RT-MAE-8923

AN ITERATIVE PROCEDURE FOR THE MSAE ESTIMATION OF PARAMETERS IN A DOSE-RESPONSE MODEL

by

C.D.S. André, C.A. Peres

and

S.C. Narula

Palavras Chaves: Exponential model; least squares; maximum
(Key words) likelihood; nonlinear regression; radiobiology

Classificação AMS: 62F10, 62F11, 62P10
(AMS Classification)

AN ITERATIVE PROCEDURE FOR THE MSAE ESTIMATION OF PARAMETERS IN A DOSE-RESPONSE MODEL

Carmen D. S. André
University of São Paulo
São Paulo, Brazil

Clovis A. Peres
University of São Paulo
São Paulo, Brazil

Subhash C. Narula
Virginia Commonwealth University
Richmond, Virginia 23284

ABSTRACT

The least squares estimates of the parameters in the multistage dose-response model are unduly affected by outliers in a data set whereas the minimum sum of absolute errors, MSAE estimates are more resistant to outliers. Algorithms to compute the MSAE estimates can be tedious and computationally burdensome. We propose a linear approximation for the dose-response model that can be used to find the MSAE estimates by a simple and computationally less intensive algorithm. A few illustrative examples show that we get comparable values of the MSAE estimates of the parameters in a dose-response model using the exact model and the linear approximation.

1. INTRODUCTION

Let y_i denote the value of the response variable corresponding to dose d_i. The multistage dose-response model

$$y_i = 1 - \exp\left(-\Sigma_{j=0}^{k} \alpha_j d_i^j\right) + e_i, \quad i=1,\ldots,n, \tag{1}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_k$ denote the nonnegative unknown parameters and e_i denotes the unobservable random error, was developed by

Armitage and Doll (1954) to assess the risk of exposure to chemicals and pollution agents. The model is based on the assumption that the mechanism of carcinogenesis can be expressed as a series of k mutations at the cellular level. The model can also be used in radiobiological areas, Peres and Narula (1989) where $1 - y$ represents the survival probability corrected by the natural survival at α_0 and radiation dose equal to zero.

The maximum likelihood and the least squares methods are often used to estimate the parameters of the model in (1). These estimates may be unduly affected by outliers. As an illustration, consider the data in the first three columns of Table I. These data are taken from Table II of Sankaranarayanan (1969b) and represent the effect of nitrogen post-treatment on mortality of Drosophilia eggs irradiated as stage-7 oocytes. In Column 4, we have replaced the value of y_5 by y_3, i.e., changed y_5 from 0.104 to 0.249. The least squares, LS estimates of the parameter of the model for the original and the altered data are given in Table I.

TABLE I

The Data and the Estimates of the Parameter for the Model $y = \exp(-\alpha d) + e$.

i	d_i	Original y_i	Altered y_i
1	0.15	0.625	0.625
2	0.30	0.407	0.407
3	0.45	0.249	0.249
4	0.60	0.157	0.157
5	0.75	0.104	0.249
LS Estimate		3.0620	2.8416
MSAE Estimate		3.0858	3.0848

Clearly, the values of the two estimates are quite different. Therefore, it is desirable to develop alternative estimation procedures that are more resistant to outliers.

During the last two decades it has been recognized that the minimum sum of absolute errors, MSAE estimates of the parameters in the multiple linear regression model are not unduly affected by the presence of outliers, Huber (1974) and Narula and Wellington (1985). The MSAE estimates of the parameter for the original and the altered data are given in Table I. The values of the two estimates are the same.

The model in (1) is intrinsically nonlinear. A number of algorithms have been proposed to compute the MSAE estimates of the parameters in a general linear regression model. However, most of these algorithms are tedious and require intensive computations. Since the model in (1) has a special form, we can approximate it by a linear model that can be solved by a simpler and computationally less intensive algorithm.

Our objective in this paper is to present an easy to understand and effective algorithm to compute the MSAE estimates of the parameters in a multistage dose-response model. The rest of the paper is organized as follows: In Section 2, we describe an algorithm to estimate the parameters using model in (1). In Section 3, we give a linear approximation for the dose-response model and develop an algorithm to compute the MSAE estimates of the parameters in model (1) using the approximation. In Section 4, we compare the MSAE estimates of the parameters in model (1) using the two procedures. We conclude the paper with a few remarks in Section 5.

