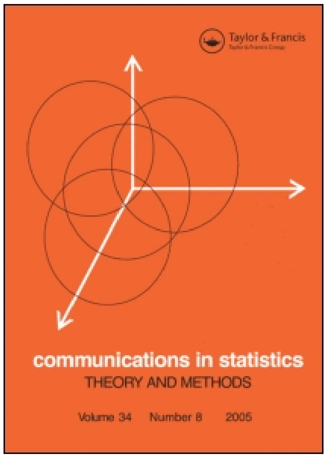

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

Bayesian analysis of categorical data informatively censored

Carlos Daniel Mimoso Paulino [a]; Carlos Alberto De Bragança Pereira [b]
[a] Dep. de Matemática, Univ. Técnica de Lisboa, Lisboa CODEX, Portugal
[b] Dep. de Estatistica, Universidade de Sào Paulo, São Paulo, Brasil

Online Publication Date: 01 January 1992
To cite this Article: Paulino, Carlos Daniel Mimoso and Pereira, Carlos Alberto De Bragança (1992) 'Bayesian analysis of categorical data informatively censored', Communications in Statistics - Theory and Methods, 21:9, 2689 - 2705
To link to this article: DOI: 10.1080/03610929208830937
URL: http://dx.doi.org/10.1080/03610929208830937

COMMUN. STATIST.—THEORY METH., 21(9), 2689-2705 (1992)

BAYESIAN ANALYSIS OF CATEGORICAL DATA INFORMATIVELY CENSORED

Carlos Daniel Mimoso Paulino

Univ. Técnica de Lisboa, Dep. de Matemática - IST
Av. Rovisco Pais, 1096 Lisboa CODEX - Portugal

Carlos Alberto de Bragança Pereira

Universidade de São Paulo, Dep. de Estatística - IME
Caixa Postal 20570 - Ag. Iguatemi, CEP 01498 - São Paulo - Brasil

Key words: Bayesian operation, Dirichlet and generalized Dirichlet distributions, incomplete categorical data, informative censoring process.

AMS classefication: 62A15, 62H17, 62C10.

ABSTRACT

This article presents a general Bayesian analysis of incomplete categorical data considered as generated by a statistical model involving the categorical sampling process and the observable censoring process. The novelty is that we allow dependence of the censoring process paramenters on the sampling categories; i.e., an informative censoring process. In this way, we relax the assumptions under which both classical and Bayesian solutions have been developed. The proposed solution is outlined for the relevant case of the censoring

pattern based on partitions. It is completely developed for a simple but typical example. Several possible extensions of our procedure are discussed in the final remarks.

1. INTRODUCTION

In order to better illustrate and clarify the problem of analysing categarical data informatively censored, we decided to use the following practical example (Paulino 1990) throughout the paper:

Example: To evaluate the influence of physical exercise on the pulmonary function of asthmatic children, $n = 167$ of such children were studied. The experimental protocol specified that each child should be submitted to cycloergometric exercises for two fixed periods of five and seven minutes, after which the status of the child with respect to bronchial spasm (positive or negative) should be recorded.

For reasons related to the condition of each child, observations for both sessions (absence of censoring) were obtained for only 23 children while 24 children did not present any report (total censoring). The remaining 120 children were only reported at a single session (partial censoring), 81 at five minutes and 39 at seven minutes. Denoting presence and absence of bronchial spasm by "+" and "−", respectively, Table I presents the results relative to the 23 uncensored children. Among the 81 children reported only at five minutes, 50 responded positively (and 31 negatively). Among the 39 children reported only at seven minutes, 27 responded positively (and 12 negatively).

Note that each sampling unit is classified into one of the following categories, defined by the possible results of the two sessions: $(+,+)$, $(+,-)$, $(-,+)$ and $(-,-)$. To these possible results we associate respectively the vectors $\mathbf{e}_1 = (1,0,0,0)$, $\mathbf{e}_2 = (0,1,0,0)$, $\mathbf{e}_3 = (0,0,1,0)$ and $\mathbf{e}_4 = (0,0,0,1)$ that form the canonical basis of $\mathbb{R}^4$. As usual, this sampling process is modelled by the Bernoulli multivariate distribuition. That is, $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n$ $(n = 167)$ are (conditionally) independent (given $\boldsymbol{\theta}$) random vectors associate to the sampling units such that, for all $k = 1, 2, \ldots, n$ and $j = 1, 2, 3, 4$,

$$\Pr\{\mathbf{W}_k = \mathbf{e}_j\} = \theta_j \ ,$$

TABLE I
Observed frequencies related to bronchial spasm induced by exercise (BIE) at 5 and 7 minutes

5m \ 7m	BIE +	BIE −	Total
BIE +	12	4	16
BIE −	5	2	7
Total	17	6	23

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ belongs to the tridimensional simplex

$$\mathcal{S}_4 = \{(s_1, s_2, s_3, s_4) : s_i > 0 \ , \quad s_1 + s_2 + s_3 + s_4 = 1\}$$

Questions of interest include the distributional evalution of $\boldsymbol{\theta}$ in order to investigate whether the frequency of bronchial spasm changes or not with exercise time; that is, whether θ_2 is near to θ_3.

Unlike the standard situation, in the present example not all $\mathbf{W}_k$'s are observed completely. For the interpretation of what is in fact observed, the sampling process defined above is insufficient. While our interest is directed to the elements of $\boldsymbol{\theta}$ (sampling process paramenters), it is necessary to include a report (or censoring) process which indicates the type of censorship each sampling unit may suffer.

