# MAXIMUM LIKELIHOOD ESTIMATION IN A LATENT VARIABLE PROBLEM

*David R. Brillinger*[1]

Department of Statistics
The University of California
Berkeley, California

*Haiganoush K. Preisler*[1,2]

Northern California Sickle Cell Center
The University of California
San Francisco, California

## I. INTRODUCTION

Latent variates are random variables which cannot be measured directly, but which play essential roles in the description of observable quantities. They occur in a broad range of statistical problems, a few of which will be surveyed shortly. This paper is concerned particularly with one such problem involving radioactive counting data. The emphasis of the work is on developing estimates of parameters, and examining goodness of fit via "uniform residuals". In the problem considered a plot of the "uniform residuals" versus replicate number led to a change of model.

The random effects model of analysis of variance may be written

[1]*Partially supported by National Science Foundation Grant CEE-7906142.*
[2]*Partially supported by National Institute of Health Grant HL-20985.*




ISBN 0-12-398750-4

$$y = X\alpha + Zu + e \tag{1.1}$$

with $y$, $X$, $Z$ observed, $\alpha$ an unknown parameter, $e$ a random disturbance and with $u$ a vector of random effects (that are of specific interest, but are not directly observable). Here $u$ is an example of a latent variable. Interest often centers on the variances of its entries, the variance components. A number of different procedures, including maximum likelihood, have been developed for estimating the parameters of the model (1.1) and for estimating the realized value of the latent variate $u$. Recent references include Harville (1977), Henderson and Henderson (1979), and Dempster *et al.* (1981). A closely related model is the random coefficient regression model, surveyed in Spjøtvoll (1977).

Another example of a latent variable model is that of controlled regression, given by

$$y_i = h(u_i|\alpha) + e_i \tag{1.2}$$

with the latent variables $u_i$ having known mean, $Eu_i = x_i$, with $h(\cdot)$ of known functional form except for the parameter $\alpha$ and with the $e_i$ random disturbances. Least square procedures have been developed for estimating $\alpha$. References include Berkson (1950) and Fedorov (1974).

There are latent variable models that have been studied especially often in particular substantive fields. For example, the state space model has reached a high level of development in engineering. It may be written

$$y_i = \alpha u_i + e_i \tag{1.3}$$

$$u_i = \beta u_{i-1} + \gamma x_i + e_i' \tag{1.4}$$

$i = 1,2,\ldots$ with $u_i$ a vector-valued latent variate, with

$y_i$, $x_i$ observable and with $e_i$, $e_i'$ random disturbances. The maximum likelihood estimation of the parameters of the model (1.3), in the case of Gaussian $e_i$, $e_i'$, is considered in Gupta and Mehra (1974). Kailath (1980) is a general reference to properties and uses of this model.

Engineers have been led to non-Gaussian models involving latent variates in problems of optical signal estimation. Suppose that $u_t$ denotes the value of a (random) signal of interest at time $t$. In a variety of physical situations what one observes are the times at which photons are absorbed for an optical field of intensity $\pi + \rho u_t$. Here $u_t$ cannot be observed directly. $\pi$ is the average background noise level. Theoretically and experimentally, assuming the process of times to be conditional Poisson seems justified. References include Macchi and Picinbono (1972) and Snyder (1975).

The field of economics has generated interesting models involving latent variates. The MIMIC (Multiple Indicators and MultIple Causes) model of Jöreskog and Goldberger (1975) may be written

$$y = \alpha u + e, \quad u = \beta x + e' \tag{1.5}$$

with $u$ real-valued. The variables $y$ and $x$ are assumed observable. Assuming replicates are available, estimates of the parameters are developed and studied in Jöreskog and Golberger (1975) and in Chen (1981).

An economic model of a different sort is considered in Amemiya and Nold (1975). They consider $y_i$ a binary variate (= 0 or 1) such that

$$\text{Prob}\{y_i = 1 | u_i\} = \exp\{\alpha x_i + u_i\}/(1 + \exp\{\alpha x_i + u_i\}) \tag{1.6}$$

with $u_i$ an unobservable variate having zero mean. Assuming

independent observations on $y_i$ are available, estimates are developed. See also Muthén (1979).

Latent variable models occur commonly in the fields of psychology and education. The first latent variable model to have been studied in real depth is the factor analysis model. It may be written

$$y = \alpha u + e \tag{1.7}$$

with $y$ an observable vector, $\alpha$ a matrix of unknown factor loadings, $u$ the vector of unobservable common factors, and $e$ disturbance. Assuming that replicate observations are available least squares and maximum likelihood (for Gaussian variates) estimation procedures have been developed. Pertinent references include Lawley and Maxwell (1971) and Jöreskog (1973).

In the case that the dependent variate $y$ is discrete, latent structure models play an important role, arising often in connection with ability tests. Suppose the ability of individual $i$ is characterized by the number $u_i$. Suppose individuals attempt $J$ test questions, the $j^{th}$ having difficulty characterized by the number $\alpha_j$. If $y_{ij} = 1$ for a correct response and = 0 otherwise, then the following model might be considered:

$$\text{Prob}\{y_{ij} = 1 | u_i\} = u_i\alpha_j/(1 + u_i\alpha_j) \ . \tag{1.8}$$

The problem of the estimation of the parameters of this and similar models is considered in Sanathanan and Blumenthal (1978), Anderson (1980) and Bock and Aitkin (1981) for example In some cases the subject's ability can be eliminated by conditioning on an appropriate statistic — this is not the case in general however.

Estimation and testing in the latent variable problems described above have been carried out by (generalized) least squares or by maximum likelihood having assumed normality. There have been few, if any, attempts to examine the goodness of fit of the overall model employed. Computing "uniform residuals", as defined below, would appear to be an effective general means to proceed in latent variable problems.

The research reported in this paper was stimulated by a problem (which will be described in detail in the next section) from nuclear medicine. In statistical essence, one had counts $y_{ijk}$, $k = 1,\ldots,K$, $j = i,\ldots,J_i$, $i = 1,\ldots,I$ that conditional on the values of latent variates $u_{ij}$ could be modelled as independent Poissons with mean $\pi + \rho_i x_{ijk} u_{ij}$ the $x_{ijk}$ being known. It was of interest to estimate the parameters $\pi$, $\rho_i$ and to examine the (common) distribution of the $u_{ij}$. The approach taken to the determination of estimates is that of maximum likelihood. Because no simplifications are apparent, numerical quadrature seems necessary to determine the estimates. (Earlier workers on similar problems seem to have come to the same conclusion. We mention Bock and Lieberman (1970), Andersen and Madsen (1977), Sanathanan and Blumenthal (1978), Bock and Aitkin (1981), Reid (1981), and Hinde (1982).) Since the data set analyzed is common to radioactive tracer experiments, and since so many such experiments are carried out in practice, it seems important to provide an analysis by means of widely available statistical programs. Building on the programs of Hinde (1982), we provide a listing of a GLIM session. (GLIM is described in Baker and Nelder (1978).)

The detailed expressions and results are developed for something less than the most general case, namely, for the situation where the $y_i$ conditional on $(u_1,\ldots,u_I)$ are independent with probability mass function $f(y_i|u_i,\alpha)$ while the $u_i$ are independent with probability density function $f(u_i|\beta)$. Generalizations are apparent, the results are not presented with needless detail. This then is the case of contagious distributions. (A review of traditional estimation procedures for contagious distributions is provided by Douglas (1980).) The parameter of interest $\theta = (\alpha,\beta)$ is seen to separate, for the case considered, into a component appearing only in the conditional distribution of the latent variate and into a component appearing only in the distribution of the observed variable conditional on the latent variate.

Since the latent variate $u$ may be viewed as a variate whose values are missing, it is clear that various of the results developed by the Missing Information Principle (Orchard and Woodbury (1972) and the EM method (Dempster *et al.* (1977)) may be of use. Likewise because $u$ sits as an argument in $f(y|u,\alpha)$, yet is actually random, various Bayesian results may be of use.

The approach taken in this paper is the brute force computation of maximum likelihood estimates. The goodness of fit of the model is examined through estimates of the c.d.f. values $F(y_i|\alpha,\beta)$ and of the values $\Phi^{-1}(F(y_i|\alpha,\beta))$. These will have approximate uniform and Gaussian distributions, respectively, when the model holds. The technique has broad applicability.

