

RT-MAE 9526

**NONLINEAR QUASI-BAYESIAN THEORY
AND INVERSE LINEAR REGRESSION**

by

*Josemar Rodrigues, Heleno Boffarine
and
Gauss M. Cordeiro*

Key words: Animal population sizes; Gauss-Newton estimation procedure; Inverse linear regression; Nonlinear quasi-Bayesian estimator; Regression models with error in both variables.

Classificação AMS: 62C10, 62J02.
(AMS Classification)

NonLinear Quasi-Bayesian Theory and Inverse Linear Regression •

Josemar Rodrigues
ICMCS-USP
C.Postal 668, CEP: 13560-970
São Carlos - SP, Brazil

Heleno Bolfarine
IME-USP
C.Postal 66281 CEP: 05389-970
São Paulo SP, Brazil

Gauss M. Cordeiro
Departamento de Estatística
CCEN/UFPE
Cidade Universitaria, CEP: 50740-540, Recife, Brazil

Abstract

In this paper, a quasi-Bayesian approach for nonlinear estimation problems is considered. The formulas given are based on the Gauss-Newton estimating procedure and require only the first and second moments of the distributions involved. The use of GLIM package to solve for the estimation problems considered is discussed. Applications are made to estimation problems in inverse linear regression, regression models with both variables subject to error and also to the estimation of the size of animal populations. Some numerical illustrations are reported. For the inverse linear regression problem, comparisons with ordinary Bayesian and other techniques are considered.

Keywords and Phrases: Animal population sizes; Gauss-Newton estimation procedure; Inverse linear regression; Nonlinear quasi-Bayesian estimator; Regression models with error in both variables.

1 Introduction

The discussion in the statistical literature on how to solve the inverse linear regression problem is characterized by disagreement and confusion. See for example, Krutchkoff (1967), Hoadley (1970) and Hunter and Lamboy (1981). Dobrigal et al. (1987) developed an approximate conditional inference for the problem. Kalotay (1971) presented a structural analysis (Fraser, 1986) of the linear calibration and Minder and Whitney (1975) advocated the likelihood approach.

In this paper, a quasi-Bayesian approach is proposed for solving the inverse linear regression and other nonlinear problems. In the case of the inverse linear regression problem the solution is original in the sense that it is considered as a nonlinear problem and the original variables are replaced by differences conveniently taken in order to eliminate the intercept parameter.

The estimation procedure suggested in this paper enables the statistician to incorporate prior information when available about the parameters involved in the estimation process, a feature not always shared by the strict Bayesian approaches proposed in the literature. This weakness of the ordinary Bayesian approach has been pointed out by some authors (Lawless, 1981).

The remainder of the paper is organized as follows. In Section 2, the quasi-Bayesian approach to nonlinear estimation is proposed. In Section 3, application is made to the inverse linear regression. Numerical comparisons with standard Bayes and other techniques are reported. Applications to regression variables with error in both variables and estimation of the size of animal populations are considered in Sections 4 and 5, respectively.

2 Nonlinear Quasi-Bayesian Approach to Estimation

The usual Bayesian approach to estimation requires a complete specification of the distributions involved. Hartigan (1969) proposed an estimation method that requires only the first and second moments instead of the complete distribution. See also Godambe (1994). As in Hartigan (1969), it is assumed that a set of observations $y' = (y_1, \dots, y_n)$ satisfies the nonlinear Bayesian model represented by

$$E[y|\theta] = f(\theta) = \mu \text{ and } Var[y|\theta] = V(\theta) , \quad (1)$$

where $\theta' = (\theta_1, \dots, \theta_p)$ is a vector of unknown parameters. With respect to the distribution of θ it is only assumed that

$$E[\theta] = \theta_0 \text{ and } Var[\theta] = \Sigma , \quad (2)$$

where θ_0 and Σ are known. The $n \times 1$ vector $f'(\theta) = (f_1(X_1; \theta), \dots, f_n(X_n; \theta))$ is a vector of known and twice continuously differentiable functions in θ , where $X'_i = (x_{i1}, \dots, x_{iq})$ represents a vector of known explanatory variables. The problem is then how to combine (1) and (2) to produce sensible estimators of θ . The natural way would be via Bayes theorem. In the next definitions, the notions of quasi-prior and posterior distributions to solve the nonlinear estimation problem proposed above are introduced. The first definition was introduced by Wedderburn (1974).

Definition 1 We define, following Wedderburn (1974) and McCullagh (1983), the logarithm of the quasi-prior function of θ by the relation

$$\frac{\partial \Pi^*(\theta)}{\partial \theta} = \sum^{-1} (\theta_0 - \theta) , \quad (3)$$

where Σ and θ_0 are known.

As mentioned above, the natural way of combining data and prior information is via Bayes theorem. Therefore, the logarithm of the quasi-posterior function of θ combines naturally, via Bayes theorem, Wedderburn's quasi-likelihood function and the quasi-prior (3).

