ON THE REPARAMETERIZATION OF MODELS TO OBTAIN UNCORRELATED MAXIMUM LIKELIHOOD ESTIMATORS

L. D. WALTER

DEPARTMENT OF STATISTICS, UNIVERSITY OF EDINBURGH

Submitted for the degree of Doctor of Philosophy in the University of Edinburgh

1973
Declaration

The following record of research work is submitted as a Thesis for the Degree of Doctor of Philosophy in the University of Edinburgh, having been submitted for no other Degree. The work was carried out under the supervision of Professor D.J. Finney, F.R.S., of the Department of Statistics, University of Edinburgh. The Thesis has been composed by me, and except where due acknowledgement is made, the work is original.

Loraine D. Walter,
This thesis is concerned primarily with the development of criteria and methods for the reparameterization of models and distributions in order that independent parameter estimators be obtained. The criteria for independence are required not to be functions of the data, which enables their use before estimation of the parameters. Methods for the application of these criteria are developed for this type of use and involve the specification of a family of parameterizations of the given model from which that member(s), if any, is to be selected which satisfies one of the criteria. The establishment of general measures of normality which can be used to indicate how well the estimator distributions for selected parameterizations are approximated by the normal is also desired.

The approach taken in this work is to suggest criteria, test them on simple cases, and then apply them to more sophisticated models (of at most two parameters) for the purposes of developing techniques and discovering problem areas.

Three criteria for independence which do not require knowledge of the estimator distribution or its moments are suggested: zero asymptotic correlation, zero mixed second derivative of the log likelihood, and each first derivative
of the log likelihood a function of one parameter only; the one used in any situation is to be determined by properties of the particular model. These are applied to three models: the two-parameter bivariate exponential distribution, the Gamma distribution, and the bioassay logistic model. The first of these is uninteresting except as a test of the criteria. For the Gamma distribution it was found that parameters which are simple functions of the original parameters satisfy the criterion of zero asymptotic correlation. Several sets of Gamma data are generated for a few parameter values and the estimates of asymptotic correlation obtained from these are compared for estimators of the two parameterizations; those for the new parameters are close to zero and uniformly less than those for the original parameters. (Chapter 3)

Parameterization of the bioassay logistic model is studied in considerable depth. (Chapters 4 and 5) The criteria and methods which worked for the other models failed in this case, since no family of parameterizations could be found which included a data-independent parameterization satisfying one of the criteria. However, there are three different parameterizations of this model already in frequent use ((α, β), (μ, σ), and (μ, β)) and these are compared both with respect to the asymptotic measures of correlation and with respect to some of the exact
properties of the estimator distributions. The asymptotic measures indicate that the \((\mu, \sigma)\) and \((\mu, \beta)\) parameterizations are better in general than \((\alpha, \beta)\), at least for well-designed assays. The exact estimator distributions are obtained by construction for a few parameter values and a particular assay structure (3 equally spaced doses, \(n\) subjects at each) for which the maximum likelihood estimates are calculated with little or no iteration using expressions derived by Wilson and Worcester (1943). The method of construction is an efficient one and although designed specifically for the type of assay described, it could be adapted for use with other assay structures. Some very interesting results arise from the study of the exact properties which includes, in addition to consideration of correlation, discussion of skewness and kurtosis, bias, and probability of indeterminacy, visual examination of marginal densities, and detailed description of indeterminate and infinite estimates.

Some criteria for normality based on the likelihood function are suggested and are applied to parameterization families of the exponential distribution. (Chapter 2) Comparisons of the exact properties of the estimator distributions for several members of the family of power parameter transformations, including the two usual parameters and the one selected by one of the criteria, are
made with some interesting results. One of these is that the selected parameterization is far from being the best; another is that of the two usual parameters, the one representing mean time to failure is better than its inverse, which represents failure rate. These criteria do not, however, provide useful measures of normality; no such measures of sufficient generality were found, so that the assessment of selected parameterizations in the two-parameter models discussed with respect to normality of estimator distributions is minimal.

This study demonstrates that the objective of developing a method of reparameterizing which is not dependent on the data is unlikely to attainable in general, even in the simplest case of two parameters. The main reason for this is that there are no rules for specifying a family of parameterizations so as to ensure that a parameterization satisfying the criterion being applied is included. It is also evident, however, that such an objective can be realized for some particular models. One general result of this investigation is the recommendation that this objective not be pursued without first being modified to permit data-dependent reparameterizations.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>(i)</td>
</tr>
<tr>
<td>Abstract of Thesis</td>
<td>(ii)</td>
</tr>
<tr>
<td>List of Figures</td>
<td>(xi)</td>
</tr>
<tr>
<td>List of Tables</td>
<td>(xiii)</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>(xiv)</td>
</tr>
<tr>
<td>Chapter 1 Introduction</td>
<td>1.1-1.10</td>
</tr>
<tr>
<td>Chapter 2 The Single-Parameter Exponential Distribution: A Simple Case</td>
<td>2.1 - 2.47</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>2.1</td>
</tr>
<tr>
<td>2.2 Specification of Distributions</td>
<td>2.5</td>
</tr>
<tr>
<td>2.2.1 Exact Sampling Distributions</td>
<td>2.6</td>
</tr>
<tr>
<td>2.2.2 Asymptotic Distributions</td>
<td>2.8</td>
</tr>
<tr>
<td>2.3 Asymmetry of the Likelihood Function</td>
<td>2.10</td>
</tr>
<tr>
<td>2.4 Moments and Coefficients of Skewness</td>
<td>2.13</td>
</tr>
<tr>
<td>2.4.1 Conclusions from Tabulations</td>
<td>2.17</td>
</tr>
<tr>
<td>2.5 Confidence Intervals</td>
<td>2.23</td>
</tr>
<tr>
<td>2.5.1 Exact Confidence Limits</td>
<td>2.24</td>
</tr>
</tbody>
</table>
2.5.2 Asymptotic Confidence Limits 2.25
2.5.3 Limits using Bartlett's Transformed Score Function 2.26
2.5.4 Comparison of the Three Confidence Intervals 2.29
2.6 Bias 2.34
2.7 Plots of the Distributions of Estimators 2.37
2.7.1 Standardization of the Estimators 2.37
2.8 Conclusions 2.44

Chapter 3. Some Simple Two-Parameter Distributions 3.1 - 3.41

3.1 Introduction 3.1
3.2 Two-Parameter Bivariate Exponential Distribution 3.4
3.2.1 Specification of the Distribution 3.4
3.2.2 Use of the Mixed Second Derivative of the Log Likelihood 3.8
3.2.3 Use of the First Derivatives of the Log Likelihood 3.11
3.3 Gamma Distribution 3.13
3.3.1 Specification of the Distribution 3.13
3.3.2 Mixed Second Derivatives 3.16
3.3.3 Generation of Data 3.18
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.31 Data Generation via a Pseudo-Random Number Generator</td>
<td>3.20</td>
</tr>
<tr>
<td>3.3.32 Data Generation via Tables of Random Numbers</td>
<td>3.21</td>
</tr>
<tr>
<td>3.3.33 $\chi^2$ Goodness-of-Fit Tests</td>
<td>3.24</td>
</tr>
<tr>
<td>3.3.4 Effects of Reparameterization on Correlation of Estimates</td>
<td>3.24</td>
</tr>
<tr>
<td>3.3.5 Effects of Reparameterization on the Distribution of Estimators</td>
<td>3.27</td>
</tr>
<tr>
<td>3.3.6 Conclusions</td>
<td>3.36</td>
</tr>
<tr>
<td>3.4 Conclusions</td>
<td>3.37</td>
</tr>
<tr>
<td>Chapter 4 Bioassay Logistic Model: Theoretical Discussion</td>
<td>4.1 - 4.23</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>4.1</td>
</tr>
<tr>
<td>4.2 The Logistic Model</td>
<td>4.1</td>
</tr>
<tr>
<td>4.2.1 Likelihood Function and Normal Equations</td>
<td>4.3</td>
</tr>
<tr>
<td>4.3 Correlation of Estimators</td>
<td>4.6</td>
</tr>
<tr>
<td>4.3.1 Comparison of Three Parameterizations of the Logistic Model</td>
<td>4.6</td>
</tr>
<tr>
<td>4.3.2 Reparameterizations for Independence</td>
<td>4.10</td>
</tr>
<tr>
<td>4.4 Normality of Estimator Distributions</td>
<td>4.17</td>
</tr>
<tr>
<td>4.5 Conclusions</td>
<td>4.22</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Problems with Constructed Distributions</td>
</tr>
<tr>
<td>5.2</td>
<td>Construction of the Exact Distributions</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Calculation of Estimates: the Wilson-Worcester Method</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Enumeration of Sufficient Statistics</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Enumeration of Samples and Calculation of Probabilities</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Description of the Computer Programme</td>
</tr>
<tr>
<td>5.3</td>
<td>Adjustment of the Constructed Distributions</td>
</tr>
<tr>
<td>5.4</td>
<td>Properties of Estimator Distributions</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Configurations</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Coefficients of Correlation</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Frequency Polygons</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Coefficients of Skewness and Kurtosis</td>
</tr>
<tr>
<td>5.4.5</td>
<td>Bias</td>
</tr>
<tr>
<td>5.4.6</td>
<td>Probability of Indeterminacy</td>
</tr>
<tr>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td></td>
</tr>
<tr>
<td>5.5 Conclusions</td>
<td>5.64</td>
</tr>
<tr>
<td>Chapter 6 Conclusions</td>
<td>6.1 - 6.22</td>
</tr>
<tr>
<td>6.1 Establishment of Criteria</td>
<td>6.3</td>
</tr>
<tr>
<td>6.2 Application of the Criteria</td>
<td>6.9</td>
</tr>
<tr>
<td>6.3 Conclusions Relating to Specific Models</td>
<td>6.13</td>
</tr>
<tr>
<td>6.4 Summary</td>
<td>6.20</td>
</tr>
<tr>
<td>Appendix I The Method of Wilson and Worcester</td>
<td>I.1 - I.5</td>
</tr>
<tr>
<td>Appendix II Indeterminacies in Estimates of the Parameters of the Logistic Model</td>
<td>II.1 - II.17</td>
</tr>
<tr>
<td>Bibliography</td>
<td>B.1 - B.2</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.71a</td>
<td>Probability Density of $\hat{\phi}_i$, $(i = 0.0, 0.33, 1.0)$, $n = 5$</td>
<td>2.40</td>
</tr>
<tr>
<td>2.71b</td>
<td>Probability Density of $\hat{\phi}_i$, $(i = -0.33, -1.0)$, $n = 5$</td>
<td>2.41</td>
</tr>
<tr>
<td>3.31</td>
<td>Contours of Constant Relative Likelihood of $(\alpha, \beta)$ and $(\theta_1, \theta_2)$ for Data from $G(2.0, 1.0)$</td>
<td>3.31</td>
</tr>
<tr>
<td>3.32</td>
<td>As Figure 3.31 using Data from $G(2.5, 2.0)$</td>
<td>3.32</td>
</tr>
<tr>
<td>4.1</td>
<td>Contours of Constant Relative Likelihood for 4 Parameterizations of the Logistic Model, log doses $-1, 0, 1$, $n = 10$, observed $p_i = (0.3, 0.5, 1.0)$</td>
<td>4.19</td>
</tr>
<tr>
<td>5.21</td>
<td>Flowchart of Routine 'Sample'</td>
<td>5.24</td>
</tr>
<tr>
<td>5.22</td>
<td>Continuation of Routine 'Sample'</td>
<td>5.25</td>
</tr>
<tr>
<td>5.23</td>
<td>Flowchart of Enumeration of Estimator Distributions for Parameters of the Logistic Model</td>
<td>5.29</td>
</tr>
<tr>
<td>5.40</td>
<td>Absolute Values of Correlation Coefficients as a Function of one Parameter</td>
<td>5.41</td>
</tr>
<tr>
<td>5.41</td>
<td>Frequency Polygon Representations of Marginals of Standardized $(\hat{\alpha}, \hat{\beta})$ and $(\hat{\mu}, \hat{\sigma})$ Distributions Compared with the Normal Curve $(n = 10, \beta = 0.84730, P_2 = 0.5)$</td>
<td>5.46</td>
</tr>
<tr>
<td>5.42</td>
<td>As Figure 5.41 with $n = 10, \beta = 0.84730$, $P_2 = 0.7$</td>
<td>5.47</td>
</tr>
<tr>
<td>5.43</td>
<td>As Figure 5.41 with $n = 10, \beta = 0.84730$, $P_2 = 0.8$</td>
<td>5.48</td>
</tr>
<tr>
<td>5.44</td>
<td>As Figure 5.41 with $n = 10, \beta = 2.19722$, $P_2 = 0.6$</td>
<td>5.49</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.45</td>
<td>Coefficients of Skewness as a Function of $P_2$ for fixed values of $\beta$ and $n$</td>
<td>5.53</td>
</tr>
<tr>
<td>5.46</td>
<td>Coefficients of Kurtosis as a Function of $P_2$ for $n = 10$ and $\beta = 0.84730$</td>
<td>5.56</td>
</tr>
<tr>
<td>5.47</td>
<td>Bias of $\hat{\alpha}$ and $\hat{\mu}$ as a Function of their true values for fixed $\beta$ and $n$</td>
<td>5.58</td>
</tr>
<tr>
<td>1</td>
<td>Contours of Constant Negative Log Likelihood for $(\mu, \sigma)$ with $n = 10$, log doses -1, 0, 1 and response triple (0, 0, 1)</td>
<td>II.8</td>
</tr>
<tr>
<td>2</td>
<td>As Figure 1 for response triple (0, 0.5, 1)</td>
<td>II.8</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.41</td>
<td>Coefficients of Skewness of $\hat{\phi}_i$</td>
<td>2.18</td>
</tr>
<tr>
<td>2.42</td>
<td>Variance ($\hat{\phi}_i$)</td>
<td>2.19</td>
</tr>
<tr>
<td>2.5</td>
<td>Comparison of 95% and 99% Confidence Limits for $\alpha$</td>
<td>2.30</td>
</tr>
<tr>
<td>2.6</td>
<td>Bias of $\hat{\phi}_i$</td>
<td>2.36</td>
</tr>
<tr>
<td>3.31</td>
<td>Intervals used in $\chi^2$ Goodness-of-Fit Tests on $n = 30$ Generated $G(\alpha, \beta)$ Variates</td>
<td>3.25</td>
</tr>
<tr>
<td>3.32</td>
<td>Results of $\chi^2$ Tests on Gamma Data</td>
<td>3.26</td>
</tr>
<tr>
<td>3.33</td>
<td>Estimates and Coefficients of Correlation for Generated $G(\alpha, \beta)$ Data</td>
<td>3.28</td>
</tr>
<tr>
<td>3.34</td>
<td>Third Differential Coefficients of Log Likelihood Function $L(\hat{\phi}_1, \hat{\phi}_2)$ evaluated at $(\hat{\phi}_1, \hat{\phi}_2)$</td>
<td>3.34</td>
</tr>
<tr>
<td>5.40</td>
<td>Configurations whose Estimator Distributions are Enumerated</td>
<td>5.36</td>
</tr>
<tr>
<td>5.41</td>
<td>Coefficients of Correlation for $(\hat{\alpha}, \hat{\beta}), (\hat{\mu}, \hat{\sigma}), (\hat{\mu}, \hat{\beta})$</td>
<td>5.40</td>
</tr>
<tr>
<td>5.42</td>
<td>Coefficients of Skewness and Kurtosis for $\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\sigma}$ with $n = 10$</td>
<td>5.52</td>
</tr>
<tr>
<td>5.43</td>
<td>Probabilities of Indeterminacy</td>
<td>5.62</td>
</tr>
<tr>
<td>6.1</td>
<td>Asymptotic and Exact Correlation Coefficients for Three Parameterizations of the Bioassay Logistic Model for Configurations with $\mu \neq 0$</td>
<td>6.17</td>
</tr>
<tr>
<td>II.1</td>
<td>Indeterminacies Classified by Sufficient Statistics</td>
<td>II.5</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank Professor D.J.Finney, F.R.S., of the Department of Statistics in the University of Edinburgh for his guidance, stimulation, and encouragement throughout the research phase of this work. I wish also to express my appreciation to Dr J.M.Last, M.D., Chairman of the Department of Epidemiology and Community Medicine in the University of Ottawa who has allowed me during the eight months of my employment with his department to spend the necessary time on the completion of this thesis, and also to the other members of this department for their cooperation during this time. Finally I thank my husband, Stephen, without whom this thesis could not have been produced, for his competent typing, particularly of the many symbols and algebraic expressions, for his helpful comments, and for his constant encouragement and understanding.
CHAPTER 1

INTRODUCTION

There has been a great deal of work done on the transformation of data, most commonly for the purpose of attaining homogeneity of variance. There is also a sizeable body of information about transformations for normality, including several different approaches. For example, Sprott and Kalbfleisch (1969) and Anscombe (1964) investigate transformations which satisfy certain conditions relating to the symmetry of the likelihood function. Cornish and Fisher (1937) have developed an expression for the density of any statistic with known cumulants which involves a series of operators acting on the normal density. By inversion, an expression can be obtained for a normal deviate in terms of the original variate and its cumulants. Bartlett (1953) illustrates the use of the Cornish-Fisher expansion of the score function, whose moments are known exactly, to obtain confidence limits for parameters.

There has been little work done, however, in the
area of reparameterizations of models to reduce or eliminate the correlation of parameter estimates. For some particular models, such as the simple linear regression model, it is known that a parameterization yielding independent estimators exists, but there are, to my knowledge, no general methods for obtaining such transformations. Guttman and Meeter (1964) have attempted to derive a method for transforming parameters to increase the speed of convergence of non-linear least-squares procedures, but difficulties were encountered with the techniques used and the results were inconclusive, since sometimes the transformation of parameters actually increased computer time. (This was because a non-linear least-squares procedure was necessary to choose the appropriate parameter transformation.) Ross (1970) has investigated parameter transformations to orthogonalize the estimate space. He recognizes that non-linear transformations are essential and introduces the concept of 'stable parameters' (parameters which are little affected by changes in other parameters in regions of high likelihood), but was unable to suggest a general method for obtaining such parameters. Anscombe (1964) has not only considered transformations to make the likelihood function symmetric, but has also looked at reparameterizations to orthogonalize the likelihood
1.3

function, where an orthogonal likelihood is one which factorizes such that each factor contains only one parameter. None of these attempts at reparameterizing to reduce the correlation of estimates is sufficiently general and successful; it would seem then that there is a great deal to be done in this area in the development of both criteria and method.

In a multiparameter model, independent parameter estimates are desirable for two main reasons. Firstly, inferences are straightforward to make and interpret if estimates are independent, since one estimate is used to make inferences and test hypotheses about one parameter, regardless of the values of the other parameters. If, however, estimates are correlated, any inferences about one parameter depend on the values of the other parameters, so that inferences must be either conditional or joint. In terms of confidence limits, this means that either regions of confidence which include all parameters whose estimates are correlated must be set up, or else a single interval for any one parameter must be made conditional on the values of the other parameters. In practice, covariances are often neglected and confidence intervals set up as if the estimates were independent, but how meaningful
are such intervals and how serious is the neglect of correlation? Secondly, if an iterative procedure is necessary to obtain the estimates, convergence is likely to be reasonably rapid and reliable estimates are usually found; if, however, estimates are highly correlated convergence can be slow. This occurs in the method of maximum likelihood, for example, when the contours of constant likelihood are long and narrow and oblique to the parameter axes. In such cases, many iterations may be required in the region of the maximum to achieve only small increases in the value of the likelihood function, with the result that a considerable amount of computer time is used and the estimates obtained may not be as close to the values of the parameters which maximize the likelihood as one would like. The variance-covariance matrix is estimated as the inverse of the matrix of second differences (or second differential coefficients) evaluated at the maximum likelihood estimates, and if the estimates are highly correlated it may happen that this matrix is so nearly singular that unreliable estimates of the variances and covariances are obtained. Since in general the estimates of the parameters of a multiparameter model cannot be expressed in terms of the observations, an iterative procedure to obtain them is usually necessary, and so these
1.5

Considerations of convergence and reliability are highly relevant.

Thus, we would like to develop a general method of reparameterizing a model or distribution function so that the parameter estimates are uncorrelated. It is important also that the distribution of estimates of the transformed parameters can be reasonably well approximated by the normal distribution, not only because assumptions of normality are commonly used in inferences but also because two variates which are uncorrelated are only necessarily independent if they are jointly normally distributed. In large samples, asymptotic assumptions of normality of the estimators may be reasonable, but in small samples estimators need not be normally distributed and a transformation of parameters could have a significant effect on the estimator distribution in such samples.

The first stage in the development of a method of reparameterization is the establishment of criteria, which in this case will involve measures of correlation and normality. If they are to have any generality, these criteria should not depend on knowledge of the estimator distribution; suitable criteria for independence should in addition be independent of the data, thereby enabling reparameter-
1.6

ization to take place before the initiation of any iteration, or else suitable for integration into an iterative maximization procedure; criteria for normality are needed mostly in the evaluation of parameterizations which satisfy some criterion of independence.

As mentioned previously, there has been considerable work done on transformations for normality and several different types of criteria have been put forward (Anscombe (1964), Bartlett (1953), Sprott and Kalbfleisch (1969)). In order to compare and assess a few of these criteria as regards their generality, ease of application, and effectiveness, the parameter of a particularly simple model for which the distribution of the parameter estimator\(^1\) is known is transformed in various ways to satisfy the selected criteria. (See Chapter 2) These parameterizations are then compared with respect to the known exact distributional properties of their parameter estimators. In this fashion, information is gathered about the suitability of some measures of normality for the evaluation of parameterizations. In general, the exact distribution of the estimators is unknown and so one cannot often secure this sort of assurance

\(^1\) In this thesis, an estimator will always be the maximum likelihood estimator unless specified otherwise.
that a reparameterization has actually accomplished what was intended.

The establishment of criteria for independence is more difficult since less work has been done in this area. The most obvious criterion is zero correlation of estimators, but since the exact correlation of a pair of estimators is seldom known, this criterion is of very limited use. Asymptotic correlation is a measure worth investigating and the likelihood function seems a probable source of measures on which criteria for independence might be based. In Chapter 3, some criteria which are not data-dependent and which involve either asymptotic measures or certain properties of the likelihood function are suggested, and the reparameterization of two models to satisfy some of these criteria is attempted. The first of these models is a simple two parameter distribution for which a parameterization having uncorrelated estimators is known to exist, included because it allows some assessment of the criteria. The second is a more complex and interesting model, namely the two parameter Gamma distribution, for which the parameterization is evaluated by such of the established measures of correlation and normality as are appropriate and applicable.
If a criterion is to be used to select a parameterization, the specification of a family of parameterizations from which the member(s) best satisfying the criterion is to be chosen, is essential. In general, this can be a family of polynomial, power, logarithmic, or almost any other non-linear transformations of the original parameters. Unfortunately there are no guidelines for choosing a suitable transformation family in any given situation, so that this step becomes a matter of judicious guess-work and trial and error; yet whether or not a satisfactory reparameterization is obtained is certainly dependent on the appropriateness of the specified family (where an 'appropriate' family is one which actually includes a parameterization satisfying the criterion being applied).

For the models which have been mentioned thus far (and which are discussed in detail in Chapters 2 and 3 as noted) reparameterizations to satisfy at least one of the criteria are found in each case. The last model to be considered is the logistic model for the quantal response bioassay of a single substance. In Chapter 4 several different families of transformations are tried for this model, but none yield a parameterization which satisfies any of the established criteria for independence and is not data-
dependent. In addition, the three commonly used parameterizations of this model are compared with respect to correlation and normality of distribution of the estimators, although the exact estimator distributions are unknown. In Chapter 5 a detailed comparison of these three parameterizations is conducted, based on the properties of the exact estimator distributions constructed for a particular assay structure and a few parameter values. The purpose of such a detailed study is twofold: firstly to gather information about criteria for independence in general, and secondly to gain some insight into possible parameterization families which might be appropriate for use with this model in particular and perhaps with other models as well.

This thesis, then, is concerned primarily with the establishment of criteria for independence which are appropriate for use with most models and which, if an appropriate family of parameterizations is specified, can be relied upon to yield a reparameterization with estimators which are independent or close to it; the establishment of criteria for normality with similar
characteristics of generality and reliability is another objective, although of secondary importance. The ideal is set out to be reparameterization before estimation, and so criteria which are not data-dependent are sought; the establishment of criteria suitable for integration into an iterative routine has been mentioned as a possible alternative ideal, but constitutes another project. It may be that the above goals are unattainable, particularly since the fulfillment of such a strong condition as 'if an appropriate family of parameterizations is specified' is essential. Although this work deals entirely with a few specific models, it is with the eventual objective of generalization in mind that this is done; once methods and criteria which work satisfactorily for particular models have been established, their general utility and applicability can be examined. If for some reason such a generalization proves impossible, then it is evident that alternative techniques must be developed.
2.1 Introduction

The single-parameter exponential distribution has two frequently used 'natural' parameterizations. That is, there are two different physically-meaningful parameters, either of which may be used in any expression of the distribution. The parameterization which a particular experimenter chooses to employ depends on the physical quantity in which he is interested. If, for example, he has observed the times to failure, \( t_1, t_2, \ldots, t_n \) of \( n \) 'identical' components (all acting independently), he might be interested in the mean lifetime of a component and so would probably estimate \( \alpha \) from the likelihood arising from the density

\[
f(t)dt = \frac{1}{\alpha} e^{-t/\alpha} dt
\]  

(2.11)

\( \alpha \) being the mean lifetime. Alternatively, he may wish information about the failure rate, \( \theta \), of the components and might use

\[
f(t)dt = \theta e^{-\theta t} dt
\]  

(2.12)

Obviously, \( \theta = 1/\alpha \) and by the invariance property of
maximum likelihood estimators, \( \hat{\theta} = 1/\hat{\alpha} \) (where \( \hat{\alpha} \) denotes the maximum likelihood estimator of \( \alpha \)) so that use of 2.11 to estimate \( \alpha \) followed by inversion to obtain \( \hat{\theta} \) is equivalent to estimation of \( \theta \) directly from 2.12. Thus, as far as estimation alone is concerned, 2.11 and 2.12 are virtually equivalent.*

Asymptotically the distributions of both \( \hat{\alpha} \) and \( \hat{\theta} \) are normal so that in large samples it would seem reasonable to use the assumptions of asymptotic normality to obtain confidence intervals for whichever of the parameters is the one of interest to the experimenter. But in small samples, that is, in any actual set of observations, is asymptotic normal theory adequate? Are the distributions of \( \hat{\alpha} \) and \( \hat{\theta} \) equally 'close' to the normal? Or is there perhaps another parameterization producing an estimator with a distribution 'more nearly normal' than that of either \( \hat{\alpha} \) or \( \hat{\theta} \)? If so, intervals could be obtained for this parameter using asymptotic theory, and the intervals inverted to get those corresponding to \( \alpha \) or \( \theta \). These intervals would then be more accurate.

In actual fact, the exact distribution of \( \hat{\alpha} \) is known and tabulated, and can be used to set up exact confidence

* 'Estimator' will always refer to the maximum likelihood estimator, unless specified otherwise.
intervals, thereby rendering asymptotic assumptions unnecessary in this case. However, in practice, the normal distribution is often used to obtain intervals for \( \alpha \) and so comparisons of the exact distributions of estimators for members of a specified family of parameter transformations, with respect to their 'closeness' to normality, could be very interesting.

It is proposed, then, in this chapter to compare certain aspects of these distributions for various numbers of observations. The family used will in general be that described in the next section, 2.2, except in section 2.5, where the family will be as specified within that section. The following aspects of the estimators and their distributions are discussed:

(a) asymmetry of the likelihood function;
(b) second and third moments and skewness coefficients;
(c) confidence intervals;
(d) bias;
(e) plots of standardized densities.

In some sections, comparisons with asymptotic distributions will be included.

It should be emphasised that the main interest of this chapter lies not in the results obtained for the particular distribution under consideration (although these
results are of some interest), but in the criteria and techniques developed. In later chapters I shall deal with multi-parameter distributions; for these, all aspects of the original problem can be examined. In the present chapter I am merely attempting to clarify one of the aspects, namely, normality of distribution of estimators, through the use of a simple single-parameter model.
2.5

2.2 Specification of Distributions

The basic distribution in this chapter is taken to be that with density as in 2.11, namely,

\[ f(x)dx = \frac{1}{\alpha} e^{-x/\alpha} dx \]

Given \( n \) independent observations from this distribution, \( x_1, x_2, \ldots, x_n \), the joint distribution of these observations has density

\[ f(x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} dx_j = \left( \frac{1}{\alpha} \right)^n e^{-T/\alpha} \prod_{j=1}^{n} dx_j \]

where \( T = \sum_{j=1}^{n} x_j \). This then gives rise to the log likelihood function

\[ L(\alpha|T) = -n \log \alpha - T/\alpha \] (2.21)

and, from the normal equation,

\[ \hat{\alpha} = \bar{x} = T/n \]

It is well known that \( T \) is a sufficient statistic for \( \alpha \) and has a Gamma distribution with parameters \((n-1, \alpha)\). That is,

\[ f(T)dT = \frac{1}{\Gamma(n)\alpha^n} T^{n-1} e^{-T/\alpha} dT \] (2.22)

If it is noted that \( T/\alpha \) is in fact distributed as \( \frac{1}{2} \chi^2_{2n} \), tables of \( \chi^2 \) can be used to obtain exact confidence limits for \( \alpha \) or any one-one function of \( \alpha \). Also, since the normal density is known to be a good approximation to \( \chi^2_k \) for \( k > 30 \),
asymptotic normal theory is likely to be reasonable for \( n > 15 \). Thus, as far as the exponential distribution is concerned, only values of \( n \) less than fifteen need be considered. In the interest of generality, however, somewhat larger values of \( n \) are also dealt with.

Now, consider the following more general formulation of the exponential density:

\[
\begin{align*}
&f_i(x)dx = \frac{1}{\phi_i^{1/i}} \exp(-x/\phi_i^{1/i}) dx, \quad i \neq 0 \\
f_0(x)dx = \exp(-\phi_0 - e^{-\phi_0 x}) dx,
\end{align*}
\]

where \( \hat{\phi}_i = T^{i/n} \) when \( i \neq 0 \) and \( \hat{\phi}_0 = \log(T/n) \).

When \( i = 1 \), 2.23 is simply the density in the familiar form of 2.11; when \( i = -1 \), it is the density as in 2.12.

This (2.23), then, is the general family of parameterizations of the exponential distribution used throughout the rest of the chapter. When the meaning is obvious the subscript 'i' will not be included.

### 2.2.1 Exact Sampling Distributions

From 2.22 and 2.23,

\[
f_i(T)dT = \frac{1}{\Gamma(n)\phi_i^{n/i}} T^{n-1} \exp(-T/\phi_i^{1/i}) dT, \quad i \neq 0
\]
2.7

and \( f_\circ(T) dT = \frac{1}{\Gamma(n)} T^{n-1} \exp(-n\phi_\circ - Te^{-\phi_\circ}) dT \).

Letting \( Y = T^i \), then \( T = Y^{1/i} \) and \( dT = \frac{1}{i} Y^{(1-i)/i} dY \) and

\[
    f_i(Y) dY = \frac{1}{|i| \Gamma(n)} Y^{(n-i)/i} \exp(-(Y/\phi)^{1/i}) dY, \\
    i \neq 0 .
\]

With \( i = 0 \), \( \phi_\circ = \log \frac{T}{n} \) and so \( T = ne^{\hat{\phi}} \) and \( dT = ne^{\hat{\phi}} d\hat{\phi} \). Then

\[
    f_i(\hat{\phi}_i) d\hat{\phi}_i = \frac{n^n}{|i| \Gamma(n)} \frac{\hat{\phi}_i^{(n-i)/i}}{\phi_i^{n/i}} \exp(-n\hat{\phi}_i^{1/i} / \phi_i^{1/i}) d\hat{\phi}_i, \\
    i \neq 0
\]

\[
    f_\circ(\hat{\phi}_\circ) d\hat{\phi}_\circ = \frac{n^n}{\Gamma(n)} \exp \left[ n(\hat{\phi}_\circ - \phi_\circ) - ne^{(\hat{\phi}_\circ - \phi_\circ)} \right] d\hat{\phi}_\circ .
\]

(2.24)

An interesting point to note is that \( \hat{\phi}_i, i \neq 0 \), is constrained by the physical properties of the model to be non-negative, whereas \( \hat{\phi}_\circ \), the estimator from the logarithmic parameterization, is unconstrained. One might expect, therefore, that the normal distribution would approximate the distribution of \( \hat{\phi}_\circ \) better than that of \( \hat{\phi}_i, i \neq 0 \).
Throughout this chapter, where it is necessary for the purposes of tabulation to assume a value for \( \phi_i \), this value is taken as 1.0 when \( i \neq 0 \) and as 0 when \( i = 0 \), and all comparisons are made on this basis. Since any comparisons made are scale-free, use of these 'simple' values is justified.

2.2.2 Asymptotic Distributions

For any value of \( i \), asymptotically

\[
\hat{\phi}_i \sim N \left( \phi_i, \{I(\phi_i)\}^{-1} \right),
\]

where \( I(\phi_i) = -E(\partial^2 L/\partial \phi_i^2) \) is the information.