The structure of the example (cross-classified categorical data) permits us to define the report process by the vectors $\mathbf{R}_k$, which indicate the kind of censoring suffered by unit k. Hence, if unit k suffers no censoring, $\mathbf{R}_k = (1, 0, 0, 0)$, indicating that the response reported for the k-th unit is an element of the set $\{(+,+),(+,-),(-,+),(-,-)\}$. If there is no report in the session of 7 minutes for the k-th unit, $\mathbf{R}_k = (0, 1, 0, 0)$, indicating that the response reported for the k-th unit is an element of the set $\{(+,\cdot),(-,\cdot)\}$. If there is no report in the session of 5 minutes for the k-th unit, $\mathbf{R}_k = (0, 0, 1, 0)$, indicating that the response reported for the k-th unit is an element of the set $\{(\cdot,+),(\cdot,-)\}$. Finally. $\mathbf{R}_k = (0,0,0,1)$ indicates that for the k-th unit there are no results of either sessions.

Using standard statistical terminology, as a first assumption on the model. we consider that the vectors $(\mathbf{R}_k, \mathbf{W}_k)$, $k = 1,\ldots,n$, form a sequence of independent and identically distributed random quantities. In adition. each of the conditional distributions of $(\mathbf{R}_k|\mathbf{W}_k)$ is multivariate Bernoulli. More precisely, we define for every $i,j = 1,2,3,4$,

$$\Pr\{\mathbf{R}_k = \boldsymbol{\epsilon}_i|\mathbf{W}_k = e_j\} = \lambda_{ij} \ ,$$

where. for every $j = 1,2,3,4$.

$$\boldsymbol{\lambda}_j = (\lambda_{1j}, \lambda_{2j}, \lambda_{3j}, \lambda_{4j})$$

is an element of the simplex $\mathcal{S}_4$. Also note that $\boldsymbol{\epsilon}_i$ is an element of the set $\{(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\}$. This definition of report process parameters imply coherence of the reporting with the sampling process in the sense that the reported data that do not contradict the missing information. For example, for any unit k, λ_{12} is the conditional probability that the k-th unit is reported in both sessions given that $\mathbf{W}_k = \mathbf{e}_2$ (representing the sample point $(+,-)$); λ_{22} is the conditional probability that the k-th unit is reported only in the session of 5 minutes (with result $(+,\cdot)$) given that $\mathbf{W}_k = \mathbf{e}_2$; λ_{32} is the probability that the k-th unit is reported only in the session of 7 minutes (with result $(\cdot,-)$) given that $\mathbf{W}_k = \mathbf{e}_2$; and λ_{42} is the probability that the k-unit is not reported in any of the sessions (with result $(\cdot,\cdot)$) given that $W_k = e_2$.

Note that, in the example, $\mathbf{R}_k$ and $\mathbf{W}_k$ have equal dimension, which does not hold in general. Also, this censoring pattern, common in medical cases, is very special and, as we will discuss later, simplifies considerably the analysis from an interpretational viewpoint.

To complete the notation, the vector of observation is denoted by $\mathbf{N} = (\mathbf{N_1}, \mathbf{N_2}, \mathbf{N_3}, n_4)$ where in a lexicografic ordering $\mathbf{N_1} = (n_{11}, n_{12}, n_{13}, n_{14})$ is for uncensored data, $\mathbf{N_2} = (n_{2+}, n_{2-})$ is for the data do not report the session of 7 minutes, $\mathbf{N_3} = (n_{3+}, N_{3-})$ is for the data that do not report the session of 5 minutes. and finally. n_4 is the number of units that do report neither of the two sessions. The censoring parameter vector is represented by

$$\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

and the parameters that we will use to describe the likelihood are represented by

$$\mu_{ij} = \theta_j \lambda_{ij} \ .$$

The probabilities of a unit being reported, respectively, in both sessions, only in the first, only in the second, and in none of them, are all obtained by making i equal to 1,2,3, and 4 in the following expression

$$\mu_{i1} + \mu_{i2} + \mu_{i3}\mu_{i4} = \mu_{i.} \ .$$

From the joint distribution of $\{(\mathbf{W}_k, \mathbf{R}_k) \ ; \ k = 1, \ldots, n\}$, the likelihood is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{N}) \ \propto$$

$$(\mu_{11})^{12}(\mu_{12})^{4}(\mu_{13})^{5}(\mu_{14})^{2}(\mu_{21}+\mu_{22})^{50}(\mu_{23}+\mu_{24})^{31}(\mu_{31}+\mu_{33})^{27}(\mu_{32}+\mu_{34})^{12}(\mu_{4})^{24}$$

This form of the likelihood reveals that not only the probabilistic model but also the parameters of interest, $\theta_j = \mu_{.j} = \mu_{1j} + \mu_{2j} + \mu_{3j} + \mu_{4j}$ $(j = 1, 2, 3, 4)$, are not identifiable.

The problem of non-identifiability - in general, a characteristic of incomplete categorical data - is in fact the true root of the inferential problems and justifies the procedures described in the non Bayesian literature.

The main purpose of this article is to develop a solution for the general problem of incomplete categorical data based on Dirichlet prior distributions and which is also tractable in several ways. In Section 2, for sake of simplicity, this solution is derived in general for censoring by partitions of the original category set. In Section 3, this solution is developed in detail for the example described above. The corresponding solution for a general censoring pattern is completely described by Paulino (1988) and follows the same line of thought.

2. BAYESIAN MODEL WITH INFORMATIVE REPORT PROCESS

From a population partioned in m categories a random sample of size n is to be selected. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ be a vector for which the element θ_i,

$i = 1, \ldots, m$, represents the positive probability that a sample unit belangs to the i-th category. Hence. θ assumes values on the $(m-1)$-dimensional simplex

$$\mathcal{S}_m = \{(s_1, \ldots, s_m) : s_i > 0 \ , \quad s_1 + \ldots + s_m = 1\} \ .$$

Let us represent by $\{e_1 : i = 1, \ldots, m\}$ the set of vectors of the canonical basis of $\mathbb{R}^m$ and define the random vectors $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n$ in such a way that $\mathbf{W}_k = \epsilon_i$ indicates that the k-th sample unit belongs to the i-th category. Hence. the vectors $\mathbf{W}_k$ are (conditionally) independent (given θ) and identically distributed as a Bernoulli multivariate probability ditribution with paramenter θ.

