ANALYSIS AND COMPARISON OF SOME STATISTICAL MODELS

by

JAMES KEITH LINDSEY

Thesis submitted for the Diploma of Membership of Imperial College and the degree of Ph.D. in the University of London

ABSTRACT

Often, in the analysis of scientific data, some relationship between observed responses and the response conditions is sought by means of a mathematical model. Random variation in the measured response, described by a probability model, permits statistical analysis of the data. Then, a given statistical model is considered more plausible than another if it makes the observed data more probable, as measured by the likelihood function.

For various reasons, a natural linear normal statistical model has been used traditionally whenever possible. This model is extended to nonlinear, transformed response normal models for biological response surface methodology, and an example from fisheries biology provided.

The natural linear model is derived for a member of the exponential family in general, using the binomial probability model as a specific example.

By considering observed responses as discrete measurements, a method of comparison and plausibility of fit of probability models is developed using the multinomial model as the basis. The proper interval widths are determined by a graphical method. A number of numerical examples are provided.

A complete analysis of data is described when various mathematical and probability models are possible. Numerical examples are given for normal, exponential, Poisson, and

binomial models. The roles played by a power transformation of a normal response are determined by using the relative likelihood function.

In special circumstances, specific inferences about a given parameter in a statistical model, in the absence of knowledge about any other, may be made using the conditional likelihood function. Examples involve the binomial and nonlinear normal distributions.

The exact Fisherian test of significance is described and applied to contingency table examples. Comparison of the assumed and the approximate likelihood functions shows how good the approximation is when an asymptotic test is used.

ACKNOWLEDGEMENTS

The author is very much indebted to his supervisor, Professor D.R. Cox for the guidance and encouragement provided during the course of this research. He would also like to thank Professor J.G. Kalbfleisch and Dr. D.F. Alderdice for the stimulation they provided.

This work was supported by a Graduate Research Fellowship from Imperial Oil Limited of Canada, except for parts of chapters II and V which were done under a research contract with the Fisheries Research Board of Canada.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS USED IN THE TEXT

		Definition on page
ANOVA	analysis of variance	12.
CLF	conditional likelihood function	97.
CRL	conditional relative likelihood	99.
H_o	null hypothesis	109.
LF	likelihood function	14.
MLE	maximum likelihood estimate	16.
MM	mathematical model	12.
PM	probability model	11.
PSD	power series distribution	37.
RL	relative likelihood	15.
SM	statistical model	12.
SS	sum of squares	26.
SSM	simple statistical model	13.

CHAPTER I

PROBABILITY MODELS, MATHEMATICAL MODELS, AND STATISTICAL INFERENCE

1. Introduction

The analysis of scientific data usually involves, among others, two basic purposes. First, the scientist may desire to ascertain if some specific hypothesis, made before the data were obtained, is viable. Second, the scientist wishes to determine what information the given set of data can provide about the unspecified parts of general hypotheses made before the data were obtained. For example, the data may be assumed to arise from a normal distribution with unknown mean. The scientist wishes to determine if his specific hypothesis that the mean has value μ_0 is tenable and also what information the data provide about the plausibility of various values of the mean within the general hypothesis of normality.

This brief outline illustrates the difference between the two purposes: whereas the first requires an absolute judgement of viability, the second only requires a comparative judgement of plausibility. Of course, a sufficiently strong judgement of the implausibility of one of the compared hypotheses will yield a judgement of inviability. But, this inviability statement results from some other hypothesis than the specified one being much more plausible and not from an absolute measure.

The analysis of scientific data always carries the assumption that it is conceivable to obtain further such data. This assumption will be necessary for developing an absolute judgement of viability. For a specific hypothesis, long run probability statements may be made about possible sets of data (and functions of such data, i.e. statistics) not yet available, and from these statements, an absolute criterion can be developed. Further discussion of this purpose in the analysis of scientific data will be left until chapter VII. Until that time, we shall be concerned with developing a comparative measure of plausibility to fulfil the second purpose.

In certain situations, information about the probability distribution of unknown parameters in the hypothesis will be available before the data are obtained. Then, this information may be incorporated into the methods in the following chapters using Bayes' theorem. Such procedures will not be discussed in what follows.

A number of numerical examples will be given to illustrate the procedures developed. Except for several examples in chapter IV, these are drawn from the biological sciences.

2. Probability Models

Many sets of scientific data may be considered to consist of either counts of discrete individuals or measurements of continuous variables made on individuals. In either case, an observation, y_k, will be made on the response variable, Y. Because of the finite limits of any measuring instrument, the measurements of a continuous response variable will actually be discrete; they may be designated as $y_k \pm \frac{1}{2}\Delta y_k$.

In order to carry out a statistical analysis of the data, one must assume that the response variable is subject to some random fluctuation, ε_k, so that

$$y_k = Y + \varepsilon_k \ . \tag{1.1}$$

Then, the observations will have a frequency distribution. Any hypothesis must specify one or more possible probability functions which may represent (approximate) the frequency distribution. These hypothesized probability functions will be called probability models (PM's). The PM will usually contain unknown parameters, $\underline{\phi}$, which must be estimated from the data and will be represented by $F(\underline{y};\underline{\phi})$. When Y is continuous,

$$F(\underline{y};\underline{\phi}) \doteq dF(\underline{y};\underline{\phi})\Delta\underline{y} = f(\underline{y};\underline{\phi})\Delta\underline{y} \ . \tag{1.2}$$

In the following chapters, only PM's from the exponential family, of the form

$$F(y;\mu) = \int_{y-\Delta y}^{y+\Delta y} \exp[A(y)B(\mu)+C(y)+D(\mu)]dy \tag{1.3}$$

will be considered, where μ is the expected (mean) response, $\mu = E(Y)$. For a discrete PM, dy = 1 and the integral sign disappears. If $B(\mu) = \theta$, the parameter θ is called natural. Common PM's belonging to the exponential family include the binomial, Poisson, normal, and exponential distributions. Equation (1.3) may be generalized by the addition of parameters not related to the mean.

3. Mathematical Models

Often, scientific data consist of observations of responses, Y_i, measured under a number of different conditions, i, either determined by experimental design (e.g. analysis of variance (ANOVA)) or by nature (e.g. some regression problems). In such cases, there will be a different PM under each condition. Usually, the probability function will be assumed to remain the same for all i, with only the parameter values varying with i. This variation in parameter values may be described by some mathematical function which will be known as the mathematical model (MM). Incorporation of the MM into the PM yields a complete statistical model (SM). Of course, PM's not incorporating a MM varying with i may still be considered as SM's with the parameter values set equal to (unknown) constants, as in equation (1.5) below.

If the MM is a linear function of the parameters, then it is called linear. If, in addition, all of these parameters are natural, it is called a natural linear MM. Common examples include normal theory linear regression and ANOVA, although these are special cases because of the presence of a variance parameter (see chapter V, section 2).

For a given PM, the least informative (about the relationship between conditions and response) MM occurs when a different parameter (vector) exists for each condition, i, for example

$$B(\mu_i) = \theta_i \ . \qquad (1.4)$$

If the various conditions do not actually affect the response, the MM will be of the form

$$B(\mu_i) = \theta \ . \qquad (1.5)$$

Usually, desirable MM's lie between these extremes and are derived by using information about the response conditions. Again, examples include the regression and ANOVA models to explain variation in the mean of the response. In succeeding chapters, various MM's involving the mean will be considered.

Natural linear SM's from the exponential family are usually statistically most useful because of the existence of sufficient statistics for all of the parameters. But, often theoretical aspects of the scientific problem will point to some other form of MM, as especially considered in chapter II. Elsewhere, natural linear SM's will be used if possible.

After the specification of all parameter values in a SM, the probability of observing any given set of data may be calculated exactly. A SM with all parameter values specified will be called a simple statistical model (SSM). Then, any SM is made up of a set of possible SSM's, labelled by the unknown parameter values. If various SM's are considered possible before the data are obtained, these form a still larger set of SSM's.

4. Likelihood Inference

The preceding two sections have outlined what is to be assumed before the data is obtained: some combinations of mathematical models with probability models to form statistical models. No prior probabilities are assumed for either various models or various parameter values.

We wish to determine which SSM's are plausible in light of the data. For each hypothesized SSM of the set, the probability of observing the given data is calculated. Then, the criterion of plausibility for making comparative statistical inferences, as described in the first section of this chapter, is to consider more plausible that SSM which makes the observed data more probable.

Consideration of the set of SSM's in this way, with the observed data given and fixed, yields a likelihood function (LF, of possible SM's and parameters) for the data,

$$L(\underline{\phi}_s, s; \underline{y}) \doteq F_s(\underline{y}; \underline{\phi}_s), \qquad (1.6)$$

where F_s specifies the SM, and $\underline{\phi}_s$ the (vector of) parameter(s) within model s.

Use of the likelihood function for making statistical inferences, as originally proposed by Fisher. has been recommended by a number of authors; Fisher (1958, 1959), Barnard, Jenkens, and Winsten (1962), Birnbaum (1962), Anscombe (1961), Feigl and Zelen (1965), Sprott and Kalbfleisch (1969) and others. The procedures are well known, at least when only one SM is available and only the parameter values are unknown. For a simple explanation in this case, see Lindsey (1970).

In making inferences using the likelihood function, all statements about SSM's are relative. We do not say that a SSM, SM, MM, or PM is implausible unless another corresponding model is much more plausible. This is especially relevant to the procedures developed in chapter IV for discrimination among PM's. Hence, the relative likelihood (RL) function,

$$R(\underline{\phi}_s, s) = L(\underline{\phi}_s, s; \underline{y}) / L(\hat{\underline{\phi}}_s, \hat{s}; \underline{y}), \qquad (1.7)$$

may always be used for making inferences, where $\hat{s}$ denotes the most plausible SM and $\hat{\underline{\phi}}$, the most plausible parameter values within this SM.

CHAPTER II

MODIFICATION OF THE NATURAL LINEAR NORMAL MODEL

1. Introduction

For many statistical problems arising from scientific data, as outlined in the first chapter, some form of natural linear MM combined with the normal PM is entirely satisfactory for obtaining the desired information from the data. Hence, normal theory regression analysis and ANOVA have been developed as prime tools for the practising statistician. Natural linear normal theory models have a number of desirable features which enhance their use: (i) robustness to departures from the assumption of a normal PM; (ii) linear likelihood equations which may easily be solved to yield explicit maximum likelihood estimates (MLE's) of the unknown parameters; (iii) sufficient statistics so that the MLE's contain all of the information in the data about the parameters. The second feature means that these models may easily be used without the need for powerful computing equipment to make the necessary calculations. With this form of SM, exact long run probability statements may easily be made about data not yet observed, no matter how much data is already available (see chapter VII) since the required distributions are well tabulated.

But, even given these considerations, the natural linear normal theory SM may often be unsatisfactory, either due to theoretical scientific reasons or to implausibility. Two distinct methods may be used to improve on the natural linear normal SM. The parameters of the PM may vary in a nonlinear manner as the responses change under different conditions making abandonment of the natural linear MM necessary. Or, the observed responses may not follow a normal PM.

Introduction of a nonlinear MM entails loss of the second and third features above. More powerful computing equipment is usually necessary to solve the nonlinear likelihood equations. Exact long run probability statements may no longer easily be made. But, often nonlinearity is at the root of a theoretical discrepancy.

Two approaches may be used if the responses do not appear to follow a normal PM. The normality assumption may be abandoned or some transformation of the observed response which agrees satisfactorily with the normal PM may be used. The first alternative will be considered in succeeding chapters. Traditionally, in adopting the second alternative, some completely specified transformation, such as $\sin^{-1}\sqrt{Y}$ for binomial type data, is derived from theoretical statistical considerations. More recently, transformations involving unknown parameters have been introduced, as described by Box and Cox (1964). These will be used throughout this chapter. Transformation of the response is occasionally used to solve the statistical problems inherent in a nonlinear MM, as when a logarithmic transformation is applied to the response in a regression problem, while retaining a linear regression MM, instead of introducing the nonlinear exponential MM.

As a specific example of the procedures used when departures from a natural linear normal SM are important, the analysis of response surfaces will be considered. Box and Wilson (1951) originally introduced response surface methodology, using a second degree polynomial regression MM, as a procedure for determining the combination of levels of various factors which produce the optimum response. Efficient methods have been developed for determining this optimum (e.g. the method of steepest ascent) from experimental results, as well as for devising efficient experimental designs to gain maximum information from a given number of points in the factor space (e.g. composite and rotatable designs). When important deviations from the polynomial MM occur, Box and Tidwell (1962) have suggested power transformations of the factor variables. This not only increases the efficiency of determination of the optimum response, but also provides a more accurate picture of the shape of the response surface as a whole. For departures from a normal PM, Box and Cox (1964) proposed a procedure for estimating transformation parameters applied to the response variable; see also Dolby (1963) and Draper and Hunter (1969).

With the introduction of these two types of nonlinear parameters, both the PM and the MM may be considered nonlinear. Thus, any exact long run probability statements, either about statistics derived from the data (e.g. significance levels and the probability statements upon which confidence intervals are based) or about the parameters (e.g. fiducial intervals), are difficult or impossible to calculate, unless sufficient data are available for asymptotic properties to hold. As stated in chapter I, we are interested in what information is available from the observed data without using asymptotic long run statements.

A number of procedures in the analysis of response surfaces, besides the estimation problems, are developed extensively here, since they will be of use in analyzing data which may be considered to arise from a nonnormal PM. The use of response surface methodology is discussed in relation to the biological field of ecology.

2. The Role of Biological Response Surface Methodology

Response surface techniques may be used to describe many biological phenomena (e.g. survival, growth rate, oxygen consumption) within a range of levels of various environmental variables (factors). For limited changes in the environmental variables (i.e. within a limited region of the factor space), quadratic systems, such as equation (2.3) below with $\alpha_j = \alpha_k = 1$ (all j, k), are often adequate for approximating the relationship between a response and levels of several of the factors. In this region, a local maximum may be reached and the quadratic relationship will approximate this. However, a large portion of the variation due to the treatments (levels of the environmental variables) is often unexplained after fitting the quadratic surface. This is associated with departures of the true biological response from that expressed by the quadratic approximation. Usually, the required higher order effects are difficult to determine because of the greatly increased cost and difficulty of a more complex biological experiment. The use of such transformed response surfaces, with the addition of one parameter for each factor variable (forming a nonlinear MM), provides much greater flexibility for more adequate representation of the actual surface than does the quadratic expression.

Often, biological response data, as gathered, does not follow a normal PM very well. In most cases, this may be corrected by the introduction of some transformation of the response variable. Here, the power transformation of Box and Cox (1964), with one estimable parameter, is used to fulfil better the assumptions of normality and constant variance of the SM. Draper and Hunter (1969) demonstrate that the MLE of this transformation serves as a type of average in performing these two functions (see also chapter V, section 10). In addition to the improved PM which this yields, the response transformation also provides the same benefits as does use of the nonlinear MM, in that the shape of the response surface becomes more flexible.

The use of such transformations, both of the response and of the environmental variables, may also indicate alternate units of measurement, which may be employed to simplify (linearize), as well as to make more biologically meaningful, the relationship. For example, if the transformation of a response of time to death is estimated as an inverse power ($\gamma = -1$ in equation (2.1) below), a more meaningful unit of measurement would appear to be death rate. In this case, the scientist is concerned with readily interpretable transformations which still allow the SM to explain the observed response well and which may provide a simpler model (if possible, linear in the transformed variables) for further theoretical and experimental work.

In summary, transformations can (i) provide useful insight as to different units of measurement for variables to be used with a linear SM; (ii) provide a more accurate (than the original linear SM) representation of the relationship under study by means of a nonlinear model, and perhaps lead to a refinement of the model; (iii) allow the response to fulfil more closely the probability assumptions (PM).

3. Notation

The general SM to be considered is of the form

$$y_i^{\gamma} = B(\underline{x}_i, \underline{\phi}) + \varepsilon_i \quad , \qquad (2.1)$$

where the PM is

$$y_i^{\gamma} \sim N[B(\underline{x}_i, \emptyset), \sigma^2] \qquad (2.2)$$

and the MM is

$$B(\underline{x}_i, \emptyset) = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij}^{\alpha_j} + \sum_{j=1}^{m} \sum_{k=j}^{m} \beta_{jk} x_{ij}^{\alpha_j} x_{ik}^{\alpha_k} \quad (i=1,\ldots,n). \qquad (2.3)$$

From these equations, the LF, maximized with respect to the variance, σ^2, is

$$L(\underline{\alpha}, \underline{\beta}, \gamma; \underline{y}) = \left(\sum_{i=1}^{n} [y_i^{\gamma} - B(\underline{x}_i, \emptyset)]^2\right)^{-n/2} \gamma^n \prod_{i=1}^{n} y_i^{\gamma-1} \Delta y_i \quad , \qquad (2.4)$$

where $\gamma^n \prod_{i=1}^{n} y_i^{\gamma-1}$ is the Jacobian of the transformation of the response variable. In the more general form, the normal LF maximized with respect to the variance is

$$L(\emptyset, \gamma; \underline{y}) = \hat{\sigma}^{-n} \gamma^n \prod_{i=1}^{n} y_i^{\gamma-1} \Delta y_i \quad , \qquad (2.5)$$

where $\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i^{\gamma} - B_i[\mu(\emptyset)])^2/n$.

4. Analysis of the Response Surface

4.1 Estimation of Parameters

With a nonlinear SM, the normal equations are nonlinear in the parameters and some iterative procedure such as Newton's method must be used to determine the solutions. Various methods have been devised which are efficient with certain models and/or certain data (see Draper and Smith (1966) for some of the more important cases). A three-step linearization procedure which has been found satisfactory for the above SM is described below. As suggested by Box and Cox (1964), a linear transformation of the response variable simplifies the LF by eliminating the Jacobian of equation (2.4). Let the geometric mean, $\prod_{i=1}^{n} y_i^{1/n}$, for the response variables, be denoted by $\dot{y}$. Then, apply the transformation

$$y \to y^{\gamma} \to \frac{y^{\gamma}-1}{\gamma \dot{y}^{\gamma-1}} .$$

The Jacobian of the complete transformation is equal to unity and the new variable is continuous at $\gamma = 0$. This linear transformation of y^{γ} does not affect the MLE of γ. A similar linear transformation may be applied to x_j:

$$x_j \to x_j^{\alpha_j} \to \frac{x_j^{\alpha_j}-1}{\alpha_j}$$

for continuity at the origin. The family of transformations then becomes

$$y \to \frac{y^{\gamma}-1}{\gamma \dot{y}^{\gamma-1}} = y^{(\gamma)} \quad (\gamma \neq 0, \pm\infty),$$

$$\underline{x} \to \frac{\underline{x}^{\underline{\alpha}}-1}{\underline{\alpha}} = \underline{x}^{(\underline{\alpha})} \quad (\underline{\alpha} \neq 0, \pm\infty).$$

Denote these transformations, which will be used in this form only for the estimation procedures of this section, by $y^{(\gamma)}$ and $\underline{x}^{(\underline{\alpha})}$.

Initial estimates, $\underline{\alpha}_0$ and γ_0, of unity have proven adequate using the estimation procedure here described.

The three step procedure is as follows:

Step I: After substituting the initial estimates of the power parameters into equation (2.3), the MLE's of the β_j's are calculated (i.e. linear least squares estimates). These estimates are conditional on the power parameters being set at the initial estimates.

Step II: Equation (2.3) is linearized with respect to the power parameters of the factor space (see Draper and Smith (1966) pp. 267-273). This is equivalent to taking terms up to first order in the Taylor series expansion of $B(\underline{x}_i,\underline{\beta},\underline{\alpha})$ about the initial estimates, $\underline{\alpha}_o$. The resulting equation is

$$y_i^{(\gamma_o)} = B(\underline{x}_i,\underline{\beta},\underline{\alpha}_o)+\sum_{j=1}^{m}(\alpha_{j1}-\alpha_{jo})B_{\alpha_j}(\underline{x}_i,\underline{\beta},\underline{\alpha}_o)$$

$$=\beta_o+\sum_{j=1}^{m}\beta_j x_{ij}^{(\alpha_{jo})}+\sum_{j=1}^{m}\sum_{k=j}^{m}\beta_{jk}x_{ij}^{(\alpha_{jo})}x_{ik}^{(\alpha_{ko})}$$

$$+\sum_{j=1}^{m}\Big[\beta_j x_{ij}^{(\alpha_{jo})}+2\beta_{jj}x_{ij}^{(2\alpha_{jo})}+\sum_{k>j}^{m}\beta_{jk}x_{ij}^{(\alpha_{jo})}x_{ik}^{(\alpha_{ko})}\Big]$$

$$\cdot(\alpha_{j1}-\alpha_{jo})\log x_{ij} \quad .$$

Now $[\beta_j x_{ij}^{(\alpha_{jo})}+2\beta_{jj}x_{ij}^{(2\alpha_{jo})}+\sum_{k>j}^{m}\beta_{jk}x_{ij}^{(\alpha_{jo})}x_{ik}^{(\alpha_{ko})}]\log x_{ij}$ do not contain the unknown parameters, α_{j1}, of this step, and thus may be considered as new independent variables. Then, MLE's may be calculated for the new parameters, $(\alpha_{j1}-\alpha_{jo})$, using only linear likelihood equations. Throughout step II, the values of $\underline{\beta}$ from step I are used.

Step III: The MM now stands as

$$y_i^{(\gamma_o)} = \beta_o+\sum_{j=1}^{m}\beta_j x_{ij}^{(\alpha_{j1})}+\sum_{j=1}^{m}\sum_{k=j}^{m}\beta_{jk}x_{ij}^{(\alpha_{j1})}x_{ik}^{(\alpha_{k1})} \quad ; \qquad (2.6)$$

where $\underline{\beta}$ is obtained from step I and $\underline{\alpha}_1$ from step II. Note that in equation (2.6), $\underline{\beta}$ is not the MLE when $\underline{\alpha} = \underline{\alpha}_1$ but when $\underline{\alpha} = \underline{\alpha}_o$. Since much of the computation time in this procedure involves calculating cross product matrices, use of the same $\underline{\alpha}_o$ in this step as in step I has been found to be more efficient than recalculating $\underline{\beta}$ for the new $\underline{\alpha}_1$ after step II. This usually causes a few more iterations to be required for convergence, but the total number of actual calculations, and therefore the computing time, is reduced.

At this stage, the right hand side of equation (2.4) can be considered as a new dependent variable and $y^{(\gamma)}$ linearized as in step II for the MM. This yields the relationship

$$B(\underline{x}_1, \underline{\beta}, \underline{\alpha}_0) = \frac{y_i^{\gamma_0}-1}{\gamma_0 \dot{y}^{\gamma_0-1}} + (\gamma_1-\gamma_0)\left[\frac{y_i^{\gamma_0}\log y_i}{\gamma_0 \dot{y}^{\gamma_0-1}} - \frac{y_i^{\gamma_0}-1}{\gamma_0^2 \dot{y}^{\gamma_0-1}} - \frac{(y_i^{\gamma_0}-1)\log \dot{y}}{\gamma_0 \dot{y}^{\gamma_0-1}}\right].$$

The parameter $(\gamma_1-\gamma_0)$ is estimated by simple linear regression yielding a new estimate of γ.

The three steps are repeated, using the values of $\underline{\alpha}$ and γ from steps II and III as new initial estimates in step I, until some convergence criterion is met. In most examples analyzed, using two and three factor MM's ($m = 2$ or 3 in equation (2.3)), convergence to three or more digits occurred within ten iterations. Exceptions involve SM's where the RL function graphs of the parameters are very flat, hence pointing to no definite power parameter values. In these cases, values near the MLE's usually result after ten iterations.

This method of calculating the MLE's is also used for estimating the parameters for the maximized RL function by placing a restraint on the iteration technique so that one of the parameters remains at a fixed value. The estimates so obtained are then substituted into the RL function.

4.2 Adequacy of the Statistical Model

Since no alternatives to the PM of equation (2.2) are considered, adequacy will only be discussed within the framework of this PM. This implies that the PM with a transformed normal response estimated from the data will be considered adequate. When more than one value of the response variable is observed under at least some of the environmental conditions, MM (1.4) may be used, i.e. a different mean for each condition. Then, the pure error sum of squares (SS) of Table 2.1 is a measure of experimental error and may be used in the estimation of the variance in the normal LF (2.5), providing the basis for all comparisons, i.e. this LF becomes the denominator of equation (1.7). The plausibility (adequacy) of the response surface MM is then determined by comparing the LF (2.4) with this. This is the RL for lack of fit in Table 2.1. Both LF's of the RL are maximized over all parameters, including Υ and σ^2.

If only one observation is made at each point in the factor space, the pure error SS is not calculable since MLE's of Υ and σ^2 cannot be obtained using this MM. Then, LF (2.4) must be used as the basis of all comparison, i.e. the response surface MM must be assumed adequate for all further comparisons. Some other more ad hoc procedures may be used to check the validity of this assumption. Measurement of the response at nearly adjacent factor points will yield an approximate estimate of the required variance. The linear polynomial response MM (equation (2.3) with $\Upsilon = 1 = \alpha_j$) can be considered a second order approximation to the Taylor series expansion of some unknown function. Box (1954) has emphasized that the experimental design should allow estimation of some higher order effects to give an indication of adequacy of the approximation. This also applies when the transformations are considered as changes of units of measurement. Box and Hunter (1965) have suggested that an indication of desired changes in the MM may be discovered by making several runs of an experiment using different points in the factor space at each run. The MLE's of the parameters in the model are given by $\hat{\underline{\phi}}$, which is calculated from the combined observations for all runs. In addition, $\hat{\underline{\phi}}_i$ may be calculated from each run i and the RL for $\hat{\underline{\phi}}$ versus each $\hat{\underline{\phi}}_i$ used to discover if any significant relationship of the run estimates of $\underline{\phi}$ to the location of the observation points exists. If so, appropriate modifications will need to be made to the SM. The relationship of expected to observed responses (e.g. plots of residuals) also provide a useful indication of possible inadequacy, especially of the PM.

4.3 Inferences about the Parameters

Often, the statistician is concerned only with two values for each parameter in making inferences about a SM: the MLE (or unbiased minimum variance, etc.) and the point estimate where the parameter disappears from the SM, e.g. $\beta_j = 0$ or $\alpha_j = 1$ in equation (2.3). In the present case for a linear MM, this is exemplified by use of the ANOVA table which contains significance tests of the hypotheses that various parameters and combinations of parameters are zero. Thus, the two values are that providing the most plausible

SM given the data, and that simplifying the SM by elimination of the parameter(s) in question. This is an important first step which will be considered further; nevertheless, to understand the SM more completely, the effect of variation of the parameter values on the model should be considered through use of the LF, by means of contour plots of the RL function and of the RL function maximized with respect to all parameters except that of interest. Consideration of various parameter values besides the two indicated above may show that some theoretically more satisfactory value is plausible.

In the case of the nonlinear SM, standard F tests are no longer valid in the ANOVA table, since probabilities for the ratios of SS's cannot be obtained from existing tables. The column of F values in the ANOVA table may be replaced by a column of maximized RL's of the various parameters and combinations of parameters being zero. Such ANOVA tables for linear SM's are useful in familiarizing oneself with the properties of the equivalent tables in the nonlinear case. The construction of such a table for a replicated experiment is shown in Table 2.1. Without replications, the residual SS cannot be split into lack of fit and pure error SS's. Note the relationship, $R = (SS_A/SS_B)^{-N/2} = (1+kF)^{-N/2}$, using equation (2.4), where F is the F ratio for A and B, and k is the ratio of degrees of freedom.

The use of orthogonal polynomials provides maximum independence in making inferences about various elements of $\underset{\sim}{\beta}$. The method of Robson (1959) for calculating the polynomials when the spacing between levels of a factor, $x_j^{\alpha_j}$, are unequal, has been used. For the linear MM, this procedure is equivalent to the method mentioned by Box in Davies (1956, p. 519) for orthogonalizing the quadratic term in the MM.

If the maximized RL of some coefficient parameter being zero or of some power parameter being unity is high, this indicates that elimination of that parameter from the SM is possible without affecting the adequacy of the model, for values of the other parameters near their MLE's (anywhere, if the parameters can be estimated independently). If elimination of a parameter is implausible, a plot of the maximized RL function will show the plausibility of various other values, and may point to a plausible interpretable point in the parameter space.

If the experimenter is primarily interested in determining the optimum response conditions, a RL function may be used to give an idea of the precision of the estimates of the factor coordinates of the centre; see Box and Hunter (1954).. If equation (2.3) is differentiated with respect to the various factors, a system of linear equations in the transformed coordinates is obtained:

$$\frac{\partial y^{\gamma}}{\partial x_{ij}^{\alpha_j}} = \beta_j + 2\beta_{jj} x_{sj}^{\alpha_j} + \sum_{k \neq j}^{m} \beta_{jk} x_{sk}^{\alpha_k} = 0 \quad .$$

These constraint equations can be substituted directly into equation (2.3) eliminating the β_j's, yielding

$$y_i^{\Upsilon} = \beta_o + \sum_{j=1}^{m} \sum_{k=j}^{m} \beta_{jk}(x_{ij}^{\alpha_j} x_{ik}^{\alpha_k} - x_{ij}^{\alpha_j} x_{sk}^{\alpha_k} - x_{sj}^{\alpha_j} x_{ik}^{\alpha_k}) \quad .$$

Since $\underline{x}_s = (x_{s1},\dots,x_{sm})$ is the MLE of the centre of the surface, the RL of various possible centre points may be calculated using this equation as a MM, and maximizing over the remaining parameters.

4.4 Plotting the Surface

Plotting the response surface can play a valuable role in interpreting the inferential results, especially if plots are made for various plausible SSM's, i.e. for various plausible sets of parametric values.

With a linear response surface model, both canonical analysis of the MM and plotting of the contours are extremely important in understanding the surface. With nonlinear MM's, canonical analysis yields little information about the shape of the surface but does provide a simple technique for calculating points to plot on the surface. This is one important reason for choosing a MM of the form (2.3). The following is a simple procedure for determining the centre of the fitted surface.

