Journal of Statistical Planning and Inference 6 (1982) 345–362
North-Holland Publishing Company

ON THE BAYESIAN ANALYSIS OF CATEGORICAL DATA: THE PROBLEM OF NONRESPONSE

D. BASU*

The Florida State University, Talahassee, FL 32306, USA

Carlos A. de B. PEREIRA**

Universidade de São Paulo, Brazil

Received 18 September 1981; revised manuscript received 28 December 1981
Recommended by S. Zacks

Abstract: It is demonstrated how a suitably chosen prior for the frequency parameters can streamline the Bayesian analysis of categorical data with missing entries due to nonresponse or other causes. The two cases where the data follow the Multinomial or the Hypergeometric model are treated separately. In the first case it is adequate to restrict the prior (for the cell probabilities) to the class of Dirichlet distributions. In the case of the Hypergeometric model it is convenient to select a prior from the class of Dirichlet-Multinomial (DM) distributions. The DM distributions are studied in some details.

Key words: Nonidentifiability, Nuisance parameters, Dirichlet-Multinomial distributions, Conjugate family of distributions.

1. Introduction

The simplest case of the problem of nonresponse is as follows. Let Π_1 be the unknown proportion of individuals in a certain population, $\mathscr{P}$, that belong to a particular category A_1. With Π_1 as the only parameter of interest, a survey is conducted using a simple random sample of size n. Of the n individuals surveyed, n_1 respond to the question "Do you belong to category A_1?" with a yes/no answer but $n_2 = n - n_1$ individuals do not respond. Denoting the category of respondents by R, and the complementary category by R', the survey data may be summarized as

	R	R'	
A_1	x_1	n_2	
A_2	x_2		
	n_1	n_2	n

(1.1)

with A_2 being the complement of A_1.

* Research partially supported by NSF Grant No. 79-04693.
** Research partially supported by CAPES e CNPq.

0378-3758/82/0000–0000/$02.75 © 1982 North-Holland

In many practical problems, it is understood that the nonresponse of an individual is highly dependent on the value of the measurement under study. For example, suppose that one is surveying a population of students in order to estimate the proportion of cannabis smokers. In this case, it should be expected that a student who smokes has a higher chance of being a nonrespondent than one who does not. In this instance, at least, a nonresponse is a source of information.

The above understanding of the problem suggests that the population must also be partitioned into the categories R and R'; that is, the class of elements which would respond to the question, if selected, and its complement. The population proportions may be displayed in a 2×2-tabulator form as

	R	R'	
A_1	p_{11}	p_{12}	Π_1
A_2	p_{21}	p_{22}	Π_2
	q	$1-q$	1

$$\Pi_2 = 1 - \Pi_1 \qquad q = p_{11} + p_{21} \tag{1.2}$$

How can the data (1.1) be analysed vis-à-vis the parameter of interest $\Pi_1 = p_{11} + p_{12}$?

If the population of size N is regarded to be infinitely large compared to the sample size n; that is, if a multinomial model for the data is adopted, then the likelihood function is

$$L = p_{11}^{x_1} p_{21}^{x_2} (1-q)^{n_2}. \tag{1.3}$$

We represent the data by $X = (x_1, x_2, n_2)$ with $n_2 = n - (x_1 + x_2)$.

Since p_{12} cannot be defined in terms of the sampling distribution of X, an orthodox non-Bayesian would characterize $\Pi_1 = p_{11} + p_{12}$ as nonidentifiable, and would have little else to say on the matter. None of the many non-Bayesian methods of nuisance parameter elimination listed in Basu (1977) apply to the present case. On the other hand, a Bayesian regards a parameter as an unknown entity that exists in its own right. It enters into the sampling distribution of a properly planned experiment but is not defined by the experiment. Nonidentifiability is, therefore, a non-problem from the Bayesian viewpoint.

With a suitable representation ξ of his/her opinion about $\boldsymbol{p} = (p_{11}, p_{21}, p_{12}, p_{22})$, the Bayesian will proceed to derive the posterior distribution by matching ξ with the likelihood function (1.3). The posterior marginal distribution of the parameter of interest Π_1 will be obtained by integration.