To obtain the "medical" censoring process. we consider M distinct partitions, $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_M$, of the set of categories. $\{1, \ldots, i, \ldots, m\}$, in such a way that there is no common element between any two of such partitions. The report process is then defined from these partitions. Each sample unit is reported as being an element of one (and only one) of these partitions. Hence, the report process is a process that. for a unit k, identifies a partition (the censoring type suffered by k) and its element that corresponds to the result effectively obtained by k. In the example of Section 1. if the partition associated to the k-th unit is $\{(+,\cdot),(-,\cdot)\} \equiv \{\{(+,+),(+,-)\},\{(-,+),(-,-)\}\}$, then this unit was not reported in the seven minutes session. Identifying the element of this partition corresponds to identifying the effective result obtained for the five minutes session. Note that the report process is necessarily coherent with the vector $\mathbf{W}_k$. For instance, in the example, the result $(+,-)$ could never produce a report $(\cdot,+)$. No attention is devoted here to the missclassification problem but only to the problem of missing data. However, it would not be difficult to adjust the present case to a more general and larger model that would cover the problem of missclassification. Also. the restriction on censoring defined by partitions can be relaxed but would bring some difficulties to our development of the solution. The derivation presented below pinpoints the main idea of the general argument which can be used for an arbitrary reporting pattern as shown by Paulino (1988). Here we do not have to use explicitely generalized Dirichlet distributions.

To build the report process. associate to each selected unit k the vector $\mathbf{R}_k$ taking values on $\{\epsilon_i \ ; \quad i = 1, \ldots, M\}$, the canonical basis of $\mathbb{R}^M$. Hence.

$\{\mathbf{R}_k = \epsilon_i\}$ means that the kind of censoring suffered by the k-th unit is defined by partition $\mathcal{P}_i$. For partition $\mathcal{P}_i\,(i = 1,\ldots,M)$, $C_{i1}, C_{i2}, \ldots, C_{im_i}$ denote its m_i elements (each such element is a set of categories). If $\mathcal{P}_1$ is the partition representing the absence of censoring, then $m_1 = m$. If $\mathcal{P}_M$ is the partition representing the absence of report (total censoring), then $m_M = 1$. In the sample observation process, besides the censoring kind defined by $\mathcal{P}_i$, an element C_{ij} of $\mathcal{P}_i$ is also observed. Returning to the example, from the 39 children reported only in the seven minutes session, 27 reported positively and 12 negatively. The reported censoring for these children is defined by the partition $\{\{(+,+),(-,+)\},\{(+,-),(-,-)\}\}$ and the reported observations $(\cdot,+)$ and $(\cdot,-)$ correspond to the two subsets that form the partition.

The vector of observations is denoted by $\mathbf{N} = (\mathbf{N}_1,\ldots,\mathbf{N}_i,\ldots,\mathbf{N}_M)$ where, in a lexicografic ordering, $\mathbf{N}_i = (n_{i1},\ldots,n_{im_i})$ for every $i = 1,2,\ldots,M$. In situations where there exists units with complete censoring, the corresponding vector, $\mathbf{N}_M$, of these units has only one element denoted by n_M. On the other hand, in situations where there exist units with absence of censoring, the corresponding vector, $\mathbf{N}_1$, of such units has m elements.

Given that $\mathbf{W}_k = \mathbf{e}_j$, to say that $\mathbf{R}_k = \epsilon_i$, is equivalent to saying that the k-th unit is classified in the only element of $\mathcal{P}_i$ that contains category j. To define the probabilistic model we consider that $\{(\mathbf{W}_k,\mathbf{R}_k)\}_{1\le k\le n}$ is a sequence of independent and identically distributed vectors. Also, for every $j = 1,\ldots,m$, the conditional distribution of $\mathbf{R}_k$ given $\mathbf{W}_k = \mathbf{e}_j$ is multivariate Bernoulli with parameter

$$\lambda_j = (\lambda_{1j},\ldots,\lambda_{Mj}) , \quad \text{where, for } i = 1,\ldots,M ,$$
$$\lambda_{ij} = \Pr\{\mathbf{R}_k = \epsilon_i \mid W_k = \mathbf{e}_j\} .$$

Using the parametrization

$$\mu_{ij} = \theta_i\lambda_{ij} = \Pr\{\mathbf{R}_k = \epsilon_i\ ,\ \mathbf{W}_k = \mathbf{e}_j\} ,$$

for the joint probabilities of $(\mathbf{W}_k, \mathbf{R}_k)$, the likelihood can be written as

$$f(\mathbf{N}\,|\,n,\mu) = (n!)\prod_{i=1}^{M}\prod_{l=1}^{m_i}\frac{1}{(n_{il})!}\left\{\sum_{j\in c_{il}}\mu_{ij}\right\}^{n_{il}} , \tag{2.1}$$

where μ represents the vector whose elements are the paramenters μ_{ij}, i.e., $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$ where, for $j = 1, \ldots, m$, $\mu_j = (\mu_{1j}, \ldots, \mu_{Mj})$.

Now we specify the prior distribution in terms of the vector μ of the Mm joint probabilities of $(\mathbf{R}_k, \mathbf{W}_k)$. For this, we adopt the Dirichlet distribution with parameter $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$, where, for $j = 1, \ldots, m$, $\mathbf{a}_j = (a_{j1}, \ldots, a_{Mj})$. This statement is denoted by $\mu|\mathbf{a} \sim D_{Mm}(\mathbf{a})$, where the scalar Mm in this expression indicates the dimension of $\mathbf{a}$. In order to avoid technical problems, here we consider that (for all possible i and j) the a_{ij}'s are positive real numbers, although in Section 3 we consider a case where some of the a_{ij}'s are zero.

