Regression Analysis of Poisson-Distributed Data

EDWARD L. FROME, MICHAEL H. KUTNER, and JOHN J. BEAUCHAMP*

The principle of maximum likelihood is used to obtain estimates of the parameters in a regression model when the experimental observations are assumed to follow the Poisson distribution. The maximum likelihood estimates are shown to be equivalent to those obtained by minimization of a quadratic form which reduces to a modified chi square under the Poisson assumption. Computationally, both of these estimation procedures are equivalent to a properly weighted least squares analysis. Approximate tests of the assumed Poisson variation and "goodness of fit" of the data to the model are proposed. Applications of the estimation procedure to linear and nonlinear regression models are discussed, and numerical examples are presented.

1. INTRODUCTION

1.1 Regression Analysis

Consider the general regression model

$$E(y_{ij}) = f(X_i, \theta), \quad \begin{matrix} i = 1, \cdots, N, \\ j = 1, \cdots, n_i, \end{matrix} \tag{1.1}$$

where $X_i = (x_{i1}, \cdots, x_{im})$ is the ith set of values of the m independent variables, n_i is the number of replications of the ith experimental condition, $\theta = (\theta_1, \cdots, \theta_p)'$ is a p-dimensional vector of unknown parameters, and $\{y_{ij}\}$, $i = 1, \cdots, N$, $j = 1, \cdots, n_i$, is a particular realization of the experiment. The regression function, $f(X, \theta)$, relates the expected value of the dependent variable to the independent variables and the parameters, and, given the experimental conditions and the data, we would like to estimate the unknown parameters. The most widely used methods of estimation have been developed using either the maximum likelihood or the least squares principle. The assumptions underlying these principles of estimation and the properties of the estimators have been given by Kendall and Stuart [14].

It is well known that maximum likelihood and least squares estimates are identical when the y_{ij}'s are independent and normally distributed with $E(y_{ij}) = f(X_i, \theta)$ and Var $(y_{ij}) = \sigma^2$. If $f(X, \theta)$ is linear in the unknown parameters, then the estimates are obtained using linear regression analysis. When $f(X, \theta)$ is not linear in all of the parameters, estimation is more difficult, and some type of iterative procedure will usually be required. An introductory account of nonlinear regression, and the extent to which standard results from linear estimation are applicable, have been given by Draper and Smith [8, Ch. 10]. Beauchamp [1] has presented a thorough discussion of generalized least squares estimation, and the properties of the estimates obtained using a weighted nonlinear regression procedure. The equivalence of maximum likelihood and least squares was demonstrated by Turner, Monroe, and Lucas [25] for a wide class of nonlinear models. They assumed independently normally distributed residuals with constant variance, or variance that depends on the independent variable in a known way.

Another situation in which maximum likelihood and least squares estimates are found to be equivalent is in the analysis of sensitivity and quantal response data, such as probit analysis. Here, the observations are assumed to be independent and to follow the binomial distribution with expectation given by equation (1.1), and it can be shown that maximum likelihood estimates are computationally equivalent to those obtained in a properly weighted least squares analysis (see [16, 17]).

1.2 Poisson Distribution

The Poisson distribution has only recently been considered in the context of regression analysis—see [12, Ch. 5]. Turner [24] established the equivalence of the maximum likelihood and weighted least squares estimator for simple linear regression through the origin, and Gart [11] has considered hypothesis testing. Multiple linear regression has been discussed by Jorgenson [13]. In what follows it will be assumed that the dependent variables in a regression analysis are counts that follow the Poisson distribution, and that the observations are independent with expectation as defined by equation (1.1). The regression model may represent, for example, the number of failures of a piece of equipment per unit time, the number of purchases of a particular commodity per family, or the number of bacteria per unit volume of suspension. It is further assumed that some general form of the model is known, that $f(X, \theta)$ is a differentiable function of θ, that N values of the independent variables are selected by the experimenter or specified by the situation, and that N is sufficiently greater than p to ensure estimability of the parameters.

In Section 2, the maximum likelihood principle will be used to estimate the parameters in the regression model.