In Section 2 we demonstrate how the choice of a Dirichlet prior for $\boldsymbol{p}$ simplifies the Bayesian operation. The more general case where the respondents are classified into k (instead of 2) categories, $A_1, \ldots, A_k$, is analyzed in a similar fashion. Since the inference is based on the data, it is of interest to study the distribution of the data under the particular prior. Section 3 introduces the Dirichlet–Multinomial distribution and some of its properties. This distribution, besides being the marginal distribution of the data, plays an important role in the rest of the paper.

Sections 4 and 5 deal with the case of sampling from a finite population; that is, the case where the statistical model is Hypergeometric or, more generally, Multivariate Hypergeometric. For the case where $k=2$, instead of p_{11}, p_{21}, p_{12}, and p_{22}, the unknown frequency counts θ_{11}, θ_{21}, θ_{12}, and θ_{22} must be considered. As in (1.2), the population parameters may be displayed as

	R	R'	
A_1	θ_{11}	θ_{12}	θ_1
A_2	θ_{21}	θ_{22}	θ_2
	ψ	$N-\psi$	N

(1.4)

with $\theta_2=N-\theta_1$, and the parameter of interest being $\theta_1=\theta_{11}+\theta_{12}$. A Dirichlet-Multinomial prior for $\boldsymbol{\theta}=(\theta_{11},\theta_{21},\theta_{12},\theta_{22})$ greatly simplifies the analysis of the data (1.1) vis-à-vis the parameter of interest θ_1.

Notation. Let x, y and z be either random variables or random vectors. When x, y, and z are mutually independent we write $x \amalg y \amalg z$. By $x \amalg y \mid z$ it is meant that x and y are conditionally independent given z. If x and y have the same distribution we write $x \sim y$.

Let $\boldsymbol{p}=(p_1,\dots,p_k)$ be a k-dimensional positive random vector such that $\sum_{i=1}^k p_i=1$. We write $\boldsymbol{p}\sim D(\alpha_1,\dots,\alpha_k)$ to indicate that the distribution of $\boldsymbol{p}$ is a Dirichlet with nonnegative real parameters $\alpha_1,\alpha_2,\dots,\alpha_k$. For $k=2$, instead of $(p_1,p_2)\sim D(\alpha_1,\alpha_2)$, we use the conventional Beta distribution notation, $p_1\sim B(\alpha_1,\alpha_2)$.

Let $\boldsymbol{x}=(x_1,\dots,x_k)$ be a k-dimensional nonnegative integer random vector with fixed $n=\sum_{i=1}^k x_i$. We write $\boldsymbol{x}\mid\boldsymbol{p}\sim \mathrm{M}(n;\boldsymbol{p})$, where $\boldsymbol{p}$ is defined as above, to indicate that the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{p}$ is Multinomial with parameters n and $\boldsymbol{p}$. For $k=2$, instead of $(x_1,x_2)\mid(p_1,p_2)\sim \mathrm{M}(n;(p_1,p_2))$, we use the conventional Binomial distribution notation, $x_1\mid p_1\sim \mathrm{Bi}(n;p_1)$. When $\boldsymbol{\theta}=(\theta_1,\dots,\theta_k)$ is a nonnegative integer random vector with $\sum_{i=1}^k\theta_i=N$ fixed, we write $\boldsymbol{x}\mid\boldsymbol{\theta}\sim \mathrm{H}(N,n,\boldsymbol{\theta})$ to indicate that the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{\theta}$ is Multivariate Hypergeometric with parameter $(N,n,\boldsymbol{\theta})$. For $k=2$, instead of $(x_1,x_2)\mid(\theta_1,\theta_2)\sim \mathrm{H}(N,n,(\theta_1,\theta_2))$, we use the conventional notation for Hypergeometric distributions, $x_1\mid\theta_1\sim \mathrm{h}(N,n,\theta_1)$. The probability function corresponding to $\mathrm{H}(N,n,\boldsymbol{\theta})$ may be expressed in the following two ways:

$$f(\boldsymbol{x}\mid\theta)=\frac{\binom{\theta_1}{x_1}\binom{\theta_2}{x_2}\cdots\binom{\theta_k}{x_k}}{\binom{N}{n}}=\frac{\binom{n}{x_1,\dots,x_k}\binom{N-n}{\theta_1-x_1,\dots,\theta_k-x_k}}{\binom{N}{\theta_1,\dots,\theta_k}}.$$

2. Nonresponse: the Multinomial model

First we consider the case of $k=2$, where the data, the population parameters, and the likelihood are described by (1.1), (1.2) and (1.3) respectively.