Using the strong properties of the Dirichlet distribution we can write a convenient description of the prior in terms of the original parameters θ and λ_j, $j = 1, \ldots, m$. That is, to say that $\mu|\mathbf{a} \sim D_{Mm}(\mathbf{a})$ is equivalent to saying that

$$\begin{cases} \theta|\mathbf{a} \sim D_m(a_1, a_2, \ldots, a_m)\ ; \\ \lambda_j|\mathbf{a} \sim D_M(a_{1j}, a_{2j}, \ldots, a_{Mj})\ ; \quad \text{for} \quad j = 1; \ldots ; m \quad \text{and} \\ \theta \amalg \lambda_1 \amalg \cdots \amalg \lambda_m|\mathbf{a} \end{cases} \tag{2.2}$$

where $a_j = a_{1j} + a_{2j} + \ldots a_{Mj}$ and the last expression denotes the fact that, for each fixed a, the vectors $\theta, \lambda_1, \ldots, \lambda_m$ are mutually independent. In the remaining part of the paper, $\amalg$ will be used for independence. Note that, with this distribution, the variances of θ are necessarily smaller than the ones of λ, which is not unrealistic.

Expression (2.1) shows that the likelihood depends on μ only through the vector

$$\mu(\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_M) = (\mu(\mathcal{P}_1), \mu(\mathcal{P}_2), \ldots, \mu(\mathcal{P}_M))\ , \tag{2.3}$$

where

$$\mu(\mathcal{P}_i) = \left(\sum_{j \in C_{i1}} \mu_{ij}, \sum_{j \in C_{i2}} \mu_{ij}, \ldots \sum_{j \in C_{im_i}} \mu_{ij} \right)\ . \tag{2.4}$$

corresponding to a linear tranformation of μ. Again from properties of the Dirichlet distribution, we conclude that vector $\mu(\mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_M)$ is distributed as Dirichlet, both a priori and a posteriori. The prior and posterior parameters are, respectively, $(\mathbf{a}(\mathcal{P}_1), \mathbf{a}(\mathcal{P}_2), \ldots, \mathbf{a}(\mathcal{P}_M))$ and $(\mathbf{A}(\mathcal{P}_1), \mathbf{A}(\mathcal{P}_2), \ldots, \mathbf{A}(\mathcal{P}_M))$, where to obtain $\mathbf{a}(\mathcal{P}_i)$, it is enough to replace a_{ij} for μ_{ij} in the right side of equation (2.4) and $\mathbf{A}(\mathcal{P}_i) = \mathbf{a}(\mathcal{P}_i) + \mathbf{N}_i$, for every $i = 1, \ldots, M$.

The sample observations are fully used to update the parametric function $\mu(\mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_M)$, i.e., there is no further information in the sample to calibrate the parametric complement which, together with $\mu(\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_M)$, form a parametrization equivalent to μ. This fact is shown by the posterior distribuition of μ, which is a member of the family of generalized Dirichlet distributions (Dickey, 1983). This distributions can be viewed as a mixture of Dirichlet distributions obtained by considering all possible hypothetical frequencies of the missing portions of the data, the mixing distribution being the distribution of that set of frequencies given N. We notice in addition that the posterior moments in this case are not difficult to obtain (Paulino 1988).

It is possible to define a parametrization that produces a very nice form of the posterior (independent Dirichlet distributions) having the particularity of identifying the parameters that are updated and the ones that are not. However, we would need a heavier notation to define it. To avoid this we restrict ourselves to the example. In the next section, the Bayesian analysis of the example is described in detail, which allows us to foresee the way of defining that parametrization and computing the posterior moments of interest in more general cases.

3. USING AN EXAMPLE TO DEVELOP THE BAYESIAN SOLUTION

In this section we use the example described in Section 1 to develop the Bayesian solution outlined in Section 2. In this example we have

$$\mu(\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4) = (\mu(\mathcal{P}_1), \mu(\mathcal{P}_2), \mu(\mathcal{P}_3), \mu(\mathcal{P}_4)) ,$$

where

$$\begin{aligned} \mu(\mathcal{P}_1) &= (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}) , \quad \mu(\mathcal{P}_2) = (\mu_{21} + \mu_{22}, \mu_{23} + \mu_{24}) , \\ \mu(\mathcal{P}_3) &= (\mu_{31} + \mu_{33}, \mu_{32} + \mu_{34}) \text{ and } \mu(\mathcal{P}_4) = \mu_{41} + \mu_{42} + \mu_{43} + \mu_{44} = \mu_{4.} . \end{aligned}$$

Let's us also consider

$$\begin{aligned} &\mu_{11} + \mu_{12} + \mu_{13} + \mu_{14} = \mu_{1.} , \ \mu_{21} + \mu_{22} + \mu_{23} + \mu_{24} = \mu_{2.} \quad \text{and} \\ &\mu_{31} + \mu_{32} + \mu_{33} + \mu_{34} = \mu_{3.} , \end{aligned}$$

that represent the marginal probabilities or reporting in $\mathcal{P}_1, \mathcal{P}_2$ and $\mathcal{P}_3$, respectively.

Now let us define a new parametrization that brings a great deal of simplification to the analysis:

$$\mathbf{M} = (\mu_{1.}, \mu_{2.}, \mu_{3.}, \mu_{4.}) \tag{3.1}$$

$$\mathbf{\Pi_1} = (\Pi_{11}, \Pi_{12}, \Pi_{13}, \Pi_{14}) = \frac{1}{\mu_{1.}}(\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}) \ , \tag{3.2}$$

$$\mathbf{\Pi_2} = \frac{1}{\mu_{2.}}(\mu_{21} + \mu_{22}, \mu_{23} + \mu_{24}) = (\Pi_{2+}, \Pi_{2-}) \tag{3.3}$$

$$\mathbf{\Pi_3} = \frac{1}{\mu_{3.}}(\mu_{31} + \mu_{33}, \mu_{32} + \mu_{34}) = (\Pi_{3+}, \Pi_{3-}) \tag{3.4}$$

$$\mathbf{\Pi_4} = (\Pi_{41}, \Pi_{42}, \Pi_{43}, \Pi_{44}) = \frac{1}{\mu_{4.}}(\mu_{41}, \mu_{42}, \mu_{43}, \mu_{44}) \ , \tag{3.5}$$