Definition 2 The logarithm of the quasi-posterior function of θ , $\Pi^*(\theta|y)$, is defined by the relation

$$\frac{\partial \Pi^*(\theta|y)}{\partial \theta} = \sum^{-1} (\theta_0 - \theta) + D'V^{-1}(\mu)(y - \mu), \quad (4)$$

where $D = \frac{\partial \mu}{\partial \theta}$ is an $n \times p$ matrix.

Expression (4) will be represented by $U(\theta)$ in the sequel.

Definition 3 The quasi-Bayes estimator of θ is considered to be the value of θ that maximizes $\Pi^*(\theta|y)$, that is, the value of θ , $\hat{\theta}^*$, that satisfies

$$\frac{\partial \Pi^*(\theta|y)}{\partial \theta} \Big|_{\theta=\hat{\theta}^*} = U(\hat{\theta}^*) = 0.$$

By replacing $\frac{-\partial U(\theta)}{\partial \theta}$ by its expected value $K(\theta) = \sum^{-1} + D'V^{-1}(\mu)D$ (called the quasi-information matrix), as recommended by Wedderburn (1974), $\hat{\theta}^*$ may be obtained by the iterative procedure that results from equation (essentially Fisher's scoring technique)

$$\hat{\theta}_{(m+1)} = \theta_{(m)} + \left(\sum^{-1} + D'V^{-1}(\mu)D \right)^{-1} U(\theta) \Big|_{\theta_{(m)}}. \quad (5)$$

Equation (5) holds iteratively until convergence, that is, until $\|\theta_{(m+1)} - \theta_{(m)}\| \leq \epsilon$, for some fixed ϵ . Note that if $f(\theta) = X\theta$ and $V(\theta) = V$, the usual linear set up, then

$$\hat{\theta}^* = C_1 \hat{\theta}_{MQ} + (I - C_1)\theta_0,$$

where I is the identity matrix of appropriate dimension,

$$\hat{\theta}_{MQ} = (XV^{-1}X)^{-1} X'V^{-1}y, \text{ and } C_1 = \left(\sum^{-1} + X'V^{-1}X \right)^{-1} X'V^{-1}X$$

is the usual linear Bayes estimate of θ , which was obtained by Hartigan (1969), among others. The maximum value $\hat{\theta}^*$ has quasi-posterior variance

$$Cov(\hat{\theta}^*) = -E \left[\frac{\partial U(\theta)}{\partial \theta} \right]^{-1} \Big|_{\hat{\theta}^*} = \left[\sum^{-1} + D'V^{-1}(\mu)D \right]^{-1} \Big|_{\hat{\theta}^*}. \quad (6)$$

3 Using GLIM Package

Another way of presenting the iterative procedure above suitable for using GLIM is now described. The iterative procedure described by (5) may also be written as

$$K(\theta_{(m)}) [\theta_{(m+1)} - \theta_{(m)}] = U(\theta_{(m)}).$$

We shall assume that D is of full rank r , and that $K(\theta)$ the quasi-information matrix is positive definite throughout the parameter space. Thus, (5) is a nonsingular $p \times p$ system of equations for $\theta_{(m+1)}$ and we can find a Cholesky decomposition of $K(\theta)$, $K(\theta) = Q(\theta)'Q(\theta)$, where $Q(\theta)$ is an $p \times p$ upper triangular matrix with all diagonal entries positive. Now, (5) may be written as

$$Q'(\theta_{(m)}) Q(\theta_{(m)}) \theta_{(m+1)} = Q'(\theta_{(m)}) \tau(\theta_{(m)}), \quad (7)$$

where

$$\tau(\theta) = Q(\theta)\theta + Q(\theta)'^{-1}U(\theta). \quad (8)$$

It results that $\hat{\theta}^*$ can be found using an iteratively weighted least squares algorithm with working independent variable $Q(\theta)$ and working dependent variable $\tau(\theta)$. The decomposition of $K(\theta)$ can be time-consuming, but this way of computing $\hat{\theta}^*$ has numerical advantages concerning rounding errors.

Although the algorithm outlined here is applicable to many positive definite quasi-information matrices $K(\theta)$, it is limited in practice since it is much more complicated to handle the required Cholesky decomposition in statistical packages like GLIM and GENSTAT. It is to be hoped that the enhanced matrix handling and inversion facilities of GLIM4 will enable to do a separate Cholesky decomposition followed by the ordinary least squares procedure to obtain $\hat{\theta}^*$.

An important special case in which we may easily use GLIM to compute $\hat{\theta}^*$ occurs when $V(\mu)$ is a diagonal matrix. When $V(\mu)$ is diagonal, we use the Cholesky decomposition of Σ , $\Sigma = P'P$, where P is an $p \times p$ upper triangular matrix. This yields the following partitioned quasi-information matrix

$$K(\theta) = A(\theta)'W(\theta)A(\theta), \quad (9)$$

where

$$W(\theta) = \begin{pmatrix} V^{-1}(\mu) & O_{n \times r} \\ O_{r \times n} & I_{r \times r} \end{pmatrix}$$

is a diagonal matrix of order $(n-p)$ with $O_{n \times p}$ and $O_{r \times n}$ being $n \times r$ and $n \times n$ zero matrices, respectively, $I_{p \times p}$ being the identity matrix of order p and $A(\theta) = (D'P^{-1})'$ being an $(n+r) \times p$ hypothetical local model matrix. The additional p rows to form the augmented matrix $A(\theta)$ represent the quasi-information we get from the quasi-prior function proposed for the vector θ .