* Edward L. Frome is assistant professor of statistics-operations research, Department of General Business, The University of Texas at Austin, Austin, Tex. 78712. Michael H. Kutner is assistant professor, Department of Statistics and Biometry, Emory University, Atlanta, Ga. 30322, and John J. Beauchamp is a statistician, Oak Ridge National Laboratory, Oak Ridge, Tenn. 37830. This research was supported by National Institute of Health Fellowship 5 F03 GM 45025, and by the U.S. Atomic Energy Commission under contract with the Union Carbide Corporation. The authors wish to acknowledge the assistance of the Emory University Computing Center in carrying out the computations, and to express their gratitude to the associate editor and referee for their critical comments and helpful suggestions.

Reprinted from: © Journal of the American Statistical Association
December 1973, Volume 68, Number 344
Theory and Methods Section
Pages 935-940

It will then be demonstrated that a properly weighted least squares analysis leads to the same iterative procedure that is obtained when the method of scoring is used to find a root of the likelihood equations. Minimum chi-square estimation will then be introduced, and we will show that minimizing an appropriately defined quadratic form results in an iterative method for obtaining best asymptotically normal estimates. It will then be demonstrated that, under certain conditions, the three methods of estimation are equivalent—in the sense that each results in the same iterative computational procedure. In Section 3, tests of the "goodness of fit" of the data to the model as well as the assumption of Poisson variation will be proposed. In Section 4, computational methods will be discussed, and examples of the application of these methods using multiple linear regression and nonlinear regression models will be presented.

2. ESTIMATION

2.1 Maximum Likelihood (ML)

The logarithm of the likelihood function of $\boldsymbol{\theta}$, given a particular realization of the experiment described in Section 1, is—neglecting a constant that does not involve the parameters—

$$\ln L(\boldsymbol{\theta}) = \sum_i [y_{i\cdot} \ln f(X_i, \boldsymbol{\theta}) - n_i f(X_i, \boldsymbol{\theta})], \quad (2.1)$$

where[1] $y_{i\cdot} = \sum_j y_{ij}$. The ML equations are obtained by differentiating (2.1) with respect to each of the parameters, i.e.,

$$\mathbf{G}(\boldsymbol{\theta}) = [\partial \ln L(\boldsymbol{\theta})/\partial\theta_r] = [\textstyle\sum_i (p_{ir}[\{y_{i\cdot}/f(X_i, \boldsymbol{\theta})\} - n_i])], \quad r = 1, \cdots, p, \quad (2.2)$$

where $p_{ir} = \partial f(X_i, \boldsymbol{\theta})/\partial\theta_r$. Since the ML equations will generally be nonlinear with respect to the unknown parameters, the method of scoring [19, p. 305] can be used to develop an algorithm to find a root of (2.2). This leads to the following system of equations:

$$\mathbf{C}(\boldsymbol{\theta}^0)\boldsymbol{\delta}^0 = \mathbf{G}(\boldsymbol{\theta}^0), \quad (2.3)$$

where $\mathbf{C}(\boldsymbol{\theta}^0)$ is the information matrix with elements

$$c_{rs} = \sum_i [p_{ir}p_{is}n_i/f(x_i, \boldsymbol{\theta})], \quad r, s = 1, \cdots, p, \quad (2.4)$$

evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0 = (\theta_1^0, \cdots, \theta_p^0)'$ (a set of initial values), and $\boldsymbol{\delta}^0 = (\boldsymbol{\theta} - \boldsymbol{\theta}^0)$. The system of equations in (2.3) is solved for $\boldsymbol{\delta}^0$, and new values of the parameters are obtained as $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0 + \boldsymbol{\delta}^0$. The procedure is repeated until a stable solution is reached.

