RT-MAE-9103

AN ALGORITHM FOR THE MSAE ESTIMATION OF THE MULTISTAGE DOSE-RESPONSE MODEL

by

Carmen D.S. Andrē, Clovis A. Peres

and

Subhash C. Narula

Palavras Chaves: Exponential model; L_1-norm; Least squares; MSAE regression; Nonlinear regression; Radiobiology.
(Key words)

Classificação AMS: 62J99
(AMS Classification)

AN ALGORITHM FOR THE MSAE ESTIMATION OF THE MULTISTAGE 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 USA 23284

ABSTRACT

The least squares procedure is often used to estimate the parameters of the multistage dose-response model. However, these estimates are unduly affected by outliers in a data set. The minimum sum of absolute errors, MSAE estimates are more resistant to outliers than the least squares estimates. 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.

Keywords: Exponential model; L_1-norm; Least squares; MSAE regression; Nonlinear regression; Radiobiology.

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 1. The values of the two estimates are practically the same.

The model in (1) is intrinsically nonlinear. It is not possible to find the closed form MSAE estimators of the parameters. A number of algorithms have been proposed to compute the MSAE estimates of the parameters for the general nonlinear regression model, Gonin and Money (1987). The available algorithms are iterative and converge to the MSAE estimates. For example, Osborne and Watson (1971) reduce the original nonlinear problem to a sequence of linear MSAE problems and then solve each problem as a linear programming problem. Tishler and Zang (1982) overcome the nondifferentiability of the original objective function by transforming the problem into a sequence of unconstrained nonlinear minimization problems. That is, the available algorithms consist of two nested iterative procedures and require intensive computations. These algorithms may also result in round-off and truncation errors. Furthermore, for bootstrap, jackknifing or simulation studies, it is useful, if not imperative, to have some computationally less intensive method to solve the problem. The solution from such a method may also be used as a starting solution of one of the available algorithms. Since the model in (1) has a special form, we can approximate it by a linear model which can be solved by a simpler and computationally less intensive algorithm than the available algorithms.

1. INTRODUCTION

Consider the multistage dose-response model

$$y_i = 1 - \exp(-\underline{d}_i' \underline{\alpha}) + \epsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

where y_i denotes the value of the response variable corresponding to dose $\underline{d}_i' = (1, d_i, d_i^2, \ldots, d_i^k)$, $\underline{\alpha}' = (\alpha_0, \alpha_1, \ldots, \alpha_k)$ is a $k + 1$ vector of unknown parameters, $k < n$, and ϵ_i denotes the unobservable random error. Armitage and Doll (1954) developed the model to assess the risk to the population of exposure to toxic 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 has often been used to compute virtually safe dose VSD by extrapolating the curve to dose levels below the experimental doses, Portier and Hoel (1983). However, they pointed out that the variability due to binomial sampling or the improper assumptions concerning the functional form of the dose-response model can result in errors in the estimation of VSD for small bioassays.

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 1. These date are taken from Table II of

---

Insert Table 1 about here

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_2 by y_1, i.e., changed y_2 from 0.479 to 0.714. The least squares, LS estimates of the parameter of the model for the original and the altered data are given in Table 1. Clearly, the values of

The rest of the paper is organized as follows: In Section 2, we briefly describe the Tishler and Zang (1982) algorithm to estimate the MSAE estimates of parameters of model (1). In Section 3, we give a linear approximation for the dose-response model and modify the Tishler and Zang (1982) algorithm to estimate its parameters. In Section 4, we give some results comparing the estimates obtained from the exact model and the linear approximation of the model. We conclude the paper with a few remarks in Section 5.

2. TISHLER AND ZANG ALGORITHM FOR THE EXACT MODEL

It is possible to compute the MSAE estimates of the parameter in (1) by using the algorithm proposed by Tishler and Zang (1982) for the general nonlinear model. To compute the MSAE estimate $\hat{\alpha}$ of α in (1), we observe that

$$|e_i| = |z_i - g(d_i,\alpha)|, \qquad i = 1, \ldots, n, \tag{2}$$

where $z_i = 1 - y_i$ and $g(d_i, \alpha) = \exp(-d_i\alpha)$. Thus, our objective is to minimize

$$G(\alpha) = \sum_{i=1}^{n} |e_i| \ . \tag{3}$$

The function $G(\alpha)$ does not have continuous first derivative. They noted that the absolute functional can be written as

$$|r| = \max(0, r) + \max(0, -r) \ , \ \forall \ r \in R \tag{4}$$

and can be approximated by

$$|r| \doteq H(\beta, r) = \begin{cases} -r \ , & \text{if } r < -\beta, \\ \dfrac{r^2 + \beta^2}{2\beta} \ , & \text{if } -\beta \leq r \leq \beta, \\ r \ , & \text{if } r > \beta. \end{cases} \tag{5}$$

