Biometrika (1995), **82**, 2, *pp*. 439–46
Printed in Great Britain

Bayesian methods for categorical data under informative general censoring

BY CARLOS DANIEL MIMOSO PAULINO
Universidade Técnica de Lisboa, Departamento de Matemática, Instituto Superior Técnico, Centro de Análise e Processamento de Sinais, Av. Rovisco Pais, 1096 Lisboa Codex, Portugal

AND CARLOS ALBERTO DE BRAGANÇA PEREIRA
Universidade de São Paulo, Departamento de Estatística, Instituto de Matemática e Estatística, Caixa Postal 66.281, CEP 09389-970, São Paulo, Brasil

SUMMARY

This paper develops a Bayesian approach to the problem of incomplete categorical data informatively censored where the reported sets are not restricted to follow any specific pattern. It generalises that introduced by Paulino & Pereira (1992) in not requiring a censoring pattern by partitions of the set of sampling categories. Some extensions are also discussed.

Some key words: Bayesian analysis; Dirichlet and generalised Dirichlet distributions; Informative censoring process; Unidentifiability.

1. INTRODUCTION

Standard models for censored categorical data are usually nonidentifiable. Paulino (1991) presents a survey on the subject. This problem has been bypassed in the non-Bayesian literature by restricting the analysis to two types of procedures: see, e.g., Little & Rubin (1987). One of them assumes that the censoring mechanism is ignorable; i.e. the unknown parameter of the distribution describing the censoring mechanism is unrelated to the parameter of interest. Classical inferences must then be interpreted as conditional on the observed censoring pattern (Paulino, 1991). The other procedure assumes that the nuisance parameters of the censoring mechanism are known. In this case, the report process, defining the censoring mechanism, can be informative but the ensuing analysis is unavoidably conditional on the assumed values of the nuisance parameters. For more details about informative versus noninformative report processes see Rubin (1976) and Dawid & Dickey (1977).

Although the effects of unidentifiability on Bayesian inference are less drastic than on classical inference, almost all Bayesian procedures have been developed under the same kind of restrictions on censoring mechanisms: see Dickey, Jiang & Kadane (1987) and references therein. These procedures are not suitable for incomplete data problems where prior beliefs violate the assumption of a noninformative incompleteness mechanism.

Paulino & Pereira (1992), extending the work of Basu & Pereira (1982), developed a Bayesian solution which neither assumes an ignorable censoring process nor proceeds conditionally on fixed parameters of this process; compare Kadane (1985). This also differs from other methods which assume nonignorable nonresponse mechanisms, e.g. Park & Brown (1994). The procedure described by Paulino & Pereira (1992) considers an informative censoring mechanism for data patterns where the reports can be structured into partitions of the set of sampling categories. There is no common element between any pair of those partitions. This is relevant for contingency tables where censoring is usually defined through incomplete classification only into marginal tables. However, this does not cover all cases, since the incomplete categorisation can occur in an entirely arbitrary collection of subsets such as the example of § 4 of the present paper.

The main purpose of the present paper is to develop a Bayesian solution to the problem under an informative general censoring pattern. This solution, based on Dirichlet priors for all the model parameters, is described in §§ 2 and 3 and illustrated in § 4. Section 5 is devoted to a summary of other less tractable solutions developed in C. D. M. Paulino's 1988 University of São Paulo doctoral thesis and based on more general prior distributions capable of expressing wider prior beliefs. In § 6, some brief conclusions are drawn.

2. LIKELIHOOD AND PRIOR DISTRIBUTION

Consider a population partitioned into m categories from which a random sample of size n is to be chosen. Let $\theta' = (\theta_1, \ldots, \theta_m)$ be the vector of positive probabilities of the categories such that $\sum_i \theta_i = 1$. The sampling process can be defined by random variables W_k $(k = 1, \ldots, n)$, where $W_k = i$ if the kth sample unit belongs to the ith category, $i \in \{1, 2, \ldots, m\}$. We then obtain a finite sequence of independent and identically distributed random variables with a multivariate Bernoulli distribution parametrised by θ.