2.2 Least Squares (LS)

Consider the following weighted sum of squares—which is to be minimized with respect to $\boldsymbol{\theta}$—

$$S(\boldsymbol{\theta}) = \sum_i w_i[z_i - f(X_i, \boldsymbol{\theta})]^2, \quad (2.5)$$

where $z_i = y_{i\cdot}/n_i$, and w_i^{-1}, $i = 1, \cdots, N$, is proportional to Var (z_i)—or is a consistent estimate of Var (z_i). $f(X, \boldsymbol{\theta})$ will in general be nonlinear in the unknown parameters, so we expand it about an initial estimate, $\boldsymbol{\theta}^0$, in a Taylor series through the linear terms. The resulting approximation is then substituted into (2.5) to obtain

$$\sum_i w_i[z_i - f(X_i, \boldsymbol{\theta}^0) - P_i^0\boldsymbol{\delta}^0]^2, \quad (2.6)$$

where $\boldsymbol{\delta}^0 = (\delta_1^0, \cdots, \delta_p^0)'$, and the vector P_i^0 is the ith row of $\mathbf{P}(\boldsymbol{\theta}^0)$, the $N \times p$ matrix of partial derivatives p_{ir} evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. The δ_r^0's are the only unknowns in (2.6), so we use the least squares principle to obtain estimates of these quantities by solving the following system of p linear equations:

$$\mathbf{P}(\boldsymbol{\theta}^0)'\mathbf{W}\mathbf{P}(\boldsymbol{\theta}^0)\boldsymbol{\delta}^0 = \mathbf{P}(\boldsymbol{\theta}^0)'\mathbf{W}[Z - F(\boldsymbol{\theta}^0)], \quad (2.7)$$

where $\mathbf{W} = \text{diag}\ (w_1, \cdots, w_N)$, $Z = (z_1, \cdots, z_N)'$, and $F(\boldsymbol{\theta}^0)$ denotes $F(\boldsymbol{\theta}) = [f(X_1, \boldsymbol{\theta}), \cdots, f(X_N, \boldsymbol{\theta})]'$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. We then obtain a revised parameter estimate, $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0 + \boldsymbol{\delta}^0$, replace the zero superscripts in (2.7) with ones, and solve for $\boldsymbol{\delta}^1$. This iterative process (Gauss-Newton method) continues until some convergence criteria are satisfied (see Section 4). Comparing (2.7) with (2.3) we see that if on each iteration we let $w_i = n_i/f(X_i, \boldsymbol{\theta}^k)$, $i = 1, \cdots, N$ (where $\boldsymbol{\theta}^k$ is the estimate of $\boldsymbol{\theta}$ obtained on the preceding iteration), then the two systems of linear equations are identical. Consequently, the iterative procedure for obtaining weighted LS estimates is computationally equivalent to using the method of scoring to find a root of the likelihood equations (provided, of course, the same initial estimate is employed).

2.3 Minimum Chi Square (MCS)

We now consider a method for obtaining best asymptotically normal (BAN) estimates—see [9, 18, 28, 29]. Let $Y_1, Y_2, \cdots, Y_n$, be a sequence of N-dimensional independent random vectors,[2] where $Y_j = (y_{1j}, \cdots, y_{Nj})'$ represents the outcome of the jth replication of the experimental conditions X_i, $i = 1, \cdots, N$ (see Section 1). The distribution of the Y's depends on the parameter $\boldsymbol{\theta}$, with expectation $E(Y|\boldsymbol{\theta}) = F(\boldsymbol{\theta})$, and variance

$$\mathbf{V}(\boldsymbol{\theta}) = E\{[Y - F(\boldsymbol{\theta})][Y - F(\boldsymbol{\theta})]'\}.$$

The quadratic form

$$n[Z_n - F(\boldsymbol{\theta})]'\mathbf{V}(\boldsymbol{\theta})^{-1}[Z_n - F(\boldsymbol{\theta})], \quad (2.8)$$

where $Z_n = n^{-1}(\sum_j y_{1j}, \cdots, \sum_j y_{Nj})'$, is called a chi square, and the value of $\boldsymbol{\theta}$ which minimizes it is called a MCS estimate. If $\mathbf{W}(Z_n)$ is a $p \times p$ positive-definite, symmetric matrix depending on Z_n only, then

$$n[Z_n - F(\boldsymbol{\theta})]'\mathbf{W}(Z_n)[Z_n - F(\boldsymbol{\theta})], \quad (2.9)$$

is called a modified chi square. Differentiating (2.9) with respect to each of the parameters and equating the results to zero yields

$$n\mathbf{P}(\boldsymbol{\theta})'\mathbf{W}(Z_n)[Z_n - F(\boldsymbol{\theta})] = \mathbf{0}, \quad (2.10)$$

[1] In all of the equations that follow, the subscript i will assume the values 1, $\cdots$, N, and the subscript j, which occurs only in expressions that involve y_{ij}, will assume the values 1, $\cdots$, n_i. As a notational convenience, only the index of summation will be indicated as a subscript on the summation operators, i.e., Σ_i and Σ_j.