Clearly, $|r|$ is approximated by $H(\beta, r)$ only in the interval $-\beta \leq r \leq \beta$, and this interval can be made arbitrarily small by reducing $\beta (> 0)$. In fact,

$$\lim_{\beta \to 0} H(\beta, r) = |r|.$$

Applying approximation (5) to $|e_i|$, $i = 1, \ldots, n$, in the expression $G(\alpha)$, we obtain

$$G(\beta,\alpha) = \sum_{i=1}^{n} H(\beta, e_i) \tag{6}$$

as an approximation to $G(\alpha)$. It is easy to see that

(i) $G(\beta, \alpha)$ has continuous first derivatives;

(ii) $G(\beta, \alpha)$ can be made arbitrarily close to $G(\alpha)$ by reducing β. In fact

$$\lim_{\beta \to 0} G(\beta, \alpha) = G(\alpha) \text{ , and}$$

(iii) $G(\beta, \alpha)$ differs from $G(\alpha)$ only in a small neighborhood of the points where $G(\alpha)$ is not differentiable. The approximation applies only to observations for which the residuals are smaller than β. The observations with large residuals ($> \beta$) affect the minimization of $G(\alpha)$ and $G(\beta, \alpha)$ in the same way.

Furthermore, it is clear that if β is very small, the function $G(\beta,\alpha)$ is very close to $G(\alpha)$ that is non-differentiable; and, if β is too large, the computed estimate will be the same as the least squares estimate.

Let $\beta^{(1)}$ and $\alpha^{(1)}$ be the initial values of β and α, respectively. Set $m = 1$. Then the algorithm can be stated as:

Step 1: Solve

$$\text{minimize } G(\beta^{(m)}, \alpha).$$

Let $\underline{\alpha}^{(m+1)}$ be the solution to this problem. Also compute

$$e_i^{(m+1)} = \exp\left(-d_i' \alpha^{(m+1)}\right) - z_i.$$

Step 2: If

$$|e_i^{(m+1)} - e_i^{(m)}| \leq 0, \quad i = 1, \ldots, n.$$

and

$$|\alpha^{(m+1)} - \alpha^{(m)}| < \delta,$$

where $\delta > 0$ is a small pre-specified tolerance, stop, and let $\hat{\alpha} = \alpha^{(m+1)}$ be the optimal solution; if not, choose $\beta^{(m+1)} < \beta^{(m)}$, set $m = m + 1$ and go to Step 1.

It may be observed that the algorithm is an iterative procedure. Furthermore, within each iteration, the algorithm has to solve a nonlinear model which involves an iterative process. Tishler and Zang (1982) prove the

convergance of the algorithm. They suggest an initial choice of $\beta = 0.1$ but do not give any guidelines to decrease β.

3. AN ALGORITHM FOR LINEAR APPROXIMATION

Clearly, if the model can be linearized, it will be possible to use a computationally less intensive algorithm to compute the MSAE estimates of the parameters. Peres and Narula (1989) have shown that it is possible to approximate the dose-response model (1) by a linear model. To do so, we rewrite (2) as

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

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

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

$$|e_i| = |z_i(\ln z_i + d_i'\alpha)|, \quad i = 1, \ldots, n .$$

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

$$|e_i| = w_i (\ln z_i + d_i' \alpha)| , \quad i = 1, \ldots, n \tag{7}$$

where $w_i = z_i (1 - (\ln z_i + d_i'\alpha)/2)$, $i = 1, \ldots, n$. Using the approximation (5) to the absolute value functional in (7), we minimize

$$G^*(\beta,\alpha) = \sum_{i=1}^{n} H^*(\beta, e_i^*) , \tag{8}$$

where

$$H^*(\beta, e_i^*) = \begin{cases} -w_i (\ln z_i + d_i'\alpha) , & \text{if } e_i^* < -\beta, \\ \dfrac{w_i^2 (\ln z_i + d_i'\alpha)^2}{2\beta} + \dfrac{\beta}{2}, & \text{if } -\beta \le e_i^* \le \beta, \\ w_i (\ln z_i + d_i'\alpha) , & \text{if } e_i^* > \beta. \end{cases}$$

To solve (8), the Tishler and Zang algorithm can be modified as follows: Let $\beta^{(1)}$ and $\alpha^{(1)}$ be the initial values of β and α, respectively. Set $m = 1$.