It is remarkable to note how many problems concerning statistical inference for latent variates have concerned T. W. Anderson. Anderson (1969a,b;1973) are concerned with variance components. Anderson (1955) is concerned with controlled variates. Anderson and Hsiao (1981) makes use of the state space model. Anderson and Rubin (1956) is concerned with factor analysis. Anderson (1954,1968) are concerned with estimation in latent structure and class analysis. Latent variable problems appear to pervade virtually all of his research work. His contributions to the solutions of those problems are substantial.

The remainder of the paper is structured as follows: First there is a description of the particular scientific problem that stimulated the research. Then a general means of constructing estimates for a broad class of latent variable problems is presented. This is followed by some detailed expressions for the case at hand, the results obtained for that case and discussion. The paper concludes with appendices listing the data, providing some details of the computations and listing a GLIM program.

## II. THE PARTICULAR PROBLEM

This section describes the experiment from nuclear medicine that motivated the research of this paper. The objective of the experiment (referred to as a red blood cell survival experiment) was to estimate the mean life span of circulating red cells.

Normally, red cells of healthy individuals live an average of 120 days after which they are removed from circulation,

mainly through the reticuloendothelial system. Hemolytic disorders, whether they are intracorpuscular, extracorpuscular, or combined are characterized by a shortened red cell life span. To determine the effectiveness of treatments claiming to lengthen the life span of defective red cells (in the present case of individuals with sickle cell disease), it is important to have good estimates of mean life span of cells and of sampling fluctuations.

To get estimates of the mean life span one needs to follow a sample of red cells in circulation (the time plot of the number surviving is the red cell survival curve) over a period of time and to observe how many have survived from those present initially. This, of course, is not possible. One is forced to observe these quantities of interest indirectly, as will be explained below.

A typical red cell survival experiment consists of taking a sample of blood from a patient, labeling the cells via a radioactive tracer, and reinjecting them into the patient. Samples of venous blood are then drawn at specified times, after the initial injection of labeled blood, and measured. (For more details on survival studies see International Committee (1972).)

The data given in Appendix I represents the number of gamma photons, for a unit time period (10 minutes) recorded by a gamma scintilation counter. At each time point a sample of blood is drawn from the patient. Then three replicates are pipetted from that sample and put in separate vials. After the last sample is drawn, all the vials are put in the scintilation counter and their level of radioactivity recorded. The

whole group of vials is run through the counter four times, thus getting four readings (the cycles in Appendix I) for each vial. (It may be worth remarking that these cycles, or columns, are not identical replicates due to radioactive decay of the label, $^{51}Cr$, that occurs between their successive measurement.) In Figure II.1 the logarithms of the counts are graphed versus time, using box plots (see Tukey (1977)) to represent the observations at each time point. (There are 12 in each case but the last, which are measurements of background radioactivity, then there are eight.) The figure shows

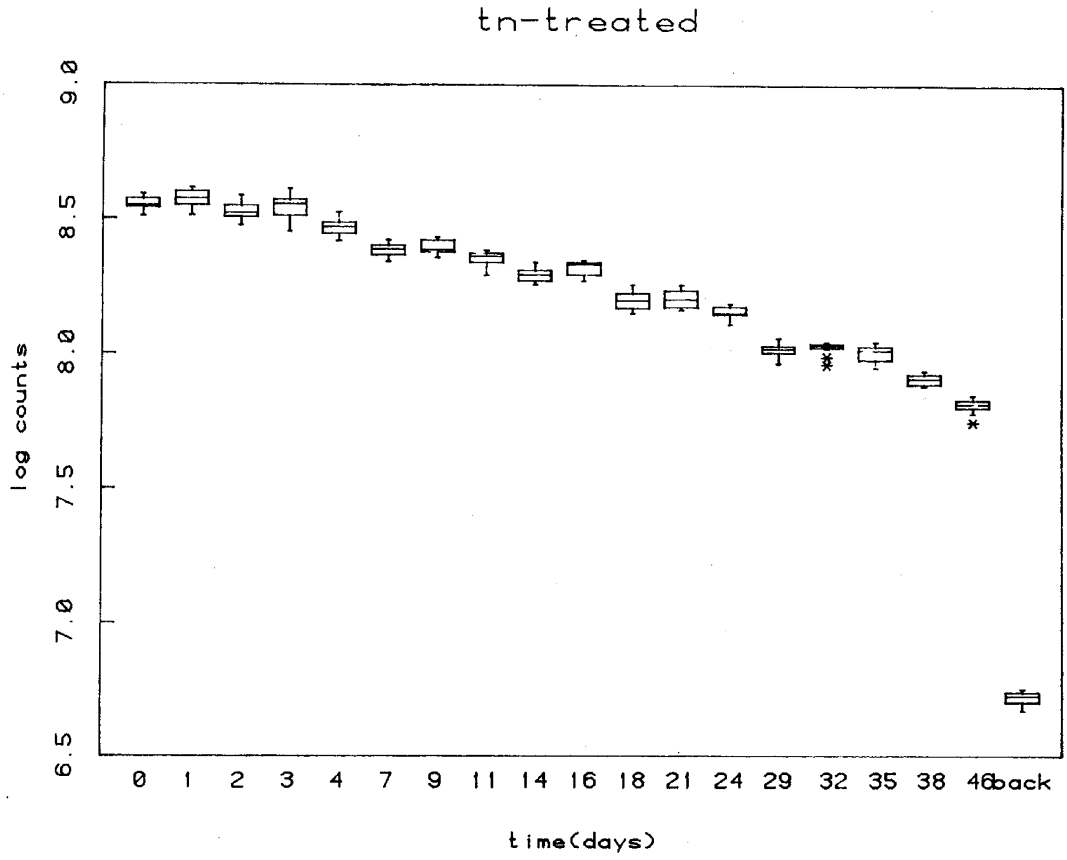


Figure II.1

a more or less steady fall off of counts with increasing time, and a background level substantially below that of the last sample drawn (at 46 days).

Let $y_{ijk}$ denote the count value for the $k^{th}$ cycle of the $j^{th}$ replicate pipetted from the sample drawn at time $t_i$. Because of the everpresent background radioactivity, $y_{ijk}$ is the sum of two random variables,

$$y_{ijk} = z_{ijk} + b_{ijk} \tag{2.1}$$

where $z_{ijk}$ is the count due to radioactivity disintegrations of elements in the $ij^{th}$ blood sample and $b_{ijk}$ is the count due to background radiation.

There is substantial theory and evidence to the effect that the $b_{ijk}$ may be modelled as independent Poissons. (See for example Evans (1955).) We will denote their mean by $\pi/K$ ( K denoting the number of cycles). The distribution of the $z_{ijk}$ cannot be characterized so easily. Variation of the $z_{ijk}$ results from: a decay in cells alive with time, fluctuations associated with the randomness of radioactivity disintegrations and the error (due to pipetting inaccuracy) in the volume of blood used for the various replicates.

For a given volume, u, of blood pipetted from the sample drawn at time $t_i$ physical theory suggests assuming that the $z_{ijk}$'s are independent Poissons with $E\{z_{ijk}|u\} = \rho_i x_k u$, $x_k$ being a known value reflecting the decay of the radioactive tracer between the cycles of counting. Here $\rho_i$ is the expected number of atoms decaying in a unit time period from a unit volume of blood, and hence is proportional to the number of surviving red cells in circulation at time $t_i$ out of the original group of labelled cells. Hence $\rho_i$ is the parameter

of greatest interest. As a function of time $t_i$ it is the survival curve of the treated red cells. The difficulty is that $\rho_i$ cannot be observed directly. What is observable is $y_{ijk}$, whose mean is $\rho_i x_k + \pi/K$, and whose distribution is affected by the variation of the $z_{ijk}$ and the volume actually pipetted. This volume is a latent variate.

Under this model the row totals, $y_{ij}$, for a given volume $u$ are also independent Poissons, with $E\{y_{ij}|u\} = \rho_i u(x_1 + \ldots + x_K) + \pi$. If we assume that the parameters are standardized so that $x_1 + \ldots + x_k = 1$ and that the volume has probability density function $f(u|\beta)$, then the probability of observing the $y_{ij}$ is given by

$$\prod_{i,j} \int \frac{1}{y_{ij}!}(\rho_i u + \pi)^{y_{ij}} \exp\{-(\rho_i u + \pi)\} f(u|\beta)du \ . \qquad (2.2)$$

In the following sections we base estimates of the $\rho_i$, $u$, $\beta$ on (2.2) and further use it to derive estimates of sampling variability.