[2] This corresponds to $n_i = n$ for $i = 1, \cdots, N$.

which is a linear form as defined in Ferguson's Theorem (see [28, Theorem 2]). If $\mathbf{W}(Z_n)$ converges in probability to $\mathbf{V}(\boldsymbol{\theta})^{-1}$ and certain regularity conditions are satisfied, then a root, say $\hat{\boldsymbol{\theta}}$, of (2.10) is BAN. To find $\hat{\boldsymbol{\theta}}$ we expand $F(\boldsymbol{\theta})$ in a first-order Taylor series about an initial estimate $\boldsymbol{\theta}^0$, substitute the resulting approximation into (2.10), and obtain

$$\mathbf{P}(\boldsymbol{\theta})'\mathbf{W}(Z_n)[Z_n - F(\boldsymbol{\theta}^0) - \mathbf{P}(\boldsymbol{\theta}^0)\boldsymbol{\delta}^0] = \mathbf{0}. \quad (2.11)$$

It can be shown [9] that if $\boldsymbol{\theta}^0$ is consistent, then replacing $\mathbf{P}(\boldsymbol{\theta})$ with $\mathbf{P}(\boldsymbol{\theta}^0)$ in (2.11) will result in a BAN estimate whenever the root to (2.10) is BAN. This leads to

$$\mathbf{P}(\boldsymbol{\theta}^0)'\mathbf{W}(Z_n)\mathbf{P}(\boldsymbol{\theta}^0)\boldsymbol{\delta}^0 = \mathbf{P}(\boldsymbol{\theta}^0)'\mathbf{W}(Z_n)[Z_n - F(\boldsymbol{\theta}^0)], \quad (2.12)$$

from which we obtain $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0 + \boldsymbol{\delta}^0$, and the procedure is repeated until convergence is obtained. If the elements of the Y's are mutually independent, then $\mathbf{V}(\boldsymbol{\theta})$ will be diagonal. If we take $\mathbf{W}(Z_n)$ to be a diagonal matrix whose elements are consistent estimates of the reciprocals of the variances, then the iterative procedure just described is identical to that described in Section 2.2, i.e., equations (2.7) and (2.12) are the same. If we further assume that the observations are Poisson, it follows that the iterative procedures for finding ML, LS, and MCS estimates are computationally equivalent when the method of scoring, the Gauss-Newton method, and the modified MCS procedures are employed. We are not able to show that the iterative procedure will converge, or that, if it converges, the root of the linear form is unique. The procedure described will produce a BAN estimate if the initial estimate, $\boldsymbol{\theta}^0$, is consistent—there is, however, no general method for finding consistent initial estimates as far as we know. Other iterative methods will also produce BAN estimates, and although these estimates are "asymptotically equivalent," there is no theory that indicates which method is "best" for small samples. We note that the results discussed in Section 1 are easily obtained using Section 2 results and the appropriate definition of $\mathbf{V}(\boldsymbol{\theta})$ under the binomial or normal assumption.

3. COVARIANCE MATRIX AND CHI SQUARE

The large sample covariance matrix of maximum likelihood estimators is the inverse of the information matrix, $\mathbf{C}(\boldsymbol{\theta})^{-1}$—see (2.4). If $\hat{\boldsymbol{\theta}}$ is a stable solution of the likelihood equations, then estimates of the elements of this matrix may be obtained by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$.