$\boldsymbol{\pi} = (\pi_{21}, \pi_{22}, \pi_{23}, \pi_{24}, \pi_{31}, \pi_{33}, \pi_{32}, \pi_{34})$, where

$$\pi_{21} = \frac{\mu_{21}}{\mu_{21} + \mu_{22}} \quad , \quad \pi_{22} = 1 - \pi_{21} = \frac{\mu_{22}}{\mu_{21} + \mu_{22}} \ , \tag{3.6}$$

$$\pi_{23} = \frac{\mu_{23}}{\mu_{23} + \mu_{24}} \quad , \quad \pi_{24} = 1 - \pi_{23} = \frac{\mu_{24}}{\mu_{23} + \mu_{24}} \ , \tag{3.7}$$

$$\pi_{31} = \frac{\mu_{31}}{\mu_{31} + \mu_{33}} \quad , \quad \pi_{33} = 1 - \pi_{31} = \frac{\mu_{33}}{\mu_{31} + \mu_{33}} \ , \tag{3.8}$$

$$\pi_{32} = \frac{\mu_{32}}{\mu_{32} + \mu_{34}} \quad , \quad \pi_{34} = 1 - \pi_{32} = \frac{\mu_{34}}{\mu_{32} + \mu_{34}} \ , \tag{3.9}$$

The meaning of these parameters is apparent. For instance, $\mathbf{\Pi_1}$ ($\mathbf{\Pi_4}$) is the vector of conditional probabilities of each category given classification into $\mathcal{P}_1$ ($\mathcal{P}_4$); $\mathbf{\Pi_2}$ ($\mathbf{\Pi_3}$) is the vector of conditional probabilities of each element of $\mathcal{P}_2$ ($\mathcal{P}_3$) given classification into $\mathcal{P}_2$ ($\mathcal{P}_3$); π_{21} (π_{31}) is the conditional probability of the first category given classification in the first element of $\mathcal{P}_2$ ($\mathcal{P}_3$). Recall that

$$\begin{aligned}
\mathcal{P}_1 &= \{(+,+),(+,-),(-,+),(-,-)\} \ , \\
\mathcal{P}_4 &= \{\{(+,+),(+,-),(-,+),(-,-)\}\} \ , \\
\mathcal{P}_2 &= \{\{(+,+),(+,-)\},\{(-,+),(-,-)\}\} \ , \text{and} \\
\mathcal{P}_3 &= \{\{(+,+),(-,+)\},\{(+,-),(-,-)\}\} \ .
\end{aligned}$$

Note that the parameter

$$\mathbf{\Pi} = (\mathbf{M}, \mathbf{\Pi_1}, \mathbf{\Pi_2}, \mathbf{\Pi_3}, \mathbf{\Pi_4}, \boldsymbol{\pi}) \tag{3.10}$$

is an one-to-one transformation of the original parameter $\boldsymbol{\mu}$ and is a 16-dimensional vector composed by the independent vectors $\mathbf{M}, \mathbf{\Pi_1}, \mathbf{\Pi_2}, \mathbf{\Pi_3}, \mathbf{\Pi_4}$ and $\boldsymbol{\pi}$. Writing the likelihood in terms of this new parameter yields the following expression:

$$L(\mathbf{\Pi}|\mathbf{N}) \propto$$
$$(\Pi_{11})^{12}(\Pi_{12})^{4}(\Pi_{13})^{5}(\Pi_{14})^{2}(\Pi_{2+})^{50}(\Pi_{2-})^{31}(\Pi_{3+})^{27}(\Pi_{3-})^{12}(\mu_{1.})^{23}(\mu_{2.})^{81}(\mu_{3.})^{39}(\mu_{4.})^{24} \tag{3.11}$$

Note that $\Pi_{2+} = 1 - \Pi_{2-}$ and $\Pi_{3+} = 1 - \Pi_{3-}$ and the observed frequency vector is $\mathbf{N} = (\mathbf{N_1}, \mathbf{N_2}, \mathbf{N_3}, n_4) = (n_{11}, n_{12}, n_{13}, n_{14}, n_{2+}, n_{2-}, n_{3+}, n_{3-}, n_4) = (12, 4, 5, 2, 50, 31, 27, 12, 24)$. Clearly the lidelihood depends neither on $\boldsymbol{\pi}$ nor on $\mathbf{\Pi_4}$. Another interesting aspect of this representation is the fact that the likehood can be factored out as a product of four functions. The first depends only on $\mathbf{\Pi_1}$, the second only on $\mathbf{\Pi_2}$, the third only on $\mathbf{\Pi_3}$, and finally, the fourth only on $\mathbf{M}$.

According to Section 2, we have a priori $\mu|\mathbf{a} \sim D_{16}(\mathbf{a})$ where $\mathbf{a}$ is a vector which elements, represented by a_{ij}, $i, j = 1, \ldots, 4$, are positive real numbers (there are cases where, in a more general singular representation, some of these elements can be taken as null). As a result, taking $a_{2+} = a_{21} + a_{22}$, $a_{2-} = a_{23} + a_{24}$, $a_{3+} = a_{31} + a_{33}$ and $a_{3-} = a_{32} + a_{34}$, the distributions of the new parameters are as follows:

$$\mathbf{M}|\mathbf{a} \sim (\mu_{1.}, \mu_{2.}, \mu_{3.}, \mu_{4.})|\mathbf{a} \sim D_4(a_{1.}, a_{2.}, a_{3.}, a_{4.}) \ ,$$

$$\mathbf{\Pi_1}|\mathbf{a} \sim D_4(a_{11}, a_{12}, a_{13}, a_{14}) \ ,$$

$$\mathbf{\Pi_2}|\mathbf{a} \sim D_2(a_{21} + a_{22}, a_{23} + a_{24}) \quad \text{or} \quad \Pi_{2+}|\mathbf{a} \sim B(a_{2+}, a_{2-}) \ ,$$