In the incomplete data problem, the $\{W_k\}$ are not fully observed for some, possibly all, sample units. For each sample unit, it is known only that it belongs to some reported nonempty subset d of $\{1, 2, \ldots, m\}$.

Let r_d be the cardinality of d. A unit is said to be fully categorised, or uncensored, if $r_d = 1$. One says that there is a partial or complete censoring depending on whether $1 < r_d < m$ or $r_d = m$. Denote by λ_{di} the probability of a unit which belongs to category i being reported to lie in class d. We assume no misclassification errors; that is $\lambda_{di} = 0$ whenever $i \notin d$. However, with minor changes in the model, the case of untruthful reports can be covered too.

Let D be the class of possible reported subsets, that is $d \in D$ if and only if $\lambda_{di} > 0$ for some $i = 1, \ldots, m$. We assume that D includes the cases of no censoring, $D_0 = \{\{i\}, i = 1, \ldots, m\}$. However, if $\lambda_{ii} = 0$ for some i, our results can be easily adapted by removing the associated terms pertaining to fully categorised units. Also, D may include the complete censoring case $\{1, 2, \ldots, m\}$. The reporting process is described by associating with each sample unit k a random variable R_k, taking values in $\{1, 2, \ldots, c\}$, where c is the cardinality of D. Thus, λ_{di} is the conditional probability that R_k takes on the value indicating the report d, given that W_k indicates the category i, $i \in d$, for any k.

The joint outcomes of the sampling and reporting process are the values of the sequence $\{(W_k, R_k), k = 1, \ldots, n\}$ of random vectors, independent and identically distributed with joint distribution indexed by the parameters θ and $\beta_i' = (\lambda_{di}, d \in D_i)$, for $i = 1, \ldots, m$, where $D_i = \{d \in D : i \in d\}$.

The values of R_k $(k = 1, \ldots, n)$ define the observable data from which we wish to draw inferences about θ, the parameter of interest. The likelihood function of θ and $\beta = (\beta_i', i = 1, \ldots, m)'$ is

$$L(\theta, \beta \mid \{R_k\}) = \prod_{d \in D} \left(\sum_{i \in d} \theta_i \lambda_{di} \right)^{n_d}$$

$$= \prod_{i=1}^{m} (\theta_i \lambda_{ii})^{n_i} \prod_{d \in D_c} \left(\sum_{i \in d} \theta_i \lambda_{di} \right)^{n_d} \qquad (1)$$

where n_d, respectively n_i, is the frequency count of reports in d, respectively $\{i\}$, and $D_c = D - D_0$ with cardinality $l = c - m$. The data may be summarised as $N = (N_0', N_c')'$, where $N_0' = (n_i, i = 1, \ldots, m)$ and $N_c' = (n_d, d \in D_c)$.

For convenience we shall sometimes use other parametrisations. One of them refers to the joint probabilities $\{\mu_{di} = \theta_i \lambda_{di}\}$, in number $r = m + \sum_{d \in D_c} r_d$, where r_d is the number of categories in d. These probabilities can be arranged in a two-way $m(m + l)$ table, defined by the sampling and reporting categories, which contains $m(m + l) - r$ structural zeros, since $\mu_{d_i} = 0$ whenever $i \notin d$.

The likelihood expression under the parametrisation $\mu = (\mu_{di})$ shows that the model is unidentifiable. The parameters of interest, $\theta_i = \sum_{d \in D_i} \mu_{di}$, are usually unidentifiable as well. They cannot be determined by the identified functions, $\{\sum_{i \in d} \mu_{di},\, d \in D\}$, unless further assumptions about $\{\lambda_{di}\}$ are made. This is why most procedures assume that the $\{\lambda_{di}\}$ are either independent of i, for each $d \in D_i$, or are known. This practice cannot be entirely justified from a Bayesian viewpoint, as shown by Paulino & Pereira (1992).