The expected number of counts for each value of X_i, $i = 1, \cdots, N$, may be estimated by $f_i = f(X_i, \hat{\boldsymbol{\theta}})$. Assuming the regression model and Poisson variation (see Section 1.2), then the statistic

$$Q_t = \sum_i \sum_j f_i^{-1}(y_{ij} - f_i)^2 \quad (3.1)$$

will be distributed approximately as a chi square with $D_t = \sum_i n_i - p$ degrees of freedom (d.f.). Q_t may be partitioned into two independent approximately chi square statistics, Q_w and Q_d, as follows:

$$Q_t = Q_w + Q_d = \sum_i \sum_j f_i^{-1}(y_{ij} - z_i)^2 + \sum_i n_i f_i^{-1}(z_i - f_i)^2, \quad (3.2)$$

where the d.f. for Q_w and Q_d are $D_w = \sum_i n_i - N$ and $D_d = N - p$, respectively. If the value of the statistic Q_t is found to be significantly large, then this may be attributed to either heterogeneity of variance or "lack of fit" of the regression model, or both. In this situation, Q_w may be compared with the chi-square distribution with D_w d.f. If Q_w is significantly large (indicating heterogeneity of variance), then the ratio $(Q_d/D_d)/(Q_w/D_w)$ may be compared with the F distribution (approximate test). A significantly large value of this ratio indicates the lack of fit of the model. If the regression model is not rejected, then the presence of heterogeneity of variance is taken into account by multiplying the elements of the estimated covariance matrix by the factor Q_w/D_w.

Although the preceding results are only approximate, they may provide some insight into the source of errors in an experiment. In practice, it may be advisable to perform a preliminary test for heterogeneity of variance by calculating Fisher's index of dispersion [12, Ch. 5] for each experimental condition, i.e., for each value of X_i compute $\sum_i z_i^{-1}(y_{ij} - z_i)^2$, which is approximately a chi square with $(n_i - 1)$ d.f., $i = 1, \cdots, N$. If these approximate chi squares indicate that the Poisson assumption is reasonable, then the method of estimation described in Section 2.1 can be used. If a significant deviation from Poisson variability is apparent and cannot be attributed to known causes, then the least squares procedure of Section 2.2 can be used with fixed weights (rather than iterative weights) defined by $w_i = n_i/s_i^2$, where s_i^2 is the sample variance. If the conditions given in Section 2.3 are satisfied, then the weighted least squares estimates will be BAN.

4. APPLICATIONS

4.1 Computations

Various approaches—based on gradient methods—to the numerical problem encountered in nonlinear regression have been discussed by Smith and Shannon [22]. A gradient method is one which calculates (on each iteration) a search vector, $D^k = (d_1^k, \cdots, d_p^k)'$, defined by

$$D^k = \mathbf{A}(\mathbf{WP})'[Z - F(\boldsymbol{\theta})], \quad (4.1)$$

where $\mathbf{A}$ is a $p \times p$ matrix, and the right side of (4.1) is evaluated at $\boldsymbol{\theta}^k$ (the current estimate of $\boldsymbol{\theta}$). We then let $\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + D^k$, and the foregoing procedure is repeated until some convergence criterion is satisfied. If $\mathbf{A} = (\mathbf{P}'\mathbf{WP})^{-1}$, then D is the Gauss-Newton vector and, if good starting values are available this approach has desirable convergence properties. When good initial estimates are not available, some other search procedure (e.g., Marquardt's [15]) may be useful. Marquardt's vector combines the best features of the steepest descent and Gauss-Newton methods and avoids singularities that

may occur when $\mathbf{P}'\mathbf{W}\mathbf{P}$ is ill-conditioned. Another approach to function optimization that uses a random search technique has been developed by Bremermann [4]. This method may be of value when there are a large number of parameters and initial estimates are not available. A method for generating (pseudo random) normal deviates—which are required in Bremermann's program—has been discussed by Chen [5].

In the applications that follow, methods for finding starting values are given, and the Gauss-Newton procedure is used (see Section 2.2). If a stable solution, $\hat{\boldsymbol{\theta}}$, is found, it will be a root of the likelihood equations when the observations are counts that follow the Poisson distribution. The starting values, final estimates of the parameters, number of iterations to convergence, and the estimated covariance matrix are given for each example. Convergence is defined to have occurred when the relative change in each of the parameters is less than 10^{-5}. The computations were carried out in single precision arithmetic, and the Gauss-Jordan method was used for matrix inversion (see [27, Ch. 6]).

