b which falls outside this previous experience because this is simply assessed by

\[ 1 - \max \{ I_b(x_i) : i = 1, \ldots, n_b \}, \]  

(11.22)

where \( x_1, \ldots, x_{n_b} \) are the feature vectors of the \( n_b \) cases of b in the basic set. For type b there is thus a probability of 0.25 of obtaining a new case outside previous experience. For types a and c the corresponding probabilities are 0.44 and 0.48.

11.5 Estimative and predictive diagnosis

A few simple assumptions have led us inevitably to a particular form of diagnostic assessment, in terms of the predictive diagnostic distribution. The statistical diagnostic methods currently widely advocated are of estimative
type and are generally supported by an argument of the following kind. If we knew the structural parameter \( \theta \) then we could easily arrive at appropriate plausibilities for the unknown disease category \( u \) of a new patient with feature vector \( y \) from prior plausibilities \( p(u) \), or possibly \( p(u|t) \). We would simply apply Bayes's theorem in the form

\[
p(u|y) \propto p(u)p(y|u, \theta).
\] (11.23)

Recognising that we do not know \( \theta \) the estimative method replaces \( \theta \) by a suitable estimate \( \hat{\theta}(z) \), for example a maximum likelihood estimate, based on the past records \( z \). We could thus rewrite the estimative method as:

\[
p(u|y, z) \propto p(u) p(y|u, \hat{\theta}(z)).
\] (11.24)

We recall the form (11.11) of the predictive method,

\[
p(u|y, z) \propto p(u) p(y|u, z)
\] (11.25)

where

\[
p(y|u, z) = \int_\theta p(y|u, \theta) p(\theta|z) d\theta.
\] (11.26)

Clearly (11.24) and (11.25) will be in good agreement if \( p(\theta|z) \) is highly concentrated at \( \hat{\theta}(z) \). While, by the usual large sample arguments, this will be the case when there is a substantial past experience, there are many areas of medicine where past experience is modest, and these are situations where data interpretation is at a premium. One way of interpreting the fallacy of the estimative method is that it takes no account of the sampling variability of the estimator \( \hat{\theta}(z) \) of \( \theta \). Whereas, through (11.26), the predictive method weights the possible distributions \( p(y|u, \theta) \) according to the plausibilities of the various \( \theta \).

A common dissatisfaction with the estimative method is that far too optimistic a picture is painted when the method is assessed on the basic set of patients and all estimates of the 'misclassification rates' for future patients are far too low. While this is to some extent attributable to such factors as exclusion of difficult cases from the set \( z \) of past records, undoubtedly one contributory factor is the use of the estimative rather than the predictive approach. The predictive method has a tendency to damp down the over-optimism to a realistic degree. That there must be this tendency can be seen from very simple considerations. If \( r = 2, d = 1 \) and the feature distributions for the two categories are normal then the difference between (11.24) and (11.25) is that, whereas the distribution of \( y \) in (11.24) is being treated as normal, that of (11.25) is the corresponding Student distribution; see fig. 11.6. The alteration to the prior odds for the new patient by the estimative method is by the factor \( A_P/B_P = 8.0 \), whereas by the predictive method the factor is \( A'_P/B'_P = 2.5 \).
Fig. 11.6 Typical difference between estimative odds $AP/BP$ and predictive odds $A'P'/B'P'$. 
Fig. 11.7 Comparison between the estimative and predictive diagnostic assessments for the 6 new cases of Cushing's syndrome.

- o estimative
- * predictive

For example 11.1 we illustrate this difference between the estimative and predictive diagnostic methods very simply. In fig. 11.7, for each of the six new cases of table 11.2, the estimative and predictive plausibility points are shown joined by a directed line which leads from the estimative assessment to the corresponding predictive point. It is clear that the estimative method tends on the whole to give the appearance of a firmer diagnostic view than the predictive method. That this tendency can be completely misleading in a practical problem will be seen in the application in the next section.

We record here for reference purposes the estimative counterpart of (11.20). For normally distributed feature vectors the estimative counterpart of (11.14) is

\[ p(y|u, \hat{\theta}(z)) = N_d(\bar{x}_u, S_u^{-1}) \]  

(11.27)
and the estimative index of atypicality for type \( u \) of patient with feature vector \( y \) can be shown to be
\[
\hat{J}_u(y) = J_{Q(y)}(1|d),
\]
where
\[
Q(y) = (y - \bar{x}_u)'S_u^{-1}(y - \bar{x}_u). \tag{11.29}
\]
If the assumption of equal covariance matrices for the different types is made then \( S_u \) in (11.27) to (11.29) is simply replaced by \( S \), as defined in (11.15).

### 11.6 An application to differential diagnosis of Conn's syndrome

That the distinction we have made between estimative and predictive diagnosis is no mere hair-splitting is well illustrated by the application of these techniques to the differential diagnostic problem of Conn's syndrome posed as example 1.7. We shall not set out in detail the computations which are a straightforward application of the techniques of the preceding sections, but simply highlight the special aspects of this particular problem and report the widely differing results of the two techniques.

First we observe that there is appreciable skewness in some of the data; see in particular the aldosterone results for the adenoma patients. Some transformation is advisable to satisfy the normality assumption. In what follows we have used a blanket logarithmic transformation for each of the features; this appears to give reasonably symmetric distributions and so makes the normality assumption at least feasible. Secondly an assumption of equality of covariance matrices appears a doubtful proposition; for example, the aldosterone results for the bilateral hyperplasia patients are tightly packed compared with the wide spread of aldosterone results for the adenoma patients. We therefore work on the basis of different covariance matrices for the two types.

With these assumptions the structural parameter \( \psi \) consists of the two eight-dimensional mean vectors and the two covariance matrices of order 8, and so is an 88-dimensional parameter. The estimative method thus attempts to estimate this 88-dimensional parameter from 31 eight-dimensional vectors and to use each of the 88 component estimates as if it was the true value. Clearly such a method, by ignoring the obvious unreliability of the estimate, must run a risk of wild assessment. A first indication of the remarkable extent of this risk is seen in fig. 11.8 where we apply both the estimative and the predictive methods to the 31 patients in the basic set of table 1.6. Since there are just two types it is simplest to show the results in terms of the odds in favour of adenoma. We show these on the basis that the 31 case records form a natural, as opposed to a designed, informative experiment so that, by (11.17) we take
\[
p(u = 1|t) = \frac{\bar{p}}{1-\bar{p}}, \quad p(u = 2|t) = \frac{\bar{p}}{1-\bar{p}}. \tag{11.30}
\]
Fig. 11.8 Comparison of estimative and predictive odds on adenoma, shown on a $\log_{10}$ scale, for the 31 past cases of Conn's syndrome.
Table 11.4 Feature vectors of four new cases of Conn's syndrome

<table>
<thead>
<tr>
<th>Case no.</th>
<th>Age (Years)</th>
<th>Na (meq/l)</th>
<th>K (meq/l)</th>
<th>CO₂ (meq/l)</th>
<th>Renin (meq/l)</th>
<th>Aldosterone (meq/l)</th>
<th>Blood Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Systolic</td>
</tr>
<tr>
<td>u1</td>
<td>50</td>
<td>143.3</td>
<td>3.2</td>
<td>27.0</td>
<td>8.5</td>
<td>51.0</td>
<td>210</td>
</tr>
<tr>
<td>u2</td>
<td>49</td>
<td>142.6</td>
<td>2.3</td>
<td>36.0</td>
<td>6.2</td>
<td>35.7</td>
<td>192</td>
</tr>
<tr>
<td>u3</td>
<td>44</td>
<td>143.1</td>
<td>4.0</td>
<td>26.8</td>
<td>5.5</td>
<td>17.7</td>
<td>170</td>
</tr>
<tr>
<td>u4</td>
<td>53</td>
<td>142.8</td>
<td>4.0</td>
<td>26.3</td>
<td>3.6</td>
<td>65.8</td>
<td>260</td>
</tr>
</tbody>
</table>

Table 11.5 Comparison of estimative and predictive odds and atypicality indices for four new cases of Conn's syndrome

<table>
<thead>
<tr>
<th>Patient no.</th>
<th>Odds a/b</th>
<th>Atypicality indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimative</td>
<td>Predictive</td>
</tr>
<tr>
<td>u1</td>
<td>10^12/1</td>
<td>64/1</td>
</tr>
<tr>
<td>u2</td>
<td>10^10/1</td>
<td>56/1</td>
</tr>
<tr>
<td>u3</td>
<td>1/245</td>
<td>1/80</td>
</tr>
<tr>
<td>u4</td>
<td>3700/1</td>
<td>1/2</td>
</tr>
</tbody>
</table>

(For a clinic anticipating a different incidence rate for the two types the appropriate modification of odds could be easily applied.) Because of the large proportional differences in the odds for the different cases the odds have been plotted on a logarithmic scale, \( \log_{10}(p(u = 1|y, z)/p(u = 2|y, z)) \) being calculated by means of (11.11), (11.14) and (11.30).

From fig. 11.8 we see that the exaggerated confidence of assessments of odds of 10^20 to 1 by the estimative method can be slashed to 10^3 to 1 by the predictive assessment. On the whole the estimative method gives much more extreme odds than the predictive method. It could be argued that we are re-applying the statistical diagnostic system to cases whose type we already know so that we might expect large odds to be assessed, so that perhaps the estimative method is the more realistic. But the differences with respect to new patients can be equally startling. Table 11.4 shows the feature vectors of four typical new patients known to have Conn's syndrome but of unknown type, case u1 being reproduced from table 1.6. The dramatic alterations of odds in some of these cases (table 11.5) is due to the fact that they are, in the technical sense of §4.3, outside the previous limited experience of both types, though not in any way atypical of the favoured disease. For such cases the predictive method acts with extra caution which is again very reasonable. Case u4 is particularly interesting; it is in fact a now-confirmed case of bilateral hyperplasia, not the clear case of adenoma assessed by the estimative method.

To sum up, the estimative method can give an exaggerated view of the evidence in favour of a particular type whereas the predictive method almost invariably moderates this view. This moderation is greatest when there is a limited past experience and in interpreting this we must remember that it has...
Diagnosis

to be measured not in terms of the number of cases but in terms of the number of cases relative to the dimension of the feature vector. Eleven cases may be adequate experience if we are dealing with a one-dimensional feature, but they constitute a very limited experience for an eight-dimensional feature. We can see this in terms of the probabilities of obtaining a new patient within the previous experience of the two types. For adenoma this is 0.45, and for bilateral hyperplasia it is only 0.08.