For the purpose of identifying which parameters are updated and which are not, we consider another parametrisation (γ, α). Here, $\gamma = (\gamma_0, \gamma_d, d \in D_c)$ contains the marginal probabilities of full classification,

$$\gamma_0 = \sum_{i=1}^{m} \mu_{ii} \equiv \sum_{i=1}^{m} \theta_i \lambda_{di},$$

and partial classification in each $d \in D_c$,

$$\gamma_d = \sum_{i \in d} \mu_{di} \equiv \sum_{i \in d} \theta_i \lambda_{di},$$

while $\alpha = (\alpha_0', \alpha_d', d \in D_c)'$ contains the conditional probabilities of each category given each kind of report,

$$\alpha_0 = (\alpha_{ii}, i = 1, \ldots, m)', \quad \alpha_{ii} = \mu_{ii}/\gamma_0, \quad \sum_{i=1}^{m} \alpha_{ii} = 1,$$
$$\alpha_d = (\alpha_{id}, i \in d)', \quad \alpha_{id} = \mu_{di}/\gamma_d, \quad \sum_{i \in d} \alpha_{id} = 1, \quad d \in D_c.$$

If P denotes the partition indicator matrix, a block diagonal matrix with diagonal blocks 1_m and 1_{r_d}, $d \in D_c$, then we can write $\gamma = P'\mu$, under a suitable ordering of μ_{di} in μ. Note also that $\theta_i = \sum_{d \in D_i} \gamma_d \alpha_{id}$ $(i = 1, \ldots, m)$. With this new parametrisation, the form of the likelihood,

$$L(\gamma, \alpha \mid \{R_k\}) = \left(\gamma_0^{n_0} \prod_{d \in D_c} \gamma_d^{n_d} \right) \left\{ \prod_{i=1}^{m} \alpha_{ii}^{n_i} \prod_{d \in D_c} \left(\sum_{i \in d} \alpha_{id} \right)^{n_d} \right\}, \tag{2}$$

where $n_0 = 1_m' N_0$, shows that the observations are used for updating just the parameters γ and α_0. Note that $\sum_{i \in d} \alpha_{id} = 1$, $d \in D_c$.

For most part of this paper we consider a prior distribution for μ defined by the Dirichlet distribution with hyper-parameter $a = (a_{di}) \in (\mathbb{R}_+)^r$, which we denote by $\mu \sim D(a)$. This is equivalent, in the other parametrisation, to:

$$\theta \sim D(a_{.*}), \quad a_{.*} = (a_{.i}, i = 1, \ldots, m)', \quad a_{.i} = \sum_{d \in D_i} a_{di},$$
$$\beta_i \sim D(a_i), \quad a_i = (a_{di}, d \in D_i)' \quad (i = 1, \ldots, m), \tag{3}$$

with $\theta, \beta_1, \ldots, \beta_m$ mutually independent; and

$$\gamma \equiv P'\mu \sim D(P'a),$$
$$\alpha_0 \equiv (\alpha_{ii}, i = 1, \ldots, m)' \sim D(a_0), \quad a_0 = (a_{ii}, i = 1, \ldots, m)', \tag{4}$$
$$\alpha_d \equiv (\alpha_{id}, i \in d)' \sim D(a_d), \quad a_d = (a_{di}, i \in d)', \quad d \in D_c,$$

with γ, α_0 and all α_d, $d \in D_c$, mutually independent.

Note that the elements of $P'a$ are the sums of the components of a_0 and a_d for all $d \in D_c$, and $a = (a_0', a_d', d \in D_c)'$.

3. A posteriori results

From the functional form of the likelihood of (γ, α), their posterior distribution is

$$\gamma \mid N \sim D(P'a + x), \quad x = (n_0, N_c')', \quad n_0 = 1_m' N_0,$$
$$\alpha_0 \mid N \sim D(a_0 + N_0), \quad \alpha_d \mid N \sim \alpha_d \sim D(a_d), \quad d \in D_c, \tag{5}$$

γ, α_0 and all α_d's being conditionally mutually independent given N.

In terms of μ, the posterior distribution is a member of the family of generalised Dirichlet distributions introduced by Dickey (1983). In a notation somewhat similar to Dickey's, it may be expressed as $\mathscr{D}_r^l(a+x_0, Q_c, N_c)$, where x_0 is a $r \times 1$ vector obtained from the null vector by replacing the first m elements by N_0. Recall that l is the number of reporting categories involving some censoring, and r is the number of joint probabilities different from zero, i.e. the dimension of μ. The symbol Q_c' indicates the $l \times r$ submatrix of P' formed by the last l rows. Note that $P' = (J, Q_c)'$, with $J' = (1_m', 0_{(r-m)}')$.