Let $B = (\zeta_{jk}\beta_{jk})$ be the matrix of coefficients in MM (2.3), where $\zeta_{jk} = 1$ if $j = k$ and $\frac{1}{2}$ if $j \neq k$; let $C = (\frac{1}{2}\beta_j)$ be the vector of 0.5 times the coefficients and let $X = (x_{sj}^{\alpha_j})$ be the vector of the centre point. The vector of the centre coordinates is found by solving the set of simultaneous linear equations

$$BX = C \tag{2.7}$$

and the response centre by

$$y_s^{\gamma} = \beta_o + X'C \quad .$$

Canonical analysis of the surface uses the eigenvalues and eigenvectors of the matrix of coefficients, B, which will be denoted by $\underline{\lambda}' = (\lambda_1,\ldots,\lambda_m)$ and by $\underline{v}_i' = (v_{1i},\ldots,v_{mi})$ (i = 1,...,m) respectively. Canonical variables are derived:

$$z_{ik} = \sum_{j=1}^{m} (x_{ij}^{\alpha j} - x_{sj}^{\alpha j}) v_{kj} \quad .$$

The point $\underline{x}_i$ in the factor space is transformed to the point $\underline{z}_i = (z_{i1},\ldots,z_{im})$ in the canonical space, which has its origin at the centre of the surface and its axes along the principal axes of the conical equation in the transformed variables, $x_{ij}^{\alpha j}$. The canonical equation of the surface is

$$y_i^{\gamma} - y_s^{\gamma} = \sum_{j=1}^{m} \lambda_j z_{ij}^2 \quad . \tag{2.8}$$

Equation (2.3) is a conical equation if the units of the coordinates are considered to be $x_{ij}^{\alpha j}$ and of the response to be y_i^{γ}. Equation (2.8) then shows the conical shape of the surface in terms of these transformed units. When the surface is considered in terms of the original units, x_{ij}, it is, of course, nonlinear and not conical.

If all of the λ_i have the same sign, some optimum (maximum or minimum) has been reached in terms of both the $x_{ij}^{\alpha j}$ and the x_{ij} factor spaces. If the eigenvalues are of differing signs, various nonlinear surfaces are possible.

Since an (m+1) dimensional space is involved in the study of a response surface, difficulties arise in viewing the entire space when $m > 2$. When $m = 2$, contours of various levels of the response may be plotted on coordinates of the two factors. In higher dimensional spaces, slices may be taken on hyperplanes to obtain a surface suitable for plotting in two dimensions. For example, with $m = 3$, the plane, $x_1 = k_1$, might be considered and contours of y plotted on the x_2x_3 coordinates along this plane. Thus, all plotting of surfaces can be considered in terms of the two factor response surface. The method outlined below provides the calculations for forty points (x'_{i1}, x'_{i2}) on each response contour, $y = y_k$, but the number of points may easily be altered; see Lindsey (1968).

When the constraints of the hyperplane are substituted into equation (2.3), two eigenvalues, λ'_1, λ'_2, for the matrix of coefficients may be obtained. Let

$$u_{1,i} = [(y_k^{\gamma}-y_s^{\gamma})/\lambda'_1]^{\frac{1}{2}}\cos[(i-1)\pi/20] \qquad (i = 1,\ldots,11) \qquad (2.9)$$

if the eigenvalues have the same sign, and

$$u_{1,i} = x_{11}^{\alpha_1}+[(y_k^{\gamma}-y_s^{\gamma})/\lambda'_1]^{\frac{1}{2}}-x_{11}^{\alpha_1}\cos[(i-1)\pi/20] \qquad (i=1,\ldots,11) \qquad (2.10)$$

if they are of opposite sign, where x_{11} is a limit on the size of the x_1 factor. Also

$$u_{2,i} = [(y_k^{\gamma}-y_s^{\gamma}-\lambda'_1u_{1,i}^2)/\lambda'_2]^{\frac{1}{2}} \qquad (i = 1,\ldots,11). \qquad (2.11)$$

Then,

$$u_{1,i} = u_{1,42-i} = -u_{1,22-i} = -u_{1,20+i} \qquad (i = 1,\ldots,11) \qquad (2.12)$$

and

$$u_{2,i} = -u_{2,42-i} = u_{2,20+i} = u_{2,22-i} \qquad (i = 1,\ldots,11). \qquad (2.13)$$

The points are given by

$$x_{i1}^{\alpha_1} = v'_{11}u_{1,i}+v'_{21}u_{2,i}+x_{s1}^{\alpha_1}$$
$$x_{i2}^{\alpha_2} = v'_{12}u_{1,i}+v'_{22}u_{2,i}+x_{s2}^{\alpha_2} \qquad (i = 1,\ldots,40) \quad . \qquad (2.14)$$

Throughout equations (2.9) to (2.14), x_{i1} and x_{i2} signify the first and second remaining variable factors after the constraints of the slice are applied to equation (2.3). If equation (2.9) is used, only response conditions below the maximum (conversely, above the minimum) will be calculated, whereas with equation (2.10) contours both above and below are used.

5. An Example

Fry and Hart (1948) performed experiments on the effects of acclimation (x_1) and experimental (x_2) temperature experience on the swimming speed of goldfish (Carassius auratus). Cruising speed of a fish is defined as the speed (in ft./min.) at which the fish can swim steadily for a considerable period of time, although fatigue will begin after some hours. A test fish was thermally adapted to water at the acclimation temperature before being transferred directly to a rotating chamber containing water at the experimental temperature where determination of speed was made.

Unfortunately, although three different fish were used for each response determination, Fry and Hart only give average results for each set of three fish, as reproduced in Table 2.2. Individual response curves plotted by Fry and Hart for the various acclimation temperatures show that the response surface does not have a conical shape; hence, use of nonlinear parameters should result in improved fit.

Although not performed specifically for response surface analysis, the experiment lends itself well to such analysis. With no hypothetical MM available, that of equation (2.3) provides a useful basis for the analysis. No special PM seems theoretically justifiable, so that some form of transformed normal PM may prove satisfactory.

Since only one (average) measurement is available at each response condition, the pure error SS cannot be calculated and the plausibility of the response surface MM (2.3) cannot be determined. Hence, lack of fit and pure error are omitted from the ANOVA Tables 2.3 and 2.4. Note that the residual SS has only 13 degrees of freedom in Table 2.4, three being used in estimation of the power parameters.

MLE's of the power parameters are calculated to be $\hat{\gamma} = 0.1080$, $\hat{\alpha}_1 = 1.6344$, and $\hat{\alpha}_2 = 1.2939$, yielding a nonlinear MM

$$y^{0.11} = 1.48 - 6\text{x}10^{-5}x_1^{1.63} + 5\text{x}10^{-3}x_2^{1.29} - 5\text{x}10^{-6}x_1^{3.26} - 1\text{x}10^{-5}x_2^{2.58} + 3\text{x}10^{-5}x_1^{1.63}x_2^{1.29} \quad .$$

From the graphs in Figures 2.1, 2.2, and 2.3 of the RL functions of the power parameters, the maximized RL's of the powers being unity are $R(\alpha_1=1) = 0.0004$, $R(\alpha_2=1) = 0.05$, and $R(\gamma=1) = 0.008$. The maximized RL of the linear SM (i.e. all power parameters unity) as opposed to the nonlinear model is $9\text{x}10^{-6}$, indicating that the simplified linear SM is not plausible. From Figure 2.3, the maximized RL of γ = 0.0, or a logarithmic transformation of the response is high ($R(\gamma=0.0) \doteq 1.0$). Comparison of the graphs of Figures 2.1, 2.2, and 2.3 indicates that α_1 and α_2 are estimated from the data more precisely than γ.

The plausibilities of eliminating coefficient parameters are listed in the ANOVA Tables 2.3 and 2.4. Only the parameter β_1 in the linear MM

$$y = 21.84+0.97x_1+4.49x_2-0.18x_1^2-0.21x_2^2+0.28x_1x_2 \quad ,$$

with $R(\beta_1=0.0) = 0.24$, could plausibly be eliminated, but this is not of interest since the linear MM has been shown to be implausible. Note that none of the coefficient parameters is estimated independently in either MM, although orthogonal polynomials have been used.

The response centre or point of optimum response (maximum cruising speed) is given as $\underline{x}_s = (26.2^\circ C,\ 27.5^\circ C)$ with y_s = 99.9 ft./min. for the nonlinear SM as compared with $\underline{x}_s = (22.8^\circ C,\ 26.6^\circ C)$ with y_s = 92.6 ft./min. for the linear SM. The shapes of the linear and nonlinear surfaces may be compared by means of the contour plots of Figures 2.4 and 2.5. RL function graphs of the centre coordinates are omitted here (as are those of the coefficient parameters) but intervals with maximized RL greater than 0.1 are (25,27) for x_1 and (26,29) for x_2 for the nonlinear MM.

The nonlinear SM resulting from the analysis appears to provide a good predictive model for the response surface, although little of theoretical biological interest seems to have resulted. Measurement of the response, the cruising speed of the goldfish with logarithms or, equivalently, use of the log normal PM may require further analysis.

CHAPTER III

NATURAL LINEAR STATISTICAL MODELS FROM THE EXPONENTIAL FAMILY

1. Introduction

As mentioned in chapter I, the normal PM is the most frequently used member of the exponential family; both normal theory linear regression and ANOVA use natural linear MM's. Extension to other members of the exponential family is usually straight forward. A number of concrete examples will be provided in chapter V, including analyses of numerical data.

In this chapter, for illustrative purposes, some natural linear MM's for discrete members of the exponential family will be considered, with detailed development for the binomial PM. The discrete exponential PM's are members of the power series distribution (PSD), examples of which include the binomial, Poisson, and logarthmic PM's.

The PSD is defined on the set of non-negative integers, T. With parameter ϕ and series function $g(\phi) = \sum_{y=0}^{\infty} a(y)\phi^y$, the PSD is defined by the probability function

$$F(Y) = \frac{a(Y)\phi^Y}{g(\phi)}, \quad Y \in T, \quad \phi \in \Phi = (\phi : 0<\phi<r) \quad , \qquad (3.1)$$

where r is the radius of convergence of the power series of $g(\phi)$. By comparison with equation (1.3), $A(Y) = Y$, $B[\mu(\phi)] = \log \phi$, $C(Y) = \log a(Y)$, and $D[\mu(\phi)] = -\log g(\phi)$. Then, the natural parameter is $\theta = \log \phi$.

In the following development of natural linear MM's for the PSD, alteration of the probability function when simplified MM's are introduced is discussed. Although not necessary for the LF's so far considered, this alteration will be needed for making more sensitive inferences about individual parameters in a plausible SM in chapter VI using conditional probability functions and also for making tests of significance in chapter VII.

2. The Natural Linear Mathematical Model

When the discrete variable Y_i is observed under a number of different conditions i, the general SM will be

$$F(\underline{Y};\underline{\phi}) = \prod_{i=1}^{n} \frac{a_i(Y_i)\phi_i^{Y_i}}{g_i(\phi_i)} \quad . \tag{3.2}$$

Reduction in the number of parameters results when some (hopefully) theoretically justifiable MM dependent on the conditions of observation is introduced. With members of the exponential family, use of a natural linear MM insures that sufficient statistics will exist for all of the parameters. In equation (3.2), each observed y_i is individually sufficient for the corresponding parameter, ϕ_i. For a given PM of the PSD family, the LF derived from equation (3.2) forms the basis for plausibility comparisons in the presence of natural linear MM's.

The simplest SM to which equation (3.2) may be reduced is

$$F(\underline{Y};\phi) = \prod_{i=1}^{n} \frac{a_i(Y_i)\phi^{Y_i}}{g_i(\phi)} = \phi^{\Sigma Y_i} \prod_{i=1}^{n} \frac{a_i(Y_i)}{g_i(\phi)} \quad , \tag{3.3}$$

where the response conditions have no effect upon the observed response. In this case, $t = \Sigma y_i$ is the sufficient statistic for the new parameter $\theta = \log \phi$. However, note that

$$\sum_{\underline{y}} \prod_{i=1}^{n} \frac{a_i(y_i)\phi^{y_i}}{g_i(\phi)} \neq \sum_{t} \phi^t \prod_{i=1}^{n} \frac{a_i(y_i)}{g_i(\phi)} \quad , \tag{3.4}$$

since various vectors $\underline{y}$ will yield the same t.

In the same way, the analogue of any normal theory natural linear SM, such as linear regression or ANOVA, may be derived. When the natural linear MM chosen has fewer parameters, and hence fewer sufficient statistics, than response conditions, an adjustment must always be made to the probability function. This is equivalent to integration of a continuous probability function using a Jacobian.

Suppose $\underline{t}' = (t_1,\ldots,t_s)$ ($s < n$, the number of response conditions) are sufficient statistics for the parameters of a natural linear MM. Then, the factor $\prod a_i(y_i)$ becomes

$$a'(\underline{t}) = \sum_{\underline{y}}{}' \prod_{i=1}^{n} a_i(y_i) \quad , \tag{3.5}$$

where the summation, Σ', is restricted by the s restraints defining $\underline{t}$. For example, using the SM (3.3) with a binomial PM,

$$a'(t) = \sum_{\underline{y}}{}' \left[\prod_{i=1}^{n} \binom{N_i}{y_i} \right] = \sum_{\underline{k}} \left[\prod_{i=1}^{n-1} \binom{N_i}{k_i} \binom{N_n}{t - \sum_{i=1}^{n-1} k_i} \right] = \binom{\sum_{i=1}^{n} N_i}{t} \quad ,$$

so that

$$F(\underline{y};\underline{\phi}) = \prod_{i=1}^{n} \binom{N_i}{y_i} \phi_i^{y_i} (1+\phi_i)^{-N_i}$$

becomes

$$F(t;\phi) = \binom{\sum_{i=1}^{n} N_i}{t} \phi^t (1+\phi)^{-\Sigma N_i} \quad .$$

ANOVA with a PSD may be exemplified by the analysis of a two-way randomized block design with one observation per block. Even with only one observation per block, the interaction effect may be included in the MM since the variance need not be estimated, as it usually must with a normal PM. Then, the natural linear MM may be

$$\log \phi_{ij} = \theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad , \qquad (3.6)$$

with $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$, where i and j label the two-way system. Thus, effects are measured on a logarthmic scale. This MM is equivalent to that incorporated in SM (3.2), since each has the same number of parameters.

Examples from PSD's might include: (i) if the observations, y_{ij}, are the numbers of tags returned in a year from a given species of fish, equal numbers of tagged fish being released of a number of ages, J, at a number of locations, I, the previous year, the PSD being assumed a Poisson PM; (ii) if the observations are numbers of eggs hatching under various levels of two environmental factors experimentally maintained, with N_{ij} eggs held at each combination of levels, the PSD being assumed a binomial PM.

Linear regression for a PSD is similarly developed, the MM being

$$\log \phi_i = \theta_i = \sum_j \beta_j x_{ij} \quad . \qquad (3.7)$$

This MM is substituted into PM (3.2) to form the corresponding natural linear SM.

As mentioned previously, such data as are suitable for analysis using these PSD SM's have traditionally been treated by some approximating normal theory SM such as that discussed in chapter II.

3. Binomial Statistical Models

Although the Poisson PM is probably the simplest PSD to consider, the binomial PM perhaps yields some more familiar results, especially related to contingency tables and logit analysis. Rasch (1961), Gart (1969, 1970a, 1970b), and Cox (1966, 1970) use some of the MM's outlined below. The familiar form of the binomial PM is

$$F(\underline{Y};\underline{p}) = \prod_{i=1}^{n} \binom{N_i}{Y_i} p_i^{Y_i}(1-p_i)^{N_i - Y_i} \quad .$$

In the form of a PSD, this becomes

$$F(\underline{Y};\underline{\phi}) = \prod_{i=1}^{n} \binom{N_i}{Y_i} \phi_i^{Y_i}(1+\phi_i)^{-N_i} \quad , \qquad (3.8)$$

with $a_i(Y_i) = \binom{N_i}{Y_i}$ and $g_i(\phi_i) = (1+\phi_i)^{N_i}$ so that $\theta_i = \log\left(\frac{p_i}{1-p_i}\right)$, the logit transformation. Natural linear MM's follow from this logit form.

An IxJ randomized block design with I = 2 corresponds to J 2x2 contingency tables. Interaction effects are not usually considered in analysis of such tables so that an appropriate natural linear MM is

$$\theta_{ij} = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \mu + \alpha_i + \beta_j \quad , \qquad (3.9)$$

with $\sum_j \beta_j = 0$ and $\alpha_1 = -\alpha_2$. Then, sufficient statistics are $y_{\cdot\cdot} = \sum_{i,j} y_{ij}$, $r_j = y_{\cdot\cdot} - y_{\cdot j} = y_{\cdot\cdot} - \sum_i y_{ij}$, and $s_1 = y_{\cdot\cdot} - y_{1\cdot} =$

$y_{..}-\sum_j y_{1j}$. Since there are J+1 < IJ parameters, $\prod_{i,j} a(y_{ij})$ must be modified for use in chapters VI and VII:

$$a'(y_{..},\underline{r},s_1) = \sum_{\underline{z}}\left[\prod_{j=1}^{J-1}\binom{N_{1j}}{y_{..}-r_j+z_j}\binom{N_{2j}}{-z_j}\cdot\binom{N_{1J}}{y_{..}-r_J-s_1-\sum z_j}\binom{N_{2J}}{s_1+\sum z_j}\right] . \tag{3.10}$$

The LF is

$$L(\mu,\underline{\alpha},\underline{\beta}) = \prod_{i,j}\binom{N_{ij}}{y_{ij}}\frac{e^{y_{ij}(\mu+\alpha_i+\beta_j)}}{(1+e^{\mu+\alpha_i+\beta_j})^{N_{ij}}} . \tag{3.11}$$

If this MM is found to be plausible when compared with one of the form in equation (3.8) (i.e. if an interaction effect is implausible), then the parameter of interest in equation (3.11) will usually be $\alpha_1 = -\alpha_2$. Chapters VI and VII contain further discussion about inferences for this parameter.

One-way ANOVA corresponds to analysis of an Ix2 contingency table for a binomial PM. Note that only one observation (one count of y_i successes in N_i trials) is necessary in each of the I blocks. The MM is

$$\theta_i = \log\left(\frac{p_i}{1-p_i}\right) = \mu + \alpha_i \quad , \tag{3.12}$$

where $\sum_i \alpha_i = 0$. Then, the sufficient statistics are $y_. = \sum_i y_i$ and $r_i = y_.-y_i$, with as many sufficient statistics (I) as response conditions i.e. no reduction in the number of parameters. The LF is

$$L(\mu,\underline{\alpha}) = \prod_i \binom{N_i}{y_i}\frac{e^{y_i(\mu+\alpha_i)}}{(1+e^{\mu+\alpha_i})^{N_i}} . \tag{3.13}$$

Interest usually centres on the vector of parameters, $\underline{\alpha}$, measuring differences among the blocks.

With I = 2, this SM applies to the 2x2 contingency table. The results developed in chapter VI for a conditional probability distribution involving only $\alpha_1 = -\alpha_2$ lead to the long run probability statements about the data in the form of Fisher's exact test for the 2x2 contingency table discussed in chapter VII.

Suppose that the n groups of individuals are observed under quantifiably different conditions, such that, of the N_i individuals in group i, y_i respond as "successes" under conditions $\underline{x}_i = (x_{i1},\ldots,x_{ip})$. Then, a natural linear MM is

$$\theta_i = \log\left(\frac{p_i}{1-p_i}\right) = \sum_{j=0}^{p}\beta_j x_{ij} \quad , \; x_{i0} = 1 \text{ (all i)}. \tag{3.14}$$

The sufficient statistic for β_j is $t_j = \sum_{i=1}^{n} y_i x_{ij}$. If $p+1 < n$, $\prod_i a(y_i)$ must be modified for use in the conditional distribution. For example, if p = 1,

$$a'(t_0,t_1) = \sum_{\underline{z}}\left[\prod_{i=1}^{n-2}\binom{N_i}{z_i}\left(\frac{(t_1-\sum_{i=1}^{n-2}z_i x_i)-x_{n-1}(t_0-\sum_{i=1}^{n-2}z_i)}{x_n-x_{n-1}}\right)^{N_{n-1}} \cdot \left(\frac{(t_1-\sum_{i=1}^{n-2}z_i x_i)-x_n(t_0-\sum_{i=1}^{n-2}z_i)}{x_{n-1}-x_n}\right)^{N_n}\right] . \tag{3.15}$$

The one-hit quantal response model used, for example, by Cox (1962) is a regression MM similar in form to equation (3.14) but with only the β_1 coefficient.

4. Natural Linear Statistical Models in the Exponential Family

To construct a natural linear SM in general, using the probability function of equation (1.3), the function of the mean, $\theta = B(\mu)$, is set equal to some linear function of new parameters which will incorporate information about the response conditions. MM's (1.4), (1.5), (2.3), (3.9), (3.12), and (3.13) are examples of such models. Obviously, a theoretical MM will often not be natural linear. In this case, although sufficient statistics for the parameters will not exist, all of the likelihood analyses described in these chapters, except for those of chapter VI, are still valid, and usually are no more difficult to perform. An example of a linear MM which is not natural is developed for the mean of the exponential PM in chapter V.

CHAPTER IV

COMPARISON OF PROBABILITY MODELS

1. Introduction

Often, more than one PM is theoretically feasible when considering SM's for an experiment. The problem of determination of the more plausible PM will be discussed in this chapter for the simple case where all observations are made under the same response conditions, i.e. where no MM need be introduced. To do this using likelihood inference, a base SM must be introduced with which all other SM's under consideration may be compared. The derivation to follow yields the multinomial PM.

Several approaches have been suggested in the literature to the problem of determining which of a number of possible SM's best describes a set of data. Cox (1961,1962) develops asymptotic Neyman-Pearson likelihood ratio tests and suggests an alternative approach involving a combination, either additive or multiplicative, of the density functions, with estimation of additional parameters. This approach is further developed by Atkinson (1970). The applicability of some of these long run probability statement methods will be further discussed in chapter VII. When prior probabilities, both for each SM and for the parameters within the models, are available, Lindley (1961, p.456) gives a posterior odds ratio of the two models using Bayes theorem. When applicable (i.e. when prior probabilities are available), this approach may be used with the methods developed below.

Since attention will be restricted, in this chapter, to the situation where all observations have been taken under the same conditions, they can be assumed to have the same (although unknown) PM. The more complex case, where the PM varies with the observations (i.e. by introduction of a MM), will be discussed in chapter V.

2.The Method of Comparison

The data have some observed frequency distribution, and, as pointed out in chapter I, are always discrete, although they may be theoretically continuous. Thus, the observation space is naturally divided into discrete intervals by the means of measurement and recording. A theoretical frequency distribution (PM) will predict what proportion of the observations should fall into each interval. The data are then the actual frequencies with which the observations fall into the various intervals and we wish to determine which theoretical PM best describes the observed frequencies. If the measuring device is too precise for the number of observations to be taken, i.e. if there will be too many intervals in the region where most observations will fall and too few observations per interval, a set of wider intervals may be specified as part of the design of the analysis before the data are collected or the procedure described in the next section may be used.

For observations y_{jk} $(k = 1,\ldots,n_j)$, the interval will be $y_{jk}\pm\frac{1}{2}\Delta y_j$, the theoretical proportion (probability) p_j, and the observed frequency n_j. If the observations are counts (i.e. the PM is discrete), $\Delta y_j = 1$ and the possible y_{jk}'s are integral. If the observations are measurements (i.e. the theoretical PM is continuous), they will have the form $y_{jk} = 12.5$ and $\Delta y_j = 0.1$. The best estimate of p_j is

$$\hat{p}_j = n_j/\sum_i n_i \quad . \tag{4.1}$$

There will be as many parameters as intervals (i.e. usually an infinite number) and many will be estimated as zero. Although the PM defined by equation (4.1) will make the observed data most probable, a theoretical model involving fewer parameters is usually desirable. Several may be suggested by the manner in which the data are to be generated and we will wish to determine which is preferable, i.e. which best approximates the $\hat{p}_j$'s. Thus, equation (4.1) provides the base PM which is required for making the likelihood inferences.

Usually, the more parameters to be estimated in the PM, the more potential for plausibility, since the flexibility of the PM in following the observed frequency distribution increases. Before the experiment, theoretical considerations or desire for simplicity should determine the desirable number of parameters allowable. Of course, if, after the experiment, no theoretical PM is shown to be plausible as compared to the base model (4.1), the data may point to some other PM whose plausibility must be affirmed by further data collection.

If the observations are independent, the probability of the observed data using the estimated proportions, $\hat{p}_j$, will be equal to the LF of the multinomial PM,

$$L_M(\hat{\underline{p}}) = \prod_j \hat{p}_j^{n_j} \quad . \tag{4.2}$$

If the observations are counts and the proposed PM is discrete, the theoretical proportion of observations falling into interval j, given the data, will be

$$\tilde{p}_j = F(Y_j;\hat{\underline{\phi}}) \quad , \tag{4.3}$$

where F is the probability function and $\underline{\phi}$ a vector of estimable parameters. For a proposed continuous PM this becomes

$$\tilde{p}_j = \int_{Y_j-\frac{1}{2}\Delta Y_j}^{Y_j+\frac{1}{2}\Delta Y_j} f(Y_j;\hat{\underline{\phi}})dY_j \doteq f(Y_j;\hat{\underline{\phi}})\Delta Y_j \quad , \tag{4.4}$$

where f is the probability density function. Specifically, for members of the exponential family, these yield equations (1.3). The probability of the observed data using the frequencies $\tilde{p}_j$, estimated for the theoretical PM, will be equal to the LF

$$\begin{aligned} L_F(\hat{\underline{\phi}}) = \prod_j \tilde{p}_j^{n_j} &= \prod_{j,k} F(y_{jk};\hat{\underline{\phi}}) && \text{(discrete)} \\ &\doteq \prod_{j,k} f(y_{jk};\hat{\underline{\phi}})\Delta y_j && \text{(continuous).} \end{aligned} \tag{4.5}$$

Then, the plausibility of this theoretical PM compared with the most plausible one is determined from the RL

$$R_F(\tilde{\underline{p}}) = \prod_j (\tilde{p}_j/\hat{p}_j)^{n_j} \quad . \tag{4.6}$$

Thus, although the measure of plausibility has been derived for comparison of theoretical PM's, it is absolute in that it can be used to determine how well the given PM fits the data. If enough observations have been made and if suitably sized intervals were previously chosen so that a number of adjacent intervals have reasonably large observed frequencies, asymptotic long run probability statements may be made, as discussed in chapter VII, by means of the log RL, $-2\log R_F$, which gives asymptotically, the Chi-squared goodness of fit test.

For a given set of data, the calculation of equation (4.6) for each theoretical PM will give a plausibility ranking of the models. If all PM's appear to be implausible, either insufficient data have been collected or some different PM should be considered. Of course, one or more PM's may be much more implausible than the others; these can be eliminated. Note that equation (4.5) is proportional to the usual LF and thus, for comparison of PM's F and G, the ratio R_F/R_G is equivalent to the usual RL. Of course, in this form, discrete and continuous PM's may be compared. The need for such a comparison most often arises in determining how well some continuous PM, such as the normal distribution, approximates to a theoretical discrete PM. This will be the prime use of this procedure in chapter V, where it is extended to SM's containing a MM.

3. Determination of Optimum Interval Width

As mentioned in the previous section, the interval widths naturally defined may need to be enlarged, especially for continuous data, if they provide too few observations per interval. Suppose that the data were generated by some reasonably smooth unknown PM, $f(\underline{y})$. If the natural intervals are too narrow, the observed frequency distribution will not appear to have a smooth shape. We wish to determine the minimum interval width which establishes this (unknown) shape of the frequency distribution. Since the multinomial distribution follows the shape of the observed frequency distribution exactly, when the intervals are too narrow, the multinomial probability of the data will be too great for the data to have arisen from a smooth PM, $f(\underline{y})$. Let $(\tilde{n}_j)$ represent the frequencies predicted by $\tilde{f}(\underline{y})$, the unknown MLE of the PM. If the natural interval widths are too narrow, the n_j's and $\tilde{n}_j$'s will be very much different. From the inequality

$$\left(\frac{\tilde{n}_1}{n_1}\right)^{n_1}\left(\frac{\tilde{n}_2}{n_2}\right)^{n_2} \leq \left(\frac{\tilde{n}_1+\tilde{n}_2}{n_1+n_2}\right)^{n_1+n_2} , \qquad (4.7)$$

as intervals are combined, the multinomial probability of the data will increase more slowly than the probability given by the unknown PM, $\tilde{f}(\underline{y})$, until they converge. When the intervals are sufficiently wide so that the observed frequency distribution is smooth enough, the relationship

$$L_M = \prod_j \hat{p}_j^{n_j} \doteq \tilde{f}(\underline{y}) \prod_j \Delta y_j \qquad (4.8)$$

will hold (using equations (4.4) and (4.5) which are numerically good enough for graphical methods). Further increases in interval widths will not change this relationship until the number of intervals becomes so small that the shape of $\tilde{f}(\underline{y})$ is distorted. Thus, $\log L_M = \sum_j \log \Delta y_j$ + a constant in this range of interval widths; log L_M may be plotted against $\sum_j \log \Delta y_j$ for various interval widths. The region of this curve in which the points begin to follow a straight line inclined at 45^o indicates the required width.

In producing the curve, a simple procedure is to plot points for various constant interval widths e.g. 1, 2, 5, 10, 20 units,etc. Then, the 45^o line is positioned on the straight region of this curve. The point for a multinomial LF from the data with tail area intervals combined (e.g. 10 unit intervals but with small tail frequencies combined in wider intervals) will lie on the same curve but displaced upwards to the right from the corresponding point with constant interval width,i.e. both log L_M and $\sum_j \log \Delta y_j$ will increase. If the point does not lie on the curve, the shape of the PM has been distorted by the unequal widths and the grouping should be discarded.

Extrapolation of the 45^o line to $\sum_j \log \Delta y_j = 0$ yields an approximate value of the unknown log $\tilde{f}(y)$ which may be compared with the calculated values for the various theoretical PM's.

This procedure was applied to the three sets of data of Table 4.8 (see example 4.6 below) for intervals of constant widths 1, 2, 3, 4, 5, 10, 20, 30, and 40 minutes. The results plotted in Figure 4.1 show the optimum intervals to lie between 5 and 10 minutes.

After optimum intervals have been determined, the RL's of various theoretical PM's may be calculated for goodness of fit and comparison. The original, ungrouped data should be used for estimating parameter values in these PM's to avoid the loss of accuracy specified by Sheppard's correction.