The log quasi-likelihood function $U(\theta)$ becomes

$$U(\theta) = A(\theta)'W(\theta) \begin{pmatrix} y - \mu \\ P'^{-1}(\theta_0 - \theta) \end{pmatrix}. \quad (10)$$

From (4) and (10), it results that

$$A'(\theta_{(m)}) W(\theta_{(m)}) A(\theta_{(m)}) \theta_{(m+1)} = A(\theta_{(m)})' W(\theta_{(m)}) \Gamma(\theta_{(m)}), \quad (11)$$

where

$$\Gamma(\theta) = \begin{pmatrix} y - \mu + D\theta \\ P'^{-1}\theta_0 \end{pmatrix}$$

is an $(n+p) \times 1$ working dependent variable to be regressed on the columns of $A(\theta)$. It is now shown that the computations of the iteratively weighted least squares procedure (10) and (11) may be performed in GLIM with a few modifications.

If we assume that there is a vector of additional observations $(y_{n+1}, \dots, y_{n+r})'$ defined by $P'^{-1}\theta_0$ with corresponding variances equal to one, and denoting $(y_1, \dots, y_{n+r})'$ as the dependent variable and specifying the columns of $A(\theta)$ by the FIT directive, it may be seen that $W(\theta)$ will work as a modified weighting function and $\Gamma(\theta)$ will have the same form as the modified dependent variable in GLIM. The columns of $A(\theta)$ must be recalculated in each iteration and therefore, we have to use only one step of the GLIM algorithm by declaring RECYCLE 1. Here, the system vector of the linear predictors %LP in the m th iteration is calculated as

$$\Gamma(\theta) = \begin{pmatrix} D'_{(m)}\theta_{(m)} \\ P'^{-1}\theta_0 \end{pmatrix}$$

and it should be redefined for the next iteration by replacing $D'_{(m)}\theta_{(m)}$ by $f(\theta_{(m)})$ after extracting $\theta_{(m)}$ in %PE. Of course, before the first iteration, %LP must contain starting values defined by replacing $D_{(m)}$ by $D_{(0)}$ and $\theta_{(m)}$ by $\theta_{(0)}$ in $\Gamma(\theta)$, where θ_0 denotes a initial guess for θ and $D_{(0)}$ is the value of D at $\theta_{(0)}$.

We have to use the OWN directive by setting up the model through 4 macros. In the first macro we let %FV = %LP. In the second macro, %DR = A and in the third the vector %V A represents the $(n+p)$ vector of variances $(v(\mu_1), \dots, v(\mu_n), 1, \dots, 1)'$. In the fourth macro we have to define an analogous of the deviance components to calculate %DI.

We are proposing a nonlinear Bayesian approach based on the quasi-likelihood approach of Wedderburn (1974), which surely is related to the work of Hartigan (1969). Thus, emphasis is placed on the relation of the approach with the GLIM Package, which can facilitate the derivation of these estimators.

4 The Inverse Linear Regression Problem

The discussion in the statistical literature on how to make inference under the inverse linear regression model is characterized by disagreement and confusion. In this section, we look at the inverse regression problem as a nonlinear problem and an alternative solution based on the results of Section 2 is proposed. The usual set up of the inverse linear regression problem

can be stated formally as follows. Two sets of independent random variables y_{11}, \dots, y_{1n} and y_{21}, \dots, y_{2n} are taken, where,

$$y_{1i} = \alpha + \beta x_i + e_{1i},$$

$$y_{2j} = \alpha + \beta X + e_{2j}, \quad (12)$$

$$e_{1i} \sim (0, \sigma^2) \text{ and } e_{2j} \sim (0, \sigma^2),$$

$i = 1, \dots, n$ and $j = 1, \dots, m$. It is assumed that x_1, \dots, x_n are known constants, σ^2 is known and α and β are unknown. The problem is to make inference about X based on $y_{11}, \dots, y_{1n}, y_{21}, \dots, y_{2m}$. Without loss of generality, the x_i 's may be chosen so that

$$\sum_{i=1}^n x_i = 0 \text{ and } \sum_{i=1}^n x_i^2 = n. \quad (13)$$

Ordinary Bayesian solution to this problem are given by Hoadley (1970) and Hunter and Lamboy (1981), among others. In their papers, the errors e_{1i} and e_{2j} are considered to be normally distributed and a noninformative prior is considered for α and β . Emphasis was given on justifying the classical or inverse estimators, rather than on finding more convenient alternative solutions. The solutions to the estimation problems most widely recommended in the literature are the classical estimator, $\hat{X}_{MLE} = \frac{(\bar{y}_2 - \bar{y}_1)}{\hat{\beta}_{MLE}}$, with $\hat{\beta}_{MLE} = \frac{1}{n} \sum_{i=1}^n y_{1i} x_i$, $\hat{\alpha}_{MLE} = \bar{y}_1$ also known as the maximum likelihood solutions and the inverse estimator which is given by

$$\hat{X}_I = \frac{n \hat{\beta}_{MLE}^2}{\sum_{i=1}^n (y_{1i} - \bar{y}_1)^2} (\bar{y}_2 - \bar{y}_1),$$

where $\bar{y}_1 = \sum_{i=1}^n y_{1i}/n$ and $\bar{y}_2 = \sum_{j=1}^m y_{2j}/m$.