From 2.22, for \( i \neq 0 \),

\[
L(\phi_i | T) = -n \log \phi_i - \frac{T}{i} \phi_i^{1/i}
\] (2.25)

\[
\frac{\partial L}{\partial \phi_i} = -n + \frac{T}{i^{(i+1)/i}} \phi_i^{(i+1)/i}
\] (2.26)

\[
\frac{\partial^2 L}{\partial \phi_i^2} = \frac{n}{i \phi_i^2} - \frac{(i+1)T}{i^2 \phi_i^{(2i+1)/i}}
\] (2.27)
Since $T$ has a Gamma distribution, $E(T) = n \phi_i^{1/i}$ and so

$$I(\phi_i) = \frac{n}{i^2 \phi_i^2}.$$  \hspace{1cm} \text{(2.28)}

Therefore, asymptotically

$$\hat{\phi}_i \sim N(\phi_i, \frac{i^2 \phi_i^2}{n}).$$

For $i = 0$,

$$L(\phi_0 \mid T) = -n \phi_0 - T \exp(-\phi_0)$$

$$\frac{\partial L}{\partial \phi_0} = -n + T \exp(-\phi_0)$$

$$\frac{\partial^2 L}{\partial^2 \phi_0} = -T \exp(-\phi_0)$$

$$E(T) = n \exp(\phi_0).$$

Therefore, $I(\phi_0) = n$ and, asymptotically,

$$\hat{\phi}_0 \sim N(\phi_0, 1/n).$$
2.3 Asymmetry of the Likelihood Function

Sprott and Kalbfleisch (1969) consider the family of parameterizations 2.23 and choose that value of \( i \) (\( i \neq 0 \)) which minimizes the asymmetry of the likelihood function. They do this by equating to zero the cubic term in the Taylor series expansion of the log likelihood about the maximum. This series expansion gives

\[
L(\phi | T) = L(\hat{\phi}) + (\phi - \hat{\phi}) \frac{\partial L}{\partial \phi} + \frac{(\phi - \hat{\phi})^2}{2!} \frac{\partial^2 L}{\partial \phi^2}
\]

\[
+ \frac{(\phi - \hat{\phi})^3}{3!} \frac{\partial^3 L}{\partial \phi^3} + ...
\]

where \( \frac{\partial^k L}{\partial \phi^k} \) is evaluated at \( \phi = \hat{\phi} \), \( k = 1, 2, ... \).

Therefore

\[
\log R(\phi | T) = \frac{(\phi - \hat{\phi})^2}{2!} \frac{\partial^2 L}{\partial \phi^2} + \frac{(\phi - \hat{\phi})^3}{3!} \frac{\partial^3 L}{\partial \phi^3} + ...
\]

where \( R(\phi | T) \) is the likelihood of \( \phi \) relative to the value of the likelihood at \( \hat{\phi} \).

Differentiation of 2.27 gives
\[
\frac{\partial^3 L}{\partial \phi_i^3} = \frac{-2n}{i \phi_i^3} + \frac{(i+1)(2i+1)}{i^3 \phi_i} T^{(3i+1)/i}
\]

And so

\[
\frac{\partial^3 \hat{L}}{\partial \hat{\phi}_i^3} = \frac{n^{3i+1}}{T^{3i}} = \frac{3i+1}{i^3}
\]

Equating 2.31 to zero and solving for \(i\) gives \(i = -1/3\).

With \(i = -1/3\), \(\log R(\phi_i \mid T)\) is approximately quadratic in the region of \(\hat{\phi}\) and so, in this region, resembles the equivalent function produced by a normal density. It might be expected that intervals for \(\phi_{-1/3}\) obtained by applying the assumptions of asymptotic normality to \(\hat{\phi}_{-1/3}\) would be more accurate than those obtained by applying these same assumptions to \(\hat{\phi}_1\). However, although it is true that a normal distribution always yields a quadratic log relative likelihood function, the reverse is not necessarily true. Sprott and Kalbfleisch point this out in their paper and suggest that this parameter transformation should only be used if it seems to apply consistently to several sets of data arising from similar situations.
There is another reason for thinking carefully about using this transformation: although it may be reasonably applied in the region of the maximum, the 95% and 99% confidence limits are going to be a long way from this maximum, in regions where the transformed likelihood may not be very well approximated by a quadratic.

This method of choosing a transformation is not generally satisfactory. As is shown later from information about the exact distribution of \( \phi_{-1/3} \), \( \phi_{-1/3} \) is actually a rather poor parameterization overall.
2.4 Moments and Coefficients of Skewness

From the exact distribution of $\hat{\phi}_i$, 2.24, second and third moments and coefficients of skewness can be obtained. The latter are particularly interesting as a measure of departure from normality.

For $i \neq 0$, moments of $\hat{\phi}_i$ can be calculated as follows:

$$E(Y = T^i) = \int_0^\infty T^i f(T) \, dT$$

$$= \int_0^\infty T^{n+i-1} \phi_i^{1/i} \exp(-T/\phi_i^{1/i}) \, dT$$

$$= \phi_i \int_0^\infty T^{n+i-1} \phi_i^{1/i} \exp(-T/\phi_i^{1/i}) \, d(T/\phi_i^{1/i})$$

$$= \phi_i \frac{\Gamma(n+i)}{\Gamma(n)}$$

$$\text{Var}(Y) = \mu_2(Y) = E(T^{2i}) - E^2(T^i)$$

$$= \phi^2 \left[ \frac{\Gamma(n+2i)}{\Gamma(n)} - \frac{\Gamma^2(n+i)}{\Gamma^2(n)} \right]$$
\[\mu_3(Y) = E(Y^3) - 3E(Y^2)E(Y) + 2E^3(Y)\]

\[= \phi^3 \left[ \frac{\Gamma(n+3i) - 3\Gamma(n+2i)\Gamma(n+i)}{\Gamma(n)} \right. + \frac{2\Gamma^3(n+i)}{\Gamma^3(n)} \right].\]

In general, \(\mu_r(\hat{\phi}_i) = (1/n)^r_i \mu_r(T^n), r = 1, 2, \ldots\).

Of course, the coefficient of skewness of \(\hat{\phi}_i\) is

\[\gamma_1(\hat{\phi}_i) = \frac{\mu_3(\hat{\phi}_i)}{\left[\mu_2(\hat{\phi}_i)\right]^{3/2}}\]

and is independent of the true value of \(\phi_i\).

For \(i = 0\), letting \(\rho = e^{\phi_0}\),

\[E(\hat{\phi}_0) = E(\log T) - \log n\]

\[= E(\log T/\rho) - \log n + \log \rho\]

\[= \frac{1}{\Gamma(n)} \int_0^{\infty} \left(\frac{T}{\rho}\right)^{n-1} e^{-T/\rho} \log(T/\rho) \, d(T/\rho) + \log (\rho/n)\]
\[ 2.15 \]

\[
\begin{align*}
\mu_2 (\hat{\phi}_0) &= \mu_2 \left( \log \left( \frac{T}{\rho} \right) \right) \\
&= \frac{d^2}{dn^2} \log \Gamma(n) \\
\mu_3 (\hat{\phi}_0) &= \frac{d^3}{dn^3} \log \Gamma(n) 
\end{align*}
\]

In general, it can be easily shown that, for \( r > 1 \), the \( r \) th. cumulant is

\[
\kappa_r (\hat{\phi}_0) = \frac{d^r}{dn^r} \log \Gamma(n)
\]

This is the well-known polygamma function, \( \psi^{(r-1)}(n) \), defined as

\[
\psi^{(r-1)}(n) = \frac{d^r}{dn^r} \log \Gamma(n) , \quad r > 1 ,
\]

\[
\psi(n) = \frac{d}{dn} \log \Gamma(n) ,
\]
which has been tabulated. In terms of polygamma functions
the moments of $\hat{\phi}_0$ are

$$E(\hat{\phi}_0) = \psi(n) - \log n + \phi_0$$

$$\mu_2(\hat{\phi}_0) = \psi^{(2)}(n)$$

$$\mu_3(\hat{\phi}_0) = \psi^{(3)}(n)$$

and

$$\gamma_1(\hat{\phi}_0) = \frac{\psi^{(2)}(n)}{\left[\psi^{(1)}(n)\right]^{3/2}}$$

Using tabulated \(^1\) values of $\psi^{(1)}(2)$ and $\psi^{(2)}(2)$ and
the recurrence relation

$$\psi^{(r)}(n+1) = \psi^{(r)}(n) + (-1)^r r! \ n^{-r-1}$$

$\mu_2$, $\mu_3$, and $\gamma_1$ can be obtained for several values of $n$.

---

1. From Handbook of Mathematical Functions, Section 6,
   Gamma Function and Related Functions.

2. From asymptotic expansions of $\psi^{(1)}(n)$ and $\psi^{(2)}(n)$ it
   can be shown that for large $n$,

   $$\gamma_1^2 = 1/n + O(n^{-2})$$

   See Johnson (1949).
A good value of $i$ would be one which minimised $\gamma_1(\hat{\phi}_i)$ for any value of $n$. Since $\gamma_1$ is expressed in terms of gamma and polygamma functions there will be no simple algebraic solution to $\gamma_1(\hat{\phi}_i) = 0$, although it might be possible to minimise $\gamma_1(\hat{\phi}_i)$ by some iterative technique. Alternatively, one can tabulate $\gamma_1$ for various values of $i$ and $n$; it is this procedure which I shall employ.

The IBM 360/50 computer was programmed to calculate $\mu_2$, $\mu_3$, and $\gamma_1$ for $i = 2.0$, ±1.0, ±0.5, ±0.33, ±0.25, 0.0 and $n = 5, 10, 15, 20, 25, 30$ with $\phi_1 = 1.0$ ($i \neq 0$) and $\phi_0 = 0.0$. The results appear in Tables 2.41 and 2.42.

2.4.1 Conclusions from Tabulations

The entries in Table 2.41 can be compared to see if there is any suggestion of a value of $i$ producing a distribution which is symmetric, or nearly so.

Of the more usual parameterizations ($i = \pm 1$), $\hat{\phi}_1$ has a much less skewed distribution than $\hat{\phi}_{-1}$. This implies that to obtain intervals for $\hat{\phi}_{-1}$, it would be better to assume asymptotic normality of $\hat{\phi}_1$ to get limits for $\hat{\phi}_1$ and invert these to obtain the desired intervals for $\hat{\phi}_{-1}$, rather than to assume normality of $\hat{\phi}_{-1}$.
TABLE 2.41

COEFFICIENTS OF SKEWNESS OF \( \phi_i \)

<table>
<thead>
<tr>
<th>( i )</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>2.479</td>
<td>1.675</td>
<td>1.344</td>
<td>1.153</td>
<td>1.025</td>
<td>0.932</td>
</tr>
<tr>
<td>1.0</td>
<td>0.894</td>
<td>0.632</td>
<td>0.516</td>
<td>0.447</td>
<td>0.400</td>
<td>0.365</td>
</tr>
<tr>
<td>0.5</td>
<td>0.237</td>
<td>0.163</td>
<td>0.132</td>
<td>0.114</td>
<td>0.101</td>
<td>0.092</td>
</tr>
<tr>
<td>0.33</td>
<td>0.010</td>
<td>0.002</td>
<td>0.000</td>
<td>-0.0006</td>
<td>-0.0008</td>
<td>-0.0009</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.101</td>
<td>-0.075</td>
<td>-0.063</td>
<td>-0.055</td>
<td>-0.049</td>
<td>-0.045</td>
</tr>
<tr>
<td>0.0*</td>
<td>-0.469</td>
<td>-0.324</td>
<td>-0.262</td>
<td>-0.226</td>
<td>-0.202</td>
<td>-0.184</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.896</td>
<td>0.590</td>
<td>0.471</td>
<td>0.404</td>
<td>0.359</td>
<td>0.326</td>
</tr>
<tr>
<td>-0.33</td>
<td>1.052</td>
<td>0.680</td>
<td>0.541</td>
<td>0.462</td>
<td>0.410</td>
<td>0.372</td>
</tr>
<tr>
<td>-0.5</td>
<td>1.434</td>
<td>0.884</td>
<td>0.694</td>
<td>0.590</td>
<td>0.522</td>
<td>0.473</td>
</tr>
<tr>
<td>-1.0</td>
<td>3.464</td>
<td>1.616</td>
<td>1.202</td>
<td>0.998</td>
<td>0.872</td>
<td>0.784</td>
</tr>
</tbody>
</table>

* \( i = 0 \) represents the logarithmic parameterization.

\( \phi_i \) is the parameter in the general exponential density

\[
f(x)dx = \phi_i^{-1/i} \exp(-x/\phi_i^{1/i}) \, dx.\]
TABLE 2.42

VARIANCE($\hat{\phi}_1$) ($\phi_1 = 1$)

<table>
<thead>
<tr>
<th>n</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>1.248 (0.800)</td>
<td>0.506 (0.400)</td>
<td>0.313 (0.267)</td>
<td>0.226 (0.200)</td>
<td>0.176 (0.160)</td>
<td>0.145 (0.133)</td>
</tr>
<tr>
<td>1.0</td>
<td>0.200 (0.200)</td>
<td>0.100 (0.100)</td>
<td>0.067 (0.067)</td>
<td>0.050 (0.050)</td>
<td>0.040 (0.040)</td>
<td>0.033 (0.033)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.049 (0.050)</td>
<td>0.025 (0.025)</td>
<td>0.016 (0.017)</td>
<td>0.012 (0.012)</td>
<td>0.010 (0.010)</td>
<td>0.008 (0.008)</td>
</tr>
<tr>
<td>0.33</td>
<td>0.022 (0.022)</td>
<td>0.011 (0.011)</td>
<td>0.007 (0.007)</td>
<td>0.005 (0.006)</td>
<td>0.004 (0.004)</td>
<td>0.004 (0.004)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.013 (0.012)</td>
<td>0.006 (0.006)</td>
<td>0.004 (0.004)</td>
<td>0.003 (0.003)</td>
<td>0.002 (0.002)</td>
<td>0.002 (0.002)</td>
</tr>
<tr>
<td>0.0*</td>
<td>0.221 (0.200)</td>
<td>0.105 (0.100)</td>
<td>0.069 (0.067)</td>
<td>0.051 (0.050)</td>
<td>0.041 (0.040)</td>
<td>0.034 (0.033)</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.016 (0.012)</td>
<td>0.007 (0.006)</td>
<td>0.004 (0.004)</td>
<td>0.003 (0.003)</td>
<td>0.003 (0.002)</td>
<td>0.002 (0.002)</td>
</tr>
<tr>
<td>-0.33</td>
<td>0.029 (0.022)</td>
<td>0.012 (0.011)</td>
<td>0.008 (0.007)</td>
<td>0.006 (0.006)</td>
<td>0.005 (0.004)</td>
<td>0.004 (0.004)</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.076 (0.050)</td>
<td>0.030 (0.025)</td>
<td>0.019 (0.017)</td>
<td>0.014 (0.012)</td>
<td>0.011 (0.010)</td>
<td>0.009 (0.008)</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.521 (0.200)</td>
<td>0.154 (0.100)</td>
<td>0.088 (0.067)</td>
<td>0.062 (0.050)</td>
<td>0.047 (0.040)</td>
<td>0.038 (0.033)</td>
</tr>
</tbody>
</table>

* i = 0 represents the logarithmic parameterization.

Numbers in brackets are asymptotic variances.

$\phi_i$ is the parameter in the general exponential density

$$f(x)dx = \phi_i^{-1/i} \exp(-x/\phi_i^{1/i}) \ dx.$$
Amongst the other i-values, \( i = 0.33 \) leads to the smallest skewness coefficients for all tabulated \( n \)-values (and is even superior in this respect to \( i = 1 \) ). The peculiar behaviour of the skewness coefficient for \( i = 0.33 \) for \( n > 15 \) is perhaps caused by underflows or loss of accuracy within the computer because the values are so close to zero. (This occurs despite the use of double precision.) The tabulated quantities seem to indicate that there is some value \( i' \), where \( 0.25 < i' < 0.33 \), such that the skewness is in fact zero, and that perhaps this holds true for all values of \( n \).

More generally, each positive value of \( i \), \( 0 < i \leq 1 \), is better than its negative counterpart and in fact, \( i = 0.0, 0.25, 0.33, 0.50 \) are all better than any of the tabulated negative \( i \)-values. (These \( i \)-values are 'better' in the sense that the corresponding parameter estimators have less skewed distributions.) It seems strange that the application of Sprott's criterion of minimisation of asymmetry of the likelihood function should lead to a parameterization which compares so unfavourably with other parameterizations, at least in respect of the measure under discussion.

\(^{1}\) In fact I calculated the skewness coefficient with \( n = 5 \), \( i = 0.3 \) and this was \(-0.0374\). This suggests that minimum skewness occurs for \( i' \), \( 0.3 < i' < 0.33 \), at least for \( n = 5 \).
Although the distribution of $\hat{\phi}_0$ is defined for all real values of $\hat{\phi}_0$ (a fact which makes a normal approximation seem more reasonable), it is quite skewed, more so than distributions resulting from several other parameterizations.

Table 2.42 includes asymptotic as well as exact variances and it is interesting to briefly compare these for some $i$. (As usual, asymptotic variances were calculated with $\phi_i = 1.0$, for $i \neq 0$, and $\phi_0 = 0.0$.) An examination of the table shows that for $i < 0$, the asymptotic variance is less than the true variance for small $n$; in particular, when $i = -1$ the asymptotic variance is much smaller than the exact one. For values of $i$ between 0 and 1, the asymptotic is slightly larger than the exact variance, but even for $n$ as small as 5 the two variances agree to two decimal places for $i$-values in this range. When $i = 1$, exact and asymptotic variances are identical. For the logarithmic parameterization the two agree reasonably well.

Over all, an examination of Tables 2.41 and 2.42 suggests that the application of asymptotic theory to $\hat{\phi}_1$ for small $n$ could produce extremely misleading results, either in the form of incorrect confidence intervals or else as meaningless statements of means and standard
deviations. If one wished to use one of $\phi_1$ or $\phi_{-1}$, this section surely indicates that one would do much better by applying asymptotic theory to $\phi_i$. The best parameterization of those tabulated, based on the criteria of this section, are those with $i$ such that $0 \leq i \leq 1$; and of these, $i = 0.33$ seems to be the best.
2.5 Confidence Intervals

In this section, a rather different approach to the reparameterization problem is taken, in that another family of transformations is considered.

A confidence interval for a parameter is often the aim of an experiment with an assumed underlying model and/or distribution. These intervals are usually based, if at all possible, on the standard normal distribution, since its percentage points have been tabulated, even if the estimator's true distribution is known to be non-normal. The central limit theorem and the known asymptotic properties of maximum likelihood estimators are usually considered sufficient justification for the use of the asymptotic normal distribution. For small numbers of observations, however, asymptotic assumptions are often not reasonable. This is one of the main reasons for searching for a 'normalizing' parameter transformation such as is being considered in this chapter. Bartlett (1953) suggests a polynomial transformation of the score function to make the distribution of this function agree, up to the third or fourth moments, with the standard normal. He then uses the transformed score function, assuming its normality, to obtain confidence intervals for the parameter of interest.
2.24

It is his family of transformations which is discussed in this section.

Since the exact distribution of $\hat{\alpha}/\alpha = \phi_1/\phi_1$ is known and tabulated, intervals obtained from it can be compared with those from the asymptotic distribution of $\hat{\alpha}/\alpha$ and those from Bartlett's transformed score function. In this chapter these comparisons are made for the $\alpha$-parameterization only since intervals for any other parameter can be obtained from those for $\alpha$.

Unfortunately, Bartlett's method is sufficiently complex that it is not of much use in dealing with complicated score functions. In the present case, the score function is quite simple and Bartlett's technique can be applied without much difficulty.

2.5.1 Exact Confidence Limits

Since

$$f(T)dT = \frac{1}{\Gamma(n)} \frac{T^{n-1} e^{-T/\alpha}}{\alpha^n} dT,$$

then

$$f(T/\alpha)d(T/\alpha) = \frac{1}{\Gamma(n)} \left(\frac{T}{\alpha}\right)^{n-1} e^{-T/\alpha} d(T/\alpha),$$
which is one half the density of a $\chi^2$ variate with $2n$ degrees of freedom. Hence

$$\frac{T}{\alpha} = n \frac{\phi_i^{1/i}}{\phi_i^{1/i}} \sim \frac{1}{2} \chi^2_{2n} ,$$

$i \neq 0$.

From tables of $\chi^2$ exact confidence limits can be calculated for $\phi_i$ for any $i \neq 0$ at the desired confidence level $P$.

Let $\mu_1$ and $\mu_2$ be the appropriate percentage points on the $\chi^2$ curve (considering equal tail areas). Then

$$\Pr \left( \frac{\mu_1}{2} \leq \frac{\hat{a}}{\alpha} \leq \frac{\mu_2}{2} \right) = 1 - P$$

and confidence limits for $\alpha$ are $(2n\hat{\alpha}/\mu_2, 2n\hat{\alpha}/\mu_1)$.

2.5.2 Asymptotic Confidence Limits

Large sample theory leads to the assumption that asymptotically,

$$\hat{\alpha} \sim N ( \alpha, \alpha^2/n ) ,$$

and so

$$\frac{\hat{\alpha} - \alpha}{\alpha / \sqrt{n}} \sim N (0,1) .$$

(2.51)
From tables of standard normal deviates, the P-level percentage points, $\pm \mu$, can be obtained. Then confidence limits for $\alpha$ can be found by solving

$$\frac{\hat{\alpha} - \alpha}{\alpha / \sqrt{n}} = \pm \mu$$

(2.52)

These limits are

$$\hat{\alpha} \pm \mu / \sqrt{n}$$

An interesting fact to note here is that for small values of $n$, $\mu / \sqrt{n}$ can be greater than one, and a negative confidence limit for $\alpha$ obtains, although $\alpha$ is constrained by definition to be non-negative. This is a point against the use of a normal approximation to a distribution defined on an interval with a finite limit.

2.5.3 Limits using Bartlett's Transformed Score Function

One reason why Bartlett transforms the score function is that its moments are always known exactly. To

---

1 Bartlett notes in his paper that his method, as far as it goes, is equivalent to that of Cornish and Fisher (1937). If one wished to adjust for agreement with cumulants of the normal of higher order than the fourth, appropriate terms could be obtained from this 1937 paper.
get rid of skewness in the distribution of the score function, he therefore proposes a transformation $T(\alpha)$ where

$$T(\alpha) = \frac{3L}{\alpha} - \frac{\kappa_3}{6I^2} \left\{ \left( \frac{\partial L}{\partial \alpha} \right)^2 - I \right\},$$

with

$$\kappa_3 \left( \frac{\partial L}{\partial \alpha} \right) = E \left\{ \left( \frac{\partial L}{\partial \alpha} \right)^3 \right\} = 3 \frac{\partial I}{\partial \alpha} + 2 E \left( \frac{\partial^3 L}{\partial \alpha^3} \right),$$

and

$$I = -E \left( \frac{\partial^2 L}{\partial \alpha^2} \right).$$

It is known that $E \left( \frac{\partial L}{\partial \alpha} \right) = 0$ and so $E \{ T(\alpha) \} = 0$, $\text{Var} \{ T(\alpha) \} = I$ (neglecting terms in $\kappa_3^2$), and $\text{cov} \{ T(\alpha) \} = 0$ (neglecting terms in $\kappa_3^2$ and $\kappa_4$).

Unfortunately, even for relatively simple score functions, this method is not as straightforward in application as it appears. To calculate Bartlett's confidence limits for $\alpha$, let $i = 1$ in 2.25 - 2.27. This gives

$$L \left( \alpha/T \right) = -n \log \alpha - T/\alpha,$$

$$\frac{\partial L}{\partial \alpha} = -n/\alpha + T/\alpha^2,$$

$$\frac{\partial^2 L}{\partial \alpha^2} = n/\alpha^2 - 2T/\alpha^3.$$
\[ I(\alpha) = \frac{n}{\alpha^2} \]
\[ \frac{\partial I}{\partial \alpha} = -2 \frac{n}{\alpha^3} \]
\[ \kappa_1(\alpha) = 2 \frac{n}{\alpha^3} \]

Therefore,
\[ T(\alpha) = \frac{-n}{3\alpha} \left\{ (\hat{\alpha}/\alpha - 1)^2 - 3(\hat{\alpha}/\alpha - 1) - 1/n \right\} \] (2.53)

Assuming, then, that \( T(\alpha) \sim N(0, I) \), confidence limits for \( \alpha \) should be obtained by solving
\[ T(\alpha) = \pm \mu \sqrt{I} \] (2.54)

where \( \pm \mu \) are the P-level percentage points of the normal. Equation 2.54, however, is quadratic in \( \alpha \) for each sign of \( \mu \), yielding two values of \( \alpha \) for each end of the interval. Bartlett suggests reducing 2.54 by replacing \( (\hat{\alpha}/\alpha - 1)^2 \) only by the value which it would take if asymptotic normal theory were applied directly to \( \hat{\alpha} \), thus maintaining the same order of approximation. From 2.52,
\[ \hat{\alpha}/\alpha - 1 = \pm \mu / \sqrt{n}. \]

Substituting in this way alters 2.54 to
\[ \frac{n}{3a} \left\{ \mu^2/n - 3(\hat{\alpha}/\alpha - 1) - 1/n \right\} = \pm \mu/\alpha. \]

Solution of this gives

\[ \alpha = n\hat{\alpha} / (\mu^2/3 \pm \mu/n + n - 1/3) \]

as the P-level confidence limits for \( \alpha \).

2.5.4 Comparison of the Three Confidence Intervals

95\% and 99\% confidence intervals for \( \alpha \) have been calculated for \( n = 5, 10, 20, 30 \), with \( \hat{\alpha} = 1.0 \), using each of the above methods. The results appear in Table 2.5.

For very small \( n \), that is \( n = 5 \) or 10, the asymptotic normal approximation is obviously not very satisfactory. The asymptotic intervals are all too far to the right, that is both ends of each interval are larger than the corresponding end points of the exact interval. As \( n \) increases, however, the asymptotic intervals agree more closely with the exact ones. With small \( n \), then, use of asymptotic theory can produce very misleading and sometimes (as in the 99\% interval with \( n = 5 \)) completely uninformative intervals. Intervals computed using Bartlett's method compare very favourably with the exact, even for \( n \).
<table>
<thead>
<tr>
<th>n</th>
<th>EXACT</th>
<th></th>
<th>ASYMPTOTIC</th>
<th></th>
<th>BARTLETT'S*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Lower</td>
<td>Upper</td>
<td>Lower</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>0.995</td>
<td>0.975</td>
<td>0.975</td>
<td>0.995</td>
<td>0.975</td>
</tr>
<tr>
<td>5</td>
<td>0.397</td>
<td>0.488</td>
<td>3.077</td>
<td>4.630</td>
<td>0.465</td>
</tr>
<tr>
<td>10</td>
<td>0.500</td>
<td>0.585</td>
<td>2.086</td>
<td>2.692</td>
<td>0.551</td>
</tr>
<tr>
<td>20</td>
<td>0.608</td>
<td>0.680</td>
<td>1.666</td>
<td>2.009</td>
<td>0.635</td>
</tr>
<tr>
<td>30</td>
<td>0.659</td>
<td>0.724</td>
<td>1.498</td>
<td>1.730</td>
<td>0.680</td>
</tr>
</tbody>
</table>

* Adjusted to the third moment only.

** Solution of the usual equation involving the asymptotic variance to obtain the 99% confidence limits gives

\[
\Pr(\alpha \geq 0.464 \text{ and } \alpha \leq -6.494) = 0.99.
\]

Since \( \alpha > 0 \) by definition, the upper 99% limit is effectively infinity.

\( \alpha \) is the parameter in the exponential density \( f(x) dx = \frac{1}{\alpha} \exp(-x/\alpha) \ dx \).
as small as 5.

Interesting comparisons can be made of the exact intervals for \( \alpha \) and those obtained by application of asymptotic theory to the parameterizations suggested in sections 2.3 and 2.4.

From section 2.3 one might assume that

\[
\hat{\phi}_{-1/3} \sim N \left( \phi_{-1/3}, \left( \frac{1}{\frac{1}{3} \phi_{-1/3}} \right)^2 \right)
\]

or equivalently,

\[
\frac{3 \sqrt{n} \left( \hat{\phi}_{-1/3} - \phi_{-1/3} \right)}{\phi_{-1/3}} \sim N \left( 0, 1 \right).
\]

Taking \( \hat{\phi}_{-1/3} = 1.0 \), confidence limits for \( \phi_{-1/3} \) are obtained by solving

\[
3 \sqrt{n} \left( 1 - \phi_{-1/3} \right) = \pm \mu
\]

where as previously, \( \pm \mu \) are the P-level percentage points of the normal distribution. If these limits are \( (l_1, l_2) \) then, since \( \alpha = \phi_{-1/3}^{-3} \), the corresponding limits for \( \alpha \) are \( (l_2^{-3}, l_1^{-3}) \).

From section 2.4 one might choose to assume, again
with $\hat{\phi}_{1/3} = 1.0$, that

$$\frac{3 \sqrt{n \left(1 - \hat{\phi}_{1/3}\right)}}{\hat{\phi}_{1/3}} \sim N(0,1)$$

and obtain limits for $\phi_{1/3}$ identical to those for $\phi_{-1/3}$ above. But

$$\alpha = \phi_{1/3}^3,$$

and so the corresponding limits for $\alpha$ are $(\ell_1^3, \ell_2^3)$.

95% and 99% confidence limits for $\alpha$ are calculated in the above two ways with $n = 5$ and $n = 10$ and are presented in the table below. To facilitate comparisons, the corresponding exact limits for $\alpha$ are also included.

<table>
<thead>
<tr>
<th>Source</th>
<th>$n = 5$</th>
<th>$n = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>$i = -\frac{1}{3}$</td>
<td>0.355, 2.158</td>
<td>0.234, 2.651</td>
</tr>
<tr>
<td>Exact</td>
<td>0.488, 3.077</td>
<td>0.397, 4.630</td>
</tr>
<tr>
<td>$i = \frac{1}{3}$</td>
<td>0.463, 2.820</td>
<td>0.377, 4.278</td>
</tr>
</tbody>
</table>

* If a value of $i$ is given it means that the intervals for $\alpha$ which follow were computed by assuming that $\phi_i$ was asymptotically normally distributed.
From the above, it is obvious that asymptotic theory applied to $\hat{\phi}_{1/3}$ produces much 'better' (that is, closer to the exact) intervals for $\alpha$ than $\hat{\phi}_{-1/3}$, for small $n$. This suggests that the distribution of $\hat{\phi}_{-1/3}$ is not very close to the normal in the tail regions even if it is approximately normal in the region of its mean.
2.34

2.6 Bias

Although maximum likelihood estimators are asymptotically unbiased, they are not in general exactly unbiased, and for small $n$ the bias can be substantial. The size of the bias relative to the true parameter value will vary between parameterizations and so is a factor which should be considered in this problem. If a particular parameterization produces estimators which have a large systematic bias in small samples, these estimators will not be very informative.

In general for $i \neq 0$,

$$\hat{\phi}_i = \left( \frac{T}{n} \right)^i$$

$$E(\hat{\phi}_i) = \frac{\phi_i \Gamma(n+i)}{n^i \Gamma(n)}$$

Therefore the bias of $\hat{\phi}_i$ is

$$E(\hat{\phi}_i) - \phi_i = \phi_i \left[ \frac{\Gamma(n+i)}{n^i \Gamma(n)} - 1.0 \right]$$

and so the bias relative to the true value $\phi_i$ is

$$b(\hat{\phi}_i) = \frac{\Gamma(n+i)}{n^i \Gamma(n)} - 1.0$$

For $i = 0$, 

\[
E(\hat{\phi}_0) = \phi_0 + \psi(n) - \log n
\]

and so
\[
b(\hat{\phi}_0) = \psi(n) - \log n,
\]

independently of the true value \(\phi_0\). From tables of \(\psi(n)\), \(b(\hat{\phi})\) can be calculated for various values of \(n\). Biases for several values of \(i\) and \(n\) are presented in Table 2.6.

It is apparent from Table 2.6 that \(i = 1\) is the best \(i\)-value as far as size of bias is concerned, since \(\hat{\alpha}\) is unbiased. For all tabulated \(n\), each value of \(i\) between 0 and 1 gives smaller absolute bias than its negative counterpart; \(i = -1\) shows up as the worst tabulated value again. The logarithmic parameterization, although better than \(\phi_{-1}\), is not very good compared to some of the others.

---

1 Handbook of Mathematical Functions, Section 6.
**TABLE 2.6**

BIAS OF $\hat{\phi}_i$ ($\phi_i = 1.0$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.200</td>
<td>0.100</td>
<td>0.0667</td>
<td>0.0500</td>
<td>0.0400</td>
<td>0.0333</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.0246</td>
<td>-0.0124</td>
<td>-0.0083</td>
<td>-0.0062</td>
<td>-0.0050</td>
<td>-0.0042</td>
</tr>
<tr>
<td>0.33</td>
<td>-0.0221</td>
<td>-0.0110</td>
<td>-0.0074</td>
<td>-0.0055</td>
<td>-0.0044</td>
<td>-0.0037</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.0189</td>
<td>-0.0094</td>
<td>-0.0063</td>
<td>-0.0047</td>
<td>-0.0038</td>
<td>-0.0031</td>
</tr>
<tr>
<td>0.0*</td>
<td>-0.1033</td>
<td>-0.0508</td>
<td>-0.0337</td>
<td>-0.0252</td>
<td>-0.0201</td>
<td>-0.0168</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.0334</td>
<td>0.0162</td>
<td>0.0106</td>
<td>0.0079</td>
<td>0.0063</td>
<td>0.0053</td>
</tr>
<tr>
<td>-0.33</td>
<td>0.0476</td>
<td>0.0228</td>
<td>0.0150</td>
<td>0.0112</td>
<td>0.0089</td>
<td>0.0074</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.0837</td>
<td>0.0396</td>
<td>0.0259</td>
<td>0.0192</td>
<td>0.0153</td>
<td>0.0127</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.2500</td>
<td>0.1111</td>
<td>0.0714</td>
<td>0.0526</td>
<td>0.0417</td>
<td>0.0345</td>
</tr>
</tbody>
</table>

* $i = 0$ represents the logarithmic parameterization.

$\phi_i$ is the parameter in the general exponential density

$$f(x)dx = \phi_i^{-1/i} \exp(-x/\phi_i^{1/i}) \, dx.$$
2.7 Plots of the Distributions of Estimators

An obvious way of comparing the estimator distributions with the normal is by eye. If $\hat{\phi}_i$ is transformed to a variate $\hat{\phi}'_i$ with zero mean and unit variance, a direct comparison between the density of this new variate and the standard normal density is possible, since the distributions have the same first two moments. This comparison will give some indication, although subjective, of how far from normality the estimator distributions are for various values of $i$, where the deviations from normality occur, and which value of $i$ leads to the most normal-looking distribution. The comparison need only be effected for small $n$ ($n \leq 15$) since it is known that $\chi^2_{2n}$, the distribution of $2\hat{\alpha}$ (when $\alpha = 1$), is well approximated by the normal for $n > 15$. In fact, curves have only been plotted for $n = 5$ because of the amount of work involved in drawing the graphs and because I would expect to obtain the same relative results for $n = 10$ and $n = 15$.