Consider now the expected log multinomial likelihood for the arbitrary "smooth" density function, $f(\underline{y})$, expectations being taken only over the m intervals with non-zero observed frequency:

$$E(\log L_M) = E(\log n!-n \log n-\sum_{j=1}^{m} \log n_j!+\sum_{j=1}^{m} n_j \log n_j)$$
$$\doteq \log n!-n \log n-\frac{m}{2} \log n\Delta-\frac{m}{4n} \cdot$$
$$+\sum_{j=1}^{m}\left[n\Delta f(y_j)-\frac{\log f(y_j)}{2}+\frac{1}{4n\Delta f(y_j)}\right] ,$$

using Stirling's approximation with $p_j = \Delta f(y_j)$ and $\Delta y_j = \Delta$. As Δ is made smaller from very wide intervals to the minimum (natural) width, the term $-\frac{m}{2} \log(n\Delta)$ stops decreasing as m stops increasing, while $\sum f(y_j)$, $\sum \log f(y_j)$, and $\sum \frac{1}{f(y_j)}$ approach constant values. Then, the expected log likelihood no longer follows the 45° line. Thus, the optimum interval width occurs when further decrease in width does not increase the number of intervals proportionally. If the number of observations, n, is increased, this optimum width will be proportionally smaller. Note that these approximate results are not altered by using more terms of the series expansions of log x!, p_j, or the expected value.

4. Examples

4.1 Fisher (1958, p. 299) discusses linkage in the progeny of self-fertilized heterozygote maize. The two factors of interest are starchy versus sugary and green versus white base leaf, where starchy and green are dominant. Thus, observations (counts) will fall into the four possible intervals made up of the various combinations of the two factors. From Fisher's data reproduced in Table 4.1, the multinomial LF is

$$L_M = \left(\frac{1997}{3839}\right)^{1997}\left(\frac{906}{3839}\right)^{906}\left(\frac{904}{3839}\right)^{904}\left(\frac{32}{3839}\right)^{32} \quad ,$$

using equations (4.1) and (4.2).

One possible theoretical PM occurs with no linkage, yielding a SSM. The expected ratios of the various outcomes are 9:3:3:1 yielding $\tilde{p}_1 = \frac{9}{16}$, $\tilde{p}_2 = \tilde{p}_3 = \frac{3}{16}$, and $\tilde{p}_4 = \frac{1}{16}$, with no unknown parameter to be estimated. The second theoretical PM of interest occurs with linkage where the expected ratios are $2+\phi:1-\phi:1-\phi:\phi$ and $\tilde{p}_1 = \frac{2+\hat{\phi}}{4}$, $\tilde{p}_2 = \tilde{p}_3 = \frac{1-\hat{\phi}}{4}$, $\tilde{p}_4 = \frac{\hat{\phi}}{4}$. Then, the two LF's for Fisher's data are

$$L_{F_1} = \left(\frac{9}{16}\right)^{1997}\left(\frac{3}{16}\right)^{1810}\left(\frac{1}{16}\right)^{32}$$

and

$$L_{F_2}(\hat{\phi}) = (2+\hat{\phi})^{1997}(1-\hat{\phi})^{1810}\hat{\phi}^{32}\,4^{-3839} \quad ,$$

respectively. The RL of no linkage is $\log R_{F_1} = -198.344$ and of linkage is $\log R_{F_2} = -1.023$ showing that the PM without linkage is very implausible in comparison to either of the other models, and that the linkage model is a very good representation of the multinomial PM. This agrees with the conclusions from Fisher's Chi-squared tests.

4.2 Cox and Brandwood (1959) provide a further example of the comparison of multinomial PM's, in this case SSM's. Between writing the Republic (R) and the Laws (L), Plato wrote a number of short dialogues, the order of five of which is uncertain: Critias (C), Philebus (F), Politicus (P), Sophist (S), and Timaeus (T). The distribution of long and short syllables at the ends of sentences is used in an attempt to order the seven works. Consideration of the last five syllables provides 32 classes (intervals). For each work, two SSM's are available: that the distribution of syllables is the same as for R and as for L. Cox and Brandwood give the proportions in each interval reproduced in Table 4.2.

The observed proportions yield the base PM for all comparisons for each work; the multinomial LF's for these PM's are given in Table 4.3, along with the log RL's of the two SSM's for each work. These RL's show that neither SSM fits well for any of the five works compared. This may be expected since we are only interested in which is better and not if either is good. The ratio of the two RL's for each work provides a comparison of the fit for the two SSM's; these ratios give the ordering R, T, S, C, P, F, L of the works.

The ratio of RL's divided by the number of sentence endings for the work yields the statistic $\bar{s}$ which Cox and Brandwood use. They justify their result by the introduction of a power parameter, λ, to which the probabilities of a sentence falling into one of the intervals is raised. Atkinson (1970) generalizes this by introducing two power parameters,

λ_1 and λ_2, one for the probabilities from each of the two SSM's. These two approaches yield two new SM's, with one and two estimable parameters, respectively. The RL's for these two models are given for each work in Table 4.3. These models both give improved fits for all five works, with the Atkinson two-parameter model consistently better. But the small size of the RL's (implausibility) for this two parameter model casts doubt on its applicability in the manner suggested by Atkinson (-2 log R asymptotically has a χ^2 distribution with 2 degrees of freedom, giving very significant lack of fit; see chapter VII). This poor fit makes the use of the λ's for ranking suspect and thus also the use of $\bar{s}$ instead of the RL.

4.3 Cox (1962) provides a sample of 30 observations (Table 4.4) generated from a Poisson PM of mean 0.8. The base PM for these data gives a multinomial LF of log L_M = -35.095. In addition to the Poisson and geometric PM's considered by Cox, the normal model with two estimable parameters is also used here. The estimated theoretical probabilities for these three models are given by

$$\tilde{p}_P = e^{-\hat{\phi}}\hat{\phi}^y/y! \quad ,$$

$$\tilde{p}_G = \hat{\phi}^y/(1+\hat{\phi})^{1+y} \quad ,$$

and

$$\tilde{p}_N = (2\pi\hat{\phi}_2)^{-\frac{1}{2}}\exp\left[\frac{-(y-\hat{\phi}_1)^2}{2\hat{\phi}_2}\right] \quad ,$$

respectively. Then, the log RL's are log R_P =-0.609, log R_G = -3.548, and log R_N = -2.369. Thus, very little is lost by representing the frequencies observed by the Poisson PM (i.e.

it is very plausible) while representation by either of the others is poor. (Note the computational error in the maximized likelihood ratio of Cox which should read $e^3 \doteq 20$). Then, the plausibility ranking of the PM's is multinomial, Poisson, normal, and geometric, with probabilities of the observations in the ratios 1:1.8:9.3:35.

Equation (1.2) has been used to calculate $\tilde{p}_N$. The exact calculation using equation (1.3) yields virtually the same result.

4.4 "Student" (1907) gives data, reproduced in Table 4.5, for the distribution of yeast cells in 400 equal-sized squares of a haemacytometer. For these data, $\log L_N = -444.527$. For the same PM's as in the preceding example, the log RL's are $\log R_P = -4.835$, $\log R_G = -9.139$, and $\log R_N = -86.412$. None of these models fits the data well; the normal PM is much worse than the others because of the large theoretical probability of observing "negative counts".

Bardwell and Crow (1964) fit a two-parameter hyper-Poisson PM to these data and compare it with the fit of a Neyman type A PM; both are improvements over the Poisson: $\log R_{HP} = -2.301$ and $\log R_{NA} = -1.789$, respectively. They provide a description of this hyper-Poisson PM which involves a confluent hypergeometric function, and outline methods for estimating the two parameters.

Asymptotic Chi-squared goodness of fit tests, as in chapter VII, applied to the various models give results agreeing with those above.

4.5 Monfort (1964) fits the log normal PM with estimated theoretical probabilities

$$\tilde{p}_{LN} = (2\pi\hat{\phi}_2)^{-\frac{1}{2}}\exp\left[-\frac{(\log y - \hat{\phi}_1)^2}{2\hat{\phi}_2}\right]\Delta y/y$$

and the gamma PM with

$$\tilde{p}_G = \left[(\hat{\phi}_2/\hat{\phi}_1)^{\hat{\phi}_2}/\Gamma(\hat{\phi}_2)\right]y^{\hat{\phi}_2-1}\exp\left[-\hat{\phi}_2 y/\hat{\phi}_1\right]\Delta y$$

to the observed frequency distributions of the fibre diameters of eight lots of combed slivers of reference wools (wool tops) given in Table 4.6. Jackson (1969) provides further analysis of these data using asymptotic Neyman-Pearson likelihood ratio tests of Cox (1961, 1962). The results for a multinomial likelihood analysis are given in Table 4.7, where R_{LN} and R_G are the RL's for the log normal and gamma PM's, respectively. The likelihood ratios for comparison of the two models are calculated from $\log R_{LN} - \log R_G = \log R_{LNG}$. Using asymptotic theory, Jackson apparently has calculated approximations to these ratios ($\log \tilde{R}_{LNG}$ in Table 4.7). This asymptotic theory does not produce a LF which very well represents the exact LF's (see chapter VII, section 4). Hence, the long run probability statements will not be very accurate.

Lots A and C point to the log normal PM and the remaining lots to the gamma. But for lot A, both models fit almost equally well, and for lots G and H, neither fits well. If we had assumed that one of the two models was correct, the poor fit of G and H might imply that insufficient data had been collected.

The eight lots of data appear to be grouped into two sets, ABCD and EFGH, of which the first set have 600 observations per lot and the second have 450. The observed frequency distributions of the first set are relatively narrow and peaked while those of the second set are much more dispersed and ill-defined, having two or more modes each. This is reflected in the poorness of fit of all of the lots of the second set. Jackson mentions that further investigation reveals that a "mixture" of two gamma PM's fits lot H quite well. This may also apply to the remaining three lots of this set.

4.6 Bliss (1967, pp. 106 and 122) provides the data reproduced in Table 4.8 from the experiments described by Campbell (1927) to determine the survival times in minutes of fourth instar silkworm larvae when fed a lethal dose of 0.10 mg. of sodium arsenate per gram of body weight. Data are provided from three experiments. These data provide an example where most counts are zeroes with many ones; this is reflected in the small size of the RL's given in Table 4.9 for the ungrouped data. For all three sets of data, the log normal, the gamma (see the previous example) and the exponential PM with

$$\tilde{p}_E = \phi e^{-\phi y}\Delta y$$

are fitted. From Table 4.9 for the grouped data (using 10 minute intervals from section 3 above), the log normal and gamma PM's are equally plausible for all sets of data, while the exponential is always much less plausible and can be eliminated.

4.7 In studies of the carcinogenic effect of ultra-violet radiation, Blum (1959) exposed male mice of a given strain to dosages of radiation of intensity 3.4×10^4 ergs/cm^2/sec. for five days per week. Five groups of mice were used, exposed to different doses per day, i.e. for different lengths of time, and the development times in days of an ear tumor determined. In order to provide an example with Δy_i varying, the development times are grouped into intervals of equal log days in Table 4.10, as provided by Bliss (1967, p.274). A comparison of the log normal and gamma PM's in Table 4.11 shows that both are equally plausible for all doses. The size of the RL seems to decrease with the size of the multinomial likelihood, and hence with increase in sample size, indicating that increased sample size provides no stronger evidence that either model is plausible for this type of data.

5. Discussion

Atkinson (1970) develops the comparison of PM's suggested by Cox (1961,1962) using a product of probability functions, each raised to a power λ_i. Essentially, a new PM, an amalgam of those hypothetically possible, is introduced. If this new PM is also hypothetically reasonable, there will be little difficulty of interpretation. If not, it may have meaning only in the null cases where it reduces to one of the original PM's. In any case, doubt will be thrown on the utility of using the λ's for ranking the models if the composite PM does not, itself, fit the data well, as in example 4.2 above.

Interpretation of a composite PM may be further illustrated by an example. If the two PM's to be compared are the exponential and the normal truncated at zero with known variance set equal to unity (before truncation), with $\lambda_1 = 1-\lambda_2 = \lambda$, then the derived PM is normal truncated at zero with unknown variance, $\frac{1}{1-\lambda} = \sigma^2$. The graph of the LF of λ is equivalent to one of $(\sigma^2-1)/\sigma^2$. Thus, only the variance, and no higher moments, is used in this comparison of PM's.

Difficulty may arise in using the MLE's of λ_i's for ranking PM's without the use of some interval about the maximum. In the comparison of two PM's, such as those above, the MLE of λ may be greater than one half, say 0.7, pointing to the first model, while the LF for this PM ($\lambda = 1$) is smaller than that for the second ($\lambda = 0$), making the second more plausible by the likelihood criterion. In other words, the λ of Atkinson's exponential combination may point to a PM which makes the observed data less probable than another to which it is compared. Of course, Atkinson's combined PM will be more plausible at $\tilde{\lambda}$ than either of the individual models, and should be considered if interpretable.

If specific alternatives to the given PM's are to be considered, and these coincide with the family of PM's defined by Atkinson's exponential product, this approach will yield useful results. If the plausibility of the given PM's is to be considered as opposed to any possible alternatives, comparison with the multinomial LF will yield the desired inference.

In comparison of PM's, attention must be paid to the size of each R_F, since a difference between R_F's is of less importance if the individual R_F's are so small as to indicate that none of the models is very plausible. This is especially important if the applicability of at least one of the PM's is not fairly certain. If this is reasonably certain, lack of fit may indicate insufficient data. For goodness of fit, not only is the difference in number of parameters estimated important, but also the number of observations made. An indication of this is given by the size of the RL in comparison to the multinomial likelihood, i.e. by the change in probability of the data with introduction of a theoretical PM as compared to the maximum probability of the data. If the RL is relatively large as in the first six examples, sufficient observations are usually available, and consideration of the size of RL in conjunction with parameter numbers (degrees of freedom) will determine goodness of fit. In the last example (4.7), the RL was not relatively large, and either insufficient data or poor

PM's may have caused lack of plausibility. But comparison among dosages seems to indicate that fit does not improve with increased sample size.

In the next chapter, examples involving MM's will be considered. For these, we must usually assume that at least one of the PM's is acceptable, since insufficient observations are usually made under each response condition to check goodness of fit of the various PM's.

CHAPTER V

ANALYSIS OF DATA INVOLVING A MATHEMATICAL MODEL

1. Introduction

In this chapter, the methods of the preceding four chapters will be applied to statistical models (as defined in chapter I, section 3). Although the methods are applicable to SM's in general, attention will be restricted to PM's from the exponential family of equation (1.3) and to MM's describing variation in the mean of the PM. Most MM's will be natural linear or nonlinear extensions of these. The exceptions to this are some MM's for the exponential PM in the examples of section 6.

Four PM's will be considered as representative, since they are probably the most commonly encountered and since they usually require no extremely complicated numerical estimation procedures (Newton's iterative method sufficing). These are the normal, exponential, binomial, and Poisson PM's. The continuous models are generalized to include such offshoots as the log normal and the Weibull PM's. Analytically, if not numerically, these methods will apply to other PM's, as well as to other MM's, than those considered.

2. Probability Models

In chapter II, a generalized form of the normal PM was considered with generalizations of a natural linear MM in the case of response surface data, when no alternative PM's were available. Cases where various other PM's are theoretically plausible will now be considered. The generalized normal PM will be used, either as one possibility or as an approximating model.

The LF's for the PM's to be considered are:

(i) normal distribution, $N(\mu,\sigma^2)$

$$L_N(\mu,\sigma^2,\underline{\Upsilon}) = \exp\left[\frac{g(y,\underline{\Upsilon})\mu}{\sigma^2}-\frac{g^2(y,\underline{\Upsilon})}{2\sigma^2}-\frac{\mu^2}{2\sigma^2} -\log\sqrt{2\pi\sigma^2}\right]\Delta g(y,\underline{\Upsilon}), \quad (5.1)$$

(ii) exponential distribution

$$L_E(\mu,\underline{\Upsilon}) = \exp\left[-g(y,\underline{\Upsilon})/\mu-\log \mu\right]\Delta g(y,\underline{\Upsilon}) \quad , \qquad (5.2)$$

(iii) binomial distribution, $B(\mu/n,n)$

$$L_B(\mu) = \exp\left[y \log\left(\frac{\mu}{n-\mu}\right)+n \log\left(1-\frac{\mu}{n}\right)+\log\binom{n}{y}\right] , \quad (np = \mu), (5.3)$$

(iv) Poisson distribution

$$L_P(\mu) = \exp(y \log \mu - \mu - \log y!) \quad . \qquad (5.4)$$

The unit of measurement is Δy, where the Δ operator acts as a differential, e.g. $\Delta g(y) = \Delta \log y = y^{-1}\Delta y$. If the observations are counts, then $\Delta y = 1$. The continuous PM's (5.1) and (5.2) are generalized by allowing a transformation of the response, $g(Y,\underline{\Upsilon})$, which may contain parameters, $\underline{\Upsilon}$, to be estimated, such as those used in chapter II for the normal PM. For example, equation (5.2) becomes the LF for the Weibull PM if $g(y,\Upsilon) = y^{\Upsilon}$, $\Delta g(y,\Upsilon) = \Upsilon y^{\Upsilon-1}\Delta y$.

For the normal PM (5.1), the parameter μ may still be considered natural (if there is no estimable parameter, $\underline{\Upsilon}$), since jointly sufficient statistics exist for μ and σ^2. Thus, natural linear MM's may be constructed involving μ.

3. Mathematical Models

Mathematical models were briefly described in chapter I, section 3 where the two simplest cases for the mean were given in equations (1.4) and (1.5), and more generally in chapter III. From such MM's, the expected value of the response may be calculated under a given response condition; for example, $E[g(Y,\underline{Y})] = B^{-1}(\theta_1)$ using MM (1.4) with the PM from LF (5.1) or (5.2).

In the examples to follow, the MM's used will be those for ANOVA and regression. If the response conditions imply a two-way factorial experiment, the reduced (from equation (1.4)) natural linear MM might be either the generalization of equation (3.6),

$$B(\mu_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij} \qquad (5.5)$$

or the generalization of equation (3.9),

$$B(\mu_{ij}) = \mu + \alpha_i + \beta_j \quad , \qquad (5.6)$$

where i and j specify levels of the two factors and $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$. For a one-way design, the MM might be the generalization of equation (3.12),

$$B(\mu_i) = \mu + \alpha_i \quad , \qquad (5.7)$$

with $\sum_i \alpha_i = 0$ and i specifying levels. If the conditions determine a two-factor response surface or a regression problem, the reduced natural linear MM is

$$B(\mu_i) = l(\underset{\sim}{\beta}, \underline{x}_i) \quad , \qquad (5.8)$$

where $l(\underset{\sim}{\beta}, \underline{x}_i)$ is some linear function of the parameter vector $\underset{\sim}{\beta}$ such as

$$l(\underline{\beta},\underline{x}_i) = \beta_o+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i1}^2+\beta_4 x_{i2}^2+\beta_5 x_{i1}x_{i2} \qquad (5.9)$$

and (x_{i1},x_{i2}) are the factor levels (transformed to orthogonal poynomials as in chapter II) for response y_{ik}. A nonlinear MM analogous to equation (5.9) is given in equation (2.3).

If parameters related to other functions than to those of the mean of the response are present in the PM, as in equation (5.1), variation of these with the response conditions may also be desirable. For example, if ten observations y_{ijk} $(k = 1,...,10)$ are taken at each level of an IxJ factorial experiment and the LF is equation (5.1) with $g(y,\gamma) = y^{\gamma}$, the MM's might be equation (5.5) with $G(\sigma_{ij}^2) = \sigma_i^2$ and $F(\gamma_{ij}) = \gamma$. This type of situation may occur when the normal PM is used to represent a response distribution for which the variance changes with the mean (e.g. binomial, Poisson). Problems related to this will be discussed in section 9 for the examples of this chapter. Otherwise, only parameters related to the mean will be considered to vary in the examples.

4. Inferences

The basic procedures for making inferences about SM's were developed in chapter II for the specific case of a nonlinear normal response model. Here, these procedures will be described more generally, with inclusion of the case where more than one PM is feasible. The steps outlined below provide one way of proceeding; of course, these will often have to be modified for specific problems.

Step I. Parameter Estimation. In order to carry out the following steps, MLE's of all parameters of each SM must be calculated.

Step II. Goodness of Fit of the PM. As mentioned at the end of the last chapter, for a SM which varies with the response conditions, insufficient observations are usually made to determine the plausibility of the various PM's. To determine goodness of fit, sufficient observations must be made at a response condition to provide a useful frequency distribution. This distribution will vary with the reponse conditions. If enough observation are available, plausibility is determined as in chapter IV using as the base SM the combined multinomial PM fitted individually for each response condition.

An alternative in designing such experiments may be to take a much larger number of observations at one or more response conditions (not yielding an extreme response) and to include these results to determine goodness of fit.

Step III. Comparison of PM's. If several PM's may possibly describe the random fluctuation of the response, these may be compared by fitting one set of parameters under each response condition for each LF. If only parameters of the mean are present, this is equivalent to using equation (1.4) with each LF. The sizes of the LF's using the corresponding MLE's of all parameters are then compared as in chapter IV. Thus, the PM is fitted to the data from each response condition individually and the combined distribution for all the data is compared with the corresponding one for another PM. This provides a comparison of PM's for the data without involving any of the theoretical MM's. Of course, this SM (e.g. involving equation (1.4)) will be equivalent to a SM including some hypothesized MM such as equation (5.5) if both have the same number of parameters (i.e. if only a transformation of the parameter space is involved and not a reduction in number).

If parameters other than functions of μ (e.g. σ^2 of equation (5.1)) are present, the above procedure requires that more than one observation be made at the different response conditions. In the numerical examples to follow, a discrepancy from the procedure just described is allowed in that the normal PM is assumed to have constant variance.

Comparison of PM's assumes that either some of the models have been shown to be plausible (in step II) or that at least one of them is theoretically justifiable. If one PM is more justifiable theoretically than another, a stronger plausibility may be required before adopting the second than if this were not the case.

Step IV. Comparison of MM's. Usually, it is first useful to see if a MM is necessary at all. Comparison of the unstructured SM from the previous step with a common PM fitted to all of the data (i.e. no parameters varying with response conditions, e.g. equation (1.5) for the mean) will show if the same MM under all response conditions is plausible. This corresponds to the F test for overall treatment effect in ANOVA.

If an unvarying MM is not plausible, the various postulated MM's may be compared for any PM selected in step III. This corresponds to the F test for lack of fit in regression analysis. Even if an unvarying MM is plausible, further analysis may yield some likely effects under the postulated MM's.

If more than one PM was still plausible after step III, the inclusion of a reduced MM may indicate that some PM's can be eliminated at this stage. This also applies if parameters other than μ are present with only one observation per response condition, making step III impossible, in which case assumptions may be made about higher order interactions.

Step V. Inferences about Parameter Values. After one or more MM's has been found plausible, inferences are made as to whether all of the parameters of each MM are necessary, e.g. some $\beta_i = 0$ in equation (5.9) or some $\alpha_i = 1$ in equation (2.3). These correspond to F tests for individual treatment effects in ANOVA tables. They should then be extended to see which values of the necessary parameters are plausible by plotting graphs of RL's (see below). If, within the MM, some parameters are of interest while others may be considered as nuisance parameters, inferences may be possible about the parameters of interest in the absence of knowledge about the nuisance parameters. If sufficient statistics are available for the nuisance parameters, this may be done using a conditional LF (see chapter VI).

As in chapter II, parameters are estimated by maximum likelihood methods. This procedure will involve an iterative process when the likelihood equations are nonlinear. Some form of Newton's method is also used here for the examples.

Since inferences about the parameters are to be made by observing the effect on the size of the LF of variation in the parameters, multidimensional models can yield very complex situations; see Sprott and Kalbfleisch (1969). Thus, as in chapter II, the simplifying and approximating technique of maximizing the LF over all but one or two parameters at a time is used. Within a given SM, this maximized LF is then compared with the LF maximized over all parameters, by means of the RL function of equation (1.7):

$$R(\phi_i) = \sup_{\forall \phi_j \text{ except } \phi_i} L(\underline{\phi}) / \sup_{\underline{\phi}} L(\underline{\phi}) \quad . \qquad (5.10)$$

Throughout the examples, log L and log R are listed instead of the LF and the RL. Steps I and II are omitted in all discussion of examples.

5. Normal Theory Models

Since the use of some form of normal theory analysis is the standard practice for attacking the types of problems under discussion, most of the examples include this analysis, usually in various forms depending upon appropriate transformations of the original response. These transformations are incorporated in the LF (5.1) through $g(y,\underline{Y})$. Some common transformations are listed:

(A) linear normal ($\gamma = 1$ in (B)) $g(y) = y$

(B) power transformed normal $g(y,\gamma) = y^{\gamma}$

(C) log normal ($\gamma = 0$ in (B)) $g(y) = \log y$

(D) square root normal ($\gamma = \frac{1}{2}$ in (B)) $g(y) = \sqrt{y}$.

The following apply only to binomial data with y successes in n trials:

(E) logit normal ($\gamma = 0$ in (G)) $g(y) = \log[y/(n-y)]$

(F) odds normal ($\gamma = 1$ in (G)) $g(y) = y/(n-y)$

(G) transformed odds normal $g(y,\gamma) = [y/(n-y)]^{\gamma}$

(H) per cent normal ($\gamma = 1$ in (I)) $g(y) = 100y/n$

(I) transformed per cent normal $g(y,\gamma) = (100y/n)^{\gamma}$

(J) arcsine normal $g(y) = \sin^{-1}\sqrt{y/n}$.

For the normal PM,

$$B(\mu) = \mu = E[g(Y,\gamma)] \tag{5.11}$$

so that, for example, from equation (5.5)

$$E[g(Y_{ij},\gamma)] = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad .$$

In the examples, neither σ^2 nor any parameter in $g(y,\underline{\gamma})$ is assumed to vary with the response conditions. The validity of this assumption about the variance is examined in section 9.

Since the normal PM contains the parameter σ^2 and, possibly, parameters in $g(y,\underline{\gamma})$, comparison of PM's or use of a MM such as (1.4), (5.5), or (5.7) is not possible with only one observation for each condition. But, these MM's can be used for all of the other PM's considered: (5.2) (unless $g(y,\underline{\gamma})$ contains unknown parameters), (5.3), and (5.4). This means, for example, that, with one of these PM's and a two-way factorial experiment with one observation per cell, the interaction effect can be examined using equation (5.5).

6. Lifetime Data

When data are observations of (life, survival, reaction, completion) times, their frequencies may often be described by one or more of the exponential, Weibull, log normal, and gamma PM's, with an appropriate MM. When ANCVA is involved, the standard normal theory analysis is usually used, perhaps with a logarithmic or power transformation of the response. In the examples to follow, the three normal PM's, (A), (B), and (C), of section 5 are used as well as two exponential PM's (from equation (5.2) with $g(y,\underline{Y}) = y$):

(K) natural linear exponential in which

$$B(\mu) = 1/\mu = 1/E(Y) \tag{5.12}$$

and (L) "expected value" exponential in which

$$B'(\mu) = 1/B(\mu) = \mu = E(Y) \tag{5.13}$$

and $B'(\mu)$ replaces $B(\mu)$ in the MM such as equation (5.5). The natural linear exponential SM equates the reciprocal of the mean with the MM whereas the "expected value" SM equates the mean with the MM.

Example 6.1 Box and Cox (1964) provide two analyses of a 3x4 factorial experiment on the survival times of animals (reproduced in Table 5.1), the factors being three poisons and four treatments, with each combination used on four animals.

Box and Cox use the interaction (5.5) and no interaction (5.6) MM's with the linear (A) and power transformed (B) ($\hat{\gamma}$ = -0.77). normal PM's. Only the interaction MM is here considered. In addition, the two exponential SM's (K) and (L) with this same MM are used here.

Step III. Assuming that one of the SM's is adequate, the fit of the four PM's may be compared using MM (1.4). Since MM (5.5) has as many parameters as response conditions, this model has as many parameters as (1.4) and the two are equivalent for comparison of PM's. In order of plausibility, the PM's are transformed normal, linear normal, and exponential (Table 5.2). Both of the normal PM's are much more plausible than the exponential PM's. Unless theoretical considerations strongly dictate otherwise, the exponential PM may be eliminated.

Step IV. A uniform SM for all response conditions is very implausible for either normal PM, but much less implausible for the exponential PM's. Only one theoretical MM is being entertained so that no further comparison is necessary. Note, from Table 5.2, that the no interaction MM is plausible for all PM's except the linear normal. Comparison of MM's (5.5) and (5.6) is equivalent to determining the plausibility of eliminating the interaction effect, γ_{ij}, from MM (5.5) in step V.

Step V. All of the SM's give the same relative analysis of effects in Table 5.2, with differences in poisons more likely than in treatments, and interaction relatively less so. But the likelihood of the various effects (i.e. plausibility of no effect or of eliminating the relevant parameters) is much less (plausibility higher) for the exponential than for the normal SM's.

For illustrative purposes, regression MM (5.9) was fitted for PM's (A), (B), and (K) to analyze linear and quadratic effects, although the design provides no reason for doing this. This MM is only plausible for the exponential PM (see regression effects in Table 5.2). Only the linear effect of poisons is plausible in the three models, although the other non-interaction effects border on plausibility for model (B). This part of the analysis is an extension of the methods of chapter II.

Example 6.2 In analyzing experiments with analgesic drugs, Ipsen (1949) gives two control readings, listed in Table 5.3, spaced 20 to 30 minutes apart, of the tail flip reaction times of ten white rats on nine different days when a strong light is focused on their tails. The same analyses are performed as in the preceding example.

Step III. The reaction times were supposed to follow the log normal PM (C), to which model (B) with $\hat{\gamma} = 0.10$ and $R(\gamma=0.0) = 0.95$ points. PM's (B) and (C) provide virtually no improvement over the linear normal (A) (see Table 5.4), but all three are more plausible than the exponential models.

Since the fits of PM's (B) and (C) are virtually identical, only the analysis for (C) is given in Table 5.4.

Step IV. With the exponential PM's, the data might plausibly have the same distribution for all response conditions, i.e. MM (1.5) (this could be related to the implausibility of these PM's), but this is very implausible for the normal PM's.

Step V. Again, all of the SM's give the same relative analysis of effects. Interaction between days and rats is most likely and differences between rats least. But the RL's for the exponential SM's show that all of the parameters of MM (5.5) except μ could plausibly be eliminated, reinforcing the conclusions of step IV.

The consistent difference in size of the RL's for the normal and exponential PM's will be further discussed in section 11.

Example 6.3 Bliss (1967, p.327) gives the times required for ten rats to run through a maze on various trials, and looks at additivity on five of the trials (Table 5.5). These data are analyzed as in the previous examples, but using no interaction MM (5.6) for the normal PM's, since there is only one trial per cell in the 5x10 factorial experiment. Interaction MM (5.5) is used with the exponential PM's.