11.7 Diagnosis under a utility structure

If diagnosis is to be regarded as a decision problem then the decision theory model must have the following components.

Parameter space. The set of unknown parameters can be identified with the index set $T$ for the class $F$ of possible feature determining experiments, the unknown type $u$ of the case under consideration playing the role of the unknown state of nature.

Action set. An action $a$ is simply a decision to act as if the type of the case were $a$, so that the action set $A$ is the set $T$ of possible types.

Utility function. To complete the specification of diagnosis as a decision problem we must therefore be in a position to attach a utility $U(a, u)$ to each possible action $a$ in the face of each possible type $u$.

The decision problem is then to maximise the expected utility of having observed the past case records $z$ and the feature vector $y$ of the new case under consideration and of having taken action $a$. We have that

$$U(a, y) = \sum_{u \in T} U(a, u)p(u|y, z),$$

(11.31)

where $p(u|y, z)$ is the diagnostic distribution and where we have dropped from the notation the obvious dependence on $z$. Thus if the diagnostic distribution has been evaluated there is no difficulty in evaluating $U(a, y)$ for each of the $r$ possible actions $a$ and in choosing the maximising action.

In medical diagnostic problems the difficulty lies in obtaining a realistic specification of the utility structure. Using utilities can be envisaged only in terms of the outcome of treatment allocated, with the almost overwhelming problem of expressing, for example, the advantage to a patient of some improvement in his condition in terms of the same utility units as the cost of treatment, not only in terms of monetary expense but also in terms of discomfort. Indeed only if there is a unique treatment associated with each type will it be sensible...
to formulate the diagnostic problem in terms of (11.31). Otherwise the problem is part of a more extensive sequential process and should be treated as such.

A popular device to by-pass this difficulty is to concentrate on minimising the overall misclassification or mistyping rate. In terms of a decision problem this is equivalent to adopting a utility structure

\[
U(a, u) = \begin{cases} 
1 & \text{if } a = u, \\
0 & \text{if } a \neq u,
\end{cases}
\]

that is, placing zero utility on a mistyping of whatever kind, and unit utility to each correct typing. With this utility structure (11.31) becomes

\[
U(a, y) = p(a|y, z),
\]

so that we obtain the very simple rule: for a new patient with feature vector take as decisive type the mode of the diagnostic distribution. While this simple rule has a strong intuitive appeal there is of course no guarantee that the real utility structure of the problem is such as to make this optimum.

History

The linear discriminant method of Fisher (1936) was the first statistical technique to be devised for the diagnostic type of problem considered in this chapter. A good account of this and other estimative methods is given in Anderson (1958), who incidentally presents though does not develop a method equivalent to the predictive method through a generalised likelihood ratio test argument. The predictive method is first presented explicitly in Geisser (1964) for the multivariate normal case, and in Dunsmore (1966). The great practical differences that there can be between the estimative and predictive methods appear to have been first noted by Aitchison and Kay (1974), who also consider the uses of indices of atypicality.

Problems

11.1 A diagnostic assessment between two psychiatric types, 1 and 2, of subject is to be made on the basis of a single binomial test on a subject who either responds or does not respond to the test stimulus. Of \( n \) naturally arising past subjects, firmly typed by other more elaborate means, \( n_1 \) and \( n_2 \) were of types 1 and 2. Suppose that the psychiatrist is persuaded to express his prior knowledge of the two response probabilities \( \theta_1 \) and \( \theta_2 \) as independent \( \text{Be}(g_1, h_1) \), \( \text{Be}(g_2, h_2) \) distributions. When the test was applied to the \( n \) past subjects \( x_i \) of the \( n_i \) of type \( i \) (\( i = 1, 2 \)) responded.

A new subject has just undergone the test and responded. How do you assess the plausibility of his being of type 1? Can you define indices of atypicality for this diagnostic problem?
11.2 Past experience for a certain type consists of two observations \( x_1 \) and \( x_2 \) on a one-dimensional normally distributed feature. Show that the probability that a new case of this type has an observed feature falling between \( x_1 \) and \( x_2 \) is \( 2\Phi(1/\sqrt{2}) - 1 = 0.52 \) on the estimative and \( I_{1/4}(1/4) = \frac{1}{2} \) on the predictive assessment. Which do you regard as the more reasonable assessment, and why?

11.3 Construct a diagnostic system for Conn’s syndrome (example 1.7) on the basis of the single feature \( K \) of table 1.6.

11.4 The ‘disintegration times’ of two types 1 and 2 of ‘elementary particles’ are assumed to be \( \text{Ex}(\theta_1) \) and \( \text{Ex}(\theta_2) \) with unknown indices \( \theta_1, \theta_2 \). In an experiment in which six naturally occurring particles were observed the following case records (particle type, disintegration time in ms) were recorded:

\[(1, 47), (2, 75), (1, 17), (1, 32), (2, 31), (1, 19).\]

In two further independent experiments the disintegration times of two recorded particles of unknown type were 40 and 70 ms. Obtain the diagnostic distributions and atypicality indices for these two particles, both on a predictive and on an estimative basis.

How do you assess the probability that a new type 2 particle is outside previous experience?

11.5 For a problem of differential diagnosis of three types on the basis of two features the case records of 18 pathologically diagnosed patients are shown in the table below. Construct and compare estimative and predictive diagnostic systems on the basis of this past experience. Apply both systems to the new patients listed below.

<table>
<thead>
<tr>
<th>Past case records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case no.</td>
</tr>
<tr>
<td>a1</td>
</tr>
<tr>
<td>a2</td>
</tr>
<tr>
<td>a3</td>
</tr>
<tr>
<td>a4</td>
</tr>
<tr>
<td>a5</td>
</tr>
<tr>
<td>a6</td>
</tr>
<tr>
<td>a7</td>
</tr>
<tr>
<td>a8</td>
</tr>
<tr>
<td>a9</td>
</tr>
<tr>
<td>a10</td>
</tr>
<tr>
<td>a11</td>
</tr>
<tr>
<td>b1</td>
</tr>
<tr>
<td>b2</td>
</tr>
<tr>
<td>b3</td>
</tr>
<tr>
<td>b4</td>
</tr>
<tr>
<td>b5</td>
</tr>
<tr>
<td>b6</td>
</tr>
<tr>
<td>b7</td>
</tr>
</tbody>
</table>
11.6 Construct a diagnostic system for Conn’s syndrome (example 1.7) based on the three plasma concentrations of the electrolytes Na, K and CO₂ as given in table 1.6.

11.7 Complete the analysis of problem 1.6.
12

Treatment allocation

12.1 The nature of a treatment allocation problem

When a treatment is applied to an object or individual it is with the express purpose of altering the future of that object or individual. Thus when we choose one of a number of possible refining processes for a batch of raw material we intend that the batch will in the future attain some desirable quality. When we select a method of machining an industrial component we have in mind some future characteristic of the component. When we prescribe a particular treatment for a patient we hope that some specific aspect of his future condition will be more agreeable than his present state of disease. Because of this preoccupation with the future state of an object or individual it will not be surprising to find that statistical prediction analysis has an important role to play in the problem of treatment allocation.

In the examples already mentioned there are three basic sets which must clearly play an important role. First we suppose that the present state or indicator $t$ of the individual unit under consideration belongs to some specifiable set $T$ of possible initial states or indicators. Secondly, there is some set $A$ of possible treatments from which we have to select a treatment $a$ to apply to the individual unit. Thirdly we must to some extent assess the effectiveness of treatment in terms of the future state or response $y$ attained by the unit after application of treatment; we thus have to be in a position to envisage the set $Y$ of possible future states or responses. In order to see the exact roles played by these sets and to obtain a clearer insight into the nature of treatment allocation problems we now consider specific examples.

Example 12.1

Quality improving process. An attempt is to be made to rationalise the method of allocating treatments to batches of raw materials of differing initial quality $t$ to obtain a final quality $y$. There are three possible treatments 1, 2 and 3 and information on the effectiveness of these treatments has been sought from 30 experimental runs of the process. The results of these runs are shown in table 12.1. If the selling price of a batch of quality $y$ is $g(y)$ and the cost of treatment $a$ is $c_a$ per batch ($a = 1, 2, 3$) what treatment allocation policy should be adopted?
### Table 12.1 Treatments, initial and final qualities in 30 experimental runs of a quality improving process

<table>
<thead>
<tr>
<th>Treatment</th>
<th>(Initial quality, final quality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(30.9, 44.2), (35.8, 48.6), (28.2, 44.3), (40.5, 50.0), (23.5, 43.0), (47.4, 52.5), (51.2, 55.0), (43.0, 51.8), (37.7, 49.6), (33.8, 46.1)</td>
</tr>
<tr>
<td>2</td>
<td>(33.3, 46.1), (31.3, 46.7), (23.9, 42.7), (42.2, 50.0), (27.4, 45.0), (50.3, 51.0), (35.8, 47.3), (45.7, 51.0), (39.8, 49.1), (37.6, 47.7)</td>
</tr>
<tr>
<td>3</td>
<td>(39.8, 46.3), (31.3, 38.6), (41.0, 50.1), (51.2, 57.3), (36.4, 43.1), (45.7, 56.8), (26.0, 37.8), (37.1, 47.3), (43.0, 52.4), (35.2, 45.0)</td>
</tr>
</tbody>
</table>

We shall consider specific forms for \( g(y) \) and the \( c_a \) later in this chapter. For the moment we are concerned solely with the nature of the treatment allocation problem. In our study of regulation, optimisation, calibration and diagnosis we have already seen that it is necessary to deal with a whole class \( F \) of possible future experiments indexed by the elements of \( T \). This feature persists in the present problem since clearly final quality \( y \) of a batch can depend appreciably on its initial quality \( t \). There is, however, the additional complication here that final quality may also depend on the treatment \( a \) allocated to the batch. Let \( f_{at} \) denote the experiment which records the final quality \( y \) of a batch of initial quality \( t \) subjected to treatment \( a \). Then in analysing this treatment allocation problem we have to envisage the class

\[
F = \{ f_{at} : a \in A, t \in T \} \quad (12.1)
\]

of future experiments, where \( A = \{1, 2, 3\} \) is the set of possible treatments and \( T \), the real line say, is the set of possible initial qualities. Then the data of table 12.1 can be expressed in the form

\[
z = \{(a_i, t_i, x_i) : i = 1, \ldots, 30\},
\]

where the triplet \((a, t, x)\) typically denotes the treatment, initial quality and final quality of a batch. The informative experiment \( e \) thus takes the form

\[
e = \{ f_{a_1 t_1}, \ldots, f_{a_{30} t_{30}} \}.
\]