For a detailed discussion on the properties of the above estimators, we suggest Krutchokoff (1967). Since the main objective is to estimate X , we may eliminate α from the estimation context by considering the model $z_i = \beta(x_i - X) + u_i, i = 1, \dots, n$, where $z_i = y_{1i} - \bar{y}_2, u_i = e_{1i} - \bar{e}_2, \bar{y}_2 = \frac{1}{m} \sum_{j=1}^m y_{2j}$ and $\bar{e}_2 = \frac{1}{m} \sum_{j=1}^m e_{2j}$.

Observe that,

$$Var[u_i] = (1 + 1/m)\sigma^2 \text{ and } Cov(u_i, u_j) = \sigma^2/m,$$

$i \neq j = 1, \dots, n$. To complete the distribution free Bayesian set up, let the prior information about $\theta' = (X, \beta)$ be specified by

$$\theta_0 = (X_0, \beta_0) \text{ and } \Sigma = \text{diag}(\sigma_X^2, \sigma_\beta^2, \sigma_\beta^2),$$

all known.

According to the notation of Section 2, it follows from (1) that

$$y = (z_1, \dots, z_n), f'(\theta) = (\beta(x_1 - X), \dots, \beta(x_n - X)) \quad (14)$$

and

$$\Sigma = \begin{pmatrix} 1 + 1/m & \dots & 1/m \\ \vdots & \ddots & \vdots \\ 1/m & \dots & 1 + 1/m \end{pmatrix}.$$

From (3) and (4), it can be shown, after some tedious algebraic manipulations, that the iterative procedure to obtain $\hat{\theta} = (\hat{X}, \hat{\beta})$ is given by

$$X_{(m+1)} = X_{(m)} - \frac{A_{21}^{(m)} B_1^{(m)} - A_{22}^{(m)} B_2^{(m)}}{\Delta^{(m)}}, \quad (15)$$

$$\beta_{(m+1)} = \beta_{(m)} - \frac{-A_{11}^{(m)} B_1^{(m)} + A_{12}^{(m)} B_2^{(m)}}{\Delta^{(m)}}, \quad (16)$$

where

$$A_{11} = \frac{1}{\sigma_x^2} + \frac{nm\beta^2}{(n+m)\sigma^2}, \quad A_{21} = A_{12}, \quad \Delta = A_{11}A_{22} - A_{12}A_{21},$$

$$A_{12} = \frac{m\beta}{(n+m)\sigma^2} \sum_{j=1}^n (x_j - X) - \frac{m}{\sigma^2(n+m)} \beta_{(m)} \sum_{j=1}^n (x_j - X_{(m)}),$$

$$A_{22} = \frac{1}{\sigma_\beta^2} + \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - X)^2 - \frac{1}{\sigma^2(n+m)} \left[\sum_{j=1}^n (x_j - X_{(m)}) \right]^2,$$

$$B_1 = \frac{\beta_0 - \beta}{\sigma_\beta^2} + \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - X)(z_j - \beta(x_j - X))$$

$$- \frac{1}{\sigma^2(n+m)} \sum_{j=1}^n (z_j - \beta(x_j - X)) \sum_{j=1}^n (x_j - X)$$

and

$$B_2 = \frac{X_0 - X}{\sigma X_0} - \frac{m\beta}{(m+n)\sigma^2} \sum_{j=1}^n (z_j - \beta(z_j - \beta(x_j - X))).$$

Let X^* and β^* be the estimators that follow from the equations (15) and (16) after the iterative procedure stabilizes. Estimators for the quasi-posterior variances that follow from (6) are

$$\text{var} [\hat{X}^*] \doteq -\frac{\hat{A}_2}{\hat{\Delta}} \text{ and } \text{var} [\hat{\beta}] \doteq -\frac{\hat{A}_{11}}{\hat{\Delta}},$$

where \hat{A}_{22} , \hat{A}_{11} and $\hat{\Delta}$ are given as in (15) and (16) with $X_{(m)}$ and $\beta_{(m)}$ replaced by the NBLs $\hat{\theta}^* = (\hat{X}^*, \hat{\beta}^*)$ obtained from the iterative method described above.

The solution to the inverse regression problem was described in detail because as noted in the last section, it may be complicated to use the GLIM package since the matrix $V(\mu)$ is not diagonal. Note also that the above estimators are natural modifications of the MLE and inverse estimators which use prior information and information from both stages of the inverse linear regression model.

Application 1: In this illustration we analyse two sets of data. One is obtained by computer simulation of (7) with the x 's values as in Hoadley (1970). The other set appears in Hunter and Lamboy (1981).