In most survival studies, of the type described above, the procedure is to substitute for $\rho_i$ a monotonically decreasing function of time and several parameters. (See for example, M. Pollycove (1975) and Berlin (1959).) Doing this substantially reduces the number of parameters to be estimated. In the present paper no assumptions are made concerning the functional form of $\rho_i$. (Although the estimation techniques employed can be used for that case as well.) The main interest of the present paper is in investigating the properties of the latent variable, $u$, and in verifying the distributional assumption set down. Giving $\rho_i$ a specific functional form is yet another assumption to be dealt with.

It is anticipated that once estimates (and corresponding standard error estimates) are determined and plotted against time, then reasonable parametric forms for $\rho_i$ will suggest themselves. We are seeking to operate with one less level of assumptions than has usually been the case.

## III. A GENERAL APPROACH

A variety of problems involving latent variables were described in Section I. A particular latent variable problem requiring solution was described in Section II. This section presents an approximate maximum likelihood solution to a broad class of such problems. In order not to obscure essential details, the most general case is not presented.

Let $U$ be a latent variate with density function $f(u|\beta)$ depending on the parameter $\beta$. Let $Y$ be an observable variate with probability mass (or density) function, given $U = u$, $f(y|u,\alpha)$ depending on the parameter $\alpha$. Then the marginal probability mass function of $Y$ is given by

$$f(y|\alpha,\beta) = \int f(y|u,\alpha)f(u|\beta)du \,. \tag{3.1}$$

The parameters $\alpha$ and $\beta$ are assumed distinct, and the probability function is assumed to separate in the manner of (3.1). (This separation occurs for the tracer data described in the previous section.)

Suppose that a sample of observations $y_1,\ldots,y_I$ is available for the distribution given by (3.1). With $\theta = (\alpha,\beta)$ and

$$\psi(y|\theta) = \frac{\partial \log f(y_i|\theta)}{\partial\theta} \tag{3.2}$$

the maximum likelihood equation for estimating $\theta$ is given by

$$\sum_{i=1}^{I} \psi(y_i|\hat{\theta}) = 0 . \tag{3.3}$$

Various conditions leading to the consistency and asymptotic normality of $\hat{\theta}$ have been set down. For example, $\hat{\theta}$ is consistent under conditions (B-1) to (B-4), (involving $\psi$), of Huber (1967) as $I \to \infty$. Further, if $\theta_0$ denotes the true parameter value, then $\sqrt{I}(\hat{\theta}-\theta_0)$ is asymptotically normal with mean 0 and covariance matrix $i(\theta_0)^{-1}$, under conditions (N-1) to (N-4) of that paper, supposing that $E\{\psi(Y|\theta)\}$ is differentiable at $\theta = \theta_0$ and that

$$i(\theta) = E\{\psi(Y|\theta)\psi(Y|\theta)^T\} . \tag{3.4}$$

Either by direct argument, or by the Missing Information Principle of Orchard and Woodbury (1972), one sees that for a p.m.f. of the form (3.1) the equations (3.3) correspond to setting the following to 0,

$$\sum_{i=1}^{I} \int \psi(y_i|u,\alpha) f(y_i|u,\alpha) f(u|\beta) du \,/\, f(y_i|\alpha,\beta) \tag{3.5}$$

$$\sum_{i=1}^{I} \int \psi(u|\beta) f(y_i|u,\alpha) f(u|\beta) du \,/\, f(y_i|\alpha,\beta) . \tag{3.6}$$

In a variety of circumstances it is not possible to carry out the integrations of (3.5) and (3.6). (This seems to be the case for the problem of Section II.) In consequence some approximation is required. Numerical quadrature is one way to proceed. Suppose that the probability element $f(u|\beta)du$ is approximated as follows

$$f(u|\beta)du \simeq \sum_{m=1}^{M} p_m \delta\{u - u_m\} \tag{3.7}$$

$\delta\{u\}$ denoting a unit mass at $u = 0$. The nodes, $u_m$, and

weights, $p_m$, will generally depend on $\beta$. Then one is led to the following approximate maximum likelihood equations

$$\sum_{i=1}^{I}\sum_{m=1}^{M} \psi(y_i|u_m,\hat{\alpha})w_m(y_i|\hat{\alpha},\hat{\beta}) = 0 \tag{3.8}$$

$$\sum_{i=1}^{I}\sum_{m=1}^{M} \psi(u_m|\hat{\beta})w_m(y_i|\hat{\alpha},\hat{\beta}) = 0 \tag{3.9}$$

where the $w_m$ are weight functions given by

$$w_m(y|\alpha,\beta) = f(y|u_m,\alpha)p_m \Big/ \sum_{n=1}^{M} f(y|u_n,\alpha)p_n . \tag{3.10}$$

Providing the approximation of (3.7) is good enough, solutions $\hat{\alpha},\hat{\beta}$ of the equations (3.8) and (3.9) will be consistent and have the same asymptotic distribution as $\hat{\theta}$ above, for Huber (1967) actually shows that the result holds with (3.3) replaced by

$$\sum_{i=1}^{I} \psi(y_i|\hat{\theta}) = o_p(\sqrt{I}) \tag{3.11}$$

as $I \to \infty$. It is apparent now that, generally, M must tend to $\infty$ with I and that if the cumulative error of the quadrature approximations is $o_p(\sqrt{I})$, then $(\hat{\alpha},\hat{\beta})$ will be asymptotically normal with mean $(\alpha_0,\beta_0)$ and, following (3.4), with covariance matrix that may be approximated by the inverse of

$$\sum_{m=1}^{M}\sum_{i=1}^{I}\begin{bmatrix} \psi(y_i|u_m,\hat{\alpha}) \\ \psi(u_m|\hat{\beta}) \end{bmatrix}\begin{bmatrix} \psi(y_i|u_m,\hat{\alpha}) \\ \psi(u_m|\hat{\beta}) \end{bmatrix}^T w_m(y_i|\hat{\alpha},\hat{\beta}) . \tag{3.12}$$

(Actually, it follows from Huber's general results, that in broad circumstances the estimate will be asymptotically normal with covariance matrix estimable by a modified form of (3.12) even when the model is untrue.

e is led

tions

(3.8)

(3.9)

(3.10)

ugh, solu-

e consis-

above,

s with

(3.11)

must tend

the

ill be

wing (3.4),

he inverse

(3.12)

that in

lly normal

of (3.12)

The problems now arise of how to determine the approximating discrete measure of (3.7) and how to solve the equations (3.8) and (3.9). Consider (3.7) first, and suppose that $\beta$ is known. One might take $p_m = 1/M$ and

$$u_m = F^{-1}\left(\frac{m}{M+1} \mid \beta\right) \tag{3.13}$$

where $F(u|\beta)$ denotes the c.d.f. of $U$. This is the usual discrete approximation to an integral. Expressions for the error of approximation may be derived directly from the results of Niederreiter (1978). Alternatively, suppose that $U$, given $\beta$, may be approximated by some function of a normal variate, (as in a Cornish-Fisher expansion), then the integral may be approximated by a Gauss-Hermite formula. Sources for the nodes and weights are given in Davis and Rabinowitz (1975). They further quote (p.174) the error of using an $M$ point formula as a multiple of the $2M^{th}$ derivative of the integrand. An approximation of quite another sort comes from noting that expression (3.5) is actually

$$\sum_{i=1}^{I} E\{\psi(y_i|U,\alpha)|\ Y = y_i\}\,. \tag{3.14}$$

The individual terms have the form $E\{h(V)\}$ for $V$ a random variable with the distribution of $U$ given $Y = y$. Since the time of Gauss people have been approximating such expected values by either making Taylor series expansions of $h(\cdot)$ and using the moments of $V$ or by evaluating $h(\cdot)$ at selected points in the neighborhood of $\mu_V = E\{V\}$. If $\sigma_V^2 = \text{var } V$, then the simplest of these approximations are

$$h(\mu_V),\ h(\mu_V) + \tfrac{1}{2}\sigma_V^2 h''(\mu_V),\ \tfrac{1}{2}[h(\mu_V + \sigma_V) + h(\mu_V - \sigma_V)]. \tag{3.15}$$

One reference is Evans (1972). Monte Carlo is yet another approximation procedure.