The density of μ on the $(r-1)$-dimensional simplex is

$$g(\mu|N) = g(\mu|A)\left\{\prod_{d \in D_c}\left(\sum_{i \in d} \mu_{di}\right)^{n_d} \Big/ \mathscr{R}(A, Q_c, -N_c)\right\}, \tag{6}$$

where $A = a + x_0$ and $g(\mu|A)$ denotes the density of the Dirichlet $D(A)$ distribution.

The quantity $\mathscr{R}(A, Q_c, -N_c)$, the $D(A)$ expected value of the product of powers of linear forms of μ indicated in (6), is a Carlson's bidimensional hypergeometric function, here expressible as a ratio of vector-argument beta functions, or Dirichlet complete integrals,

$$\mathscr{R}(A, Q_c, -N_c) \equiv \mathscr{R}(A, Q, -N_c^+) = B(Q'A + N_c^+)/B(Q'A), \tag{7}$$

where N_c^+ is the vector N with N_0 replaced by the null vector and $Q' = (I_0', Q_c)'$, with $I_0 = (I_{(m)}, 0_{(m,r-m)})$, $I_{(m)}$ being the identity matrix of order m. This simple form of $\mathscr{R}$ is a consequence of Q being a partition-indicator matrix whose columns indicate the $m+l$ parts of a partition of the cells with $\mu_{di} > 0$. In particular, $Q'\mu \sim D(Q'A)$ whenever $\mu \sim D(A)$.

The posterior distribution of μ can be expressed as a finite mixture of Dirichlet distributions by applying the multinomial expansion by every power appearing in the second factor of (6). Letting $\{y_{di}\}$ be the hypothetical frequencies, arranged into vectors y_d, $d \in D_c$, underlying the observed total counts n_d over categories $i \in d$, the density (6) is a mixture of the densities $D(a+y)$, where

$$a + y = (a_0' + N_0', a_d' + y_d', d \in D_c)'.$$

The mixing distribution comes from the predictive distribution of $\{y_d\}$ and corresponds to the l conditionally independent Dirichlet–Multinomial distributions, $DM(n_d, a_d)$, for y_d given n_d.

This representation for the updated distribution of μ shows that the posterior distribution of θ is the same mixture of the corresponding Dirichlet distributions with parameter $S(y) = (s_i, i = 1, \ldots, m)'$, where

$$s_i = a_{ii} + n_i + \sum_{d \in D_i \cap D_c} (a_{di} + y_{di}).$$

Taking into account the form of Q, the posterior mixed moments of μ can be expressed by

$$E\left(\prod_{d \in D}\prod_{i \in d} \mu_{di}^{b_{di}} \middle| N\right) = \frac{B(A+b)}{B(A)} \frac{B\{Q'(A+b) + N_c^+\}}{B\{Q'(A+b)\}} \frac{B(Q'A)}{B(Q'A + N_c^+)}. \tag{8}$$

The computation of the posterior mixed moments of θ can be carried out from (8). Alternatively, we may make use of the posterior distribution (5) in the relations

$$\theta_i = \gamma_0 \alpha_{ii} + \sum_{d \in D_i \cap D_c} \gamma_d \alpha_{id} \quad (i = 1, \ldots, m).$$

Thus, it is easy to show that the posterior mean of θ_i $(i = 1, \ldots, m)$ is given by

$$E(\theta_i|N) = \frac{a_.}{a_. + n} E(\theta_i) + \frac{n}{a_. + n}\left(\frac{n_i}{n} + \sum_{d \in D_i \cap D_c} \frac{n_d}{n} \frac{a_{di}}{\sum_{i \in d} a_{di}}\right), \tag{9}$$

where $a_.$ is the sum of the elements of a.