Step III. With only one observation per response condition, comparison of the normal PM with any other is not possible.

Step IV. As in the previous example, model (B) with $\hat{\gamma}$ = 0.04 points to the log normal PM (C) suggested by Bliss. Both are marked improvements on the linear normal PM (A), and both give the same analysis, hence only that for (C) is given in Table 5.6.

With MM (5.6), some of the expected times are negative; these yield MLE's of zero if exponential LF (5.2) is used (due to the interaction effects shown in step V below), making use of the exponential PM's with this MM meaningless.

When MM (5.5) is used with the exponential PM's (K) and (L), the fits are much better than with the linear normal PM but not as good as with the log normal, both using MM (5.6).

A uniform SM for all response conditions is implausible for all of the PM's.

Step V. All analyses show that differences between trials are more likely than between rats, although neither can be plausibly eliminated. The exponential SM's involving MM (5.5) show that an interaction effect between rats and trials is very plausible.

7. Poisson Count Data

When data are observations which appear to follow a Poisson distribution, and which involve a linear MM, a linear normal SM is often used with either a square root or a logarithmic transformation of the response. In the following examples, normal PM's (A), (C), and (D) are used. The Poisson PM

(M) of equation (5.3) is also used with a natural linear MM for which

$$B(\mu) = \log \mu = \log E(Y) \quad . \qquad (5.14)$$

Again, the validity of the assumption of constancy of variance for varying reponse conditions for the normal PM's will be discussed in section 9.

For a Poisson PM, the parameters of the no interaction MM (5.6) may be estimated exactly without using an iterative method.

Example 7.1 Bartlett (1936b) provides an example of a 6x6 factorial experiment testing the effectiveness of four toxic emulsions in controlling leatherjackets. 36 plots of one sq. yd. each were divided into six blocks and the emulsions applied to four plots in each block, the remaining two plots of the block being controls. Some days after application, two sample counts were made of the number of leatherjackets remaining on one sq. ft. each, as tabulated in Table 5.7. Since the counts may be considered to have a Poisson PM, a square root transformation, model (D), was suggested before performing normal theory ANOVA. This is compared below with linear normal (A), log normal (C), and Poisson (M) PM's using interaction MM (5.5).

Step III. Comparison of the four PM's in Table 5.8 shows that the suggested square root normal model and the Poisson provide the same fit, while the linear normal model is much worse and the log normal model somewhat better. In this case, because of theoretical considerations, use of the Poisson PM may be preferable, although the log normal PM is better for this set of data. This comparison of models is further discussed in section 9, where residuals are considered.

Step IV. With no PM is elimination of the variation due to response conditions plausible, as shown in Table 5.8. Only MM (5.5) is considered for all of the PM's.

Step V. With all four PM's, the order of likelihood of effects is the same, with differences between treatments most plausible, and interaction effect next. All analyses show some difference between the blocks in the experiment.

Example 7.2 A randomized block experiment on the numbers of surviving sugar beet plants (Table 5.9) under four fertilizer treatments in four blocks was performed in Iowa; see Snedecor and Cochran (1967, p.344). Again, a square root transformation is suggested before performing normal theory ANOVA. The same PM's as in the previous example are used below, but with no interaction MM (5.6) since there is only one observation per cell. In addition, the interaction MM (5.5) is used with the Poisson PM.

Step IV. With no PM is the variation in response conditions unimportant. From Table 5.10, the no interaction MM with the three normal PM's provides virtually the same fit, whereas the Poisson PM is somewhat worse with this MM. The interaction MM with the Poisson PM provides a much improved fit.

Step V. All no interaction SM's show that the difference between fertilizers is likely while that between blocks is not. In addition, the interaction Poisson SM shows that the interaction effect is likely.

The same analyses applied to the data of Bartlett (1936a) on the numbers of poppy plants in oats yield similar results.

Example 7.3 P. Wickett (Fisheries Research Board of Canada) has recorded numbers of pink salmon caught by British Columbia fisheries over a forty year period. Since this species of salmon has a two year return cycle, two genetically separable groups return on alternate years. The data for the group coming on even years are given in Table 5.11, together with measures of rainfall (x_1) and sunshine (x_2). Regression analysis was applied to these data with MM's (5.9) and (2.3) using log normal (C) and Poisson (M) PM's and the procedures of chapter II.

Step I. With MM (2.3), the normal theory estimates of the nonlinear parameters are $\hat{\alpha}_1 = 4.5$ and $\hat{\alpha}_2 = 1.1$.

Step IV. The two MM's for the normal PM (C) may be compared by $R(\alpha_1=1, \alpha_2=1) = 0.15$, showing that the nonlinear MM provides little improvement. Thus, the corresponding nonlinear Poisson SM was not attempted, and only the analyses for the two linear SM's are given in Table 5.12. The normal theory model gives a much more plausible fit than the Poisson model; in neither case is elimination of all parameters except β_o plausible (i.e. use of MM (1.5)). Most of the poorness of fit of the Poisson SM seems to come from introduction of the MM. Because of the need to estimate the variance, no statement can be made about this effect for the normal SM.

Step V. Use of the linear term for rainfall (x_1) in the regression is less plausible than the others in both SM's. But the log normal SM only shows inclusion of β_2 and β_4 (i.e. the sunshine terms) to be very plausible. The ranking of plausibilities for the various terms in the regression equation for the two PM's is identical, but the RL's for the Poisson are very much smaller. This has been true of all examples involving Poisson PM's, and will be further discussed in section 11.

8. Binomial Count Data

Analysis of binary data in the context of this chapter presents special problems, since the parameters to be estimated are the actual probabilities (expected frequencies) in the binomial PM. Thus, comparison of PM's will be the same as for the multinomial example 3.1 of chapter IV; the use of the normal PM will actually be only an alternative means (to maximum likelihood estimation) of estimating the probability parameters and not a direct representation of the observed frequency distribution. Comparison will always show the binomial PM MLE's to be better than the estimates obtained from the normal PM. We shall be interested to see how well the normal theory approximates the binomial model estimates.

Using normal theory, from LF (5.1), $E[g(Y_i,\underline{Y})] = \mu_i$ is the function $g(np_i,\underline{Y})$ of the p_i of LF (5.3). Inferences may be made about the normal theory SM's by calculating the MLE's from equation (5.1) for $\mu_i = g(np_i,\underline{Y})$, where $B(\mu_i) = \mu_i$ is any MM, and then comparing PM's by using these estimates, $\tilde{p}_i = g^{-1}(\hat{\mu}_i,\hat{\underline{Y}})/n$ in the binomial LF (5.3).

Alternatively, in steps IV and V of section 4, if a normal approximation seems plausible from the preceding steps, the more conventional inference procedure of looking at the normal LF (5.1) may be used. But, this often yields likelihood ratios of much different size than does the previous procedure.

A number of transformations, $g(y,\underline{Y})$, of the response may be used in the normal theory PM (5.1), as listed in section 5, (E) to (J). Then, for example, if $g(\hat{y}) = B(\hat{\mu}) = \hat{\mu}$ is the predicted value of $g(y)$ in model (E). $\tilde{p} = e^{\hat{\mu}}/(1+e^{\hat{\mu}}) = g^{-1}(\hat{\mu})/n$ and this is used in LF (5.3).

The natural linear binomial SM (N) has

$$B(\mu) = \log[\mu/(n-\mu)] = \log(E(Y)/[n-E(Y)]) \qquad (5.15)$$

as shown in chapter III, section 3, and is analogous to the logit normal model (E).

Example 8.1 If one pair of marginal totals are fixed in a 2x2 contingency table, the table may be analyzed using MM (5.7) $(i = 1,2)$ and the binomial PM (N). Since $\frac{p_1(1-p_2)}{p_2(1-p_1)} = e^{\alpha_1-\alpha_2}$ and $\alpha_1+\alpha_2 = 0$, the plausibility of $\alpha_1 = \alpha_2 = 0$ or no effect of treatments is analogous to the χ^2 test for independence and to Fisher's exact test (although this is conditional on all marginal totals being fixed). For the contingency table of Table 5.13, $\log R(\alpha_1=0) = -0.36$ providing no evidence against the hypothesis of independence. The conditional likelihood analysis of chapter VI for this example and the exact long run test of significance of chapter VII agree with this result.

Example 8.2 Cox (1966) analyzes the data given in Table 5.14 on the effect of rocking on whether or not babies cry. Each day for 18 days, one baby was chosen at random to be rocked out of a group varying in number, providing an 18x2 factorial experiment. With the binomial PM (N), interaction (5.5) and no interaction (5.6) MM's are used to analyze the results.

Step IV. Since MM (5.5) has as many parameters as response conditions, it fits the data exactly; MM (5.6) provides a poor approximation. In neither case is elimination of the variation due to response conditions plausible.

Step V. Neither MM provides strong evidence (Table 5.15) that rocking reduces the number of babies crying. Differences between days (which are not of interest) are very plausible;

MM (5.5) gives an indication of interaction between days and rocking.

The effect of rocking is considered further in chapters VI and VII where a conditional likelihood analysis and an exact significance test are applied. These agree with the above results. The comparison of these results with the asymptotic result of Cox is discussed in chapter VII.

Example 8.3 Snedecor and Cochran (1967, p.300) give the data reproduced in Table 5.16 from a 5x5 factorial experiment in which five replications were made of a comparison of four treatments and no treatment of soybean seeds. Out of 100 seeds planted in each plot, the numbers failing to emerge were counted. Snedecor and Cochran use linear normal theory PM (H) ANOVA with no interaction MM (5.6). The analyses for logit normal (E), odds normal (F), arcsine normal (J), and binomial (N) PM's with no interaction MM (5.6) are also given in Table 5.17, using the binomial LF (5.3) with that for the normal LF (5.1) in parentheses. For all of the PM's, the parameters of the interaction MM (5.5) may be estimated and hence $\tilde{\hat{p}} = g^{-1}(\hat{\mu})/n$, but the variance of PM (5.1) cannot. The values of log R from LF (5.3) when this $\tilde{\hat{p}}$ is substituted in LF (5.3) are given in Table 5.18; those for LF (5.1) cannot be calculated.

Step III. All of the normal SM's provide good approximations to the corresponding binomial SM's. Using MM (5.5), all SM's fit exactly, since there are as many parameters as response values.

Step IV. Variation in the response conditions is plausible for all SM's. In every case, the interaction MM fits better than the no interaction model.

Step V. All analyses using LF (5.3) provide the same result: differences between treatments are more plausible than between replications. Use of MM (5.5) shows that the interaction between replications and treatments is more plausible than the differences between treatments. This does not appear in the analysis of Snedecor and Cochran. Except for the odds normal PM (F), all normal theory analyses using LF (5.1) and MM (5.6) give similar results to those from LF (5.3). The unusual results given by PM (F) are difficult to explain.

Example 8.4 Lindsey, Alderdice, and Pienaar (1970) provide two normal theory analyses (the same as that of chapter II) of data of Alderdice and Forrester (1968) from an experiment to determine the effects of variation in salinity (x_1 °/oo S) and temperature (x_2 °C) of sea water on the proportion of eggs of the English sole (Parophrys vetulus) hatching. The original data, showing counts of eggs hatching, are provided in Table 5.19. The two analyses use the per cent normal PM's (H) and (I) ($\hat{Y} = 0.85$) with linear (5.9) and nonlinear (2.3) regression MM's respectively. Analyses using logit normal (E) and binomial (N) PM's with the two MM's are also given in Tables 5.20 and 5.21 respectively. In the tables, the log RL's from normal LF (5.1) are given in parentheses beside those obtained by substituting normal theory estimates, $\tilde{p}$, in binomial LF (5.3), as in the previous example.

This example provides an extension of the analysis of response surfaces given in chapter II to the case where an alternative PM is available. Once the parameters of an acceptable SM have been estimated, the procedures outlined in chapter II for exploring the response surface are applicable.

Step I. For the nonlinear MM (2.3), the MLE's of the power parameters are $\hat{\alpha}_1 = 0.34$ and $\hat{\alpha}_2 = -0.55$ for the logit normal PM and $\hat{\alpha}_1 = 0.19$ and $\hat{\alpha}_2 = -0.24$ for the per cent normal PM. This first pair of estimates is used in the nonlinear (2.3) binomial SM instead of calculating the exact MLE's.

Step III. A comparison of PM's using MM (1.4) shows that the per cent normal models (H) and (I) are very good approximations to the binomial model (N), while the logit normal model (E) is not.

Step IV. When MM's (5.9) and (2.3) are used in the comparison, none of these normal theory approximations appears to be very good, as shown in Table 5.22. Although the complete analyses are not given here, two other PM's were fitted, the transformed odds normal (G) ($\hat{\gamma} = 0.18$) with MM (2.3) and the arcsine normal (J) with both MM's (5.9) and (2.3). The corresponding results for these SM's are included in Table 5.22. Although neither of these PM's fits very well, when the desired MM's are introduced, they provide better fits than the binomial SM's. This implies that some MM derived from these and used with the binomial PM instead of equation (5.15) (i.e. abandonment of the natural model) should yield an improved fit.

Step V. As seen from Tables 5.20 and 5.21, all SM's provide essentially the same analysis of effects, with only the effect of x_1 linear being relatively unlikely. The RL's using binomial LF (5.3) are in all cases much smaller than those from normal LF (5.1).

In Figure 5.1, contours for p = 0.3 and 0.8 hatch have been plotted for the linear MM (5.9) with the binomial (N) and per cent normal (H) PM's, showing that they give very similar surfaces. Figure 5.2 gives the corresponding nonlinear SM's, MM (2.3) with PM's (I) and (N), which give somewhat different surfaces (much different from those of Figure 5.1). The surfaces for the two best fitting nonlinear SM's are plotted in Figure 5.3, the arcsine normal SM (J) and the transformed odds normal SM (G). All of the surfaces appear to have a similar shape along the plane $x_2 = 4.0$ where the surface rises very steeply, and to differ elsewhere in the factor space.

9. Mathematical Models for the Variance

All of the preceding examples have involved MM's describing variation in a mean parameter with change in the response conditions. For the normal SM's, this has involved an assumption of constant variance. But implicit in the MM's for the other PM's has been an assumption of changing variance, since $\sigma^2 = \mu^2$ for the exponential PM, $\sigma^2 = \mu-\mu^2/n$ for the binomial PM, and $\sigma^2 = \mu$ for the Poisson PM.

If the data in an example actually come from the theoretical PM and the normal model is only an approximation, a linear normal SM will not take this changing variance into account. The use of a transformed response has traditionally been one method of attempting to improve the situation, the method used in the examples. Two alternatives to this are immediate. If sufficient observations are available, a different variance may be estimated for each response condition. Or, the relationship stated above between the mean and variance may be substituted into the normal SM, eliminating the variance parameter. In this case, a normal PM truncated at zero must be used.

The use of an asymptotic Chi-squared likelihood ratio test (see chapter VII, section 4) often actually gives an exact test of significance for the second alternative (using the relationship between the mean and variance in a normal PM) when it is applied to a likelihood ratio such as the Poisson or exponential. This is further discussed in chapter VII.

The assumption of constancy of variance of a normal SM may be assessed by fitting a SM with the variance changing (the first alternative). For the data of example 6.1, an additional MM, for the variance σ^2_{ij}, was introduced into the normal PM (B) with MM (5.5). This SM gives a RL of fit of $\log R_{VN} = -90.51$ with $\hat{\gamma} = 0.52$. But, $R(\gamma=1) = 0.85$ so that normal PM (A) is acceptable with this combination of MM's. The lack of fit for this SM may be compared with those given in Table 5.2 for the other two normal SM's. Although the improvement is large, especially over the linear model, the greatly increased number of parameters may not warrant use of this model. Apparently, the transformation of the response and the changing variance play somewhat the same role in improving the linear PM (A), with changing variance being better at the cost of adding more parameters (see the next section).

A more ad hoc method is to plot residuals. In Figure 5.4, the deviations from expected values for the various response conditions of example 6.1 are plotted against the expected values using the interaction MM (5.5) with the linear normal PM (A). The plot is identical using the exponential PM (K). The variance appears to increase with increase in the size of the mean (expected response). In contrast, the residuals for the corresponding transformed normal SM (B), plotted in the transformed units in Figure 5.5, show no noticeable change in variance with the response conditions. Such a transformation of the response does not change in the same way the expected values of the observed responses when transformed back to the original units, as a comparison of Figures 5.6 and 5.7 shows. In these diagrams, expected values are plotted against observed in the original units; only the scale of the expected values is altered.

Similarly, such plots may be used to compare the deviations for Poisson models. For example 7.1, in plots for the linear normal (A) and Poisson (M) PM's, the deviations increase with the mean, confirming the relatively good fit of the Poisson. A plot for the log normal PM (C) shows little change in variance with the transformed means. For example 7.3, a plot of residuals for the Poisson PM (M) with linear regression shows no increase in variance with the mean. In fact, for the one large response observed (in 1962, see table 5.11), y_{17}= 17,381 is estimated as $\hat{y}_{17}$= 17,694, whereas many smaller responses have a larger deviation than 313. This confirms the relatively very bad fit reported in the example.

For normal theory approximations to binomial SM's, the (transformed) variables can be plotted against the residuals to determine if constancy of variance is valid as above.

When more than one observation is made for each response condition, comparison of PM's in step III is made using a MM with different mean for each condition (i.e. MM (1.4)). For a normal PM, this leaves only the assumptions of normality and constancy of variance for comparison. Hence, fit of such a SM in step III depends only on these two assumptions, and the results of this step for the examples mentioned in this section are confirmed by the plots of residuals. Use of a changing variance SM eliminates the second of these assumptions, providing an assessment of this normality assumption alone.

The random effects model in ANOVA is a more common example of a MM for the variance.

10.Roles of Transformations of Normal Responses

As discussed in chapter II, transformations of the response using a normal PM may perform three functions: (i) to provide a distributional form of PM which better describes the data (e.g. log normal etc.),i.e. to discover new units of measurement making the response variable more nearly normal; (ii) to fulfil better the assumption of constancy of variance; and (iii) to provide a better fit for the KM. The relative importance of these three functions may be analyzed by an extension of the method of the previous section for assessing the assumption of constant variance. If more than two observations are available under each response condition, three SM's may be fitted, and the RL graph for the transformation parameters plotted in each case. Each SM has a common transformation under all response conditions. For the first SM, a different mean and variance is estimated under each response condition; for the second, constant variance with a different mean under each condition. The third SM is the one of interest, using the specified MM.

The transformed per cent normal PM (I) of example 8.4 was analyzed in this way for the data of Table 5.19. The three RL graphs of the power transformations are plotted in Figure 5.8. For the first SM, only the normal distributional form is assumed; the graph is relatively flat with $R(\gamma=1) = 0.75$, showing that the data fulfil this assumption well (as opposed to the transformed normal alternative). For the second SM, the assumption of constant variance is added; the RL graph of γ changes markedly, with $R(\gamma=1) = 0.004$, indicating that this is a poor assumption without the transformation. For the third SM, the further assumptions of the nonlinear response surface MM (2.3) are added; the graph shifts to give a different range of plausible values of the power transformation. Thus, for these data, the transformation performs the second and third functions listed above.

When this analysis is applied to the data of example 6.1, considering the no interaction MM (5.6) as the one of interest, the RL graphs of γ from the second and third SM's are almost identical, revealing that the transformation is not important in reducing interaction effect. For these data, the prime function of the transformation is to fulfil the assumption of constant variance.

11. Discussion

The main problem in extending the principles of normal theory regression analysis and ANOVA to other PM's has lain in the lack of suitable accompanying distributions, such as the F and Student's t distributions, for developing appropriate exact significance tests. In some very simple problems, such as 2x2 contingency tables(which, it has been seen, may be regarded as 2 block one-way ANOVA), exact significance tests may be developed by enumerating all of the possible outcomes of the experiment within some defined sample space, as in chapter VII. But for more complex problems, this rapidly becomes impossible even with a high speed computer. Various approximating techniques have been employed such as asymptotic (large sample) normal approximations (even with small samples) and Monte Carlo methods. In this chapter, likelihood methods of inference have been outlined for a first attack on the data to determine what information they can give about various theoretical SM's before any approximating techniques are used.

One drawback to this method, which has appeared throughout the examples, has been a divergence in the size of likelihood ratios for the same MM when different PM's are used. The key to these differences may come from the discussion of section 9, where the link between the mean and variance is mentioned. When such a link occurs, change in the MM will have more effect on the LF than when no such link exists.

For example, considering the SM's of section 7, in example 7.1, elimination of the effect of treatments from the MM only directly affects the mean for the normal PM's, whereas for the Poisson PM, the poorer fit of the MM (describing the mean) causes the variance, and hence the whole Poisson PM, to fit much more poorly.

Thus, this apparent drawback may be very useful in detecting overall departures from the SM through variation in the MM. This is true because of the existence of the base multinomial PM through which all likelihood ratios from a given set of data are directly comparable.

Comparison of sizes of likelihood ratios between different sets of data (i.e. data measuring different types of responses) is not clearly justified, especially if the sets have very different numbers of observations. But, the same may be said of a simple test of significance (in Fisher's sense): the level at which either a very improbable event has occurred or the hypothesis is wrong will vary with the data set and especially with the number of observations. Of course, it two sets of data are derived independently from the same type of experiment (measuring the same response, although perhaps at different points in the factor space), the corresponding LF's may be combined directly by multiplication.

Another problem involves maximization over parameters in the LF. If the inference desired involves comparison of

SM's, then we are interested in the best that each SM can do and maximization is valid. In this case, differences in the number of parameters maximized over is taken into account by desire for simplicity, theory, or other considerations. The introduction of more parameters to be maximized over will naturally provide more potential for superior fit.

When parameter values are being considered within a given SM, the method of maximizing the LF over parameters not of immediate interest must be viewed as an approximation. If at all possible, use of the entire LF is desirable. Another procedure for making inferences about parameter values for a given SM when nuisance parameters are present will be discussed in chapter VI, using a conditional LF.

When the methods of chapter IV are applied to data involving responses measured under a number of conditions, so that the PM is actually a product of different probability functions as in this chapter, consideration of the average RL per distribution (per condition) i.e. the geometric mean of the RL, is useful. This is given in each table of RL's for the preceding examples in the form $\frac{1}{I}\log R$ where I is the number of conditions. They show that the difference in plausibility (step II) between PM's is usually very small, because of the very small number of observations per condition.

In spite of many apparent problems, proceeding beyond standard normal theory SM's often is useful. Besides providing fresh insight into the data, fitting a theoretically more justifiable PM than one involving the normal distribution often allows more complicated MM's, by eliminating the variance parameter, and provides the potential for better predictive power.

CHAPTER VI

INFERENCES USING THE CONDITIONAL LIKELIHOOD FUNCTION

1. Introduction

In analyzing a set of data, the stage is usually reached where some SM is considered acceptable (plausible) and we wish to consider the plausibility of various parameter values within the model. No problem arises in using the LF for comparison of plausibilities if only one estimable parameter is present in the SM. If more than one such parameter is present and all are of equal interest (perhaps linked in some way), again theoretically, no problem arises. Various SSM's to be considered may be compared by use of the likelihood ratio. If we wish to see how plausibility varies with change in the parameter values, methods of visualization become increasingly difficult with larger numbers of parameters. As used in previous chapters, one method of overcoming this problem is consideration of a silhouette of the likelihood surface, obtained by maximizing the LF over some of the parameters. But the approximating interpretation of this must be kept in mind.

Often a special circumstance appears when a SM has been adopted for the explicit consideration of only some of its estimable parameters. Although the remaining nuisance parameters are essential for considering the SM as a whole (steps I to IV of chapter V, section 4) and the LF is then used, after the SM has been found to be acceptable, we wish to make inferences about the parameters of interest in the absence of knowledge of what values the nuisance parameters might have. Since sufficient statistics contain all of the information in the data about a parameter, in order to do this exactly, all of the nuisance parameters must have sufficient statistics. Then, inferences about the parameter(s) of interest will be made, conditional on this information about the nuisance parameters, through the use of the conditional probability function in the form of a conditional likelihood function (CLF).

Fraser (1967) introduced the concept of marginal likelihood in connection with structual inference, for use in the elimination of nuisance parameters. Sprott (1968) and Kalbfleisch and Sprott (1970) have extended this concept by considering likelihoods based on distributions conditional on sufficient statistics. This is somewhat similar to obtaining some forms of uniformly most power similar confidence tests for multiparameter distributions by conditioning on sufficient statistics, as in Lehman (1959, p.134), and also to Fisher's exact test for the 2x2 contingency table

Because of the strong restrictions, CLF's will only be applicable in certain special cases. In addition, their use will be further restricted by the difficulty of deriving the appropriate conditional distribution for many of the MM's considered in chapters III and V. This difficulty usually results from the condition (as in ANOVA) that a number of parameters sum to zero, e.g. MM's (5.5), (5.6), (5.7).

Of the MM's considered in previous chapters, those for ANOVA may often present cases where a CLF would be desirable. In many such problems of two-way ANOVA, only the effect one way is of direct interest (see any of the two-way ANOVA examples of chapter V). Then, for example, we should be interested in making conditional likelihood inferences about all of the α_i's of no interaction MM (5.6) in the absence of knowledge about μ or any of the β_j's. Unfortunately, in such cases as this, derivation of the conditional distribution is usually difficult. The MM of a regression problem does not usually contain nuisance parameters, although in cases such as response surface methodology it may.

When transformation of the response is introduced into a continuous PM, we may be primarily interested in what values of the transformation parameters are plausible. The same applies for the nonlinear parameters of a regression MM such as equation (2.3). Then, all other parameters will be considered as nuisance parameters.

Two types of examples will be provided in succeeding sections: conditional likelihood inferences about parameters of interest in ANOVA involving a binomial PM (of chapter V, section 8), an extension of the likelihood analogue of Fisher's exact test, and about the power parameters of normal SM's.

As a prelude to the first of these, and an extension of the results of chapter III, the CLF's will be derived for various MM's using the binomial PM.

2. The Conditional Likelihood Function

Suppose that the SM represented by $f(\underline{y};\underline{\phi})$ has parameters of interest $\underline{\phi}_1$ with nuisance parameters $\underline{\phi}_2$, and that, for fixed $\underline{\phi}_1$, minimal sufficient statistics, $\underline{t}_2$, possibly functions of $\underline{\phi}_1$, exist for the elements of $\underline{\phi}_2$. Then, the marginal distribution for these statistics is given by

$$f_M(\underline{t}_2;\underline{\phi}) = \int_{\underline{t}_2=\text{const.}} f(\underline{y};\underline{\phi})d\underline{y} \quad , \qquad (6.1)$$

where the integral becomes a summation for discrete SM's. The conditional distribution of the observations given these sufficient statistics will then not be a function of the nuisance parameters:

$$f_C(\underline{y};\underline{\phi}_1/\underline{t}_2) = f(\underline{y};\underline{\phi})/f_M(\underline{t}_2;\underline{\phi}) \quad . \qquad (6.2)$$

From this conditional distribution, a CLF may be obtained which may be handled in the same way as the ordinary LF previously used, although the interpretation of inferences will be different, as described in the previous section. Complications arise, treated in Kalbfleisch and Sprott (1970), when a transformation of the response contains a parameter of interest, as in the SM's arising from LF's (5.1) and (5.2). Only relatively straightforward examples will be used here.

The CLF is equivalent to the conditional distribution of equation (6.2):

$$L_C(\underline{\phi}_1) = f_C(\underline{y};\underline{\phi}_1/\underline{t}_2) \quad . \qquad (6.3)$$

As with the ordinary LF, all comparisons use ratios of CLF's. Analogous to equation (1.7), a conditional relative likelihood (CRL) function may be defined:

$$R_C(\underline{\phi}_1) = L_C(\underline{\phi}_1)/L_C(\hat{\underline{\phi}}_1) \quad , \qquad (6.4)$$

where $\hat{\underline{\phi}}_1$ denotes the parameter vector maximizing the CLF. Again, with many parameters of interest, looking at profiles of the CRL surface by maximization over all but one or two parameters of $\underline{\phi}_1$ may be a useful approximate tool, although exploration of the complete surface will be more meaningful.

Note that no base CLF, análogous to the multinomial LF used in the ordinary likelihood inferences of previous chapters, is defined. When theoretical problems arise in derivation of a CLF for a continuous PM, retention of the base multinomial LF may clarify the problem. Then, equation (6.1) becomes

$$f_M(\underline{t}_2;\underline{\phi}) = \sum_{\underline{t}_2=\text{const.}} f(\underline{y};\underline{\phi})\Delta\underline{y} \quad .$$

Theoretically, this is useful in such problems as the derivation of a CLF for the power transformation of the response.

3. Conditional Likelihoods for Binomial Statistical Models

Various natural linear SM's for the PSD and specifically, for the binomial PM have been extensively considered in chapter III. For the PSD, if each element of the parameter vector $\underline{\phi}$ has a corresponding minimal sufficient statistic in the statistic vector $\underline{t}$, and if ϕ_1 is the parameter of interest, then the CLF is

$$L_C(\phi_1) = \frac{a'(\underline{t})\phi_1^{t_1}}{\sum_z [a'(z,t_2,\ldots,t_k)\phi_1^{z}]} \quad . \qquad (6.5)$$

This may easily be extended to a parameter vector of interest, $\underline{\phi}_1$. The SM's considered for the binomial distribution will all be of this form, with sufficient statistics for all parameters, since natural linear MM's are used.

For the analysis of J 2x2 contingency tables (a 2xJ randomized block design with binomial error) with no interaction MM (3.9), the parameter $\alpha_1 = -\alpha_2$ is of interest. Then,

$$L_C(\alpha_1) = \frac{e^{2\alpha_1 y_{1\cdot}}}{\sum_z \left[a'(y_{\cdot\cdot},\underline{r},y_{\cdot\cdot}-z)e^{2\alpha_1 z}\right]} \qquad (6.6)$$

from equation (6.3), where $a'(y_{\cdot\cdot},\underline{r},s_1)$ is from equation (3.10) and $\phi_1 = e^{\alpha_1}$. If this SM is implausible and the interaction MM (3.6) is necessary, then the CLF becomes

$$L_C(\alpha_1) = \frac{e^{2J\alpha_1 y_{1\cdot}}}{\sum_z \left[\prod_j \binom{N_{1j}}{y_{1j}-y_{1\cdot}+z}\binom{N_{2j}}{y_{2j}-y_{2\cdot}+y_{\cdot\cdot}-z} e^{2J\alpha_1 z}\right]} \;. \qquad (6.7)$$

In this case, $a(\underline{y})$ need not be modified since the interaction MM has as many parameters as response conditions. These two CLF's will be used to make further inferences about some of the parameters in examples of chapter V, section 8.