Our treatment allocation problem can then be framed in the following terms. A new batch of initial quality \( t \) awaits treatment. On the basis of the information contained in \( e \) which of the three future experiments \( f_{1t}, f_{2t}, f_{3t} \) is it best to conduct? The term 'best' has clearly to be interpreted here in terms of some kind of utility structure, constructed from the information about selling price and costs. Since the advantage of increasing quality from \( t \) to \( y \) is \( g(y) - g(t) \), and since the cost of treatment \( a \) is \( c_a \), we can associate with each triplet \((a, t, y)\) the utility \( U(a, t, y) \) of transforming a batch of initial
quality \( t \) to final quality \( y \) by way of treatment \( a \). Then

\[
U(a, t, y) = g(y) - g(t) - c_a \quad (a \in A, t \in T, y \in Y).
\] (12.2)

**Example 12.2**

*Treatment of a skin allergy.* In a clinical trial to compare the effectiveness of two barrier creams in the prevention of recurrence of a certain skin allergy, 100 out of 200 previous sufferers were chosen at random and allocated to cream 1, the remainder being assigned to cream 2. Table 12.2 summarises the results of this trial for male and female sufferers separately.

<table>
<thead>
<tr>
<th>Cream</th>
<th>Response</th>
<th>No recurrence</th>
<th>Some recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28M, 22F</td>
<td>21M, 29F</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18M, 32F</td>
<td>35M, 15F</td>
<td></td>
</tr>
</tbody>
</table>

**Table 12.2 Responses of 200 cases of skin allergy to the two barrier creams**

M denotes male, F denotes female.

In this example there are just two possible treatments, cream 1 and 2, so that we may take \( A = \{1, 2\} \). From the information we have the only possible indicator of appropriate treatment is sex, so that \( T = \{M, F\} \). The response to treatment is measured simply in terms of success (no recurrence) or failure (some recurrence); we then take \( Y = \{0, 1\} \), where 0 denotes a failure, 1 a success. With this specification the future experiment \( f_{1M} \), for instance, records whether or not there is a recurrence of the skin allergy for a male patient using cream 1. The class \( F \) consists of four possible experiments \( f_{1M}, f_{1F}, f_{2M}, f_{2F} \), and the informative experiment \( e \) consists of 200 independent such experiments, 49 of type \( f_{1M} \), 51 of type \( f_{1F} \), 53 of type \( f_{2M} \), and 47 of type \( f_{2F} \). Again the data of table 12.2 can be expressed in the form

\[
z = \{(a_i, t_i, x_i): i = 1, \ldots, 200\}.
\]

The treatment allocation problem, say for a new male sufferer, then consists of deciding whether the future experiment to be performed should be \( f_{1M} \) or \( f_{2M} \).

### 12.2 The prognostic distributions and utility structure

Suppose that for the class \( F \) of experiments as defined in (12.1) we can postulate some parametric family of distributions indexed by a parameter \( \theta \in \Theta \). More precisely, for the future experiment \( f_{at} \) the possible density functions on \( Y \) are

\[
p(y | a, t, \theta) \quad (\theta \in \Theta).
\] (12.3)
Treatment allocation

For instance, in example 12.1 we may consider postulating

\[ p(y | a, t, \theta) = N(\alpha_a + \beta_a t, \tau_a) \quad (a = 1, 2, 3). \] (12.4)

the assumption that for each treatment there is a normal linear regression of final quality \( y \) on initial quality \( t \). This has the form (12.3) with \( \theta = (\alpha_1, \beta_1, \tau_1, \alpha_2, \beta_2, \tau_2, \alpha_3, \beta_3, \tau_3) \). We can then regard the duty of the informative experiment with its data of the form

\[ z = \{ (a_i, t_i, x_i) : i = 1, \ldots, n \} \] (12.5)

to transform some prior assessment \( p(\theta) \) into a posterior assessment \( p(\theta | z) \).

It is then natural to use this assessment to arrive at the predictive distribution for the future experiment \( f_{a_t} \):

\[ p(y | a, t, z) = \int_\Theta p(y | a, t, \theta) p(\theta | z) d\theta. \] (12.6)

Our interest in (12.6) is to assess the effect of treatment \( a \) on an individual unit in present or initial state \( t \), and in order to choose between treatments we need to obtain such a predictive distribution for each possible treatment. In order to emphasise the dependence of these predictive distributions on treatment we term (12.6) the \textit{prognostic distribution} for treatment \( a \) with respect to initial state \( t \).

If there is a well specified utility structure \( U(a, t, y) \), assigning a utility to each possible triplet \( (a, t, y) \), then the treatment allocation problem can be simply resolved. For an individual unit with indicator \( t \) we can evaluate for each possible treatment \( a \) the expected utility

\[ U(a, t) = \int_y U(a, t, y) p(y | a, t, z) dy, \] (12.7)

expectation being taken with respect to the appropriate prognostic distribution. The optimum treatment \( a^*(t) \) corresponding to indicator \( t \) maximises this expectation:

\[ U[a^*(t), t] = \max_{a \in A} U(a, t). \] (12.8)

12.3 Two applications

We are now in a position to apply the decisive theory approach of the preceding section to the treatment allocation problems of §12.1.

Example 12.1 (continued)

The scatter diagram of fig. 12.1 shows that the model suggested in §12.2 is not an unreasonable one, so that we set
The regression-type computations for each treatment are shown in table 12.3. Each of the three prognostic distributions is then easily obtained on the basis of vague prior information by a straightforward application of case 5 of table 2.3 along the lines set out in §2.5 exemplified in §10.3. Note that relevant information on \( (\alpha_a, \beta_a, \tau_a) \) in the data of table 12.1 comes only from the 10 results on treatment \( a \). The prognostic distributions thus take the form

\[
p(y|a, t, z) = \text{No}(\alpha_a + \beta_a t, \tau_a).
\]

with mean \( \hat{\alpha}_a + \hat{\beta}_a t \) and are set out in table 12.4.
Table 12.3 Regression calculations for the three quality-improving treatments

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta_i )</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>( \delta_i )</td>
<td>31.51</td>
<td>35.89</td>
<td>12.52</td>
</tr>
<tr>
<td>( \mu_i )</td>
<td>0.457</td>
<td>0.321</td>
<td>0.904</td>
</tr>
<tr>
<td>( \eta_0 )</td>
<td>37.20</td>
<td>36.73</td>
<td>38.67</td>
</tr>
<tr>
<td>( S_0(t, t) )</td>
<td>666.72</td>
<td>598.48</td>
<td>466.38</td>
</tr>
<tr>
<td>( v_0 )</td>
<td>5.497</td>
<td>3.497</td>
<td>32.633</td>
</tr>
</tbody>
</table>

Table 12.4 Prognostic distributions for quality improvement

| Treatment | Prognostic distribution \( p(y | a, t, z) = S(t, k, c) \) |
|-----------|--------------------------------------------------|
| 1         | \[ 8; 31.5 + 0.457 t; 0.756 + 0.00103 (t - 37.20)^3 \] |
| 2         | \[ 8; 35.9 + 0.321 t; 0.481 + 0.000730 (t - 36.73)^3 \] |
| 3         | \[ 8; 12.5 + 0.904 t; 4.487 + 0.00875 (t - 38.67)^3 \] |

To resolve this treatment allocation problem we require to know the treatment costs \( c_a (a = 1, 2, 3) \) and the form of \( g(y) \) in (12.2). We examine the problem for two cases.

(i) \( c_1 = 4, \ c_2 = 5, \ c_3 = 3 \) and \( g(y) = y \), so that the selling price is directly proportional to the quality. Then

\[
U(a, t) = \int_{-\infty}^{\infty} y p(y | a, t, z) dy - t - c_a
\]

\[
= \hat{\alpha}_a + \hat{\beta}_a t - t - c_a
\]

\[
= \hat{\alpha}_a - c_a (1 - \hat{\beta}_a) t.
\]

(12.9)

Fig. 12.2 shows the graphs of \( U(a, t) \) plotted against \( t \) for each of the three treatments. The optimum treatment allocation rule must then clearly take the form: the optimum treatment is

\[
\begin{align*}
& 2 \quad (t < 24.8), \\
& 1 \quad (24.8 < t < 40.2), \\
& 3 \quad (t > 40.2).
\end{align*}
\]

(ii) \( c_1 = 0.5, \ c_2 = 0.3, \ c_3 = 0.6 \), with

\[
g(y) = \begin{cases} 
2 & (y \geq 48), \\
1 & (y < 48), 
\end{cases}
\]

so that the selling price depends only on whether the quality reaches a standard, 48, or not. Clearly for a batch with \( t \geq 48 \) the standard has already been attained.
Treatment allocation

Fig. 12.2 Graphs of $U(a, t)$ plotted against $t$ for each of the three treatments: case (i).

in the raw state and there is thus no sense in treating the batch. We therefore now allow the possibility of a fourth 'treatment' $a = 0$, corresponding to the action of no treatment. Since treatment 0 is clearly optimum for $t \geq 48$ we can restrict comparison of the four treatments to the interval $t < 48$. Then for $a = 1, 2, 3$,

$$U(a, t) = 2 \int_{-\infty}^{48} p(y | a, t, z) dy + \int_{-\infty}^{48} p(y | a, t, z) dy - g(t) - c_a$$

$$= -c_a + \int_{48}^{\infty} p(y | a, t, z) dy \quad (t < 48), \quad (12.10)$$
and

\[ U(0, t) = 0 \quad (t < 48). \quad (12.11) \]

The integral in (12.10) can be readily evaluated in terms of the incomplete beta function by (A26) of appendix I (and its simple extension for the case \( a < b \)). Again the simplest presentation of the solution is in graphical form. In fig. 12.3 the four graphs of \( U(a, t) \) against \( t \) are shown and from this the following optimum solution emerges: the optimum treatment is

- \( 0 \) (\( t \leq 36.1 \)),
- \( 1 \) (\( 36.1 < t < 39.1 \)),
- \( 2 \) (\( 39.1 < t < 48 \)),
- \( 0 \) (\( 48 \leq t \)).
Example 12.2 (continued)

If the two creams have the same cost, which can then be set equal to 0, and if the utility structure simply records 1 for a success and 0 for a failure the resolution of the treatment allocation problem is almost trivial. For

$$U(a, t, y) = \begin{cases} 1 & (y = 1) \\ 0 & (y = 0) \end{cases}$$

for each cream \(a\) and for each sex \(t\), and so

$$U(a, t) = p(y = 1 | a, t, z).$$

If we use vague priors independently on the four success probabilities \(\theta_{at}\) then, from Table 2.3, the predictive density functions (and hence expected utilities) are obtained simply as the observed success rates in the four \((a, t)\) categories. These are given in Table 12.5, and the treatment allocation rule can be expressed simply as: for males use cream 1, for females use cream 2.