This expression shows that the posterior mean of θ_i $(i = 1, \ldots, m)$ is a weighted mean of the prior mean and the sum of a fraction of the sampling proportions of each report allocated to the ith category. Each fraction is defined by the posterior, or equivalently prior, mean of the conditional probability of the category i given the report in each d that includes it. Note also that the second term in (9) is proportional to the sum, for all appropriate d's, of the predictive mean of the 'fictitious' frequencies $\{y_{di}, d \in D_i\}$ $(y_{ii} = n_i)$, given the observed data.

With some further algebra, the calculation of the posterior covariance matrix of θ is also straightforward.

In many situations, such as when the categories form a contingency table, there is interest in computing generalised posterior moments of θ, that is

$$E\left(\prod_{i=1}^{m} \theta_i^{b_i} \middle| N\right),$$

where some b_i's are nonpositive integers. These quantities are also given by a ratio of Carlson functions but they are more complicated to compute because the argument matrix of the hypergeometric function in the numerator is not a partition indicator. If there are only a few censored data, one practical computational alternative is to use the mixture representation of the posterior density of θ, yielding

$$E\left(\prod_{i=1}^{m} \theta_i^{b_i} \middle| N\right) = \sum \prod_{d \in D_c} \binom{n_d}{y_d} \{B(a+y)/B(A)\} \, [B\{S(y)+b\}/B\{S(y)\}] \times \{B(Q'A+N_c^+)/B(Q'A)\}^{-1}, \tag{10}$$

where

$$b = (b_i, i = 1, \ldots, m)', \quad \binom{n_d}{y_d} = n_d! \Big/ \prod_{i \in d} y_{di}!$$

and $\sum$ refers to all y_d's such that $\sum_{i \in d} y_{di} = n_d$, $d \in D_c$.

Jiang, Kadane & Dickey (1992) explore computational methods for hypergeometric functions arising in Bayesian analyses conditional on the censoring parameters, or using ignorable censoring mechanism. We believe that some of these methods are useful in our set-up. In particular, the procedure leading to (10) can be seen as an application of the so-called expansion method. In fact, the numerator $\mathscr{R}$ hypergeometric function is represented in (10) by a linear combination of other $\mathscr{R}$ functions, which are expressible in a closed form as ratios of beta functions. For extensive incomplete data, spread over many varied sets, this exact method will probably be cumbersome to apply. In this case, the use of approximations such as the Monte Carlo method, as suggested by Jiang et al. (1992), will certainly be a valuable alternative.

4. Illustration

We use a simple example to illustrate the application of the method described in § 3. This is concerned with the determination of the degree of sensitivity to dental caries, categorised in three risk levels: low, medium and high. The technique uses the coloration obtained on reaction of a spittle sample with a chemical. Based on the colour obtained, each subject is to be assigned a corresponding risk level.

Due to problems with this simplified procedure, it is expected that certain subjects may not be fully categorised, caused by inability to distinguish adjacent categories. This is confirmed by the observed outcomes. Out of 97 subjects analysed only 51 were fully categorised. A total of 28 subjects were only classified as low or medium risk, and 18 as medium or high risk.

Labelling the low, medium and high levels by 1, 2 and 3, respectively, we shall assume, as seems natural, that the possible reporting subsets are just those observed in this experiment, namely $\{i\}$, for $i = 1, 2, 3$, $\{1, 2\}$ and $\{2, 3\}$. The observed data were $n_1 = 14$, $n_2 = 17$, $n_3 = 20$, $n_4 = 28$ and $n_5 = 18$, where the indices 4 and 5 label the subsets $\{1, 2\}$ and $\{2, 3\}$, respectively. In the notation of the preceding sections, $N_0 = (14, 17, 20)'$ and $N_c = (28, 18)'$.

The probabilities of the outcomes of the sampling and reporting processes are displayed in

Table 1. *Joint and marginal probabilities*

Sampling categories	Reporting sets 1	2	3	4	5	Total
1	μ_{11}	0	0	μ_{41}	0	θ_1
2	0	μ_{22}	0	μ_{42}	μ_{52}	θ_2
3	0	0	μ_{33}	0	μ_{53}	θ_3
Total	μ_{11}	μ_{22}	μ_{33}	γ_4	γ_5	1

$\gamma_0 = \mu_{11} + \mu_{22} + \mu_{33}$.