For the analysis of an Ix2 contingency table with MM (3.12) (one-way ANOVA), the difference parameters, $\underline{\alpha}$, may be of interest with μ a nuisance parameter. The CLF is

$$L_C(\underline{\alpha}) = \frac{\prod e^{\alpha_i y_i}}{\sum_{\underline{r}} \left[\prod_{i=1}^{I-1}\binom{N_i}{r_i} e^{\alpha_i r_i}\binom{N_I}{y_\cdot-\sum r_i} e^{\alpha_I (y_\cdot - \sum r_i)}\right]} \;. \qquad (6.8)$$

Again, no modification of $a(\underline{y})$ is needed. For I = 2, this reduces to the CLF for the 2x2 contingency table,

$$L_C(\alpha_1) = \frac{e^{2\alpha_1 y_1}}{\sum_{r_1}\left[\binom{N_1}{r_1}\binom{N_2}{y_\cdot - r_1} e^{2\alpha_1 r_1}\right]} \;. \qquad (6.9)$$

Equations (6.6) and (6.7) also yield this result when J = 1 as does the conditional distribution used to make Fisher's exact test of significance for the 2x2 table.

For the binomial regression SM with MM (3.14) and p = 1, suppose the parameter of interest is β_o, after removal of the effect of the independent variable x_{i1}. Then,

$$L_C(\beta_o) = \frac{e^{\beta_o t_o}}{\sum_z \left[a'(z,t_1)e^{\beta_o z}\right]} , \qquad (6.10)$$

where $t_o = \sum y_i$ and $a'(t_o,t_1)$ is from equation (3.15). This may be extended to a multiple regression situation where only some of the coefficient parameters are of interest.

4. Examples from the Binomial Distribution

4.1 The traditional SM to which a conditional argument has been applied is that for the 2x2 contingency table for which Fisher proposed his exact test of significance. The CLF in this simple case illustrates the main points of conditional likelihood inference.

In Figure 6.1, the RL function for α_1 maximized with respect to μ of MM (5.7) has been plotted for the data of chapter V, example 8.1. In the same figure, the CRL function for α_1 is plotted using equation (6.9). Log $R_C(\alpha_1=0) = -0.34$ whereas log $R(\alpha_1=0) = -0.36$ and the MLE, $\hat{\alpha}_1 = -0.45$ virtually the same as the conditional MLE.

Use of the maximized (over μ) RL function for α_1 assumes that μ takes on (is known to be) the value maximizing the LF with α_1 set. Use of the CRL function assumes that the value of μ is unknown. This makes very little difference for inferences in this problem, as illustrated by the very similar graphs for the two functions in the figure.

4.2 The example (8.2, chapter V) from Cox (1966) concerning the effect of rocking on babies crying is a case where a one-way treatment effect is important in a two-way ANOVA. In Figure 6.2, the RL function for α_1 maximized with respect to μ and β of no interaction MM (5.6) has been plotted, along with the CRL function. Again, the CRL graph is very similar in shape to that for the RL, but log $R_C(\alpha_1=0)$ = -2.04, $\hat{\alpha}_{1C} = 0.62$ whereas log $R(\alpha_1=0) = -2.33$, $\alpha_1 = 0.70$, as indicated by the displacement of the graph in the figure. Thus, use of the CRL function makes an effect due to rocking less plausible and gives a smaller best estimate of difference in effect.

But the no interaction MM was found to be implausible in chapter V. The corresponding plots for interaction MM (5.5) yield uniform (R=1) graphs for α_1 in both cases, i.e. for this model, the data provide no information about the difference in effect of rocking. This reveals another reason for plotting the LF. The analysis in chapter V using MM (5.5) gave log $R(\alpha_1=0) \doteq 0.0$ which might lead one to conclude that α_1 lies very near zero. In fact, the data do not indicate this, and only a plot reveals that the LF is flat.

The first part of this analysis agrees with the results provided by Gart (1970b).

4.3 Gart (1970a) analyzes the results of an experiment by Innes et al testing the carcinogenic effect on mice of a fungicide, Avadex. As given in Table 6.1, individuals of two strains of mice, split by sex to give four strata, are either treated or used as controls, and the number with tumors after 85 weeks recorded. This yields a 2x4 ANOVA table. These data are analyzed as in chapter V, section 8 using the binomial PM (5.3) with the interaction (5.5) and no interaction (5.6) MM's. This analysis, provided in Table 6.2, shows plausible effects both of the treatments and within the sex-strain grouping. Although the RL for no effect of treatment is slightly larger than that for sex-strain differences, this effect may be considered more plausible because the change of MM here only involves loss of one parameter whereas for the sex-strain effect, three parameters are involved. No interaction effect is indicated using MM (5.5).

Primary interest lies in the carcinogenic effect of the treatment. Thus, the RL function for the parameter of interest, maximized with respect to the other parameters, has been plotted in Figure 6.3 for both of the MM's . For comparison, the corresponding CRL functions have been plotted in the same figure. All four functions are nearly identical, with those for the interaction MM slightly wider.

With other sets of simmilar data analyzed, the same result occurs. The conditional and ordinary LF's for a given MM are very similar, although the difference in graphs between two MM's may be greater than in this example. The largest discrepancy found between the two analyses for a given MM occurs in the previous example (Figure 6.2), but here the corresponding graphs for the other (interaction) MM are completely different, being flat.

5. Conditional Likelihoods for Nonlinear Normal Models

A second type of parameter of interest in the SM's discussed in preceding chapters may be the power transformation, either of the response for a continuous PM or of the factors of a regression or response surface MM. Conditional likelihoods for transformation of the response raise special difficulties about the metric of the sample space, since the transformation is incorporated in the differential of the LF; see equation (5.1). Kalbfleisch and Sprott (1970) discuss some of the problems encountered in this situation. Only the second case, transformation of the independent variable, will be considered here, for a normal PM.

As a simple example of the development, consider the MM

$$B(\mu_i) = \beta_o + \beta_1 x_i^{\alpha} \quad , \tag{6.11}$$

with the normal PM of LF (5.1) and $g(y,\underline{Y}) = y$. The nuisance parameters are $\underset{\sim}{\beta}$ and σ^2. For given α, the sufficient statistics for these parameters are $t_o = \sum y_i$, $t_1 = \sum y_i x_i^{\alpha} / \sum x_i^{2\alpha}$, and $t_2 = \sum (y_i - \hat{\beta}_o - \hat{\beta}_1 x_i^{\alpha})^2$. The marginal distribution is

$$f_M(\underline{t};\underset{\sim}{\beta},\alpha,\sigma^2) = \frac{\sqrt{n}\, t_2^{\frac{n-4}{2}} \sqrt{\sum (x_i^{\alpha} - \overline{x^{\alpha}})^2}}{\pi\, \Gamma(\frac{n-2}{2})\, 2^{n/2} \sigma^n}$$

$$\cdot \exp\left[-\frac{t_2}{2\sigma^2} - \frac{n(\beta_o + \beta_1 \overline{x^{\alpha}} - \hat{\beta}_o - \hat{\beta}_1 \overline{x^{\alpha}})^2}{2\sigma^2} - \frac{\sum (x_i^{\alpha} - \overline{x^{\alpha}})^2 (\beta_1 - \hat{\beta}_1)^2}{2\sigma^2} \right]$$

and the conditional distribution is

$$f_C(\underline{y};\alpha/\underline{t}) = \frac{\Gamma\left(\frac{n-2}{2}\right)}{2\pi^{\frac{n-2}{2}} t_2^{\frac{n-3}{2}}} = \frac{\Gamma\left(\frac{n-2}{2}\right)}{2\pi^{\frac{n-2}{2}}\left[\sum(y_i-\hat{\beta}_o-\hat{\beta}_1 x_i^{\alpha})^2\right]^{\frac{n-3}{2}}},$$

so that the CLF is

$$L_C(\alpha) = \left[\sum(y_i-\hat{\beta}_o-\hat{\beta}_1 x_i^{\alpha})^2\right]^{-\frac{n-3}{2}}.$$

This compares with the ordinary maximized LF

$$\max_{\underline{\beta},\sigma^2} L(\alpha,\underline{\beta},\sigma^2) = \left[\sum(y_i-\hat{\beta}_o-\hat{\beta}_1 x_i^{\alpha})^2\right]^{-n/2}$$

derived from equation (2.4). In general, the power in the LF will change from $-\frac{n}{2}$ to $-\frac{n-p-1}{2}$ in the CLF, where p is the number of coefficient parameters conditioned out. For the two factor response surface model in the example of chapter II, p = 6 and the CLF is a function of the two parameters, α_1 and α_2. This function can then be maximized over one of the parameters and plotted as in chapter II, or contour plots of the surface can be made as below.

The effect of use of this CLF on inferences may be more clearly seen than in the previous two sections. Here, the power in the LF is altered by a factor corresponding to the number of nuisance parameters eliminated; the graph for the parameter(s) of interest is widened without changing the MLE's. Little effect is produced if the number of parameters eliminated is much smaller than the number of observations.

The nonlinear SM of chapter II, section 5 is refitted to the data of that section, but with $\gamma = 1$, giving MLE's of $\hat{\alpha}_1 = 1.61$ and $\hat{\alpha}_2 = 1.36$. Contours of the RL function are plotted in Figure 6.4 along with those for the CRL function. For these data, n = 22 and p = 6 so that the power changes by a factor of 15/22 producing considerable effect. For an analogous MM with data from chapter V, section 8.4, the effect would be much smaller since n = 68 and the factor is 61/68.

6. Discussion

The results of section 4 give some indication that conditional inferences may make very little difference when the nuisance parameters eliminated are of the same form as the parameters of interest (linear parameters describing the mean in this case). This is related to the fact that variation in one parameter will not greatly affect the MLE of another, i.e. the parameters are estimated almost independently. In this context, it may be noted that the analogous CLF for the Poisson PM with the no interaction MM (5.6) is identical to the corresponding LF maximized with respect to the nuisance parameters. When the parameter of interest is of a different type from the nuisance parameter, a definite difference in inferences is often discernable, as in section 5. Another simple example involves data from a normal PM with mean, μ, and variance, σ^2, where μ is the nuisance parameter. The conditional MLE of σ^2 is $\sum(y_i-\bar{y})^2/(n-1)$ as opposed to the MLE, $\sum(y_i-\bar{y})^2/n$; the CLF is correspondingly modified.

Normal theory ANOVA is not changed by introduction of the conditional argument if the variance is known. Conditioning out a (constant unknown) variance makes some difference in the width of the likelihood graph without affecting the estimates of mean parameters. From equation (2.5), the normal LF with no power transformation, maximized with respect to the variance and to any other nuisance parameters, ϕ_2, is

$$L(\phi_1) = \left[\sigma^2(\phi_1,\hat{\phi}_2)\right]^{-n/2} .$$

Corresponding to this the CLF is

$$L_C(\phi_1) = \left[\sigma^2(\phi_1,\hat{\phi}_2)\right]^{-\frac{n-p-1}{2}} ,$$

where the vector ϕ_2 has p elements. The lack of effect is associated with independence of the sufficient statistics for the mean parameters in contrast to joint sufficiency with the variance statistic. The same is true for the Poisson SM mentioned above. Apparently, the ANOVA MM's for the other PM's approach independence closely, although not exactly, yielding CLF's very similar to the LF's.

CHAPTER VII

TESTS OF SIGNIFICANCE

1. Introduction

The previous chapters have provided a means of attacking one of the problems outlined in chapter I, where the plausibilities of a number (possibly infinite) of hypothesized SSM's are to be compared. No assumption is made about preference for any subset of these SSM's before collection of the data. If this problem of chapter I is extended to include an assumption of preference for members of the set of possible SSM's, by means of prior distributions or loss functions, then the appropriate posterior probabilities (from Bayes' theorem) or losses may be calculated extending the likelihood analyses derived in the previous chapters. This process will not be discussed here.

Often, an assumption (null hypothesis, H_o) will be available of strict preference for one or a few of the possible SSM's, and this will be of special interest (i.e. consideration of the first purpose of chapter I, section 1). This subset will often be derived by setting some parameter(s) defining the SM constant, e.g. no effect in an ANOVA MM. If this set does not consist of a unique element, i.e. define a unique SSM, a conditional argument, as in chapter VI, is used in the following procedure. In this case, the preference assumption is a composite H_o. When H_o defines a single SSM, it is called simple.

For a simple H_o, the probability of any outcome in the sample space, including the one observed, may be calculated. If the observed outcome has a smaller probability of occurrence than most possible outcomes, either a rare event has been observed or the assumed SSM is wrong. One method of determining where the observed outcome stands on the scale of possible outcomes under the given SSM is to sum up the probabilities of all outcomes with at least as small a probability as that of the one observed, giving a significance level, as described in Fisher (1959).

Extreme care must be taken in defining what constitutes an outcome for continuous PM's, i.e. in defining the intervals into which observations may fall. Some natural unit of interval width, Δy_i, should be used as in chapter IV, since a nonlinear transformation of this will change the probability scale of possible outcomes. Of course, intervals with small probability may be combined as in discrete cases. An example of a nonlinear width was given in chapter IV, example 4.7 where $\Delta' y_i = \log \Delta y_i$ was used.

For a composite H_0, the same procedure must be repeated for each SSM in the subset. This provides a collection of significance levels, one for each SSM. The level of largest size will be the significance level of the subset of SSM's, since sufficient evidence must be obtained about each SSM being wrong before considering the entire subset to be wrong. This is discussed in Fisher (1959, pp.89-93).

Sometimes, all of the information provided by the data about variation in plausibility of the various SSM's within the subset may be summarized in sufficient statistics for the parameters determining the various members of the subset, i.e. the nuisance parameters. Then, setting these sufficient statistics equal to their observed values defines a subspace and a conditional probability statement may be made about each element of this subspace. The significance level is calculated using the probabilities of outcomes in the sample subspace analogous to when the subset contains a single SSM. Thus, a composite H_o can be considered simple in the subspace. This is equivalent to the argument for deriving a CLF in chapter VI, section 3.

Unless sufficient statistics are available for the parameters set constant by H_o, placing the observed outcome in the scale of possible outcomes according to probability of occurrence will usually be very difficult. The search for statistics to provide this result without great loss of information has resulted in the approximations of Neyman-Pearson theory. These problems will be further discussed in section 5.

Occasionally, as well as giving a fixed parameter value (e.g. zero mean), H_o states that the possible parameter values lie in a range (e.g. non-negative mean) smaller than that allowed by the SM. Then, to make the exact test, the possible outcomes with smaller probability to be considered are restricted to those unfavourable to this assumption, giving a one-sided test. This is only possible when sufficient statistics are available for the parameters so that a monotone relationship between parameters and statistics can be used to restrict outcomes considered for smaller probability.

Two points should be noted. The use of tests of significance is an attempt to provide an absolute measure of implausibility about certain selected SSM's and not to compare plausibilities of various SSM's, as with a likelihood ratio, nor to provide probabilities of long run error as in Neyman-Pearson theory. The significance level obtained is a probability statement about possible outcomes, not about parameter values or SM's.

Care must be taken if several H_0's are available about a SM, each involving a different set of parameters in the SM (e.g. no effect each way in two-way ANOVA or stepwise multiple regression where different answers may result depending on whether parameters are added or removed). After an inference has been made about one H_0 (subset of SSM's), further inferences must take this result into account. This must be especially emphasized where conditional significance levels are involved. Only when the two sets of sufficient statistics are distributed independently, not jointly, can order of the inferences be ignored. This precaution is not necessary when making ordinary likelihood inferences, since these always involve comparison of SSM's ("preferences").

For example, in multiple regression with two independent variables and a normal PM, suppose the first H_o, that $\beta_1 = 0$, is plausible. Then, testing the second H_o, that $\beta_2 = 0$, in the original MM, without taking the first result into account, and finding this second H_o plausible, of course, does not mean that both parameters may be plausibly eliminated (unless the sufficient statistics are distributed independently, i.e. orthogonal polynomials are used and the variance is known). Again, suppose $\beta_1 = 0$ is plausible and, taking this into account, $\beta_2 = 0$ is implausible. Then, testing $\beta_2 = 0$ first will not necessarily yield the same result.

Often, the subset of SSM's is not derived by setting some parameter(s) constant; preference is only stated for a given PM (e.g. H_o: the data arise from a Poisson PM). Thus, the test of significance is used for determining the goodness of fit of the SM no matter what values the parameters have. All of the above arguments also apply in this case.

Only occasionally is calculation of exact significance levels for SSM's feasible. For many situations involving normal PM's, the necessary distributions have been tabulated, at least for the SSM where the parameter(s) disappear from the model. For discrete PM's involving small counts, tabulation of the probabilities of all possible outcomes is often possible, using an electronic computer, and hence derivation of the significance level. In other cases, asymptotic approximations are usually adopted. These involve representing the subset of SSM's by an equivalent set of distributions which has been or easily can be tabulated.

Thus, two situations arise where approximate methods are needed for tests of significance: (i) sufficient statistics are available for the parameter(s) about which a null hypothesis is made, but the distribution has not been and cannot easily be tabulated; (ii) no sufficient statistics for these parameters are available making determination of possible outcomes with smaller probability than that of the one observed difficult. In the first case, an aymptotic distribution is required, assuming that enough observations are available. In the second case, an approximate statistic using most of the information in the data about the parameter(s) is required. In either case, an approximate LF results; the usefulness of the approximate test for a particular set of data may be assessed by comparing the approximate LF graph with that for the observed exact LF.

2. Exact Tests of Significance

Fisher (1959. pp.86-89) discusses an exact test of significance for a difference parameter in the 2x2 contingency table using a conditional argument as in chapter VI, example 4.1. This analysis may be extended to the case of J 2x2 contingency tables in analogous manner to the likelihood inference procedure of examples 4.2 and 4.3 of the same chapter. Thus the probabilities of outcomes are defined by a binomial PM with no interaction MM (5.6) where the β parameters refer to differences between tables and $\alpha_1 = -\alpha_2 = \alpha_o$ defines the subset of preferred SSM's. As in the derivation of CLF (6.6), the subset of SSM's may be reduced to one SSM by conditioning on the sufficient statistics for μ and $\underset{\sim}{\beta}$, giving the conditional probability function,

$$f_C(y_{1\cdot};\alpha_1) = \frac{a'(y_{\cdot\cdot},\underline{r},y_{\cdot\cdot}-y_{1\cdot})e^{2\alpha_1 y_{1\cdot}}}{\sum_z\left[a'(y_{\cdot\cdot},\underline{r},y_{\cdot\cdot}-z)e^{2\alpha_1 z}\right]} , \quad (7.1)$$

from which CLF (6.6) was derived. After replacing α_1 by the value α_o, the probability of any outcome in the sample subspace may be calculated, and the conditional significance level derived. Since tabulation of these probabilities is necessary in order to calculate the denominator of the CLF (6.6), if the CLF is plotted, very little further calculation is needed to determine the significance level. In the same way, a conditional significance level may be calculated for a hypothesized value of α_1 in MM(5.5) after deriving the corresponding conditional probability function.

Since even data consisting of continuous measurements are actually discrete, theoretically, exact significance levels can always be calculated by this summation procedure in the discrete sample space. In practice, for continuous PM's, significance levels are calculated by integration. In this way, the tables used in normal theory analysis have been produced. A simple example from normal theory occurs when H_o sets the mean as μ_o for unknown variance. Then, the conditional distribution used for an exact test is Student's t.

If sufficient statistics are available for all of the parameters in a SM, the conditional argument will yield an exact test for the H_o that the data come from the given SM without specifying any parameter values. For a discrete SM, the significance level is obtained by summing probabilities in the sample subspace as before. Since integration is difficult in the continuous case because of lack of sufficient statistics for the SM after conditioning, summation may be used, considering the data discrete. This procedure is only used for lack of fit of a single SM. If other SM's are also being considered, as in chapter IV, they may be compared with the first using likelihood ratios.

3. Exact Tests for Contingency Tables

The three numerical examples of the previous chapter for ANOVA using a binomial PM will be reconsidered for the H_o, $\alpha_1 = \alpha_2 = 0$. In each case, additional parameters are present in the SM so that the preference assumption yields a subset of SSM's instead of a unique SSM. Sufficient statistics are available both for the nuisance parameters, so that the conditional argument can be used, and for the parameters of interest, so that determination of outcomes less probable in the sample subspace is simple.

The first example (8.1, chapter V and 4.1, chapter VI) involves one 2x2 contingency table; we wish to determine the implausibility of no effect between treatments. Summation over the sample subspace determined by conditioning on the sufficient statistic, $y.$, for the nuisance parameter, μ, yields the probability 0.608 of an event occurring which is at least as improbable as the one observed under H_o, so that the given outcome is relatively quite probable under H_o. This is Fisher's exact test.

The second example (8.2, chapter V and 4.2, chapter VI) involves eighteen 2x2 contingency tables for which we wish to determine the implausibility of no (common) effect between treatments of the babies. Here, two MM's are possible, as considered in chapter V. Each will lead to a different test, since variation in the subset of SSM's determined by H_o will be different i.e. the nuisance parameters will be different. For the interaction MM (5.5), the sample subspace is determined by the sufficient statistics, $y_i.$ and $y_{i1}-y_{i2}-y_{\cdot 1}+y_{\cdot 2}$ ($i = 1,\ldots,18$), for the parameters μ, $\underset{\sim}{\beta}$, and $\underset{\sim}{\gamma}$. Unfortunately, this subspace contains only one point, the observed outcome, yielding a significance level of 1.0. For the no interaction MM (5.6), the sample subspace is determined by $y_i.$ ($i = 1,\ldots,$ 18) for the parameters μ and $\underset{\sim}{\beta}$. Under this H_o, the probability is 0.062 of observing an outcome at least as improbable as that observed, again showing that H_o is not too implausible for the observed outcome.

Cox (1966) adds to this H_o that α_1 must be non-negative (i.e. that rocking cannot produce a negative effect). A sufficient statistic is available so that outcomes in the sample subspace unfavourable to this assumption may be determined. Summation over this restricted subspace yields a (one-sided) level of 0.0449, showing the H_o to be more implausible under this additional assumption. This H_o with an additional assumption should not be confused with the H_o that α_1 is non-positive. In this case, the significance level must be calculated for each non-positive α_1 in the ordinary (two-sided) way and the largest level chosen for the composite H_o.

The final example (4.3, chapter VI) involves four 2x2 contingency tables, where the test is for no carcinogenic effect of Avadex on mice. As in the previous example, two MM's are possible. The likelihood analysis of chapter VI revealed little difference between the two MM's (Table 6.2). But, the significance level of H_o for the no interaction MM is 0.0096 whereas it is 0.036 for the interaction model.

4. Asymptotic Tests of Significance

If the required sufficient statistics are available, any test of significance, theoretically, can be reduced to a test involving one (perhaps by conditioning) or more SSM's. In this section, cases are considered where such a reduction can be made but where the resulting (perhaps conditional) SSM neither has been nor can be readily tabulated. Then, some asymptotic distribution is needed which represents the SSM exactly when the number of observations becomes very large and is a good approximation for smaller numbers. Thus, a normal distribution or some distribution derived from it is usually most useful.

The two most common approximations involve the (C)LF. (i) If H_o is true,

$$\frac{\partial}{\partial\phi}\log L(\phi)\Bigg/\left[-E\left(\frac{\partial^2}{\partial\phi^2}\log L(\phi)\right)\right]^{\frac{1}{2}}\Bigg|_{\phi_o} \sim N(0,1) \qquad (7.2)$$

for large n. (ii) Asymptotically, $-2 \log R(\phi_o)$ will have a Chi-squared distribution with one degree of freedom when H_o is true. This assumes that the (conditional) MLE

$$\hat{\phi} \sim N\left(\phi_o;\ 1\Big/n\ E\left[\left(\frac{\partial}{\partial\phi}\log L(\phi)\right)^2\right]\Bigg|_{\phi_o}\right) \quad ; \qquad (7.3)$$

see Wilks (1962, p.408). The extensions to multi-parameter situations are straightforward. For the conditional argument, L and R are replaced by L_C and R_C. Other approximations have been derived for special circumstances.

To check the appropriateness of the approximation for a given set of data, the LF derived from the assumptions (normal in the above cases) may be plotted along with the exact LF. Unfortunately, in many situations where such a test of significance is desirable, the exact (C)LF is difficult to tabulate, as for the CLF of chapter VI, examples 4.2 and 4.3. In this case, the original assumptions (7.3) of the Chi-squared approximation may be used under the null hypothesis, as by Cox (1966). But the approximate LF (conditional in this case) cannot be compared easily with the exact one. Sometimes, point comparison of LF's is useful, as was done in chapter IV, example 3.5.

A simple example of checking appropriateness occurs with the asymptotic Chi-squared test for the 2x2 contingency table approximating the exact test given in the previous section. The approximate LF can be plotted for various values of α_1 by using the fact that χ^2_1 is the square of a normal, N(0,1), variate with variance proportional to $\hat{p}_1(1-\hat{p}_1)/e^{2\alpha_1}(\hat{p}_1+e^{2\alpha_1}-\hat{p}_1 e^{2\alpha_1})^2$. The expected values in the table are calculated by fixing the marginal totals and setting the log odds ratio equal to the determined constant, $2\alpha_1$. For the data of the first example of the previous section, the approximate test gives a significance level of 0.401 compared with the exact 0.608. A graph of the approximate LF has been included in Figure 6.1, showing the difference which explains the divergence in significance levels.

As a second example, consider the data of chapter IV, example 3.3. Suppose H_o is that $\mu = 1$ in the Poisson PM, yielding a unique SSM. Then, the exact level of significance is 0.52 providing no evidence against H_o. Using an asymptotic Chi-squared likelihood ratio test, the approximate level is 0.55, with log $R(\mu=1) = -0.288$. The approximate LF derived from equation (7.3) comes from

$$30\hat{\mu} = 26 \sim N(30\mu, 30\mu) \quad .$$

(See chapter V, section 9). This and the exact LF are plotted in Figure 7.1. The approximate log $R(\mu=1) = -0.342$. The divergence in LF's explains the difference in significance levels. This asymptotic test approximates the Poisson PM by a normal PM with $\mu = \sigma^2$.

These methods can be extended to the asymptotic Chi-squared goodness of fit test usually used for discrete data. Expected values of the frequencies are calculated from the SM, using MLE's of any unknown parameters. The Chi-squared test applied to deviations from observed values provides an asymptotic equivalent to the exact test of section 2. In this case, however, no parameters are left for an approximate or an exact LF to determine how good is the approximation.

Often, an asymptotic approximation is useful for determination of outcomes less probable than the one observed. Since many distributions are asymptotically symmetric (normal), distance of the MLE of the mean from the value given by H_0 for any possible outcome can be used. Those outcomes giving $\hat{\mu}$ further from μ_0 than the observed $\hat{\mu}$ are considered less probable. If the distribution actually is symmetric, this yields an exact test. This asymptotic procedure usually is useful only for a mean parameter.

Sprott and Kalbfleisch (1969) discuss comparisons of approximate and exact LF's and show how transformation of the parameter may provide a better approximation.

5. Tests of Significance without Sufficient Statistics

Most of the examples considered in the previous chapters used SM's with sufficient statistics available for all parameters. The only major exception is the normal SM with power transformations of the response or of the regression variables. If H_0 specifies a value of the transformation parameter(s), sufficient statistics are available for the remaining parameters and the conditional argument applies. But no sufficient statistics are available for the transformation parameters making determination of outcomes less probable than the one observed difficult. Two procedures are available: (i) the two asymptotic tests, (7.2) and (7.3), of the preceding section are usually available; (ii) some statistic may be found which is almost sufficient and an approximate test applied. In either case, an approximate LF is available to compare with the exact one.

Box and Cox (1964) use the first method for the data of chapter V, example 6.1 for a power transformation of the response, but without using a conditional argument.

Williams (1962) uses the second method for a SM involving nonlinear regression parameters. In a simple example, suppose $Y_i \sim N(e^{-\alpha x_i},1)$, i.e. a normal PM with MM

$$B(\mu_i) = e^{-\alpha x_i} \quad .$$

Then, the statistic, $\sum y_i x_i e^{-\alpha x_i}$, has distribution $N(\sum x_i e^{-2\alpha x_i}, \sum x_i^2 e^{-2\alpha x_i})$ which may be used to calculate an approximate significance level for some H_0 about α. The distribution of the statistic may be used to plot an approximate LF for comparison with the assumed SM for the data to determine how much of the information in the data is used, i.e. how closely the statistic approaches sufficiency. If coefficient parameters are present in the MM, a conditional argument is used. This procedure may be extended to obtain an approximate test when a nonlinear MM such as equation (6.11) or (2.3) is used. analogous to the CLF of chapter VI, section 5.

6. Discussion

When a scientific experiment is to be performed, the assumptions about underlying mechanisms which will produce the observed outcome can take many forms. The experiment may be performed to test the validity of some specific null hypothesis about the mechanism, in which case a test of significance is used. Or, it may be simply to provide (further) information about the mechanism, in which case a likelihood analysis is applicable. This likelihood analysis may provide the basis for some null hypothesis which can be tested by further experimentation. But, even when a null hypothesis is available, further exploration of the underlying mechanism by means of the LF is always useful to gain maximum utility from the data.

Although likelihood analysis is a relative criterion of plausibility, an absolute basis for comparison always exists, the (multinomial) SM making the observed data most probable. In contrast, a test of significance is always an absolute measure of implausibility in itself. Only in certain special cases, when independently distributed sufficient statistics exist, as described in section 1, can more than one test be applied for a given set of data. Thus, if two PM's are suspected of being plausible for an experiment, such as the Poisson and geometric of chapter IV, example 3.3, a combined null hypothesis that the correct SM is Poisson or geometric does not make sense. One SM is chosen for H_o and the appropriate (conditional) test of significance applied to give an implausibility measure for it. The plausibility of the other SM may then be compared to this by means of the likelihood ratio as in chapter IV. The same reasoning applies to parameter values. A null hypothesis will not state that $\phi = \phi_1$ or $\phi = \phi_2$. One value is chosen, as above for SM's. In the special case where H_o is $\phi_1 = \phi_{1o}$ and $\phi_2 = \phi_{2o}$ and independent sufficient statistics exist for ϕ_1 and ϕ_2, the two parts of H_o may be tested independently by a conditional argument using different sample subspaces.