2.7.1 Standardization of the Estimators

The new variate, standardized with respect to the exact moments, is
\[
\hat{\phi}_i' = \frac{\hat{\phi}_i - E(\hat{\phi}_i)}{\sqrt{\text{var}(\hat{\phi}_i)}}
\]

Let \( E_i \) be the expectation of \( \hat{\phi}_i \), \( b_i \) its bias, \( V_i \) its variance, and \( f_i \) its density; let \( f_i' \) represent the density of \( \hat{\phi}_i' \). Then from sections 2.4 and 2.6, for \( i \neq 0 \) and \( \phi_i = 1.0 \), it is known that

\[
E_i = b_i + 1.0
\]

\[
V_i = \frac{1}{n^2i} \left[ \frac{\Gamma(n+2i)}{\Gamma(n)} - \frac{\Gamma^2(n+i)}{\Gamma^2(n)} \right]
\]

and so

\[
\hat{\phi}_i' = \frac{\hat{\phi}_i - b_i - 1.0}{\sqrt{V_i}}
\]

\[
d\phi_i' = \sqrt{V_i} \, d\hat{\phi}_i'
\]

and

\[
f_i' = \sqrt{V_i} \, f_i
\]

When \( i = 0 \) and \( \phi_0 = 0.0 \), from 2.4 and 2.6,

\[
E_0 = b_0
\]

\[
V_0 = \psi(1)(n)
\]

and, as above,

\[
f_0' = \sqrt{V_0} \, f_0
\]
The tabulated values for bias and variance (Tables 2.42 and 2.6) along with equations 2.24 (specification of $f_1$) can be used to construct the density functions $f_1$ for various values of $i$.

With $n = 5$, $\phi_0 = 0.0$, and $\phi_1 = 1.0$ ($i \neq 0$), the probability densities $f_1'$ have been plotted for $i = 1.0, 0.33, 0.0$ in Figure 2.71(a) and for $i = -1.0, -0.33$ in Figure 2.71(b). Figure 2.71(b) includes a standard normal curve to facilitate comparisons; in Figure 2.71(a) $f_{'.33}$ serves this purpose as it is so close to the $N(0,1)$ curve as to be indistinguishable from it on a graph of this scale.

Since the first two moments of $f_1'$ are identical to those of the standard normal, it is the effect of differences between the third (and higher order) moments of the exact distributions and those of $N(0,1)$ that show up in Figures 2.71(a) and (b). It is therefore not surprising that the following conclusions reached from these graphs are basically the same as those derived from the examination of the coefficients of skewness (section 2.4.1).

Generally the densities with $i \geq 0$ appear much
FIGURE 2.71 (a)

PROBABILITY DENSITY OF $\phi_i'$

$n = 5$

- $i = 0.33$
- $i = 0.0$
- $i = 1.0$
closer to the normal than the others; $f'_{.33}$ seems to be particularly good, except that it terminates at $\hat{\phi}_{.33} = 0.0$ (as do all the densities under consideration except $f'_{.0}$). Negative $i$-values seem to produce very peaked curves, which implies that perhaps kurtosis is not very close to zero. Also, curves with $i < 0$ have long tails on the right and terminate very sharply on the left, much more so than in their positive counterparts.

Assessment of the plots of $f'_{.33}$ and $f'_{-.33}$ in particular can be linked with the findings presented section 2.5.4. It is apparent from Figure 2.71(b) that $f'_{-.33}$ is not very well approximated by a quadratic in the regions of the curve used to compute confidence limits - namely the tail areas -, although it does appear approximately quadratic around its maximum. This explains why the confidence intervals calculated on the assumption of the asymptotic normality of $\hat{\phi}_{-1/3}$ are so far from the exact ones. On the other hand, $f'_{.33}$ appears very similar to a normal curve over its entire range and, despite the finiteness of its range,

1 Although the point of termination is not obvious in Figure 2.71(a), this statement is known to be true.
confidence limits computed from the corresponding asymptotic normal distribution are quite close to the exact. This illustrates the reality, at least for very small $n$, of one of the doubts expressed in section 2.3 about the $\phi_{-1/3}$ parameterization.
2.8 Conclusions

One aim of this chapter was to find the 'best' parameterization of the simple one-parameter exponential distribution; that is, the parameterization producing the maximum likelihood estimator whose distribution is most nearly normal. Let us consider only the members of the general family of section 2.2 for which some measures at least, have been tabulated in this chapter, and ignore the fact that tabulation of the distribution of \( \hat{\alpha} \) permits exact confidence intervals to be obtained. Then the parameterization with \( i = 0.33 \) seems to be the 'best' by virtue of its minimal skewness and by the subjective assessment of its density compared with the normal as in section 2.7. The assumption of asymptotic normality of \( \hat{\phi}_{1/3} \) produced accurate confidence limits for \( \alpha \) and the bias of \( \hat{\phi}_{0.33} \) was small. Only the \( \alpha \)-parameterization itself can compare at all favourably with \( \phi_{0.33} \). \( \hat{\alpha} \) is unbiased, its exact and asymptotic variances agree, and it is a meaningful quantity. Asymptotic assumptions about \( \hat{\alpha} \) however, are not very reasonable, since for small \( n \) its distribution is very skew and not very normal-looking. Yet \( i = 1 \) is on all counts a better parameterization than \( i = -1 \), and if the choice must lie between these two physically meaningful parameterizations, then \( i = 1 \) is undoubtedly superior. It is
surprising, in fact, that the $i = -1$ parameterization is used at all, for surely it is always better to obtain intervals for $\alpha$ and invert, if $\theta$ is the parameter of interest. The log parameterization is mediocre by all criteria, its one strong point being that its estimator, $\hat{\phi}_0$, can assume any real value and is not constrained to lie in an interval with a finite limit. A real oddity is the value of $i = -0.33$ suggested by Sprott and Kalbfleisch, for it compares unfavourably with several other $i$-values.

The other purpose of this chapter was to examine measures which could be used to indicate which parameterization, from a specified family of parameterizations, had the estimator which was the most nearly normally distributed.

Size of bias is not very useful as a criterion for selecting a transformation, although it is an interesting associated quantity. If an experimenter wishes to estimate $\alpha$, say, as a parameter which is meaningful to him, he will do so regardless of bias. If there is a good reason why he should use another parameterization, for example one whose estimator has a less skewed distribution, and this new parameter can be shown to be inherently interesting to him as well, he could probably be persuaded to use the latter. If, however, the bias of this new estimator were much greater
than that of the original, he might well use the new parameterization only to obtain intervals which he could invert to get intervals for original parameter. Thus bias is likely to be of import only as a somewhat subsidiary quantity associated with a transformation.

Accuracy of confidence intervals based on asymptotic assumptions is perhaps the best measure to use in assessing a parameterization, since confidence intervals often form an important part of experimental results. It is for the calculation of these limits that normal theory is assumed, and accurate confidence limits for a parameter based on this theory can be obtained only when the distribution of the parameter's estimator is close to the normal. Bartlett's method of transforming to a parameter whose estimator distribution has zero skewness, to a second order approximation, produces very accurate confidence limits, even for small \( n \). Therefore his method seems a very reasonable one to use, if the score function and its moments are not too complicated. In general, parameterizations whose estimator distributions are not very skewed and are close to the normal overall yield the most accurate intervals. One would expect, then, that the amount of skewness and the proximity to normality should be the two most important criteria in choosing a transformation. Unfortunately,
however, one cannot always measure or examine these. Bartlett's method is not always straightforward to apply; minimisation of skewness, even if only approximate, as in this chapter, requires knowledge of the second and third moments of the distribution of the estimator, which can only be obtained if either the exact distribution itself is known or else there is a simple algebraic expression for the estimator in terms of the observations; in order to make general comparisons of certain densities with the normal one must know which densities are likely to be reasonably close to the normal and therefore worth considering, and what the explicit expressions for these densities are.

Regardless of the criteria used, how good the best transformation is depends to a certain extent on the family of parameterizations under consideration, and the choice of a good family is not always obvious.

One important point brought out in this chapter is that one must be careful to use criteria consistent with the purpose of the transformation. For example, choosing a parameterization because its estimator distribution has nice properties in the region of the maximum (or mean) does not insure that it will yield accurate confidence limits, which are calculated from the tail areas.
CHAPTER 3

SOME SIMPLE TWO-PARAMETER DISTRIBUTIONS

3.1 Introduction

From the ideas concerning distributions of estimators arising from single-parameter models, we move on to consider multi-parameter situations and the additional problem of correlation between estimators.

The original aim was to find, for multi-parameter models with correlated parameter estimators, a reparameterization yielding estimators which were independent or at least less correlated. Certain questions come to mind immediately upon consideration of this aim. Firstly, what measure of dependence should be used, particularly when iterative techniques are necessary to find the solution to the normal equations? Is the usual asymptotic variance-covariance matrix, obtained by evaluation of second derivatives or second differences of the log likelihood function, reliable as a measure of dispersion when one is dealing with small samples? Even if the maximum likelihood estimators can be expressed as simple functions of the observations, covariances may not be so easy to obtain. And lastly, as in Chapter 2, one has the problem of specifying a suitable
3.2

family of parameterizations to consider.

The answer to the first question will often be dictated by the particular model under consideration and, to some extent, by the family of parameterizations being studied. It is difficult - if not impossible - to find a simple, universal measure of dependence which, when used in conjunction with any desired family of parameterizations (or even, indeed, with a carefully specified family), will yield a 'better' set of parameters.

The second question will not be investigated, although asymptotic theory will be used, on the assumption that the reduction of asymptotic correlation is probably better than nothing when exact results are not obtainable. In fact, the effects of reparameterizations suggested by asymptotic theory are examined for some simulated data.

In the background at all times is the desire for normality, since one would like to use asymptotic normality assumptions to make inferences about the parameters. Thus, a reparameterization which leads to estimators whose distribution is more poorly approximated by a normal distribution than that for the original estimators will not be regarded as favourably as one which yields estimators which appear more normally distributed. One would wish,
therefore, to ensure that any suggested reparameterization would not adversely affect normality. This matter is considered in this chapter. Here, too, there is the problem of what measures to use to determine how well a distribution is approximated by the normal.

In this chapter, the two-parameter exponential and the Gamma distributions are examined; the latter presents many more complications than the former and is, as a result, discussed in greater depth. As in Chapter 2, these particular distributions are studied primarily for the purposes of developing and illustrating techniques, and not for the purpose of finding reparamaterizations, although suggested parameter transformations could prove interesting.
3.2 Two-Parameter Bivariate Exponential Distribution

The first distribution to be discussed is the very simple two-parameter bivariate exponential. Its most usual parameterization yields independent estimators; a parameterization producing dependent estimators is suggested and measures of dependence which could possibly be used as criteria for selecting an alternative parameterization giving less correlated estimators are examined. A family of parameterizations which includes the usual one is specified.

3.2.1 Specification of the Distribution

The density of this distribution is most usually expressed as

\[ f(x_1, x_2) = \alpha_1 \alpha_2 e^{-\alpha_1 x_1 - \alpha_2 x_2} \]

Marginal densities are then

\[ f_1(x_1) = \alpha_1 e^{-\alpha_1 x_1} \]

\[ f_2(x_2) = \alpha_2 e^{-\alpha_2 x_2} \]
and so
\[ f(x_1, x_2)dx_1dx_2 = f_1(x_1)f_2(x_2)dx_1dx_2. \]
Thus, \( x_1 \) and \( x_2 \) are independent.

If \( n \) independent pairs \((x_{i1}, x_{i2})\), \( i = 1, 2, \ldots, n \)
are observed, then, letting \( T_1 = \sum_{i=1}^{n} x_{i1} \) and
\[ T_2 = \sum_{i=1}^{n} x_{i2}, \]
the solution to the normal equations is
\[ \hat{\alpha}_1 = n/T_1, \quad \hat{\alpha}_2 = n/T_2. \]
It follows then, that \( T_1 \) and \( T_2 \) are independent and so
\( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) are independent too.

Equation 3.21 can also be written as
\[ f(x_1, x_2)dx_1dx_2 = \theta_1 e^{-\theta_1 x_1/\theta_2} - \theta_2 x_2 \, dx_1dx_2 \quad (3.22) \]
and the likelihood of \((\theta_1, \theta_2)\), after \( T_1 \) and \( T_2 \) have been
observed, is
\[ \theta_1^n e^{-\theta_1 T_1/\theta_2} - \theta_2 T_2. \]
Let \( L(\theta_1, \theta_2) \) represent the log likelihood function of \((\theta_1, \theta_2)\). Then

\[
L(\theta_1, \theta_2) = n \log \theta_1 - \frac{\theta_1 T_1}{\theta_2} - \theta_2 T_2,
\]

\[
\begin{align*}
\frac{\partial L}{\partial \theta_1} &= \frac{n}{\theta_1} - \frac{T_1}{\theta_2}, \\
\frac{\partial L}{\partial \theta_2} &= -\frac{\theta_1 T_1}{\theta_2^2} - T_2.
\end{align*}
\]

The solution to the normal equations is therefore

\[
\hat{\theta}_1 = \frac{n^2}{T_1 T_2}, \quad \hat{\theta}_2 = \frac{n}{T_2}.
\]

Since

\[
f_1(x_{1i}, x_{2i}, \ldots, x_{ni}) \prod_{i=1}^{n} dx_i = (\theta_1/\theta_2)^n e^{-\theta_1 T_1/\theta_2} \prod_{i=1}^{n} dx_i
\]

and

\[
f_2(x_{1i}, x_{2i}, \ldots, x_{ni}) \prod_{i=1}^{n} dx_i = \theta_2^n e^{-\theta_2 T_2} \prod_{i=1}^{n} dx_i,
\]

1. \( L \) alone will be used to represent the log likelihood of \((\theta_1, \theta_2)\) when there is no ambiguity.
it is known from Chapter 2, that $T_1$ and $T_2$ have Gamma distributions and that

$$E\left( \frac{1}{T_1} \right) = \frac{\theta_1}{\theta_2(n-1)}, \quad E\left( \frac{1}{T_2} \right) = \frac{\theta_2}{(n-1)},$$

and

$$E\left( \frac{1}{T_2^2} \right) = \frac{\theta_2^2}{(n-1)(n-2)}.$$

The covariance of $\hat{\theta}_1$ and $\hat{\theta}_2$ can then be calculated.

$$\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = n^3 E\left( \frac{1}{T_1 T_2^2} \right) - n^3 E\left( \frac{1}{T_1 T_2} \right) E\left( \frac{1}{T_2} \right)$$

$$= n^3 E\left( \frac{1}{T_1} \right) \left\{ E\left( \frac{1}{T_2^2} \right) - E^2\left( \frac{1}{T_2} \right) \right\}$$

$$= n^3 \frac{\theta_1 \theta_2}{(n-1)^3(n-2)}.$$

This covariance is not in general zero and so $\hat{\theta}_1$ and $\hat{\theta}_2$ are correlated.

Consider now the general family of parameter transformations from $(\theta_1, \theta_2)$ to $(\alpha_1, \alpha_2)$ specified by

$$\theta_1 = \alpha_1^{i_1} \alpha_2^{j_1}, \quad \theta_2 = \alpha_1^{i_2} \alpha_2^{j_2}. \quad (3.24)$$
Using suitable measures and transformation family 3.24, can a reparameterization of 3.22 be found which yields independent maximum likelihood estimators, given that at least one such parameterization exists?

The density 3.22 should be relatively easy to work with, having two major advantages over most two-parameter densities. Firstly, maximum likelihood estimators for both parameters are simple functions of the sufficient statistics $T_1$ and $T_2$. Secondly, since the distributions of $T_1$ and $T_2$ are known, the moments of the maximum likelihood estimators, including their covariance, can be easily calculated.

Due to the lack of generality of any method using exact covariance, attention is centred hereinafter on measures which are likely to be more frequently obtainable.

3.2.2 Use of the Mixed Second Derivative of the Log Likelihood

To obtain asymptotic estimates of variance and covariance, the inverse of the information matrix is evaluated at the maximum likelihood estimates. The covariance, estimated from the non-diagonal elements in this
inverse matrix, is directly proportional to the expectation of the mixed second derivative of the log likelihood,

\[ E \left[ \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \right], \] evaluated at \((\hat{\theta}_1, \hat{\theta}_2)\). If a parameterization, \((\alpha_1, \alpha_2)\) say, can be found such that \( E \left[ \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} \right] \) is zero, then, asymptotically at least, \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) will have zero covariance, although the estimate of this covariance need not be zero. A stronger condition, yielding a solution less often, is that \( \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} \) be zero. If a solution to this can be obtained, it should be used in preference to a solution to \( E \left[ \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} \right] = 0 \) as it ensures that the asymptotic estimate of covariance is identically zero. Although zero covariance does not in general imply independence, if the assumptions of normality usually employed in making inferences are reasonably justified, then zero covariance does in fact allow an independent inference to be made about each parameter.
Consider the family of parameterizations specified by 3.24. Then use of equations 3.23 in

\[
\frac{\partial L}{\partial \alpha_i} = \frac{\partial L}{\partial \theta_1} \frac{\partial \theta_1}{\partial \alpha_i} + \frac{\partial L}{\partial \theta_2} \frac{\partial \theta_2}{\partial \alpha_i}, \quad i = 1, 2
\]

yields

\[
\frac{\partial L}{\partial \alpha_1} = \frac{i_1 n}{\alpha_1} - (i_1 - i_2)T_1 \alpha_1^{i_1 - i_2 - 1} a_2^{j_2 - j_1} - i_2 T_2 \alpha_1^{i_2 - 1} a_2^{j_2} \quad (3.25)
\]

and

\[
\frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} = + (i_1 - i_2)(j_1 - j_2)T_1 \alpha_1^{i_1 - i_2 - 1} a_2^{j_2 - j_1 - 1}
- (i_2 j_2)T_2 \alpha_1^{i_2 - 1} a_2^{j_2 - 1} \quad (3.26)
\]

In order that 3.26 be identically zero, the coefficients \((i_1 - i_2)(j_1 - j_2)\) and \((i_2 j_2)\) must both be zero. There are two permissible ways in which this may be accomplished:

\[
(1) \quad i_2 = 0; \quad i_1 \neq 0; \quad j_1 = j_2 = j, \quad \text{where } i_1 \text{ and } j \text{ can take any real non-zero values};
\]

\[
(2) \quad j_2 = 0; \quad j_1 \neq 0; \quad i_1 = i_2 = i, \quad \text{where } j_1 \text{ and } i \text{ can take any real non-zero values}. \quad (3.27)
\]

By the symmetry in the specification 3.24, (1) and (2) are equivalent. Therefore only members of the specified family which are of the form

\[
\theta_1 = \alpha_1^i \alpha_2^j, \quad \theta_2 = \alpha_2^j, \quad \text{where } i \text{ and } j \text{ are non-zero, yield parameterizations with the desired property. In the particular case where } i = j = 1, \text{ the original parameterization of 3.21 is obtained.}
3.2.3 Use of the First Derivatives of the Log Likelihood

If the mixed second derivative of the log likelihood function is difficult to work out or work with, the first derivatives may sometimes prove useful in the choice of a reparameterization. If the first derivative of the log likelihood with respect to a parameter \( \alpha_1 \) can be made a function of \( \alpha_1 \) only, and similarly for \( \alpha_2 \), then \( \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} \) will be zero and the asymptotic results discussed in 3.2.2 will follow. As with the condition that \( \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} \) be zero which has just been discussed, these conditions are quite strong and so it will often be impossible to find a parameterization which satisfies them.

The aim, then, is to choose from the family specified in 3.24 those members, if any, for which

\[
\frac{\partial L}{\partial \alpha_k} = h_k(\alpha_k), \quad k = 1, 2 \tag{3.28}
\]

Equation 3.25 gives \( \frac{\partial L}{\partial \alpha_1} \) and, by symmetry,

\[
\frac{\partial L}{\partial \alpha_2} = \frac{j_1 n}{\alpha_2} (j_1 - j_2) T_{1} \alpha_1^{i_1 - i_2} \alpha_2^{j_1 - j_2 - 1} - j_2 T_2 \alpha_1^{i_2} \alpha_2^{j_2 - 1}.
\]
The two ways in 3.27, section 3.2.2, are the only two permissible ways of satisfying 3.28. Using (1) with \( i_1 = j = 1 \) gives

\[
\theta_1 = \alpha_1 \alpha_2, \quad \theta_2 = \alpha_2
\]

as in section 3.2.2.
3.3 Gamma Distribution

The exponential distribution was very straightforward and simple to work with, particularly since a parameterization producing independent estimators was known. The Gamma distribution is more complex and reveals difficulties which are likely to be encountered in other distributions.

In the following sections a family of parameter transformations is specified and the methods discussed in 3.2 are tried. Several sets of simulated data are fed into a maximization routine, along with both the original and the reparameterized likelihoods, and the estimated covariances (calculated from second differences) and correlations for the two parameterizations computed. An attempt is made to compare the distributions of the estimators of the original and transformed parameters with respect to their normality.

3.3.1 Specification of the Distribution

The Gamma distribution is a univariate two-parameter distribution with density expressed as
\[ f(x) \, dx = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{-x/\beta} \, dx \]

Then if \( x_1, x_2, \ldots, x_n \) are independent observations from this distribution, their joint density is

\[ f(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} dx_i = \left[ \frac{1}{\Gamma(\alpha) \beta^\alpha} \right]^n x_i^{\alpha-1} \exp \left[ -\frac{n}{\beta} x_i/\beta \right] \prod_{i=1}^{n} dx_i \]

Letting \( T_1 = \prod_{i=1}^{n} x_i \) and \( T_2 = \sum_{i=1}^{n} x_i \), the likelihood function can be written as

\[ \text{lik}(\alpha, \beta ; T_1, T_2) = \left[ \frac{1}{\Gamma(\alpha) \beta^\alpha} \right]^n T_1^{\alpha-1} e^{-T_2/\beta} \]

and \((T_1, T_2)\) are jointly sufficient for \((\alpha, \beta)\). Then the log likelihood is

\[ L(\alpha, \beta) = -n \log \Gamma(\alpha) - n\alpha \log \beta + (\alpha - 1) \log T_1 - T_2/\beta \]

and

\[ \frac{\partial L}{\partial \alpha} = -n \psi(\alpha) - n \log \beta + \log T_1, \]

where \( \psi(\alpha) = \frac{d}{d\alpha} \log \Gamma(\alpha) \)

\[ \frac{\partial L}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{T_2}{\beta^2} \]

(3.31)
Since there is no closed solution to $\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial \beta} = 0$, $\hat{\alpha}$ and $\hat{\beta}$ must be obtained by an iterative procedure and hence their exact distribution (including their correlation, of course) is not known. However the asymptotic covariance of $(\hat{\alpha}, \hat{\beta})$ can be calculated. Differentiating equations 3.31 gives

$$\frac{\partial^2 L}{\partial \alpha^2} = -n \psi^{(1)}(\alpha),$$

where $\psi^{(n)}(\alpha) = \frac{d^{n+1}}{d\alpha^{n+1}} \log \Gamma(\alpha), n \geq 1.$

$$\frac{\partial^2 L}{\partial \beta^2} = \frac{n\alpha}{\beta^2} - \frac{2T_2}{\beta^3}$$

$$\frac{\partial^2 L}{\partial \alpha \partial \beta} = -\frac{n}{\beta}$$

and so, since $E(T_2) = n E(x_1) = n\alpha \beta$, the information matrix is

$$I = \begin{pmatrix} n \psi^{(1)}(\alpha) & n/\beta \\ n/\beta & n\alpha/\beta^2 \end{pmatrix}$$

yielding an asymptotic variance-covariance matrix

$$V = \frac{1}{|I|} \begin{pmatrix} n\alpha/\beta^2 & -n/\beta \\ -n/\beta & n \psi^{(1)}(\alpha) \end{pmatrix}$$

The asymptotic covariance of $(\hat{\alpha}, \hat{\beta})$ is therefore

$\psi^{(n)}(\alpha)$ is the polygamma function already encountered in Chapter 2.
which will not in general be zero, except in the limit when 
\( n \) is large relative to 
\[
\frac{\beta}{\alpha \psi^{(1)}(\alpha) - 1}
\]

Thus we shall try to find a parameterization \((\theta_1, \theta_2)\)
which yields less correlated maximum likelihood estimators.

The particular family of parameterizations to be considered
in subsequent sections is defined by the relationships
\[
\begin{align*}
\alpha &= \theta_1 \\
\beta &= \theta_1^{i} \theta_2^{j} , \; j \neq 0 \end{align*}
\]

3.3.2 Mixed Second Derivatives

Substituting from 3.31 and 3.32 into
\[
\frac{\partial L}{\partial \theta_k} = \frac{\partial L}{\partial \alpha} \frac{\partial \alpha}{\partial \theta_k} + \frac{\partial L}{\partial \beta} \frac{\partial \beta}{\partial \theta_k}, \; k = 1, 2
\]
gives
\[
\begin{align*}
\frac{\partial L}{\partial \theta_1} &= -n \psi(\theta_1) - ni \log \theta_1 - nj \log \theta_2 \\
&\quad + \log T_1 - ni + \frac{i T_2}{\theta_1^{i+1} \theta_2^{j+1}} \\
\frac{\partial L}{\partial \theta_2} &= j \left\{ \frac{n \theta_1}{\theta_2} + \frac{T_2}{\theta_1^{i+1} \theta_2^{j+1}} \right\}
\end{align*}
\]
and
\[
\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} = - \frac{n j}{\theta_2} - \frac{ij T_2}{\theta_1 \theta_2^{j+1}} \quad (3.34)
\]

The expression 3.34 can be identically zero only if \( j = 0 \) which is not permissible.

Consider then the expectation of expression 3.34. Since
\[
E(T_2) = n \alpha \beta = n \theta_1 i + \theta_2 j
\]
then
\[
- E \left[ \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \right] = \frac{n j (i+1)}{\theta_2}
\]
If \( i = -1 \), this expectation is zero for any permissible value of \( j \). If, for simplicity, \( j \) is taken to be 1, then from the relationships 3.32,
\[
\alpha = \theta_1 \\
\beta = \theta_2 / \theta_1
\]
is given as the parameter transformation for which the estimators of \( \theta_1 \) and \( \theta_2 \) have zero asymptotic covariance.

For this parameterization equations 3.33 become
\[
\frac{\partial L}{\partial \theta_1} = -n \psi(\theta_1) + n \log \theta_1 - n \log \theta_2 \\
+ \log T_1 + n - T_2 / \theta_2
\]
\[
\frac{\partial L}{\partial \theta_2} = \frac{\theta_1}{\theta_2} \left[ \frac{T_2}{\theta_2} - n \right]
\]
and so
\[ \hat{\theta}_2 = \frac{T_2}{n} \]
and \( \hat{\theta}_1 \) is then a function of \( T_1 \) and \( \hat{\theta}_2 \).

This reparameterization satisfies the condition of zero asymptotic covariance, but in order to see if it is effective in reducing the actual correlation of estimators comparisons can be made using data from Gamma distributions. An iterative maximization procedure can be applied to sets of data from \( G(\alpha, \beta) \) for both the original and the new parameterizations, and the resultant estimates of correlation compared. The estimates of correlation are calculated either from second differences or from exact second differential coefficients, and so need not be zero even for the new parameterization. If the absolute values of the correlation coefficients for the two parameterizations are compared for several sets of data with different true parameter values, some tentative conclusions can perhaps be drawn about the effectiveness of the parameterization chosen by this criterion, as applied to this particular distribution.

3.3.3 Generation of Data

Use can be made of the relationships between Gamma, \( \chi^2 \) and normal distributions to generate data.
from a Gamma distribution; the sum of the squares of \( n \) standard normal deviates is a \( \chi^2 \) variate with \( n \) degrees of freedom; if \( x \) is distributed as \( \Gamma(\frac{n}{2}, \beta) \), then \( 2x/\beta \) is distributed as \( \chi^2 \) on \( n \) degrees of freedom. Therefore, \( n \) random standard normal deviates can be generated by a pseudo-random number generator (or read from a table of random Gaussian deviates); these can be squared and summed to obtain one \( \chi^2_n \) variate \( y \), say, which, when multiplied by \( \beta/2 \) becomes a Gamma variate with parameter pair \( (\frac{n}{2}, \beta) \). This method is, however, somewhat restrictive, as it only permits the simulation of Gamma distributions with first parameters which are multiples of 0.5, since \( n \), the degrees of freedom of the \( \chi^2 \), must be integral. For the purposes of this chapter this is not a severe restriction since enough reasonable (and interesting) parameter values are still available. Since \( \beta \) is only a scale parameter it will be considered as fixed at 1.0 for the rest of this chapter.

In total, fifteen 'acceptable' sets of data were generated, three for each of five values of \( \alpha \) (0.5, 1.0, 2.0, 5.0, 10.0) with each set comprising thirty variates from \( \Gamma(\alpha, 1.0) \). Two different methods of producing the necessary Gaussian deviates were employed: one, based on tables of random Gaussian deviates, was used with small \( \alpha \) values (0.5, 1.0, 2.0) and the other, utilizing the
computer and a pseudo-random number generator, was used for the two larger values of \( a \). A Chi-square goodness of fit test was performed on each set of data, and those sets with very large \( \chi^2 \) values were rejected. When a data set was rejected thusly another was generated to replace it, so that in the end there were three acceptable sets of variates for each parameter value.

3.3.31 Data Generation via a Pseudo-Random Number Generator

To produce a set of data of size 30 from \( G(a,1.0) \), 30 x \( (2a) \) standard normal deviates are needed. The method of Newman and Odell (1971) was programmed for the IBM 360/50 to produce these. This method requires random variates from the uniform distribution on \((-0.5, 0.5)\) and these were obtained by using a variation of the IBM Scientific Subroutine Package routine RANDU. Beginning with the first Gaussian deviate produced, each group of \( 2a \) consecutive deviates was used to make one \( \chi^2_{2a} \) variate, which was then transformed into a variate from \( G(a,1.0) \). The procedure is repeated for each data set required, with a new starting value every time. Initially, this starter

1. The decomposition method was used, the programme being kindly supplied by Mr. Brian Thompson of the Department of Statistics in Edinburgh.
was chosen from a table of random digits as a five (or more) digit odd integer (the value used was 746,909); for further data sets generated in the same run, the starter was taken as the last random integer generated by the modified RANDU for the previous data set.

This method was used to generate data from G(5.0, 1.0) and G(10.0, 1.0), although certainly the Gamma distributions with smaller \( \alpha \)-values could have been simulated in this way also.

### 3.3.32 Data Generation via Tables of Random Numbers

Tables of random digits and random Gaussian deviates\(^1\) were used to generate sets of data from Gamma distributions \( G(0.5, 1.0) \), \( G(1.0, 1.0) \), and \( G(2.0, 1.0) \). This method can be described as follows for each data set of size 30:

1. A starting point is chosen in both the table of random digits and the table of Gaussian

---

1. Rand Corporation's 'A Million Random Digits with 100,000 Normal Deviates' were the tables actually used.
deviates by the procedure suggested in the book of tables (p.xxiii). The starters are a single digit in the digits table and a signed real number with three decimal places in the table of standard normal deviates.

2. A two-digit integer is formed from digits 3 and 4 of the five-digit number containing the starter; the second normal deviate is found by proceeding this number of columns from the first Gaussian deviate (the starter). For example, if this first deviate is in line 500, column 1 and the first two-digit number from the digits table is 42, then the second normal deviate is that in line 504, column 3. (There are 10 columns per line.)

3. The next two-digit integer is that in the same column as the previous one, but in the

---

1. If these two digits are both zero the corresponding digits of the number in the next row, same column are used.
next row, and it is used in the same manner as the first to determine the position of the next normal deviate. When the bottom of a five-digit column is reached, the next column is started.

4. In general, the position of the $k^{th}$ normal deviate is that of the $(k-1)^{st}$ one plus the number of columns represented by the $(k-1)^{st}$ two-digit integer.

5. The above procedure is repeated until $30n$ normal deviates have been obtained, where $n$ is twice the value of the first parameter of the Gamma distribution being simulated.

6. The resultant sequence of normal deviates is divided into consecutive groups of size $n$. Within each group the deviates are squared and these squares summed to form thirty $\chi^2_n$ variates.

7. Each of these $\chi^2$ variates is then multiplied by $\beta/2$ to obtain the appropriate Gamma variate.
8. For each set of data generated, a new pair of starting values is used.

3.3.3 $\chi^2$ Goodness of Fit Tests

On each set of data generated a $\chi^2$ goodness of fit test was performed. Any set with such a large $\chi^2$ value that the probability of obtaining a larger one is less than 0.05 was discarded. Intervals used for these tests are shown in Table 3.31; Table 3.32 displays the results.