2. AN ALGORITHM USING THE EXACT MODEL

Schlossmacher (1973) proposed an iterative weighted least squares procedure to compute the MSAE estimates of the parameters of a multiple linear regression model. He commented that one may also use this procedure to compute the MSAE estimates of the parameters in the nonlinear regression models.

The basic idea of the algorithm can be stated as follows: Our objective is to minimize

$$G(\alpha) = \Sigma_{i=1}^{n} |e_i| .$$

This objective can be achieved by minimizing

$$G_w(\alpha) = \Sigma_{i=1}^{n} w_i e_i^2 \qquad (2)$$

where $w_i = 1/|e_i|$, $i = 1, \ldots, n$. However, w_i's are not known at the start of the procedure. Therefore, one may use the following iterative procedure to minimize $G_w(\alpha)$.

Step 1: Set $m = 0$ and solve the weighted least squares problem (2) using $w_i(m) = 1$, $i = 1, \ldots, n$. Compute

$$r_i(m) = 1 - y_i - \exp\left(-\Sigma_{j=0}^{k} \hat{\alpha}_j(m)\, d_i^j \right)$$

the observed residuals at the mth iteration.

Step 2: Solve the weighted least squares problem (2) using $w_i(m+1) = 1/|r_i(m)|$, $i = 1, \ldots, n$; if any $r_i(m) = 0$, set $w_i(m+1) = 0$. Set $m = m + 1$ and compute $r_i(m)$.

Step 3: If $|r_i(m+1) - r_i(m)| = 0$, $i = 1, \ldots, n$, stop; otherwise, go to Step 2.

It is clearly an iterative algorithm. Furthermore, since (1) is a nonlinear model, in Steps 1 and 2 we have to use some iterative procedure to determine the weighted least squares estimates of the parameters. Therefore, the algorithm consists of two nested iterative procedures requiring intensive computations.

3. AN ALGORITHM USING THE APPROXIMATE MODEL

Peres and Narula (1989) have shown that it is possible to approximate a dose-response model by a linear model. Clearly, once the model has been linearized, we can avoid the iterative procedure required to solve the nonlinear model in Steps 1 and 2 of the algorithm in Section 2; thus making the algorithm computationally less intensive.

To compute the MSAE estimates of the parameters in (1), we observe that

$$|e_i| = |z_i - g(d_i, \alpha)|$$

where $z_i = 1 - y_i$ and $g(d_i, \alpha) = \exp\left(-\Sigma_{j=0}^{k} \alpha_j d_i^j\right)$.

Thus

$$|e_i| = |z_i\{1 - g(d_i, \alpha)/z_i\}|$$
$$= |z_i\{1 - \exp(-(\ln z_i - \ln g(d_i, \alpha)))\}|.$$

When model (1) is the correct model, then $|\ln z_i - \ln g(d_i, \alpha)| < 1$, and, in fact, it is close to zero. Therefore, if we expand $\{1 - \exp(-(\ln z_i - \ln g(d_i, \alpha)))\}$ by the Taylor series expansion around $\ln z_i - \ln g(d_i, \alpha) = 0$ and retain only the first term, Peres and Narula (1989) have shown that

$$|e_i| = |z_i(\ln z_i - g(d_i, \alpha))|, \; i = 1,\ldots,n.$$

However, to obtain even better approximation, we may include the second term in the approximation which gives us

$$|e_i| = |v_i(\ln z_i - \ln g(d_i, \alpha)|, \; i = 1,\ldots,n \qquad (3)$$

where $v_i = z_i \{1 - (\ln z_i - (\ln g(d_i, \alpha))/2)\}$. Using (3), our objective is to find α's that minimize

$$G^*(\alpha) = \Sigma_{i=1}^{n} |v_i e_i^*|, \qquad (4)$$

where $e_i^* = \ln z_i - \ln g(d_i, \alpha) = \ln z_i + \Sigma_{j=0}^{k} \alpha_j d_i^j$. Note that e_i^*'s denote the errors from the linear model

$$\ln z_i = \ln g(d_i, \alpha) + e_i^*,$$
$$= -\Sigma_{j=0}^{k} \alpha_j \, d_i^j + e_i^*. \tag{5}$$