One possible procedure for determining estimates $\hat{\alpha}$, $\hat{\beta}$ may now be described. Have the equations (3.8), (3.9) in mind. Suppose that a computer program, such as GLIM, is available for determining (fixed) weight maximum likelihood estimates for the p.m.f. $f(y|u,\alpha)$ and, separately, for the p.d.f. $f(u|\beta)$. Given estimates $\hat{\alpha}_*$, $\hat{\beta}_*$ at an iterative step determine improved estimates using the programs to determine solutions to (3.8), (3.9) with the weights replaced by $w_m(y_i|\hat{\alpha}_*,\hat{\beta}_*)$. If this procedure converges, it will converge to estimates $\hat{\alpha}$, $\hat{\beta}$ satisfying equations (3.8), (3.9). (For examples and theorems relating to the convergence of such a procedure, see Boyles (1980,1982) and Wu (1983).) In the case of $f(y|u,\alpha)$ the program will be run for data involving M copies of each $y_i$, specifically for the data, $y_i$, $u_m$ with weight $w_m(y_i|\hat{\alpha}_*,\hat{\beta}_*)$, $m = 1,\ldots,M$ and $i = 1,\ldots,I$. Hinde (1982) does this.

In summary, the estimation procedure proposed is an approximation to maximum likelihood. This has the advantage over other estimation procedures (such as: method of moments, minimum chi-squared, weighted least squares, transformation to normality) that by making the approximation in (3.7) arbitrarily good, one may come arbitrarily close to m.l. estimates. Estimation procedures that might prove useful in some circumstances are the Fourier method (Feuerverger and McDunnough, 1981), the empirical prob. gen. func. (Turner, 1982), and the approximation of the likelihood by a saddle-point procedure (Barndorff-Nielson and Cox, 1979). However, none of

these two proved helpful to us in the case of the problem described in Section II.

The approach suggested, being a form of weighted estimation, has the further advantage that a robust/resistant version may be implemented directly. Such implementations are discussed in Green (1982).

Once the estimates $\alpha$, $\beta$ are in hand a number of interesting questions may be addressed and further things computed. For example, one may ask: Is the distribution of $U$ concentrated at a point? Are the distributions $f(y|u,\alpha)$, $f(u|\beta)$ validated by the data?

In the next section it will be indicated how the first question may be addressed by the deviance statistic proceduced by GLIM. Turning to the broad question of examining overall goodness of fit, one knows that if a random variate $Y$ has c.d.f. $F(y)$, then the variate $F(Y)$ has the uniform distribution. In the present setup, the c.d.f. is given by

$$F(y|\alpha,\beta) = \int F(y|u,\alpha)f(u|\beta)du \tag{3.16}$$

with $F(y|u,\alpha)$ the conditional c.d.f. of $Y$ given $U = u$. Expression (3.16) may be approximated by

$$\hat{F}(y|\alpha,\beta) = \sum_{m=1}^{N} F(y|u_m,\alpha)p_m . \tag{3.17}$$

"Uniform residuals" may now be computed as $\hat{F}(y_i|\hat{\alpha},\hat{\beta})$ once the estimates $\hat{\alpha}$, $\hat{\beta}$ are at hand. These may be plotted on uniform probability paper, graphed against possible explanatory variables and the like as is done with the traditional regression residuals. Examples of this are presented later in the paper. The distribution of these residuals is complicated.

Loynes (1980) derives the asymptotic distribution in a related situation.

The approach of this paper allows estimates of the value of $U$ corresponding to a given value of $Y$ to be constructed. Specifically, one can compute

$$\hat{u}_i = \sum_{i=1}^{I} \sum_{m=1}^{M} u_m w_m(y_i|\hat{\alpha},\hat{\beta}) \tag{3.18}$$

as an estimate of $E\{U|Y = y_i\}$. These values may be used to examine the properties of $U$.

Other types of general residuals are discussed in Pregibon (1981,1982) and Green (1982); however, the above "uniform residuals" and their Gaussian transform seem apt for our purposes.

## IV. SOME DETAILS

For the data set of concern in this paper, the following seems a plausible model: $Y$ is conditionally Poisson with

$$f(y|u,\alpha) = \frac{1}{y!}(\mu_u)^y \exp\{-\mu_u\} \tag{4.1}$$

where $\mu_U = \pi + \rho U$, $\alpha = (\pi,\rho)$, and $U$ is a random variable with mean 1 and density $f(u|\beta)$. In particular the cases of $U$ normal, lognormal and gamma will be considered. For observations corresponding to background (Poisson) noise $U$ is identically 0. For other observations $U$ represents the volume of solution pipetted (standardized to have mean 1).

Supposing $\text{var } U = \sigma^2$ ($\sigma$ hence is the coefficient of variation of $U$), for the above model:

$$EY = \pi + \rho \tag{4.2}$$

$$\text{var } Y = \pi + \rho + \rho^2\sigma^2 . \tag{4.3}$$

If $\rho\sigma$ is small, then $Y$ will be approximately Poisson. In

general, however, its distribution will be complicated. (In the case that $U$ is normal, the distribution of $Y$ will be Poisson normal (or Hermite), see Steyn (1976).) For the given data it is of interest to: see if $Y$ is approximately Poisson, to estimate the coefficient of variation of the pipetting error, and to examine the goodness of fit of various assumed densities for $U$, among other things.

The computations to be described were carried out by means of the statistical program GLIM-3. This program is now widely available. It uses iteratively reweighted least squares to fit generalized linear models with $EY_i = h(x_i'\alpha)$ and $Y_i$ from a one parameter exponential family. It is convenient to use for fitting the p.m.f. (4.1). GLIM also contains a high-level syntax for handling variables with factorial structure, vectors and non-full rank models. Its powerful directives shortened the length of the program considerably (they act like subroutine calls), and allowed simple simulation of the whole situation for checking programs and logic. In the development of the analysis and in carrying out alternative analyses it was most helpful to be able to use GLIM's feature allowing the "dumping" of the analysis at the stage reached, followed by "restoring" at a later time. One disadvantage is that there is no directly accessible matrix inversion routine; however, by setting up an artificial regression problem the inverse of the matrix (3.12) could be determined. (The specifics may be seen in the program listing in Appendix III.)

GLIM produces maximum likelihood estimates in standard cases. It measures the degree of fit via a deviance function

$$D(\theta|y_1,\ldots,y_I) = \sum_{i=1}^{I} d(\theta|y_i) \qquad (4.4)$$

where

$$d(\theta|y_i) = -2[\underline{1}(y_i|\theta) - \sup_{\theta'} \underline{1}(y_i|\theta')]$$

with $\underline{1}(y_i|\theta) = \log f(y_i|\theta)$. Minimizing the deviance is equivalent to maximizing the likelihood. As noted by Pregibon (1981,1982) following the usual approximation to the null distribution of the likelihood ratio statistic, the distribution of the drop in deviance resulting from introducing unnecessary parameters may be approximated by $\chi^2_f$, $f$ being the number of unnecessary parameters introduced.

For the data set analysed in this paper, some of the observations (the background measurements) correspond to unconditional Poisson variates. For these observations, the expected value (3.4) was evaluated directly and inserted into expression (3.12). This appeared to give substantially improved estimates for the variances.

## V. RESULTS

The data employed in the analysis are listed in Appendix I. The basic variate modelled is the row total, $y_{ij}$, with $i = 1,\ldots,19$ running over the 18 time values (0,1,...,46 days) and the background measurements and with $j$ running over replicates. (There are three replicates except in the background case when there are only two.) There are 56 observations all told.

The first model to be fit was one of $y_{ij}$, conditional on the "volumes", being Poisson with mean $\pi + \rho_i u_{ij}$, and with the $u_{ij}$ independent normals of mean 1 and variance $\sigma^2$.

(Because of the double subscript this isn't quite the model discussed in the previous sections, however, the extension of that discussion is direct.) The analysis was carried out by GLIM. (A program listing is provided in Appendix III.) Numerical integration was carried out by Gaussian quadrature with $M = 12$ nodes. The analysis was also carried out for $M = 3,4,\ldots,11$ and the results found not to change much for $M \geq 8$.)

In the approach adopted, as in Hinde (1982), a simple Poisson model (corresponding to $M = 1$) is first fit. This has the advantages of allowing one to see how near Poisson the data is and of producing initial values for the recursive fitting procedure to follow. The deviance statistic for the simple Poisson fit was 276.59. It fell to 132.49 with the full model, a substantial reduction. As mentioned earlier, the fall in deviance may be modelled as $\chi^2_{36}$ in the null case, for sufficiently large sample size and appropriate regularity conditions. It was found that the estimates of the parameters $\pi$ and the $\rho_i$ did not change much at all; however, the estimates of the standard errors of the $\hat{\rho}_i$ became larger. (These values are reported later in the paper.)