3.3.4 Effects of Reparameterization on Correlation of Estimates

For each of the fifteen accepted sets of data generated in section 3.3.3, maximum likelihood estimates of the parameters and estimates of the variance-covariance matrices (from second differences) and correlation coefficients can be obtained for both parameterizations by use of a general programme for maximum likelihood. The estimates of the

1. The programme used was MAXLIKE 10 written by Professor D.J. Finney and Dr. D.N. Lawley in the IMP(AA) language and available at the Edinburgh Regional Computing Centre.
### TABLE 3.31

**INTERVALS USED IN $\chi^2$ GOODNESS OF FIT TESTS ON $n=30$ GENERATED $G(\alpha, \beta)$ VARIATES**

<table>
<thead>
<tr>
<th>Interval</th>
<th>$\chi^2_5$ Test: $\alpha = 0.5, \beta = 1.0$</th>
<th>Expected No.**</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 0.02</td>
<td>4.77</td>
<td></td>
</tr>
<tr>
<td>0.02 - 0.10</td>
<td>5.58</td>
<td></td>
</tr>
<tr>
<td>0.10 - 0.25</td>
<td>5.25</td>
<td></td>
</tr>
<tr>
<td>0.25 - 0.50</td>
<td>4.89</td>
<td></td>
</tr>
<tr>
<td>0.50 - 1.00</td>
<td>4.80</td>
<td></td>
</tr>
<tr>
<td>&gt; 1.00</td>
<td>4.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>30.00</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interval</th>
<th>$\chi^2_6$ Test: $\alpha = 2.0, \beta = 1.0$</th>
<th>Expected No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 0.70</td>
<td>4.68</td>
<td></td>
</tr>
<tr>
<td>0.70 - 1.1</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td>1.1 - 1.5</td>
<td>4.23</td>
<td></td>
</tr>
<tr>
<td>1.5 - 2.0</td>
<td>4.56</td>
<td></td>
</tr>
<tr>
<td>2.0 - 2.5</td>
<td>3.57</td>
<td></td>
</tr>
<tr>
<td>2.5 - 3.5</td>
<td>4.53</td>
<td></td>
</tr>
<tr>
<td>&gt; 3.5</td>
<td>4.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>30.00</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interval</th>
<th>$\chi^2_6$ Test: $\alpha = 10.0, \beta = 1.0$</th>
<th>Expected No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 6.75</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td>6.75 - 8.0</td>
<td>4.14</td>
<td></td>
</tr>
<tr>
<td>8.0 - 9.0</td>
<td>3.90</td>
<td></td>
</tr>
<tr>
<td>9.0 - 10.0</td>
<td>3.87</td>
<td></td>
</tr>
<tr>
<td>10.0 - 11.5</td>
<td>5.07</td>
<td></td>
</tr>
<tr>
<td>11.5 - 13.5</td>
<td>4.62</td>
<td></td>
</tr>
<tr>
<td>&gt; 13.5</td>
<td>4.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>30.00</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interval</th>
<th>$\chi^2_6$ Test: $\alpha = 5.0, \beta = 1.0$</th>
<th>Expected No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 2.7</td>
<td>4.11</td>
<td></td>
</tr>
<tr>
<td>2.7 - 3.5</td>
<td>4.14</td>
<td></td>
</tr>
<tr>
<td>3.5 - 4.2</td>
<td>4.05</td>
<td></td>
</tr>
<tr>
<td>4.2 - 5.0</td>
<td>4.50</td>
<td></td>
</tr>
<tr>
<td>5.0 - 6.0</td>
<td>4.65</td>
<td></td>
</tr>
<tr>
<td>6.0 - 7.5</td>
<td>4.59</td>
<td></td>
</tr>
<tr>
<td>&gt; 7.5</td>
<td>3.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>30.00</strong></td>
<td></td>
</tr>
</tbody>
</table>

* A $\chi^2_5$ test was done on data sets with $\alpha=0.5$ because suitable intervals for a $\chi^2_6$ test could not be found from tables.

** This is expected number of observations lying in the given interval.
### TABLE 3.32

RESULTS OF $\chi^2$ TESTS ON GAMMA DATA

<table>
<thead>
<tr>
<th>Hypothesized Distribution</th>
<th>Data Set No.</th>
<th>Observed $\chi^2$ Value</th>
<th>Probability of Larger Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) $G(0.5,1.0)$</td>
<td>A1</td>
<td>4.14</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>1.88</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>13.74</td>
<td>0.02 **</td>
</tr>
<tr>
<td></td>
<td>A4</td>
<td>7.78</td>
<td>0.17</td>
</tr>
<tr>
<td>(B) $G(1.0,1.0)$</td>
<td>B1</td>
<td>4.87</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>2.02</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>6.75</td>
<td>0.34</td>
</tr>
<tr>
<td>(C) $G(2.0,1.0)$</td>
<td>C1</td>
<td>3.49</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>3.05</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>3.53</td>
<td>0.74</td>
</tr>
<tr>
<td>(D) $G(5.0,1.0)$</td>
<td>D1</td>
<td>2.73</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>2.00</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>3.31</td>
<td>0.77</td>
</tr>
<tr>
<td>(E) $G(10.0,1.0)$</td>
<td>E1</td>
<td>4.00</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>7.63</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>6.73</td>
<td>0.35</td>
</tr>
</tbody>
</table>

* All these values are from a $\chi^2$ distribution, except those based on data for (A), which are from a $\chi^2_5$ distribution.

** This data set was not used.
parameters and the correlation coefficients are shown in Table 3.33, along with the means of the correlations for different data sets with the same true parameter values. From the table it is evident that \( \hat{\alpha} \) and \( \hat{\beta} \) are generally highly negatively correlated, with the absolute value of the correlation increasing as the true value for the parameter \( \alpha \) increases. The correlation coefficients of \( (\hat{\theta}_1, \hat{\theta}_2) \), on the other hand, are close to zero, sometimes taking positive values and sometimes negative, and show no trend with changing true values. The small absolute values of the coefficients of correlation of \( (\hat{\theta}_1, \hat{\theta}_2) \) suggest that inferences could be made about \( \theta_1 \) and \( \theta_2 \) independently with reasonable justification; one would, however, be sceptical regarding independent inferences about \( \alpha \) and \( \beta \).

3.3.5 Effects of Reparameterization on the Distribution of Estimators

The \( \theta_1, \theta_2 \) parameterization, then, does appear to have the desired property of lack of correlation between the estimators of the two parameters. But what distributional properties do these estimators possess? Is their distribution closer to the normal than that of \( (\hat{\alpha}, \hat{\beta}) \)? It would be nice to be able to answer these questions at least partially.
### TABLE 3.33

**ESTIMATES AND COEFFICIENTS OF CORRELATION FOR GENERATED G(\(\alpha,\beta\)) DATA**

<table>
<thead>
<tr>
<th>((\alpha,\beta))</th>
<th>DATA SET NO.</th>
<th>(\hat{\alpha} = \hat{\theta}_1)</th>
<th>(\hat{\beta})</th>
<th>(\hat{\theta}_2)</th>
<th>(r(\hat{\alpha},\hat{\beta}))</th>
<th>(r(\hat{\theta}_1,\hat{\theta}_2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0.5, 1.0))</td>
<td>A1</td>
<td>0.470</td>
<td>0.766</td>
<td>0.360</td>
<td>-0.6009</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>0.589</td>
<td>0.877</td>
<td>0.516</td>
<td>-0.6687</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>0.328</td>
<td>1.300</td>
<td>0.426</td>
<td>-0.5259</td>
<td>-0.0012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MEAN, (\bar{r} = -0.5985)</td>
<td>0.0028</td>
</tr>
<tr>
<td>((1.0, 1.0))</td>
<td>B1</td>
<td>0.803</td>
<td>1.341</td>
<td>1.076</td>
<td>-0.7189</td>
<td>-0.0002</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>0.998</td>
<td>1.151</td>
<td>1.148</td>
<td>-0.7683</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>0.785</td>
<td>1.090</td>
<td>0.856</td>
<td>-0.6995</td>
<td>0.0126</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MEAN, (\bar{r} = -0.7289)</td>
<td>0.0111</td>
</tr>
<tr>
<td>((2.0, 1.0))</td>
<td>C1</td>
<td>3.209</td>
<td>0.586</td>
<td>1.881</td>
<td>-0.9187</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>1.827</td>
<td>1.163</td>
<td>2.125</td>
<td>-0.8657</td>
<td>0.0116</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>1.598</td>
<td>1.187</td>
<td>1.897</td>
<td>-0.8495</td>
<td>0.0129</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MEAN, (\bar{r} = -0.8780)</td>
<td>0.0130</td>
</tr>
<tr>
<td>((5.0, 1.0))</td>
<td>D1</td>
<td>6.262</td>
<td>0.789</td>
<td>4.941</td>
<td>-0.9541</td>
<td>-0.0060</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>4.831</td>
<td>1.055</td>
<td>5.097</td>
<td>-0.9430</td>
<td>-0.0087</td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>6.910</td>
<td>0.740</td>
<td>5.115</td>
<td>-0.9686</td>
<td>-0.0073</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MEAN, (\bar{r} = -0.9552)</td>
<td>-0.0073</td>
</tr>
<tr>
<td>((10.0, 1.0))</td>
<td>E1</td>
<td>8.294</td>
<td>1.293</td>
<td>10.725</td>
<td>-0.9638</td>
<td>0.0189</td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>6.785</td>
<td>1.438</td>
<td>9.757</td>
<td>-0.9648</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>10.486</td>
<td>1.037</td>
<td>10.877</td>
<td>-0.9829</td>
<td>-0.0066</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MEAN, (\bar{r} = -0.9705)</td>
<td>0.0057</td>
</tr>
</tbody>
</table>
The Taylor series expansion of a two-parameter log likelihood $L(\theta_1, \theta_2)$ about the maximum likelihood estimates ($\hat{\theta}_1, \hat{\theta}_2$) including up to cubic terms can be expressed as

$$\log \frac{\text{lik}(\theta_1, \theta_2)}{\text{lik}(\hat{\theta}_1, \hat{\theta}_2)} = \frac{1}{2} \left( (\theta_1 - \hat{\theta}_1)^2 \frac{\partial^2 L}{\partial \theta_1^2} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} \right.$$ 

$$+ \left( (\theta_2 - \hat{\theta}_2)^2 \frac{\partial^2 L}{\partial \theta_2^2} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} + (\theta_1 - \hat{\theta}_1)(\theta_2 - \hat{\theta}_2) \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} \right)$$ 

$$+ \frac{1}{3!} \left( (\theta_1 - \hat{\theta}_1)^3 \frac{\partial^3 L}{\partial \theta_1^3} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} + (\theta_1 - \hat{\theta}_1)(\theta_2 - \hat{\theta}_2)^2 \frac{\partial^3 L}{\partial \theta_1 \partial \theta_2^2} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} \right.$$ 

$$+ (\theta_1 - \hat{\theta}_1)(\theta_2 - \hat{\theta}_2)^2 \frac{\partial^3 L}{\partial \theta_1 \partial \theta_2^2} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} + (\theta_2 - \hat{\theta}_2)^3 \frac{\partial^3 L}{\partial \theta_2^3} \bigg|_{\hat{\theta}_1, \hat{\theta}_2} \right)$$

(3.35)

Consider this expansion for the $\theta_1, \theta_2$ parameterization. $\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \bigg|_{\hat{\theta}_1, \hat{\theta}_2}$ is in general so close to zero, as has been shown, as to be negligible. Thus, if cubic and higher order terms of (3.35) are ignored then for each fixed value
of the log relative likelihood (the left hand side of 3.35),
3.35 is the equation of an ellipse with axes parallel to
the $\theta_1$, $\theta_2$ axes and centre $(\hat{\theta}_1, \hat{\theta}_2)$. A similar set of
ellipses would be obtained for the log likelihood arising
from the distribution of $(\hat{\theta}_1, \hat{\theta}_2)$ if this were normal
with zero covariance and variances the inverses of second
differential coefficients evaluated at $(\hat{\theta}_1, \hat{\theta}_2)$. If,
however, an actual set of contours for $(\theta_1, \theta_2)$ is plotted
(cf. Figures 3.31a and 3.32a), it is obvious that these
contours are not ellipses, although their axes are parallel
to the $\theta_1, \theta_2$ axes, and so they do not arise from a normal
likelihood. Now, although a normal type likelihood (one
with elliptical contours of constant log likelihood) does
not necessarily arise from a normal distribution, a non-
normal likelihood certainly can not come from a normal
distribution. Therefore, the distribution of $(\hat{\theta}_1, \hat{\theta}_2)$
is not normal, but how much so? Is it closer to the normal
than is the distribution of $(\hat{\alpha}, \hat{\beta})$?

Examination of the cubic coefficients in the series
expansion 3.35 for the two parameterizations will give some
indication of which pair has the more normal likelihood
function, although as mentioned above this does not imply
FIGURE 3.31

CONTOURS OF CONSTANT RELATIVE LIKELIHOOD OF \((\alpha, \beta)\) AND \((\theta_1, \theta_2)\)
FOR DATA FROM \(G(2.0,1.0)\)*

FIGURE 3.31a
\(\theta_1, \theta_2\) PARAMETERIZATION
\[\hat{\theta}_1 = 1.83\]
\[\hat{\theta}_2 = 2.13\]

FIGURE 3.31b
\(\alpha, \beta\) PARAMETERIZATION
\[\hat{\alpha} = 1.83\]
\[\hat{\beta} = 1.16\]

* Generated data set C2 of Table 3.33 was used. \(\alpha = \theta_1 = 2.0; \beta = 1.0; \theta_2 = 2.0.\)
FIGURE 3.32

CONTOURS OF CONSTANT RELATIVE LIKELIHOOD OF \((\alpha, \beta)\) AND \((\theta_1, \theta_2)\)
FOR DATA FROM \(G(2.5, 2.0)^*\)

* Generated data was used. \(\alpha = \theta_1 = 2.5; \beta = 2.0; \theta_2 = 5.0\).
a more normal density. Table 3.34 gives all the third differential coefficients evaluated at the maximum likelihood estimates for both parameterizations; it also gives for purposes of comparison, the coefficients for \(( \hat{\theta}_1, \hat{\theta}_2 )\) in terms of those for \(( \hat{\alpha}, \hat{\beta} )\) and from these it is evident that whenever \(\hat{\alpha} > 1\) all of the non-zero coefficients for \(( \hat{\theta}_1, \hat{\theta}_2 )\) will be less than those for \(( \hat{\alpha}, \hat{\beta} )\), and that when \(\hat{\alpha} < 1\) the reverse is true, except that \(\frac{\partial^3 L}{\partial \theta_1^3}\) is always less than \(\frac{\partial^3 L}{\partial \alpha^3}\). One can conclude, then, that generally (that is, when \(\hat{\alpha} > 1\)) the \(( \theta_1, \theta_2 )\) parameterization yields a relative likelihood function which, when approximated to this order, deviates less from the normal relative likelihood than does that of \(( \alpha, \beta )\). Fourth - and perhaps higher - order coefficients could also be considered to provide more information about the likelihood.

An alternative, however, is to compare plots of the relative likelihood functions for the two parameterizations with the same sets of data. Figures 3.31 a-b and 3.32 a-b show contours of constant relative log likelihood for \(( \theta_1, \theta_2 )\) and \(( \alpha, \beta )\) for two data sets. It appears

1. This fact was well illustrated in Chapter 2 in the single parameter exponential distribution.
<table>
<thead>
<tr>
<th>COEFFICIENT</th>
<th>(\alpha, \beta) PARAMETERIZATION at (\hat{\alpha}, \hat{\beta})</th>
<th>(\theta_1, \theta_2) PARAMETERIZATION at (\hat{\theta}_1, \hat{\theta}_2)</th>
<th>in terms of (\hat{\alpha}, \hat{\beta})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\frac{\partial^3 L}{\partial \phi_1^3})</td>
<td>(-n\psi^{(2)}(\hat{\alpha}))</td>
<td>(-n\left(\psi^{(2)}(\hat{\theta}_1) + \frac{1}{\hat{\theta}_1^2}\right))</td>
<td>(\frac{\partial^3 L}{\partial \alpha^3}</td>
</tr>
<tr>
<td>(\frac{\partial^3 L}{\partial \phi_1^2 \partial \phi_2})</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\frac{\partial^3 L}{\partial \phi_1 \partial \phi_2^2})</td>
<td>(\frac{n}{\hat{\beta}^2})</td>
<td>(-\frac{n}{\hat{\theta}_2^2})</td>
<td>(-\frac{1}{\hat{\beta}^2}\left(\frac{\partial^3 L}{\partial \alpha \gamma^2}</td>
</tr>
<tr>
<td>(\frac{\partial^3 L}{\partial \phi_2^3})</td>
<td>(\frac{4n\hat{\alpha}}{\hat{\beta}^3})</td>
<td>(\frac{4n\hat{\theta}_1}{\hat{\theta}_2^3})</td>
<td>(\frac{1}{\hat{\alpha}^3}\left(\frac{\partial^3 L}{\partial \beta^3}</td>
</tr>
</tbody>
</table>
evident from these that the contours for ( $\theta_1, \theta_2$ ) correspond much more closely to the elliptical contours of a normal relative log likelihood than do those for ( $\alpha, \beta$ ).

In the above, only the relative likelihoods for the two parameterizations are compared; in order to compare the actual distributions of the maximum likelihood estimators, one would have to simulate these distributions by calculating estimates for many sets of data from the same Gamma distribution and then computing the relative frequencies of these estimates. This process would then have to be repeated for several different parameter values if the conclusions were to have any generality. Since an iterative procedure is necessary to obtain most of these estimates, this could be very costly and time-consuming and possibly not very fruitful and so will not be attempted here.

From the foregoing discussion, table, and graphs it appears that ( $\theta_1, \theta_2$ ) produces the more normal relative log likelihood, and so one is tempted to infer that the distribution of ( $\hat{\theta}_1, \hat{\theta}_2$ ) is likely to be at least as well-approximated by the normal as is that of ( $\hat{\alpha}, \hat{\beta}$ ).
3.3.6 Conclusions

The fact that $\hat{\theta}_1$ and $\hat{\theta}_2$ are asymptotically uncorrelated plus the fact, as stated above, that reparameterization to $(\theta_1, \theta_2)$ does not seem to adversely affect the estimator distribution, suggest that the $(\theta_1, \theta_2)$ parameterization is superior, at least in some respects, to the usual parameterization of the Gamma distribution.
3.4 Conclusions

In general, the exact distribution of a pair of maximum likelihood estimators is not known and often not even the exact moments can be determined, as is the case with estimators of the Gamma distribution. The only way to obtain information about the actual properties of estimator distributions in this situation is by massive simulations. As this is not feasible in this study, asymptotic properties of the estimator distributions and characteristics of the likelihood function which are always known serve both as criteria for choosing a parameterization and as bases for assessing this parameterization. The assumption is made that if a reparameterization improves asymptotic properties it probably also improves the corresponding exact properties.

For the two distributions considered in this chapter, parameterizations for which the asymptotic correlation of the estimators either is identically zero or else has expectation zero have been found. In both cases this success was at least partly dependent on the specification of an appropriate family of reparameterizations, which is not in general easy, and is often a matter of some judicious guessing and some trial and error. In the case of the exponential distribution, the estimators of the repara-
meterized exponential are known to be independent. The Gamma distribution, however, has been reparameterized to make the expected asymptotic correlation of the estimators zero, but the estimate of the asymptotic correlation need never be zero. It has been shown for several sets of generated data that the coefficients of correlation for estimators of the reparameterized Gamma cluster around zero and are consistently lower (in absolute value) than the coefficients for the original estimators.

Plots of contours of constant relative log likelihood for data from the Gamma distribution provide further evidence that for the new parameterization inferences can reasonably be made about either parameter independently of the other since the axes of the contours are essentially parallel to the parameter axes; for the original parameters the banana-shaped contours make any independent inferences about the two parameters of the pair appear of doubtful value.

An important part of any attempt at reparameterization for independence is a check of some sort on the extent to which the resultant estimator distribution can be approximated by the normal distribution, particularly as compared
to the distribution of the original estimators. There are two reasons for the interest in normality. Firstly, only if the distribution can be considered to be approximately normal does zero correlation imply that independent statements can be made about the two parameters. Secondly, assumptions of normality are those most frequently used in interval estimation when the exact estimator distribution is not known. If the new parameter pair seems to be at least as good as the old in respect of the distribution of its estimators, then it is certainly acceptable; otherwise, its value is questionable. The assessment of how close to normality an estimator distribution is must be made from the likelihood function when neither explicit expressions for the maximum likelihood estimators nor the exact distribution of the estimators are available. For the Gamma distribution, the evidence from the likelihood function suggests that the distribution of the new estimators is at least as well-approximated by the normal as is the original.

In conclusion, the recommended method of approaching the problem of finding a parameterization whose estimators are uncorrelated can be summarized as follows:

(1) Decide upon a family of parameter transformations which seem appropriate for the distribution and parameters at hand.
(2) If possible, choose a member of the family which fulfills the criterion (or criteria) concerning a suitable measure of correlation. Some such criteria which have been tried are

(a) zero asymptotic correlation, that is

\[ \frac{\partial^2 L}{\partial \phi_1 \partial \phi_2} = 0 \]

(b) zero expected asymptotic correlation, that is

\[ \mathbb{E}\left( \frac{\partial^2 L}{\partial \phi_1 \partial \phi_2} \right) = 0 \]

(c) each normal equation a function of one estimator only, that is

\[ \frac{\partial L}{\partial \phi_i} = f_i(\phi_i) , \quad i=1,2 \]

If unsuccessful, try another family of transformations.

(3) Check that the distribution of the chosen parameterization is reasonably well-approximated by the normal distribution. This can be done by the methods of Chapter 2 if the exact estimator distribution (or its moments) is known, or by use of the knowledge of the likelihood function as in this chapter if the distribution is not
known. If the parameterization is not satisfactory in this respect, try another member of the same family, if another fulfills one of the criteria for zero correlation, or else try another family.

In the next chapter, this method is applied to a model which is more complex than any considered thus far.
CHAPTER 4

BIOASSAY LOGISTIC MODEL: THEORETICAL DISCUSSION

4.1 Introduction

The two-parameter models commonly used to describe a bioassay with quantal responses have three parameterizations which are in frequent use (one of these actually comprises one parameter from each of the other two) and there is no theoretical evidence that any one is in general superior to the others. In this chapter, these three parameterizations of the logistic model are compared with respect to correlation of estimators and normality of estimator distributions, using the ideas, measures, and techniques developed in the previous two chapters. In addition, attempts are made at finding an alternative parameterization which satisfies some of the criteria for independence discussed in Chapter 3.

4.2 The Logistic Model

The logistic sigmoid is one of the models used
to describe the relationship between a dose and the probability of response at that dose in a quantal response bioassay. Under the logistic model, this relationship is assumed to have the form

\[ P_i = \frac{1}{1 + e^{-Y_i}} \quad (4.21) \]

where \( P_i \) is the true response probability at a dose \( x_i \) and \( Y_i = \alpha + \beta x_i \). (In fact, \( x_i \) is more usually the logarithm of the dose.) Thus

\[ Y_i = \alpha + \beta x_i = \log_e(P_i/(1-P_i)) = \logit(P_i) \]

\( Y_i \) can also be expressed as

\[ Y_i = \frac{x_i - \mu}{\sigma} \quad \text{or} \quad Y_i = \beta(x_i - \mu) \]

where \( \mu \) is the ED50, the dose (or log dose) eliciting the response with probability 0.5. The three parameterizations, \((\alpha , \beta )\), \((\mu , \sigma )\), \((\mu , \beta )\), are conceptually quite different and the one which is actually used in any situation may be determined by the manner in which the dose-response relationship is viewed: \((\alpha , \beta )\) emphasizes that the assumed relationship between log dose and the

\[ 1 \text{ In a quantal response type of bioassay only the presence or absence of a particular specified response can be observed.} \]
transformed response probability is a straight line with slope $\beta$ and that the estimation procedure is simply a weighted linear regression; $(\mu, \sigma)$ suggests an underlying tolerance distribution with mean $\mu$ and variance $\sigma^2$;¹ $(\mu, \beta)$ is made up of the two parameters which are most likely to be of interest in an actual experiment.

4.2.1 Likelihood Function and Normal Equations

In the assay of a single substance utilizing $N$ subjects, there are $k$ doses, $x_1, x_2, \ldots, x_k$, of the substance administered to $n_1, n_2, \ldots, n_k$ subjects, where $\sum n_i = N$, and the numbers of subjects responding to the $k$ doses, $r_1, r_2, \ldots, r_k$, are observed (where a response has been determined before the experiment). If the probability of a subject responding to dose $x_i$ is $P_i$ (where $P_i$ is the function of dose expressed in equation 4.21), then the probability of observing $r_i$ responses out of the possible $n_i$ at this dose is

$$Pr(r_i) = \binom{n_i}{r_i} P_i^{r_i} (1 - P_i)^{n_i - r_i}$$

¹ The tolerance distribution which underlies the logistic model is of little interest. (See Finney (1964), p455, for a brief discussion.) In actual fact, the results obtained by using this model differ little in general from those obtained using the probit model for which the underlying assumed distribution of individual tolerances is normal.
and so the probability of observing the set of responses 
\((r_1, r_2, \ldots, r_k)\) is

\[
Pr(r_1, r_2, \ldots, r_k) = \prod_{i=1}^{k} \binom{n_i}{r_i} p_i^{r_i} (1-p_i)^{n_i-r_i} (4.22).
\]

Substituting for \(p_i\) from 4.21, 4.22 can be written as

\[
\prod_{i=1}^{k} \binom{n_i}{r_i} \frac{\exp(-(n_i-r_i)Y_i)}{[1+e^{-Y_i}]^{n_i}} = \frac{\exp(-\Sigma n_iY_i + \Sigma r_iY_i)}{\prod [1+e^{-Y_i}]^{n_i}} \prod \binom{n_i}{r_i}.
\]

(4.23)

It is evident that \(\Sigma r_iY_i = \alpha \Sigma r_i + \beta \Sigma r_ix_i\) is the only quantity in 4.23 which involves both the parameters and the observations, and so \(\Sigma r_i\) and \(\Sigma r_ix_i\) are sufficient statistics for \(\alpha\) and \(\beta\) (or any other pair of parameters).

Letting \((\theta_1, \theta_2)\) be a general parameterization of the logistic model, the log likelihood function derived from 4.22 is

\[
L(\theta_1, \theta_2) = \Sigma r_i \log P_i + \Sigma (n_i-r_i) \log(1-P_i) (4.24)
\]

and hence, letting \(p_i = r_i/n_i\), the normal equations are

\[
0 = \frac{\partial L}{\partial \theta_j} = \Sigma n_i \frac{p_i}{P_i} \frac{\partial P_i}{\partial \theta_j} - \Sigma n_i \frac{(1-P_i)}{(1-P_i)} \frac{\partial P_i}{\partial \theta_j}
\]
\[ \frac{n_i}{\theta_j} \frac{\partial P_i}{\partial \theta_j} \begin{pmatrix} \frac{P_i - \hat{P}_i}{\hat{P}_i(1 - \hat{P}_i)} \end{pmatrix} \quad j = 1, 2 \quad (4.25) \]

where \( \frac{\partial P_i}{\partial \theta_j} \) is evaluated at \((\hat{\theta}_1, \hat{\theta}_2)\), the maximum likelihood estimates. Since \( P_i \) is non-linear in the parameters, 4.25 does not in general yield explicit algebraic expressions for \( \theta_1 \) and \( \theta_2 \), so that for given data numerical values of the estimates must usually be obtained by application of a numerical search procedure or iterative maximization technique to 4.24. This means that little is known about the exact distribution of the estimators, with the result that asymptotic theory is applied to the distribution of the estimators readily and often uncritically. Dowden (1971) shows how this practice can sometimes lead to incorrect results.

In all subsequent discussion of this model, it is assumed, unless otherwise specified, that \( n_i = n \), \( i = 1, 2, \ldots, k \) and that the \( x_i \) are equally spaced log doses.
4.3 Correlation of Estimators

4.3.1 Comparison of Three Parameterizations of the Logistic Model

As noted in section 4.2.1, explicit expressions for the maximum likelihood estimators in terms of the sufficient statistics are generally not obtainable, which means that the moments of the estimator distributions, including the covariance, are unknown and cannot be estimated from the data. If, however, assumptions of asymptotic normality are made, the asymptotic correlations can be obtained from the information matrices and can be compared for the three pairs of estimators. If $I_{\theta_1, \theta_2}$ represents the information matrix for $(\hat{\theta}_1, \hat{\theta}_2)$, then, with $Q_i = 1 - P_i$,

\[
I_{\alpha, \beta} = \begin{pmatrix}
 n\Sigma P_i Q_i & n\Sigma x_i P_i Q_i \\
 n\Sigma x_i P_i Q_i & n\Sigma x_i^2 P_i Q_i \\
\end{pmatrix}
\]

\[
I_{\mu, \beta} = \begin{pmatrix}
 n\beta^2 \Sigma P_i Q_i & -n\beta \Sigma (x_i - \mu) P_i Q_i \\
-n\beta \Sigma (x_i - \mu) P_i Q_i & n\Sigma (x_i - \mu)^2 P_i Q_i \\
\end{pmatrix}
\]

(4.30)
\[
I_{\mu, \sigma} = \begin{pmatrix}
\frac{n}{\sigma^2} \sum P_i Q_i & \frac{n}{\sigma^3} \sum (x_i - \mu) P_i Q_i \\
\frac{n}{\sigma^3} \sum (x_i - \mu) P_i Q_i & \frac{n}{\sigma^4} \sum (x_i - \mu)^2 P_i Q_i 
\end{pmatrix}
\]

Letting \( A = \sum P_i Q_i \), \( \bar{x} = \frac{\sum x_i P_i Q_i}{A} \), and with \( r_a(\hat{\theta}_1, \hat{\theta}_2) \) representing the asymptotic correlation of \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \), from 4.30 the following can be obtained:

\[
r_a(\hat{\alpha}, \hat{\beta}) = \frac{-\bar{x}}{(\frac{1}{A} \sum x_i^2 P_i Q_i)^{\frac{1}{2}}} = \frac{-\bar{x}}{\left( \bar{x}^2 + \frac{1}{A} \sum (x_i - \bar{x})^2 P_i Q_i \right)^{\frac{1}{2}}}, \quad (4.31)
\]

\[
r_a(\hat{\mu}, \hat{\beta}) = \frac{(\bar{x} - \mu)}{(\bar{x}^2 + \frac{1}{A} \sum (x_i - \bar{x})^2 P_i Q_i)^{\frac{1}{2}}}, \quad (4.32)
\]

\[
r_a(\hat{\mu}, \hat{\sigma}) = -r_a(\hat{\mu}, \hat{\beta}) \quad . \quad (4.33)
\]

From the above it is evident that \( r_a(\hat{\alpha}, \hat{\beta}) \) depends on the size of \( \bar{x} \); while \( r_a(\hat{\mu}, \hat{\beta}) \) is determined by \( (\bar{x} - \mu) \), and that the asymptotic correlation of \( (\hat{\alpha}, \hat{\beta}) \) actually exceeds that of \( (\hat{\mu}, \hat{\beta}) \) in absolute value whenever \( |\bar{x}| > |\bar{x} - \mu| \), while \( r_a(\hat{\mu}, \hat{\beta}) \) is the larger when \( |\bar{x} - \mu| > |\bar{x}| \). It is intuitively obvious that \( \bar{x} \) can be near zero only when the log doses are clustered around zero. However, in the interest of performing
a useful and statistically valid assay, one should choose log doses to be centred on the ED50 (or a guess of the ED50 based on prior knowledge); if this happens to be far from zero, \( \bar{x} \) will also be far from zero and the asymptotic correlation of \( \hat{a} \) and \( \hat{b} \) will be large, approaching its limiting value of 1 as \(|\bar{x}|\) increases. In an assay with doses centred on a guessed value of \( \mu \), the size of \( \bar{x} - \mu \) depends on how good this guess is while the estimated value of \( \bar{x} - \mu \) depends additionally on the data. It seems clear then that in a well-designed assay where there is some prior knowledge about \( \mu \) the estimate of \( \bar{x} - \mu \) is more likely to be close to zero than is the estimate of \( \bar{x} \), unless \( \mu = 0 \) in which case all three parameterizations yield zero asymptotic correlations of estimators. If, however, there is no information about \( \mu \) available before the experiment and log doses are selected to be centred on zero, then the experiment is considered unsuccessful if \( \hat{\mu} \) turns out to be far from the central dose. The conclusion which would be drawn from such results would be either that \( \mu \) was really near 0 and the experimenter was unlucky, or else that the chosen doses were not suitable and another selection of doses suggested by the experimental results should be used
in future experiments. In this case, the fact that \((\hat{\alpha}, \hat{\beta})\) might well be less highly correlated than \((\hat{\mu}, \hat{\beta})\) is irrelevant.

Additional evidence in favour of the \((\mu, \beta)\) parameterization (or, equivalently, the \((\mu, \sigma)\) parameterization) with respect to asymptotic correlation of estimators is provided by Anscombe (1964). In this paper, he is actually concerned with orthogonalizing the likelihood function and therefore choosing that transformation of parameters which makes the mixed second derivative of the likelihood function evaluated at the maximum likelihood estimates vanish.\(^1\) Anscombe shows that in an 'ideal' experiment with all \(n_i = n\) and doses which are equally spaced with \(x_{i+1} - x_i = h\) \((i = 1, 2, \ldots, k-1)\) not too large and which cover such a range that at all doses larger than \(x + (k-1)h\) a 100% response would be obtained and at all doses less than \(x\) a response of 0% would be obtained, then

\[ \frac{\partial^2 L}{\partial \mu \partial \beta} \bigg|_{\hat{\mu}, \hat{\beta}} = 0, \]

and so \(\hat{\mu}\) and \(\hat{\beta}\) have first order orthogonality.\(^2\) It can be shown that \(\hat{\alpha}\) and \(\hat{\beta}\)

\(^1\) In fact for the logistic model this criterion is equivalent to zero estimated asymptotic correlation.

\(^2\) Anscombe applies the Euler-Maclaurin theorem to

\[ \frac{\partial^2 L}{\partial \mu \partial \beta} \bigg|_{\hat{\mu}, \hat{\beta}} = n \beta \Sigma (x_i - \hat{\mu}) P_i \delta_i. \]

This theorem can also be applied to \(E\{ \frac{\partial^2 L}{\partial \mu \partial \beta} \}\) with the same results.
do not possess this property. Although such an ideal experiment is not usually possible, approximations of varying degrees are frequently obtainable and the above results provide some encouragement for use of the \((\mu , \beta )\) parameterization in such cases.

In summary, then, it does seem that of the three parameter pairs in frequent use with this model \((\mu , \beta )\) or \((\mu , \sigma )\) are superior to \((\alpha , \beta )\) with respect to the measure of correlation considered in this section.

4.3.2 Reparameterizations for Independence

Although \((\mu , \beta )\) is generally superior to \((\alpha , \beta )\) as far as asymptotic correlation of estimators is concerned, there may be some other pair of parameters which is better still in this regard. It is possible of course that there is no parameterization which satisfies the criterion of zero asymptotic correlation and which can be obtained from any simple family of transformations which one might consider. Nevertheless, following the method for finding a parameterization outlined in section 3.6, a few families of non-linear transformations of \((\mu , \beta )\) are specified below and those members, if any, satisfying the criterion
of zero asymptotic correlation are selected.