In the full response model, it is well known that the family of Dirichlet distributions of the correct dimension is the natural conjugate family for the Bayesian analysis. That is, if y_1 and y_2 were the observations in R', and

$$\boldsymbol{p}=(p_{11},p_{21},p_{12},p_{22})\sim D(\alpha_{11},\alpha_{21},\alpha_{12},\alpha_{22}) \tag{2.1}$$

a priori, then the posterior distribution would be

$$D(\alpha_{11}+x_1,\alpha_{21}+x_2,\alpha_{12}+y_1,\alpha_{22}+y_2).$$

To introduce a Bayesian solution to the nonresponse case, it is useful to consider the following reparametrization:

$$q=p_{11}+p_{21},\qquad q_{11}=\frac{p_{11}}{q},\qquad q_{12}=\frac{p_{12}}{1-q} \tag{2.2}$$

with the reverse transformation being

$$\begin{aligned} p_{11}&=qq_{11}, & p_{12}&=(1-q)q_{12},\\ p_{21}&=q(1-q_{11}), & p_{22}&=(1-q)(1-q_{12}). \end{aligned} \tag{2.3}$$

The following general result for Dirichlet distributions is a key to the solution. Let $m\in\{2,\dots,k-1\}$ be fixed.

Lemma 1. *The following set of conditions is necessary and sufficient to have* $(p_1,\dots,p_k)\sim D(\alpha_1;\dots,\alpha_k)$:

$$y=\sum_{i=1}^{m}p_i\sim B\left(\sum_{i=1}^{m}\alpha_i,\ \sum_{i=m+1}^{k}\alpha_i\right), \tag{i}$$

$$\frac{1}{y}(p_1,\dots,p_m)\sim D(\alpha_1,\dots,\alpha_m), \tag{ii}$$

$$\frac{1}{1-y}(p_{m+1},\dots,p_k)\sim D(\alpha_{m+1},\dots,\alpha_k),$$

and

$$y\perp\!\!\!\perp\frac{1}{y}(p_1,\dots,p_m)\perp\!\!\!\perp\frac{1}{1-y}(p_{m+1},\dots,p_k). \tag{iii}$$

The proof of this result is straightforward and therefore is omitted.

Suppose that, a priori, (2.1) is considered. By Lemma 1, this is equivalent to

$$\begin{aligned} &q\sim B(\alpha_{.1},\alpha_{.2}), && q_{11}\sim B(\alpha_{11},\alpha_{21}),\\ &q_{12}\sim B(\alpha_{12},\alpha_{22}), && q\perp\!\!\!\perp q_{11}\perp\!\!\!\perp q_{12} \end{aligned} \tag{2.4}$$

where $\alpha_{.j} = \alpha_{1j} + \alpha_{2j}$ $(j = 1, 2)$.

The reparametrization (2.2) changes the likelihood (1.3) to

$$L = q^{n_1}(1-q)^{n_2} q_{11}^{x_1}(1-q_{11})^{x_2}. \tag{2.5}$$

By matching the prior (2.4) with (2.5), we derive the posterior distribution of (q, q_{11}, q_{12}):

$$q \amalg q_{11} \amalg q_{12} \mid X, \tag{2.6)(i}$$

$$q_{11} \mid X \sim B(\alpha_{11} + x_1, \alpha_{21} + x_2), \qquad q_{12} \mid X \sim q_{12} \sim B(\alpha_{12}, \alpha_{22}), \tag{2.6)(ii}$$

and

$$q \mid X \sim q \mid n_1 \sim B(\alpha_{.1} + n_1, \alpha_{.2} + n_2). \tag{2.6)(iii}$$