The goodness of fit of the model was examined by means of the "uniform residuals" $\hat{F}_i(y_{ij}|\hat{\alpha},\hat{\beta})$ computed via formula (3.17). Table 5.1 provides a stem and leaf display of these values. Were the model correct, their distribution would be (approximately) uniform on the interval (0,1). The stem and leaf display suggests that there is some departure from uniformity. (In order to assess the likely extent of sampling fluctuations in this display, five simulations of the

postulated model using the estimated values $\hat{\alpha}$, $\hat{\beta}$ as the population values were carried out. The stem and leaves were much nearer to uniformity.) The first graph in Appendix III (labeled Uniform Probability Plot) is a plot of the "uniform residuals" versus rank. Were the assumptions made satisfied, the relationship would be near linear. Again, there is evidence of lack of fit. As well as these uniform displays, corresponding "normal residuals" were computed by applying the inverse normal cumulative to the "uniform residuals". The logic was that one is more used to examining departures from normality than uniformity. Table 5.1 and Appendix III give the corresponding displays. Once again there is substantial evidence of lack of fit. Further, an outlier turns up in the normal displays.

Two attempts to find an improved model by changing the postulated distribution of $U$ were made. Namely, $U$ was assumed to be lognormal and gamma (with mean 1 and variance $\sigma^2$), respectively. There was virtually no change in the values of the estimates or of the deviance. This may well be due to the small value of the coefficient of variation $\sigma$ in the

*Table V.1*

| *"Uniform residuals"* | *"Normal residuals"* |
|---|---|
| *0 : 04458* | *Low : -3.49* |
| *1 : 23679* | *-1 : 876* |
| *2 : 77889* | *-1 : 42100* |
| *3 : 5679* | *-0 : 966665* |
| *4 : 346779* | *-0 : 4433221110* |
| *5 : 234779* | *0 : 01122234444444* |
| *6 : 1444555789* | *0 : 55566667889* |
| *7 : 02344689* | *1 : 0012* |
| *8 : 25558* | *1 :* |
| *9 : 8* | *2 : 1* |
| *10 : 0* | *2 : 6* |

present context. It would appear that these different distributions for U, normal, lognormal and gamma, cannot be readily distinguished in this situation.

The situation was further examined by a form of residual plot, namely a plot of the points $(j, \hat{F}_i(y_{ij}|\hat{\alpha},\hat{\beta})$, i.e., a plot of the "uniform residuals" versus the replicate number. If the lab technician was treating the replicates in different fashions, this plot might make that apparent. The plot obtained is the last one in Appendix III. This figure does suggest that there is a difference between replicates - the volume pipetted is tending to be smaller for the third replicate.

A modified model was therefore examined, namely one involving $EU_{ij} = \gamma_j$, with $\sum_j \gamma_j = J$, and with the coefficient of variation of $U_{ij}$ remaining $\sigma$. The distribution of $U_{ij}$ was taken to be normal. With the modified model the deviance dropped to 123.16, a drop of 9.33 from the previous deviance, with the addition of two parameters. Large sample theory suggests that the distribution of the drop in the null case might be approximated by $\chi^2_2$. The degree of fit was examined by computing "uniform" and "normal" residuals as above. Table V.2 provides the stem and leaf displays. While evidence of lack of fit remains, it is less strong. The probability plots lead to the same conclusion.

Figure V.1 is a graph of $\log \hat{\rho}_i \pm 2$ s.d. versus time $t_i$ based on the model involving replicate effects. This picture is the one of interest to the scientists. It may be usefully compared to Figure II.1 based on the basic data alone. The estimates and their estimated standard errors are given in Table V.3.

*Table V.2*

| *"Uniform residuals"* | *"Normal residuals"* |
|---|---|
| *0 : 02479* | *Low : -3.11* |
| *1 : 0288* | *-2 : 2* |
| *2 : 4455* | *-1 : 85* |
| *3 : 244447* | *-1 : 332* |
| *4 : 6788899* | *-0 : 9977775* |
| *5 : 0225799* | *-0 : 444431111000* |
| *6 : 023455778* | *0 : 011122233334444* |
| *7 : 567788* | *0 : 577788899* |
| *8 : 23348* | *1 : 002* |
| *9 : 489* | *1 : 5* |
| | *2 : 0* |
| | *2 : 6* |

The estimate of the standard error of $\hat{\rho}_8$ is noticeably larger than the rest. This phenomenon occurred in some of the simulations as well (and not always for the same i ) and no doubt represents the fact that one has only three replicates at each time point and is estimating a separate $\rho_i$ at each time point. One would be wise to form a pooled estimate of standard error.

It was indicated in Section III of the paper, that having obtained estimates of the values of the unknown parameters, $\alpha$, $\beta$ the conditional expected values $E\{U_{ij}|y_{ij}\}$ could be estimated via formula (3.18). Table V.4 gives the stem and leaf display of these estimated values, having removed the replicate effects. The distribution has a Gaussian shape. There is one outlier and a suggestion of skewness to the left.

Table V.3

| Estimate | S.E. | Parameter |
|---|---|---|
| 3304 | 40.65 | pi |
| 0.1699e-01 | 0.2029e-02 | sig |
| 1.005 | 0.5837e-02 | gam1 |
| 1.007 | 0.5102e-02 | gam2 |
| 0.1741e+05 | 187.1 | rho(1) |
| 0.1785e+05 | 217.0 | rho(2) |
| 0.1690e+05 | 270.1 | rho(3) |
| 0.1724e+05 | 107.2 | rho(4) |
| 0.1577e+05 | 135.9 | rho(5) |
| 0.1415e+05 | 306.6 | rho(6) |
| 0.1437e+05 | 262.5 | rho(7) |
| 0.1367e+05 | 2129. | rho(8) |
| 0.1264e+05 | 490.0 | rho(9) |
| 0.1309e+05 | 255.3 | rho(10) |
| 0.1123e+05 | 106.2 | rho(11) |
| 0.1137e+05 | 386.8 | rho(12) |
| 0.1058e+05 | 112.5 | rho(13) |
| 8809 | 148.5 | rho(14) |
| 8878 | 313.1 | rho(15) |
| 8628 | 181.5 | rho(16) |
| 7531 | 287.6 | rho(17) |
| 6527 | 108.4 | rho(18) |
| zero | aliased | rho(19) |

Table V.4

Stem and Leaf of Corrected Volumes

```
Low :  .952
 96 :  7
 97 :  578
 98 :  1569
 99 :  00013444459999
100 :  00012233455556677
101 :  00011123448
102 :  38
103 :  9
```