In general, any set of transformations from \((\mu, \beta)\) to \((\phi_1, \phi_2)\) can be written as

\[
\begin{align*}
\mu &= g_1(\phi_1, \phi_2) \\
\beta &= g_2(\phi_1, \phi_2)
\end{align*}
\]

If \(g_{ij} = \frac{\partial g_i}{\partial \phi_j}\), \(i,j = 1,2\) and if \(L\) represents the log of the likelihood function, then by application of the chain rule

\[
\frac{\partial L}{\partial \phi_j} = \frac{\partial L}{\partial \mu} g_{1j} + \frac{\partial L}{\partial \beta} g_{2j}, \quad j = 1,2
\]

and further, letting \(L_{\theta_i \theta_j} = -E\{\partial^2 L / \partial \theta_i \partial \theta_j\}\) the elements of the information matrix of \((\hat{\phi}_1, \hat{\phi}_2)\) are

\[
-E\{\partial^2 L / \partial \phi_i \partial \phi_j\} = L_{\phi_i \phi_j}
\]

\[
= L_{\mu \mu} g_{1i} g_{1j} + L_{\mu \beta} (g_{1i} g_{2j} + g_{1j} g_{2i}) \\
+ L_{\beta \beta} g_{2i} g_{2j}, \quad i,j = 1,2
\]  
(4.34)
and $L_{\mu \mu}$, $L_{\mu \beta}$, and $L_{\beta \beta}$ are the known elements of $I_{\mu, \beta}$ in 4.30. Any pair of functions $g_1, g_2$ which makes 4.34 zero obviously satisfies the criterion of zero asymptotic correlation.

The first specific family which is considered is one which includes the transformation from $\beta$ to $\gamma = \log \beta$.

$$\mu = g_1 = \phi_1^i \phi_2^j$$

$$\beta = g_2 = \lambda \phi_1^k e^{\phi_2^2}$$

From 4.34

$$L_{\phi_1^i \phi_2^j} = \lambda^2 n \phi_1^{2i-1} e^{2\phi_2} \left[ A \phi_1^{2i} \phi_2^{2j-1} (ij+i^2 + jk + k \phi_2) 
- A \phi_1^{i} \phi_2^{j-1} x (i \phi_2 + jk + 2k \phi_2) + kAX^2 
+ k \Sigma (x_i - \bar{x})^2 p_i q_i \right]$$

When $k = 0$ many terms involving $\bar{x}$ vanish and 4.35 simplifies to

$$L_{\phi_1^i \phi_2^j} = \lambda^2 nA \phi_1^{i-1} \phi_2^{j} e^{2\phi_2} \left[ \phi_1^i \phi_2^{j-1} (j + \phi_2^2) - \bar{x} \right]$$

1 This transformation is of particular interest because it is suggested by Anscombe (1964) as satisfying a certain criterion of normality. See section 4.4.
which is in general zero only when
\[ \phi_1 \phi_2 j^{-1} (j + \phi) = \bar{x}, \quad (4.36) \]
since \( \lambda \) cannot be zero and \( i \) and \( k \) cannot simultaneously be zero. Clearly there are no values of \( i \) and \( j \) which satisfy 4.36 independently of particular data. However if from a given set of data estimates \( \hat{\mu}, \hat{\beta} \) and \( \hat{x} \) are obtained, \( L_{\phi_1 \phi_2} \) evaluated at \( (\hat{\mu}, \hat{\beta}) \) will be zero (with \( k = 0 \)) if \( \hat{\phi}_1 = 0, i > 1 \) or \( \hat{\phi}_2 = 0, j > 0 \), or if 4.36 is satisfied at \( (\hat{\phi}_1, \hat{\phi}_2) \), that is, if
\[ \hat{\mu} \left\{ j / \log \hat{\beta} + 1 \right\} = \hat{x}, \]
which happens when \( j = \frac{\hat{x} - \hat{\mu}}{\mu} \log \hat{\beta} \), for all \( i \neq 0, \hat{\mu} \neq 0 \).

The second family of transformations considered comprises power transformations of both parameters such that
\[ \mu = g_1 = \phi_1 \phi_2 \]
\[ \beta = g_2 = \phi_1 \phi_2 \]

\[ (4.36) \]
In this case

\[ L_{\phi_1\phi_2} = n\phi_1^{2k-1}\phi_2 2^l 2^l - 1 \left( A\phi_1^{2i}\phi_2 2^j (ij + i\ell + jk + \ell k) \right. \]

\[ - Ax\phi_1^{i}\phi_2^{j} (i\ell + jk + 2\ell k) \]

\[ \left. + \ell k (Ax^2 + \Sigma (x_i - \bar{x})^2 Q_1) \right) \]  \hspace{1cm} (4.37)

As with 4.35, 4.37 simplifies when \( k = 0 \) to

\[ L_{\phi_1\phi_2} = niA\phi_1^{i-1}\phi_2 2^l + j - 1 \left( \phi_1^{i}\phi_2^{j} (j + \ell) - \bar{x}\ell \right) \]

which is in general zero only when \(^1\)

\[ \phi_1^{i}\phi_2^{j} (j + \ell) = \bar{x}\ell \]  \hspace{1cm} (4.38)

and there are no non-trivial values of \( i, j, \) and \( \ell \) for which 4.38 will always hold. If maximum likelihood estimates are substituted for the parameter values, \( L_{\phi_1\phi_2} | \hat{\phi}_1 \hat{\phi}_2 \) will be zero (with \( k = 0 \)) if \( \hat{\phi}_1 = 0, i > 1 \), or \( \hat{\phi}_2 = 0, 2\ell + j > 1 \), or if

\(^1 \)k and i cannot simultaneously be zero; nor can k and \( \ell \), or \( \ell \) and j, or i and j.
\[ \hat{\mu}(j + \ell) = \hat{x} \ell \]

which occurs when \( j/\ell = \hat{x}/\hat{\mu} - 1 \), for all \( i \), \( \hat{\mu} \neq 0 \).

If \( \ell \) rather than \( k \) is chosen to be zero in 4.37, similar relationships involving \( i \), \( j \), and \( k \) are obtained.

Thirdly the set of transformations leaving \( \mu \) unaltered but replacing \( \beta \) by some unspecified function of \( (\mu, \phi) \) is briefly considered. The purpose of looking at such parameterizations is to see if there is a transformation of \( (\mu, \beta) \) which has the desired property of independent estimators while retaining the parameter \( \mu \), often of considerable interest in an assay experiment. Although this situation will not be discussed in detail here, mention should be made of the fact that the conclusions reached were the same as for the other two transformation families considered, namely that there is no member of the family which consistently satisfies the criterion of zero asymptotic correlation independently of data.

The conclusion evident from consideration of these three families is that there is not likely to be a simple parameter transformation which satisfies the criterion for independence being applied here and which is not dependent on data. It might be interesting to examine the effects
on the estimated asymptotic correlation of some of the data-dependent transformations suggested in this section. This will not be done here, however, since the primary aim at present is to find a parameter transformation which is independent of the data and so can be effected before iteration takes place.
4.4 Normality of Estimator Distributions

As noted previously, the exact distributions of the maximum likelihood estimators are unknown and there are no explicit expressions for their exact moments; asymptotically, the distributions of all estimator pairs are normal. Thus, the only ways in which certain aspects of the estimator distribution for various parameterizations can be compared are through properties of the likelihood function or through construction of the exact distributions for some parameter values. In this section the likelihood functions are considered; exact constructed distributions are discussed in depth in Chapter 5.

In Chapters 2 and 3 there is some discussion of the ways in which likelihoods can be used to provide some information about the validity of assumptions of asymptotic normality; their limitations are also discussed in these chapters and need not be mentioned again here. There are two aspects of the likelihood function which can prove useful in assessing the distributions of estimators: the third derivative of the log likelihood function evaluated at the maximum likelihood estimates can be examined as a measure of departure from symmetry in the region of the
maximum; also, the shape of the contours of constant relative likelihood\(^1\) of various parameterizations for given data can be compared with each other and with elliptical normal contours. In order to consider the latter of these two, contours of constant relative likelihood of \((\alpha, \beta)\), \((\mu, \beta)\) and \((\mu, \sigma)\) are shown in Figures 4.1 a–c for a 3-dose assay, log doses \(-1, 0, 1\), with \(n = 10\) subjects at each dose, and with the observed proportions of subjects responding \((0.3, 0.5, 1.0)\); the estimated asymptotic standard errors for all maximum likelihood estimates are also given in the Figures. The data set employed was chosen deliberately to be asymmetric about 0.5, and results in contours identical to those for other data sets with the same sufficient statistics. One interesting point evident from Figure 4.1b is that the axes of contours for \((\mu, \beta)\) appear almost parallel to the parameter axes, an illustration perhaps of the first order orthogonality of \(\hat{\mu}\) and \(\hat{\beta}\); the axes of the \((\alpha, \beta)\) contours look more oblique to the parameter axes, which indicates that \((\hat{\alpha}, \hat{\beta})\) are probably more highly correlated than \((\hat{\mu}, \hat{\beta})\), at least for this

\(^1\) The relative likelihood of \((\theta_1, \theta_2)\) is the likelihood of this pair of parameter values relative to the likelihood function evaluated at \((\hat{\theta}_1, \hat{\theta}_2)\).
CONTOURS OF CONSTANT RELATIVE LIKELIHOOD* FOR FOUR PARAMETERIZATIONS OF THE LOGISTIC MODEL, LOG DOSES \(-1, 0, 1\), n=10, OBSERVED \(p=(0.3, 0.5, 1.0)\)

* 0.05, 0.25, 0.5 contours are shown.
Standard errors quoted are estimated asymptotic standard errors.
4.20

data. Contours for both \((\mu, \beta)\) and \((\alpha, \beta)\) approximate
to concentric ellipses quite well in the region of the
maximum and in fact in the region including two standard
deviations from the maximum likelihood estimates. Contours
for \((\mu, \sigma)\) on the other hand are not concentric, being
much closer together at small values of \(\sigma\) than at larger
ones, which is because \(\sigma\) is constrained to be non-zero.
This fact suggests a transformation to \(\gamma = -\ln \sigma\), for
\(\gamma\) is not so constrained. Anscombe (1964) suggests that
in his ideal experiment described in section 4.3.1, a
transformation to \((\mu, \gamma = \ln \beta)\) achieves some degree of
normality of the likelihood function in the region of the
maximum. Assuming it is desirable to leave \(\mu\) unaltered
because of particular interest in it, he has selected
the parameter \(\gamma\), a function of \(\beta\) only, such that
\(\frac{\partial^3 L}{\partial \gamma^3} \) is approximately zero at \((\hat{\mu}, \hat{\gamma})\). Contours for
\((\mu, \gamma = \ln \beta)\) are illustrated in Figure 4.1d for the same
set of data as used for the contours of other parameter
pairs. The asymptotic correlation of \((\hat{\mu}, \hat{\gamma})\) is the same
as that of \((\hat{\mu}, \hat{\beta})\), so that this transformation neither

1 The asymptotic correlations estimated from this data set
are \(r(\alpha, \beta) = 0.30\) and \(r(\mu, \beta) = 0.17\).

2 All third derivatives except \(\frac{\partial^3 L}{\partial \mu^2 \partial \gamma}\) are approximately
zero at \((\mu, \gamma)\). Anscombe suggests a "modification" of \(\gamma\)
which improves orthogonality and makes all third derivatives
approximately zero. Specification of this and a discussion
of drawbacks to its use are found in his paper.
improves nor makes worse this aspect of the asymptotic estimator distributions. Contours for \((\mu, \gamma)\) are, like those for \((\mu, \beta)\) and \((\alpha, \beta)\), reasonably concentric and elliptical in the area including two standard deviations about the estimates, but they do not appear any better than those of \((\mu, \beta)\) in this respect. Thus, for this data set at least, Figures 4.1a-d offer no evidence in favour of any one of the four parameterizations under discussion, although \((\mu, \sigma)\) might be said, on the basis of this limited information, to be a poor choice.
4.5 Conclusions

In the logistic model, as in many other models, explicit expressions for the estimators of the parameters cannot in general be obtained, which means that neither the moments nor the density functions of the exact estimator distributions are known. The result is that only properties of the asymptotic distribution and the likelihood function can be used to compare the three common parameterizations of the logistic model and to provide criteria for the selection of a parameterization. On the basis of such criteria as are available and applicable to this case, \((\mu, \beta)\) seems to be the best of the three usual parameterizations; \((\mu, \beta)\) and \((\mu, \sigma)\) are equivalent as far as asymptotic correlation is concerned and appear to be superior to \((\alpha, \beta)\) in this respect, while \((\mu, \beta)\) appears superior to \((\mu, \sigma)\) from considerations of normality of likelihood, although evidence of this is insufficient to provide a basis for any firm statements.

Attempts to reparameterize for independence proved fruitless since a family of transformations which included a parameterization satisfying a criterion for independence of estimators independently of data could not be found.
Some transformations are suggested which are dependent on data but these are not further assessed.

A transformation which improves normality of the likelihood is suggested but the contours of constant relative likelihood for this parameterization, \((\mu, \gamma = \ln \beta)\), do not appear to approximate more closely those of a normal likelihood than do the contours of the original parameterization, \((\mu, \beta)\), at least not for the data considered. In addition, this parameterization has the same asymptotic correlation of estimators as does \((\mu, \beta)\) so does not result in any improvement in this property.

It seems that there is no obvious reparameterization which satisfies criteria for independence without regard to the data and that \((\mu, \beta)\) is probably as good a parameter pair as any. Because of the inconclusiveness of the discussion in this chapter, the next chapter is addressed to the same questions as are discussed herein, but attempts to answer them are based on comparisons of exact estimator distributions constructed for several parameter values.
5.1

CHAPTER 5

BIOASSAY LOGISTIC MODEL : EXACT DISTRIBUTIONS OF MAXIMUM LIKELIHOOD ESTIMATORS

5.1 Introduction

Attempts to compare the asymptotic properties of the distributions of the maximum likelihood estimators for various parameterizations of the bioassay logistic model proved futile: not only was it impossible to find a parameterization with zero, or even minimal asymptotic correlation but it was also found that no general comparative statements could be made about the asymptotic correlations of the usual parameterizations.

Clearly then, some other basis of comparison is needed. Unfortunately explicit algebraic expressions for the maximum likelihood estimators of the parameters of the logistic model do not in general exist; even when such expressions do exist, as they do for some specific configurations\(^1\) or for some particular data, they are usually complicated

\(^1\) A configuration will be considered as comprising a set of experimental doses, the number of subjects at those doses, and a pair of true parameter values. (These then determine the true response probabilities which are therefore also part of a configuration.)
non-linear functions of the sufficient statistics so that derivations of exact theoretical densities or moments are impractical if not impossible. However, the actual estimator distributions for any given configuration can be constructed, either approximately or exactly. For the approximate construction a random sample of possible experimental outcomes must be chosen\(^1\) and maximum likelihood estimates calculated from each sample outcome. Then for each estimator a frequency histogram or polygon can be drawn, using appropriate groupings of estimates and their relative frequencies. One would expect this figure to approximate the exact distribution.

Since the actual estimator distributions are discrete, they can be given exactly in the form of a complete enumeration of the estimates and their probabilities. For each estimator the estimates can then be grouped and the distribution displayed as a frequency polygon. Also, the properties of the distribution can be calculated directly from the enumeration.

Because the approximate method of construction is incomplete, any conclusions drawn from distributions so

\(^1\) The method of sampling must be determined from the known properties of the model and configuration. For the details of such a method see Berkson (1955).
constructed will be less reliable than conclusions drawn from exactly constructed distributions. On the other hand, the exact construction requires a greater amount of calculation, an important factor if the number of subjects is large and use of an iterative technique to obtain estimates is necessary. Wilson and Worcester (1943) have derived explicit expressions for the maximum likelihood estimators when the logistic model is assumed and there are exactly 3 equally spaced doses with the same number of subjects, say \( n \), at each dose. Therefore, for any configuration of this type with reasonably small \( n \) (\( n \leq 15 \), say) all values of the estimators can easily be calculated, particularly if use is made of the sufficient statistics. It is a relatively simple task to compute the probabilities of the estimates, thereby completing the description of the distribution. For the purposes of this chapter I feel that it is important to look at some exact distributions and I think it is reasonable to restrict the discussion in the first instance to exact distributions which can be easily obtained by the application of the Wilson-Worcester method.

The work contained in this chapter really constitutes a pilot study. As in Chapter 4 the logistic model is assumed as the expression of the relationship between the dose and the proportion of subjects responding; that is
\[ p_i = \frac{1}{1 + e^{-(\alpha + \beta x_i)}} = \frac{1}{1 + e^{-(x_i - \mu)/\sigma}} \] (5.11)

For a few different configurations exact estimator distributions are constructed for \( \hat{\alpha}, \hat{\beta}, \hat{\mu}, \) and \( \hat{\sigma} \) and some of their properties examined, but it is evident that many more configurations than could possibly be included herein, because of practical considerations, such as limitations on computer time, need to be considered before any conclusions can be drawn. Also it would obviously be desirable to eventually include some other parameterizations in such a comparative study.

In the following pages the difficulties involved in the study of discrete exact distributions as opposed to continuous asymptotic ones are discussed. Then the method of construction is outlined and for some specified configurations the exact estimator distributions are obtained. The bivariate distributions for some estimator pairs are summarised by their moments and functions of their moments and, in some instances, by frequency polygons, and are compared for different estimator pairs with respect to some of these properties.¹

¹ Certain of these properties and comparisons are considered only for the marginals of the bivariate distributions of \((\hat{\alpha}, \hat{\beta})\) and \((\hat{\mu}, \hat{\sigma})\).
5.5

5.1.1 Problems with Constructed Distributions

There are some problems in dealing with the constructed estimator distributions and although their use is superior to use of asymptotic distributions in many respects, the greater difficulties encountered must not be overlooked.

Firstly, the exact estimator distributions are all discrete and so visually can at best be represented by frequency polygons or histograms which are not usually easy to compare. This problem of discreteness is especially acute for small values of \( n \) (for example, \( n = 5 \)) when there is only a small number of points with non-zero probability. In cases such as these, any visual comparisons are virtually meaningless. Of course, as \( n \) gets larger the difficulty decreases, since the number of points with non-zero probability increases to the point where the distribution can for most purposes be considered to be continuous. Asymptotic theory assumes continuity, of course, with the result that distributions are easy to visualize, work with, and compare, although they may well be very inaccurate for small values of \( n \).

Secondly, an exact distribution obtained by construction is dependent on a particular underlying configuration and cannot be generalized to other configurations. At best, one can systematically vary some component(s) of the
configuration and note resulting changes in the estimator distributions.

Thirdly, some permissible observations (experimental outcomes) produce indeterminate or infinite estimates for some parameters. Although these observations do not convey much information about the nature of the response curve, they do not always have negligible probability (in particular, the sum of the probabilities of all such observations for one pair of estimators can be sizeable) and so cannot simply be ignored. Thus some special consideration must be given to the estimates resulting from these observations. Either a method could be devised for the inclusion of such estimates, perhaps in a somewhat altered form, such as the $1/2n$ rule of Berkson (1955); or else perhaps the distribution conditional on the estimate being defined and finite could be studied. Once a procedure for dealing with these estimates has been decided upon, the observations producing them must be determined so that their probabilities can be computed.\footnote{The observations producing infinite or indeterminate estimates for various parameters are described in some detail in Appendix II; methods for dealing with these estimates are also discussed in more depth there.} Another problem here is that since we are dealing with pairs of parameters, it is quite possible to have the estimate of one parameter defined and the other
infinite or indeterminate, which means that if one were interested in marginal distributions a decision would have to be made as to whether to look at the marginal conditioned on its variate only being defined or else at the marginal of the bivariate distribution conditioned on both its variates being defined, since these will differ in some instances. As the number of subjects per dose increases the probability of indeterminate or infinite estimates for any configuration decreases (at least for a reasonable dose set).

One method of comparing a distribution with the normal distribution is of course, by eye. While this type of comparison is not at all rigorous it gives some idea of what sorts of differences exist between two distributions and where these occur. However, in order to be able to draw any conclusions about the exact distributions for different parameterizations, many frequency curves for several configurations must be plotted and comparisons made: the situation for the various parameterizations of this model cannot be summarized by a few theoretical densities and their theoretical moments; even measures such as skewness and kurtosis have to be calculated numerically for each different configuration and parameterization.
5.8

Lastly, the larger the value of $n$ the more possible values of estimators there are and the more work there is involved in constructing distributions.

Despite these difficulties, the exact distributions of the maximum likelihood estimators of the parameters of the bioassay logistic model have been constructed by complete enumeration for a few configurations of the type for which the Wilson-Worcester method applies, and certain interesting aspects of these distributions have been studied and some between-parameterization comparisons made. The rest of this chapter then is concerned entirely with the construction and utilisation of these distributions.¹

5.2 Construction of the Exact Distributions

If the method of Wilson and Worcester is to be used to obtain the maximum likelihood estimates, consideration must be limited to configurations of three equally spaced doses with $n$ subjects at each. Although within this framework any configuration could be used, it would be especially interesting to study the distributions arising from configurations which differ from one another

¹ In this chapter I deal only with maximum likelihood estimators; I do not attempt to compare different methods of estimation. Such comparisons are dealt with in various ways by Berkson (1955), Silverstone (1957) and Cramer (1964).
in some systematic way in order to get some idea of the effects on the distributions of such changes in the configurations.

If \((P_1, P_2, P_3)\) are the true response probabilities for the set of log doses \((x_1, x_2, x_3)\), a desired change in the \(P_i\) can be effected either by an appropriate linear transformation of the \(x_i\) with no change in the parameters, or by a linear transformation of the parameters with no change in the doses.\(^1\) In a similar fashion other desired changes can be accomplished. In his study, Cramer (1964) fixed the parameters at convenient values and changed the central dose and dose spacing at will, forcing changes in the \(P_i\), in order to obtain the configurations he wished to study. In this chapter, for reasons of computational convenience the log doses are fixed at \((-1, 0, 1)\), and the central dose probability and one parameter are altered so that sets of configurations which can possibly provide some useful information are obtained.

Once a value of \(n\) has been specified, all possible values of the estimators can be computed. There are

\(^1\) In fact, because doses are equally spaced, only two of the three \(P_i\) can be changed at will, so that such a pair uniquely defines the configuration.
(n+1)^3 possible outcomes of the assay and one could laboriously calculate the set of estimates from each of them individually; however, the existence of a pair of sufficient statistics means that such a procedure is unnecessary, since for any fixed $n$ the possible values of the sufficient statistics are all that are needed for the complete enumeration of the estimates of all parameters. The probability associated with any given set of estimates is determined by the probabilities of the observations producing these estimates, and the values of these probabilities are dependent on the $P_i$. Therefore for each set of response probabilities, $(P_1, P_2, P_3)$, the probability of any pair of sufficient statistics is obtained by the summation of the probabilities of all those observations with the given sufficient statistics. (This probability is, of course, that associated with the set of estimates corresponding to these sufficient statistics.)

In theory, the progression from observations to estimates to probabilities is simple and straightforward; in practice, it is not quite so, particularly if one wishes to be systematic and efficient. Firstly there are those samples which produce indeterminate or infinite estimates and so must be identified and omitted from routine calculations to be dealt with separately in some determined manner.
Secondly, one must ensure that all possible pairs of sufficient statistics are enumerated and that all samples producing any one of them have their probabilities included in the calculation of the probability of the resultant pair of estimates. Thirdly, one must try to use the symmetry in the relationship between estimates and sufficient statistics in order to make calculations as efficient as possible in terms of time and labour.

It is obvious that systematic enumeration of the pairs of sufficient statistics is essential since each pair yields a set of estimates. Also an algorithm is needed which, for each pair of sufficient statistics, gives all samples with these statistics; the probabilities of these samples must then be computed and summed to give the probability of the pair of estimates obtained from these statistics. The following subsections are concerned with the development of the algorithms necessary for the efficient construction of the distributions.

5.2.1 Calculation of Estimates: the Wilson-Worcester Method

In general, the sufficient statistics for parameters $(\theta_1, \theta_2)$ of the logistic model for a quantal response type $k$ of bio-assay are $(\sum_{i=1}^{k} n_i p_i, \sum_{i=1}^{k} n_i x_i p_i)$ where $(x_1, x_2, \ldots, x_k)$ are the doses administered (or their logs),
(p_1, p_2, \ldots, p_k) are the observed proportions responding to the doses, and (n_1, n_2, \ldots, n_k) are the numbers of individuals subjected to the doses. For an assay with three log doses, \((x-c, x, x+c)\), where \(c\) is a constant, and \(n\) subjects at each dose, the sufficient statistics are simply \((\sum \ p_i, \ p_3-p_1)\). Letting \(D = \sum p_i\), \(B = p_3-p_1\), and \(A = 2D - 3\), and assuming the relationship between dose and response to be as in 5.11, Wilson and Worcester (1943) show the maximum likelihood estimators to be

\[
\hat{\beta} = \frac{2}{c} \tanh^{-1} C
\]
\[
\hat{\mu} = x - \frac{2}{\beta} \tanh^{-1} x
\]
\[
\hat{\alpha} = 2 \tanh^{-1} x - \beta x
\]

where \(X\) is the root of the cubic equation

\[
f(X) = 3X^3 - 4AX^2 + (6A^2 - 4B^2)X - 2A = 0 \quad (5.22)
\]

and

\[
C = \frac{2B}{X^2 - AX + 2} \quad (5.23)
\]

A good approximation to the root of 5.22 is given by Wilson and Worcester as

\[1\] The method of Wilson and Worcester is given in more detail in Appendix I.
\[
X_0 = \frac{4A}{9} - \frac{b}{a} + \frac{729}{a} (b/a)^3
\]

where

\[
a = 1458 - 189A^2 - 972B^2
\]
\[
b = A(162 - 20A^2 - 432B^2).
\]

However, for computational convenience, a slightly different but equivalent procedure is followed in that the root of

\[
f(Y) = \frac{1}{3} f(X - \frac{4A}{9}) - Y^3 + \frac{a}{729}Y + \frac{b}{729} = 0 \quad (5.22a)
\]

is approximated by

\[
Y_0 = -\frac{b}{a} + \frac{729}{a} (b/a)^3
\]

and if \(|f(Y_0)| < 10^{-6}\), \(Y_0\) is accepted as the root of 5.22a; if \(|f(Y_0)| \geq 10^{-6}\) the Newton-Raphson iterative technique is used to get a solution closer to the true root, with the initial approximation to the root being taken as \(Y_1 = -\frac{b}{a}\). \(f(Y)\) and the first derivative of \(f(Y)\) with respect to \(Y\) are evaluated at \(Y_1\) and the next approximation to the root is

\[
Y_2 = Y_1 - \frac{f(Y_1)}{d_1}
\]

where

\[
d_1 = \frac{df(Y)}{dY} \bigg|_{Y=Y_1} = 3Y_1^2 + \frac{a}{729}
\]
In general, then, in the \((i+1)\) th iteration the root is approximated by

\[
Y_{i+1} = Y_i - \frac{f(Y_i)}{d_i}
\]

and if \(|Y_{i+1} - Y_i| < 10^{-6}\) then \(Y_i\) is accepted as the root of 5.22a. If after twenty iterations convergence has still not been obtained, a message to this effect is printed by the computer. In the studies conducted in the present chapter this never occurred.

Once the root of 5.22a has been obtained as, say, \(Y\), then the root of 5.22 is obviously \(X = Y + 4A/9\), which is the value used in the expressions for the maximum likelihood estimators, 5.21.

When log doses are \((-1, 0, 1)\), as is assumed throughout this chapter, equations 5.21 simplify to

\[
\begin{align*}
\hat{\beta} &= 2 \tanh^{-1} C \\
\hat{\mu} &= -\frac{2}{\hat{\beta}} \tanh^{-1} X \\
\hat{\alpha} &= 2 \tanh^{-1} X 
\end{align*}
\]

(5.21a)
5.2.2 Enumeration of Sufficient Statistics

In the development of an algorithm to efficiently find all possible distinct pairs of sufficient statistics, three points should be noted. Firstly, the constraints on the possible values of \((D,B)\) imposed by the nature of the relationship between \(D\) and \(B\) and by the nature of the observations of which they are functions must be recognised. Secondly, some pairs produce indeterminate or infinite estimates and so must be identified and handled separately. Thirdly, because of the symmetry in the relationships between sufficient statistics and estimators (cf equations 5.23, 5.24, 5.21a) not all pairs of sufficient statistics are needed in order that all pairs of estimates be obtained.

The constraints on the values of \((D,B)\) are considered first. The set of values which \((D,B)\) may take is limited both by the relationship between \(D\) and \(B\) and by the nature of the observations, which are discrete and finite in number. For any observed triple of proportions responding, \((p_1, p_2, p_3)\), \(D = \sum p_i\) and \(B = p_3 - p_1\), and so both \(D\) and \(B\) must be of the form \(k/n\), where \(k\) is integral; they must also satisfy \(0 \leq D \leq 3\) and \(-1 \leq B \leq 1\); finally, for any given value of \(B\), say \(B'\), \(D\) is constrained to lie in the
5.16
closed interval $[|B'|, 3 - |B'|]$ . However, as shown in Appendix II, at least one element of the pair $(\hat{\alpha}, \hat{\beta})$ is infinite when $|B| = 1$, regardless of the value of $D$, or when $D = |B|$ or $D = 3 - |B|$ for any value of $B$.

The matter of symmetry requires first proof of the presence of such relationships between sufficient statistics and estimators and then a description of these relationships. Since $\tanh^{-1}(-y) = - \tanh^{-1} y$, it is evident from equations 5.23 and 5.21a that a change in sign of $X$ or $C$ results at most in changes in sign of the estimators. Since $X$ and $C$ are functions of the sufficient statistics, the effects on $X$ and $C$ of reflecting values of $D$ and $B$ about their midvalues (0 for $B$ and 3/2 for $D$) can be examined. If it can be shown that such changes in $D$ or $B$ result in at most changes in sign of $X$ or $C$, then a symmetric relationship between sufficient statistics and estimators has been established.

When the root of 5.22 is acceptably approximated by $X$ as expressed in equation 5.24, it is easy to derive the symmetric relationships between sufficient statistics and $X$ and $C$, noting that a change from $D$ to $3 - D$ is equivalent to the change from $A$ to $-A$, since $A = 2D - 3$ . The effects of the reflections of the
sufficient statistics as mentioned above on both X and C are summarised in the following table:\(^1\)

<table>
<thead>
<tr>
<th>SUFFICIENT STATISTICS</th>
<th>ESTIMATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>D B</td>
<td>X C</td>
</tr>
<tr>
<td>D' B'</td>
<td>X' C'</td>
</tr>
<tr>
<td>D' -B'</td>
<td>X' -C'</td>
</tr>
<tr>
<td>3 - D' B'</td>
<td>-X' C'</td>
</tr>
<tr>
<td>3 - D' -B'</td>
<td>-X' -C'</td>
</tr>
</tbody>
</table>

It remains to be shown then that these relationships hold even when the root of 5.22 is obtained by iteration. This entails proof that the tabulated changes in a given pair of sufficient statistics produce the same changes in X and C as those presented in the table. If X' is the root of 5.22 obtained by iteration for the pair of sufficient statistics \((D',B')\) and \(Y_i'\) is the \(i\)th approximation to the root of 5.22a, then the \(i\)th approximation to this root for the pair \((D',-B')\) is also \(Y_i'\) and so the root of 5.22 for \((D',-B')\) is also X'. For \((3-D',B')\), however, the \(i\)th approximation to the root of 5.22a is \(-Y_i'\) and the root of 5.22 is therefore \(-X'\). Thus the same relationships exist between the sufficient statistics and X whether X is obtained as in

\(^1\) The entries in this table are valid only for log doses \(-1, 0, 1\).
5.24 or by iteration, and so the symmetries cited in the above table hold for all pairs of sufficient statistics.

The conclusion to be drawn from this discussion is that from the admissible values of \( B > 0 \) and \( D \leq 3/2 \) all estimates can be calculated if the relationships in the above table are utilised.

Considering all the information presented in this section, it is evident that if for each admissible value of \( B \) such that \( 0 \leq B < 1 \), all admissible values of \( D \) satisfying \( B < D \leq 3/2 \) are listed, then by use of the symmetry, all finite pairs of estimates \((\hat{\alpha}, \hat{\beta})\) can be calculated from them. (Here an admissible value of a sufficient statistic is one that is of the form \( k/n \), where \( k \) is integral.)