4.2 Linear Regression

A special case of (1.1) occurs when the expected number of counts can be represented as a linear combination of the parameters and the independent variables. The regression model is $F(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$, where $\mathbf{X}$ is an $N \times p$ matrix with rows X_i. In this situation, the matrix of partial derivatives is $\mathbf{P} = \mathbf{X}$, and we take

$$\mathbf{W} = \text{diag } (n_i/X_i\boldsymbol{\theta}), \quad i = 1, \cdots, N.$$

Example 1: Simple Linear Regression. The problem of estimating the concentration of viruses or bacteria from dilution counts has been discussed by Gart [11]. Here we have $m = p = 1$, and $E(y_{ij}) = x_i\theta$, where x_i is the ith dilution factor for which n_i parallel counts are made, and y_{ij} is the observed colony count for the jth plate of the ith dilution. The parameter θ represents the mean particle density per unit volume of suspension. The ML estimator is easily obtained from (2.2) as

$$\hat{\theta} = \sum_i \sum_j y_{ij} / \sum_i n_i x_i,$$

with variance Var $(\hat{\theta}) = \theta / \sum_i n_i x_i$. Roberts and Coote [20] have discussed the interpretation of the chi-square statistics (see Section 3), and have shown how a further partitioning can be achieved. This is the only situation considered that results in a closed form expression for the ML estimator. Therefore, numerical examples will be presented to illustrate further areas of application of the methods proposed in Sections 2 and 3.

Example 2: Multiple Linear Regression. Consider the linear model with $m = p = 2$, so that $X_i = (x_{i1}, x_{i2})'$, where x_{i1} and x_{i2} are the times spent in regimes one and two during the ith cycle of operation of a piece of electronic equipment, and y_i is the number of failures that occur during the ith cycle. Using the data from Jorgenson ([13, Table 1]), and the initial estimates $\theta_1^0 = .1725$ and $\theta_2^0 = .0657$, we obtain $\hat{\theta}_1 = .1666$ and $\hat{\theta}_2 = .0904$ after five iterations. The estimated elements of the parameter covariance matrix are var $(\hat{\theta}_1) = .001216$, var $(\hat{\theta}_2) = .00397$, cov $(\hat{\theta}_1, \hat{\theta}_2) = -.001956$; and $S(\hat{\boldsymbol{\theta}}) = 8.897$. The multiple linear regression model of the Poisson process has also been used by Weber [26] in the study of accident rate potential.

4.3 Nonlinear Regression

In Section 4.2, we saw that the matrix of partial derivatives of $f(X, \boldsymbol{\theta})$ did not depend on $\boldsymbol{\theta}$. In nonlinear models the partial derivatives will involve the parameters as well as the independent variables, and consequently the matrix $\mathbf{P}$ is reevaluated after each iteration. Other than this the computations are identical to those used in the multiple linear regression case. To clarify the difference, and to demonstrate the general procedure that is followed for nonlinear models, examples from two different areas of application are now considered.

Example 3: Simple Exponential Regression. In this example the independent variable $X_i = x_i$ is time $(m = 1)$, and the counts describe the decay of the neutron-density in a medium size assembly of beryllium with $f(X_i, \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 x_i)$, $i = 1, \cdots, N$. The data used here is taken from Cornell ([7, p. 107]), and we use as preliminary estimates of the parameters the values obtained by Cornell using the method of partial totals: $\theta_1^0 = 100043$ and $\theta_2^0 = .2539$. The matrix of partial derivatives in this case is calculated using $p_{i1} = \exp(-\theta_2 x_i)$ and $p_{i2} = -x_i\theta_1 \exp(-\theta_2 x_i)$, $i = 1, \cdots, N$. The iterative procedure converged after two iterations, indicating that the partial total estimates were excellent first approximations. The final values for the parameters are $\hat{\theta}_1 = 100099$, $\hat{\theta}_2 = .2537$, and the estimated covariance matrix is

$$\mathbf{C}(\hat{\boldsymbol{\theta}})^{-1} = \begin{bmatrix} 42882 & .06169 \\ .06169 & .18827 \times 10^{-6} \end{bmatrix},$$

and $S(\hat{\boldsymbol{\theta}}) = 13.208$.

The application of maximum likelihood estimation to the analysis of the multi-exponential function, which is frequently used in tracer kinetic studies, has been discussed by Sandor *et al.* [21]. In the more general case of simultaneous compartmental analysis, the partial totals method of Beauchamp and Cornell [2, 3] may be used to obtain consistent initial estimates of the parameters.