Step 1: Solve

$$\text{minimize } G^*(\beta^{(m)}, \alpha) .$$

Let $\alpha^{(m+1)}$ be the solution to this problem. Also compute

$$w_i^{(m+1)} = z_i (1 - (\ln z_i + d_i' \alpha^{(m)})/2) ,$$

$$e_i^{*(m+1)} = w_i^{(m+1)} (\ln z_i + d_i' \alpha^{(m)})$$

$$e_i^{(m+1)} = \exp(-d_i' \alpha^{(m)}) - z_i , \text{ and}$$

$$\sum_{i=1}^{n} |e_i^{(m+1)}| .$$

Step 2: If

$$|e_i^{(m+1)} - e_i^{*(m)}| \doteq 0, \quad i = 1, \ldots, n$$

and

$$|\alpha^{(m+1)} - \alpha^{(m)}| < \delta$$

where $\delta > 0$ is a small pre-specified tolerance, stop, and let $\hat{\alpha} = \alpha^{(m+1)}$ be the optimal solution; if not, choose $\beta^{(m+1)} < \beta^{(m)}$, set $m = m + 1$ and go to Step 1.

Based on our limited computational experience, we suggest the following guidelines to choose β. In the first step, select $\beta^{(1)}$ such that the interval $[-\beta^{(1)}, \beta^{(1)}]$ contains between $k + 1$ (the number of parameters of the model) and n residuals. In the subsequent iterations, choose $\beta^{(m)}$ such that the $k + 1$ smallest residuals (in terms of the absolute value) lie within the interval $[-\beta^{(m)}, \beta^{(m)}]$. However, if at some iteration, the sum of absolute

errors increases, repeat the iteration with a larger value of β and continue the process.

4. COMPUTATIONAL EXPERIENCE

It is possible to compute the MSAE estimates of the parameters of a multistage dose-response model from using (6) or from the linearized model using (8). In an effort to compare these estimates, the Tishler and Zang algorithm (Section 2) was implemented on the Burrough B-6900 computer at the Centro de Computacão Electronica da USP and the algorithm for the linearized model (Section 3) was implemented in Turbo Pascal on an IBM compatible microcomputer. We computed the MSAE estimates for a few data sets taken from Sankaranarayanan (1969a, 1969b). The results for the model $y_i = \exp(-\alpha d_i) + \varepsilon_i$ are summarized in Table 2, and for the model $y_i = \exp(-\alpha_1 d_i - \alpha_2 d_i^2) + \varepsilon_i$ in Table 3.

Insert Tables 2 and 3 about here

From Tables 2 and 3 we observe that the MSAE estimates obtained using the exact model and the linear approximation are compatible.

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. If desired, the estimates from the linear approximation of the model can be used as a starting solution in an algorithm for the exact model. The proposed approximation and the algorithm can also be used in bootstrap, jackknifing or simulation studies.

REFERENCES

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

Gonin, R. and Money, A. H. (1987). A review of computational methods for solving the nonlinear L_1-norm estimation problem. Statistical Data Analysis Based on the L_1-norm and Related Methods (Y. Dodge, editor), North-Holland, Amsterdam, 117-129.

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.

Osborne, M. R. and Watson, G. A. (1971). On an algorithm for discrete nonlinear L_1 approximation. Computing Journal, 14, 184-188.

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.

Portier, C. and Hoel, D. (1983). Low dose rate extrapolation using the multistage model. Biometrics, 39, 897-906.

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.

Tishler, A. and Zang, I. (1982). An absolute deviation curve fitting algorithm for nonlinear models. TIMS Studies in Management Sciences, 19, 81-103.