5.2.3 Enumeration of Samples and Calculation of Probabilities

In the previous section was outlined an efficient method of obtaining all finite pairs of maximum likelihood estimates of \((\alpha, \beta)\) for an assay with log doses \((-1, 0, 1)\), \( n \) subjects at each, under the assumption that the logistic model adequately describes the physical situation. The next step is to enumerate for each of the possible pairs of sufficient statistics all the samples producing the given pair of statistics.
Before attempting to develop an algorithm for this enumeration, it might be helpful to examine a complete list of the samples which yield a particular value for one of the sufficient statistics. These samples can be classified according to the value of the other sufficient statistic. For example, with $n=10$ and $B = 0.7$ such a listing of samples would be:

<table>
<thead>
<tr>
<th>$D$</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>$(0, 1, 0.7)$</td>
</tr>
<tr>
<td>0.9</td>
<td>$(0, 2, 0.7); (1, 0, 0.8)$</td>
</tr>
<tr>
<td>1.0</td>
<td>$(0, 3, 0.7); (1, 1, 0.8)$</td>
</tr>
<tr>
<td>1.1</td>
<td>$(0, 4, 0.7); (1, 2, 0.8); (2, 0, 0.9)$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>$(0, 7, 0.7); (1, 5, 0.8); (2, 3, 0.9); (3, 1, 0.8)$</td>
</tr>
<tr>
<td>1.5</td>
<td>$(0, 8, 0.7); (1, 6, 0.8); (2, 4, 0.9); (3, 2, 0.9)$</td>
</tr>
<tr>
<td>1.6</td>
<td>$(0, 9, 0.7); (1, 7, 0.8); (2, 5, 0.9); (3, 3, 0.9)$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>$(3, 9, 1.0)$</td>
</tr>
</tbody>
</table>

It is evident from the above that in general the samples yielding a given admissible pair of sufficient statistics $(D, B)$ are of the form

$$(i/n, D - B - 2i/n, B + i/n)$$

(5.25)
where $i$ is integral. Using the fact that all three elements of a sample must lie between 0 and 1, the range of $i$ for any given $(D,B)$ can be determined and the permissible samples listed. The restrictions on sample elements thus require that all three of the following inequalities hold:

\begin{align*}
0 & \leq i \leq n & (5.26) \\
0 & \leq D - B - 2i/n \leq 1 & (5.27) \\
0 & \leq B + i/n \leq 1 & (5.28)
\end{align*}

Because of symmetry, discussion can be confined initially to values of $B \geq 0$ and $D \leq 3/2$, which means that $B + i/n \geq 0$ will be true whenever 5.26 is satisfied. Writing the inequalities in terms of $i$ gives, from 5.27,

$$
\frac{n}{2} (D - B - 1) \leq i \leq \frac{n}{2} (D - B)
$$

and from 5.28,

$$
i \leq n(l - B)
$$

Therefore, the range of $i$ is given by

$$
\max\left\{ \frac{n}{2} (D-B-1), 0 \right\} \leq i \leq \min\left\{ \frac{n}{2} (D-B), n(1-B) \right\}
$$

(5.29)

where $i$ can take only integral values.
5.21

Thus, for any pair \((D,B)\) the range of \(i\) can be calculated and used to define the limits of an algorithm to list all the possible samples of the form 5.25 which yield the sufficient statistics \((D,B)\); such an algorithm is easily programmable for the computer.

To complete the enumeration, samples yielding pairs of sufficient statistics symmetrically related to \((D,B)\) are needed. If \((p_1,p_2,p_3)\) is any sample with sufficient statistics \((D,B)\) then the samples listed in the following table have sufficient statistics \((D',B')\) which are related to \((D,B)\) in the manner indicated:

<table>
<thead>
<tr>
<th>Samples</th>
<th>(D')</th>
<th>(B')</th>
</tr>
</thead>
<tbody>
<tr>
<td>((p_1,p_2,p_3))</td>
<td>(D)</td>
<td>(B)</td>
</tr>
<tr>
<td>((p_3,p_2,p_1))</td>
<td>(D)</td>
<td>(-B)</td>
</tr>
<tr>
<td>((p_1,p_2+3-2D,p_3))</td>
<td>(3-D)</td>
<td>(B)</td>
</tr>
<tr>
<td>((p_3,p_2+3-2D,p_1))</td>
<td>(3-D)</td>
<td>(-B)</td>
</tr>
</tbody>
</table>

As long as \(p_2 + 3 - 2D \leq 1\) all of these samples are admissible (that is, they satisfy 5.26-5.28), although they may not all be distinct. (For example, if \(B = 0\), then \((D,B)\) and \((D,-B)\) are identical.) It is easy to incorporate a procedure for obtaining the samples listed in the above table into the algorithm just described which lists samples yielding \((D,B)\). Unfortunately, however, not all samples yielding sufficient statistics symmetrically related to \((D,B)\) can be obtained in this way. This
becomes evident if one considers an admissible sample with
\( D' = 3 - D \) such as
\[
(i/n, D' - B - 2i/n, B + i/n)
\]
which would have to derive from
\[
(i/n, D - B - 2i/n, B + i/n)
\]
if it were to be included by the above procedure. It is possible, however, that for some value(s) of \( i \) the latter sample would not be admissible because \( D - B - 2i/n < 0 \) and so 5.25a would not be obtained from the set of samples for \((D, B)\). These additional samples can be enumerated by the same algorithm as was used to obtain samples for \((D, B)\) by substituting into 5.25a all integral values of \( i \) in the range defined by
\[
\frac{n}{2} (D-B) < i \leq \min\{\frac{n}{2}(D'-B), n(1-B)\}
\]
These are the values of \( i \) for which \( D - B - 2i/n < 0 \) but for which the sample 5.25a is admissible\(^1\), as determined by 5.29. This set will be empty when \( n(1-B) \leq n/2 (D-B) \). Additional samples for \((3-D, -B)\) are obtained from those for \((3-D, B)\) by a simple interchange of the first and third

---
\(^1\) Note that the only way in which the situation described here can occur is if the inequality 5.27 is satisfied for \((D', B)\) and not for \((D, B)\); the only way that this can be is if \( D-B-2i/n < 0 \), since if \( D-B-2i/n > 1 \) certainly \( D'-B-2i/n > 1 \), since \( D' \geq D \).
elements of each sample triple.

Since it is known that the probability of any sample \((p_1, p_2, p_3)\) is

\[
\prod_{i=1}^{3} \left( \frac{n_{p_i}}{n} \right)^{p_i} (1 - p_i)^{n(1-p_i)}
\]

where \(p_i\) is the true response probability at log dose \(x_i\), the probability of the pair of sufficient statistics \((D, B)\) is easily obtained as the sum of the probabilities of all samples yielding this pair.

Even for pairs of sufficient statistics which yield infinite or indeterminate estimates, the algorithms developed can be used to provide lists of samples and to compute the probabilities of the sufficient statistics.

Figures 5.21 and 5.22 contain flow charts showing the algorithm used for obtaining all samples yielding a given pair of sufficient statistics \((D, B)\) where \(B \geq 0\) and \(D \leq 3/2\) as well as all samples yielding any pair of sufficient statistics symmetrically related to the pair \((D, B)\). Also, the calculation of the probability of each pair of sufficient statistics is indicated.
FIGURE 5.21

FLOWCHART OF ROUTINE 'SAMPLE'

Enumeration of Samples for and Calculation of Probabilities of Sufficient Statistics \((D, B)\) and Those Statistics Symmetrically Related to \((D, B)\)

\[
\begin{align*}
D' &\leftarrow 3 - D \\
A &\leftarrow 3 - 2D \\
Pr(D, B) &\leftarrow 0 \\
Pr(D', B) &\leftarrow 0 \\
Pr(D, -B) &\leftarrow 0 \\
Pr(D', -B) &\leftarrow 0 \\
\end{align*}
\]

\[\begin{align*}
i_i &\leftarrow \left\lfloor \frac{D - B - 1}{2} \right\rfloor \\
j_i &\leftarrow \left\lceil \frac{D - B}{2} \right\rceil \\
i &\leftarrow \text{int} \left( i_i + 1 \right) \\
j &\leftarrow \text{int} \left( j_i \right) \\
j &\leftarrow \text{min} \left( j_i, j_i \right) \\
j &\leftarrow \text{int} \left( j_i \right)
\end{align*}\]

\[\begin{align*}
S &\leftarrow (p_1, p_2, p_3) \\
Pr(D, B) &\leftarrow Pr(D, B) + Pr(S)
\end{align*}\]

\[\begin{align*}
S &\leftarrow (p_2, p_3 + A, p_3) \\
Pr(D', B) &\leftarrow Pr(D', B) + Pr(S)
\end{align*}\]

* For glossary of flowchart symbols see Figure 5.22.

VARIABLE DESCRIPTION

\(Pr(D, B)\) = a \(k \times 1\) matrix whose ith element is the probability of \((D, B)\) for the ith configuration \((p_{1i}', p_{2i}', p_{3i}')\)

\(S\) = a sample triple

\(Pr(S)\) = a \(k \times 1\) matrix whose ith element is the probability of \(S\) for the ith configuration
5.25

FIGURE 5.22
CONTINUATION OF ROUTINE 'SAMPLE'

Enumeration of Additional Samples for Sufficient Statistics (D' = 3-D, ±B)

GLOSSARY OF FLOWCHART SYMBOLS

- input
- assignment
- decision (contents true or false)
- output
- computation

\[ i' \leftarrow \text{int} \left( f_i \right) \]
\[ k_1 \leftarrow \text{int} \left( c_0 \right) \]
\[ k_2 \leftarrow \text{min} \left( k_1, f_i \right) \]
\[ f_i' \leftarrow \text{int} \left( k_2 \right) \]

\[ i' > j \]

YES

NO

\[ f_i' \leftarrow \frac{i'}{m} \]
\[ p_{i'} \leftarrow D' - B = \frac{i'}{m} \]
\[ p_3 \leftarrow B + \frac{i'}{m} \]
\[ s \leftarrow (p_1', p_2', p_3') \]
\[ Pr(D', B) + Pr(D', B) + Pr(s) \]

\[ B = 0 \]

YES

NO

\[ i' > j \]

\[ s \leftarrow (p_1', p_2', p_3') \]
\[ Pr(D', -B) + Pr(D', -B) + Pr(s) \]

GO TO MAIN PROGRAMME, FIG. 5.23, FOR NEXT VALUES OF D AND B.
5.2.4 Description of the Computer Programme

The algorithms described in the previous sections are included in a FORTRAN IV programme which is designed to efficiently construct the exact distributions for any given configuration. Although in the programme emphasis is placed on the distribution of the pair \((\hat{\alpha}, \hat{\beta})\) some quantities involving \((\hat{\mu}, \hat{\sigma})\) are also computed so that only a few additional calculations are necessary to complete the description of the \((\hat{\mu}, \hat{\sigma})\) distribution.

Since the programme comprises one large loop which is repeated for each new \(n\), it can handle any number of values of \(n\), with up to four configurations for each\(^1\). For a given \(n\), the sufficient statistics \((D, B)\) are initiated at \((1/n, 0)\) and the method described by Wilson and Worcester (1943) is used to calculate the maximum likelihood estimates. When this method on its own does not provide sufficiently accurate estimates, it is supplemented with the Newton-Raphson iterative

\(^1\) Here a configuration consists of a set of true response probabilities only, since the log doses are fixed at \((-1, 0, 1)\) and \(n\) is already determined.
numerical procedure for finding the root of a function, as described in section 5.2.1. All samples producing this pair of estimates are then constructed and their probabilities computed from the given true response probabilities, and added together to give the probability of the estimates. Simultaneously, the probabilities of pairs of estimates symmetrically related to the given pair are also calculated. B and D are systematically incremented and the procedure is repeated until all sets of finite estimates and their probabilities for all input configurations have been obtained. Then the probabilities of various classes of indeterminate and infinite estimates are computed, the classification being that described in Appendix II. Finally, the programme utilises all pairs \((\hat{\alpha}, \hat{\beta})\) and \((\hat{\mu}, \hat{\sigma})\) for which both \(\hat{\alpha}\) and \(\hat{\beta}\) are finite and defined, and their probabilities to compute the unconditional expectations of the first four powers of \(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\sigma}\) and of the products \(\hat{\alpha}\hat{\beta}\) and \(\hat{\mu}\hat{\sigma}\). All probabilities and expectations are calculated for all input configurations. Figure 5.23 is a flow chart showing a

---

1 This procedure is included in the form of a subroutine to which the programme branches under the conditions described in section 5.2.1.

2 The total probability of all estimates over which these expectations are taken will not usually be 1 unless the probability of either \(\alpha\) or \(\beta\) being infinite (or indeterminate) is negligible. Expectations involving \(\mu\) or \(\sigma\) do not of course include values of these estimators which are infinite or indeterminate.
slightly simplified version of the overall programme (excluding sections which compute samples and their probabilities, as these are illustrated in Figures 5.21 and 5.22) for a given value of n.

The estimates are printed in the same order in which they are calculated, that is according to the value of D within B. The values of the estimators \( \hat{\alpha} \), \( \hat{\beta} \), \( \hat{\mu} \), \( \hat{\sigma} \) are printed first, \(^1\) followed by the probabilities of this set of values for the various configurations. The types of indeterminacy and their probabilities are listed next, and then the expectations, again for all configurations.

\(^1\) When \( \hat{\beta} = 0 \), \( \hat{\mu} \) and \( \hat{\sigma} \) are infinite and sequences of asterisks replace them in the listing.
**FIGURE 5.23**

FLOWCHART OF ENUMERATION OF ESTIMATOR DISTRIBUTIONS FOR PARAMETERS OF THE LOGISTIC MODEL

\[ TP(i) = (\hat{\beta}, \hat{\alpha}, \hat{\gamma}), \{B_1, B_2, \ldots, B_k\} \]

- **B** = 0
- **D** = 0
- **D** = **D** + \(\frac{1}{\alpha}\)

**VARIABLE DESCRIPTION**

- **k** = number of configurations
- **TP(i)** = i th configuration
- \(\hat{\theta}(D, -B)\) = MLE of \(\theta\) for \(D, -B\)
- \(\hat{\theta}(D, -B)\) = MLE of \(\theta\) for \(D, -B\)
5.3 Adjustment of Constructed Distributions

As mentioned previously, the estimator distributions include estimates which are infinite or indeterminate; the problem of what to do with these estimates is discussed in some detail in Appendix II. It was decided there that estimator distributions conditional on one or both estimates being defined and finite would be considered, and that the properties of interest, such as moments, correlation coefficients, and skewness, would therefore be calculated from these distributions. Bivariate distributions are then conditioned on both estimates being defined, while any univariate distributions considered are either the marginals of these conditional bivariate distributions or else marginals of more the encompassing bivariate distributions which include all pairs of estimates for which the variable of interest (that is, the variate of the marginal distribution) is defined and finite. In the case of \((\hat{\alpha}, \hat{\beta})\), the distributions of \(\hat{\alpha}\) derived by these two methods can differ substantially for some configurations and the one used must be that appropriate for the purpose. It is a peculiarity of the dose structure (in particular, it is because central log dose is 0) that \(\hat{\alpha}\), for some configurations, has quite a high probability of being defined when \(|\hat{\beta}|\) is
infinite, producing this difference. This situation obtains, with $x = 0$, for samples of types $(0, p, 1)$ and $(1, p, 0)$, $p \neq 0, 1$ as noted in Table II.1 in Appendix II.

In this chapter, marginal distributions studied are the marginals of the conditional bivariate distributions (that is, those distributions conditioned on both variates being defined and finite) unless otherwise specified.
5.4 Properties of Estimator Distributions

There are many properties of the distributions which could be studied. In line with the aim of obtaining uncorrelated estimators, correlation coefficients of the various pairs of estimators are the obvious quantities to compare. As well, the marginal distributions can be compared with each other and with the normal distribution by means of the appropriate measures of skewness and kurtosis and visually through frequency polygons or histograms. Other properties of the estimator distributions such as bias and probability of indeterminacy \(^1\) are of some interest also and will be considered briefly.

5.4.1 Configurations

Although an unavoidable disadvantage of this type of method of studying a distribution is the dependency of any

---

1. For the remainder of this chapter, 'indeterminacy' will be taken to mean a pair of estimates, one or both of which is either infinite or undefined.
results on the particular configurations used, if configurations are selected which are related but which differ systematically then some hypotheses about the properties of interest can perhaps be formulated. For the purpose of this investigation, groups of configurations such that all members of a group have the same \( \beta \)-value but different \( \alpha \)-values are used. (Ideally, the set of \( \alpha \)-values within each group should be the same.) A set of response probabilities which are symmetric about 0.5, \((p, 0.5, 1-p)\) where \(p\) can be assigned any value between 0 and 0.5, is chosen as the basis of a group, the choice of \(p\) determining \( \beta = \logit(1-p) \). The other configurations in the group are obtained by incrementing the response probability at central dose by a fixed amount (this is equivalent to changing \(\alpha\)) and changing the other two response probabilities so that \(\beta\) remains constant.

The first basic configuration is the set of true response probabilities \((0.30, 0.50, 0.70)\), which gives \(\beta = 0.84730\). Other desired configurations with this \(\beta\)-value are obtained by increasing the central dose response probability in steps of 0.05 up to 0.8 and, in each case, adjusting the other probabilities accordingly. A second group is initiated by \((0.222, 0.50, 0.778)\), for which
$\beta = 1.25276$, other members being obtained by increases in $P_2$ of 0.1 up to 0.8. As well as these two groups, a few other configurations, some symmetric and some asymmetric, are also used to generate distributions, but none have values of $\beta$ smaller than 0.84730. Although distributions are generated for every one of these configurations, they are not all compared in all the ways discussed, primarily because some of these, such as frequency polygons, require more work than is warranted by the results expected.

For some configurations distributions are enumerated for $n = 5, 10, and 15$, but the extreme discreteness of the distributions with $n = 5$ (in the sense that there are so few points of non-zero probability) has led to the exclusion of this value of $n$ in many cases. Most comparisons are made with $n = 10$ and some with $n = 15$ also. Initially, at least, effort is concentrated on the different configurations with $n = 10$, on the assumption that if interesting results obtain then other values of $n$ can be considered, but if there are no results of interest for this $n$-value, then those for other $n$ are unlikely to be interesting. This particular value of $n$ is chosen because it is large enough that the estimator distributions have a reasonable number of points with non-zero probability.
but not so many that the carrying out of the sorts of calculations and manipulations contained herein becomes too formidable a task.

The actual configurations for which estimator distributions are enumerated are as in Table 5.40.

5.4.2 Coefficients of Correlation

There are some configurations for which certain of the pairs of estimators are uncorrelated. For example, if the set of log doses chosen is symmetric about \( \mu \), the ED50, then \( \hat{\mu} \) and \( \hat{\beta} \) are uncorrelated, as are \( \hat{\mu} \) and \( \hat{\sigma} \); if further, \( \mu = 0 \) then \( \hat{\alpha} \) and \( \hat{\beta} \) are also uncorrelated. The implication of this is that if it were possible to choose doses which are symmetric about the ED50 then the \((\mu, \beta)\) or \((\mu, \sigma)\) parameterizations would yield uncorrelated estimates while \( \hat{\alpha} \) and \( \hat{\beta} \) would in general be correlated. Proof of these statements regarding the correlation of estimators when \( x = \mu \) follows.

If the log doses chosen are \( \mu - c, \mu, \mu + c \), then because of the symmetry in the logistic model the true response probabilities at these doses must be \( P, 0.5, 1 - P \), and so the probability of any observed sample \((p_1, p_2, p_3)\) with sufficient statistics \((D, B)\) is exactly the same
### TABLE 5.40

**CONFIGURATIONS** whose estimator distributions are enumerated

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$\mu$</th>
<th>$P_1$, $P_2$, $P_3$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84730</td>
<td>1.18022</td>
<td>0.0</td>
<td>0.0</td>
<td>0.30, 0.50, 0.70</td>
<td>5;10;15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.20068</td>
<td>-0.23685</td>
<td>0.344, 0.55, 0.74</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.40546</td>
<td>-0.47853</td>
<td>0.391, 0.60, 0.778</td>
<td>10;15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.61904</td>
<td>-0.73060</td>
<td>0.443, 0.65, 0.813</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.84730</td>
<td>-1.0</td>
<td>0.50, 0.70, 0.845</td>
<td>10;15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.09861</td>
<td>-1.29660</td>
<td>0.563, 0.75, 0.875</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.38630</td>
<td>-1.63614</td>
<td>0.632, 0.80, 0.903</td>
<td>10;15</td>
</tr>
<tr>
<td>1.25276</td>
<td>0.79824</td>
<td>0.0</td>
<td>0.0</td>
<td>0.222, 0.50, 0.778</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.40546</td>
<td>-0.32365</td>
<td>0.30, 0.60, 0.84</td>
<td>5;10;15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.84730</td>
<td>-0.67635</td>
<td>0.40, 0.70, 0.891</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.38630</td>
<td>-1.10660</td>
<td>0.533, 0.80, 0.933</td>
<td>10</td>
</tr>
<tr>
<td>2.19722</td>
<td>0.45512</td>
<td>0.0</td>
<td>0.0</td>
<td>0.10, 0.50, 0.90</td>
<td>5;10;15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.40546</td>
<td>-0.18453</td>
<td>0.143, 0.60, 0.931</td>
<td>10;15</td>
</tr>
</tbody>
</table>

* A configuration is completely defined by the log doses, $n$-value, and either the parameters or the true response probabilities, $P_1$, $P_2$, $P_3$. The log doses are $-1, 0, 1$ for all the above cases.
as the probability of \((1-p_3, 1-p_2, 1-p_1)\) with sufficient statistics \((3-D, B)\), namely
\[
\left(\begin{array}{c} n \\
\text{n}_1 \\
\text{n}_2 \\
\text{n}_3 \\
\end{array}\right)\left(\begin{array}{c} n \\
\text{n}_1 \\
\text{n}_2 \\
\text{n}_3 \\
\end{array}\right) p^{n(p_1+1-p_3)} (0.5)^n (1-p)^n(1-p_3).
\]
Thus the probability of \((D, B)\) is the same as that of \((3-D, B)\); also, of course, the probability of \((D, -B)\) is the same as that of \((3-D, -B)\).

Now the covariance of \((\hat{\mu}, \hat{\beta})\) is
\[
E\{(\hat{\mu} - E(\hat{\mu})) (\hat{\beta} - E(\hat{\beta}))\}
\]
and from 5.21
\[
\hat{\mu} = \mu - 2 \tanh^{-1} X/\beta
\]
where \(X\) is the function of the observations derived in Section 5.2.1. Since \(\hat{\beta}\) and \(X\) are independent of the value of \(x\), the symmetric relationships involving \(\hat{\beta}\) and \(X\) and the sufficient statistics as cited in Section 5.2.2 are still valid when \(x = \mu\). Thus if \(\hat{\beta}'\) and \(X'\) are the values taken by the quantities \(\hat{\beta}, X\) when the sufficient statistics are \((D, B)\), then \(\hat{\beta}', -X'\) are their values for \((3-D, B)\); similarly if \(-\hat{\beta}', X'\) are the values for \((D, -B)\) then \(\hat{\beta} = -\hat{\beta}'\) and \(X = -X'\) when \((3-D, -B)\) are the sufficient statistics. Because \(Pr[(D,B)] = Pr[(3-D,B)]\) and \(Pr[(D,-B)] = Pr[(3-D,-B)]\), the two pairs \((\hat{\beta}', X')\) and \((-\hat{\beta}', -X')\) have the same probability and also the two pairs
(-\*\,X') and (-\*\,-X') have equal probabilities. This means that the distributions of \(\tanh^{-1}X/\hat{\beta} \), \(\tanh^{-1}X\), and \(\hat{\beta} \tanh^{-1}X\) are all symmetric about zero and so, letting \(\beta_0 = E(\hat{\beta})\),

\[
E(\mu) = \mu
\]

and

\[
\text{Cov}(\mu, \hat{\beta}) = E\{-2 \tanh^{-1}X/\hat{\beta} (\hat{\beta} - \beta_0)\} = 0
\]

If the same arguments are applied to the covariance of \((\hat{\alpha}, \hat{\beta})\) the result is

\[
\text{Cov}(\hat{\alpha}, \hat{\beta}) = E\{(\hat{\alpha} - E(\hat{\alpha}))(\hat{\beta} - \beta_0)\}
\]

\[
= E\{(\hat{\alpha} + \mu \beta_0)(\hat{\beta} - \beta_0)\}
\]

\[
= E\{(2 \tanh^{-1}X - \mu \hat{\beta} + \mu \beta_0)(\hat{\beta} - \beta_0)\}
\]

\[
= -\mu E\{(\hat{\beta} - \beta_0)^2\} = -\mu \text{var} \hat{\beta}
\]

which is zero only when \(\mu = 0\).

Unfortunately in the real world it is not usually possible to choose doses which are symmetric about the ED50 since this is in most cases unknown and to be estimated. Comparisons of the exact correlation coefficients for the
various parameterizations and configurations included in this study would indicate the effect on correlation of departures from the ideal symmetric situation, that is, the effect of moving the central dose further and further from \( \mu \). (In actual fact, in this study it is \( \mu \) that changes while the central dose remains constant; the two situations are, however, equivalent.) The coefficients of correlation, \( \rho(\hat{\alpha}, \hat{\beta}) \), \( \rho(\hat{\mu}, \hat{\sigma}) \), and \( \rho(\hat{\mu}, \hat{\beta}) \), calculated for most of the configurations listed in Table 5.40, are shown in Table 5.41, while in Figures 5.40(a)-(c) the absolute value of \( \rho \) is plotted against \( P_2 \), the central dose response probability, for fixed \( \beta \) and \( n \).

Consider first, for fixed \( \beta \) and \( n \), the effect on the correlation of each pair of estimators of increasing \( P_2 \) from 0.5.\(^1\) From the Table and Figure 5.40(a) it is evident that, for the smallest value of \( \beta \) and \( n = 10 \), \( \rho(\hat{\mu}, \hat{\beta}) \) increases from 0 at \( P_2 = 0.5 \), attains a maximum somewhere between \( P_2 = 0.65 \) and \( P_2 = 0.75 \), and thereafter decreases with further incrementation of \( P_2 \); Figure 5.40(b) indicates that the same pattern presents itself for \( \rho(\hat{\mu}, \hat{\beta}) \) with \( n = 15 \) and \( \beta = 0.84730 \), the maximum value of \( \rho(\hat{\mu}, \hat{\beta}) \) obtaining between \( P_2 = 0.6 \) and \( P_2 = 0.8 \). For the same \( \beta \) and \( n = 10 \), \( \rho(\hat{\alpha}, \hat{\beta}) \) also exhibits this behaviour, with its maximum occurring for \( P_2 \) between 0.7 and 0.8; within

\(^1\)This is equivalent to moving \( \mu \), the ED50, and \( x \), the central dose, further apart.
### TABLE 5.41

**COEFFICIENTS OF CORRELATION FOR \((\hat{\alpha}, \hat{\beta})\), \((\hat{\mu}, \hat{\sigma})\), \((\hat{\mu}, \hat{\beta})\)**

<table>
<thead>
<tr>
<th>Configuration</th>
<th>(\beta)</th>
<th>(P_2)</th>
<th>(n = 10)</th>
<th>(n = 15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(\rho(\hat{\alpha}, \hat{\beta}))</td>
<td>(\rho(\hat{\mu}, \hat{\sigma}))</td>
<td>(\rho(\hat{\mu}, \hat{\beta}))</td>
</tr>
<tr>
<td>0.84730</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.55</td>
<td>0.097</td>
<td>-0.307</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.190</td>
<td>-0.542</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>0.272</td>
<td>-0.694</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.333</td>
<td>-0.788</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.362</td>
<td>-0.846</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.344</td>
<td>-0.885</td>
<td>0.134</td>
</tr>
<tr>
<td>1.25276</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.236</td>
<td>-0.446</td>
<td>0.237</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.403</td>
<td>-0.701</td>
<td>0.387</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.439</td>
<td>-0.826</td>
<td>0.463</td>
</tr>
<tr>
<td>2.19722</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.302</td>
<td>-0.222</td>
<td>0.014</td>
</tr>
</tbody>
</table>
FIGURE 5.40

ABSOLUTE VALUES OF CORRELATION COEFFICIENTS AS A FUNCTION OF ONE PARAMETER*

**Figure 5.40(a)**

\[ \beta = 0.84730 \]
\[ n = 10 \]

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>(( \hat{\mu}, \beta ))</th>
<th>(( \hat{\alpha}, \beta ))</th>
<th>(( \hat{\mu}, \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Figure 5.40(b)**

\[ \beta = 0.84730 \]
\[ n = 15 \]

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>(( \hat{\mu}, \sigma ))</th>
<th>(( \hat{\alpha}, \beta ))</th>
<th>(( \hat{\mu}, \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Figure 5.40(c)**

\[ \beta = 1.25276 \]
\[ n = 10 \]

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>(( \hat{\mu}, \sigma ))</th>
<th>(( \hat{\mu}, \beta ))</th>
<th>(( \hat{\alpha}, \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Figure 5.40(d)**

\[ \alpha = 0.40546 \] (\( P_2 = 0.6 \))
\[ n = 10 \]

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>(( \hat{\alpha}, \beta ))</th>
<th>(( \hat{\mu}, \sigma ))</th>
<th>(( \hat{\mu}, \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.0</td>
<td>2.0</td>
<td>( \beta )</td>
</tr>
</tbody>
</table>

* Figures 5.40a-c : \( |\rho| \) as a function of \( P_2 \) for fixed \( \beta \) and \( n \).
* Figure 5.40d : \( |\rho| \) as a function of \( \beta \) for fixed \( \alpha \) (or \( P_2 \)) and \( n \).
the range of values for $P_2$ included, $\rho(\hat{\alpha},\hat{\beta})$ is monotone increasing with increasing $P_2$ when $\beta = 0.84730$ and $n = 15$, although it seems highly likely that if the range of $P_2$ were enlarged $\rho(\hat{\alpha},\hat{\beta})$ would reach a maximum and thereafter decrease in this situation also. It is impossible to decide from the behaviour of $-\rho(\hat{\mu},\hat{\sigma})$ as evidenced in Table 5.41 and Figures 5.40(a)-(c) whether $-\rho(\hat{\mu},\hat{\sigma})$ is monotone increasing, asymptotic to 1, or whether it too would exhibit the same pattern as $\rho(\hat{\alpha},\hat{\beta})$ and $\rho(\hat{\mu},\hat{\beta})$ if the range of $P_2$ were expanded.

From Figures 5.40(a)-(c) the absolute values of the coefficients of correlation of the three estimator pairs can be compared for fixed values of $\beta$ and $n$. The most obvious consistent feature of these plots is that $-\rho(\hat{\mu},\hat{\sigma})$ is always the largest of the three coefficients for any given configuration, except those for which all correlation coefficients are zero. Figure 5.40(a) shows that for $\beta = 0.84730$ and $n = 10$ when $P_2$ is not far from 0.5, $\rho(\hat{\alpha},\hat{\beta})$ is slightly smaller than $\rho(\hat{\mu},\hat{\beta})$ but for some $P_2$ between 0.55 and 0.6 $\rho(\hat{\mu},\hat{\beta})$ equals $\rho(\hat{\alpha},\hat{\beta})$ and thereafter $\rho(\hat{\mu},\hat{\beta})$ is the smaller with the difference between the two increasing with increasing $P_2$. From Figure 5.40(b) with $n = 15$ the curves $\rho(\hat{\mu},\hat{\beta})$ and $\rho(\hat{\alpha},\hat{\beta})$ intersect when $P_2$ is

---

1 The notable exception to this is when $\beta = 2.19722$. It is unfortunate that there is no information on more configurations with this value of $\beta$. 
approximately 0.71, with \( \rho(\hat{\mu}, \hat{\beta}) \) the larger up to that point. When \( \beta = 1.25276 \) and \( n = 10 \) the correlation coefficients of \((\hat{\mu}, \hat{\beta})\) and \((\hat{\alpha}, \hat{\beta})\) are very close together over the range of \( P_2 \) considered here.

It is worthwhile to utilize the small amount of information which is available for \( \beta = 2.19722 \) to examine the effect on the correlation of increasing \( \beta \) while keeping \( \alpha \) (and \( P_2 \)) constant\(^1\). In Figure 5.40(d) \(|\rho|\) is plotted against \( \beta \) for \( \alpha = 0.40546 \) (\( P_2 = 0.6 \)). As \( \beta \) increases, \(-\rho(\hat{\mu}, \hat{\sigma})\) decreases and \( \rho(\hat{\alpha}, \hat{\beta}) \) increases, both monotonely, while \( \rho(\hat{\mu}, \hat{\beta}) \) increases and then decreases. For the three values of \( \beta \) with \( \alpha = .40546 \), \( \rho(\hat{\mu}, \hat{\beta}) \) is always less than \(-\rho(\hat{\mu}, \hat{\sigma})\) and is at least as small as \( \rho(\hat{\alpha}, \hat{\beta}) \) to two places of decimals.

In the first part of this discussion of correlation coefficients it was shown that if experimental doses could be chosen which were symmetric about the ED50 then either \((\mu, \beta)\) or \((\mu, \sigma)\) would be a better parameter pair than \((\alpha, \beta)\) since only with \( \mu = 0 \) would \((\hat{\alpha}, \hat{\beta})\) be uncorrelated. Since one can rarely, if ever, choose doses which are exactly symmetric about \( \mu \), a brief look was taken at the effects of varying degrees of asymmetry, in terms of the effects on the correlation coefficients of the different

\(^1\) Since \( \mu = -\alpha/\beta \) an increase in \( \beta \) with \( \alpha \) constant represents a decrease in \( |\mu| \) and thus a decrease in the difference between \( \mu \) and \( x \).
estimator pairs. The largest effect is evidenced by 
\((\hat{\mu}, \hat{\sigma})\), for in the range of \(P_2\)-values studied 
\(-\rho(\hat{\mu}, \hat{\sigma})\) is monotone increasing with increasing
asymmetry (for fixed \(\beta\) and \(n\)), and for all configurations except those with \(\beta = 2.19722\) is at least as large
as the largest of \(\rho(\hat{\mu}, \hat{\beta})\) and \(\rho(\hat{\alpha}, \hat{\beta})\), and in most
cases is considerably larger. Overall, there is little
to choose between \((\alpha, \beta)\) and \((\mu, \beta)\), both of which
exhibit the same pattern for some \(\beta-\) and \(n\)-values
of increasing correlation of estimators with increasing
asymmetry up to a particular value of \(P_2\) after which
the correlation decreases. From the evidence presented
in this section, however, \((\mu, \beta)\) is the slightly
favoured parameterization since \(\rho(\hat{\mu}, \hat{\beta})\) is never much
larger than \(\rho(\hat{\alpha}, \hat{\beta})\) and is sometimes considerably
smaller. Obviously more sets of parameters are needed
if these conclusions are to be confirmed. Hopefully,
with more information it would be possible to show that
either \((\alpha, \beta)\) or \((\mu, \beta)\) was, in some average sense,
the superior parameterization, at least as far as the
correlation of its estimators was concerned.

5.4.3 Frequency Polygons

An additional computer programme was written to
display in two-way tables bivariate distributions which have been standardized to zero means and unit variances. For certain of the configurations, the distributions of \((\hat{\alpha}, \hat{\beta})\) and \((\hat{\mu}, \hat{\sigma})\) are standardized and the probabilities of pairs of standardized estimates lying between \((-3, -3)\) and \((+3, +3)\) are tabulated in intervals of size 0.2 by 0.2, with overflow 'intervals' at each end of the range for values exceeding \(|3|\).\(^1\) The intervals can then be combined as desired and the probabilities in the resulting enlarged intervals plotted in the form of histograms or polygons which can be compared with the appropriate standard normal curve.