Tables 1 and 2 lists the statistics and the 95% confidence interval based on Fieller's Theorem, Chebyshev inequality and the normal approximation. Table 2 shows how the prior information improves the NBLs, \hat{X} , when the prior information is close to the true values of X and β . Note also (Table 2) that to get the inverse estimator as a NBLs estimator, it is enough to consider $X_0 = 0, \beta_0 = 0.0$ and $\sigma_x^2 = 2.5$, without the t-distribution requirement for X (Hoadley, 1970). Table 1 provides inferences which are very close to the results in Hunter and Lamboy (1981), Dobrigal et al. (1987) and Sprott and Viveiros (1984) in the absence of all assumptions and sophistication they have considered.

We end this section by noticing that it should be of interest to a Bayesian (as well as to a non-Bayesian) statistician to use the nonlinear Bayesian least squares approach considered in this paper. This is so, for its simplicity in combining prior and present data information without requiring the full Bayesian apparatus and also for the closeness of the inferences. By using this approach, complex problems such as the linear calibration problem could be handled in a much simpler way.

5 The Linear Regression Model With Error In Both Variables

In this section, it is considered the usual regression model with error in both variables, i.e.,

$$y_i = \alpha + \beta x_i + e_i,$$

where x_i is not observed directly, but instead

$$X_i = x_i + u_i,$$

$i = 1, \dots, n$. We assume that $x_i \sim N(0, \sigma_{xx})$, $e_i \sim N(0, \sigma_{ee})$ and that $u_i \sim N(0, \sigma_{uu})$, being all independent.

It can be shown, after some algebraic manipulations that

$$E[y_i|X_i] = \alpha + \beta Z_i = \mu_i \quad (17)$$

and

$$\text{Var}[y_i|X_i] = \beta_1^2 k_x \sigma_{uu} + \sigma_{ee}, \quad (18)$$

where $Z_i = \mu_x + k_x (X_i - \mu_x)$, $i = 1, \dots, n$ and $k_x = \sigma_{xx}/(\sigma_{xx} + \sigma_{uu})$. For simplicity, it is assumed that the parameters μ_x , σ_{xx} , and σ_{ee} are known. But, this assumption may easily be relaxed.

For α and β , it is considered a priori that

$$E[\alpha] = \alpha_0, \quad \text{and} \quad \text{Var}[\alpha] = \sigma_\alpha,$$

and

$$E[\beta] = \beta_0, \quad \text{and} \quad \text{Var}[\beta] = \sigma_\beta.$$

This situation is easily handled by using the GLIM package, by following the steps considered in Section 3. Of course, the iterative procedure (5) may also be used. It can be shown that it reduces to the following iterative equations:

$$\begin{cases} \alpha_{(m+1)} = \alpha_{(m)} - \frac{A_{11}^{(m)} B_1^{(m)} + A_{12}^{(m)} B_2^{(m)}}{\Delta^{(m)}} \\ \beta_{(m+1)} = \beta_{(m)} - \frac{A_{21}^{(m)} B_1^{(m)} + A_{22}^{(m)} B_2^{(m)}}{\Delta^{(m)}} \end{cases}, \quad (19)$$

where

$$\begin{aligned} A_{11} &= \frac{1}{\sigma_\beta^2} + \frac{\sum_{i=1}^n z_i^2}{v}, \quad A_{12} = A_{21} = - \frac{\sum_{i=1}^n z_i \bar{y}_i}{v} \\ A_{22} &= \frac{1}{\sigma_\alpha^2} + \frac{n}{v}, \quad B_1 = \frac{\alpha_0 - \alpha}{\sigma_\alpha^2} + \frac{n(y - \alpha - \beta \bar{z})}{v}, \\ B_2 &= \frac{\beta_0 - \beta}{\sigma_\beta^2} + \frac{\sum_{j=1}^n (y_j - \alpha - \beta z_j z_j)}{v}, \end{aligned}$$

$$\Delta = A_{11}A_{22} - A_{21}A_{12} \quad \text{and} \quad v = \beta^2 k_x \sigma_{uu} + \sigma_{ee}.$$

The quasi-posterior variances of α^* and β^* follows from (6) and are given by

$$\text{Var}[\alpha^*] = - \frac{A_{22}^*}{\Delta^*} \quad \text{and} \quad \text{Var}[\beta^*] = - \frac{A_{11}^*}{\Delta^*},$$

where A_{22}^* , A_{11}^* and Δ^* are given as in (19) with $\alpha_{(m)}$ and $\beta_{(m)}$ replaced by α^* and β^* , the solutions of the interactive system equations (19).

6 Estimation of the Size of Animal Populations

In this section, it is presented a simple nonparametric nonlinear Bayes estimation procedure for estimating the size of an animal population. It has long been recognized that a model which allows capture probabilities to vary by animals, trap response and time is a complicated and usefull model. However, from the classical approach, there is no estimation procedure associated with this model. For a comprehensive material we suggest Ottis et al. (1978) and Burnhan and Overton (1978). Ordinary Bayesian approach for the estimation of the size of animal populations is considered in Castledine (1981). However, he did not consider the possibility of incorporating trap response in the model. An extension of the model considered by Castledine (1981) that allows for trap response is considered in Rodrigues et al. (1988). See also Leite et al. (1987). The basic model and the structure of the prior information is presented next.