As expected, n_1 is sufficient to predict q, and q_{12} is independent of the data. Since $\alpha_{.2} = \alpha_{12} + \alpha_{22} \leq \alpha_{.2} + n_2$, the posterior distribution of the original parameter $\boldsymbol{p}$ is again Dirichlet if and only if $n_2 = 0$. It is, however, a mixture of Dirichlet distributions, and

$$(p_{11}, p_{21}, (1-q)) \mid X \sim D(\alpha_{11} + x_1, \alpha_{21} + x_2, \alpha_{.2} + n_2).$$

Note that these properties of the posterior allow one to define a 'nice' conjugate family of distributions for the nonresponse case. That is, the prior given by (2.4) would be conjugate if, instead of $q \sim B(\alpha_{.1}, \alpha_{.2})$, we had $q \sim B(\alpha_{.1}, \beta)$, where $\beta \geq \alpha_{.2}$.

To proceed with the estimation of Π_1, the parameter of interest, we recall (2.3) to write $\Pi_1 = q q_{11} + (1-q) q_{12}$, and consider $\alpha = \alpha_{11} + \alpha_{21} + \alpha_{12} + \alpha_{22}$, and $\alpha_{i.} = \alpha_{i1} + \alpha_{i2}$ $(i = 1, 2)$. Under the squared error loss function, the Bayes estimator of Π_1 is given by

$$\hat{\Pi}_1 = E\{\Pi_1 \mid X\} = E\{q q_{11} + (1-q) q_{12} \mid X\}.$$

In view of the posterior distribution (2.6), we finally have

$$\begin{aligned} \hat{\Pi}_1 &= E\{q \mid X\} E\{q_{11} \mid X\} + E\{(1-q) \mid X\} E\{q_{12} \mid X\} \\ &= \frac{1}{\alpha + n}\left(\alpha_{1.} + x_1 + \frac{\alpha_{12}}{\alpha_{.2}} n_2\right). \end{aligned} \tag{2.7}$$

We notice that (see Example in Section 3) $(\alpha_{12}/\alpha_{.2}) n_2$ is the conditional expectation of y_1 - the sample frequency of nonrespondents that belong to A_1 - given the data. Therefore, $\hat{\Pi}_1$ is an intuitive estimator since in the case of full response we would have y_1 in place of $(\alpha_{12}/\alpha_{.2}) n_2$.

The generalization of the above analysis to the case of k categories, $A_1, \ldots, A_k$ $(k \geq 2)$, is straightforward. Tables (1.1) and (1.2) are replaced respectively by

$$\begin{array}{c|cc|c} & R & R' & \\ \hline A_1 & x_1 & & \\ \vdots & \vdots & n_2 & \\ A_k & x_k & & \\ \hline & n_1 & n_2 & n \end{array} \tag{2.8}$$

$$\begin{array}{c|cc|c} & R & R' & \\ \hline A_1 & p_{11} & p_{12} & \Pi_1 \\ \vdots & \vdots & \vdots & \vdots \\ A_k & p_{k1} & p_{k2} & \Pi_k \\ \hline & q & 1-q & 1 \end{array} \tag{2.9}$$

The parameter of interest is now $\boldsymbol{\Pi}=(\Pi_1, \ldots, \Pi_k)$, and the data is $X=(x_1, \ldots, x_k, n_2)$. In place of (2.1), a priori, we consider that

$$\boldsymbol{p}=(p_{11}, \ldots, p_{k1}, p_{12}, \ldots, p_{k2}) \sim D(\alpha_{11}, \ldots, \alpha_{k1}, \alpha_{12}, \ldots, \alpha_{k2}). \tag{2.10}$$

Analogous to (2.2) and (2.3) the following reparametrization is considered:

$$q=\sum_{i=1}^{k} p_{i1}, \quad q_{i1}=\frac{p_{i1}}{q}, \quad q_{i2}=\frac{p_{i2}}{1-q} \qquad (i=1, \ldots, k),$$

$$Q_1=(q_{11}, \ldots, q_{k1}), \qquad Q_2=(q_{12}, \ldots, q_{k2}). \tag{2.11}$$