Frequency polygon representations of the marginal distributions of \(\hat{\alpha}, \hat{\beta}, \hat{\mu},\) and \(\hat{\sigma}\) (based on the distributions of \((\hat{\alpha}, \hat{\beta})\) and \((\hat{\mu}, \hat{\sigma})\)) with intervals of size 0.6 are shown in Figures 5.41 - 5.44 for some configurations with \(n = 10\). The distributions of \(\hat{\mu}\) and \(\hat{\sigma}\) are markedly peaked, particularly for the more asymmetric configurations; \(\hat{\alpha}\) seems to have a distribution quite reasonably approximated by the normal for all configurations; and the distribution of \(\hat{\beta}\) does not appear too far from the normal, although since only two values of \(\beta\) are represented

\(^1\) The region which is broken down into intervals is extended as necessary if the total probability of either standardized estimator being greater greater than 3 or less than -3 exceeds 0.002.
FREQUENCY POLYGON REPRESENTATIONS OF MARGINALS OF STANDARDIZED $(\hat{\alpha}, \hat{\beta})$ AND $(\hat{\mu}, \hat{\sigma})$ DISTRIBUTIONS COMPARED WITH THE STANDARD NORMAL CURVE

Ordinate represents $10 \times$ the relative frequency in intervals of size 0.6.

* $P_I(\hat{\phi}_1, \hat{\phi}_2) = \Pr(\hat{\phi}_1 \text{ or } \hat{\phi}_2 \text{ is infinite or undefined})$. 
FIGURE 5.42

FREQUENCY POLYGON REPRESENTATIONS OF MARGINALS OF STANDARDIZED \((\hat{\alpha}, \hat{\beta})\) AND \((\hat{\mu}, \hat{\sigma})\) DISTRIBUTIONS COMPARED WITH THE STANDARD NORMAL CURVE

\[ P_I(\hat{\alpha}, \hat{\beta}) = 0.005^* \]

\[ P_I(\hat{\mu}, \hat{\sigma}) = 0.044^* \]

Ordinate represents \(10 \times\) the relative frequency in intervals of size 0.6.

\* \(P_I(\hat{\phi}_1, \hat{\phi}_2) = \Pr(\hat{\phi}_1 \text{ or } \hat{\phi}_2 \text{ is infinite or undefined})\).
5.48

FIGURE 5.43

FREQUENCY POLYGON REPRESENTATIONS OF MARGINALS OF STANDARDIZED \((\hat{\alpha}, \hat{\beta})\) AND \((\hat{\mu}, \hat{\sigma})\) DISTRIBUTIONS COMPARED WITH THE STANDARD NORMAL CURVE

Ordinate represents \(10 \times\) the relative frequency in intervals of size 0.6.

\[^*\text{Pr}(\hat{\phi}_1 \text{ or } \hat{\phi}_2 \text{ is infinite or undefined})\]
5.49

FIGURE 5.44
FREQUENCY POLYGON REPRESENTATIONS OF MARGINALS OF STANDARDIZED $\hat{\alpha}$, $\hat{\beta}$, AND $\hat{\mu}$, $\hat{\sigma}$ DISTRIBUTIONS COMPARED WITH THE STANDARD NORMAL CURVE

Ordinate represents $10 \times$ the relative frequency in intervals of size 0.6.  
* $P_1(\hat{\phi}_1, \hat{\phi}_2) = \Pr(\hat{\phi}_1$ or $\hat{\phi}_2$ is infinite or undefined).
in the drawings (and one only barely) it is impossible

to make any absolute statements about the behaviour of \( \hat{\beta} \). Yet the distribution of \( \hat{\beta} \) is much better approxi-
imated by a normal than is that of \( \hat{\sigma} \) for all cases considered; and \( \hat{\alpha} \) is certainly better than \( \hat{\mu} \) as repre-
sentated by the polygons shown.

One might conclude from this that the marginal distributions of \( \hat{\alpha} \) and \( \hat{\beta} \) are much closer to the normal distribution than are those of \( \hat{\mu} \) and \( \hat{\sigma} \). If this were true, one would hope that the bivariate distribution of \( (\hat{\alpha}, \hat{\beta}) \) would also be closer to the bivariate normal distribution than that of \( (\hat{\mu}, \hat{\sigma}) \). With the limited amount of data available, however, the conclusions as stated are somewhat speculative but do provide a hypothesis which could be tested by further study.

5.4.4 Coefficients of Skewness and Kurtosis

Coefficients of skewness and kurtosis can be used to assess specific differences between the estimator distributions and the normal. The necessary moments of the marginals of the various bivariate distributions are calculated for all configurations with \( n = 10 \).¹

¹ These bivariate distributions are conditioned on both variates being defined and finite.
Thus, for the distribution of \((\hat{\alpha}, \hat{\beta})\), for example,

\[
\begin{align*}
\mu_{10} &= E\{(\hat{\alpha} - E(\hat{\alpha}))^i\}, \\
\mu_{01} &= E\{(\hat{\beta} - E(\hat{\beta}))^i\}, \quad i = 2, 3, 4
\end{align*}
\]

giving coefficients of skewness

\[
\gamma_1(\hat{\alpha}) = \frac{\mu_{30}}{(\mu_{20})^{3/2}} \\
\gamma_1(\hat{\beta}) = \frac{\mu_{03}}{(\mu_{02})^{3/2}}
\]

and coefficients of kurtosis

\[
\gamma_2(\hat{\alpha}) = \frac{\mu_{40}}{(\mu_{20})^2} - 3 \\
\gamma_2(\hat{\beta}) = \frac{\mu_{04}}{(\mu_{02})^2} - 3
\]

These coefficients are all recorded in Table 5.42.

The skewness coefficients are plotted against probability of response at central dose for fixed \(\beta\) in Figures 5.45(a) and (b). From the Table and Figure
TABLE 5.42

COEFFICIENTS OF SKEWNESS AND KURTOSIS FOR $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$, $\hat{\sigma}$ WITH $n = 10$ *

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$P_2$</td>
<td>$Y_1(\hat{\alpha})$</td>
</tr>
<tr>
<td>0.84730</td>
<td>0.50</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.55</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.38</td>
</tr>
<tr>
<td>1.25276</td>
<td>0.50</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.28</td>
</tr>
<tr>
<td>2.19722</td>
<td>0.50</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.39</td>
</tr>
</tbody>
</table>

* Coefficients are calculated from marginals of bivariate distributions of $(\hat{\alpha}, \hat{\beta})$ & $(\hat{\mu}, \hat{\sigma})$. 
Figure 5.45

Coefficients of skewness as a function of $P_2$ for fixed values of $S$ and $n$.

Figure 5.45b

- $n = 10$
- $\beta = 1.25276$

Figure 5.45a

- $n = 10$
- $\beta = 0.84730$
5.45(b), when $\beta = 1.25276$, it is evident that the skewness of $\hat{\alpha}$ is uniformly less than that of $\hat{\mu}$ in absolute value and the distribution of $\hat{\beta}$ is less skew than that of $\hat{\sigma}$, indicating that $(\alpha, \beta)$ might well be the superior parameterization as far as this measure is concerned. Figure 5.45(a), however, tells a different and more complicated story. In this graph, the skewness coefficients of $\hat{\alpha}$ and $\hat{\beta}$ follow patterns similar to those exhibited in Figure 5.45(b) as $P_2$ increases; the coefficient of $\hat{\mu}$ starts out in the same way in both, that is, it decreases, taking negative values as $P_2$ increases from 0.5 and then, having attained some minimum, increases, approaching 0 again. In Figure 5.45(b), however, it is noted that $\gamma_1(\hat{\mu})$ never reaches 0 in the range of $P_2$ considered here (except of course when $P_2 = 0.5$), but in 5.45(a), when $\beta = 0.84730$, it takes the value 0 for some $P_2$ between 0.65 and 0.70 and then continues increasing.

The behaviour of $\gamma_1(\hat{\sigma})$ too is quite different for the different values of $\beta$. For $\beta = 1.25276$, $\gamma_1(\hat{\sigma})$ is always positive and monotone decreasing with increasing $P_2$; for $\beta = 0.84730$ it is always negative and its absolute value increases monotonely with increasing $P_2$. When $\beta = 0.84730$, then, it seems that there is a group of configurations for which the distributions of $\hat{\mu}$ and $\hat{\sigma}$ are less skew than those of $\hat{\alpha}$ and $\hat{\beta}$ but that eventually (for large enough $P_2$) this order is reversed.
On the basis of the information here presented and the shape of the curves plotted, one would anticipate that in general \((\alpha, \beta)\) would probably be the better parameterization, since there seems to be a maximum value for the skewness coefficients of \(\hat{\alpha}\) and \(\hat{\beta}\) for a particular value of \(\beta\).

From the coefficients of kurtosis in Table 5.42 and from Figure 5.46 in which these coefficients are plotted against \(P_2\) for \(\beta = 0.84730\) and \(n = 10\), it is evident that \((\hat{\alpha}, \hat{\beta})\) is the superior pair of estimators in this respect, since for any configuration \(\gamma_2(\hat{\alpha})\) and \(\gamma_2(\hat{\beta})\) are less in absolute value than are either \(\gamma_2(\hat{\mu})\) or \(\gamma_2(\hat{\sigma})\). Examination of Table 5.42 shows the pattern to be the same for other values of \(\beta\).

The large values of \(\gamma_2(\hat{\sigma})\) corroborate the findings of Section 5.4.3 when frequency polygons were compared and those for \(\hat{\sigma}\) were found to be very peaked.

It is interesting to note that for all of the configurations studied at least three of the four marginal estimator distributions are leptokurtic \((\gamma_2 > 0)\); in fact, \(\gamma_2(\hat{\mu})\) and \(\gamma_2(\hat{\sigma})\) are always positive; only for the largest \(\alpha\)-value is \(\gamma_2(\hat{\alpha})\) negative and only for the largest \(\beta\)-value is \(\gamma_2(\hat{\beta})\) negative.
FIGURE 5.46
COEFFICIENTS OF KURTOSIS AS A FUNCTION OF $p_2$ FOR $n = 10$ AND $\beta = 0.84730$
5.4.5 Bias

Although, as mentioned in Chapter 2, bias is not an important factor in discriminating between parameterizations, it is of some interest as one of the properties of any particular estimator. It might be profitable, then, to look at changes in the bias of one estimator of a pair as its true value changes and the other parameter remains constant. The configurations used in this chapter do not include enough different β-values to enable changes in the bias of \( \hat{\beta} \) or \( \hat{\sigma} \) to be reasonably examined. For a fixed value of \( \beta \), however, the biases of \( \hat{\alpha} \) and \( \hat{\mu} \) can be plotted against the true values of \( \alpha \) and \( \mu \) respectively. In actual fact, Figure 5.47 displays the biases of \( \hat{\alpha} \) and \( -\hat{\mu} \) for \( \beta = 0.84730 \) and \( \beta = 1.25276 \) when \( n = 10 \), and for \( \beta = 0.84730 \) when \( n = 15 \), plotted against the true values of \( \alpha \) and \( -\mu \).\(^1\)

The relationships illustrated in Figure 5.47 are interesting, but do not really provide

\(^1\) Since \( \mu \) and the bias of \( \hat{\mu} \) are less than or equal to zero for most configurations, \( -\mu \) and the bias of \( -\mu \) have been used to aid comparisons with the positive quantities \( \alpha \) and bias of \( \hat{\alpha} \).
FIGURE 5.47

BIAS OF $\hat{\alpha}$ AND $-\hat{\mu}$ AS A FUNCTION OF THEIR TRUE VALUES FOR FIXED $\beta$ AND $n$

- $n = 10$, $\beta = 0.84730$
- $n = 10$, $\beta = 1.25276$
- $n = 15$, $\beta = 0.84730$

Bias($\hat{\phi}_1$) vs $\phi_1 = \alpha, -\mu$
sufficient evidence on which to base any conclusions.

5.4.6 Probability of Indeterminacy

Although, like bias, the probability of indeterminacy is an interesting property of a pair of estimators, it too is not considered to be a relevant factor in the comparison of the various parameterizations. This is because it so directly depends on the configuration that no general statements can be made that one parameterization is any better than another in this respect.

For example, if the true response probabilities are quite close together (as in (.35,.5,.65) ) there is quite a high probability of getting $\hat{\beta} = 0$, which makes $\hat{\mu}$ and $\hat{\sigma}$ infinite, whereas if the response probabilities are very spread out (as in (.1,.5,.9) ) there is a small chance of obtaining $\hat{\beta} = 0$, but quite a high probability of $\hat{\beta}$ being infinite, in which case there will likely be a higher probability of $(\hat{\alpha},\hat{\beta})$ being indeterminate than $(\hat{\mu},\hat{\sigma})$. The relationship between the probabilities of indeterminacy for the various parameterizations can be demonstrated algebraically.

---

1 Indeterminacy is taken to mean a pair of estimates, one or both of which is infinite or undefined.
as follows, where $P_D(\hat{\theta}_1, \hat{\theta}_2)$ is the probability that the
\[ \text{two estimators } \hat{\theta}_1 \text{ and } \hat{\theta}_2 \text{ are both defined and finite:} \]
\[
P_D(\hat{\mu}, \hat{\beta}) = P_D(\hat{\alpha}, \hat{\beta}) - \Pr(\hat{\beta} = 0) \]
\[
P_D(\hat{\mu}, \hat{\sigma}) = P_D(\hat{\alpha}, \hat{\beta}) - \Pr(\hat{\beta} = 0) \]
\[+ \Pr(\hat{\alpha} \text{ finite, } |\hat{\beta}| \text{ infinite}) \]
\[+ \Pr(|\hat{\mu}| = 1, \hat{\sigma} = 0 \text{ when } |\hat{\alpha}| \text{ and } |\hat{\beta}| \text{ are infinite}) \].

Thus $(\hat{\mu}, \hat{\beta})$ always has the highest probability of
indeterminacy, and which of $(\hat{\mu}, \hat{\sigma})$ or $(\hat{\alpha}, \hat{\beta})$ has the
smaller probability of indeterminacy depends on the
relative size of $\Pr(\hat{\beta} = 0)$ and the probability of $\hat{\mu}$
and $\hat{\sigma}$ being defined and finite when one or both of $|\hat{\alpha}|$
or $|\hat{\beta}|$ is infinite, which obviously depends on the
configuration.

The probabilities of indeterminacy are recorded

\[1 \text{ When the central dose, } x, \text{ is not zero, } \Pr(\hat{\alpha} \text{ finite, } |\hat{\beta}| \text{ infinite}) \text{ should be replaced by} \]
\[
\sum_{n=1}^{n-1} \left\{ \Pr(0,p,1) + \Pr(1,p,0) \right\} \]
\[
\text{in this expression since only when } x = 0 \text{ are the two}
equivalent.\]
for the three estimator pairs and all configurations with $n = 10$ and $n = 15$ in Table 5.43.

As expected, for any $\beta$ value the probability of indeterminacy for all three estimator pairs increases as $P_2$ moves farther from 0.5. Also not unexpectedly, the smaller the true value of $\beta$ the less is the probability of indeterminacy in $(\hat{\alpha}, \hat{\beta})$ and the greater is this probability for $(\hat{\mu}, \sigma)$ and $(\hat{\mu}, \hat{\beta})$.

The value in this quantity, the probability of indeterminacy, is the same as that in bias: it is interesting but not important in discrimination between parameterizations since if $\mu$ is the parameter of interest no matter whether the estimate $\hat{\mu}$ is obtained directly or indirectly the probability of its being indeterminate does not alter. On the other hand, if it is the relationship between dose and response that is of interest and not any specific parameter, both bias and probability of indeterminacy are worth some consideration when the decision as to which parameterization is to be used to describe this relationship is made. But some idea of the true value of $\beta$ is necessary to make a choice between $(\mu, \sigma)$ and $(\alpha, \beta)$, and such information
<table>
<thead>
<tr>
<th>CONFIGURATION</th>
<th>$\beta$</th>
<th>$P_2$</th>
<th>$n = 10$</th>
<th>$n = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$P_I(\hat{\alpha}, \hat{\beta})$</td>
<td>$P_I(\hat{\mu}, \hat{\sigma})$</td>
<td>$P_I(\hat{\mu}, \hat{\beta})$</td>
</tr>
<tr>
<td>0.84730</td>
<td>0.50</td>
<td>0.0009</td>
<td>0.0308</td>
<td>0.0317</td>
</tr>
<tr>
<td></td>
<td>0.55</td>
<td>0.0009</td>
<td>0.0317</td>
<td>0.0325</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.0011</td>
<td>0.0334</td>
<td>0.0345</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>0.0021</td>
<td>0.0372</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.0054</td>
<td>0.0437</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.0150</td>
<td>0.0542</td>
<td>0.0692</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.0394</td>
<td>0.0712</td>
<td>0.1103</td>
</tr>
<tr>
<td>1.25276</td>
<td>0.50</td>
<td>0.0068</td>
<td>0.0044</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.0060</td>
<td>0.0051</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.0108</td>
<td>0.0080</td>
<td>0.0188</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.0540</td>
<td>0.0182</td>
<td>0.0720</td>
</tr>
<tr>
<td>2.19722</td>
<td>0.50</td>
<td>0.1220</td>
<td>0.0002</td>
<td>0.1220</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.1069</td>
<td>0.0007</td>
<td>0.1069</td>
</tr>
</tbody>
</table>

* $P_I(\hat{\phi}_1, \hat{\phi}_2) = 1 - P_D(\hat{\phi}_1, \hat{\phi}_2) = \Pr(\text{one or both of } \hat{\phi}_1, \hat{\phi}_2 \text{ is infinite or undefined})$. 
is not always available. The one fact that is known is that \((\mu, \beta)\) is never the best parameter pair as far as the property under discussion is concerned.
5.5 Conclusions

The first object of the investigation carried out in this chapter was to gain some knowledge of the distributional properties of the estimators of various parameter pairs by detailed examination of the exact estimator distributions for some specific configurations. This has been successfully accomplished, for specified distributions have been efficiently and satisfactorily constructed and extensively studied. The second object was to formulate hypotheses about the relative merits of the three most common parameterizations of the model using this newly acquired knowledge, and the third object was to suggest parameterizations which might be in some way superior to the usual ones, particularly with respect to correlation of estimators. The extent to which these last two objectives can be accomplished in a study such as this is restricted by the number of configurations which can be profitably and economically studied in depth and also by the confinement of configurations to the type for which the method of Wilson and Worcester is applicable. The attainment of the fourth objective, the testing of formulated hypotheses, would require a much more detailed, extensive, and planned study than can be accomplished herein. So far in this chapter
only the first aim has been discussed, as emphasis has been placed on the construction of the distributions and the studying of their various properties. The remainder of this section will therefore be concerned with the second and third aims: the utilisation of the knowledge gained in this investigation to formulate hypotheses, both about the parameterizations examined and about what other sorts of parameterizations might be of interest. The limitations of this study in the attainment of these objectives will also be discussed.

The property which is the most relevant to the original purpose of this study is, of course, the correlation of pairs of estimators. On the basis of the information gained about correlation, one could hypothesize that either of the parameterizations \((\alpha, \beta)\) or \((\mu, \beta)\) would be preferable to \((\mu, \sigma)\) in most cases, but which of these two would be better is impossible to even speculate from the available evidence. Although additional data might indicate no clear-cut decision between these two parameterizations could be made, even such a negative conclusion would be useful.

All other studied aspects of the estimator distributions are concerned with the marginal distributions only. However, any conclusions drawn from these distributions
can be used to weight those drawn from the bivariate distributions or to discriminate between parametizations which appear equally good as far as correlation of estimators is concerned. The marginal distributions are compared with the normal distribution, since this is considered a desirable distribution, and even though a distribution with normal margins is not necessarily bivariate normal, it is likely that the closer the marginals are to normality the closer the bivariate distribution will be to the normal. There are many quantities which could have been considered in addition to those included, the most obvious being more high order cumulants such as $\kappa_2 \kappa_1, \kappa_{12}, \kappa_{31},$ etc., all of which should be zero if a pair of variates is in fact bivariate normal. However, for the present investigation I feel that the quantities considered are adequate for the data available. From the properties considered, it seems evident that overall the marginals of the distribution of $(\hat{\alpha}, \hat{\beta})$ are better (that is, closer to the normal) than those of the other bivariate distributions: they look more normal, and although they are not consistently more or less skew than other marginals, neither $\hat{\alpha}$ nor $\hat{\beta}$ are very skewed at any time whereas sometimes one or both of $\hat{\mu}$ and $\hat{\sigma}$ have coefficients of skewness much higher
than the maximum skewness attained by $\hat{a}$ or $\hat{\beta}$; further $\gamma_2(\hat{a})$ and $\gamma_2(\hat{\beta})$ are both consistently less than the corresponding coefficients for both $\hat{\mu}$ and $\hat{\sigma}$. Of course, this is not to say that $(\alpha, \beta)$ is the best possible parameterization, but it does appear to be the best of the three studied in detail. This conclusion could certainly be considered as one of the hypotheses which could be tested after the accumulation of more data, although to test such a hypothesis rigorously would require the development of suitable measures and tests.

The information presented in this chapter also permits hypotheses to be made about types of parameterizations which might be superior, at least in some respects, to the three studied. Since it is easier to formulate such hypotheses from information about properties such as skewness and kurtosis and the overall appearance of a distribution than from figures on correlations, the parameter transformations suggested are such as might lead to estimator distributions which are more normal, with the hope that the estimator pairs might be less correlated as well. As mentioned before, it is of some importance anyway to have estimator distributions close to the normal. Since $\hat{\beta}$ has at least some
of the properties of a lognormally distributed variate in that it is always positively skewed and usually leptokurtic, a log transformation of $\beta$ might be worth trying. Instead of $\beta$, consider a parameter $\theta$, where

$$\theta = \lambda_1 \log (\beta - \lambda_2)$$

with $\lambda_2$ taking a value which makes $(\hat{\beta} - \lambda_2)$ greater than zero; it could be chosen, for example, to be less than the smallest possible value attained by $\hat{\beta}$ for a particular configuration. A similar sort of transformation might be tried for $\alpha$ also. It would be interesting to collect information such as that in this chapter for $(\alpha, \theta)$ and $(\phi, \theta)$, where $\phi$ is a similar transformation of $\alpha$, for some values of $\lambda_1$. The parameterizations tried in Chapter 4 might also be examined by the methods of this chapter.

As has been mentioned previously, the distributions studied do not really provide enough information to enable any firm conclusions to be drawn, although some hypotheses for further investigation have been formulated. The study has, however, served to illustrate where the data base needs to be augmented. The first thing to note in the
discussion of this point is that only positive values of $\alpha$ and $\beta$ were considered in this chapter since it is obvious that the distribution of $\hat{\alpha}$ arising from the true parameters ($\alpha, \beta$) will be the same as that arising from ($-\alpha, \beta$), if $n$ is the same, except that all estimates of $-\alpha$ will be of opposite sign to those for $\alpha$, and so will all odd-ordered moments and the bias. (The same is true for $\mu$, of course.) As far as $\beta$ is concerned, there is no need to study configurations with negative values of $\beta$ since a change in the sign of $\beta$ is just equivalent to a reversal in the ordering of the doses, so that all assays can be considered as having positive slope. Nevertheless, there is a shortage of information on $\beta$ in particular. The problem, then, is to determine what additional configurations can most profitably be included in a further study, with the testing of suggested hypotheses in mind. Firstly, values of $P_2$ closer to 1.0, although they will result in high probabilities of indeterminacy for some parameterizations, should be added to broaden the spectrum of the results. Secondly, there should be at least one group of configurations with $\beta < 0.8473$; the group with $\beta = 2.19722$ should be enlarged; and one configuration group with
\( \beta > 2.19722 \) should be added. The next step might be to include more values of \( n \) (say 5 and 15) since this could be easily handled by the existing programme. If, say, the response sets \((.4,.5,.6), (.3,.5,.7), (.222,.5,.778), (.1,.5,.9), \) and \((.05,.5,.95)\) were taken as the bases and if for each of these, additional configurations with central dose response probabilities \(.6,.7,.8,.9\) were used for \( n = 5, 10, 15 \), then a good set of distributions might well be obtained for all estimators. The next step would then be to include configurations with different dose structures from that required by the Wilson-Worcester technique and then with different numbers of subjects at each dose. These would involve developing a new method of deriving the estimates. However, using the sufficient statistics and any symmetries in the model, an iterative procedure could be used to obtain all the values of the estimators as is usually done with individual observations from this sort of experiment. The main drawback to this sort of approach is the cost in terms of computer time necessary to complete by iteration so many maximizations. Alternatively, a sampling procedure such as that mentioned in section 5.1 could be employed to simulate the exact distribution, a standard maximization routine being used to obtain the maximum likelihood estimates.
Finally, then, although no positive firm conclusions have been made, valuable information, in the form of a pilot study, has been collected and directions for further investigation - as regards both additional configurations and possible better parameterizations - have been suggested. Whether follow-up studies are done or not, the information herein and the programmes developed are of some interest and value.
CHAPTER 6  CONCLUSIONS

As set out in the introduction (Chapter 1), the primary objectives of this study were the establishment of general criteria for independence which are not data-dependent and the development of techniques for using these criteria to obtain parameterizations with independent estimators. The approach taken in this work required in addition the establishment of criteria for normality suitable for use in the assessment of parameterizations selected by the criteria for independence. When the exact estimator distributions or their moments are known, the most suitable measures of correlation and normality are easily obtained. In the more usual case where neither the distribution function nor the moments are known exactly the best criteria are not so obvious and the question of whether a quantity suggested as a measure of independence or normality does in fact measure the appropriate characteristic must be considered. Whatever the criteria, a method of using them which ensures that a parameterization with the desired properties can always - or at least usually - be obtained is necessary; the actual application of the criteria to models proved to be the most difficult aspect of the problem. Although this work has not, in fact, resulted in the estab-
ishment of such a method, it constitutes a pilot study in this field and yields interesting results for some of the particular models considered.

This chapter contains a brief description and assessment of the criteria for independence and normality which have been suggested, followed by a discussion of the way in which the criteria for independence were applied, the problems encountered, and their possible solutions, and finally a summary of the results obtained for some of the specific models studied.
6.3

6.1 Establishment of Criteria

Throughout this work independence has been considered to be equivalent to zero correlation. Strictly speaking, this is not generally true, since two uncorrelated variables are only necessarily independent if they are jointly normally distributed. Because there is no single concise quantity from which independence may be deduced except in the case of the normal distribution, the development of criteria for independence would require as a first step the establishment of a satisfactory measure on which such criteria could be based; alternatively, measures of correlation could provide the basis for criteria and the distributions of the estimators for parameterizations satisfying these criteria could be examined for their proximity to the normal. The latter is the route taken in this study, on the assumption that estimators which are uncorrelated and approximately normally distributed are likely to provide a reasonable basis for independent inferences. In addition, lack of correlation alone should eliminate some of the problems encountered in iterative procedures.

Since generality is desirable, the criteria for independence considered are derived from properties
of the likelihood functions and asymptotic distributions, and so do not require knowledge of the estimator distributions or their moments. The assumption is made that a reparameterization which decreases the asymptotic correlation of the estimators has a similar effect on their exact correlation. Testing of the validity of this assumption is discussed briefly below. The criteria suggested for two-parameter models are:

(1) zero asymptotic correlation or, in other words, zero expectation of the mixed second derivative of the log likelihood
(2) zero mixed second derivative of the log likelihood
(3) each first derivative of the log likelihood a function of one parameter only.

These criteria are obviously not independent; in fact (2) and (3) are equivalent and a choice between the two can only be made on the basis of convenience, since in some situations it may be easier to solve two equations involving first derivatives than one equation involving the second derivative, in which case (3) would be preferred. (2) implies (1) but the reverse is not true, (2) being the stronger condition guaranteeing not only that the true asymptotic correlation is zero but that its estimate is zero also. Of the three criteria, (2) or (3) is to be preferred to (1), but since (1) is the weakest and so the one most frequently satisfied,
it is used most often. All of these criteria can easily be generalized to the multi-parameter case.

By way of assessing these criteria, a model whose parameter estimates were known to be independent (the two-parameter exponential distribution) was expressed in terms of two parameters whose estimators were dependent. and criteria (2) and (3) were applied in an attempt to reparameterize. The fact that the set of parameterizations selected included the original parameterization indicates that the application of these criteria based on asymptotic measures will at least sometimes yield a parameterization with the desired exact properties. Although this result obtained for one example only is not conclusive evidence, it does suggest that the measures on which the criteria are based do in fact measure the appropriate exact characteristic. In the more complex situation where neither an independent parameterization nor the estimator distribution is known, the only way to assess the criteria would be to simulate the distribution of the estimators of the suggested parameters for several parameter values and to estimate the exact correlation from the simulations. Because of the large amount of computer time required for such simulation and because
of the importance attached initially to the development of a workable method, no large-scale simulations were attempted.

Since all maximum likelihood estimators are asymptotically normally distributed, properties of the asymptotic distributions are obviously of no use in comparing parameterizations with respect to the normality of their estimator distributions. Therefore the general measures of normality suggested are of necessity based on properties of the likelihood function, even though it is known that a normal-type likelihood (that is, a likelihood function similar in shape to that arising from a normal distribution) does not necessarily correspond to a normal estimator distribution. One such measure was the set of third derivatives of the log likelihood function evaluated at the maximum, since if these are all zero, then the likelihood is approximately quadratic in the region of the maximum, as would be the likelihood arising from a normal distribution. The implications of non-zero values of these quantities are, however, unclear, particularly since their actual magnitude depends on the scale. Perhaps a function involving additionally some of the other coefficients in the Taylor series expansion of the log likelihood function would provide
a scale-free measure of asymmetry, as the measure of skewness does when exact moments are known. To test the effectiveness of the criterion of zero third derivatives, it was applied to a specific family of parameterizations of a single-parameter model whose estimator distribution was known (the exponential distribution). Comparison of exact coefficients of skewness for several parameterizations in the family indicated that the member selected by this criterion had in fact a more skew estimator distribution than several other members. This result suggests that this criterion is not a reliable criterion for normality.

The method most frequently used in this study to assess a selected parameterization with respect to the normality of its estimator distribution was visual examination of the contours of constant relative likelihood. This method, however, is of limited general application because it can be used only for given data and at most two parameters; in any particular instance, that is with given data, it may however be very useful for assessing a parameterization.

One type of normalizing transformation which has been discussed and could prove useful, is Bartlett's (1953) transformation of the score function to make the third
cumulant (and fourth, etc., as desired) agree to a given 
order with that of a normal variate. Although not investi-
gated in this work, one possible use of this method might 
be to transform some of the parameters in a parameterization 
selected to satisfy a criterion for independence; the new 
quantities so obtained would still satisfy the independence 
criterion but would, in addition, be approximately normally 
distributed. When applied to the parameter of the exponential 
distribution this transformation method led to confidence 
limits for the original parameter which were very close indeed 
to the exact limits. There are two difficulties involved 
in the application of this method. Firstly, the transformed 
parameters will always involve the data. For this reason 
such a transformation might best be performed after estimates 
have been computed and primarily for the purpose of obtaining 
confidence limits for the parameters selected by the independ¬
ence criterion. Secondly, the method is not always easy to 
apply, even with relatively simple score functions, as was 
illustrated by the example considered.

One conclusion reached regarding general measures 
of departure from normality was that such measures are neither 
plentiful nor well-documented. As these were not of primary 
concern, little time was devoted to their study.
6.2 Application of the Criteria

In this study, the suggested criteria for independence were applied only to a few specific two-parameter models (the two-parameter exponential and Gamma distributions, and the bioassay logistic model). There are two main reasons why this approach was taken. Firstly, it was hoped that some of the methodological problems involved in the application of such criteria would become evident so that their solution could be attempted for the simplest type of model. Secondly, if a method of application could be developed which was satisfactory for two-parameter models, it could quite possibly be generalized for use with models with more parameters; if on the other hand it became obvious that, even after some of the problems had been resolved, the application of these criteria would not consistently yield a reparameterization of a given two-parameter model, then such a method would not likely be satisfactory for use with general multi-parameter models, and clearly some other method would have to be sought. In this study, the problems encountered were such that one of the conclusions reached was that the attainment of the original objective was unlikely principally because the requirement of data-independence imposes constraints which create difficulties.
One major problem area was found to be the specification of a family of parameterizations from which those parameterizations satisfying some criterion for independence were to be selected. There are no general guidelines for choosing such a family, so that if one should happen to be lucky (or able to make a good guess) then the family specified for a particular model may actually include a parameterization which satisfies the criterion and is independent of data; otherwise another family must be tried. Moreover, for any given model there is no guarantee that an appropriate parameterization family exists, where an appropriate family is one which contains a parameterization satisfying the criterion being applied, or that one can find such a family if it does exist. For the Gamma distribution, for example, the first family tried yielded a parameterization, whereas for the logistic model several families were tried, all unsuccessfully. For the latter model some parameter transformations which were members of the specified families were found but they were dependent on data. Thus success in the search for a reparameterization is dependent at least partially on the existence and specification of an appropriate family. This dependency achieves greater significance as the number of parameters increases and the task of specifying an appropriate family becomes more
formidable. Since this step is a necessary one in the reparameterization procedure, until satisfactory solutions to the problems it presents have been obtained no further progress can be made in this direction.

Other difficulties can be envisaged as the number of parameters increases. For a k-parameter model there are \( \binom{k}{2} \) asymptotic correlations which must be made to vanish. Irrespective of the mechanics of obtaining a data-independent solution to such a system of equations, the probability of success can be expected to decrease with increasing \( k \). Alternatively, more sophisticated multivariate measures and techniques would have to be developed to handle larger numbers of parameters.