$$\mathbf{\Pi_3}|\mathbf{a} \sim D_2(a_{31} + a_{33}, a_{32} + a_{34}) \quad \text{or} \quad \Pi_{3+}|\mathbf{a} \sim B(a_{3+}, a_{3-}) \ ,$$

$$\mathbf{\Pi_4}|\mathbf{a} \sim D_4(a_{41}, a_{42}, a_{43}, a_{44}) \ ,$$

$$(\pi_{21}, \pi_{22})|\mathbf{a} \sim D_2(a_{21}, a_{22}) \quad \text{or} \quad \pi_{21}|\mathbf{a} \sim B(a_{21}, a_{22}) \ ,$$

$$(\pi_{23}, \pi_{24})|\mathbf{a} \sim D_2(a_{23}, a_{24}) \quad \text{or} \quad \pi_{23}|\mathbf{a} \sim B(a_{23}, a_{24}) \ ,$$

$$(\pi_{31},\pi_{33})|\mathrm{a} \sim D_2(a_{31},a_{33}) \quad \text{or} \quad \pi_{31}|\mathrm{a} \sim B(a_{31},a_{33}) \ ,$$

$$(\pi_{32},\pi_{34})|\mathrm{a} \sim D_2(a_{32},a_{34}) \quad \text{or} \quad \pi_{32}|\mathrm{a} \sim B(a_{32},a_{34}) \ , \text{ and}$$

$$\mathbf{M} \amalg \mathbf{\Pi}_1 \amalg \mathbf{\Pi}_2 \amalg \mathbf{\Pi}_3 \amalg \mathbf{\Pi}_4 \amalg \pi_{21} \amalg \pi_{23} \amalg \pi_{31} \amalg \pi_{32}|\mathrm{a} \ . \tag{3.12}$$

These distributions are obtained by using the mentioned properties of the Dirichlet distribution.

As the likelihood is adequately factorized, the independence structure indicated above is invariant under Bayesian operations. On the other hand, given the noninformativeness of the likelihood about $\mathbf{\Pi}_4$ and $\boldsymbol{\pi}$, the distributions of these paramenters are also immutable under such operations. The full posterior description is as follows:

$$\mathbf{M}|(\mathrm{a},\mathrm{N}) \sim D_4(a_{1.}+n_{1.},a_{2.}+n_{2.},a_{3.}+n_{3.},a_{4.}+n_{4.})$$

$$\mathbf{\Pi}_1|(\mathrm{a},\mathrm{N}) \sim D_4(a_{11}+n_{11},a_{12}+n_{12},a_{13}+n_{13},a_{14}+n_{14}) \ ,$$

$$\Pi_{2+}|(\mathrm{a},\mathrm{N}) \sim B(a_{2+}+n_{2+},a_{2-}+n_{2-}) \ ,$$

$$\Pi_{3+}|(\mathrm{a},\mathrm{N}) \sim B(a_{3+}+n_{3+},a_{3-}+n_{3-}) \ ,$$

$$\mathbf{\Pi}_4|(\mathrm{a},\mathrm{N}) \sim \mathbf{\Pi}_4|\mathrm{a} \sim D_4(a_{41},a_{42},a_{43},a_{44}) \ ,$$

$$\pi_{21}|(\mathrm{a},\mathrm{N}) \sim \pi_{21}|\mathrm{a} \sim B(a_{21}+a_{22}) \ ,$$

$$(\pi_{23}|(\mathrm{a},\mathrm{N}) \sim \pi_{23}|\mathrm{a} \sim B(a_{23},a_{24}) \ ,$$

$$\pi_{31}|(\mathrm{a},\mathrm{N}) \sim \pi_{31}|\mathrm{a} \sim B(a_{31},a_{33}) \ ,$$

$$\pi_{32}|(\mathrm{a},\mathrm{N}) \sim \pi_{32}|\mathrm{a} \sim B(a_{32},a_{34}) \ ,$$

$$\mathbf{M} \amalg \mathbf{\Pi}_1 \amalg \Pi_{2+} \amalg \Pi_{3+} \amalg \mathbf{\Pi}_4 \amalg \pi_{21} \amalg \pi_{23} \amalg \pi_{31} \amalg \pi_{32}|(\mathrm{a},\mathrm{N}) \ . \tag{3.13}$$

Recall that $n_{1.} = n_{11}+n_{12}+n_{13}+n_{14} = 23$, $n_{2.} = n_{2+}+n_{2-} = 81$, $n_{3.} = n_{3+}+n_{3-} = 39$ and $n_{4.} = n_4 = 24$.

The above parametrization clearly illustrates the parametric part that is updated by the sample and the one that is not. Furthermore, it sugests more general solutions than those generated by Dirichlet priors. For example, with respect to (3.12), the parameters of the prior distributions of $\mathbf{\Pi}_2$, $\mathbf{\Pi}_3$ and $\mathbf{M}$ do not have necessarily to reflect the linear relations required by the

Dirichlet adopted for $\boldsymbol{\mu}$. The use of (3.12) with such parameters replaced by other adequate positive real numbers implies that the distribution of $\boldsymbol{\mu}$ will no longer be Dirichlet but a mixture of Dirichlet distributions. Thus, in the place of $a_{21}+a_{22}$, $a_{23}+a_{24}$, $a_{31}+a_{33}$, $a_{32}+a_{34}$, $a_{1.}$, $a_{2.}$, $a_{3.}$ and $a_{4.}$, we could have any other positive real numbers without destroying the prior-to-posterior structure [the passage of (3.12) to (3.13)]. This shows the general aspect of the solution presented here. In many practical examples, the vector $\mathbf{M}$ is considered to have a degenerate distribution (as in case of deterministic report process that include the complete data standard situation), which can be handled within the setting described.