Let p_{ij} be the probability of capturing the j th animal in the i th ocasion, and define for $j = 1, \dots, N, i = 1, \dots, k$ the indicator variables

$$X_{ij} = \begin{cases} 1, & \text{if the } j\text{th untaged animal is captured on the} \\ & \text{ith ocassion} \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

In order to gain information about the animal population size, N , the following assumptions are considered:

Assumption 1: The animal population is closed and is of size N ;

Assumption 2: The random variables X_{ij} are mutually independent for given p_{ij} and $M_i, j = 1, \dots, N - M_i, i = 1, \dots, k$, where $M_i = \sum_{j=1}^{N-1} X_j$ and $X_j = \sum_{l=1}^{N-M} X_{jl}$. Observe from (18) that X_j is the number of untaged animals captured on the j th ocassion and M_i is the number of animals seen at least once during trapping untill time i . The basic data is X_1, \dots, X_n .

Assumption 3: (Prior assumptions)

(i) $p_{ij} \sim F$ (F : unknown distribution) where

$$E[p_{ij}] = p \text{ (unknown), and } Var[p_{ij}] = \sigma_p^2 \text{ (unknown)}$$

for $i = 1, \dots, k, j = 1, \dots, N - M_i$,

$$E[p] = p_0 \text{ (known) and } Var[p_{ij}] = \sigma_p^2 \text{ (unknown)}$$

(ii) $N \sim G$ (G : unknown distribution), where

$$E_g[N] = N_0 \text{ (known) and } Var_G[N] = \sigma_N^2.$$

From assumption 2, it is not difficult to see that

$$\text{Cov}[X_i, X_j | M_i, M_j] = 0, \quad j \neq i.$$

As in Hartigan (1969), let us consider the following nonparametric structure with respect to N :

Prior:

$$E_G[N] = N_0 \quad \text{and} \quad \text{Var}_G^{-1}[N] = \sigma_{N_0}^{-2}$$

Data:

$$E[X_i | M_i, N, p_{ij}] = \sum_{j=1}^{N-M_i} p_{ij} \quad \text{and} \quad \text{Var}^{-1}[X_i | M_i, N, p_{ij}] = \left[\sum_{j=1}^{N-M_i} p_{ij} (1 - p_{ij}) \right]^{-1}.$$

Because the above model is too complicated for our purposes, we are going to simplify it by using the prior assumption 3(i). After some algebraic manipulations we get the following simpler model:

Prior:

$$E_F[N] = N_0 \quad \text{and} \quad \text{Var}_F^{-1}[N] = \sigma_{N_0}^{-2},$$

$$E[p] = p_0 \quad \text{and} \quad \text{Var}^{-1}[p] = \sigma_p^2,$$

all known;

Data:

$$E[X_i | M_i, N, p] = (N - M_i)p \quad \text{and} \quad \text{Var}[X_i | M_i, N, p] = (N - M_i)(1 - p)p.$$

Now, according to the above specifications, it follows from (4) that the NBLS of p and N , are given by the iterative solution of

$$p_{(m+1)} = p_{(m)} - \frac{A_{11}^{(m)} B_1^{(m)} + A_{12}^{(m)} B_2^{(m)}}{\Delta^{(m)}},$$

and

$$N_{(m+1)} = N_{(m)} - \frac{A_{21}^{(m)} B_1^{(m)} + A_{22}^{(m)} B_2^{(m)}}{\Delta^{(m)}},$$

given initial values N_0 and p_0 , where, without the superscript m , we have:

$$A_{11} = \frac{1}{\sigma_{N_0}^2} + \frac{p}{1-p} \sum_{j=1}^k \frac{1}{N - M_j}, \quad A_{12} = A_{21} = \frac{-k}{1-p},$$

$$A_{22} = \frac{1}{\sigma_{p_0}^2} + \sum_{j=1}^k \frac{(N - M_j)}{p(1-p)}, \quad \Delta = A_{11}A_{22} - A_{12}A_{21},$$

$$B_1 = \frac{p - p_0}{\sigma_{p_0}^2} + \sum_{j=1}^k \frac{(X_j - (N - M_j)p)}{p(1-p)}, \quad B_2 = \frac{N - N_0}{\sigma_{N_0}^2} + \sum_{j=1}^k \frac{(X_j - (N - M_j)p)}{(1-p)(N - M_j)}.$$

Of course, the NBLS of N and p can also be obtained by using the GLIM system as described in Section 3. The quasi-posterior variance of N^* and p^* follows from (6) and are given by

$$\text{Var}(N^*) = \frac{A_{22}}{\Delta} \quad \text{and} \quad \text{Var}(p^*) = \frac{A_{11}}{\Delta}.$$

Table 1: Data from a measurement process for amount of molybdenum in samples (Hunter and Lamboy, 1981).