Using the linear model (5), we can compute the MSAE estimates of the parameters of model in (1) by minimizing

$$G_w^* (\alpha) = \Sigma_{i=1}^{n} w_i \; v_i^2 \; (e_i^*)^2 , \tag{6}$$

where $w_i = 1/|e_i|$, $i = 1,\ldots,n$. Since w_i's and v_i's are not known, we can use the following iterative procedure:

Step 1: Set $m = 0$ and solve the weighted least squares problem (6) using $w_i(m) = 1$, and $v_i(m) = z_i$, $i = 1,\ldots,n$. Compute

$$r_i^*(m) = \ln z_i + \Sigma_{j=0}^{k} \hat{\alpha}_j(m) \; d_i^j,$$

and

$$r_i(m) = z_i - \exp(-\Sigma_{j=0}^{k} \hat{\alpha}_j(m) \; d_i^j),$$

the observed residuals for models in (5) and (1), respectively.

Step 2: Solve the weighted least squares problem (6) using $w_i(m+1) = z_i \;(1 - r_i^*(m)/2)$, $i=1,\ldots,n$; if any $r_i(m) = 0$, set $w_i(m+1) = 0$. Set $m = m + 1$ and compute $r_i(m)$ and $r_i^*(m)$ as in Step 1.

Step 3: If $|r_i(m+1) - r_i(m)| \cong 0$, $i=1,\ldots,n$, stop; otherwise, go to Step 2.

The preceding algorithm is also iterative. However, in Steps 1 and 2 we solve the linear model (5) using any weighted least squares regression procedure; thus making it computationally less demanding than the algorithm in Section 2.

4. COMPUTATIONAL EXPERIENCE

The algorithm of Section 2 was implemented on the Burrough B-6900 computer at the Centro de Computacão Electronica de USP using BMDP-3R routine. The algorithm of Section 3 was implemented in Pascal on an IBM compatible microcomputer.

To compare the MSAE estimates of the parameters of a multistage dose-response model obtained by using the exact model (1) and the approximate linear model (5), we computed the MSAE estimates for a few data sets taken from Sankaranarayanan (1969a, 1969b). The results for the model $y_i = \exp(-\alpha d_i) + e_i$, are summarized in Table II, and for the model $y_i = \exp(-\alpha_1 d_i - \alpha_2 d_i^2) + e_i$ in Table III.

From Tables II and III we observe that the MSAE estimates obtained using the exact model and the linear approximation are comparable.

TABLE II

The MSAE Estimates of the Parameter α in the Model $y_i = \exp(-\alpha d_i) + e_i$ and the Sum of Absolute Errors Using the Exact Model and the Linear Approximation

Data Set*	Model Used	MSAE Estimate of α	Sum of Absolute Error
Nitrogen Post-treatment Table II	Exact Model	3.0858	0.00417
	Linear Approx.	3.0858	0.00417
Oxygen Post-treatment Table II	Exact Model	2.5295	0.00943
	Linear Approx.	2.5295	0.00943

*Table number refers to Table in Sankaranarayanan (1969b).

TABLE III

The MSAE Estimates of the Parameters α_1 and α_2 in the Model $y_i = \exp(-\alpha_1 d_i - \alpha_2 d_i^2) + e_i$ and the Sum of Absolute Errors for the Data From Table II of Sankaranarayanan (1969a) Using the Exact Models and the Linear Approximation

Model Used	MSAE Estimate of α_1	MSAE Estimate of α_2	Sum of Absolute Errors
Exact Model	0.21235	0.07536	0.00822
Linear Approx.	0.21601	0.07434	0.00807

4. CONCLUDING REMARKS

The MSAE estimates of the parameters in a multistage dose-response model are more resistant to outliers than the least squares estimates. We have shown how we can obtain these estimates by a simple and computationally less intensive algorithm using a linear approximation for the model. Furthermore, the algorithm can be implemented by using any program for solving a weighted least squares multiple linear regression problem or by appropriately modifying the data and using a program for least squares linear regression problem.

ACKNOWLEDGMENTS

We thank Prof. Gabriela Stangenhaus of the University of Campinas for helpful suggestions during the research. We also thank FAPESP and FINEP for the grants that partially supported the work of Prof. Narula.