From the above discussion, initial assumptions may fall within several levels of complexity. They may specify two or more SSM's from which the most plausible is to be selected. They may allow any parameter values within a specified SM, perhaps with a null hypothesis about one value. Likelihood analysis may allow this to be reduced to the preceding situation for the next experiment. Or they may allow several SM's, perhaps with a null hypothesis about one. Likelihood analysis may then reveal parameter values within a selected plausible SM for analysis as in the preceding situations for a future experiment. Thus, statistical analysis of scientific experiments becomes a succession of steps, eliminating implausible null hypotheses and searching for plausible mechanisms illuminated by the available data.

APPENDIX I

REFERENCES

Alderdice, D.F. and C.R. Forrester (1968) Some effects of salinity and temperature on early development and survival of the English sole (Parophrys vetulus). J. Fish. Res. Bd. Canada 25: 495-521.

Anscombe, F.J. (1961) Estimating a mixed-exponential response law. J. Amer. Statist. Assoc. 56: 493-502.

Atkinson, A.C. (1970) A method of discriminating between models. J. Roy. Statist. Soc. B 32:

Bardwell, G.E. and E.L. Crow (1964) A two-parameter family of hyper-Poisson distributions. J. Amer. Statist. Assoc. 59: 133-141.

Barnard, G.A., G.M. Jenkens, and C.B. Winsten (1962) Likelihood inference and time series. J. Roy. Statist. Soc. A 125: 321-372.

Bartlett, M.S. (1936a) The square root transformation in analysis of variance. J. Roy. Statist. Soc. Suppl. 3:68-78.

Bartlett, M.S. (1936b) Some notes on insecticide tests in the laboratory and in the field. J. Roy. Statist. Soc. Suppl. 3: 185-194.

Birnbaum, A. (1962) On the foundations of statistical inference. J. Amer. Statist. Assoc. 57: 1-38.

Bliss, C.I. (1967) Statistics in Biology. McGraw-Hill Co. 558+xiv pp.

Blum, H.F. (1959) Carcinogenesis by Ultraviolet Light. Princeton University Press.

Box, G.E.P. (1954) The exploration and exploitation of response surfaces: some general considerations and examples. Biometrics 10: 16-60.

Box, G.E.P. and D.R. Cox (1964) An analysis of transformations. J. Roy. Statist. Soc. B 26: 211-252.

Box, G.E.P. and J.S. Hunter (1954) A confidence region for the solution of a set of simultaneous equations with an application to experimental design. Biometrika 41: 190-199.

Box, G.E.P. and W.G. Hunter (1965) The experimental study of physical mechanisms. Technometrics 7: 23-42.

Box, G.E.P. and P.W. Tidwell (1962) Transformation of the independent variable. Technometrics 4: 531-550.

Box, G.E.P. and K. Wilson (1951) On the experimental attainment of optimum conditions. J. Roy. Statist. Soc. B 13: 1-45.

Campbell, F.L. (1926) Speed of toxic action of arsenic in the silkworm. J. Gen. Physiol. 9: 433-443.

Cox, D.R. (1961) Tests of separate families of hypotheses. Proc. 4th Berkeley Symp. 1: 105-123.

Cox, D.R. (1962) Further results on tests of separate families of hypotheses. J. Roy. Statist. Soc. B 24: 406-424.

Cox, D.R. (1966) A simple example of a comparison involving quantal data. Biometrika 53: 215-220.

Cox, D.R. (1970) The Analysis of Binary Data. Methuen. 142+viii pp.

Cox, D.R. and L. Brandwood (1959) On a discriminatory problem connected with the works of Plato. J. Roy. Statist. Soc. B 21: 195-200.

Davies, O.L. (ed. 1956) Design and Analysis of Industrial Experiments. Oliver & Boyd. 636+xiii pp.

Dolby, J.L. (1963) A quick method for choosing a transformation. Technometrics 5: 317-325.

Draper, N.R. and W.G. Hunter (1969) Transformations: some examples revisited. Technometrics 11: 23-40.

Draper, N.R. and H. Smith (1966) Applied Regression Analysis. John Wiley and Sons. 407+ix pp.

Fiegl, P. and M. Zelen (1965) Estimation of exponential survival probabilities with concomitant observations. Biometrics 21: 826-838.

Fisher, R.A. (1958, 13th ed.) Statistical Methods for Research Workers. Oliver & Boyd. 356+xvi pp.

Fisher, R.A. (1959) Statistical Methods and Scientific Inference. Oliver & Boyd. 178 pp.

Fraser, D.A.S. (1967) Data transformations and the linear model. Ann. Math. Statist. 38: 1456-1465.

Fry, F.E.J. and J.S. Hart (1948) Cruising speed of goldfish in relation to water temperature. J. Fish. Res. Bd. Canada 7: 169-175.

Gart, J.J. (1969) An exact test for comparing matched proportions in crossover designs. Biometrika 56: 75-80.

Gart, J.J. (1970a) The comparison of proportions: a review of significance tests, confidence intervals, and adjustments for stratification. Proc. 7th Biometrics Conf. Hanover.

Gart, J.J. (1970b) Point and interval estimation of the common odds ratio in the combination of 2x2 tables with fixed marginals. Biometrika 57: 471-475.

Ipsen, J. (1949) Biometric analysis of graded response with incomplete measurements in assay of analgetic drugs. Acta Pharmacol. 5: 321-346.

Jackson, O.A.Y. (1969) Fitting a gamma or log normal distribution to fibre-diameter measurements on wool tops. J. Roy. Statist. Soc. C 18: 70-75.

Kalbfleisch, J.D. and D.A. Sprott (1970) Applications of likelihood methods to models involving large numbers of parameters. J. Roy. Statist. Soc. B 32: 175-208.

Lehman, E.L. (1959) Testing Statistical Hypotheses. John Wiley and Sons. 369 pp.

Lindley, D.V. (1961) The use of prior probability distributions in statistical inference. Proc. 4th Berkeley Symp. 1: 453-468.

Lindsey, J.K. (1968) Program for the analysis of nonlinear response surfaces. Fish. Res. Bd. Canada Tech. Report 87: 1-94.

Lindsey, J.K. (1970) Exact statistical inferences about the parameter for an exponential growth curve following a Poisson distribution. J. Fish. Res. Bd. Canada 27: 172-174.

Lindsey, J.K., D.F. Alderdice, and L.V. Pienaar (1970) The analysis of nonlinear models: the nonlinear response surface. J. Fish. Res. Bd. Canada 27: 765-791.

Monfort, F. (1964) Fitting gamma distributions to fibre-diameter measurements on wool-tops. J. Text. Inst. 55: T185-T191.

Rasch, G. (1961) On general laws and the meaning of measurement in psychology. Proc. 4th Berkeley Symp. 4: 321-333.

Robson, D.S. (1959) A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. Biometrics 15: 187-191.

Snedecor, G.W. and W.G. Cochran (1967, 6th ed.) Statistical Methods. Iowa State University Press. 593+xiv pp.

Sprott, D.A. (1968) Some exact methods of inference applied to discrete distributions. Proc. AAAS, Dallas.

Sprott, D.A. and J.D. Kalbfleisch (1969) Examples of likelihoods and comparisons with point estimates and large sample approximations. J. Amer. Statist. Assoc. 64: 468-484.

"Student" (1907) On the error of counting with a haemacytometer. Biometrika 5: 351-360.

Wilks, S.S. (1962) Mathematical Statistics. John Wiley and Sons. 644+xvi pp.

Williams, E.J. (1962) Exact fiducial limits in nonlinear estimation. J. Roy. Statist. Soc. B 24: 125-139.

APPENDIX II

TABLES

	SS	d.f.	F	RL
Treatments	$SS_6=\sum(\bar{y}_{i.}^{(Y)}-\bar{y}_{..}^{(Y)})^2$	I-1	MSS_6/MSS_2*	$(SS_1/SS_2)^{-N/2}$
Regression	$SS_5=SS_6-SS_3$	K-1	MSS_5/MSS_4	$(SS_1/SS_4)^{-N/2}$
Residual	$SS_4=SS_2+SS_3$	N-K		
Lack of Fit	$SS_3=\sum(\bar{y}_{i.}^{(Y)}-\hat{y}_{i}^{(Y)})^2$†	I-K	MSS_3/MSS_2	$(SS_4/SS_2)^{-N/2}$
Pure Error	$SS_2=\sum(\bar{y}_{ij}^{(Y)}-\bar{y}_{i.}^{(Y)})^2$	N-I		
Total	$SS_1=\sum(y_{ij}^{(Y)}-\bar{y}_{..}^{(Y)})^2$	N-1		

*$MSS_6 = SS_6/(I-1)$ etc.

†$\hat{y}_i$: value of y_{ij} predicted by equation (2.3).

Table 2.1 Analysis of variance for MM (2.3) with N observations, I points in the factor space (treatments), $\underline{\beta}$ a vector of K elements, and $\alpha_k=1$ (all k). The subscript j denotes replications.

y ft./min.	x_1 °C	x_2 °C
44	5	5
58	5	10
68	5	15
68	5	20
41	5	25
57	10	10
44	15	5
71	15	15
79	15	20
77	15	25
100	20	20
55	25	15
79	25	20
96	25	25
100	25	30
68	25	35
98	30	30
30	35	20
58	35	25
70	35	30
76	35	35
84	35	38

Table 2.2 Cruising speeds, y, of goldfish for 22 pairs of acclimation (x_1) and test (x_2) temperatures. Fry and Hart (1948).

	SS	df	MSS	F	RL
Regression	7003.62	5	1400.72	19.14	5.21×10^{-10}
Linear	2918.64	2	1459.32	19.94	1.06×10^{-6}
x_1 linear	163.50	1	163.50	2.23	2.37×10^{-1}
x_2 linear	2246.67	1	2246.67	30.69	7.64×10^{-6}
Quadratic	4790.80	2	2395.40	32.73	1.68×10^{-8}
x_1 quad.	4310.92	1	4310.92	58.90	4.22×10^{-8}
x_2 quad.	2202.10	1	2202.10	30.08	8.82×10^{-6}
x_1x_2 inter.	2574.30	1	2574.30	37.63	1.67×10^{-6}
Residual	1170.96	16	73.18		
Total	8174.59	21			

Table 2.3 Analysis of variance for the data of Table 2.2 according to MM(2.3) with $\alpha_k=1$ (all k) and $\gamma=1$.

	SS	df	MSS	Approx F	RL
Regression	8514.18	5	1702.84	54.36	1.79×10^{-15}
Linear	1147.44	2	573.72	18.31	3.98×10^{-7}
$x_1^{\alpha_1}$ linear	805.85	1	805.85	25.72	6.10×10^{-6}
$x_2^{\alpha_2}$ linear	1041.85	1	1041.85	33.26	8.63×10^{-7}
Quadratic	6005.69	2	3002.84	95.86	6.77×10^{-14}
$x_1^{\alpha_1}$ quad.	5141.46	1	5141.46	164.13	3.33×10^{-13}
$x_2^{\alpha_2}$ quad.	3280.43	1	3280.43	104.72	2.98×10^{-11}
$x_1^{\alpha_1}x_2^{\alpha_2}$ inter.	4041.49	1	4041.49	129.02	3.78×10^{-12}
Residual	407.21	13	31.32		
Total	8921.42	21			

Table 2.4 Analysis of variance for the data of Table 2.2 according to MM(2.3) with $\hat{\alpha}_1=1.63$, $\hat{\alpha}_2=1.29$, and $\hat{\gamma}=0.11$.

	Observed	No Linkage	Linkage
Starchy green	n_1 = 1997	2159.4375	1953.775
Starchy white	n_2 = 906	719.8125	925.475
Sugary green	n_3 = 904	719.8125	925.475
Sugary white	n_4 = 32	239.9375	34.275
$\tilde{\phi}$			0.03571
χ^2		287.69	2.0154
log R_F	log L_M=-4069.297	-198.344	-1.023

Table 4.1 Analysis of linkage in the progeny of self-fertilized heterozygote maize. Fisher (1958, p.299).

Republic	Laws	Critias	Philebus	Politicus	Sophist	Timaeus
1.1	2.4	3.3	2.5	1.7	2.8	2.4
1.6	3.8	2.0	2.8	2.5	3.6	3.9
1.7	1.9	2.0	2.1	3.1	3.4	6.0
1.9	2.6	1.3	2.6	2.6	2.6	1.8
2.1	3.0	6.7	4.0	3.3	2.4	3.4
2.0	3.8	4.0	4.8	2.9	2.5	3.5
2.1	2.7	3.3	4.3	3.3	3.3	3.4
2.2	1.8	2.0	1.5	2.3	4.0	3.4
2.8	0.6	1.3	0.7	0.4	2.1	1.7
4.6	8.8	6.0	6.5	4.0	2.3	3.3
3.3	3.4	2.7	6.7	5.3	3.3	3.4
2.6	1.0	2.7	0.6	0.9	1.6	2.2
4.6	1.1	2.0	0.7	1.0	3.0	2.7
2.6	1.5	2.7	3.1	3.1	3.0	3.0
4.4	3.0	3.3	1.9	3.0	3.0	2.2
2.5	5.7	6.7	5.4	4.4	5.1	3.9
2.9	4.2	2.7	5.5	6.9	5.2	3.0
3.0	1.4	2.0	0.7	2.7	2.6	3.3
3.4	1.0	0.7	0.4	0.7	2.3	3.3
2.0	2.3	2.0	1.2	3.4	3.7	3.3
6.4	2.4	1.3	2.8	1.8	2.1	3.0
4.2	0.6	4.7	0.7	0.8	3.0	2.8
2.8	2.9	1.3	2.6	4.6	3.4	3.0
4.2	1.2	2.7	1.3	1.0	1.3	3.3
4.8	8.2	5.3	5.3	4.5	4.6	3.0
2.4	1.9	3.3	3.3	2.5	2.5	2.2
3.5	4.1	2.0	3.3	3.8	2.9	2.4
4.0	3.7	4.7	3.3	4.9	3.5	3.0
4.1	2.1	6.0	2.3	2.1	4.1	6.4
4.1	8.8	2.0	9.0	6.8	4.7	3.8
2.0	3.0	3.3	2.9	2.9	2.6	2.2
4.2	5.2	4.0	4.9	7.3	3.4	1.8
3778	3783	150	958	770	919	762

Table 4.2 Percentage distribution of 32 possible sentence endings in seven works of Plato. The number of sentences in the work is given at the bottom of the column. Cox and Brandwood (1959).

	R	L	C	F	P	S	T
Log L	-12839	-12328	-501	-3117	-2551	-3145	-2616
Log R_R	0.0	-821.2	-26.4	-242.0	-158.0	-106.4	-94.6
Log R_L	-1004	0.0	-31.5	-53.0	-59.4	-146.2	-183.8
Log $R_{RL'}$	1003.6	-821.2	5.1	-189.0	-98.6	39.9	89.3
Log $R(\lambda_i)$			-15.9	-49.5	-45.8	-28.5	-34.3
Log $R(\lambda)$			-19.8	-53.4	-51.7	-68.9	-81.3

Table 4.3 Analysis of the data of Table 4.2 for ordering the works of Plato.

y_i	0	1	2	3	4
f_i	12	11	6	1	0

Table 4.4 Frequency distribution of observations generated from a Poisson distribution of mean 0.8. Cox (1962).

y_i	0	1	2	3	4	5	6
f_i	213	128	37	18	3	1	0

Table 4.5 Frequency distribution of yeast cells in 400 equal-sized squares of a haemacytometer. "Student" (1907).

	A	B	C	D	E	F	G	H
Log L_M	-1236	-1341	-1481	-1569	-1223	-1270	-1268	-1328
Log R_G	-8.0	-5.5	-16.2	-6.7	-13.3	-13.9	-21.1	-27.4
Log R_{LN}	-7.3	-7.7	-9.2	-11.1	-16.0	-15.8	-23.9	-41.2
Log R_{LNG}	0.73	-2.15	6.95	-4.46	-2.69	-1.80	-2.84	-13.8
Log $\tilde{R}_{LNG}$	3.37	-3.61	9.29	-3.13	-2.87	-3.66	-1.73	-13.9

Table 4.7 Analysis of the data of Table 4.6, comparing the fit of the gamma (G) and log normal (LN) PM's.

Diameter	A	B	C	D	E	F	G	H
9-11	5	2		1				
11-13	34	13	6	9	3	2	3	1
13-15	89	39	16	16	4	2	2	4
15-17	105	61	45	31	8	7	1	3
17-19	125	96	66	42	12	11	11	4
19-21	95	110	84	58	35	20	18	9
21-23	75	99	83	67	35	21	21	9
23-25	43	73	88	77	49	46	39	17
25-27	14	38	61	65	39	47	39	24
27-29	11	31	40	54	38	40	39	32
29-31	2	22	37	58	51	41	40	15
31-33	1	12	29	38	39	36	43	24
33-35		3	18	31	44	33	43	37
35-37		1	6	19	28	28	38	43
37-39	1		8	13	22	26	23	34
39-41			5	8	13	28	18	34
41-43			2	7	7	22	13	28
43-45			2	3	6	9	24	29
45-47			1	1	6	13	6	33
47-49				1	1	3	8	20
49-51			2		5	3	9	18
51-53					3	5	1	12
53-55			1		1	3	5	10
55-57						1	4	3
57-59				1	1	1		4
59-61								1
61-63							2	
63-65						2		
65-67								1
67-69								1
Total	600	600	600	600	450	450	450	450

Table 4.6 Frequency distribution of eight wool tops with diameters measured in microns. Monfort (1964).

A (minutes)									
210	215	217	218	226	229	233	235	238	239
240	241(2)	244(2)	248	250	251	253(3)	254(2)	255	260(2)
261(2)	262	265	267	270(2)	271(3)	273(2)	275(2)	276	278
279(2)	280(3)	282	283(2)	285(2)	286	287	289	292	293(2)
294(2)	296	298(2)	299	300	309	310	314	345	366
B									
203	212	215	218	222(2)	226	228	230	233	235
237	240(2)	241	243	245	246	250	252	254	255
256	258	259	260	263	266(2)	267(4)	270(2)	271	273
274	275(2)	276	277(2)	280(2)	281(2)	283(2)	284	285(4)	286(2)
287	289	291	299	302	304	307	310	319	322
327	330	338	339						
C									
210	215(2)	218	226	228	230	233(2)	235	236	240(2)
241(2)	243	244	248	251	253	254(2)	255	256	258(2)
259	260(3)	261	267(3)	270	271(3)	273	274	275(3)	279
280(4)	282	283(2)	284	285(3)	286(2)	287	289	291	292
293(2)	294	296	298(2)	299	300	302	307	309	310(2)
314	319	330	339	345	366				

Table 4.8 Survival times of three groups of silkworm larvae when a lethal dose of sodium arsenate is applied. If more than one larva survived a given time, the number is given in parentheses. Bliss (1967), pp. 106 and 122.

	Ungrouped (1 minute)			Grouped (10 minutes)		
	A	B	C	A	B	C
Log L_M	-268.10	-272.44	-313.29	-162.52	-168.58	-195.51
Log R_{LN}	-64.85	-66.18	-73.03	-9.67	-7.82	-6.72
Log R_G	-64.94	-66.00	-73.04	-9.66	-7.65	-6.70
Log R_E	-193.62	-189.17	-215.36	-138.85	-131.65	-148.42

Table 4.9 Comparison of the log normal (LN), gamma (G), and exponential (E) PM's for the data of Table 4.8.

Dose	0.76	1.80	2.60	3.30	5.00
Log L_M	-51.07	-200.41	-232.13	-54.71	-113.79
Log R_{LN}	-60.51	-204.77	-229.65	-64.16	-130.97
Log R_G	-60.01	-205.05	-229.86	-64.39	-130.55

Table 4.11 Comparison of the log normal (LN) and gamma (G) PM's for the data of Table 4.10.

Development Time		Dose in 10^7 ergs/cm^2.				
Days	Δ_i	0.76	1.80	2.60	3.30	5.00
91.2	6.5					2
97.7	7.0				2	1
104.7	7.5		2	5		
112.2	8.0		2	2	1	1
120.2	8.6		5	9	3	4
128.8	9.3		3	8	7	3
138.1	9.8		10	10	1	13
147.9	10.6	1	7	19	3	6
158.5	11.3		16	11	2	4
169.8	12.1		7	14	1	8
181.9	13.0		17	6	2	2
194.9	14.0	2	7	7	2	4
208.9	15.0		6	3	1	
223.9	16.0	1	2	2		3
239.9	17.2	1	1			
257.1	18.3	5		1		
275.4	19.7	3		1		
295.1	21.1	2	1			
316.2	22.1	4				
338.8	24.3	2				
363.1	26.0	1				
389.1	27.8					
416.9	29.8	1				
Total		23	86	98	25	51

Table 4.10 Frequency distribution of development times of ear tumors in male mice exposed to various doses of ultraviolet radiation. Blum (1959).

Poison	Treatment			
	A	B	C	D
I	0.31 0.46 0.45 0.43	0.82 0.88 1.10 0.72	0.43 0.63 0.45 0.76	0.45 0.66 0.71 0.62
II	0.36 0.40 0.29 0.23	0.92 0.49 0.61 1.24	0.44 0.31 0.35 0.40	0.56 0.71 1.02 0.38
III	0.22 0.18 0.21 0.23	0.30 0.38 0.37 0.29	0.23 0.24 0.25 0.22	0.30 0.31 0.36 0.33

Table 5.1 Survival times of four animals under each combination of three poisons and four treatments. Box and Cox (1964).

PM	Linear Normal (A)	Transform Normal (B)	Nat Linear Expon. (K)	"Exp. Val" Expon. (L)
Overall Effect	-31.74	-47.65	-4.83	-4.83
Regression	-13.10	-24.79	-3.29	
Treatments	-18.38	-28.60	-1.41	-2.04
Linear	-1.14	-4.67	-0.26	
Quadratic	-1.97	-2.32	-0.33	
Poisons	-19.89	-37.80	-2.29	-2.73
Linear	-10.29	-20.55	-2.54	
Quadratic	-1.35	-2.80	-0.65	
Interaction	-6.52	-3.49	-0.22	-0.61
Linear	-0.00	-0.33	-0.05	
PM Lack of Fit	-124.37	-99.04	-162.39	-162.39
Av.Lack of Fit	-10.04	-8.27	-13.55	-13.55

Table 5.2 "Analysis of variance" for the data of Table 5.1 using maximized log RL's.

Rat No.	Response on April								
	10	12	14	16	18	20	23	25	27
1	7.0	7.2	9.7	6.3	8.0	10.9	9.1	5.5	6.7
	6.4	10.5	6.9	5.7	7.2	12.3	7.4	6.1	5.9
2	10.4	8.0	9.0	7.2	10.1	9.9	8.6	6.6	6.4
	10.4	7.9	7.2	7.8	6.5	9.9	7.6	6.5	7.8
3	5.7	6.4	11.0	6.0	5.5	11.8	7.0	8.1	5.6
	4.5	7.8	6.7	6.0	7.2	9.0	7.2	6.5	9.6
4	4.5	5.7	6.6	8.0	7.3	13.0	7.9	8.1	9.7
	8.1	6.8	8.6	9.1	8.8	11.4	8.0	6.7	8.5
5	8.2	9.0	6.0	7.1	5.9	7.8	8.4	5.8	6.4
	8.2	9.0	8.6	7.0	7.1	7.2	6.9	5.3	9.2
6	8.1	7.5	9.0	9.2	8.2	12.8	6.2	8.0	9.4
	10.7	7.5	9.8	5.8	8.9	10.8	6.9	7.4	7.8
7	8.2	6.2	7.5	5.2	8.6	7.4	8.7	5.0	6.2
	7.0	5.8	8.1	5.6	6.5	8.2	6.5	6.0	6.4
8	7.2	8.0	7.8	6.6	8.1	10.0	8.5	7.0	7.5
	7.6	7.1	7.8	5.2	8.0	9.8	8.3	6.6	8.7
9	9.0	9.5	6.8	5.9	6.5	9.1	7.5	6.9	7.5
	9.8	7.4	6.6	6.6	8.6	9.9	8.3	7.6	7.7
10	9.3	4.5	7.3	9.5	7.8	10.4	10.7	7.5	8.7
	10.3	8.4	10.5	9.2	8.5	10.1	9.0	7.5	7.8

Table 5.3 Tail-flip reaction times in seconds of ten white rats in two trials on each of nine different days. Ipsen (1949).

PM	Linear Normal (A)	Log Normal (C)	Nat Linear Expon. (K)	"Exp. Val" Expon. (L)
Overall Effect	-126.97	-121.72	-2.81	-2.81
Between Days	-73.90	-65.48	-1.05	-1.09
Between Rats	-34.80	-34.14	-0.45	-0.45
Interaction	-77.06	-76.34	-1.27	-1.26
PM Lack of Fit	-1333.73	-1334.43	-1666.68	-1666.68
Av Lack of Fit	-14.82	-14.83	-18.52	-18.52

Table 5.4 "Analysis of variance" for the data of Table 5.3 using maximized log RL's.

Rat No.	Trial Number				
	2	5	8	11	14
1	145	140	40	18	10
2	50	45	16	10	8
3	300	180	35	22	20
4	240	45	85	70	30
5	165	95	35	50	18
6	110	83	45	48	23
7	305	59	54	40	20
8	345	100	110	55	35
9	422	253	75	73	20
10	32	50	8	7	6

Table 5.5 Time in seconds required by ten rats to run a maze on five different trials. Bliss (1967) p. 327.

MM	No Interaction (5.6)		Interaction (5.5)	
PM	Linear Normal (A)	Log Normal (C)	Nat Linear Expon. (K)	"Exp. Val" Expon. (L)
Overall Effect	-35.20	-55.71	-26.67	-26.67
Between Trials	-29.89	-47.37	-93.28	-17.07
Between Rats	-14.44	-32.26	-5.03	-7.01
Interaction			→ $-\infty$ *	→ $-\infty$ *
Lik. Lack of Fit	-263.30	-214.42	-246.04	-246.04
Av. Likelihood	-5.26	-4.29	-4.92	-4.92

*No convergence; log R approaches negative infinity.

Table 5.6 "Analysis of variance" for the data of Table 5.5 using maximized log RL's.

Block	Control		Toxic Emulsion			
	1	2	3	4	5	6
I	33	30	8	12	6	17
	59	36	11	17	10	8
II	36	23	15	6	4	3
	24	23	20	4	7	2
III	19	42	10	12	4	6
	27	39	7	10	12	3
IV	71	39	17	5	5	1
	49	20	26	8	5	1
V	22	42	14	12	2	2
	27	22	11	12	6	5
VI	84	23	22	16	17	6
	50	37	30	4	11	5

Table 5.7 Counts of numbers of leatherjackets surviving on two one sq. ft. areas in each of 36 one sq. yd. plots divided into six blocks with two control and four sprayed plots per block. Bartlett (1936b).

PM	Linear Normal (A)	Log Normal (C)	Sq. Root Normal (D)	Poisson (M)
Overall Effect	-81.34	-93.00	-92.07	-433.48
Between Treat.	-71.51	-84.27	-84.29	-309.71
Between Blocks	-16.12	-20.36	-18.92	-20.45
Interaction	-36.08	-40.64	-37.27	-61.19
PM Lack of Fit	-173.48	-138.28	-142.44	-142.88
Av Lack of Fit	-4.82	-3.84	-3.96	-3.97

Table 5.8 "Analysis of variance" for the data of Table 5.7 using maximized log RL's.

Treatment	Block			
	1	2	3	4
None	183	176	291	254
Superphosphate (P)	356	300	301	271
Potash (K)	224	258	244	217
P + K	329	283	308	326

Table 5.9 Numbers of surviving sugar beet plants under four fertilizer treatments in four blocks. Snedecor and Cochran (1967) p. 344.

MM	No Interaction (5.6)				Int. (5.5)
PM	Linear Normal (A)	Log Normal (C)	Sq. Root Normal (D)	Poisson (M)	Poisson (M)
Overall Effect	-8.79	-8.46	-8.64	-50.00	-77.19
Between Blocks	-1.15	-1.23	-1.18	-3.87	-4.79
Between Fert.	-8.36	-7.98	-8.20	-46.13	-48.69
Interaction					-27.19
Lik. Lack of Fit	-76.57	-77.61	-76.99	-86.54	-59.35
Av. Likelihood	-4.79	-4.86	-4.81	-5.42	-3.71

Table 5.10 "Analysis of variance" for the data of Table 5.9 using maximized log RL's.

Year	No. Caught x10^{-3}(y)	Rainfall (x_1)	Sunshine (x_2)
1930	8496	19.5	217
1932	1242	5.4	219
1934	2532	20.5	175
1936	3928	17.2	242
1938	1504	15.7	201
1940	680	11.6	210
1942	1234	26.6	291
1944	2230	6.8	153
1946	2040	29.4	233
1948	2800	17.4	210
1950	2800	24.3	231
1952	2406	13.8	234
1954	1752	10.7	193
1956	3719	11.0	172
1958	2026	10.1	262
1960	2488	18.5	178
1962	17381	22.2	354
1964	5152	17.8	278
1966	7031	19.8	293
1968	9706	13.2	301

Table 5.11 Numbers of pink salmon caught in British Columbia fisheries with rainfall and sunshine conditions on even years over a forty year period. P. Wickett, Fisheries Research Board, Nanaimo, Canada.

PM	Log Normal (C)	Poisson (M)
Regression	-8.99	-20805.00
$\beta_1 = 0$	-0.14	-59.03
$\beta_2 = 0$	-4.20	-6876.88
$\beta_3 = 0$	-1.16	-2812.49
$\beta_4 = 0$	-3.17	-2390.09
$\beta_5 = 0$	-0.72	-408.06
Lack of Fit of Reg.	-173.74	-6864.69
Lik. of PM Lack of Fit		-98.23

Table 5.12 Regression analysis for the data of Table 5.11 using maximized log RL's.

i	y_i	$n_i - y_i$	n_i
1	4	5	9
2	4	2	6

Table 5.13 A 2x2 contingency table.

Day	Experimental Babies		Control Babies	
	Not Crying	Total	Not Crying	Total
1	1	1	3	8
2	1	1	2	6
3	1	1	1	5
4	0	1	1	6
5	1	1	4	5
6	1	1	4	9
7	1	1	5	8
8	1	1	4	8
9	1	1	3	5
10	0	1	8	9
11	1	1	5	6
12	1	1	8	9
13	1	1	5	8
14	1	1	4	5
15	1	1	4	6
16	1	1	7	8
17	0	1	4	6
18	1	1	5	8

Table 5.14 Number of babies not crying after a test period of either rocking (experimental) or not on eighteen days. Cox (1966).