### Table 12.5 Expected utilities for skin allergy problem

| Sex | Cream | \(p(y = 1 | a, t, z)\) |
|-----|-------|----------------------|
| M   | 1     | \(\mathbb{1}\)       |
| M   | 2     | \(\mathbb{1}\)       |
| F   | 1     | \(\mathbb{1}\)       |
| F   | 2     | \(\mathbb{1}\)       |

12.4 Treatment allocation to meet a required specification

One form of problem, considered by Guttman and Tiao (1964) under the heading of 'selecting a best population', is trivially resolved in terms of the treatment allocation theory already developed. Instead of formulating their problem in terms of allocating treatments Guttman and Tiao regard the elements of \(A\) as different populations with the associated problem of selecting which of the possible populations is best for a particular utility specification. Moreover in their formulation no initial state \(t\) of a population is envisaged but we shall retain this aspect in our presentation here. The main feature in such problems is the simple utility structure which records success and utility 1 for a treated individual unit whose characteristic \(y\) or state after treatment falls in a prescribed subset \(S\) of \(Y\), and which records failure and utility 0 otherwise. Thus for the case where treatment costs are equal we may set

$$U(a, t, y) = \begin{cases} 1 & (y \in S) \\ 0 & (y \notin S) \end{cases}$$
Treatment allocation

whatever the initial state $t$ of the unit and whatever the treatment $a$ applied to the unit. For example the units may be electrical components which will operate satisfactorily within a certain amperage range, but fail to function or burn out otherwise.

With this utility structure

$$U(a, t) = \int_S p(y|a, t, z)dy$$

so that, in terms of the terminology of §4.2, the optimum treatment is that which results in $S$ providing maximum Bayesian cover as computed from the predictive density function. Indeed the analysis here can be regarded as a special case of (12.2) where no treatment is undertaken if $t \in S$ and where

$$g(y) = \begin{cases} 1 & (y \in S), \\ 0 & (y \notin S), \end{cases}$$

and $c_a = 0$ for every $a$. Thus no new technical problems arise.

History

While treatment allocation problems are an obvious field of application for prediction analysis Guttman and Tiao (1964) appear to have been the first authors to recognise this explicitly. Aitchison (1970) considers a wider range of problems and also investigates the feasibility of attempting to estimate the undeclared utility structure of a treatment allocator from a sequence of treatment allocations made by the allocator.

Problems

12.1 Reconsider example 12.1 with $c_1 = 7, c_2 = 5, c_3 = 2$ and

$$g(y) = 25 - (y - 50)^2.$$  

12.2 Reconsider example 12.2 with $c_1 = c_2 = c_3 = 0.2$ and

$$g(y) = \begin{cases} 1 & (44 \leq y \leq 50), \\ 0 & \text{otherwise}, \end{cases}$$

for the cases where

(i) one of the treatments 1, 2, 3 must be used,

(ii) the possibility of no treatment is allowed.

12.3 A suggested model to explain the variability of the responses of individual units in different initial states and subjected to different treatments is as follows.
For any specified treatment there is a minimum response which is an unknown linear function of initial state and the magnitude of the response in excess of this minimum has an exponential distribution. The table below shows the responses of 75 individual units randomly allocated to the three treatments. The three treatments have equal costs and treatment of a unit is successful if and only if the response is in the interval (12, 25). What treatment allocation policy would you advise?

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Response to treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>3.3</td>
</tr>
<tr>
<td>1</td>
<td>5.6</td>
</tr>
<tr>
<td>2</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
</tr>
<tr>
<td>5</td>
<td>3.7</td>
</tr>
<tr>
<td>6</td>
<td>5.7</td>
</tr>
<tr>
<td>7</td>
<td>12.0</td>
</tr>
<tr>
<td>8</td>
<td>12.9</td>
</tr>
<tr>
<td>9</td>
<td>10.3</td>
</tr>
<tr>
<td>10</td>
<td>11.1</td>
</tr>
<tr>
<td>11</td>
<td>13.2</td>
</tr>
<tr>
<td>12</td>
<td>13.0</td>
</tr>
<tr>
<td>13</td>
<td>14.8</td>
</tr>
<tr>
<td>14</td>
<td>16.9</td>
</tr>
<tr>
<td>15</td>
<td>16.9</td>
</tr>
<tr>
<td>16</td>
<td>21.6</td>
</tr>
<tr>
<td>17</td>
<td>19.0</td>
</tr>
<tr>
<td>18</td>
<td>21.1</td>
</tr>
<tr>
<td>19</td>
<td>23.5</td>
</tr>
<tr>
<td>20</td>
<td>24.4</td>
</tr>
<tr>
<td>21</td>
<td>22.1</td>
</tr>
<tr>
<td>22</td>
<td>28.9</td>
</tr>
<tr>
<td>23</td>
<td>24.4</td>
</tr>
<tr>
<td>24</td>
<td>27.6</td>
</tr>
<tr>
<td>25</td>
<td>31.7</td>
</tr>
</tbody>
</table>
**APPENDIX I**

**Notation for standard distributions**

The standard distributions are listed in increasing order of dimensionality and in alphabetical order for a given dimensionality. For each distribution the effective sample space or domain of non-zero probability is specified and any restrictions on the indexing parameters are stated.

The following notation is used in the table.

- $\mathbb{R}^d$: $d$-dimensional real space
- $\mathbb{S}^d$: the space of positive definite symmetric matrices, defined as the subspace $\left( \mathbb{R}^d(\mathbb{S}^d) \right)^2$ of points $(w_{11}, w_{12}, \ldots, w_{2d}, w_{33}, \ldots, w_{3d}, \ldots, w_{dd})$ for which the symmetric matrix $w = [w_{ij}]$ is positive definite.
- $1$: $d$-dimensional vector of units.
- $I_d$: $d$-dimensional identity matrix.
- $\Gamma$ and $\beta$ denote the gamma and beta functions as usually defined and related by $\beta(g,h) = \frac{\Gamma(g)\Gamma(h)}{\Gamma(g+h)} (g > 0, h > 0)$. (A1)

$\Gamma_d$ and $\beta_d$ are the Siegel (1935) generalisations of the gamma and beta functions defined by $\Gamma_d(g) = \pi^{d(d-1)/4} \Gamma(g) \prod_{j=1}^{d-1} \Gamma(g - \frac{1}{2}(d-1)) (g > 0)$. (A2)

$\beta_d(g,h) = \frac{\beta_d(g)\Gamma_d(h)}{\Gamma_d(g+h)} (g > 0, h > 0)$. (A3)

Two other generalised notations are used, the multinormal coefficient $\binom{n}{u} = \frac{n!}{u_1! \ldots u_d! (n - \Sigma u_i)!} (n \text{ positive integer}; u_1, \ldots, u_d \text{ non-negative integers such that } \Sigma u_i \leq n)$ (A4)

and the Dirichlet function $D(g,h) = \frac{\Gamma(g_1) \ldots \Gamma(g_d)\Gamma(h)}{\Gamma(\Sigma g_i + h)} (g_i > 0, i = 1, \ldots, d; h > 0)$. (A5)
Distributions within $\mathbb{R}^1$: random variable $u$; parameter restrictions $b \in \mathbb{R}, c > 0, g > 0, h > 0, k > 0, 0 < l < 1, n$ positive integer