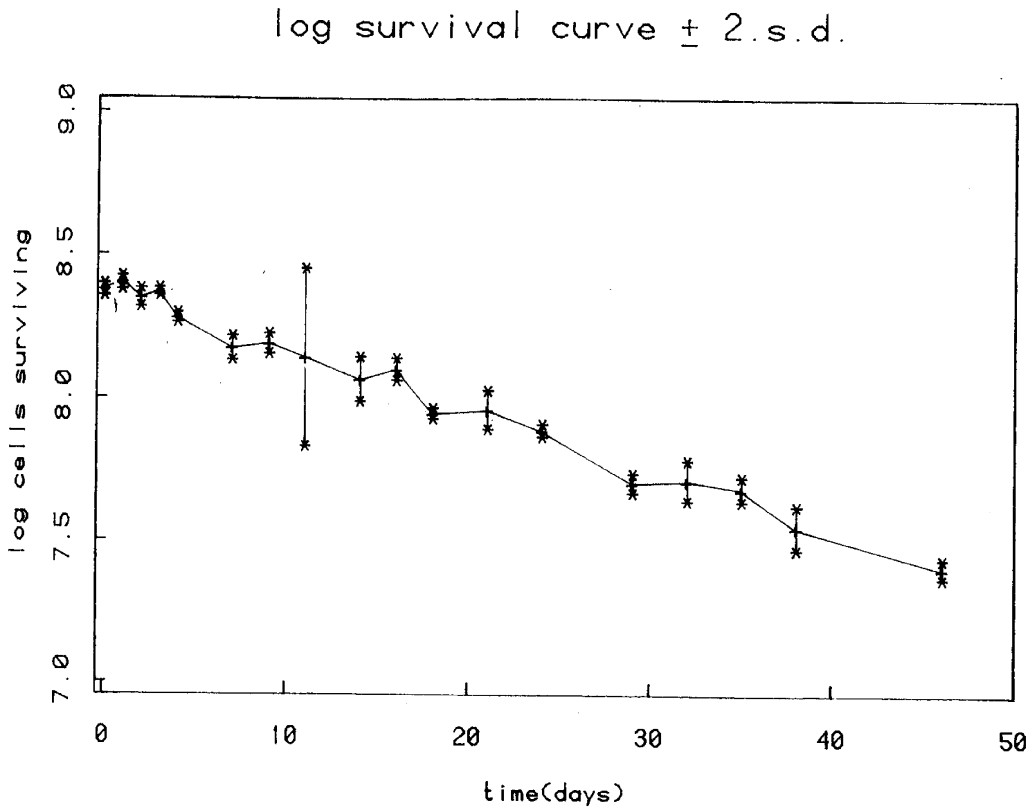


FIGURE V.1

## VI. DISCUSSION AND CONCLUDING REMARKS

It has been suggested that when one is studying counts of sampled radioactivity data, and specifically when blood is alloquoted to be placed in a counter, it is reasonable to expect the variance of the error in such cases to be proportional to the squares of the corresponding mean counts (see for example, Preisler (1977) and Jennrich and Ralston (1979)). The present set of data, described in Section II, allow this suggestion to be verified at a basic level. The replicate counts at each time point and the methodology adopted allowed the estimation of the distribution of the unobserved volumes sampled — the variation of the volumes may be viewed as leading

to inflation over Poisson variation. A quantity of specific interest in this connection is the coefficient of variation of the pipetting error. For the data analyzed, the estimate is 1.7% (±.2%), see Table V.3. Substituting this value in the formula, (4.3), for the variance of the counts and taking note of the estimated values of the $\rho_i$ (Table V.3), one sees that the term dominating the variance is $\rho^2\sigma^2$. That is, the variance of the counts is approximately proportional to the square of the corresponding counts.

The computations of the paper were unable to distinguish the actual distribution of the volumes sampled, beyond finding it to be approximately normal for the apparent parameter values. (For these values the normal, lognormal and gamma could not really be expected to be distinguishable unless a great deal of data was available in any case.)

"Uniform residuals" proved very useful in checking other aspects of the model however. Specifically goodness of fit was examined by plotting these residuals versus various explanatory variables. For example, the plot versus replicate number suggested that what had been treated as identical might better be treated as different. As to why this difference between replicates exists, we will have to leave to the experimenters.

The work of the paper shows that maximum likelihood estimation is in fact a viable approach to a broad class of latent variable problems and that goodness of fit may be examined in such a fashion that improved models suggest themselves. GLIM proved an effective tool for carrying out the needed computatations. Standard error estimation and simulations proved feasible within its structure.

ACKNOWLEDGMENTS

We would like to thank G. R. Ihaka and S.-T. Chiou of the Berkeley Statistics Department for help with GLIM, and for making some of the plots, respectively. We would also like to thank Dr. M. Guis and Dr. W. Mentzer for providing us with the data and helping us understand the experimental procedures involved in red blood cell survival experiments.

of the
for
like
s with
cedures

APPENDIX I

| TN-Treated Counts in 10 minute intervals | | | | |
|---|---|---|---|---|
| Cycle: | 1 | 2 | 3 | 4 |
| Times(days) | | | | |
| 0 | 5352 | 5092 | 5381 | 5123 |
| 0 | 5149 | 5121 | 5113 | 4960 |
| 0 | 5328 | 5156 | 5244 | 5150 |
| 1 | 5472 | 5417 | 5516 | 5147 |
| 1 | 5435 | 5263 | 5447 | 5262 |
| 1 | 5331 | 4974 | 5177 | 5001 |
| 2 | 5109 | 4956 | 4797 | 5135 |
| 2 | 5358 | 4930 | 4957 | 5182 |
| 2 | 5231 | 4929 | 4969 | 5070 |
| 3 | 5230 | 5128 | 5136 | 5238 |
| 3 | 5462 | 5266 | 5282 | 5496 |
| 3 | 5035 | 4686 | 4712 | 4896 |
| 4 | 4951 | 4759 | 4793 | 4798 |
| 4 | 4755 | 4521 | 4630 | 4632 |
| 4 | 5035 | 4665 | 4769 | 4873 |
| 7 | 4502 | 4537 | 4381 | 4301 |
| 7 | 4457 | 4434 | 4381 | 4187 |
| 7 | 4386 | 4278 | 4189 | 4308 |
| 9 | 4559 | 4348 | 4295 | 4359 |
| 9 | 4538 | 4590 | 4476 | 4387 |
| 9 | 4540 | 4372 | 4246 | 4325 |
| 11 | 4323 | 4289 | 4238 | 4178 |
| 11 | 4378 | 4290 | 4265 | 4176 |
| 11 | 4319 | 4331 | 4166 | 3984 |
| 14 | 4120 | 4053 | 3971 | 3932 |
| 14 | 4189 | 4005 | 3892 | 3846 |
| 14 | 4076 | 4012 | 3852 | 3903 |
| 16 | 4181 | 4211 | 4155 | 3904 |
| 16 | 4195 | 4123 | 4050 | 3942 |
| 16 | 4158 | 4187 | 4161 | 3909 |
| 18 | 3799 | 3856 | 3718 | 3670 |
| 18 | 3743 | 3659 | 3531 | 3520 |
| 18 | 3608 | 3568 | 3458 | 3472 |
| 21 | 3849 | 3739 | 3526 | 3654 |
| 21 | 3855 | 3785 | 3547 | 3645 |
| 21 | 3763 | 3623 | 3499 | 3540 |
| 24 | 3434 | 3451 | 3319 | 3355 |
| 24 | 3569 | 3538 | 3461 | 3560 |
| 24 | 3591 | 3475 | 3447 | 3456 |
| 29 | 3027 | 3062 | 2870 | 3033 |
| 29 | 3045 | 3159 | 2935 | 3030 |
| 29 | 3011 | 3064 | 2955 | 3151 |
| 32 | 3055 | 3065 | 3105 | 3086 |
| 32 | 3091 | 2946 | 3046 | 3085 |
| 32 | 3094 | 2866 | 3039 | 3074 |
| 35 | 3038 | 2869 | 3116 | 2945 |
| 35 | 3064 | 2956 | 3029 | 3087 |
| 35 | 3022 | 2820 | 2997 | 2860 |
| 38 | 2697 | 2638 | 2788 | 2682 |
| 38 | 2729 | 2643 | 2730 | 2797 |
| 38 | 2781 | 2642 | 2721 | 2663 |
| 46 | 2516 | 2424 | 2312 | 2441 |
| 46 | 2477 | 2500 | 2382 | 2552 |
| 46 | 2509 | 2462 | 2437 | 2482 |
| back | 829 | 853 | 839 | 787 |
| back | 846 | 831 | 803 | 819 |

## APPENDIX II. SOME DETAILS OF THE COMPUTATIONS

A variety of minor issues arose in the course of writing the computer programs. For example, in the computation of expression (4.1) it was found appropriate to divide it by $y^y \exp\{-y\}/y!$ in order to have numerical stability. This did not necessitate changes in the estimating equations.

Because of the large magnitudes of the observations being analyzed, normal approximations were made to the Poisson c.d.f., $F(y|u_m,\alpha)$, of (3.17) and to $f(u|\beta)$ when U was modelled as gamma. (The Wilson-Hilferty approximation was used in the latter case.)

The convergence criterion that GLIM employs in the fitting of system models is, apparently, stop when the deviance changes by less than $10^{-4}$. This approach was adhered to in fitting the latent variable model, although a criterion based on how near the right-hand sides of the estimating equations have got to 0 would undoubtedly be better. In trials involving many iterations (up to 100) it was found that things did not change much after 15 iterations.

Not all the GLIM macros are listed in Appendix III; however, the structure of the work should be apparent. Other subfiles were created for alternate numbers of integration nodes, m, for the replicate effect, lognormal and gamma cases. Virtually all of the work was done within GLIM. One disadvantage of GLIM-3 is its not containing an explicit matrix inversion algorithm. One was needed in computing estimates of the standard errors via (3.12) — this was done by setting up an artificial regression problem. The boxplots of Figure II.1 were done within "S".