BIBLIOGRAPHY

Armitage, P. and Doll, R. (1954). The age distribution of cancer and a multistage theory of carcinogenesis. British Journal of Cancer, 8, 1-12.

Huber, P. J. (1974). Comment on "Adaptive robust procedures: A partial review and some suggestions for future applications and theory," by R. V. Hogg, J. of the American Statistical Association, 69, 926-927.

Narula, S. C. and Wellington, J. F. (1985). Interior analysis for the minimum sum of absolute errors regression. Technometrics, 27, 181-188.

Peres, C. A. and Narula, S. C. (1989). An approximate closed form solution for a quadratic dose-response model. Communications in Statistics - Simulation and Computation, 18, 643-653.

Sankaranarayanan, K. (1969a). The effect of oxygen and nitrogen post-treatments on the mortality of Drosophilia eggs irradiated as stage-7 oocytes. Mutation Research, 7, 357-368.

Sankaranarayanan, K. (1969b). The effect of oxygen and nitrogen post-treatments on the survival of irradiated stage-14 oocytes and possible basis for sensitivity differences between stage-7 and stage-14 oocytes of Drosophilia Melanogaster. Mutation Research, 7,, 369-383.

Schlossmacher, E. J. (1973). An iterative technique for absolute deviations curve fitting. Journal of the American Statistical Association, 68, 857-859.

RELATÓRIO TÉCNICO
DO
DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS EM 1987

8701 - ACHCAR, J.A. & BOLFARINE, H.; Constant Hazard Against a Change-Point Alternative: A Bayesian Approach with Censored Data, São Paulo, IME-USP, 1987, 20p.

8702 - RODRIGUES, J.; Some Results on Restricted Bayes Least Squares Predictors for Finite Populations, São Paulo, IME-USP, 1987, 16p.

8703 - LEITE, J.G., BOLFARINE, H. & RODRIGUES, J.; Exact Expression for the Posterior Mode of a Finite Population Size: Capture-Recapture Sequential Sampling, São Paulo, IME-USP, 1987, 14p.

8704 - RODRIGUES, J., BOLFARINE, H. & LEITE, J.G.; A Bayesian Analysis in Closed Animal Populations from Capture Recapture Experiments with Trap Response, São Paulo, IME-USP, 1987, 21p.

8705 - PAULINO, C.D.M.; Analysis of Categorical Data with Full and Partial Classification: A Survey of the Conditional Maximum Likelihood and Weighted Least Squares Approaches, São Paulo, IME-USP, 1987, 52p.

8706 - CORDEIRO, G.M. & BOLFARINE, H.; Prediction in a Finite Population under a Generalized Linear Model, São Paulo, IME-USP, 1987, 21p.

8707 - RODRIGUES, J. & BOLFARINE, H.; Nonlinear Bayesian Least-Squares Theory and the Inverse Linear Regression, São Paulo, IME-USP, 1987, 15p.

8708 - RODRIGUES, J. & BOLFARINE, H.; A Note on Bayesian Least-Squares Estimators of Time-Varying Regression Coefficients, São Paulo, IME-USP, 1987, 11p.

8709 - ACHCAR, J.A., BOLFARINE, H. & RODRIGUES, J.; *Inverse Gaussian Distribution: A Bayesian Approach*, São Paulo, IME-USP, 1987, 20p.

8710 - CORDEIRO, G.M. & PAULA, G.A.; *Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models*, São Paulo, IME-USP, 1987, 26p.

8711 - SINGER, J.M.; PERES, C.A.& HARLE, C.E.; *On the Hardy-Weinberg Equilibrium in Generalized ABO Systems*, São Paulo, IME-USP, 1987, 16p.

8712 - BOLFARINE, H. & RODRIGUES, J.; *A Review and Some Extensions on Distribution Free Bayesian Approaches for Estimation and Prediction*, São Paulo, IME-USP, 1987, 19p.

8713 - RODRIGUES, J.; BOLFARINE, H. & LEITE, J.G.; *A Simple Nonparametric Bayes Solution to the Estimation of the Size of a Closed Animal Population*, São Paulo, IME-USP, 1987, 11p.