Table 1. The experimenter feels that the conditional probability of an incomplete classification may not be the same for subjects of different actual risk degrees. Concretely, she believes that a subject of medium risk tends to have a greater probability of being classified into $\{1, 2\}$ than one with a low risk, and into $\{2, 3\}$ than one with a high risk.

Suppose that the prior opinion of the experimenter can be adequately described by the distribution (3) with

$$a_1' \equiv (a_{11}, a_{41}) = (3, 1), \quad a_2' \equiv (a_{22}, a_{42}, a_{52}) = (1, 1, 1), \quad a_3' \equiv (a_{33}, a_{53}) = (3, 1).$$

Hence, the prior distribution of θ is $D(4, 3, 4)$ independently of

$$\beta_1 = (\lambda_{11}, \lambda_{41})', \quad \beta_2 = (\lambda_{22}, \lambda_{42}, \lambda_{52})', \quad \beta_3 = (\lambda_{33}, \lambda_{53})'.$$

The corresponding prior distribution (4) has

$$P'a = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad a = (7, 2, 2)',$$

$$a_0 = (3, 1, 3)', \quad a_4 = (1, 1)' = a_5.$$

The parameters $\gamma \equiv P'\mu = (\gamma_0 \ \gamma_4 \ \gamma_5)'$, α_0, α_4 and α_5 are a posteriori mutually independent, with

$$\gamma | N \sim D(58, 30, 20), \quad \alpha_0 | N \sim D(17, 18, 23), \quad \alpha_4 | N \sim \alpha_5 | N \sim Be(1, 1).$$

We obtain the following posterior quantities for the probabilities of the various risk levels:

$$E(\theta_i | N) = \begin{cases} 0{\cdot}2963 & (i = 1), \\ 0{\cdot}3982 & (i = 2), \\ 0{\cdot}3055 & (i = 3), \end{cases} \quad \operatorname{var}(\theta_i | N) = \begin{cases} 0{\cdot}79 \times 10^{-2} & (i = 1), \\ 1{\cdot}07 \times 10^{-2} & (i = 2), \\ 0{\cdot}45 \times 10^{-2} & (i = 3), \end{cases}$$

$$\operatorname{cov}(\theta_1, \theta_2 | N) = -0{\cdot}70 \times 10^{-2}.$$

A sensitivity analysis shows that the calculated quantities are fairly sensitive to prior assumptions about the value of a. This is to be expected in the light of the results found for the special case considered by Paulino & Pereira (1992).

5. SOME FURTHER RESULTS

The Bayesian solution described above rests upon the prior distribution (3), which displays linear relations among the hyper-parameters of $\{\beta_i\}$ and θ. If, in (3), $a_{.*}$ is replaced by $c = (c_1, \ldots, c_m)' \in (\mathbb{R}_+)^m$, the prior distribution for μ becomes the generalised Dirichlet distribution $\mathscr{D}_r^m(a, Z, c - a_{.*})$, where a, as before, is a vector formed by all the prior parameters of $\{\beta_i\}$, and Z is the matrix indicating the partition $\{D_i, i = 1, \ldots, m\}$. When $c_i > a_{.i}$, this distribution expresses a prior belief that the grouped probabilities $\{\sum_{d \in D_i} \mu_{di}, i = 1, \ldots, m\}$ are better than their components.

In this case the posterior distribution of μ is still a generalised Dirichlet distribution, but with the argument matrix augmented by the matrix Q_c. As a consequence, the posterior moments of θ, defined by a ratio of Carlson's hypergeometric functions, present computational difficulties similar to those for the generalised moments of θ referred to in the preceding section.

This alternative prior distribution for μ is not appropriate if the prior knowledge about the probabilities of each type of classification is greater than about the component probabilities $\{\mu_{di}, i \in d\}$. This can be achieved by a higher multiplicity generalised Dirichlet distribution with a matrix argument defined by $[Z, P]$. The term multiplicity is here used in the sense of Dickey (1983). In this case, the observed censoring pattern is included in the argument matrix of the prior distribution of μ, and therefore the posterior distribution belongs to the same family.