The work was carried out on the Berkeley Statistics Department's VAX 11/750 running UNIX.

## APPENDIX III

```
cat fort3

$c Gauss-Hermite integration - 12 nodes and weights
$mac nodes $calc %m =12 $var %m z p $data z p $read !
5.5009 .0000002 4.27183 .000048 3.22371 .00220338
2.25946 .0291167 1.34038 .146967 .444403 .321664
-5.5009 .0000002 -4.27183 .000048 -3.22371 .00220338
-2.25946 .0291167 -1.34038 .146967 -.444403 .321664
$endm $return
!
!
$subfile newmodel !
$c General macros for fitting a mixed Poisson
$mac expand ! Makes %m copies of the data and associated variates
$calc %o = %n*%m : lp = %lp ! Need to save %lp
: u = %lt(rho,%i) ! Initial values of volumes pipetted
$unit %o $calc j = %gl(%n,1) : k = %gl(%m,%n) !
: ey = y(j) : erho = rho(j) : ep = p(k) : mu = lp(j) $endm
!
$mac model ! Sets up Poisson model.
$var %n f ! f is the marginal p.m.f.
$err p $link i $yvar ey $calc %lp = mu : %fv = %lp $recy $wei ew $endm
!
$mac setup $use nodes $use expand $use model $endm
!
$mac weight ! Computes the weights and marginal p.m.f.
$calc mu = %lp :ew = %log(ep) - mu + ey + ey*%log(mu/ey) : ew = %exp(ew)
:f = 0 : f(j) = f(j) + ew : ws = f(j) : ew = ew/ws $endm
!
$mac test ! Tests for convergence by change in deviance and no. iterates
$use weight $calc %c = .001 : %e = -2*%cu(%log(f))
: %t = %ge(%d-%e,%c)+%ge(%e-%d,%c) : %d = %e : %s=%if(%le(%t,0),0,%s-1)
$print 'current deviance = ' *8 %e ' %s = '*2 %s $endm
!
$macro fitting ! Carries out the overall fitting.
$use weight $use test $use estb $fit erho.eu $endm
!
$mac uresid ! Computes the "uniform residuals" ur
!Makes normal approximation to Poisson cumulative
$var %o eur $calc eur = %np((ey-mu)/%sqrt(mu))
: ur = 0 : ur(j) = ur(j) + eur*ep $endm
!
$mac ufitted ! Computes the fitted volumes uf
$calc uf = 0 : uf(j) = uf(j)+eu*ew $endm
!
$return
!
!
$subfile normalerr
$c Assuming normal distribution for volume, estimates sig = %b
$mac estb $calc eu = u(j)*(1+%b*z(k)) : %u = %cu(u)
: %z = %cu(ew*u(j)*(eu-1)**2) : %b = %sqrt(%z/%u)
: eu = u(j)*(1+%b*z(k))
$print ' estimate of sig is ' *7 %b $endm
!
$mac errb ! Computes values required for information matrix.
$calc r1 = (z(k)**2-1)/%b $endm
!
$mac serror ! Computes s.e.'s, prints estimates and s.e.'s
$extract %pe $use errb !
$calc s1 = (ey/%lp-1)*ew : s2 =s1*eu : r1 = r1*ew
$calc %y = %lp(%n) $unit %n !
$calc pi = 0 : r = 0 : sig = 0 !
: pi(j) = pi(j)+s1 : r(j) = r(j)+s2 : sig(j) = sig(j)+r1 !
: pi = u*pi+(1-u)/%sqrt(%y) : sig = sig*u !
$alias $wei $err n $scale 1 $yvar py
$calc py = pi*%pe(1) + r*%pe(rho+1) + sig*%b
$fit pi.%gm + rho.r + sig - %gm $d e $endm
!
!
$finish
%
```

```
$c A GLIM SESSION TO FIT A NORMAL "VOLUME"
.$calc %n  =56 ; %i=19 $c SETS UP SAMPLE SIZE AND NUMBER OF RHO'S
.$units %n $factor rho %i
.$data y $dinput 8 $data rho $dinput 8 $c BRINGS IN COUNT DATA AND "i"
.$err p $link i $yvar y $c SETS UP SIMPLE POISSON (CONSTANT VOLUME)
.$alias $fit rho $d e$
        Scaled
 Cycle  Deviance      DF
    3    276.6        37

       Estimate       S.E.      Parameter
    1    3304.        40.64     %gm
    2  0.1742e+05    92.52      rho(1)
    3  0.1784e+05    93.28      rho(2)
    4  0.1690e+05    91.58      rho(3)
    5  0.1722e+05    92.15      rho(4)
    6  0.1576e+05    89.47      rho(5)
    7  0.1414e+05    86.41      rho(6)
    8  0.1437e+05    86.86      rho(7)
    9  0.1368e+05    85.51      rho(8)
   10  0.1265e+05    83.48      rho(9)
   11  0.1309e+05    84.35      rho(10)
   12  0.1123e+05    80.60      rho(11)
   13  0.1137e+05    80.89      rho(12)
   14  0.1058e+05    79.25      rho(13)
   15   8811.        75.43      rho(14)
   16   8881.        75.58      rho(15)
   17   8631.        75.03      rho(16)
   18   7533.        72.55      rho(17)
   19   6528.        70.21      rho(18)
   20    Zero     Aliased       rho(19)
  Scale Parameter taken as   1.0000

.$input 3 no12  newmodel normalerr $c MACROS FOR FITTING
.$calc %b = .02 ; %s = 15 $c INITIAL VALUE FOR SIG AND MAXIMUM NUMBER OF ITERATIONS
.$use setup $wh %s fitting
 ----- Current model abolished
 current deviance =    276.61716 %s =    14.
  estimate of sig is   0.0199987
        Scaled
 Cycle  Deviance      DF
    2    492.4       653

 ----- Current display inhibited

 current deviance =    132.48734 %s =   0.
  estimate of sig is   0.0189142
.$
        Scaled
 Cycle  Deviance      DF
    2    51.54       611

.$use serror $c DISPLAY ESTIMATES AND STANDARD ERRORS
 ----- Current model abolished
        Scaled
 Cycle  Deviance      DF
    1   0.1826e-09    36

       Estimate       S.E.      Parameter
    1    3304.        40.64     pi
    2  0.1891e-01   0.1766e-02  sig
    3  0.1744e+05    318.2      rho(1).r
    4  0.1781e+05    156.8      rho(2).r
    5  0.1690e+05    323.9      rho(3).r
    6  0.1727e+05    94.05      rho(4).r
    7  0.1577e+05    157.5      rho(5).r
    8  0.1414e+05    222.6      rho(6).r
    9  0.1437e+05    241.4      rho(7).r
   10  0.1366e+05    363.9      rho(8).r
   11  0.1265e+05    463.1      rho(9).r
   12  0.1308e+05    996.7      rho(10).r
   13  0.1123e+05    97.33      rho(11).r
   14  0.1137e+05    201.7      rho(12).r
   15  0.1058e+05    138.2      rho(13).r
   16   8807.        270.6      rho(14).r
   17   8878.        240.3      rho(15).r
   18   8630.        130.9      rho(16).r
   19   7531.        412.7      rho(17).r
   20   6526.        153.5      rho(18).r
   21    Zero     Aliased       rho(19).r
  Scale Parameter taken as   1.0000
```

```
.$use uresid $c COMPUTE "UNIFORM  RESIDUALS"
.$calc n=%gl(%n,1) $sort us ur $plot us n $c UNIFORM PROBABILITY PLOT
   1.08     ‡
   1.02     ‡                                                    u
  0.960     ‡                                                   u
  0.900     ‡                                                   u
  0.840     ‡                                                uu2
  0.780     ‡                                              uuu
  0.720     ‡                                           2u2u
  0.660     ‡                                       2u22u
  0.600     ‡                                    2u
  0.540     ‡                                uu2
  0.480     ‡                            2uu
  0.420     ‡                        u2
  0.360     ‡                    u2
  0.300     ‡                22
  0.240     ‡              u
  0.180     ‡          u2
  0.120     ‡       2
 0.600e-01  ‡uu2
    0.      ‡u
   .........‡.........‡.........‡.........‡.........‡.........‡
            0.       16.0       32.0       48.0       64.0       80.0
.$calc nr=%nd(us) : n=%nd((%gl(%n,1)-.5)/(%n+1)) $c NORMAL PROBABILITY PLOT
.$plot nr n$
   3.20     ‡
   2.80     ‡                                              n
   2.40     ‡
   2.00     ‡                                            n
   1.60     ‡
   1.20     ‡                                        nnnn
  0.800     ‡                                   22n2
  0.400     ‡                              n32322n
    0.      ‡                         3232
 -0.400     ‡                    322
 -0.800     ‡               nn2
 -1.20      ‡             nnn
 -1.60      ‡    n nnn
 -2.00      ‡
 -2.40      ‡
 -2.80      ‡
 -3.20      ‡
 -3.60      n
 -4.00      ‡
   .........‡.........‡.........‡.........‡.........‡.........‡
          -2.40      -1.20        0.       1.20       2.40       3.60
.$calc n = %gl(3,1) $c SET UP REPLICATE VARIATE FOR "RESIDUSAL PLOT"
.$plot ur n $c "UNIFORM RESIDUALS" VERSUS REPLICATE
   1.08     ‡
   1.02     ‡                    u
  0.960     u
  0.900     u
  0.840     u                    3                                  u
  0.780     u                    u                                  u
  0.720     2                    4                                  u
  0.660     3                    u                                  4
  0.600     2                    u                                  u
  0.540     2                    u                                  u
  0.480     ‡                    3                                  u
  0.420     u                    u                                  u
  0.360     u                    u                                  u
  0.300     2                                                       2
  0.240     ‡                                                       u
  0.180     u                                                       2
  0.120     ‡                    u                                  u
 0.600e-01  u                    u                                  2
    0.      ‡                                                       u
   .........‡.........‡.........‡.........‡.........‡.........‡
          1.000      1.50       2.00       2.50       3.00       3.50
.$stop
_Z
```