<table>
<thead>
<tr>
<th>Notation</th>
<th>Name</th>
<th>Domain restrictions</th>
<th>Density function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Be}(g, h)$</td>
<td>Beta</td>
<td>$0 &lt; u &lt; 1$</td>
<td>$u^{g-1}(1-u)^{h-1}$</td>
</tr>
<tr>
<td>$\text{BeBi}(n, g, h)$</td>
<td>Beta–binomial</td>
<td>$u = 0, 1, \ldots, n$</td>
<td>$\left(\begin{array}{c} n \ u \end{array}\right) \frac{\text{B}(g+u, h+n-u)}{\text{B}(g, h)}$</td>
</tr>
<tr>
<td>$\text{Bi}(n, l)$</td>
<td>Binomial</td>
<td>$u = 0, 1, \ldots, n$</td>
<td>$\left(\begin{array}{c} n \ u \end{array}\right) l^u(1-l)^{n-u}$</td>
</tr>
<tr>
<td>$\text{Ch}(g, h)$</td>
<td>Scaled chi-squared</td>
<td>$u &gt; 0$</td>
<td>$\frac{(\frac{1}{2}h)^{\frac{3}{2}u}u^{(u/2)-1}\exp(-\frac{1}{2}hu)}{\Gamma(\frac{3}{2}g)}$</td>
</tr>
<tr>
<td>$\text{El}(b, c)$</td>
<td>Exponential left-sided</td>
<td>$u &lt; b$</td>
<td>$c \exp{-c(b-u)}$</td>
</tr>
<tr>
<td>$\text{Er}(b, c)$</td>
<td>Exponential right-sided</td>
<td>$u &gt; b$</td>
<td>$c \exp{-c(u-b)}$</td>
</tr>
<tr>
<td>$\text{Ex}(h)$</td>
<td>Exponential</td>
<td>$u &gt; 0$</td>
<td>$he^{-hu}$</td>
</tr>
<tr>
<td>$\text{Ga}(g, h)$</td>
<td>Gamma</td>
<td>$u &gt; 0$</td>
<td>$\frac{h^u u^{g-1}\exp(-hu)}{\Gamma(g)}$</td>
</tr>
<tr>
<td>Notation</td>
<td>Name</td>
<td>Domain restrictions</td>
<td>Density function</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Ge((l))</td>
<td>Geometric</td>
<td>(u = 0, 1, 2, \ldots)</td>
<td>(f_{u} = u^{-1}(1 - 1))</td>
</tr>
<tr>
<td>InBe((k, g, h))</td>
<td>Inverse-beta</td>
<td>(u &gt; 0)</td>
<td>(h_{u}^{k}u_{u}^{h-1}B(k, g)(h + u)^{k+g})</td>
</tr>
<tr>
<td>NeBi((n, l))</td>
<td>Negative-binomial</td>
<td>(u = 0, 1, 2, \ldots)</td>
<td>((n + u - 1)u^{n}(1 - 1)^{n})</td>
</tr>
<tr>
<td>No((b, c))</td>
<td>Normal</td>
<td>(u \in \mathbb{R}^1)</td>
<td>((2\pi)^{-1/2}c^{1/2}\exp\left{-1/2c(u - b)^{2}\right})</td>
</tr>
<tr>
<td>Pa((g, h))</td>
<td>Pareto</td>
<td>(u &gt; g)</td>
<td>(h_{u}^{g})</td>
</tr>
<tr>
<td>Po(h)</td>
<td>Poisson</td>
<td>(u = 0, 1, 2, \ldots)</td>
<td>(\exp(-h)h_{u}^{u}/u!)</td>
</tr>
<tr>
<td>Si((k, g, h))</td>
<td>Siegel</td>
<td>(u &gt; 0)</td>
<td>(h_{u}^{(x/2)-1}B(1/2, 1/2)(1 + h^{-1}u)^{(h+1)/2})</td>
</tr>
<tr>
<td>St((k, b, c))</td>
<td>Student</td>
<td>(u \in \mathbb{R}^1)</td>
<td>(1/B(1/2, 1/2)(kce^{-1}(u - b)^{2})^{(k+1)/2})</td>
</tr>
</tbody>
</table>
Appendix 1

Inter-relationships of distributions in $\mathbb{R}^1$

\begin{align*}
\text{Ch}(g, h) &= \text{Ga}(\frac{1}{2}g, \frac{1}{2}h) \quad \text{(A6)} \\
\text{Er}(0, c) &= \text{Ex}(c) \quad \text{(A7)} \\
\text{Ex}(h) &= \text{Ga}(1, h) \quad \text{(A8)} \\
\text{Ge}(l) &= \text{NeBi}(1, l) \quad \text{(A9)} \\
\text{Si}(k, g, h) &= \text{InBe}(\frac{1}{2}g, \frac{1}{2}k, h) \quad \text{(A10)}
\end{align*}

Relationships of distributions in $\mathbb{R}^1$ to $N, \chi^2, t$ and $F$ notation

\begin{align*}
\text{No}(b, c) &= N(b, 1/c), \text{ a normal distribution with mean } b \text{ and variance } 1/c. \quad \text{(A11)} \\
\text{Ch}(g, 1) &= \chi^2(g), \text{ a chi-squared distribution with } g \text{ degrees of freedom.} \quad \text{(A12)} \\
\text{St}(k, 0, 1) &= t(k) \quad \text{a } t\text{-distribution with } k \text{ degrees of freedom.} \quad \text{(A13)} \\
\text{Si}(k, g, k/g) &= F(k, g), \text{ an } F\text{-distribution with } k \text{ and } g \text{ degrees of freedom.} \quad \text{(A14)}
\end{align*}

Quantiles

The $q$-quantile of any one-dimensional distribution is defined as

(i) the value $*$ such that \( \int_{\mathcal{U}} p(u) \, du = q \) for continuous distributions;
(ii) the value $*$ such that \( \sum_{u < *} p(u) \leq q \), \( \sum_{u < *} p(u) > q \) for discrete distributions.

We shall denote the $q$-quantiles for particular distributions in the following simple way:

\begin{align*}
\text{BeBi}(n, g, h; q), \quad \text{Po}(h; q), \quad \text{No}(b, c; q), \text{ etc.}
\end{align*}

In particular as special cases

\begin{align*}
\text{Ch}(g, 1; q) &= \chi^2(g; q), \quad \text{(A15)} \\
\text{No}(0, 1; q) &= \Phi^{-1}(q), \text{ where } \Phi \text{ is the } N(0, 1) \text{ distribution function,} \quad \text{(A16)} \\
\text{St}(k, 0, 1; q) &= t(k; q), \quad \text{(A17)} \\
\text{Si}(k, g, k/g; q) &= F(k, g; q). \quad \text{(A18)}
\end{align*}
Appendix 1

Cumulative Distributions

Incomplete beta function — as tabulated by Pearson (1934)

\[ I_a(g, h) = \int_0^a \frac{u^{g-1}(1-u)^{h-1}}{B(g, h)} \, du. \quad (A19) \]

Incomplete gamma distribution — as tabulated by Pearson (1922)

\[ J_a(g) = \int_0^a u^{g-1} \exp(-u) \, du. \quad (A20) \]

Hypergeometric cumulative distribution — as tabulated by Lieberman and Owen (1961)

\[ P_{hy}(N, n, k, a) = \sum_{u=\max(0, n+k-N)}^a \frac{k! n!}{(k-u)! (n-u)! u!} \times \frac{(N-k)!(N-n)!}{N!(N-k-n+u)!} \]

where \( N, n, k = 0, 1, 2, ... ; a \leq n \leq N, a \leq k \leq N. \quad (A21) \]

Some useful relationships

\[ \sum_a^N \text{Bi}(n, l) = I_1(a, n-a+1) \quad (a > 0) \quad (A22) \]

\[ \sum_a^N \text{BeBi}(n, l) = I_1(a, n) \quad (a > 0) \quad (A23) \]

\[ \int_0^a \text{lnBe}(k, g, h) = I_{h/(h+a)}(g, k) \quad (a > 0) \quad (A24) \]

\[ \int_0^a \text{Si}(k, g, h) = I_{h/(h+a)}(\frac{1}{2}k, \frac{1}{2}g) \quad (a > 0) \quad (A25) \]

\[ \int_0^a \text{St}(k, b, c) = \frac{1}{2}I_{(1+b)(a-c)^{-1}}(a-b, 1-1^{-1}(\frac{1}{2}k, \frac{1}{2}) (a > b) \quad (A26) \]

\[ \sum_a^N \text{Po}(h) = J_2(h) \quad (a > 0) \quad (A27) \]

\[ \int_0^a \text{Ch}(g, h) = J_{2\text{ha}}(\frac{1}{2}g) \quad (a > 0) \quad (A28) \]

\[ \int_0^a \text{Ga}(g, h) = J_{2\text{ha}}(g) \quad (a > 0) \quad (A29) \]

\[ \sum_a^N \text{BeBi}(n, g, h) = P_{hy}(a + g + h - 1, a + g, a) (a > 0) \quad (A30) \]
Distributions within \( \mathbb{R}^2 \): random vector \((u, w)\); parameter restrictions \(b \in \mathbb{R}^1, c > 0, g > 0, h > 0, k > 0\).

<table>
<thead>
<tr>
<th>Notation</th>
<th>Name</th>
<th>Domain restrictions</th>
<th>Density function</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{ElGa}(b, c, g, h) )</td>
<td>Exponential</td>
<td>( u &lt; b ), ( w &gt; 0 )</td>
<td>( p(u</td>
</tr>
<tr>
<td>( \text{ErGa}(b, c, g, h) )</td>
<td>Exponential</td>
<td>( u &gt; b ), ( w &gt; 0 )</td>
<td>( p(u</td>
</tr>
<tr>
<td>( \text{NoCh}(b, c, g, h) )</td>
<td>Normal-scaled chi-squared</td>
<td>( u \in \mathbb{R} ), ( w &gt; 0 )</td>
<td>( p(u</td>
</tr>
<tr>
<td>( \text{StSi}(k; b, c; g, h) )</td>
<td>Student-Siegel</td>
<td>( u \in \mathbb{R} ), ( w &gt; 0 )</td>
<td>( D(\frac{1}{2}, \frac{1}{2}, \frac{3}{2}; g) k \text{E}^{\frac{2}{2}} { (k \sigma)^{-1} (u - b)^2 + \frac{\sigma^2}{w} }^{(k \sigma^2 + 1)/2} )</td>
</tr>
<tr>
<td>Notation</td>
<td>Name</td>
<td>Domain restrictions</td>
<td>Density function</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Di(g, h)</td>
<td>Dirichlet</td>
<td>(0 &lt; u_i &lt; 1) for (i = 1, 2, \ldots, d) and (\sum u_i &lt; 1)</td>
<td>(\frac{u_1^{g_1-1} \cdots u_d^{g_d-1} (1 - \sum u_i)^{h-1}}{D(g, h)})</td>
</tr>
<tr>
<td>Di*(g, h)</td>
<td>Ordered Dirichlet</td>
<td>(0 &lt; u_1 &lt; u_2 &lt; \ldots &lt; u_d &lt; 1)</td>
<td>(\frac{u_1^{g_1-1} (u_1 - u_d)^{g_d-1} \cdots (u_d - u_{d-1})^{g_{d-1}} (1 - u_{d-1})^{h-1}}{D(g, h)})</td>
</tr>
<tr>
<td>DiMu(n, g, h)</td>
<td>Dirichlet-multinomial</td>
<td>(u_i = 0, 1, \ldots, n) for (i = 1, 2, \ldots, d) and (\sum u_i &lt; n)</td>
<td>(\frac{n!}{u_1^{g_1} \cdots u_d^{g_d} (1 - \sum u_i)^{n - \sum u_i}})</td>
</tr>
<tr>
<td>Mu(n, I)</td>
<td>Multinomial</td>
<td>(u_i = 0, 1, \ldots, n) for (i = 1, 2, \ldots, d) and (\sum u_i &lt; n)</td>
<td>(\frac{n!}{u_1^{g_1} \cdots u_d^{g_d} (1 - \sum u_i)^n})</td>
</tr>
<tr>
<td>NeMu(n, I)</td>
<td>Negative-multinomial</td>
<td>(u_i = 0, 1, 2, \ldots, I) for (i = 1, 2, \ldots, d)</td>
<td>(\frac{n!}{u_1^{g_1} \cdots u_d^{g_d} (1 - \sum u_i)^{n}})</td>
</tr>
<tr>
<td>No_d(b, c)</td>
<td>Normal</td>
<td>(u \in \mathbb{R}^d)</td>
<td>(1)</td>
</tr>
<tr>
<td>St_d(k, b, c)</td>
<td>Student</td>
<td>(u \in \mathbb{R}^d)</td>
<td>(\frac{1}{D[\frac{1}{2}, \frac{1}{2}(k - d + 1); kc]^1/2 {1 + (u - b)'(kc)^{-1}(u - b)}^{(r + 1)/2}})</td>
</tr>
</tbody>
</table>
Appendix I