The use of this latter prior distribution for μ implies violation of the independent relations in (3) and (4), as referred to in Paulino's thesis. The posterior distribution for the parameters (γ, α) no longer presents a functional form as neat as that in (5). As expected, the computation of posterior characteristics becomes more complicated. In particular, neither of Carlson's hypergeometric functions whose ratio defines the posterior moments of θ has a simple explicit form.

The preceding prior distribution for μ still implies a conditional generalised Dirichlet distribution for θ, given $\{\beta_i\}$, with argument matrix, say B, depending on the values of $\{\beta_i\}$. We may take advantage of this fact to study the dependence of inferences about θ on β. Given the form of the conditional likelihood, the conditional posterior distribution of θ belongs to the same family as the prior. The conditional posterior moments of θ are given by the corresponding moments of a $D(c+N_0)$ distribution multiplied by a ratio of $\mathscr{R}$ integrals associated with the matrix B. These results are thus a generalisation of those of Dickey et al. (1987).

As to the implementation of these methods, two issues remain open: elicitation of this type of prior distribution and development of computational strategies for the evaluation of characteristics of interest. For the latter, the conclusions of Jiang et al. (1992) will be helpful.

6. Conclusions

The method described in §§ 2 and 3 and illustrated in § 4 is a Bayesian solution to the problem of incomplete categorical data, with a general incompleteness mechanism. The problem can be solved without assumptions to make the parameters of interest identifiable. These assumptions are used in most publications, both classical and Bayesian, mainly because the reporting process turns out to be ignorable. The description of our method makes it evident that many computations of interest are easier to execute than in the case of the noninformative reporting process analysed by Dickey et al. (1987).

When the observed data can be structured into partitions of the set of sampling categories, this method simplifies further, as detailed by Paulino & Pereira (1992).

The generalisations and variants of our method referred to in § 5 are less straightforward to apply, requiring use of approximations for the $\mathscr{R}$ integral and its ratios, like those explored by Jiang et al. (1992).

Acknowledgement

We are grateful to FAPESP, the State of S. Paulo Research Foundation, JNICT, Junta Nacional de Investigação Científica e Tecnológica, and Fundação Calouste Gulbenkian, Portugal, for partial support of our research and of the first author's visits to São Paulo. We thank the referee for many helpful suggestions.

References

Basu, D. & Pereira, C. A. B. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse. *J. Statist. Plan. Inf.* **6**, 345–62.

DAWID, A. P. & DICKEY, J. M. (1977). Likelihood and Bayesian inference from selectively reported data. *J. Am. Statist. Assoc.* **72**, 845–50.

DICKEY, J. M. (1983). Multiple hypergeometric functions: probabilistic interpretations and statistical uses. *J. Am. Statist. Assoc.* **78**, 628–37.

DICKEY, J. M., JIANG, T. J. & KADANE, J. B. (1987). Bayesian methods for censored categorical data. *J. Am. Statist. Assoc.* **82**, 773–81.

JIANG, T. J., KADANE, J. B. & DICKEY, J. M. (1992). Computation of Carlson's multiple hypergeometric functions $\mathscr{R}$ for Bayesian applications. *J. Statist. Computat. Graph.* **1**, 231–51.

KADANE, J. B. (1985). Is victimization chronic? A Bayesian analysis of multinomial missing data. *J. Economet.* **29**, 47–67.

LITTLE, R. J. A. & RUBIN, D. B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley.

PARK, T. & BROWN, M. B. (1994). Models for categorical data with nonignorable nonresponse. *J. Am. Statist. Assoc.* **89**, 44–52.

PAULINO, C. D. M. (1991). Analysis of incomplete categorical data: a survey of the conditional maximum likelihood and weighted least squares approaches. *Brazilian J. Prob. Statist.* **5**, 1–42.

PAULINO, C. D. M. & PEREIRA, C. A. B. (1992). Bayesian analysis of categorical data informatively censored. *Commun. Statist.* A **21**, 2689–705.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–92.

[*Received October* 1993. *Revised October* 1994]