The parameters of interest are defined by $\theta_j = \mu_{.j} = \mu_{1j} + \mu_{2j} + \mu_{3j} + \mu_{4j}$, $j = 1, \ldots, 4$. which in terms of the new parametrization are written as follows:

$$
\begin{aligned}
\theta_1 &= (\Pi_{11})(\mu_{1.}) + (\pi_{21})(\Pi_{2+})(\mu_{2.}) + (\pi_{31})(\Pi_{3+})(\mu_{3.}) + (\Pi_{41})(\mu_{4.}) \\
\theta_2 &= (\Pi_{12})(\mu_{1.}) + (\pi_{22})(\Pi_{2+})(\mu_{2.}) + (\pi_{32})(\Pi_{3-})(\mu_{3.}) + (\Pi_{42})(\mu_{4.}) \\
\theta_3 &= (\Pi_{13})(\mu_{1.}) + (\pi_{23})(\Pi_{2-})(\mu_{2.}) + (\pi_{33})(\Pi_{3+})(\mu_{3.}) + (\Pi_{43})(\mu_{4.}) \\
\theta_4 &= (\Pi_{14})(\mu_{1.}) + (\pi_{24})(\Pi_{2-})(\mu_{2.}) + (\pi_{34})(\Pi_{3-})(\mu_{3.}) + (\Pi_{44})(\mu_{4.})
\end{aligned}
\tag{3.14}
$$

Since each element of the sums of the right-hand side of the equalities (3.14) is a product of independent random variables, the posterior mean of $\boldsymbol{\theta}$ is easily obtained. We only derive the posterior mean of θ_1 and by analogy, we present the respective means of θ_2, θ_3 and θ_4.

$$
\begin{aligned}
E\{\theta_1|\mathbf{a},\mathrm{N}\} &= \left(\frac{a_{11}+n_{11}}{a_{1.}+n_{1.}}\right)\left(\frac{a_{1.}+n_{1.}}{a_{..}+n_{..}}\right) + \frac{a_{21}}{a_{2+}}\left(\frac{a_{2+}+n_{2+}}{a_{2.}+n_{2.}}\right)\left(\frac{a_{2.}+n_{2.}}{a_{..}+n_{..}}\right) \\
&\quad + \frac{a_{31}}{a_{3+}}\left(\frac{a_{3+}+n_{3+}}{a_{3.}+n_{3.}}\right)\left(\frac{a_{3.}+n_{3.}}{a_{..}+n_{..}}\right) + \frac{a_{41}}{a_{4.}}\left(\frac{a_{4.}+n_{4.}}{a_{..}+n_{..}}\right) \\
&= \frac{1}{a_{..}+n_{..}}\Big\{(a_{11}+n_{11}) + \frac{a_{21}}{a_{2+}}(a_{2+}+n_{2+}) \\
&\qquad + \frac{a_{31}}{a_{3+}}(a_{3+}+n_{3+}) + \frac{a_{41}}{a_{4.}}(a_{4.}+n_{4.})\Big\} \, .
\end{aligned}
$$

An interesting interpretation of this result is the following: the term inside brackets is obtained by adding to the prior parameter the observed frequency of the first category plus a franction of each of the frequencies pertaining to the sets of confounded categories that include the first one; the fractions are defined by the expected values of π_{21}, π_{31} and Π_{41}, the category conditional probabilities given the reports. Equivalently, the posterior mean is a weighted mean of the prior mean and the sum of the sampling proportions allocated to the respective category; each term of this sum defines the conditional mean of the corresponding unobserved frequencies given the data.

If for the posterior parameters, distinct from the corresponding prior ones, we write A for a, the following expressions follow:

$$E\{\theta_1|\mathrm{a},\mathrm{N}\} = \frac{1}{A_{..}}\left\{(A_{11}) + \frac{a_{21}}{a_{2+}}(A_{2+}) + \frac{a_{31}}{a_{3+}}(A_{3+}) + \frac{a_{41}}{a_{4.}}(A_{4.})\right\} .$$

$$E\{\theta_2|\mathrm{a},\mathrm{N}\} = \frac{1}{A_{..}}\left\{(A_{12}) + \frac{a_{22}}{a_{2+}}(A_{2+}) + \frac{a_{32}}{a_{3-}}(A_{3-}) + \frac{a_{42}}{a_{4.}}(A_{4.})\right\} .$$

$$E\{\theta_3|\mathrm{a},\mathrm{N}\} = \frac{1}{A_{..}}\left\{(A_{13}) + \frac{a_{23}}{a_{2-}}(A_{2-}) + \frac{a_{33}}{a_{3+}}(A_{3+}) + \frac{a_{43}}{a_{4.}}(A_{4.})\right\} .$$

$$E\{\theta_4|\mathrm{a},\mathrm{N}\} = \frac{1}{A_{..}}\left\{(A_{14}) + \frac{a_{24}}{a_{2-}}(A_{2-}) + \frac{a_{34}}{a_{3-}}(A_{3-}) + \frac{a_{44}}{a_{4.}}(A_{4.})\right\} . \quad (3.15)$$

A thorough Bayesian study would allow for the influence of the data on a range of distributions expressing various expert opinions. To illustrate aspects of such study, we will consider 3 prior distributions. The first is a uniform distribution for $\boldsymbol{\mu}$, which corresponds to take $a_{ij} = 1$ for all $i,j = 1,2,3,4$. This is equivalent to consider

$$\begin{cases} \boldsymbol{\theta}|\mathrm{a} \sim D_4(4;4;4;4) \ ; \\ \boldsymbol{\lambda}_j|\mathrm{a} \sim D_4(1;1;1;1) \ ; \quad \text{for all} \quad j = 1;2;3;4 \quad \text{and} \\ \boldsymbol{\theta} \amalg \boldsymbol{\lambda}_1 \amalg \boldsymbol{\lambda}_2 \amalg \boldsymbol{\lambda}_3 \amalg \boldsymbol{\lambda}_4|\mathrm{a} \end{cases} \quad (3.16)$$