Description	
$X(NBLS)$	5.3
X_{HL} (Hunter and Lamboy, 1981)	5.3
β (NBLS)	0.98
β_{HL} (Hunter and Lamboy)	0.98
95% Fieller interval for X based on \bar{X} .	(4.536, 6.224)
95% Fieller interval for X based on X_{HL} .	(4.761, 5.853)
95% Chebyshev interval for X based on \bar{X} .	(4.07, 6.532)
Posterior variance for X (Hunter and Lamboy, 1981)	0.0766
$\text{Var}(\bar{X})$ (NBLS)	0.0769
95% Dobrigal interval for X (Dobrigal et al., 1987)	(4.808, 5.807)
95% Sprott Interval for X (Sprott and Viveros, 1984).	(4.7228, 5.8916)

Table 2: Data from a computer simulations of (7) with $e \sim N[0, \sigma^2]$. The X values are as in Hoadley (1970).

$\alpha = 0.0,$	$\beta = 1.0,$	$\sigma^2 = 0.47,$	$m = 1.0,$	$X = 1.0,$	$n = 9$
Description		$\beta_0 = 0.0, \sigma_{\beta}^2 = 100$ $X_0 = 0.0, \sigma_X^2 = 100$	$\beta_0 = 1.0, \sigma_{\beta}^2 = 0.5$ $X_0 = 1.0, \sigma_X^2 = 0.5$		
$X(NBLS)$		0.58	0.82		
X_I (Inverse Estimator)		0.58	0.58		
ML (Maximum Likelihood)		0.65	0.65		
\hat{H} (Hoadley, 1970)		0.35	-		
$\hat{\beta}$ (NBLS)		1.17	1.14		
$\hat{\beta}_{ML}$ (Maximum Likelihood)		1.17	1.17		
95% Chebyshev Interval for X based on X .		(0.186, 1.425)	(0.62, 1.02)		
95% shortest posterior interval for X (Hoadley, 1970).		(-1.82, 2.52)	-		
$Var \hat{X}$	(NBLS)	0.3746	0.227		
Posterior Variance for X (Hoadley, 1970).		1.16	-		

References

- Castledine, B.J. (1981) A Bayesian analysis of multiple capture-recapture sampling for a closed animal population. *Biometrika*, 68, 197-210.
- Dobrigal, A.; Fraser, D.A.S. and Gebotys, R. (1987) Linear calibration and conditional inference. *Comm. Statistics-Theory Meth.*, 16, 1037-1048.
- Duncan, D.B. and Horn, S.D. (1972) Linear dynamic recursive estimation from the viewpoint of regression analysis, *J. Amer. Statist. Assoc.*, 67, 815-821.
- Fraser, D.A.S. (1986) *The Structure Inference*. Wiley: New York.
- Godambe, V.P. (1994) *Linear Bayes and optimal estimation*. Technical Report. University of Waterloo.
- Hartigan, J.A. (1969) Linear Bayes methods. *J.R. Statist. Soc. B*, 31, 446-454.
- Hoadley, B. (1970) A Bayesian look at inverse linear regression. *Technometrics*, 12, 727-736.
- Hunter, W.G. and Lamboy, W.F. (1981) A Bayesian analysis of the linear calibration problem. *Technometrics*, 23, 323-328.
- Kalotay, A.J. (1971) Structural solution to the linear calibration problem. *Technometrics*, 9, 425-440.

- Krutchkoff, R.G. (1967)** Classical and inverse regression methods of calibration. *Technometrics*, 9, 425-439.
- Lawless, J.F. (1981)** Discussion of Hunter and Lamboy's paper. *Technometrics*, 23, 334-335.
- Lawless, J.F. (1981)** Statistical models and methods for lifetime data. Wiley: New York.
- Leite, J.G.; Bolfarine, H. and Rodrigues, J. (1987)** Exact expressions for the posterior mode of finite population size: Capture and recapture sequential sampling. *Rebrape*, 1, 91-100.
- McCullagh, P. (1983)** Quasi-likelihood functions. *Ann. Statist.*, 11, 59-67.
- Minder, C.E. and Whitney, J.B. (1975)** A likelihood analysis of the linear calibration problem. *Technometrics*, 17, 463-471.
- Overton, W.S.; Davies, D.E. (1969)** Estimating the number of animals in wildlife populations. *Wildlife Management techniques*. 3rd. ed. The Wildlife Society, Wash., D.C. Ed. R. H. Tiles, Jr., 403-455.
- Otis, D.L.; Burnham, K.P.; White, G.C.; Anderson, D. R. (1978)** Statistical Inference from capture data on closed animal populations. *Wildlife Monographs*, 62, 1-35.
- Rodrigues, J.; Bolfarine, H.; Leite, J. (1988)** A Bayesian Analysis in closed animal populations from capture recapture experiments with trap response. *Communications in Statistics*, 17, 2, 407-430.
- Sprott, D.A. and Viveros, R. (1984)** The estimation of ratios and related quantities. Technical Report Dept. of Statistics, Waterloo.
- Wedderburn, R.W.M. (1974)** Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, 61, 439-447.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9501 - FAVA, L. On Interpenetrating Order Statistics. São Paulo, IME-USP. 1995. 32p. (RT-MAE-9501)

9502 - FERRARI, P.A., KESTEN, H., MARTÍNEZ, S. R-Positivity, Quasi-Stationary Distributions and Ratio Limit Theorems for a Class of Probabilistic Automata. São Paulo, IME-USP. 1995. 39p. (RT-MAE-9502)