REFERENCES

Amemiya, T., and Nold, F. (1975). *Review Econ. Statist. 57*, 255.
Andersen, E. B. (1980). "Discrete Statistical Models with Social Science Applications." North-Holland, Amsterdam.
Anderson, E. B. (1981). *Scand. J. Statist. 9*.
Andersen, E. B., and Madsen, M. (1977). *Psychometrika 42*, 357.
Anderson, T. W. (1954). *Psychometrika 19*, 1.
Anderson, T. W. (1955). *J. Amer. Statist. Assoc. 50*, 163.
Anderson, T. W. (1968). *In* "Latent Structure Analysis" (P. F. Lazarsfeld and N. W. Henry, eds.). Houghton Mifflin, Boston.
Anderson, T. W. (1969a). *In* "Essays in Probability and Statistics" (R. C. Bose *et al.*, eds.). University of North Carolina Press, Chapel Hill.
Anderson, T. W. (1969b). *In* "Multivariate Analysis II" (P. R. Krishnaiah, ed.). Academic Press, New York.
Anderson, T. W. (1973). *Ann. Statist. 1*, 135.
Anderson, T. W., and Hsiao, C. (1981). *J. Amer. Statist. Assoc. 76*, 598.
Anderson, T. W., and Rubin, H. (1956). *In* "Proc. Third Berkeley Symp. Math. Statist. Prob." p.111. University of California Press, Berkeley.
Baker, R. J., and Nelder, J. A. (1978). "The GLIM System, Release 3." Numerical Algorithms Group, Oxford.
Barndorff-Nielsen, O., and Cox, D. R. (1979). *J. Roy. Statist. Soc. Ser. B*, 279.
Berkson, J. (1950). *J. Amer. Statist. Assoc. 45*, 164.
Berlin, N. I., Waldman, T. A., and Weissman, S. M. (1959). *Physiological Reviews 39*, 577.
Bock, R. D., and Aitkin, M. (1981). *Psychometrika 46*, 443.
Bock, R. D., and Lieberman, M. (1970). *Psychometrika 35*, 179.
Boyles, R. A. (1980). *Report No. 13, Div. of Statistics*, University of California, Davis.
Boyles, R. A. (1982). *Technical Report*, Lawrence Livermore National Laboratory, Livermore, California.
Chen, C.-F. (1981). *J. Amer. Statist. Assoc. 76*, 704.
Davis, P. J., and Rabinowitz, P. (1975). "Methods of Numerical Integration." Academic Press, New York.
Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). *J. Roy. Statist. Soc. Ser. B 39*, 1
Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). *J. Amer. Statist. Assoc. 76*, 341.
Douglas, J. B. (1980). "Analysis with Standard Contagious Distributions." International Cooperative Publishing House, Fairland.
Evans, D. H. (1972). *Technometrics 14*, 23.
Evans, R. D. (1955). "The Atomic Nucleus." McGraw Hill, New York.
Fedorov, V. V. (1974). *Biometrika 61*, 49.
Feuerverger, A., and McDunnough, P. (1981). *J. Amer. Statist. Assoc. 76*, 379.
Green, P. J. (1982). unpublished.
Gupta, N. K., and Mehra, R. K. (1974). *IEEE Trans. Automatic Control AC-19*, 771.
Harville, D. A. (1977). *J. Amer. Statist. Assoc. 72*, 320.

Henderson, C. R., and Henderson, C. R. (1979). *Commun. Statist.-Theor. Math. A8*, 751.
Hinde, J. (1982). *In* "GLIM" (R. Gilchrist, ed.). Springer-Verlag, New York.
Huber, P. (1967). *In* "Proc. Fifth Berkeley Symp. Math. Statist. Prob." Vol. 1, 221.
International Committee for Standardization in Haematology (1972). *Amer. J. Clinical Pathology 58*, 71.
Jennrich, R. I., and Ralston, M. L. (1979). *Ann. Rev. Biophys. Bioeng. 8*, 195.
Jöreskog, K. G. (1973). *In* "Multivariate Analysis III" (P. R. Krishnaiah, ed.). Academic Press, New York.
Jöreskog, K. G., and Goldberger, A. S. (1975). *J. Amer. Statist. Assoc. 70*, 631.
Kailath, T. (1980). "Linear Systems." Prentice Hall, Englewood Cliffs, New Jersey.
Lawley, D. N., and Maxwell, A. E. (1971). "Factor Analysis as a Statistical Method," Second edition. Butterworth, London.
Loynes, R. M. (1980). *Ann. Statist. 8*, 285.
Macchi, O., and Picinbono, B. C. (1972). *IEEE Trans. Information Theory IT-18*, 562.
Muthén, B. (1979). *J. Amer. Statist. Assoc. 74*, 807.
Niederreiter, H. (1978). *Bull. Amer. Math. Soc. 84*, 957.
Orchard, T., and Woodbury, M. A. (1972). *In* "Proc. Sixth Berkeley Symp. Math. Statist. Prob." p.697. University of California Press, Berkeley.
Pollycove, M., and Tono, M. (1975). *Seminars in Nuclear Medicine 5*, 11.
Pregibon, D. (1981). *Ann. Statist. 9*, 705.
Pregibon, D. (1982). *Biometrics 38*, 485.
Preisler, H. K. (1977). Ph.D. Thesis, University of California, Berkeley.
Reid, D. D. (1981). *In* "Statistical Distributions in Scientific Work," Vol. 6 (C. Taillie *et al.*, eds.). Reidel.
Sanathanan, L., and Blumenthal, S. (1978). *J. Amer. Statist. Assoc. 73*, 794.
Snyder, D. L. (1975). "Random Point Processes." Wiley, New York.
Spjøtvoll, E. (1977). *Math. Oper. Statist., Ser. Statistics 8*, 69.
Steyn, H. S. (1976). *J. Amer. Statist. Assoc. 71*, 233.
Tukey, J. W. (1977). "Exploratory Data Analysis." Addison-Wesley, Reading.
Turner, T. R. (1982). Master of Statistics Project, University of New South Wales.
Wu, C.-F. (1983). *Ann. Statist. 11*, 95

# Studies in Econometrics, Time Series, and Multivariate Statistics

EDITED BY

SAMUEL KARLIN

*Department of Mathematics*
*Stanford University*
*Stanford, California*

TAKESHI AMEMIYA

*Department of Economics*
*Stanford University*
*Stanford, California*

LEO A. GOODMAN

*Departments of Statistics and Sociology*
*University of Chicago*
*Chicago, Illinois*

1983

ACADEMIC PRESS

*A Subsidiary of Harcourt Brace Jovanovich, Publishers*
New York London
Paris San Diego San Francisco São Paulo Sydney Tokyo Toronto