In the preceding examples, the number of observations for a given value of the independent variable was one (i.e., $n_i = 1$, $i = 1, \cdots, N$). Consequently, the chi square statistic in Section 3 [equation (3.2)] is given by $Q_t = S(\hat{\boldsymbol{\theta}})$, with d.f. $= N - p$. In the next example, we will have $n_i > 1$, $i = 1, \cdots, N$, so that both the assumed distribution and the model may be tested independently using the procedure presented in Section 3.

Example 4: Multi-target Survival Curve. In this example, the observed response is the number of colonies produced in the spleen of recipient animals—by bone marrow cells from irradiated donor animals. A quantitative model of

the effect of radiation damage on stem cell survival is given by

$$E(y_{ij}) = f(X_i, \theta) = \theta_1 x_{i1}[1 - \{1 - \exp(-\theta_2 x_{i2})\}^{\theta_3}],$$

where x_{i1} is the concentration of injected cells, x_{i2} is the radiation dose, and θ_1, θ_2, and θ_3 are parameters of biological interest. This is known as the multi-target survival curve and is a special case of a more general model derived from the "target-hit" statistical theory of radiation damage. The application of the maximum likelihood estimation procedure and interpretation of the results have been presented in detail by Frome and Beauchamp [10]. Further application and discussion of the chi-square partitioning have been presented by Comas [6]. Using the data from Till and McCulloch [23] (see the table) and graphically obtained starting values ($\theta_1^0 = 8.0$, $\theta_2^0 = .01$, and $\theta_3^0 = 3.1$), the estimates of the parameters $\hat{\theta}_1 = 7.64$, $\hat{\theta}_2 = .00934$, and $\hat{\theta}_3 = 2.892$, were obtained after six iterations. The estimated covariance matrix is

$$\mathbf{C}(\hat{\theta})^{-1} = \begin{bmatrix} .8206 & -.1239 \times 10^{-3} & -.5017 \\ & .1590 \times 10^{-6} & .2544 \times 10^{-3} \\ & & .5589 \end{bmatrix}.$$

The chi-square values are calculated using equation (3.2) and are $Q_w = 24.44$ with 49 d.f. and $Q_d = 7.595$ with 4 d.f.

SURVIVAL CURVE DATA

i	Concentration[a] x_{i1}	Radiation (rads) x_{i2}	n_i	Number of colonies counted y_{ij}	z_i	$f(X_i, \hat{\theta})$
1	1.25	0	6	11 10 11 11 9 8	10.000	9.546
2	1.75	96	7	12 8 9 9 8 9 11	9.429	10.429
3	3.00	192	4	11 10 11 14	11.500	9.375
4	7.20	288	9	8 8 9 12 6 10 13 10 6	9.111	10.114
5	24.00	432	11	12 12 14 10 7 10 8 11 8 7 6	9.546	9.216
6	75.00	576	15	7 5 9 4 9 10 7 8 9 7 12 7 11 7 11	8.200	7.596
7	120.00	672	4	2 3 3 4	3.000	4.974

[a] Unit concentration of colony forming units = 10^4 bone marrow cells.
Source: Data from [23].

In presenting the applications in this section, we have attempted to demonstrate—with simple numerical examples—general areas of application of the method of estimation presented in Section 2. Since it has been our intention to demonstrate the general nature of the computational procedure and its relation to well-known statistical methods, we have chosen examples from several different disciplines. We have not attempted to discuss substantive aspects of the problems presented, such as selecting the regression model, experimental design (choosing the X_i's), or interpretation of the results. More detailed discussion of these problems and additional applications can be found in the cited references.

[*Received March 1972. Revised February 1973.*]

REFERENCES

[1] Beauchamp, John J., *Simultaneous Nonlinear Estimation*, Ph.D. dissertation, Florida State University, 1966.

[2] ——— and Cornell, R.G., "Simultaneous Estimation by Partial Totals for Compartmental Models," *Journal of the American Statistical Association*, 63 (June 1968), 573–83.