The second prior distribution is intended to stand for the opinion of a conceptual expert familiar with medical studies of this kind. It is defined in terms of (2.2) as

$$\begin{cases} \boldsymbol{\theta}|\mathrm{a} \sim D_4(10;5;5;10) \ ; \\ \boldsymbol{\lambda}_1|\mathrm{a} \sim D_4(1;3;2;4) \\ \boldsymbol{\lambda}_2|\mathrm{a} \sim D_4(1;0.5;2;1.5) \\ \boldsymbol{\lambda}_3|\mathrm{a} \sim D_4(1.5;2;0.5;1) \\ \boldsymbol{\lambda}_4|\mathrm{a} \sim D_4(4;2;3;1) \ ; \quad \text{and} \\ \boldsymbol{\theta} \amalg \boldsymbol{\lambda}_1 \amalg \boldsymbol{\lambda}_2 \amalg \boldsymbol{\lambda}_3 \amalg \boldsymbol{\lambda}_4|\mathrm{a} \end{cases} \quad (3.17)$$

which corresponds to consider for $\boldsymbol{\mu}$ a Dirichlet with parameter $(1,3,2,4,1,0.5,2,1.5,1.5,2,0.5,1,4,2,3,1)$. The third distribution considered here consists of a uniform distribution for $\boldsymbol{\theta}$ and a distribution for $\boldsymbol{\lambda}_j$, that

TABLE II
Posterior means and (posterior standard deviations)

Par. \ Prior	1	2	3
θ_1	.331(.100)	.504(.074)	.482(.094)
θ_2	.246(.093)	.131(.044)	.185(.073)
θ_3	.240(.080)	.173(.064)	.222(.079)
δ	.006(.126)	-.042(.084)	-.037(.120)

is, for all j, degenerate at point $(1,0,0,0)$ (this is equivalent to an analysis conditioned on the fully categorized subsample): i.e.,

$$\begin{cases} \boldsymbol{\theta}|\mathbf{a} \sim D_4(1;1;1;1) \; ; \\ \boldsymbol{\lambda}_j|\mathbf{a} \sim D_4(1;0;0;0) \; ; \quad \text{for all} \quad j = 1;2;3;4 \end{cases} \tag{3.18}$$

Table II displays the mean and the standard deviation of elements of $\boldsymbol{\theta}$ and the parametric function $\delta = \theta_2 - \theta_3$. To compute the standard deviations we have used the expression

$$V\{\theta_i|\mathbf{a},\mathrm{N}\} = EV\left\{\{\theta_i|\mathbf{a},\mathrm{N},\mathrm{M}\}|\mathbf{a},\mathrm{N}\}\right\} + VE\left\{\{\theta_i|\mathbf{a},\mathrm{N},\mathrm{M}\}|\mathbf{a},\mathrm{N}\right\} \; . \tag{3.19}$$

The results in Table 2 shows that posterior means and standard deviations are sensitive to choice of the prior.

Although tedious these calculations are mere algebraic exercises. Note that the elements of the sum that defines θ_i are conditionally independent given $\mathbf{M}$, which simplifies significantly the calculations. The interest on δ here is purely illustrative. Nevertheless, we could imagine that δ is a variable indicating the influence of the exercise time on the presence of bronchial spasm.

The magnitude of the values of $E(\delta|\mathbf{a},\mathrm{N})$ and $V(\delta|\mathbf{a},\mathrm{N})$ for any of the priors used suggest the equality of θ_2 and θ_3, even without further considerations about the exact or approximate distribution of δ. We must refer to the fact that this analysis yields absolute values for the posterior means and standard deviations different than the ones obtained from the solution based on a non-informative report mechanism of Dickey et al. (1987).

4. FINAL REMARKS

The example used in this paper reveals a simple widely applicable report pattern. Following the same arguments, solutions for more complex patterns can be obtained as described by Paulino (1988).

The Bayesian solution developed here leant on the prior distribution (2.2), a Dirichlet for μ, producing a posteriori a generalized Dirichlet distribution (Dickey, 1983, Dickey et al., 1987 and Paulino, 1988). This distribution has a simple representation described by (3.13). This representation enlightens the independence structure implied by the Bayesian model adopted.

The linear relations among the prior parameters in expression (2.2) can be avoided in such a way that the distribution of μ is generalized Dirichlet. Using results of Dickey (1983) and Paulino (1988), it is not difficult to verify that the posterior is also generalized Dirichlet. The computation of the posterior moments, though made more difficult, is possible by means of numerical methods.

Another main restriction of the prior used is its independence structure. Paulino (1988) analyses special cases of dependence and obtains solutions that are analytically difficult to be handled, as expected. The computation of the posterior moments of $\boldsymbol{\theta}$, for large samples, can eventually be simplified by using approximate methods as the ones suggested by Kadane (1985) and Dickey et al. (1987). However, in some cases, the option for the posterior mode as Bayes estimator greatly simplifies the analyses due to the applicability of the EM algorithm (Paulino, 1988).

ACKNOWLEDGEMENTS

When preparing this work the authors have received the partial support of CNPq and IBM Brazil through the Rio Scientific Center.

BIBLIOGRAPHY

Basu, D. & Pereira, C.A. de B. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse. *Journal of Statistical Planning and Inference 6*. 345-62

Dickey, J.M. (1983). Multiple hipergeometric functions: probabilistic interpretations and statistical uses. *Journal of the American Statistical Association 78*. 628-37.

Dickey, J.M., Jiang, J.M., & Kadane, J.B. (1987). Bayesian methods for censored categorical data. *Journal of the American statistical Association 82*. 773-81.

Kadane, J.B. (1985). Is victimization chronic? A Bayesian analysis of multinomial missing data. *Journal of Econometrics 29*. 47-67.

Little, R.J.A. & Rubin, D.B. (1987). *Statistical Analysis with missing data.* John Wiley & Sons. New York.

Paulino, C.D.M. (1988). *Analysis of incompete categorical data: foundations, methods, and applications* (in Portuguese). Doctoral Thesis. Instituto de Matemática e Estatística. Universidade de São Paulo.

Paulino, C.D.M. (1991). Analysis of categorical data with full and partial classification: a survey of the conditional maximum likelihood and weighted least squares approaches. *The Brazilian Journal of Probability and Statistics 5(1)*. 1-42.

Received November 1989; Revised February 1992