8714 - BUENO, V.C.; *Generalizing Importance of Components for Multistate Monotone Systems*, São Paulo, IME-USP, 1987, 12p.

8801 - PEREIRA, C.A.B. & WECHSLER, S.; *On the Concept of P-value*, São Paulo, IME-USP, 1988, 22p.

8802 - ZACKS, S., PEREIRA, C.A.B. & LEITE, J.G.; *Bayes Sequential Estimation of the Size of a Finite Population*, São Paulo,IME-USP, 1988, 23p.

8803 - BOLFARINE, H.; *Finite Population Prediction Under Dynamic Generalized Linear Models*, São Paulo, IME-USP, 1988, 21p.

8804 - BOLFARINE, H.; *Minimax Prediction in Finite Populations*, São Paulo, IME-USP, 1988, 18p.

8805 - SINGER, J.M. & ANDRADE, D.F.; *On the Choice of Appropriate Error Terms for Testing the General Linear Hypothesis in Profile Analysis*, São Paulo, IME-USP, 1988, 23p.

8806 - DACHS, J.N.W. & PAULA, G.A.; *Testing for Ordered Rate Rations in Follow-up Studies with Incidence Density Data*, São Paulo, IME-USP, 1988, 18p.

8807 - CORDEIRO, G.M. & PAULA, G.A.; *Estimation, Significance Tests and Diagnostic Methods for the Non-Exponential Family Nonlinear Models*, São Paulo, IME-USP, 1988, 29p.

8808 - RODRIGUES, J. & ELIAN, S.N.; The Coordinate - Free Estimation in Finite Population Sampling, São Paulo, IME-USP, 1988, 5p.

8809 - BUENO, V.C. & CUADRADO, R.Z.B.; On the Importance of Components for Continuous Structures, São Paulo, IME-USP, 1988, 14p.

8810 - ACHCAR, J.A., BOLFARINE, H & PERICCHI, L.R.; Some Applications of Bayesian Methods in Analysis of Life Data, São Paulo, IME-USP, 1988, 30p.

8811 - RODRIGUES, J.; A Bayesian Analysis of Capture-Recapture Experiments for a Closed Animal Population, São Paulo, IME-USP, 1988, 10p.

8812 - FERRARI, P.A.; Ergodicity for Spin Systems, São Paulo, IME-USP, 1988, 25p.

8813 - FERRARI, P.A. & MAURO, E.S.R.; A Method to Combine Pseudo-Random Number Generators Using Xor, São Paulo, IME-USP, 1988, 10p.

8814 - BOLFARINE, H. & RODRIGUES, J.; Finite Population Prediction Under a Linear Functional Superpopulation Model a Bayesian Perspective, São Paulo, IME-USP, 1988, 22p.

8815 - RODRIGUES, J. & BOLFARINE, H.; A Note on Asymptotically Unbiased Designs in Survey Sampling, São Paulo, IME-USP, 1988, 6p.

8816 - BUENO, V.C.; Bounds for the a Availabilities in a Fixed Time Interval for Continuous Structures Functions, São Paulo, IME-USP, 1988, 22 p.

8817 - TOLOI, C.M.C. & MORETTIN, P.A.; Spectral Estimation for Time Series with Amplitude Modulated Observations: A Review, São Paulo, IME-USP, 1988, 16p.

8818 - CHAYES, J.T.; CHAYES, L.; GRIMMETT, G.R.; KESTEN, H. & SCHONMANN, R.H.; The Correlation Length for the High Density Phase of Bernoulli Percolation, São Paulo, IME-USP, 1988, 46p.

8819 - DURRETT, R.; SCHONMANN, R.H. & TANAKA, N.I.; The Contact Process on a Finite Set, III: The Critical Case, São Paulo, IME-USP, 1988, 31p.

8820 - DURRET, R.; SCHONMANN, R.H. & TANAKA, N.I.; Correlation Lengths for Oriented Percolation, São Paulo, IME-USP, 1988, 18p.

8821 - BRICMONT, J.; KESTEN, H.; LEBOWITZ, J.L. & SCHONMANN, R.H.; A Note on the Ising Model in High Dimensions, São Paulo, IME-USP, 1988, 21p.