Conversely

$$p_{i1}=qq_{i1}, \quad p_{i2}=(1-q)q_{i2} \qquad (i=1, \ldots, k),$$

$$\Pi=qQ_1+(1-q)Q_2. \tag{2.12}$$

With the reparametrization (2.11) the likelihood is given by

$$L=q^{n_1}(1-q)^{n_2}\prod_{i=1}^{k} q_{i1}^{x_i}. \tag{2.13}$$

Again, by Lemma 1, to consider (2.10) a priori is equivalent to considering the following set of conditions:

$$q \amalg Q_1 \amalg Q_2, \qquad q \sim B(\alpha_{\cdot 1}, \alpha_{\cdot 2}),$$

$$Q_1 \sim D(\alpha_{11}, \ldots, \alpha_{k1}), \qquad Q_2 \sim D(\alpha_{12}, \ldots, \alpha_{k2}), \tag{2.14}$$

where $\alpha_{\cdot j}=\sum_i \alpha_{ij}$ $(j=1,2)$.

By matching (2.14) with (2.13), we obtain the posterior distribution which is defined by the conditions

$$q \amalg Q_1 \amalg Q_2 \mid X, \qquad q \mid X \sim q \mid n_1 \sim B(\alpha_{\cdot 1}+n_1, \alpha_{\cdot 2}+n_2),$$

$$Q_1 \mid X \sim D(\alpha_{11}+x_1, \ldots, \alpha_{k1}+x_1),$$

$$Q_2 \mid X \sim Q_2 \sim D(\alpha_{12}, \ldots, \alpha_{k2}). \tag{2.15}$$

Again, $\boldsymbol{p} \mid X$ is distributed as Dirichlet if and only if $n_2 = 0$. It is, however, a mixture of Dirichlet distributions and

$$(p_{11}, \ldots, p_{k1}, (1-q)) \mid X \sim D(\alpha_{11} + x_1, \ldots, \alpha_{k1} + x_1, \alpha_{\cdot 2} + n_2).$$

As before, we might consider a conjugate family of distributions by taking $\beta \geq \alpha_{\cdot 2}$ for $\alpha_{\cdot 2}$ in (2.14). Other important types of mixtures of Dirichlet distributions are considered by I.J. Good (1967,76) and J.F. Crook & I.J. Good (1980).

The Bayes estimator for the parameter of interest $\boldsymbol{\Pi} = (\Pi_1, \ldots, \Pi_k)$, analogous to (2.7), has the following form:

$$\hat{\boldsymbol{\Pi}} = E\{\boldsymbol{\Pi} \mid X\} = \frac{1}{\alpha + n}[(\alpha_{1\cdot}, \ldots, \alpha_{k\cdot}) + XM] \tag{2.16}$$

where M is a $(k+1) \times k$-matrix with the $(k+1)$th row being $(\alpha_{12}, \ldots, \alpha_{k2})/\alpha_{\cdot 2}$, the diagonal elements being the unity, and the remaining elements being zero.

The next section deals with the study of the distribution of the data X. The covariance matrix of $\hat{\boldsymbol{\Pi}}$ is presented at the end of the section.

3. The Dirichlet–Multinomial distribution: properties

When the discrete data follow the Multinomial model, the family of Dirichlet distributions is widely used by Bayesians since it is a conjugate family large enough to accommodate various shades of prior opinion. The study of the mixture of Multinomial distributions by a Dirichlet distribution therefore becomes relevant because the (marginal) distribution of the data is then a mixture of this kind. Generalizing the definition of the Beta–Binomial (Ferguson [1967]) this mixture is called here the Dirichlet–Multinomial distribution. More specifically, for $k \geq 2$, let $\boldsymbol{x} = (x_1, \ldots, x_k)$ be a nonnegative integer random vector such that $\sum_{i=1}^{n} x_i = n$ is fixed, and let $\boldsymbol{p} = (p_1, \ldots, p_k)$ be a nonnegative real random vector with $\sum_{i=1}^{k} p_i = 1$.