MM	No Interaction (5.6)	Interaction (5.5)
Overall Effect	-15.08	-23.23
Rocking	-2.33	-0.00
Between Days	-13.28	-5.10
Interaction		-8.15
Lack of Fit of MM	-8.15	0.00

Table 5.15 "Analysis of variance" for the data of Table 5.14 using the binomial PM (N) and maximized log RL's.

Treatment	Replication				
	1	2	3	4	5
Control	8/100	10/100	12/100	13/100	11/100
Arasan	2/100	6/100	7/100	11/100	5/100
Spergon	4/100	10/100	9/100	8/100	10/100
Semesan Jr.	3/100	5/100	9/100	10/100	6/100
Fermate	9/100	7/100	5/100	5/100	3/100

Table 5.16 Numbers of soybean seeds failing to emerge in five replications of five treatments in 25 plots. Snedecor and Cochran (1967) p. 300.

PM	Logit Normal (E)	Odds Normal (F)	Per Cent Normal (H)	Arcsine Normal (J)	Binomial (N)
Overall	-9.85 (-10.49)	-8.96 (-0.009)	-9.08 (-9.85)	-9.48 (-11.13)	-9.46
Replications	-3.81 (-5.93)	-3.22 (-0.003)	-3.35 (-4.62)	-3.68 (-5.81)	-3.72
Treatments	-5.92 (-6.70)	-5.25 (-0.006)	-5.37 (-7.01)	-5.73 (-7.65)	-5.75
Lack of Fit	-7.91	-8.02	-7.88	-7.66	-7.50

Table 5.17 "Analysis of variance" for the data of Table 5.16 using no-interaction MM (5.6) and maximized log RL's for binomial LF (5.3) with those for normal LF (5.1) in parentheses.

PM	Logit Normal (E)	Odds Normal (F)	Per Cent Normal (H)	Arcsine Normal (J)	Binomial (N)
Overall	-17.76	-16.98	-16.96	-17.13	-16.96
Replications	-5.29	-4.54	-4.34	-4.24	-4.37
Treatments	-7.37	-6.32	-6.21	-6.36	-6.72
Interaction	-7.91	-8.02	-7.88	-7.66	-7.50

Table 5.18 "Analysis of variance" for the data of Table 5.16 using interaction MM (5.5) and maximized log RL's for binomial LF (5.3).

Sal. $^o/_{oo}S$ (x_1)	Temp. ^oC (x_2)	Tank Number 1 Hatch	1 Total	2 Hatch	2 Total	3 Hatch	3 Total	4 Hatch	4 Total
15	4	236	666	203	724	183	764	212	723
15	8	600	656	697	747	615	746	641	703
15	12	407	566	343	603	365	560	302	394
25	4	203	717	177	782	155	852	138	590
25	8	591	621	564	640	714	754	532	570
25	12	475	622	465	645	506	608	415	532
35	4	1	738	3	655	10	742	3	743
35	8	526	616	419	467	410	484	374	606
35	12	272	362	352	478	392	590	382	459
10	10	303	681	329	710	262	611	301	700
10	6	277	757	234	681	263	647	287	801
40	10	387	450	389	553	388	564	318	604
40	6	276	662	247	542	248	527	149	591
20	10	351	391	559	650	527	603	476	548
20	6	585	643	620	671	437	497	667	771
30	10	447	491	462	530	475	545	499	556
30	10	522	573	615	680	539	581	517	561
30	6	563	666	600	704	562	656	615	723

Table 5.19 Effects of salinity and temperature on the proportion of eggs of English sole hatching. Alderdice and Forrester (1968).

PM	Logit Normal (E)	Per Cent Normal (H)	Binomial (N)
Overall Effect	-8847.16(-119.76)	-8798.68(-119.43)	-8758.89
Regression	-7433.16 (-61.56)	-7347.59 (-73.92)	-7931.69
Linear	-2774.49 (-33.38)	-5662.38 (-45.42)	-2700.97
x_1 linear	83.87 (-1.28)	58.06 (-0.32)	-3.06
x_2 linear	-2947.22 (-32.98)	-5673.77 (-45.26)	-2690.94
Quadratic	-4687.92 (-44.91)	-4845.57 (-56.94)	-4461.06
x_1 quadratic	-2838.86 (-23.57)	-1915.72 (-37.45)	-2438.41
x_2 quadratic	-2963.75 (-38.45)	-14095.26 (-46.32)	-3022.58
x_1x_2 Inter.	-269.25 (-12.63)	-364.21 (-7.48)	-317.62
MM Lack of Fit	-1414.00 (-58.21)	-1451.08 (-45.51)	-827.20
PM Lack of Fit	-12.48	-1.95	0.00
Between Tanks			-330.21

Table 5.20 "Analysis of variance" for the data of Table 5.19 using linear MM (5.9) and maximized log RL's for binomial LF (5.3) with those for normal LF (5.1) in parentheses.

PM	Logit Normal (E)	Transform Per Cent Normal (I)	Binomial (N)
Overall Effect	-8847.16(-119.76)	-8762.57(-122.81)	-8758.89
Regression	-8083.79 (-84.43)	-8249.97(-101.06)	-8381.18
Linear	-5029.55 (-61.43)	-9277.06 (-79.61)	-4371.96
$x_1^{\alpha_1}$ linear	16.46 (-0.31)	-36.46 (-4.94)	-13.49
$x_2^{\alpha_2}$ linear	-5082.42 (-61.41)	-4169.53 (-78.85)	-4297.84
Quadratic	-4115.59 (-57.05)	-5303.45 (-74.00)	-3600.82
$x_1^{\alpha_1}$ quadratic	-3061.11 (-41.98)	-2508.81 (-58.29)	-2623.97
$x_2^{\alpha_2}$ quadratic	-2150.39 (-44.51)	-13617.72 (-58.80)	-1898.36
$x_1^{\alpha_1}x_2^{\alpha_2}$ Inter.	-413.22 (-24.41)	-553.45 (-17.11)	-366.98
MM Lack of Fit	-763.37 (-35.33)	-512.60 (-21.75)	-377.71
PM Lack of Fit	-12.48	-1.53	0.00
Between Tanks			-330.21

Table 5.21 "Analysis of variance" for the data of Table 5.19 using nonlinear MM (2.3) and maximized log RL's for binomial LF (5.3) with those for normal LF (5.1) in parentheses.

MM	Linear (5.9)				
PM	Logit Normal (E)	Per Cent Normal (H)		Arcsine Normal (J)	Binomial (N)
SM Lack of Fit	-1426.48	-1453.03		-919.45	-827.20
PM Lack of Fit	-12.48	-1.95		-5.71	0.00
MM	Nonlinear (2.3)				
PM	Logit Normal (E)	Trans. Per Cent Normal (I)	Trans. Odds Normal (G)	Arcsine Normal (J)	Binomial (N)
SM Lack of Fit	-775.85	-514.13	-264.86	-279.55	-377.70
PM Lack of Fit	-12.48	-1.53	-17.61	-5.71	0.00

Table 5.22 Comparison of various SM's and PM's fitted to the data of Table 5.19 using maximized log RL's for binomial LF (5.3).

Sex-strain Stratum	Treated		Control	
	With	Total	With	Total
X males	4	16	5	79
X females	2	16	3	87
Y males	4	18	10	90
Y females	1	15	3	82

Table 6.1 Counts of mice of four sex-strains with tumors after 85 weeks either fed Avadex or used as controls. Gart (1970a).

MM	No Interaction (5.6)	Interaction (5.5)
Overall Effect	-7.14	-7.57
Between Sex-strains	-3.62	-2.98
Treatment	-3.48	-2.36
Interaction		-0.43
MM Lack of Fit	-0.43	0.00

Table 6.2 "Analysis of variance" for the data of Table 6.1 using the binomial PM (N) and maximized log RL's.

APPENDIX III FIGURES

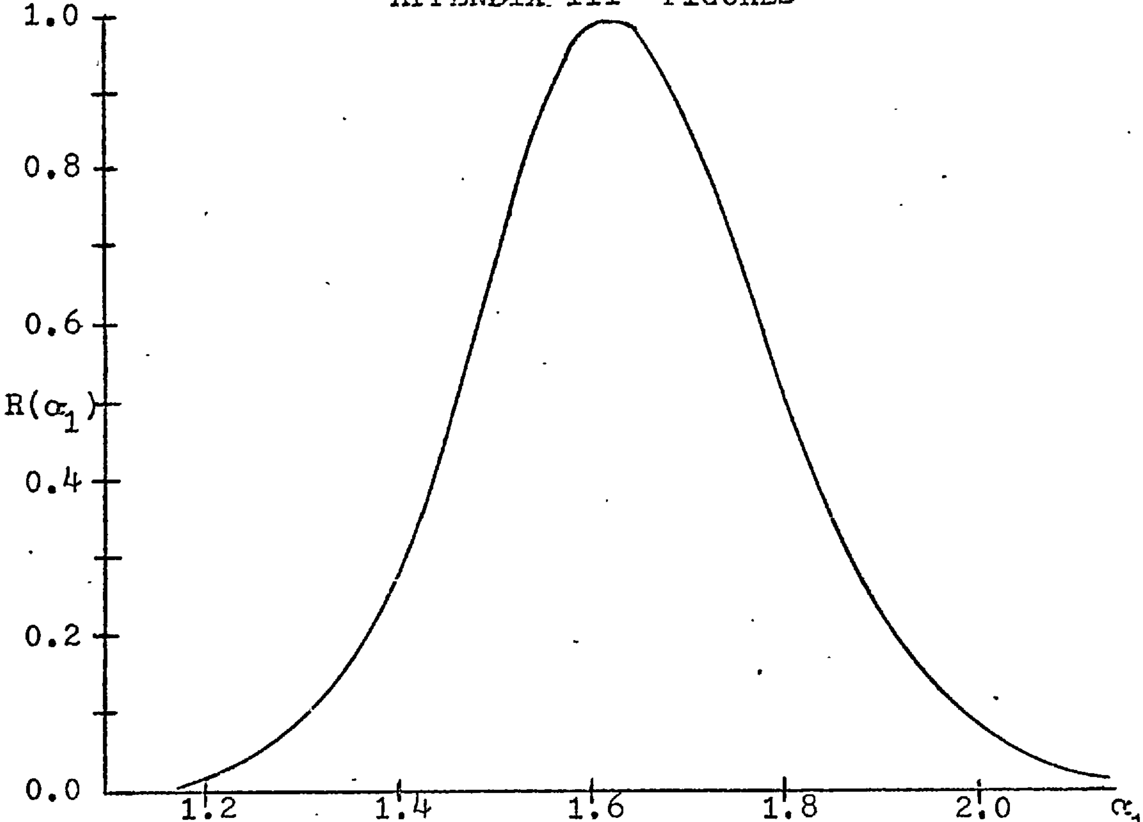

Figure 2.1 Maximized RL graph for power parameter α_1 of MM (2.3) for the data of Table 2.2.

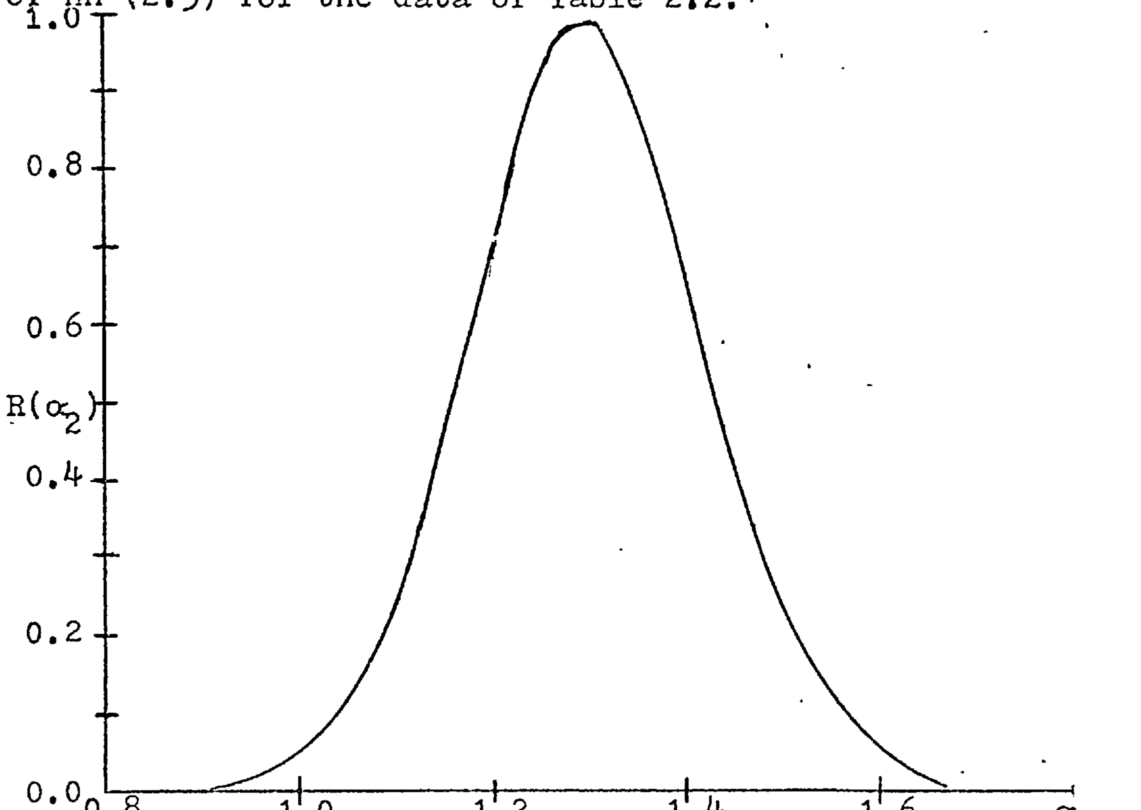

Figure 2.2 Maximized RL graph for power parameter α_2 of MM (2.3) for the data of Table 2.2.

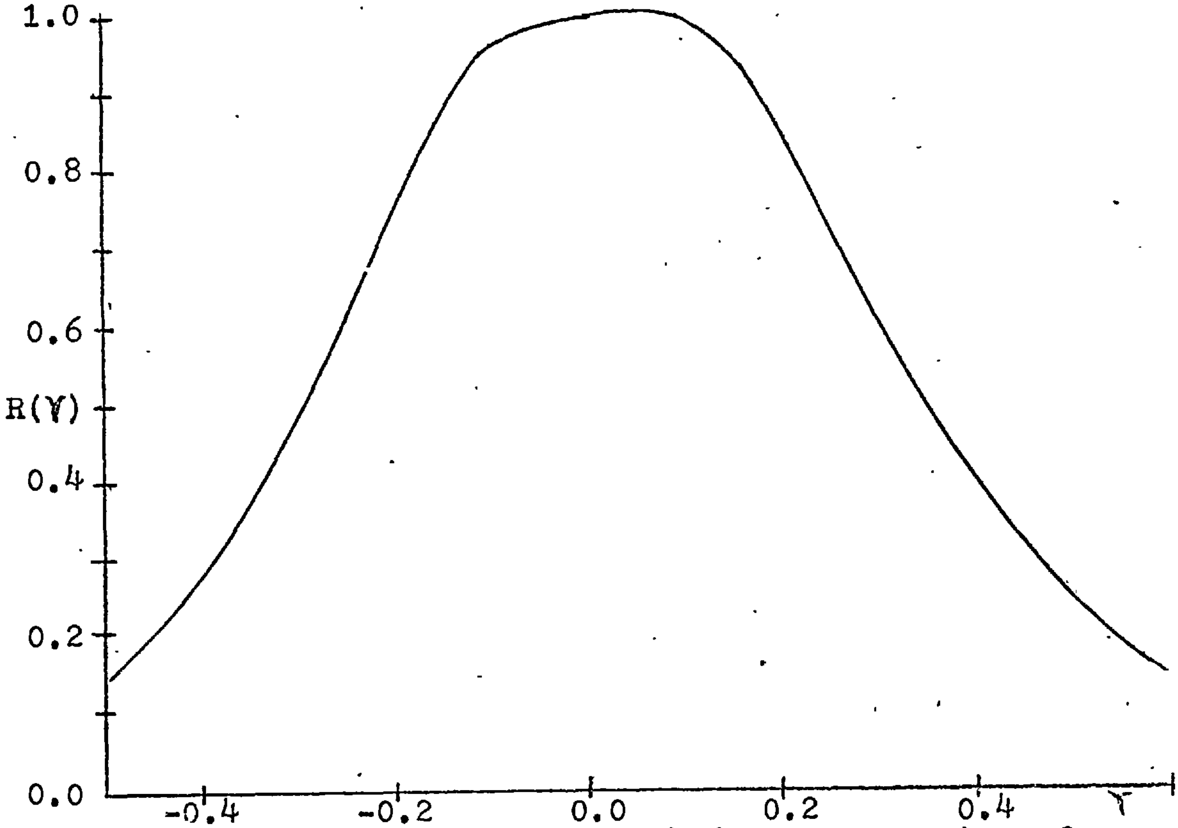

Figure 2.3 Maximized RL graph for response transformation γ of PM (2.2) for the data of Table 2.2.

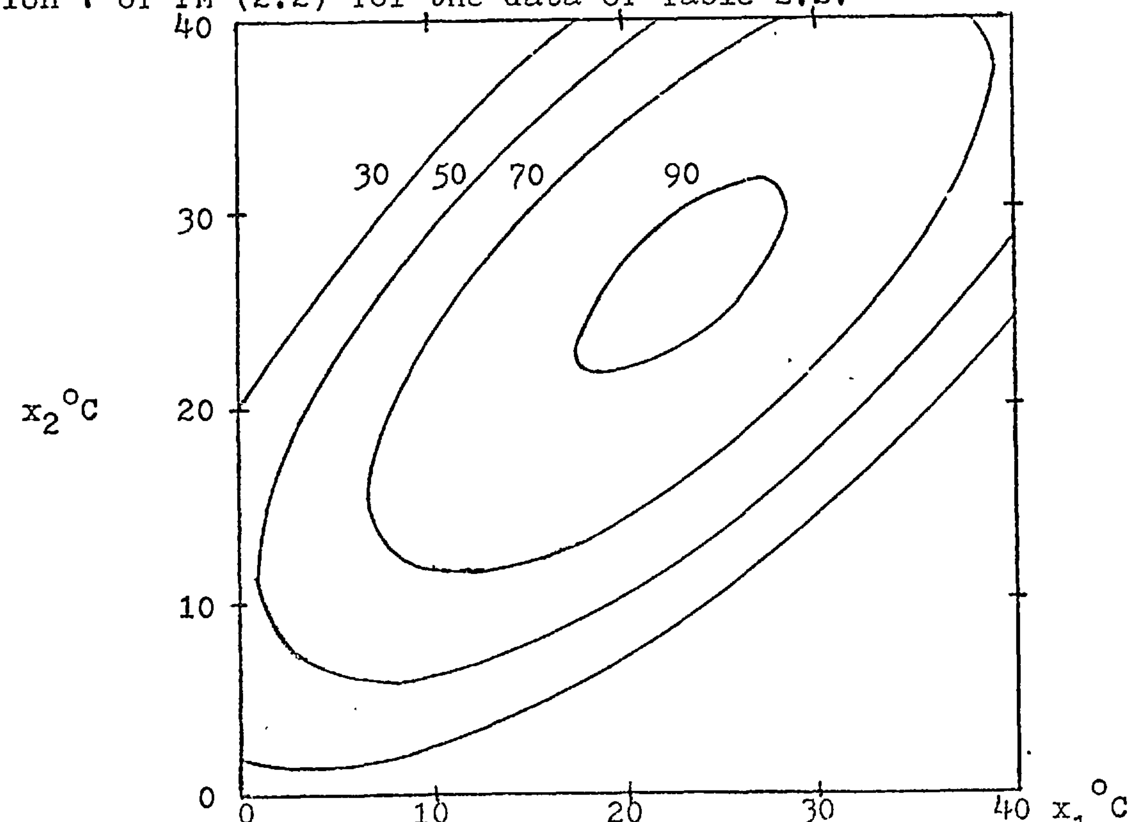

Figure 2.4 Response surface contours for cruising speed of goldfish at 20 ft/min intervals for the data of Table 2.2 using linear MM (2.3) with $\alpha_1 = \alpha_2 = \gamma = 1$.

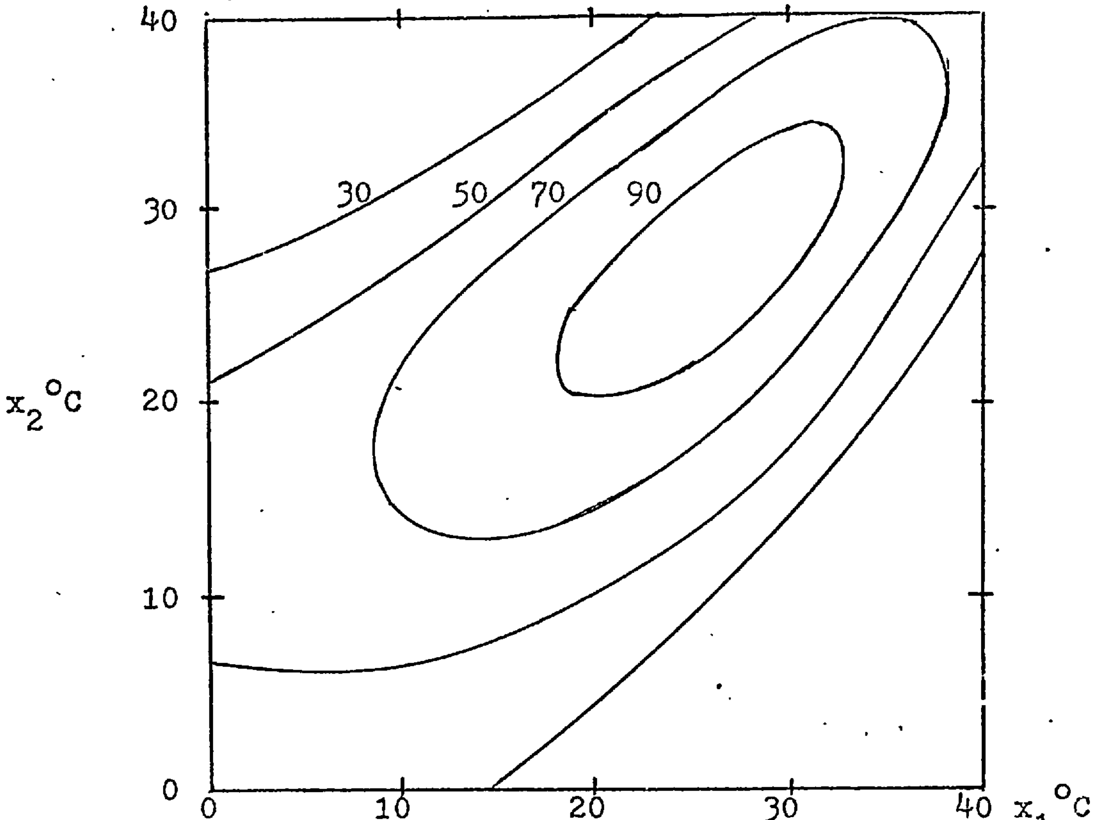

Figure 2.5 Response surface contours for cruising speed of goldfish at 20 ft/min intervals for the data of Table 2.2 using nonlinear MM (2.3).

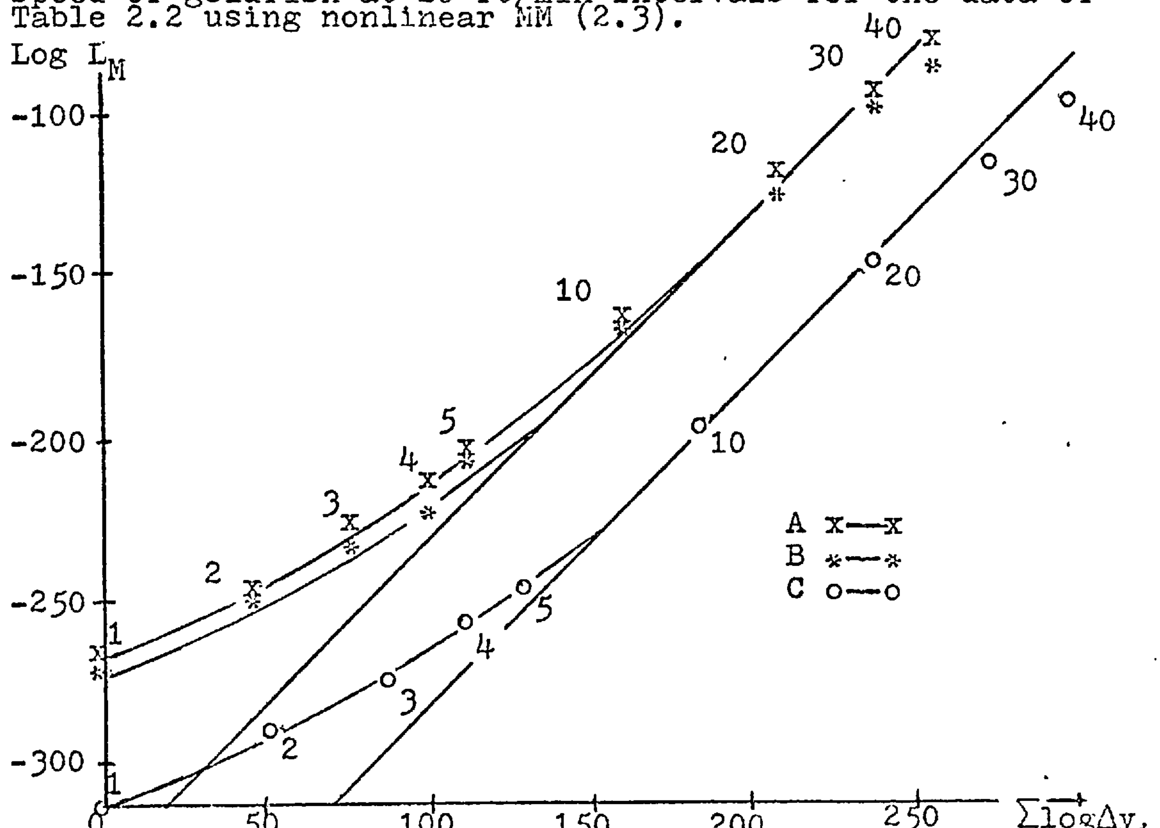

Figure 4.1 Graphs to determine optimum interval widths in minutes for the three sets of data of Table 4.8.

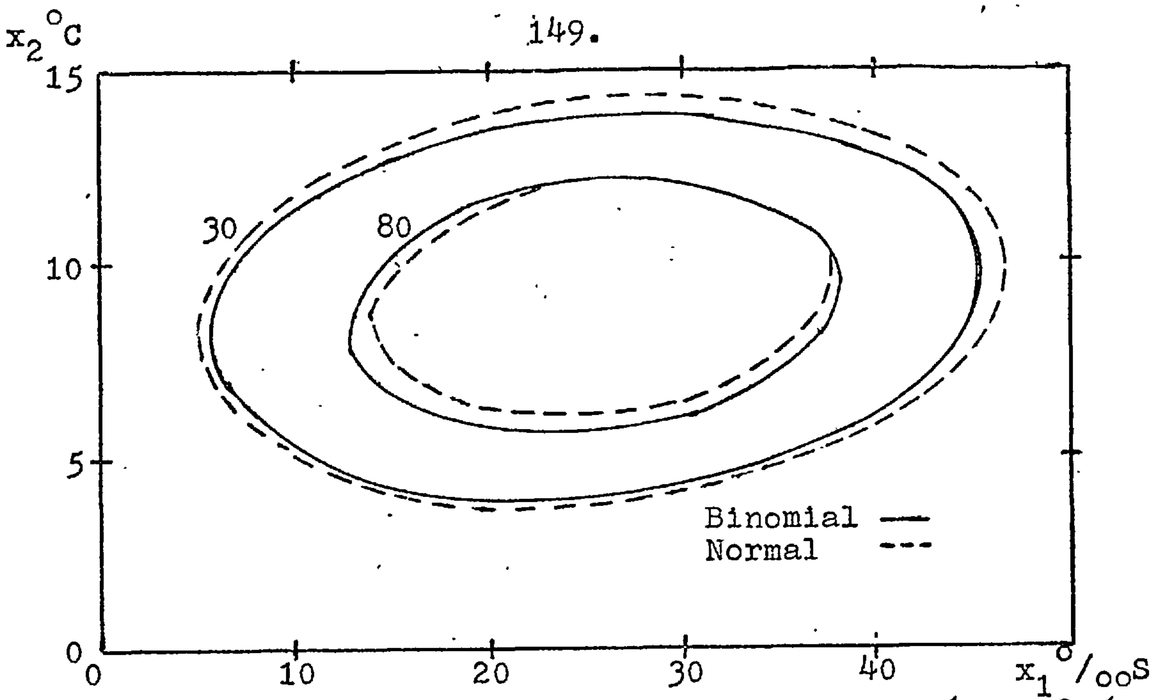

Figure 5.1 Response surface contours of 30% and 80% hatch for the linear MM (5.9) with binomial (N) and per cent normal (H) PM's for the data of Table 5.19.