8822 - KESTEN, H. & SCHONMANN, R.H.; Behavior in Large Dimensions of the Potts and Heisenberg Models, São Paulo, IME-USP, 1988, 61p.

8823 - DURRET, R. & TANAKA, N.I.; Scaling Inequalities for Oriented Percolation, São Paulo, IME-USP, 1988, 21p.

8901 - RODRIGUES, J.; Asymptotically Design - Unbiased Predictors to Two-Stage Sampling, São Paulo, IME-USP, 1989, 9p.

8902 - TOLOI, C.M.C. & MORETTIN, P.A.; Spectral Analysis for Amplitude Modulated Time Series, São Paulo, IME-USP, 1989, 24p.

8903 - PAULA, G.A.; Influence Measures for Generalized Linear Models with Restrictions in Parameters, São Paulo, IME-USP, 1989, 18p.

8904 - MARTIN, M.C. & BUSSAB, W.O.; An Investigation of the Properties of Raking Ratio Estimators for Cell Frequencies with Simple Random Sampling, São Paulo, IME-USP, 1989, 11p.

8905 - WECHSLER, S.; Yet Another Refutation of Allais' Paradox, São Paulo, IME-USP, 1989, 6p.

8906 - BARLOW, R.E. & PEREIRA, C.A.B.; Conditional Independence and Probabilistic Influence Diagrams, São Paulo, IME-USP, 1989, 21p.

8907 - BARLOW, R.E.; PEREIRA, C.A.B. & WECHSLER, S.; The Bayesian Approach to Ess, São Paulo, IME-USP, 1989, 20p.

8908 - PEREIRA, C.A.B. & BARLOW, R.E.; Medical Diagnosis Using Influence Diagrams, São Paulo, IME-USP, 1989, 13p.

8909 - BOLFARINE, H.; A Note on Finite Population Prediction Under Asymmetric Loss Functions, São Paulo, IME-USP, 1989, 8p.

8910 - BOLFARINE, H.; Bayesian Modelling in Finite Populations, São Paulo, IME-USP, 1989, 8p.

8911 - NEVES, M.M.C & MORETTIN, P.A.; A Generalized Cochrane-Orcutt-Type Estimator for Time Series Regression Models, São Paulo, IME-USP, 1989, 29p.

8912 - MORETTIN, P.A., TOLOI, C.M.C., GAIT, N. & MESQUITA, A.R.; Analysis of the Relationships Between Some Natural Phenomena: Atmospheric Precipitation, Mean Sea Level and Sunspots, São Paulo, IME-USP, 1989, 35p.

8913 - BOLFARINE, H.; Population Variance Prediction Under Normal Dynamic Superpopulation Models, São Paulo, IME-USP, 1989, 7p.

8914 - BOLFARINE, H.; Maximum Likelihood Prediction in Two Stage Sampling, São Paulo, IME-USP, 1989, 4p.

8915 - WECHSLER, S.; Exchangeability and Predictivism, São Paulo, IME-USP, 1989, 10p.

8916 - BOLFARINE, H.; Equivariant Prediction in Finite Populations, São Paulo, IME-USP, 1989, 13p.

8917 - SCHONMANN, R.H.; Critical Points of Two Dimensional Bootstrap Percolation Like Cellular Automata, São Paulo, IME-USP, 1989, 6p.

8918 - SCHONMANN, R.H.; On the Behavior of Some Cellular Automata Related to Bootstrap Percolation, São Paulo, IME-USP, 1989, 26p.

8919 - PEREIRA, P.L.V.; Local Nonlinear Trends, São Paulo, IME-USP, 1989, 8p.

8920 - ANDJEL, E.D., SCHINAZI, R.B. & SCHONMANN, R.H.; Edge Processes of One Dimensional Stochastic Growth Models, São Paulo, IME-USP, 1989, 20p.

8921 - MARKWALD, R, MOREIRA, A.R.B. & PEREIRA, P.L.V.; Forecasting Level and Cycle of the Brazilian Industrial Production Leading Indicators versus Structural Time Series Models, São Paulo, IME-USP, 1989, 23p.

8922 - NEVES, E.J. & SCHONMANN, R.H.; Critical Droplets and Metastability for a Glauber Dynamics at Very Low Temperatures, São Paulo, IME-USP, 1989, 33p.