Definition. If $\boldsymbol{p} \sim D(\alpha_1, \ldots, \alpha_k)$ and $\boldsymbol{x} \mid \boldsymbol{p} \sim M(n; \boldsymbol{p})$, then the distribution of $\boldsymbol{x}$ is called Dirichlet-Multinomial (DM) with parameter $(n; \alpha_1, \ldots, \alpha_k)$, and we write $\boldsymbol{x} \sim \mathrm{DM}(n; \alpha_1, \ldots, \alpha_k)$. When $k = 2$, in place of $(x_1, x_2) \sim \mathrm{DM}(n; \alpha_1, \alpha_2)$, we write $x_1 \sim \mathrm{BB}(n; \alpha_1, \alpha_2)$ to indicate that x_1 is distributed as Beta–Binomial.

It is easy to check that the probability function (p.f.) associated with the DM distribution is given by

$$f(\boldsymbol{x}) = \frac{n!\,\Gamma(\alpha)}{\Gamma(\alpha + n)} \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + x_i)}{x_i!\,\Gamma(\alpha_i)}, \tag{3.1}$$

where $\alpha = \sum_{i=1}^{k} \alpha_i$.

Some of the important properties of the DM distributions are given below. Let $\boldsymbol{x} = (x_1, \ldots, x_k) \sim \mathrm{DM}(n; \alpha_1, \ldots, \alpha_k)$.

Proposition 1. *If* $(i_1, \ldots, i_k)$ *is a permutation of* $(1, \ldots, k)$, *then* $(x_{i_1}, \ldots, x_{i_k}) \sim \mathrm{DM}(n; \alpha_{i_1}, \ldots, \alpha_{i_k})$.

Proposition 2. *If* $m \in \{1, 2, \ldots, k\}$ *is fixed, then for* $\beta = \sum_{i=1}^{m} \alpha_i$

$$\left(x_1, \ldots, x_m, n - \sum_{i=1}^{m} x_i\right) \sim \mathrm{DM}(n; \alpha_1, \ldots, \alpha_m, \alpha - \beta),$$

and

$$n_1 = \sum_{i=1}^{m} x_i \sim \mathrm{BB}(\mathrm{n}; \beta, \alpha - \beta).$$

These two results are immediate consequences of analogous properties of the Multinomial and the Dirichlet distributions.

Proposition 3. *For* m *and* n_1 *defined as above, we have that*

$$(x_1, \ldots, x_m) \mid n_1 \sim \mathrm{DM}(n_1; \alpha_1, \ldots, \alpha_k).$$

Proof. Note that the conditional probability function of $(x_1, \ldots, x_m) \mid n_1$ is obtained by dividing the p.f. of $(x_1, \ldots, x_m, n - n_1)$ by the p.f. of n_1, which is the p.f. of a $\mathrm{DM}(n; \alpha_1, \ldots, \alpha_k)$. □

The result we present next is an important characterization of the DM distribution which will be used in the sequel.

Let $(x_1, \ldots, x_k)$ be a nonnegative integer random vector with $\sum_{i=1}^{k} x_i = n$ fixed. Choose an integer $m \in \{2, \ldots, k-1\}$, and denote $n_1 = \sum_{i=1}^{m} x_i$ with $n_2 = n - n_1$. Consider now the following set of conditions:

$$(x_1, \ldots, x_m) \perp\!\!\!\perp (x_{m+1}, \ldots, x_k) \mid n_1 \tag{3.2)(i}$$

$$\begin{aligned} &(x_1, \ldots, x_m) \mid n_1 \sim \mathrm{DM}(n_1; \alpha_1, \ldots, \alpha_m), \\ &(x_{m+1}, \ldots, x_k) \mid n_1 \sim \mathrm{DM}(n_2; \alpha_{m+1}, \ldots, \alpha_k), \end{aligned} \tag{3.2)(ii}$$

and

$$n_1 \sim \mathrm{BB}\left(n; \sum_{i=1}^{m} \alpha_i, \alpha - \sum_{i=1}^{m} \alpha_i\right). \tag{3.2)(iii}$$