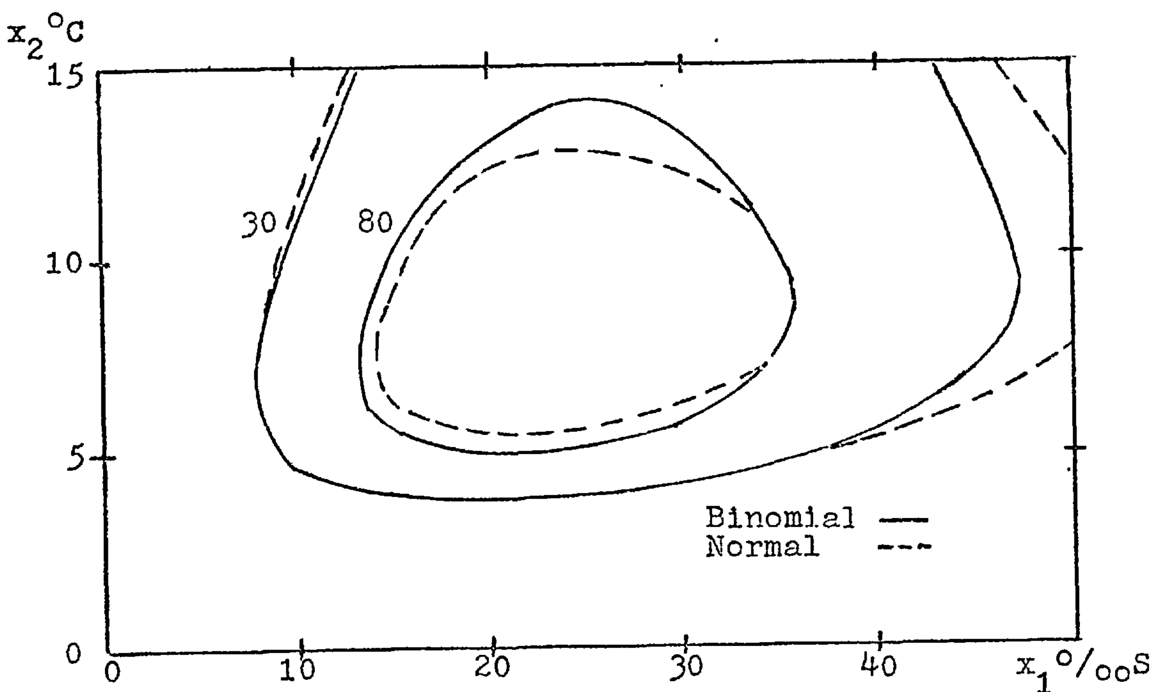

Figure 5.2 Response surface contours of 30% and 80% hatch for the nonlinear MM (2.3) with binomial (N) and transformed per cent normal (I) PM's for the data of Table 5.19.

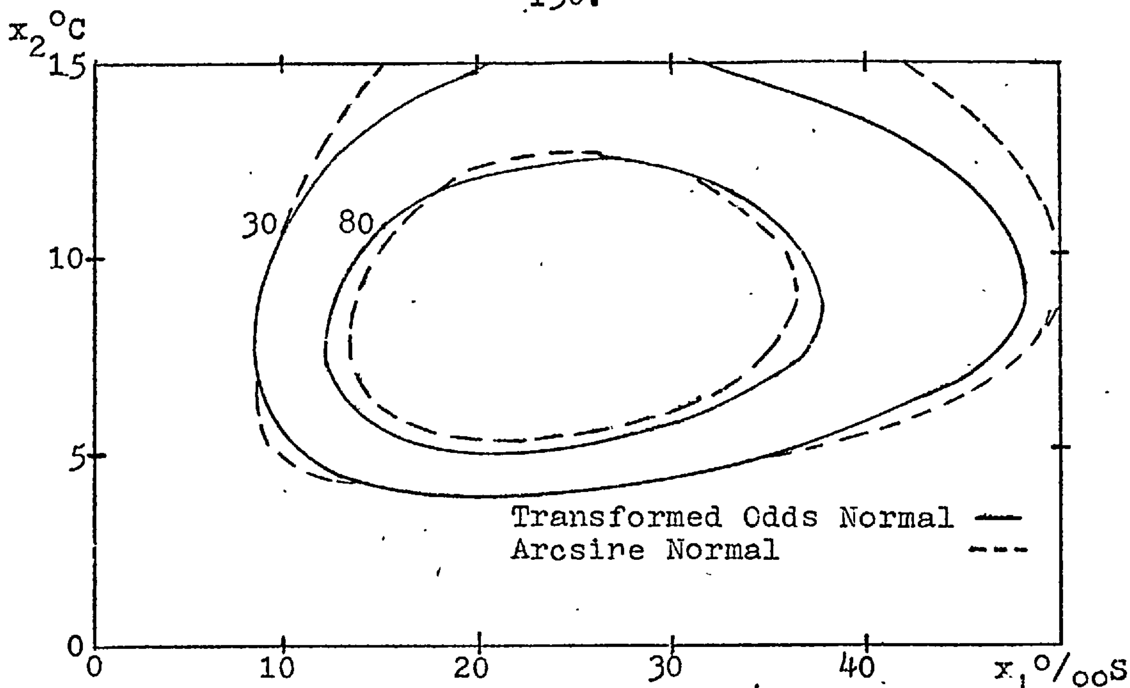

Figure 5.3 Response surface contours of 30% and 80% hatch for the nonlinear MM (2.3) with transformed odds normal (G) and arcsine normal (J) PM's for the data of Table 5.19.

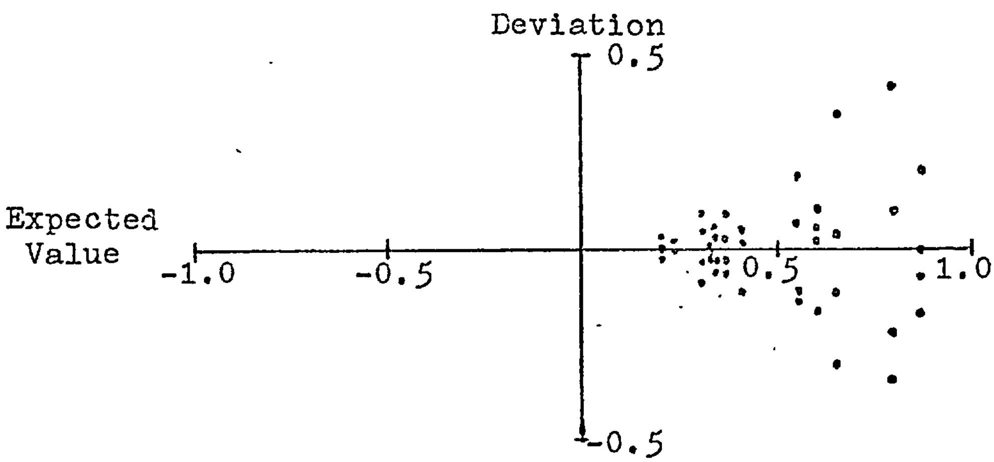

Figure 5.4 Deviations from expected values for the data of Table 5.1 using the linear normal (A) or exponential (K and L) PM's.

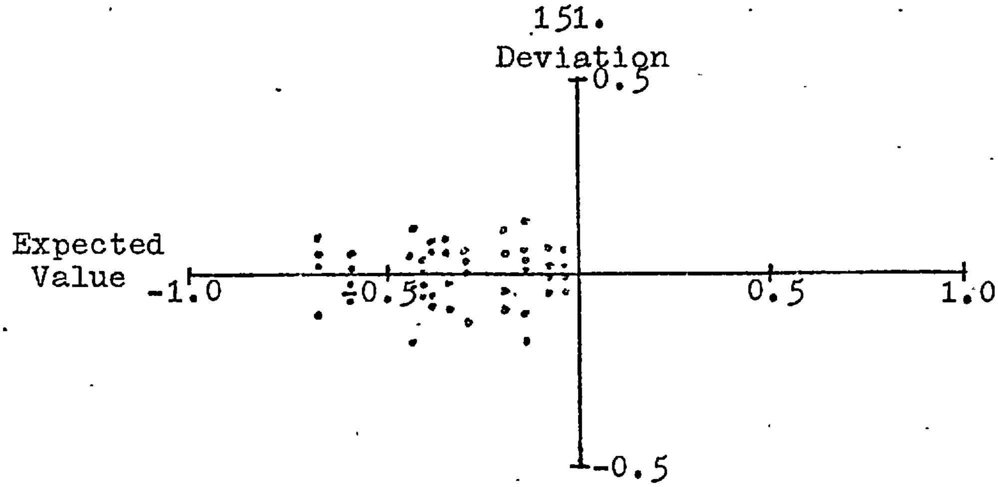

Figure 5.5 Deviations from expected values in transformed units for the data of Table 5.1 using the power transformed normal (B) PM.

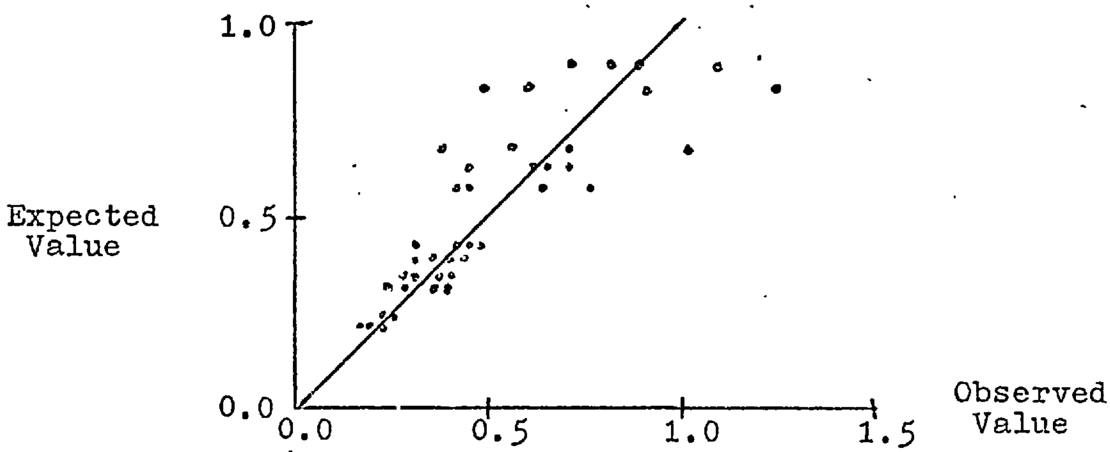

Figure 5.6 Observed and expected values for the data of Table 5.1 using the linear normal (A) or exponential (K and L) PM's.

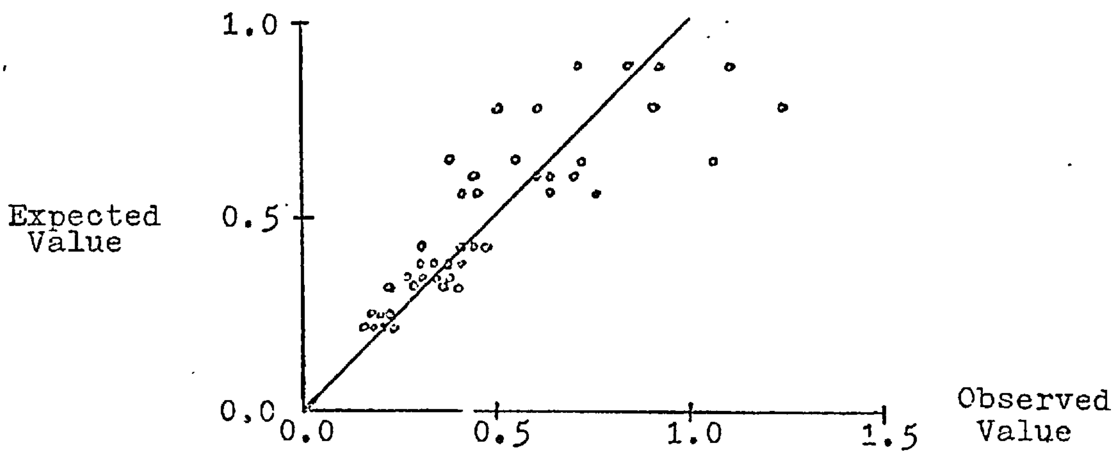

Figure 5.7 Observed and expected values for the data of Table 5.1 using the transformed normal (B) PM.

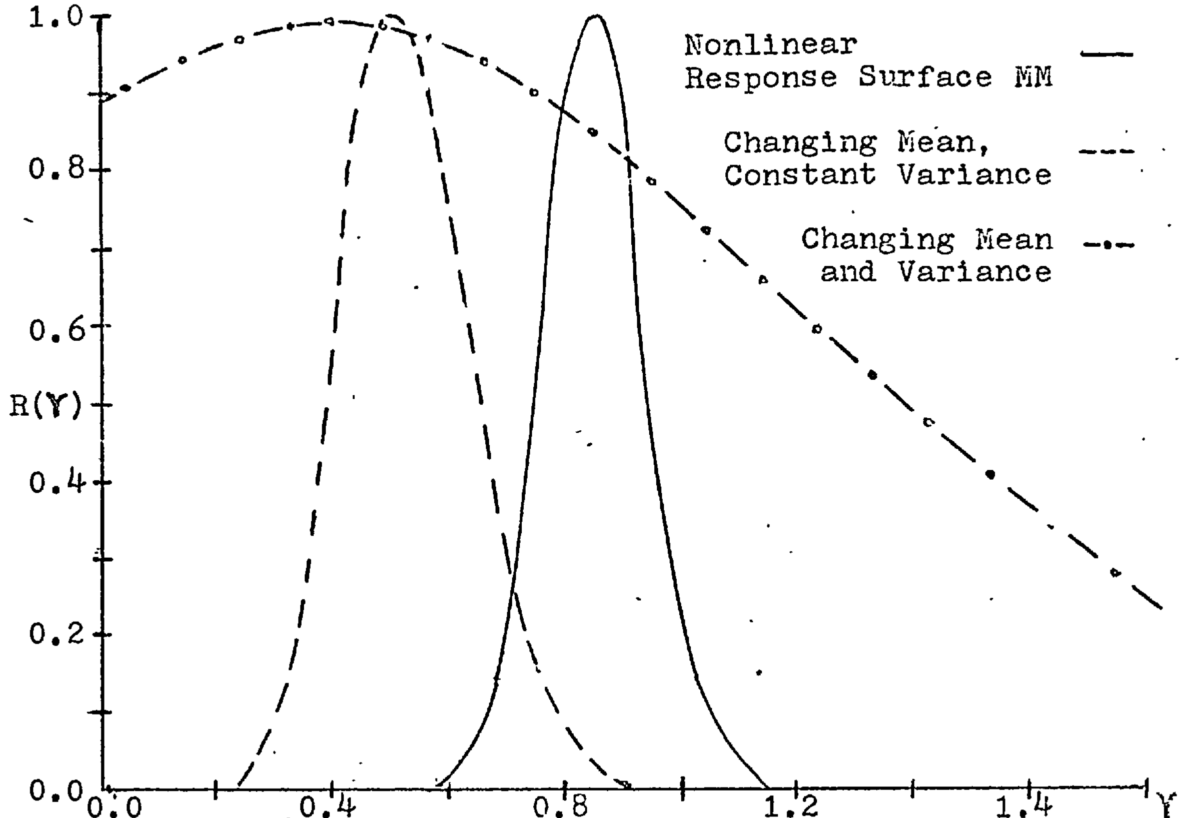

Figure 5.8 Maximized RL graphs showing the roles of the power transformation γ of the transformed per cent normal (I) PM for the data of Table 5.19.

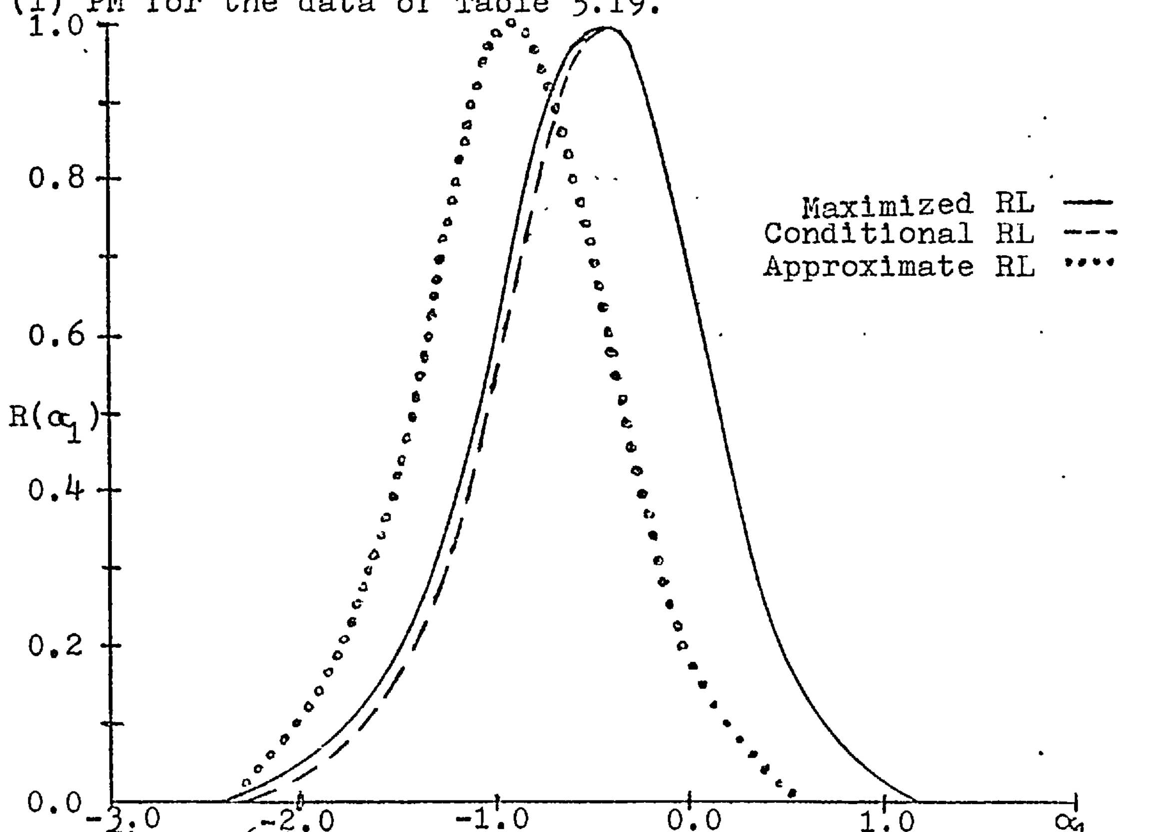

Figure 6.1 Relative likelihood graphs for parameter α_1 of MM (5.7) for the data of Table 5.13.

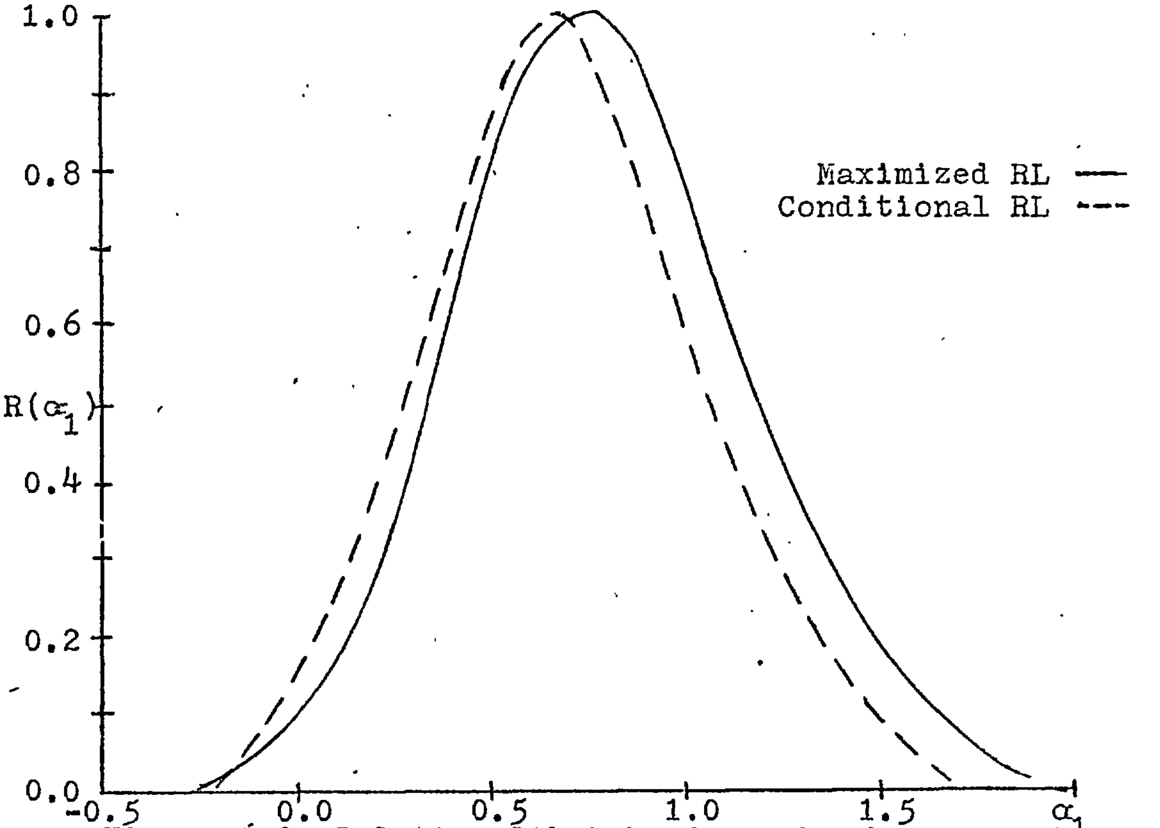

Figure 6.2 Relative likelihood graphs for parameter α_1 of MM (5.6) for the data of Table 5.14.

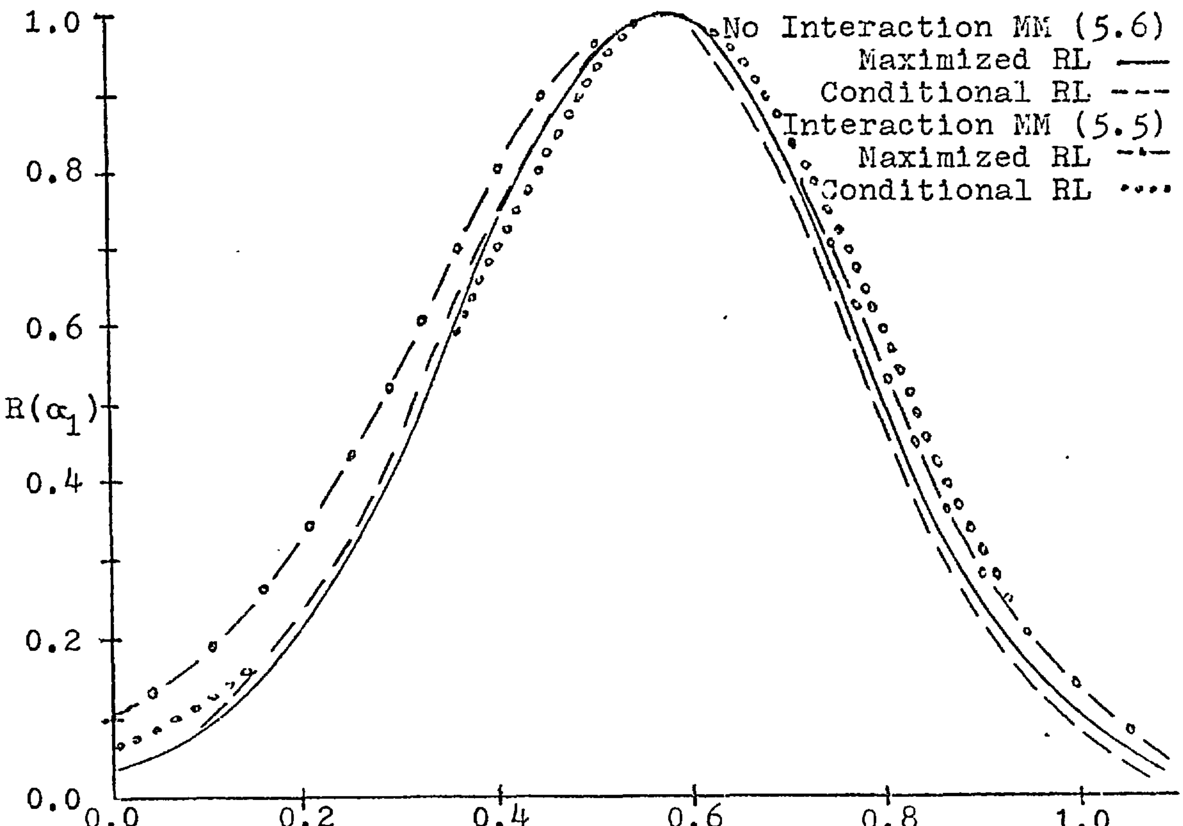

Figure 6.3 Relative likelihood graphs for parameter α_1 of MM's (5.5) and (5.6) for the data of Table 6.1.

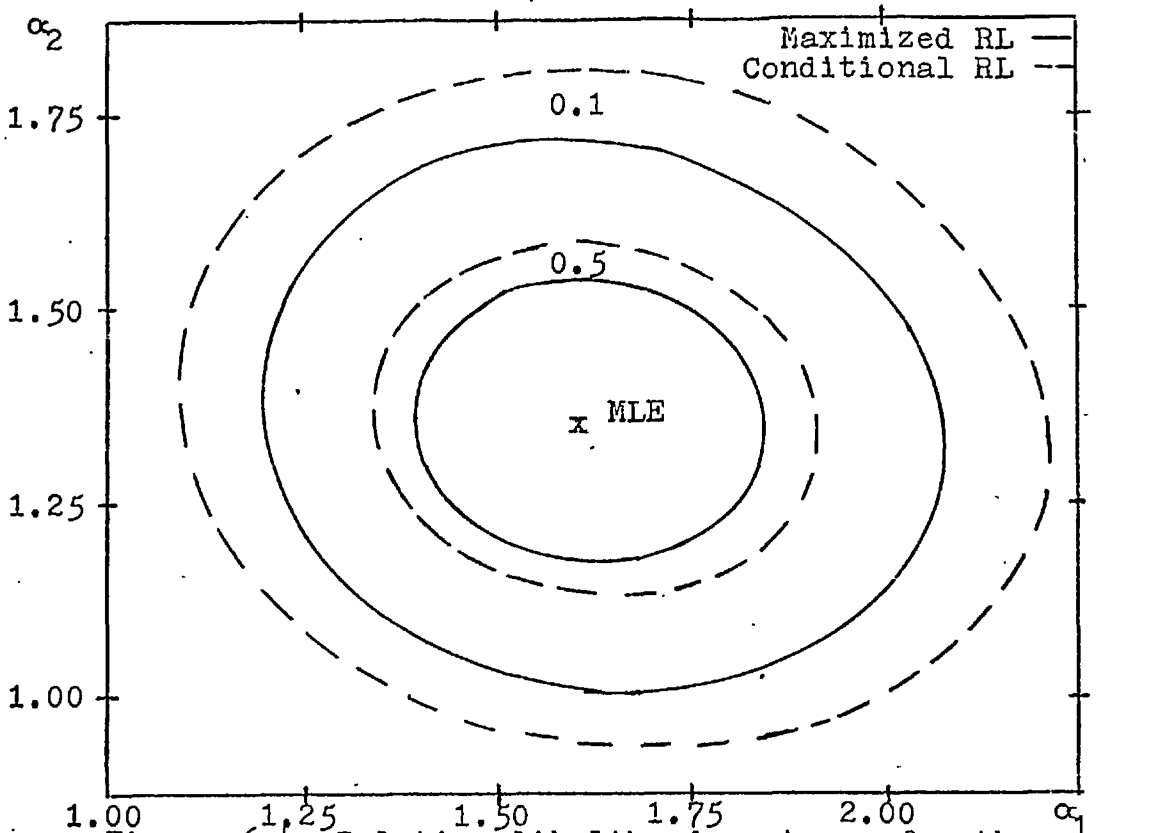

Figure 6.4 Relative likelihood contours for the power transformations α_1 and α_2 of the two factors of MM (2.3) for the data of Table 2.2 with $\gamma = 1$ in PM (2.2).

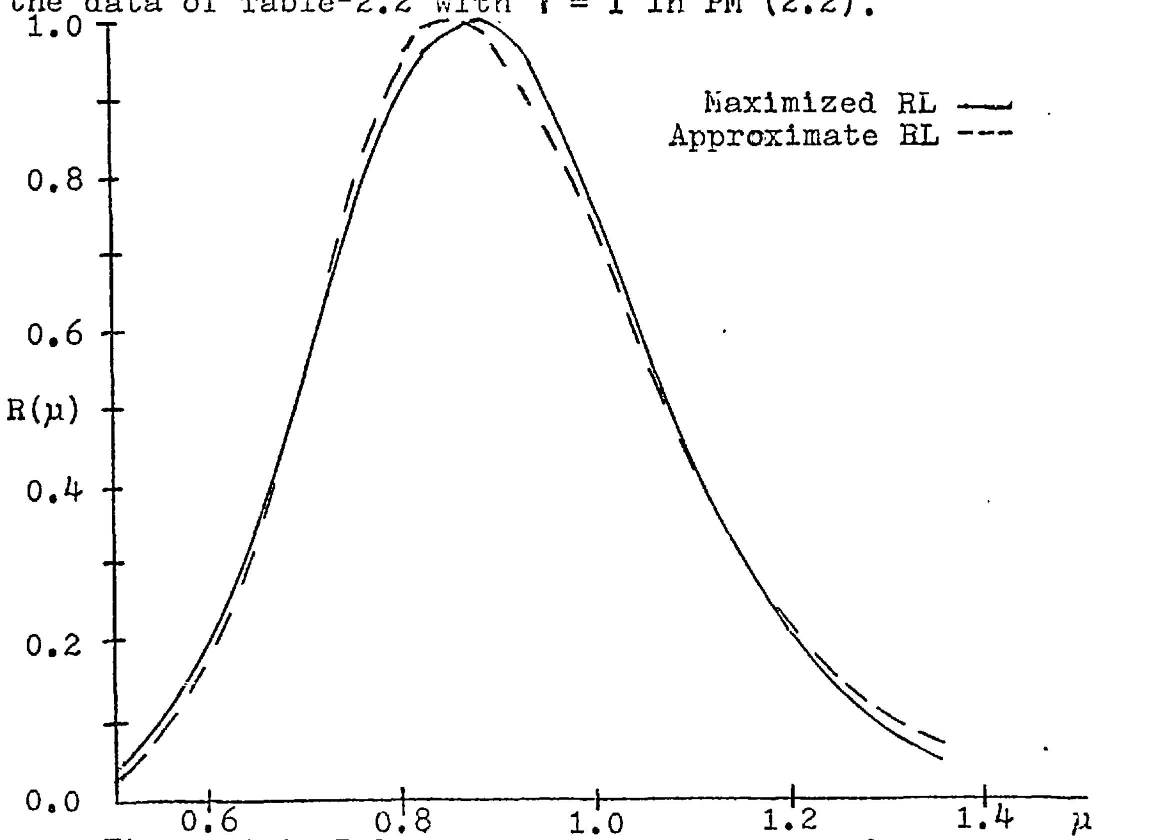

Figure 7.1 Relative likelihood graphs for parameter μ of Poisson LF (5.4) for the data of Table 4.4.