[3] ——— and Cornell, R.G., "Spearman Simultaneous Estimation for a Compartmental Model," *Technometrics*, 11 (August 1969), 551–60.

[4] Bremermann, H., "A Method of Unconstrained Global Optimization," *Mathematical Biosciences*, 9 (December 1970), 1–15.

[5] Chen, E.H., "Random Normal Number Generator for 32-bit Word Computers," *The Journal of the American Statistical Association*, 66 (June 1971), 400–03.

[6] Comas, F., "The Radiosensitivity of Rat Marrow Cells," *International Journal of Radiation Biology*, 17, No. 6 (1970), 549–57.

[7] Cornell, R.G., "A Method for Fitting Linear Combinations of Exponentials," *Biometrics*, 18 (March 1962), 104–13.

[8] Draper, N.R. and Smith, H., *Applied Regression Analysis*, New York: John Wiley & Sons, Inc., 1966.

[9] Ferguson, T.S., "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *The Annals of Mathematical Statistics*, 29 (December 1968), 1046–58.

[10] Frome, E.L. and Beauchamp, J.J., "Maximum Likelihood Estimation of Survival Curve Parameters," *Biometrics*, 24 (September 1968), 595–605.

[11] Gart, J.J., "The Analysis of Poisson Regression with an Application in Virology," *Biometrika*, 51 (December 1964), 517–21.

[12] Haight, F.A., *Handbook of the Poisson Distribution*, New York: John Wiley & Sons, Inc., 1967.

[13] Jorgenson, D.W., "Multiple Regression Analysis of a Poisson Process," *Journal of the American Statistical Association*, 56 (June 1961), 235–45.

[14] Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics, Vol. 2*, New York: Hafner Publishing Company, 1946.

[15] Marquardt, D.W., "An Algorithm for Least Squares Estimation of Nonlinear Parameters," *Journal of Society of Applied Mathematics*, 11 (June 1963), 431–41.

[16] Moore, R.H. and Zeigler, R.K., "The Use of Nonlinear Regression Methods for Analyzing Sensitivity and Quantal Response Data," *Biometrics*, 23 (September 1967), 563–67.

[17] Nelder, J.A., "Weighted Regression, Quantal Response Data, and Inverse Polynomials," *Biometrics*, 13 (February 1968), 979–85.

[18] Neyman, J., "Contributions to the Theory of the χ^2 Test," *Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, 1949.

[19] Rao, C.R., *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons, Inc., 1965.

[20] Roberts, E.A. and Coote, G.C., "The Estimation of Concentration of Viruses and Bacteria from Dilution Counts," *Biometrics*, 21 (December 1965), 600–15.

[21] Sandor, T., Conroy, M.F. and Hollenberg, N.K., "The Application of the Method of Maximum Likelihood to the Analysis of Tracer Kinetic Data," *Mathematical Biosciences*, 9 (December 1970), 149–59.

[22] Smith, F.B. and Shannon, D.F., "An Improved Marquardt Procedure for Nonlinear Regression," *Technometrics*, 13 February 1971), 63–74.

[23] Till, J.E. and McCulloch, E.A., "A Direct Measurement of Radiation Sensitivity of Normal Mouse Bone Marrow Cells," *Radiation Research*, 14 (February 1961), 213–22.
[24] Turner, M.E., "Straight Line Regression Through the Origin," *Biometrics*, 16 (September 1960), 483–85.
[25] ———, Monroe, R.J. and Lucas, H.L., "Generalized Asymptotic Regression and Non-linear Path Analysis," *Biometrics*, 17 (March 1961), 120–43.
[26] Weber, D.C., "Accident Rate Potential: An Application of Multiple Regression Analysis of a Poisson Process," *Journal of the American Statistical Association*, 66 (June 1971), 431–41.
[27] Weeg, G.P. and Reed, G.D., *Introduction to Numerical Analysis*, Waltham, Massachusetts: Blaisdell Publishing Company, 1966.
[28] Wijsman, R.A., "On the Theory of BAN Estimates," *The Annals of Mathematical Statistics*, 30 (March 1959), 185–91.
[29] ——— "Correction on 'On the Theory of BAN Estimates,'" *The Annals of Mathematical Statistics*, 30 (December 1959), 1268–70.