Relationships of distributions within $\mathbb{R}^d$ and $\mathbb{R}^1$

When $d = 1$,

\[
\begin{align*}
D_i(g, h) &= B_e(g, h) \\
D_iMu(n, g, h) &= B_eBi(n, g, h) \\
Mu(n, l) &= B_i(n, l) \\
N_eMu(n, l) &= N_eBi(n, l) \\
N_0_d(b, c) &= No(b, c) \\
St_0(k, b, c) &= St(k, b, c)
\end{align*}
\]  (A31)

\[
\begin{align*}
Di(g, h) &= Be(g, h) \\
D_iMu(n, g, h) &= BeBi(n, g, h) \\
Mu(n, l) &= B_i(n, l) \\
N_eMu(n, l) &= N_eBi(n, l) \\
N_0_d(b, c) &= No(b, c) \\
St_0(k, b, c) &= St(k, b, c)
\end{align*}
\]  (A32)

\[
\begin{align*}
W_i(g, h) &= C_i(g, h).
\end{align*}
\]  (A33)

\[
\begin{align*}
W_i(g, h) &= C_i(g, h).
\end{align*}
\]  (A34)

\[
\begin{align*}
N_0 Wi_d(b, c; g, h) &= NoCh(b, c, g, h) \\
StSi_d(k; b, c; g, h) &= StSi(k; b, c; g, h).
\end{align*}
\]  (A35)

\[
\begin{align*}
N_0 Wi_d(b, c; g, h) &= NoCh(b, c, g, h) \\
StSi_d(k; b, c; g, h) &= StSi(k; b, c; g, h).
\end{align*}
\]  (A36)

Relationships of distributions within $\mathbb{S}^d$ and $\mathbb{R}^1$

When $d = 1$,

\[
\begin{align*}
S_i_d(k, g, h) &= S_i(k, g, h), \\
W_i_d(g, h) &= C_i(g, h).
\end{align*}
\]  (A37)

\[
\begin{align*}
S_i_d(k, g, h) &= S_i(k, g, h), \\
W_i_d(g, h) &= C_i(g, h).
\end{align*}
\]  (A38)

Relationships of distributions within $\mathbb{R}^d \times \mathbb{S}^d$ and $\mathbb{R}^2$

When $d = 1$,

\[
\begin{align*}
N_0 Wi_d(b, c, g, h) &= NoCh(b, c, g, h) \\
StSi_d(k; b, c; g, h) &= StSi(k; b, c; g, h).
\end{align*}
\]  (A39)

\[
\begin{align*}
N_0 Wi_d(b, c, g, h) &= NoCh(b, c, g, h) \\
StSi_d(k; b, c; g, h) &= StSi(k; b, c; g, h).
\end{align*}
\]  (A40)
Distributions within $\mathbb{S}^d$: random positive definite symmetric matrix $w$; parameter restrictions $k > d - 1, g > d - 1, h \in \mathbb{S}^d$

<table>
<thead>
<tr>
<th>Notation</th>
<th>Name</th>
<th>Domain restrictions</th>
<th>Density function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_d(k, g, h)$</td>
<td>Siegel</td>
<td>$w \in \mathbb{S}^d$</td>
<td>$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$B_d(\frac{1}{2}k, \frac{1}{2}g)h^{\frac{1}{2}k}I_d + h^{-1}w^{(k + g)/2}$</td>
</tr>
<tr>
<td>$W_{id}(g, h)$</td>
<td>Wishart</td>
<td>$w \in \mathbb{S}^d$</td>
<td>$\frac{|h|^2/2</td>
</tr>
</tbody>
</table>

Distributions within $\mathbb{R}^d \times \mathbb{S}^d$: random $(u, w)$; parameter restrictions: $b \in \mathbb{R}^d, c > 0, g > d - 1, c \in \mathbb{S}^d, k > d - 1$

<table>
<thead>
<tr>
<th>Notation</th>
<th>Name</th>
<th>Domain restrictions</th>
<th>Density function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{Wid}(b, c, g, h)$</td>
<td>Normal–Wishart</td>
<td>$u \in \mathbb{R}^d \cap w \in \mathbb{S}^d$</td>
<td>$p(u</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p(w) =$ Wishart $(g, h)$</td>
</tr>
<tr>
<td>$StS_{id}(k; b, c; g, h)$</td>
<td>Student–Siegel</td>
<td>$u \in \mathbb{R}^d \cap w \in \mathbb{S}^d$</td>
<td>$\frac{\Gamma_d(\frac{1}{2}(k + g + 1))</td>
</tr>
<tr>
<td>Type of problem</td>
<td>Future experiment</td>
<td>Predictive density function</td>
<td>Action set $A$</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Decisive prediction</td>
<td>Single, $f$</td>
<td>$p(y</td>
<td>x)$</td>
</tr>
<tr>
<td>Simple sampling</td>
<td>Single, $f$</td>
<td>$p(y</td>
<td>x)$</td>
</tr>
<tr>
<td>inspection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation Optimisation</td>
<td>Class $F$ indexed by $T$</td>
<td>$p(yIT, z)$</td>
<td>$T$</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Class $F$ indexed by $T$</td>
<td>$p(U</td>
<td>y, z)$ (calibrative) $p(U</td>
</tr>
<tr>
<td>Treatment</td>
<td>Class indexed by initial state $T$ and treatment $A$</td>
<td>$p(y</td>
<td>a, t, z)$ (prognostic)</td>
</tr>
<tr>
<td>allocation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In addition to books and research papers referred to directly in the text this Bibliography includes a selection of other references (indicated by an asterisk) to provide the interested reader with material for further study.

Bibliography


Bibliography

Bibliography


*Mitra, S.K. (1957). Tables for tolerance limits for a normal population based on sample

mean and range or mean range. J. Amer. Statist. Ass. 52, 88–94.


60, 339–43.


Owen, D.B. (1962). Handbook of Statistical Tables. Reading, Massachusetts: Addison-

Wesley.

*Owen, D.B., Crasswell, K.J. and Hanson, D.L. (1964). Non-parametric upper confidence

bounds for Pr(Y < X) and confidence limits for Pr(Y < X) when X and Y are normal.


Proschan, F. (1953). Confidence and tolerance intervals for the normal distribution. J.

Amer. Statist. Ass. 48, 550–64.


University Press.


Stanford: University Press.


Berkeley: University of California Press.


27, 865 (abstract).


*Scott, A.J. and Symons, M.J., (1971). A note on shortest prediction intervals for log-

linear regression. Technometrics 13, 889–94.


# Author index

The full references for the authors listed here are given in the bibliography, which also includes a selection of other references to provide the interested reader with material for further study.