LIST OF TABLES

i	d_i	Original y_i	Altered y_i
1	0.15	0.714	0.714
2	0.30	0.479	0.714
3	0.45	0.321	0.321
4	0.60	0.215	0.215
5	0.75	0.150	0.150
LS Estimate		2.4930	2.1567
MSAE Estimate		2.5299	2.5299

TABLE 1

Data Set*	Model Used	MSAE Estimate of α	Sum of Absolute Error
Nitrogen Post-treatment Table II	Exact Model	3.0871	0.00418
	Linear Approx.	3.0859	0.00416
Oxygen Post-treatment Table II	Exact Model	2.5292	0.00943
	Linear Approx.	2.5299	0.00943

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

TABLE 2

Model Used	MSAE Estimate of α_1	MSAE Estimate of α_2	Sum of Absolute Errors
Exact Model	0.21332	0.07558	0.00813
Linear Approx.	0.21503	0.07480	0.00809

TABLE 3

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1990

9001 - JR., HODGES, J.L. RAMSEY, P.H. & WECHSLER, S.; Improved Significance Probabilities of the Wilcoxon Test, 30p.

9002 - PAULA, G.A.; Bias Correction for Exponential Family Nonlinear Models, 10p.

9003 - SCHONMANN, R.H. & TANAKA, N.I.; One Dimensional Caricature of Phase Transition, 16p.

9004 - ZACKS, S. & BOLFARINE, H.; Maximum Likelihood Prediction in Finite Populations, 15p.

9005 - ZACKS, S. & BOLFARINE, H.; Equivariant Prediction of the Population Variance Under Location-Scale Superpopulation Models, 8p.

9006 - BOLFARINE, H.; Finite Population Prediction Under Error in Variables Superpopulation Models, 14p.

9007 - BOLFARINE, H.; Ratio and Regression Estimators Under Error-In-Variables Superpopulation, 16p.

9008 - BOTTER, D.A. & MOTTA, J.M.; Experiments With Three-Treatment Three-Period Crossover Design: Analysis Through the General Linear Models, 15p.

9009 - SCHONMANN, R.H.; Finite Size Scaling Behavior of a Biased Majority Rule Cellular Automaton, 11p.

9010 - BUENO, V.C.; A Martingale Version of Multistate Monotone Systems, 5p.

9011 - CORDEIRO, G.M. & FERRARI, S.L.P.; A Modified Score Tests Statistic Having Chi-Squared Distribution to Order, 12p.

9012 - NEVES, E.J. & SCHONMANN, R.H.; Behavior of Droplets for a Class a Glauber Dynamics at very Low Temperature, 33pg.

9013 - FERRARI, P.A.; Schock Fluctuations in Asymmetric Simple Exclusion, 26p.

9014 - MAGALHÃES, M.N. & TANAKA, N.I.; Reverse Process in M/M/1 Queues, 7p.

9015 - CORDEIRO, G.M. & PAULA, G.A.; Estimation Large Sample Parametric Tests and Diagnostic for Non-Exponential Family Nonlinear Models, 31p.

9016 - CORDEIRO, G.M., FERRARI, S.L.P. & PAULA, G.A.; Improved Score Tests for Generalized Linear Models, 25p.

9017 - ESTON, V.R., GALVES, A. JACOBI, C.M., & LANGENVIN, R.; Dominance Switch Between Two Interacting Species and Large Deviations, 8p.

9018 - IRONY, T.Z., PEREIRA, C.A.B. & BARLOW, R.E.; Bayesian Models for Quality Assurance, 22p.

9019 - BOLFARINE, H. & ZACKS, S., Bayes and Minimax Prediction in Finite Populations, 16p.

9020 - BOLFARINE, H.; RODRIGUES, J. & ZACKS, S.; Some Asymptotic Results in Finite Populations, 13p.

9021 - BOLFARINE, H., CORDANI, L.K.; Estimation of a Structural Linear Regression Model With a Known Reliability Ratio, 9pg.

9022 - PAULINO, C.D.M., PEREIRA, C.A.B.; Bayesian Analysis of Categorical Data Informatively Censored. São Paulo, 18pg

9023 - FERRARI, P.A., MARTINEZ, S. & PICCO, P.; Some Properties of Quasi Stationary Distributions in the Birth and Death Chains, 11pg.

9024 - BOLFARINE, H., SANDOVAL, M.C.; Prediction of the finite population distribution function under Gaussian superpopultion models, 13 pg.

9025 - BUENO, V.C. & MORROS, I.: Component importance through compensator transforms, 7p

1991

9101 - RAVISHANKAR, K.: Convergence of a random motion to a point, 11p.

9102 - FERRARI, P.A. & RAVISHANKAR, R.: Shocks in asymmetric exclusion automata, 17p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
IME-USP - Caixa Postal 20.570
01498 - São Paulo, Brasil