9503 - FERRARI, S.L.P., CORDEIRO, G.M., URIBE-OPAZO, M.A., CRIBARI-NETO, F. Improved Score Tests for One-Parameter Exponential Family Models. São Paulo, IME-USP. 1995. 21p. (RT-MAE-9503)

9504 - ARIZONO, H., BUENO, V.C. Bounds for System Reliability of Used Components under MIFR/ \mathcal{J}_t Assumptions. São Paulo, IME-USP. 1995. 10p. (RT-MAE-9504)

9505 - SINGER, J.M., ANDRADE, D.F. Regression Models for the Analysis of Pretest/Posttest Data. São Paulo, IME-USP. 1995. 8p. (RT-MAE-9505)

9506 - SILVA, A.J., ZUAZOLA, P.L.I., WECHSLER, S. Partial exchangeability and extendibility. São Paulo, IME-USP. 1995. 21p. (RT-MAE-9506)

9507 - MADRUGA, M.G., OCHI-LOHMANN, T.H., OKAZAKI, K., PEREIRA, C.A.B., RABELLO-GAY, M.Z. Bayesian Dosimetry II: Credibility Intervals for Radiation Dose. São Paulo, IME-USP. 1995. 12p. (RT-MAE-9507)

9508 - FERRARI, S.L.P.; URIBE-OPAZO, M.A.; CRIBARI-NETO, F. Second order asymptotics for score tests in exponential family nonlinear models. São Paulo, IME-USP. 1995. 07p. (RT-MAE-9508)

9509 - WECHSLER, S.; BALLAS, D.; CARDOSO, R.H.A. Bayesian Optimization of Quality-of-life. São Paulo, IME-USP. 1995. 06p. (RT-MAE-9509)

9510 - BOTTER, D.A.; CORDEIRO, G.M. Bartlett Corrections for Generalized Linear Models with Dispersion Covariates. São Paulo, IME-USP. 1995. 31p. (RT-MAE-9510)

9511 - PAULA, G.A.; ROJAS, O.V. On Restricted Hypotheses in Extreme Value Regression Models. São Paulo, IME-USP. 1995. 14p. (RT-MAE-9511)

9512 - ARIZONO, H.; BUENO, V.C. Bounds for System Reliability of Used Components Under NBU₃ Assumptions. São Paulo, IME-USP. 1995. 10p. (RT-MAE-9512)

9513 - FERRARI, S.L.P.; BOTTER, D.A.; CORDEIRO, G.M.; CRIBARI-NETO, F. Second and Third Order Bias Reduction for One-Parameter Family Models. São Paulo, IME-USP. 1995. 07p. (RT-MAE-9513)

9514 - BELITSKY, V.; KOHAYAKAWA, Y. Inference on the Flip Rates from Density Dynamics for a Class of Spin-Flip Systems. São Paulo, IME-USP. 1995. 21p. (RT-MAE-9514)

9515 - BELITSKY, V. The change of sign of derivatives of the density function of a spin flip system with the order of derivation. São Paulo, IME-USP. 1995. 27p. (RT-MAE-9515)

9516 - PAULA, G.A. On Approximation of the Level Probabilities for Testing Ordered Parallel Regression Lines. São Paulo, IME-USP. 1995. 15p. (RT-MAE-9516)

9517 - PAULA, G.A. The Concept of Leverage in Restricted Regression. São Paulo, IME-USP. 1995. 09p. (RT-MAE-9517)

9518 - YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H. Bayes Estimation of the Number of Component Processes of a Superimposed Process. São Paulo, IME-USP. 1995. 19p. (RT-MAE-9518)

9519 - SIMONIS, A. Metastability of the d-Dimensional Contact Process. São Paulo, IME-USP. 1995. 17p. (RT-MAE-9519)

9520 - FONTES, L.R.G.; ISOPI, M.; SIDORAVICIUS, V. Analyticity of the Density and Exponential Decay of Correlations in 2-d Bootstrap Percolation. São Paulo, IME-USP. 1995. 17p. (RT-MAE-9520)

9521 - FERRARI, S.L.P.; BOTTER, D.A.; CRIBARI-NETO, F. Local power of three classic criteria in generalised linear models with unknown dispersion. São Paulo, IME-USP. 1995. 5p. (RT-MAE-9521)

9522 - CORDEIRO, G.M.; FERRARI, S.L.P.; A general method for approximating to the distribution of some statistics. São Paulo, IME-USP. 1995. 18p. (RT-MAE-9522)

9523 - BOLFARINE, H.; LIMA, C.; SANDOVAL, M.C. Linear Calibration in Functional Regression Models. São Paulo, IME-USP. 1995. 24p. (RT-MAE-9523)

9524 - WECHSLER, S.; BRANCO, M.D. The exercise of coherent statistical inference. São Paulo, IME-USP. 1995. 8p. (RT-MAE-9524)

9525 - ROJAS, O.V.; BOLFARINE, H.; PAULA, G.A. On restricted hypothesis in structural regression models. São Paulo, IME-USP. 1995. 16p. (RT-MAE-9525)

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

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05389-970 - São Paulo, Brasil