<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson, T.W.</td>
<td>104, 235</td>
</tr>
<tr>
<td>Ando, A.</td>
<td>39</td>
</tr>
<tr>
<td>Aoki, M.</td>
<td>178</td>
</tr>
<tr>
<td>Bain, L.J.</td>
<td>107</td>
</tr>
<tr>
<td>Barnett, B.N.</td>
<td>158</td>
</tr>
<tr>
<td>Beatty, G.H.</td>
<td>119, 128</td>
</tr>
<tr>
<td>Berelson, J.</td>
<td>200, 208</td>
</tr>
<tr>
<td>Bowker, A.H.</td>
<td>208</td>
</tr>
<tr>
<td>Bratcher, T.L.</td>
<td>39</td>
</tr>
<tr>
<td>Brown, J.A.C.</td>
<td>35</td>
</tr>
<tr>
<td>Brown, J.J.</td>
<td>84</td>
</tr>
<tr>
<td>Brownlee, K.A.</td>
<td>208</td>
</tr>
<tr>
<td>Campling, G.E.G.</td>
<td>146, 158</td>
</tr>
<tr>
<td>Chetty, V.K.</td>
<td>39, 178</td>
</tr>
<tr>
<td>Chew, V.</td>
<td>39, 107</td>
</tr>
<tr>
<td>Davies, O.L.</td>
<td>178</td>
</tr>
<tr>
<td>De Finetti, B.</td>
<td>39</td>
</tr>
<tr>
<td>De Groot, M.H.</td>
<td>65</td>
</tr>
<tr>
<td>Draper, N.R.</td>
<td>159</td>
</tr>
<tr>
<td>Dunsmore, I.R.</td>
<td>178, 189, 208, 235</td>
</tr>
<tr>
<td>Eisenhart, C.</td>
<td>208</td>
</tr>
<tr>
<td>Ellison, B.E.</td>
<td>119, 128</td>
</tr>
<tr>
<td>Ferguson, T.S.</td>
<td>65</td>
</tr>
<tr>
<td>Ferriss, J.B.</td>
<td>84</td>
</tr>
<tr>
<td>Fisher, R.A.</td>
<td>39, 235</td>
</tr>
<tr>
<td>Fraser, R.</td>
<td>84</td>
</tr>
<tr>
<td>Geisser, S.</td>
<td>39, 235</td>
</tr>
<tr>
<td>Guthrie, D.</td>
<td>158</td>
</tr>
<tr>
<td>Guttmann, L.</td>
<td>39, 107, 123, 128, 143, 159, 246, 247</td>
</tr>
<tr>
<td>Habberma, D.</td>
<td>220</td>
</tr>
<tr>
<td>Halpern, M.</td>
<td>200, 202, 203, 208</td>
</tr>
<tr>
<td>Hartley, H.O.</td>
<td>115, 116, 205</td>
</tr>
<tr>
<td>Hoadley, A.B.</td>
<td>202, 208</td>
</tr>
<tr>
<td>Howe, W.G.</td>
<td>119, 120, 123, 128, 142</td>
</tr>
<tr>
<td>Hunt, H.H.</td>
<td>39</td>
</tr>
<tr>
<td>Jeffreys, H.</td>
<td>39</td>
</tr>
<tr>
<td>Johns, M.V.</td>
<td>158</td>
</tr>
<tr>
<td>Kabe, D.G.</td>
<td>39</td>
</tr>
<tr>
<td>Kalotay, A.J.</td>
<td>208</td>
</tr>
<tr>
<td>Kaufman, G.M.</td>
<td>39</td>
</tr>
<tr>
<td>Kay, A.W.</td>
<td>84</td>
</tr>
<tr>
<td>Kay, J.W.</td>
<td>84, 220, 235</td>
</tr>
<tr>
<td>Krutchkoff, R.G.</td>
<td>202, 208</td>
</tr>
<tr>
<td>Laplace, P.S.</td>
<td>39</td>
</tr>
<tr>
<td>Lever, A.F.</td>
<td>84</td>
</tr>
<tr>
<td>Lieberman, G.J.</td>
<td>96, 97, 118, 119, 128, 208, 253</td>
</tr>
<tr>
<td>Lindley, D.V.</td>
<td>39, 158, 178</td>
</tr>
<tr>
<td>Linnig, F.J.</td>
<td>208</td>
</tr>
<tr>
<td>Mandel, J.</td>
<td>208</td>
</tr>
<tr>
<td>Maritz, J.S.</td>
<td>138, 143</td>
</tr>
<tr>
<td>Martinelli, S.</td>
<td>200, 208</td>
</tr>
<tr>
<td>Mises, R. von</td>
<td>137</td>
</tr>
<tr>
<td>Nakamizo, T.</td>
<td>178</td>
</tr>
<tr>
<td>Neville, A.M.</td>
<td>84</td>
</tr>
<tr>
<td>Neyman, J.</td>
<td>143</td>
</tr>
<tr>
<td>Odén, A.</td>
<td>208</td>
</tr>
<tr>
<td>O'Muircheartaigh, I.G.</td>
<td>84</td>
</tr>
<tr>
<td>Owen, D.B.</td>
<td>96, 97, 142, 253</td>
</tr>
<tr>
<td>Paulson, E.</td>
<td>107</td>
</tr>
<tr>
<td>Pearson, E.S.</td>
<td>39, 115, 116, 205</td>
</tr>
<tr>
<td>Pearson, K.</td>
<td>98, 116, 127, 253</td>
</tr>
<tr>
<td>Proschan, F.</td>
<td>107</td>
</tr>
<tr>
<td>Author</td>
<td>Pages</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Raiffa, H.</td>
<td>39, 65, 136</td>
</tr>
<tr>
<td>Resnikoff, G.J.</td>
<td>118</td>
</tr>
<tr>
<td>Robbins, H.</td>
<td>137, 138, 143</td>
</tr>
<tr>
<td>Roberts, H.V.</td>
<td>39</td>
</tr>
<tr>
<td>Robertson, J.I.S.</td>
<td>84</td>
</tr>
<tr>
<td>Sawagari, Y.</td>
<td>178</td>
</tr>
<tr>
<td>Scheffé, H.</td>
<td>208</td>
</tr>
<tr>
<td>Schlafner, R.</td>
<td>39, 65, 136</td>
</tr>
<tr>
<td>Schucany, W.R.</td>
<td>39</td>
</tr>
<tr>
<td>Sculthorpe, Diane</td>
<td>39, 65, 84, 107, 128, 133, 135, 143</td>
</tr>
<tr>
<td>Shukla, G.K.</td>
<td>208</td>
</tr>
<tr>
<td>Siegel, C.L.</td>
<td>249</td>
</tr>
<tr>
<td>Sunahara, Y.</td>
<td>178</td>
</tr>
<tr>
<td>Symington, T.</td>
<td>84</td>
</tr>
<tr>
<td>Teicher, H.</td>
<td>136</td>
</tr>
<tr>
<td>Thatcher, A.R.</td>
<td>39</td>
</tr>
<tr>
<td>Tiao, G.C.</td>
<td>39, 246, 247</td>
</tr>
<tr>
<td>Wald, A.</td>
<td>65, 119, 128</td>
</tr>
<tr>
<td>Wallis, W.A.</td>
<td>119, 128</td>
</tr>
<tr>
<td>Weeks, D.L.</td>
<td>107</td>
</tr>
<tr>
<td>Weissberg, A.</td>
<td>119, 128</td>
</tr>
<tr>
<td>Wetherill, G.B.</td>
<td>146, 158, 159</td>
</tr>
<tr>
<td>Wilks, S.S.</td>
<td>107, 143</td>
</tr>
<tr>
<td>Williams, E.J.</td>
<td>200, 208</td>
</tr>
<tr>
<td>Winkler, R.L.</td>
<td>39</td>
</tr>
<tr>
<td>Wolfowitz, J.</td>
<td>119, 128</td>
</tr>
<tr>
<td>Zellner, A.</td>
<td>39, 178</td>
</tr>
<tr>
<td>Subject index</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>action set 45, 54, 147, 163, 199, 234, 258</td>
<td></td>
</tr>
<tr>
<td>arrival parameter 186–7, 217–8</td>
<td></td>
</tr>
<tr>
<td>atypicality 77, 224–7, 231, 233, 235–6</td>
<td></td>
</tr>
<tr>
<td>Bayesian cover 67–70, 84–6</td>
<td></td>
</tr>
<tr>
<td>Bayesian coverage 77–8, 84, 86, 93, 131–2, 153</td>
<td></td>
</tr>
<tr>
<td>Bayesian mean coverage 77–8, 93–4</td>
<td></td>
</tr>
<tr>
<td>Bayesian prediction, most plausible 70–2, 84–6, 131, 199</td>
<td></td>
</tr>
<tr>
<td>Bayes's theorem 18, 23, 26, 136, 187–8, 219, 228</td>
<td></td>
</tr>
<tr>
<td>beta distribution 20, 28, 34–5, 99, 123, 158, 148, 150–2, 155, 158, 167, 178, 235, 250, 256</td>
<td></td>
</tr>
<tr>
<td>beta-binomial distribution 24, 28, 34, 48, 63, 150, 155, 165, 250, 253, 256</td>
<td></td>
</tr>
<tr>
<td>binomial distribution 20, 27–8, 33, 88, 94, 96, 114–5, 128–9, 148, 250, 253, 256</td>
<td></td>
</tr>
<tr>
<td>binomial trials 39, 83, 90, 108, 112</td>
<td></td>
</tr>
<tr>
<td>bivariate normal distribution 130, 217, 221</td>
<td></td>
</tr>
<tr>
<td>calibration 12–3, 162, 181–211, 215, 239, 258</td>
<td></td>
</tr>
<tr>
<td>designed 184, 189–90, 195, 198, 218</td>
<td></td>
</tr>
<tr>
<td>natural 184, 186–90, 192, 197, 218</td>
<td></td>
</tr>
<tr>
<td>calibration curve 15</td>
<td></td>
</tr>
<tr>
<td>calibrative distribution 185–200, 206–7, 215, 258</td>
<td></td>
</tr>
<tr>
<td>case records 216–18</td>
<td></td>
</tr>
<tr>
<td>classification 171</td>
<td></td>
</tr>
<tr>
<td>confidence interval 79, 114–16, 201–4</td>
<td></td>
</tr>
<tr>
<td>conjugate prior distribution 19–20, 26, 39–40, 51, 166, 175</td>
<td></td>
</tr>
<tr>
<td>cost function 164, 170, 174, 176–7</td>
<td></td>
</tr>
<tr>
<td>cover 70–1, 73, 77, 79–80, 84–6, 88, 90–1, 110</td>
<td></td>
</tr>
<tr>
<td>coverage 82–3, 94, 99, 117–19, 124, 131</td>
<td></td>
</tr>
<tr>
<td>decisive prediction 4, 45–68, 70, 131–2, 146, 258</td>
<td></td>
</tr>
<tr>
<td>diagnosis 12–13, 39, 77, 162, 212–37, 239, 258</td>
<td></td>
</tr>
<tr>
<td>designed 218, 220, 231</td>
<td></td>
</tr>
<tr>
<td>differential 75–6, 212</td>
<td></td>
</tr>
<tr>
<td>estimative 227–33</td>
<td></td>
</tr>
<tr>
<td>natural 218, 220–2, 231</td>
<td></td>
</tr>
<tr>
<td>predictive 227–33</td>
<td></td>
</tr>
<tr>
<td>diagnostic assessment 12</td>
<td></td>
</tr>
<tr>
<td>diagnostic distribution 215–24, 226, 234, 236, 258</td>
<td></td>
</tr>
<tr>
<td>Dirichlet distribution 21, 28, 139–41, 221, 255–6</td>
<td></td>
</tr>
<tr>
<td>ordered 139–40, 255</td>
<td></td>
</tr>
<tr>
<td>Dirichlet-multinomial distribution 24, 28, 255–6</td>
<td></td>
</tr>
<tr>
<td>discriminant analysis 12</td>
<td></td>
</tr>
<tr>
<td>distribution-free prediction 4, 138–43</td>
<td></td>
</tr>
<tr>
<td>distribution-free tolerance intervals 140–3</td>
<td></td>
</tr>
<tr>
<td>empirical Bayes methods 4, 137–8, 143–5</td>
<td></td>
</tr>
<tr>
<td>exponential left-sided distribution 22, 250, 254</td>
<td></td>
</tr>
</tbody>
</table>
| exponential left-sided gamma distri-
subject index