As stated earlier, the nature and magnitude of these problems are such that one can justifiably conclude that the attainment of the original objective in general is unlikely, unless a very different approach, having a greater chance of success, could be found. One possible alternative would be to consider the same objective but with one modification, namely removal of the constraint of data-independence. To attain this objective one could attempt to develop
either a method of reparameterization which could be applied after maximum likelihood estimates had been obtained if the primary reason for reparameterizing were to permit independent inferences to be made, or, preferably, a method which could be integrated into an iterative maximization routine. Zero estimated asymptotic correlations (or equivalently zero mixed second derivatives evaluated at the estimates) would seem a suitable criterion to use. Such methods as these might be more amenable to generalization, particularly if one family could be found to yield satisfactory parameterizations most of the time.
6.3 Conclusions Relating to Specific Models

Although comprehensive conclusions relating to the various models considered are given in the appropriate chapters, some of the more interesting results for the exponential and Gamma distributions, and the logistic model are worth summarizing here.

Estimator distributions for several parameterizations of the exponential distribution, all of which belong to the same general family, were compared with respect to their proximity to normality. Because the moments of the estimator distributions are known, the use of exact quantities relating to normality, such as skewness, plots of standardized densities, and confidence limits, was possible. Tabulation of certain of these measures and plots of the densities for some of the parameterizations and various numbers of observations made evident some of the trends within the parameterization family and permitted hypotheses about which parameterization was likely to be the 'best', in general, to be formulated. As mentioned in section 6.1, one parameterization included in the comparison was that selected by the criterion of zero third derivative of the log likelihood (evaluated
at the maximum), and it was found to be inferior to several other members of the same family, including one of the original parameterizations, on the basis of these exact properties. A particularly interesting comparison was that between the two usual parameterizations of the exponential distribution, which revealed that use of the parameter representing the mean lifetime (if the variable were, for example, time to failure) was superior to use of the inverse of this, the failure rate, as parameter, as far as making inferences based on assumptions of normality was concerned.

For the Gamma distribution with parameters \((\alpha, \beta)\) a parameterization satisfying a criterion for independence\(^1\) was found to be \((\theta_1 = \alpha, \theta_2 = \alpha\beta)\). To examine the effect of such a reparameterization on the estimated asymptotic correlation, a comparison was made of these estimates for the two parameterizations and several sets of generated Gamma data. In all cases the estimated asymptotic correlation coefficients of \((\hat{\theta}_1, \hat{\theta}_2)\) were found to be very close to zero and much lower than those of \((\hat{\alpha}, \hat{\beta})\). It would be interesting to simulate the estimator distributions for these two parameterizations for various parameter and \(n\) values and to compare these with respect to correlation

\(^1\) This criterion was that of zero expectation of the mixed second derivative of the log likelihood function.
and normality. The apparently successful reparameterization of a common distribution was to me both interesting and worthwhile and it seems possible that such a reparameterization might be of use in some situations.

Two different approaches to reparameterization of the bioassay logistic model have been considered in this work. Firstly, properties of the asymptotic estimator distributions were used to compare the three usual parameterizations \((\alpha, \beta), (\mu, \sigma),\) and \((\mu, \beta)\) and to provide criteria for the selection of new parameterizations. Secondly, for various parameter values and a particular assay structure exact estimator distributions were constructed and compared for the three mentioned parameterizations. This second approach was necessitated by the fact that the first yielded no satisfactory reparameterization. It is important here to draw together the results from these two approaches.

The asymptotic theory predicts that in a well-designed experiment, that is one such as that described by Anscombe (1964) as being an 'ideal' experiment or, more simply, one whose doses are centred on or near \(\mu,\)
the ED50, either the \((\mu, \beta)\) or the \((\mu, \sigma)\) parameterization is likely to yield less correlated estimates than is \((\alpha, \beta)\). On the basis of this theory, then, one would expect to find \((\hat{\mu}, \hat{\beta})\) and \((\hat{\mu}, \hat{\sigma})\) more highly correlated than \((\hat{\alpha}, \hat{\beta})\) for most configurations with central log dose 0 and nonzero values of \(\mu\). For the configurations studied in detail in Chapter 5, however, the results were not entirely as predicted by the asymptotic theory. To aid examination of where and why the exact and asymptotic results differ, both exact and asymptotic correlation coefficients for the studied configurations with non-zero values of \(\mu\) are given in Table 6.1. Several points should be noted from this Table. Firstly, as would be predicted by asymptotic theory, the asymptotic correlations for all three estimator pairs get larger with increasing asymmetry of the configuration, and except when \(\mu\) is close to 0, the asymptotic correlations of \((\hat{\mu}, \hat{\beta})\) and \((\hat{\mu}, \hat{\sigma})\) are greater than that of \((\hat{\alpha}, \hat{\beta})\). Secondly, also as expected, the exact correlation of \((\hat{\mu}, \hat{\sigma})\) in all cases but one (that being the one with smallest \(|\mu|\) ) exceeds that of \((\hat{\alpha}, \hat{\beta})\). Thirdly, and unexpectedly, the exact correlation of \((\hat{\mu}, \hat{\beta})\) is always either less than or little more than that of \((\hat{\alpha}, \hat{\beta})\). In addition, although asymptotic and exact coefficients agree quite closely for the \((\alpha, \beta)\) and
### Table 6.1

Asymptotic and Exact Correlation Coefficients for Three Parameterizations of the Bioassay Logistic Model for Configurations with $\mu \neq 0$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>0.84730</th>
<th>1.25276</th>
<th>2.19722</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.479</td>
<td>-1.0</td>
<td>-1.636</td>
</tr>
<tr>
<td></td>
<td>-0.324</td>
<td>-0.676</td>
<td>-1.107</td>
</tr>
<tr>
<td></td>
<td>-0.185</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a, $\beta$) Parameterization

<table>
<thead>
<tr>
<th>$r_a(\hat{\alpha}, \hat{\beta})$ *</th>
<th>0.126</th>
<th>0.251</th>
<th>0.370</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho(\hat{\alpha}, \hat{\beta})$, n=10</td>
<td>0.190</td>
<td>0.333</td>
<td>0.344</td>
</tr>
<tr>
<td>$\rho(\hat{\alpha}, \hat{\beta})$, n=15</td>
<td>0.169</td>
<td>0.336</td>
<td>0.450</td>
</tr>
</tbody>
</table>

($\mu$, $\beta$) Parameterization

<table>
<thead>
<tr>
<th>$r_a(\hat{\mu}, \hat{\beta})$ *</th>
<th>0.433</th>
<th>0.717</th>
<th>0.869</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho(\hat{\mu}, \hat{\beta})$, n=10</td>
<td>0.176</td>
<td>0.219</td>
<td>0.134</td>
</tr>
<tr>
<td>$\rho(\hat{\mu}, \hat{\beta})$, n=15</td>
<td>0.261</td>
<td>0.365</td>
<td>0.321</td>
</tr>
</tbody>
</table>

($\mu$, $\sigma$) Parameterization

<table>
<thead>
<tr>
<th>$r_a(\hat{\mu}, \hat{\sigma})$ *</th>
<th>-0.433</th>
<th>-0.717</th>
<th>-0.869</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho(\hat{\mu}, \hat{\sigma})$, n=10</td>
<td>-0.542</td>
<td>-0.789</td>
<td>-0.885</td>
</tr>
<tr>
<td>$\rho(\hat{\mu}, \hat{\sigma})$, n=15</td>
<td>-0.627</td>
<td>-0.849</td>
<td>-0.922</td>
</tr>
</tbody>
</table>

* $r_a(\hat{\phi}_1, \hat{\phi}_2)$ is the asymptotic correlation of $(\hat{\phi}_1, \hat{\phi}_2)$. 

[6.17]
6.18

\((\mu, \sigma)\) parameterizations, for many configurations there are substantial differences between the asymptotic and exact correlations of \((\hat{\mu}, \hat{\beta})\), the asymptotic being consistently larger (with only one exception).

It is possible that the differences between exact and asymptotic correlations demonstrated in Table 6.1 are not only the result of small values of \(n\), but also of the fact that the exact coefficients are derived from conditional estimator distributions and that sometimes the event upon which these distributions are conditioned, namely the determinacy of the estimator pair, has quite a low probability (for example as low as 0.88 for \((\hat{\mu}, \hat{\beta})\) for some configurations studied). This seems a reasonable explanation when the fact that the greatest discrepancies occur for the \((\mu, \beta)\) parameterization, which always has the largest probability of indeterminacy, is noted. Regardless of the explanation, one conclusion to be reached is that for the \((\mu, \beta)\) parameterization in particular the asymptotic correlation is not a good approximation to the exact for small \(n\). It is interesting that the parameterization involving \(\sigma\), although asymptotically equivalent to \((\mu, \beta)\) as regards correlation of estimators, appears much poorer when judged on the basis
of exact measures. Since both asymptotic and exact results point to the \((\mu, \beta)\) parameterization as the best of the three considered, despite the fact that the exact results are based on configurations which are particularly unfavourable to it, one might expect its superiority to be even more marked for better designs.

With regard to normality of exact estimator distributions, the \((\alpha, \beta)\) parameterization appeared to be the best for the configurations studied in detail. The inferiority of \((\mu, \beta)\) in this regard may be partly due to the fact that these configurations were not ideal for the estimation of \(\mu\). From examination of the relevant properties of the likelihood function none of the parameterizations could be considered superior to the others in respect of normality, although the likelihood function for the \((\mu, \gamma = \ln \beta)\) parameterization was found to be approximately symmetric in the region of the maximum.
6.4 Summary

A few brief statements can summarize the extent to which various objectives of this study have been attained. Firstly, criteria for independence which are not data-dependent have been established, and on the basis of work done herein appear to be satisfactory in the selection of a parameterization when an appropriate family can be specified. However, the specification of such a family in general proved a difficult problem with no evident solution. When the magnitude of this problem and its insolubility were realized the original objective requiring data-independence was not pursued further. The course of action recommended was modification of the objective making data-dependent parameterizations acceptable, but as mentioned elsewhere the development of suitable methods and appropriate modifications to the criteria to allow attainment of such an objective would constitute another entire project. Since for some models parameterizations satisfying the established criteria for independence were obtained, one might consider reparameterizing in the manner outlined in this work whenever possible (particularly with two-parameter models for which this method is generally simple to apply) with recourse to methods derived for attain-
ment of the modified objective, when such methods have been developed, if this failed.

The establishment of criteria for normality is an additional, although secondary, objective. Some difficulties were encountered in the development of useful measures for assessing normality of estimator distributions since these measures, in order to be general and data-independent, had to be based on the likelihood function, despite the fact that the inferences regarding an estimator distribution which could be made on the basis of a particular value (or set of values) for such a measure are far from clear. Thus, no adequate methods of assessing this property were found, except in cases where the exact moments or distributions were known.

The in-depth study of the exact estimator distributions for various parameterizations of the logistic model was inspired by the failure of the established reparameterization methods to yield a satisfactory parameterization when applied to this model. This study became a small distinct project, its primary objective being the determination of which of the three usual parameterizations, if any, was superior. with regard to the properties under consideration in this
thesis; this seemed a sufficiently important and interesting question to warrant such treatment. It is only unfortunate that time did not permit a more comprehensive examination in some of the directions indicated in Chapter 5. Nevertheless, the body of information presented in that chapter could provide a good basis for further investigation of the model and its parameterizations.
APPENDIX I

THE METHOD OF WILSON AND WORCESTER

Although the model which Wilson and Worcester (1943) use to describe a quantal response bioassay looks different from the model used in Chapters 4 and 5, the two are actually equivalent. With \( x_i \) representing the logarithm of the \( i \)th dose and \( P_i \) the true probability of response at that dose, they express the dose-response relationship as

\[
P_i = \frac{1}{2} \left[ 1 + \tanh \rho (x_i - \gamma) \right]
\]

which can alternatively be written

\[
P_i = \left[ 1 + \exp(-2\rho (x_i - \gamma)) \right]^{-1}
\]

which is the logistic model. The pair of parameters to be estimated is then \((\gamma, \rho)\) which is in fact \((\mu, \beta/2)\) in terms of the parameterizations considered in Chapter 5.

With \( n \) subjects at each of \( k \) doses and observed proportions of responsive subjects \( p_1, p_2, ..., p_k \), the first derivatives of the log likelihood function of \((\gamma, \rho)\) are

\[
\frac{\partial L}{\partial \gamma} = -n \rho \sum_{i=1}^{k} (2p_i - 1) + n \rho \sum_{i=1}^{k} \tanh \rho (x_i - \gamma)
\]

\[
\frac{\partial L}{\partial \rho} = -n \sum_{i=1}^{k} (2p_i - 1) x_i + n \sum_{i=1}^{k} \tanh \rho (x_i - \gamma) x_i
\]
\[
\frac{\partial \mathbf{L}}{\partial \rho} = n \sum_{i=1}^{k} (2p_i - 1)(x_i - \gamma) - n \sum_{i=1}^{k} (x_i - \gamma) \tanh \rho(x_i - \gamma)
\]

If there are only 3 equally spaced log doses \(x-c, x, x+c\), where \(c\) is a constant, then the normal equations are

\[
\begin{align*}
3 \sum (2p_i - 1) &= \tanh \hat{\rho}(x-c) + \tanh \hat{\rho}(x) + \tanh \hat{\rho}(x+c) \quad (I2) \\
2(p_3 - p_1) &= -\tanh \hat{\rho}(x-c) + \tanh \hat{\rho}(x+c) \quad (I3)
\end{align*}
\]

It is evident from these that \(D = \sum p_i\) and \(B = p_3 - p_1\) are sufficient statistics for \((\gamma, \rho)\).

If the relationship between the hyperbolic tangent of a sum and the sum of the hyperbolic tangents of the components of the sum is utilized and if \(A = 2D - 3\), \(X = \tanh \hat{\rho}(x-\gamma)\), and \(C = \tanh \hat{\rho}c\), then I2 and I3 can be neatly written as

\[
A = \frac{X - C}{1 - XC} + X + \frac{X + C}{1 + XC} \quad (I4)
\]

\[
2B = -\frac{X - C}{1 - XC} + \frac{X + C}{1 + XC} \quad (I5)
\]
From these equations it is found that

\[ C = \frac{2B}{X^2 - AX + 2} \]

and substitution of this expression into I4 yields the cubic equation

\[ 3X^3 - 4AX^2 + (6 + A^2 - 4B^2)X - 2A = 0 \] (I6)

which, for given \( A \) and \( B \) can be solved numerically for \( X \).

In fact, in their paper Wilson and Worcester give the approximate solution to I6 as

\[
X_0 = \frac{4A}{9} - \frac{b}{a} + \frac{729}{a} \left( \frac{b}{a} \right)^3
\]

with

\[ a = 1458 - 189A^2 - 972B^2 \]
\[ b = 162A - 20A^3 - 432AB^2 \]

which if not accurate enough in itself provides a good starting value for any iterative method of solution.

If \( X \) is the solution to I6, then the estimators are

\[ \hat{\gamma} = \hat{\mu} = x - \frac{c \tanh^{-1} x}{\tanh^{-1} C} \]
\[ \hat{\rho} = \hat{\beta}/2 = \frac{1}{c} \tanh^{-1} C \]
\[ \hat{\alpha} = 2 \tanh^{-1} x - \frac{2x}{c} \tanh^{-1} C \]
These explicit expressions for the maximum likelihood estimators derived by Wilson and Worcester are invaluable in the sort of study undertaken in Chapter 5. Since iteration is needed only in the solution of a single cubic equation, and then only for some pairs of sufficient statistics, the maximum likelihood estimates can be obtained with much less computer time than an iterative maximization procedure would require. The approximate solution to \( I_6 \) is good enough that iteration is not always necessary; \( X_0 \) and \( f(X_0) \) can be calculated, and only if \( |f(X_0)| \) is larger than is acceptable is an iterative procedure necessary, and even then \( X_0 \) is such a good starting value that not many iterations should be needed to attain a solution with the desired accuracy.

In a small pilot study preceding the investigation of Chapter 5, the method of stratified sampling outlined by Berkson (1955) was used to obtain 100 triples of data from a bioassay with response probabilities \((0.1, 0.5, 0.9)\), central log dose zero, and \( c \), the spacing between log doses, 2.593. In each case the approximate root of \( I_6 \), \( X_0 \), was determined (to four decimal places) and \( |f(X_0)| \) computed. Of the 100 triples, 13 yielded indeterminate estimates for \(( \alpha, \beta )\); of the remaining 87, 37 had \( |f(X_0)| \geq 0.0001 \)
and of these only 8 gave \[ |f(X_\circ)| \geq 0.001. \] So the approximate root \( X_\circ \) does in general seem to be a good approximation to the real root. Although the proportion of possible outcomes for which \( X_\circ \) is close enough to the real root is quite possibly dependent on the configuration, it is likely that for most configurations anyway \( X_\circ \) would be considered a good approximation.
APPENDIX II

INDETERMINACIES IN ESTIMATES OF THE PARAMETERS OF THE LOGISTIC MODEL

For the purposes of this thesis, an indeterminacy is considered to be a pair of estimates one or both of which is either undefined or infinite; as will be made evident later, if the three parameter pairs \((a, \beta)\), \((\mu, \beta)\), and \((\mu, \sigma)\) are considered, then every pairing of infinite, undefined, and defined, finite estimates occurs for at least one possible triple of observations.

Since the probability of indeterminacy is entirely dependent on the configuration and since it is obviously to an experimenter's advantage to secure finite, defined estimates of his parameters, his prior knowledge about the nature of the true response curve will certainly guide him to a choice of doses which can be expected to give him such estimates with a high probability. (It is not always possible to do this, and anyway, even the best-designed experiment can have such an improbable outcome that an indeterminacy obtains.) In this study, however, doses have been fixed and a range of true parameter values deliberately
chosen, and for some of these values the probability of indeterminacy for some parameter pairs is sizeable. Therefore, the determination of a satisfactory method for dealing with these indeterminacies is quite important.

The problems posed by these indeterminacies are in fact results of the general characteristics of the distributions of the maximum likelihood estimators: for all configurations, the distributions are discrete with unknown densities, and comprise countable numbers of finite-valued points of non-zero probability along with the infinite and indeterminate values. This means that moments and other measures like bias cannot be calculated; nor can meaningful, complete frequency plots be constructed. If, then, the distributions of estimators are to be examined and if the usual methods of describing and comparing these distributions are to be employed, the indeterminacies must be eliminated in some way. One solution which seems reasonable is to exclude these indeterminacies and compare the distributions of the remaining estimates. These distributions would, of course, all have to be standardized so that the sum of the probabilities of all admissible pairs of estimates would be one, if they are to be compared with each other and with the normal. In other words, bivariate
distributions of different pairs of estimators could all be conditioned on both variates being defined and finite and the properties of these distributions studied and compared. An alternative solution which will also be discussed is to systematically adjust some of the observed proportions in samples producing indeterminate estimates so that indeterminacies can no longer occur.

In this appendix indeterminacies are categorized according to the sufficient statistics producing them; probabilities of indeterminacies in the different pairs of parameters are briefly compared; the indeterminacies are described in terms of the best-fitting logistic curves and an attempt is made at the physical interpretation of these; for some particularly interesting cases contours of constant log likelihood are plotted. Finally, there is a discussion of some of the possible ways of dealing with the indeterminacies and a justification of the method employed in this thesis.
Categorization, Description, and Interpretation of Indeterminacies

In a k-dose experiment, indeterminacies generally occur when \((k-1)\) or more of the observed proportions are zero or one, and for certain identifiable symmetric response sets. With \(k = 3\) and log doses \(x-c, x, x+c\), Table II.1 gives the infinite and undefined estimates of all of the parameters \(\alpha, \beta, \mu,\) and \(\sigma\) classified according to the sufficient statistics producing them. The samples yielding these statistics are also shown in the Table, and it is obvious that these are the samples which one would expect to provide little or no information about the nature of the response curve; in general they indicate that either the choice of doses was unsatisfactory or else the experimenter was unlucky. It is interesting that all samples of types 1 and 2 and some of those type 3 can be fitted exactly by degenerate logistic curves.

---

1. Not all such samples produce indeterminacies. See later discussion.

2. These are functions which are zero or one at some of the given doses. Since a logistic function is defined as an asymptotic function with asymptotes 0 and 1, these extreme values should theoretically never be attained for finite dose. In real life, however, such values are bound to obtain, especially for small \(n\), so that unless the degenerate members of the family are acceptable the model will be inadequate in circumstances such as those listed in Table II.1.
TABLE II.1

INDETERMINACIES CLASSIFIED BY SUFFICIENT STATISTICS*  

<table>
<thead>
<tr>
<th>TYPE</th>
<th>SUFFICIENT STATISTICS</th>
<th>SAMPLES</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\beta})</th>
<th>(\hat{\mu})</th>
<th>(\hat{\sigma})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>(</td>
<td>B</td>
<td>=-1; \text{any } D)</td>
<td>(a) ((0,0,1);(1,1,0))</td>
<td>(\pm \infty)</td>
<td>(\pm \infty)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) ((1,0,0);(0,1,1))</td>
<td>(\pm \infty)</td>
<td>(\pm \infty)</td>
<td>(x-c&lt;\hat{\mu}&lt;x)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) ((0,p,1);(1,p,0), \text{ } p\neq 0,1.)</td>
<td>(\pm \infty^\dagger)</td>
<td>(\pm \infty)</td>
<td>(x)</td>
<td>0</td>
</tr>
<tr>
<td>2.</td>
<td>(</td>
<td>B</td>
<td>=D \text{ or }</td>
<td>B</td>
<td>=3-D,</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) ((p,0,0);(p,1,1), \text{ } p\neq 0,1)</td>
<td>(\pm \infty)</td>
<td>(\pm \infty)</td>
<td>(x-c)</td>
<td>0</td>
</tr>
<tr>
<td>3.</td>
<td>(B=0; \text{any } D)</td>
<td>(a) ((0,0,0);(1,1,1))</td>
<td>(\log</td>
<td>\frac{\sum p_i}{3-\sum p_i}</td>
<td>)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) ((p,p',p), \text{except } p=p'=0,1)</td>
<td>-many possible values-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Sufficient statistics are \(B=p_3-p_1; D=\sum_{i=1}^{3} p_i\) where \((p_1,p_2,p_3)\) is a set of observed proportions responding to log doses \(x-c, x, x+c\).

\(\dagger\hat{\alpha}=\logit p\text{ if }x=0.\)

\(\ddagger\ddagger\hat{\mu}\text{ has many possible values when }\hat{\alpha}=0.\)
Silverstone (1957) gives a good discussion of these curves for indeterminacies in \((\alpha, \beta)\).

When \(|B| = 1\) and all three observed proportions are either zero or one, each sample can be fitted exactly by a different set of degenerate curves, which is an illustration of the fact that the sample gives no information about the nature of the true response curve in one dose interval; that is, while the substance being assayed appears to be completely ineffective in one dose range, it is completely lethal (in the sense that it produces universal response) in another, with unknown effect in between. For example, the data set \((0,0,1)\) can be exactly fitted by the curves
\[
\hat{p} = 0 \quad \text{when} \quad x' \leq x
\]
\[
\hat{p} = 1 \quad \text{when} \quad x' \geq x + c
\]
where \(x'\) is the log dose, and \(\hat{p}\) is undefined between log doses \(x\) and \(x + c\). Any curve of the form below passes through the sample points and so provides the best fit:
Of course, the straight line transform of these curves is a set of vertical straight lines such that \( x < \hat{\mu} < x + c \).

The situation of \( \hat{\mu} \) is well-illustrated by contours of constant log likelihood for \( (\mu, \sigma) \) with log doses \(-1,0,1\) in Figure 1.

Although the contours obviously converge to \( \hat{\sigma} = 0 \), none of them actually enclose or touch any point on the line \( \sigma = 0 \). The contours are all open-ended towards the \( \mu \)-axis converging to the segment of straight line with \( \sigma = 0 \) and \( 0 < \mu < 1 \) as the maximum likelihood estimates.

For \((0,p,1)\), where \( p \neq 0,1 \) an exact fit is provided by

\[
\hat{p} = 0 \quad \text{when} \quad x' < x
\]

\[
= 1 \quad \text{when} \quad x' > x
\]

\[
= p \quad \text{when} \quad x' = x.
\]
Contours of constant negative log likelihood for $(\mu, \sigma)$ with $n=10$, log doses $-1, 0, 1$ and response triple $(0, p, 1), 0 \leq p \leq 1$. 
This sample provides a little more information about the nature of the response curve than did the previous sample, if only to suggest concentration on doses closer to x. These observations imply that there is no effect as far as one can tell from any dose less than x and complete response at doses larger than x, which means an infinite rate of response around dose x, and thus the transform of the logistic curve to a straight vertical line. Of course, it is the logistic model which forces this inference from the data. The contours of constant log likelihood for this sample have been plotted in Figure 2 for p = 0.5 and log doses \(-1, 0, 1\). These appear very similar to those for \((0, 0, 1)\) except that the contours for \((0, 0.5, 1)\) are all open only at the point \((0, 0)\), which is the maximum likelihood estimate. In both cases the contours do not enclose the maximum likelihood estimate as is usually the case. All samples of type 1 in the Table are the same in this respect.

A unique degenerate logistic curve can be made to pass through the sample points for each sample of type 2, although these all become vertical straight lines when transformed. The logistic curves are similar to those which fit \((0, p, 1)\), \(p \neq 0, 1\); for example, \((0, 0, p)\), \(p \neq 0, 1\) is fitted exactly by
\[ \hat{p} = 0, \text{ when } x' < x + c \]
\[ = p, \text{ when } x' = x + c \]
\[ = 1, \text{ when } x' > x + c \]

The samples which intuitively contain the least information about the nature of the response curve are 
(0, 0, 0) and (1, 1, 1) since the only action they suggest is that a different series of doses - either higher or lower than those used - be tried, although how different one cannot tell. It is not surprising, then, that although there is in each case a logistic curve which fits the data exactly, it is completely undefined beyond doses \( x-c \) and \( x+c \) and can be any one of an infinite number of degenerate curves. This means that for each parameter there is a multiplicity of values maximizing the likelihood.

For other samples of type 3, namely those of the form 
( \( p, p', p \) ), where \( p \) and \( p' \) are not both 0 or 1, the best fitting logistic function is a horizontal line given by

\[ \hat{p} = \frac{3}{\sum_{i=1}^{3} p_i/3} \]

When \( p' = p \), this curve passes through the sample points, but when \( p' \neq p \) the fit is not exact. In fact there will be several different samples of this type which have the
same sufficient statistics, and hence one logistic curve will provide the best fit in all cases. For example, 

\[(0.4, 0.4, 0.4), (0.3, 0.6, 0.3), (0.2, 0.8, 0.2)\]

each give \(\hat{\beta} = 0\) and \(\hat{\alpha} = \log(1.2/1.8)\). If \(\hat{p}\) is different from 0.5, \(|\hat{\mu}|\) is infinite, while if \(\hat{p} = 0.5\), there are many possible values for \(\hat{\mu}\); in either case, no information is obtained about \(\mu\), the dose eliciting a 50% response. In all the samples of type 3 the horizontal line logistic curve suggests that a wider range of doses might profitably be examined or else that the model is unsuitable.

It is interesting that samples \((0,1,0)\) and \((1,0,1)\) belong to the last group and give finite and defined estimates of \(\alpha\) and \(\beta\). Of course, such samples have the same sufficient statistics as others in the group and are not, in fact, any less informative than most other members of the group; they are, however, much less probable than most and one would be very unlucky indeed to obtain such observations unless \(n\) were very small or the model were unsuitable.

Some other samples which one might intuitively expect to yield indeterminacies but which in fact provide defined and finite estimates of all parameters are \((0,1,p)\).
(1, 0, p), (p, 0, 1), and (p, 1, 0), p \neq 0, 1.

Examination of (0, 1, p), for example, reveals that its sufficient statistics are the same as those for the sample (p_1, 1-2p_1, p + p_1), where p_1 can have any non-zero value which gives a sample with at most one zero or one, and there is always at least one such p_1. Thus, (0, 1, p), although not fitted exactly by any logistic curve, degenerate or otherwise, is best fitted by one which has finite values for estimates of all four parameters.

Probabilities of Indeterminacies

In Chapter 5, Section 5.4.6, the relative sizes of probabilities of indeterminacies of estimator pairs (\hat{\alpha}, \hat{\beta}), (\hat{\mu}, \hat{\sigma}), and (\hat{\mu}, \hat{\beta}) have been discussed for log doses \(-1, 0, 1\). In this section, these will be briefly considered for the more general set of log doses x-c, x, x+c.

As shown in 5.4.6, (\hat{\mu}, \hat{\beta}) always has the highest probability of indeterminacy, while the particular configuration determines which of the other two pairs has the higher probability of indeterminacy in any instance. For fixed \alpha, when \beta is large \hat{\beta} = \pm \infty (\hat{\sigma} = 0) is more
likely to obtain than if \( \beta \) were smaller, while \( \hat{\beta} = 0 \) is more probable the smaller the value of \( \beta \). Of course, even when \( \hat{\sigma} = 0 \), \( \hat{\mu} \) need not be defined, so that sometimes when \( \hat{\beta} = \pm \infty \), \( (\hat{\mu}, \hat{\sigma}) \) and \( (\hat{\alpha}, \hat{\beta}) \) are both indeterminate, although whenever \( \hat{\beta} = 0 \), \( \hat{\alpha} \) is finite and defined while \( (\hat{\mu}, \hat{\sigma}) \) is indeterminate. Thus, for any fixed \( \alpha \), \( \beta \) is the factor which determines which pair of estimators is more likely to be indeterminate, since this depends on which is the larger of \( \Pr (\hat{\beta} = 0) \) and \( \Pr (\hat{\beta} = \pm \infty , (\hat{\mu}, \hat{\sigma}) \) defined and finite). Letting \( PD (\hat{\theta}_1, \hat{\theta}_2) \) represent the probability that both \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are defined and finite, the relationship between \( PD (\hat{\mu}, \hat{\sigma}) \) and \( PD (\hat{\alpha}, \hat{\beta}) \) can be expressed as follows:

\[
P_D(\hat{\mu}, \hat{\sigma}) = P_D(\hat{\alpha}, \hat{\beta}) - \Pr(\hat{\beta} = 0)
\]

\[
+ \sum_{np=1}^{n-1} \left\{ \Pr((0,p,1)) + \Pr((1,p,0)) + \Pr((0,0,p)) \\
+ \Pr((1,1,p)) + \Pr((p,0,0)) + \Pr((p,1,1)) \right\}
\]

It is interesting to note from Table II.1 that in all cases with exactly two of the four estimators finite

1. In addition, as the central dose probability gets farther from 0.5, the probability of indeterminacy increases for all three pairs.
and defined, these two are always either $\hat{\alpha}$ and $\hat{\beta}$ or $\hat{\mu}$ and $\hat{\sigma}$.

Thus, any admissible pair of estimators chosen from the four given, other than $(\hat{\alpha}, \hat{\beta})$ and $(\hat{\mu}, \hat{\sigma})$, would be indeterminate for all samples listed in the Table and so would have a higher probability of indeterminacy than either $(\hat{\alpha}, \hat{\beta})$ or $(\hat{\mu}, \hat{\sigma})$.

Methods of Dealing with Indeterminacies

Although every sample producing an indeterminacy can be fitted by a logistic curve, the nature of these indeterminate estimates and of the whole estimator distributions, as discussed earlier, means that some special consideration must be given to the indeterminate estimates or to the samples producing them. Two possible methods of eliminating indeterminacies will be considered herein.

One way of dealing with indeterminate estimates is to exclude them and study the bivariate distributions which include only defined, finite values of the estimators; if these distributions are conditioned on both variates being defined and finite, they can be compared with each other and with the normal distribution. Intuitively this seems a
reasonable suggestion, since in general the samples which lead to indeterminacies provide little or no information about the nature of the response curve, and will be used by the experimenter merely as a guide as to which doses to use in a repeat experiment, if the experiment can be repeated. Unless a repetition is impossible, inferences about the nature of the response curve will not be attempted from such uninformative samples, but will be made to as great an extent as possible from the subset of all possible samples which are informative and which do produce finite and defined estimates. It seems sensible, therefore, to consider the distributions which include only those estimates which are defined and finite.

An alternative method for dealing with indeterminacies - and the only other one to be considered here - is due to Berkson (1955). He suggests substituting $\frac{1}{2n}$, where $n$ is the number of subjects per dose, whenever 0 is observed and $(1-\frac{1}{2n})$ for an observed proportion of 1, his main argument for this procedure being that a method of estimation must provide a useful estimate under all circumstances and that if it fails to do so then this is a fault in the estimator which must be rectified. Such substitutions do eliminate indeterminacies in $(\alpha, \beta)$ and do provide more
information about a response curve - the one that provides the best fit if the observations after adjustment had been obtained - but may lead to incorrect inferences since $0$ and $1$ observations need not imply that $1/2n$ and $(1-1/2n)$ respectively would have been observed if that were possible. Substitution of $1/4n$ has as much theoretical justification as $1/2n$ but could lead to very different inferences in some instances. There seems to be no justification for trying to get information from the samples in Table II.1 by altering the observations, when it is evident that the actual observations do not provide such information. In fact it is the poor choice of doses, unsuitability of the model, or bad luck which account for such observations being obtained and not the method of estimation; indeterminacy of estimates is merely a reflection of the lack of information contained in the sample. More detailed discussion of the effects and advantages of, and the drawbacks to the use of this method can be found in Berkson (1955) and in Silverstone (1957).

The use of conditional distributions, then, rather than the adjustment method of Berkson, is the solution to the problem presented by indeterminacies which is employed for the study in Chapter 5. One reason for this is that
the physical nature of an assay allows such observations as \((0, p_1, p_2)\) to obtain and for many values of \(p_1\) and \(p_2\) such samples do not yield indeterminacies and do provide information about the response curve that is just as respectable as that obtained from any other sample; there seems to be no reason for altering such a sample by Berkson's rule. Secondly, it seems reasonable to confine interest to the distribution of those estimates which provide information and can be used in inferences about the response curve since these are the only estimates of use to the experimenter. Thirdly, there is no justification for adjusting sample values to values not observed for the purpose of obtaining defined and finite estimates which may well lead to incorrect inferences about the response curve.

1. Silverstone (1957) gives an example of the numerical effect on the estimates of \(\alpha\) and \(\beta\) of application of Berkson's 2n-rule to the sample \((0.2, 4.5, 1)\) with \(n = 10\). For both parameters, the estimates obtained after adjustment according to Berkson's rule are larger by almost one standard error than the estimates obtained from the original sample.
B.1

BIBLIOGRAPHY