bution 22, 29-30, 87, 254
exponential right-sided distribution 22, 29-32, 105-6, 108, 129, 250, 252, 254
exponential right-sided gamma distribution 22, 254
exponential two-parameter distribution 22, 29, 31, 39, 105, 124
exponential weighing 1
F-distribution 104, 252
feature vector 215, 217-18, 220-1, 223, 225, 228, 233, 235
fiducial approach 39
frequentist decision theory 132-3, 143
frequentist linear utility theory 133-5
choice of 4-6, 162
detection of 6, 12
gamma distribution 20, 26-30, 32-3, 41-2, 44, 51, 86, 88, 99-100, 105-6, 117, 128, 136, 144-5, 160, 166, 250, 252-4
gamma distribution 107, 128, 251-2
guarantee 110, 113-14, 124
guarantee function 110-12, 118-19
hypergeometric distribution 48, 63
hypergeometric distribution 96-7, 253
inadmissible prediction 57
incomplete beta function 48-9, 63, 73, 75, 79, 98-9, 105, 115-6, 141-2, 166, 171, 226, 245, 253
incomplete gamma function 116, 124-7, 253
index set 162, 171-2, 185, 198, 239
induced utility 64, 133
informative experiment 1-2, 4-6, 8, 13, 18-9, 23, 27, 31-2, 37, 40-3, 45, 51, 68, 77, 79-80, 83, 86, 91-2, 94, 97-8, 101, 103-4, 107, 113-16, 118, 122, 128, 133, 136-8, 140, 142-3, 146, 151, 154, 162, 166, 168-9, 172, 174-5, 177, 181, 199, 205, 209-11, 213, 218, 220, 239, 241
informative prediction 4, 47, 65, 68-87, 131, 135, 137, 153
informative prediction, Bayesian 68-9
interval prediction 50, 55-7, 62, 66
all-or-nothing 70
linear utility 56
inverse 92-3, 99, 101
inverse beta distribution 24, 26-8, 30, 48, 51, 63, 78, 144, 165, 251-3
inverse prediction 183
linear discriminant 235
linear least squares prediction 1
lognormal distribution 35
multinomial distribution 21, 28, 41, 255-6
multinormal distribution 21, 27, 29, 31, 73, 75, 84, 104, 107, 109, 123, 128, 172, 220-1, 226, 230-1, 235, 255-7
multivariate regression 39
negative binomial distribution 24, 28, 48, 63, 165-7, 171, 251-3, 256
negative multinomial distribution 42, 255-6
normal chi-squared distribution 21, 28, 35-6, 42, 168, 172, 175, 190, 254, 256
normal- Wishart distribution 21, 29, 221, 256-7
optimisation 6, 162, 173-8, 183, 239, 258
order statistics 31, 140-1
parameter space 45, 136, 163, 198, 234
Pareto distribution 39-40, 44, 66, 107, 129, 251
Pareto-gamma distribution 44
point prediction 45, 50, 54-7, 69, 78, 199, 200
all-or-nothing 46, 55
linear loss 46, 62, 164
quadratic loss 47
Poisson distribution 17, 20, 27-8, 33, 42, 66, 97, 98, 114, 116, 128-9, 145, 166, 170, 208, 251, 253
<table>
<thead>
<tr>
<th>Subject index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Poisson process</strong></td>
</tr>
<tr>
<td><strong>posterior distribution</strong></td>
</tr>
<tr>
<td><strong>posterior plausibility function</strong></td>
</tr>
<tr>
<td><strong>prediction interval</strong></td>
</tr>
<tr>
<td><strong>Bayesian</strong></td>
</tr>
<tr>
<td><strong>prediction region</strong></td>
</tr>
<tr>
<td><strong>estimated</strong></td>
</tr>
<tr>
<td><strong>mean of</strong></td>
</tr>
<tr>
<td><strong>mode of</strong></td>
</tr>
<tr>
<td><strong>quantile of</strong></td>
</tr>
<tr>
<td><strong>predictive regression density function</strong></td>
</tr>
<tr>
<td><strong>predictor, simple</strong></td>
</tr>
<tr>
<td><strong>preposterior analysis</strong></td>
</tr>
<tr>
<td><strong>prior distribution</strong></td>
</tr>
<tr>
<td><strong>ignorance</strong></td>
</tr>
<tr>
<td><strong>information</strong></td>
</tr>
<tr>
<td><strong>plausibility function</strong></td>
</tr>
<tr>
<td><strong>predictive distribution</strong></td>
</tr>
<tr>
<td><strong>prognostic distribution</strong></td>
</tr>
<tr>
<td><strong>range of previous experience</strong></td>
</tr>
<tr>
<td><strong>region of previous experience</strong></td>
</tr>
<tr>
<td><strong>regression</strong></td>
</tr>
<tr>
<td><strong>binomial</strong></td>
</tr>
<tr>
<td><strong>gamma</strong></td>
</tr>
<tr>
<td><strong>linear</strong></td>
</tr>
<tr>
<td><strong>normal</strong></td>
</tr>
<tr>
<td><strong>Poisson</strong></td>
</tr>
<tr>
<td><strong>polynomial</strong></td>
</tr>
<tr>
<td><strong>regulation</strong></td>
</tr>
<tr>
<td><strong>finite index set</strong></td>
</tr>
<tr>
<td><strong>point</strong></td>
</tr>
<tr>
<td><strong>set</strong></td>
</tr>
<tr>
<td><strong>response surface</strong></td>
</tr>
<tr>
<td><strong>quadratic</strong></td>
</tr>
<tr>
<td><strong>sample size, choice of</strong></td>
</tr>
<tr>
<td><strong>sample space</strong></td>
</tr>
<tr>
<td><strong>class of measurable subsets</strong></td>
</tr>
<tr>
<td><strong>sampling inspection</strong></td>
</tr>
<tr>
<td><strong>fixed size</strong></td>
</tr>
<tr>
<td><strong>sequential</strong></td>
</tr>
<tr>
<td><strong>set prediction</strong></td>
</tr>
<tr>
<td><strong>all-or-nothing</strong></td>
</tr>
<tr>
<td><strong>set predictor</strong></td>
</tr>
<tr>
<td><strong>Siegel distribution</strong></td>
</tr>
<tr>
<td><strong>multivariate</strong></td>
</tr>
<tr>
<td><strong>simple prediction</strong></td>
</tr>
<tr>
<td><strong>simple predictor</strong></td>
</tr>
<tr>
<td><strong>all-or-nothing</strong></td>
</tr>
<tr>
<td><strong>linear loss</strong></td>
</tr>
<tr>
<td><strong>quadratic loss</strong></td>
</tr>
<tr>
<td><strong>Simpson's rule</strong></td>
</tr>
<tr>
<td><strong>stochastic control</strong></td>
</tr>
<tr>
<td><strong>structural parameter</strong></td>
</tr>
<tr>
<td><strong>Student distribution</strong></td>
</tr>
<tr>
<td><strong>multivariate</strong></td>
</tr>
<tr>
<td><strong>Student-Siegel distribution</strong></td>
</tr>
<tr>
<td><strong>multivariate</strong></td>
</tr>
<tr>
<td><strong>sufficient statistic</strong></td>
</tr>
<tr>
<td><strong>t-distribution, non-central</strong></td>
</tr>
<tr>
<td><strong>T²-distribution</strong></td>
</tr>
<tr>
<td><strong>terminal analysis</strong></td>
</tr>
<tr>
<td><strong>time series analysis</strong></td>
</tr>
<tr>
<td><strong>tolerance regions</strong></td>
</tr>
<tr>
<td><strong>guaranteed coverage</strong></td>
</tr>
<tr>
<td><strong>mean coverage</strong></td>
</tr>
<tr>
<td><strong>similar guaranteed coverage</strong></td>
</tr>
<tr>
<td><strong>similar mean coverage</strong></td>
</tr>
<tr>
<td><strong>treatment allocation</strong></td>
</tr>
<tr>
<td><strong>type of case</strong></td>
</tr>
<tr>
<td><strong>uniform distribution</strong></td>
</tr>
<tr>
<td><strong>uniformly diffuse prior</strong></td>
</tr>
<tr>
<td><strong>utility function</strong></td>
</tr>
<tr>
<td><strong>additive</strong></td>
</tr>
<tr>
<td><strong>all-or-nothing</strong></td>
</tr>
<tr>
<td><strong>expected</strong></td>
</tr>
</tbody>
</table>
Subject index

67–8, 149–51, 154–5, 163, 166, 168, 170, 199, 234, 241, 246
piecewise-linear 56, 66, 131–3, 143–4, 164–6
quadratic 164–5, 167, 169, 178
vague prior distribution 75, 84, 159, 190–1, 206, 246
vague prior information 19–22, 34–5, 39, 42–3, 59
Wishart distribution 21, 29, 75, 104, 221, 256, 257
Example and problem index

airflow through doorways 179
airlock in components 142–3
antibiotic assay 6–9, 205–8
autoanalyser calibration 183–5, 195–7
ball-bearing diameters 42, 84–5
barrier cream 240, 246, 247
batch acceptance 146–61
bird-nesting 34, 129
breakdown times 85
chemical process fault diagnosis 15–16, 237
choice of sample size 154–6
chromatographic assay 209
clearance circle assay 6–9, 210
cloth blemishes 108
component length control 173
component lifetimes 41, 85, 142–3
Conn’s syndrome 10–13, 85, 231–4, 236, 237
crop prediction 103, 142
crushing strength of mortar 179
cushing’s syndrome 73–7, 212–15, 221–7, 230
design components 120–3
destructive testing 146–61
disintegration times 236
double-sampling inspection 160
drug response 176–7
extrusion process 41, 65
fertiliser effectiveness 43, 85
foetal measurement 13, 84
income tax 14, 66, 129
industrial process yield 5–6, 177
insecticide strength 179–80
item characteristics 87
item inspection 208
laminate design 4–5, 170–1
light bulb lifetimes 108, 129
machine tool replacement 2–3, 50–4, 57–60, 66, 70–2, 77–9
medical diagnosis 10–3, 212–5, 221–7, 230, 236
medical prognosis 2, 35–8
metabolic excretion rates 73–7, 85–6, 109
particle emission 166–8
patient reactions 208–9
perishable commodity supply 65
psychiatric diagnosis 235
quality control 3–4, 96–7, 115–16
quality improving process 238–40, 241–5, 247
radiocarbon dating 15, 208
recall time of patients 13–4, 65
relapse times 129
response to stimulus 109, 208–9
seed germination 98
self-destructive components 86, 108
sequential sampling inspection 156–9, 160
setting strength of adhesive 14, 179
shoe manufacture 130
skin allergy 240, 246, 247
sterilising liquid 67
storage tank size 145
suppressor drug dose level 168–70
survival times 2, 35–8
theatre ticket agency 66
treatment allocation 247–8
ultrasonic cephalometry 13, 84
university applicants 178
vitamin assay 209–10
water content of soil 181–3, 192–5, 197, 203–5