Theorem 1. *The above set of conditions* (3.2) *are necessary and sufficient to have*

$$(x_1, \ldots, x_k) \sim \mathrm{DM}(n; \alpha_1, \ldots, \alpha_k). \tag{iv}$$

Proof. By Propositions 1, 2, and 3, (iv) ⇒ (ii) and (iii). To prove the remaining implications we need only note that (3.1) may be factored as

$$f(\boldsymbol{x})=\left[\frac{n!\,\Gamma(\alpha)}{\Gamma(\alpha+n)}\,\frac{\Gamma(\beta+n_1)\Gamma(\alpha-\beta+n_2)}{n_1!\,n_2!\,\Gamma(\beta)\Gamma(\alpha-\beta)}\right]$$
$$\times\left[\frac{n_1!\,\Gamma(\beta)}{\Gamma(\beta+n_1)}\prod_{i=1}^{m}\frac{\Gamma(\alpha_i+x_i)}{x_i!\,\Gamma(\alpha_i)}\right]\times\left[\frac{n_2!\,\Gamma(\alpha-\beta)}{\Gamma(\alpha-\beta+n_2)}\prod_{i=m+1}^{k}\frac{\Gamma(\alpha_i+x_i)}{x_i!\,\Gamma(\alpha_i)}\right]$$

where, as before, $\alpha=\sum_{i=1}^{k}\alpha_i$, and $\beta=\sum_{i=1}^{m}\alpha_i$. The first factor is the p.f. of a $\text{BB}(n;\beta,\alpha-\beta)$, the second is the p.f. of a $\text{DM}(n_1;\alpha_{m+1},\ldots,\alpha_k)$, and the third is the p.f. of a $\text{DM}(n_2;\alpha_{m+1},\ldots,\alpha_k)$. □

Example. Recalling the Bayes estimator $\hat{\Pi}_1$ presented in (2.7), we notice that $(x_1,x_2)\amalg(y_1,y_2)\mid n_1$, and then $y_1\mid X\sim y_1\mid n_2\sim\text{BB}(n_2;\alpha_{12},\alpha_{22})$ which implies (see (3.3) below) that $E\{y_1\mid X\}=E\{y_1\mid n_2\}=n_2(\alpha_{12}/\alpha_{\cdot 2})$.

An interesting property of the DM distribution is given below where we consider the finite sequence $(z_1,\ldots,z_k)$ with $z_j=\sum_{i=1}^{j}x_i$ $(j=1,\ldots,k)$. Clearly, $z_1=x_1$, $z_m=n_1$, and $z_k=n$.

Corollary. *If* $(x_1,\ldots,x_k)\sim\text{DM}(n;\alpha_1,\ldots,\alpha_k)$, *then* $(z_1,\ldots,z_k)$ *forms a Markov chain.*

It is intuitive that we might give a characterization of the DM distribution in terms of $(z_1,\ldots,z_k)$. This, however, would go beyond our needs.

To present the mean vector and the covariance matrix of the DM distribution we introduce the vector $a=(\alpha_1,\ldots,\alpha_k)$, and the matrix

$$\mathscr{A}=\begin{pmatrix}\alpha_1 & 0 & \cdots & 0\\ 0 & \alpha_2 & \cdots & 0\\ \vdots & & & \vdots\\ 0 & 0 & \cdots & \alpha_k\end{pmatrix}.$$

From Proposition 1 and 2, we notice that $x_i\sim\text{BB}(n;\alpha_i,\alpha-\alpha_i)$, and $x_i+x_j\sim\text{BB}(n;\alpha_i+\alpha_j,\alpha-\alpha_i-\alpha_j)$ for $i,j=1,\ldots,k$ with $i\neq j$. From easy computations when using the definition of BB we have that

$$\begin{aligned}
E\{x_i\}&=n\frac{\alpha_i}{\alpha},\\
\text{Var}\{x_i\}&=\left[\alpha_i-\frac{\alpha_i^2}{\alpha}\right]\frac{\alpha+n}{\alpha(\alpha+1)}n,\\
\text{Var}\{x_i+x_j\}&=\left[\alpha_i+\alpha_j-\frac{(\alpha_i+\alpha_j)^2}{\alpha}\right]\frac{\alpha+n}{\alpha(\alpha+1)}n\\
&=\text{Var}\{x_i\}+\text{Var}\{x_j\}+2\,\text{cov}\{x_i,x_j\}.
\end{aligned}\tag{3.3}$$

From this last equation, it follows that

$$\operatorname{cov}\{x_i, x_j\} = \frac{\alpha_i \alpha_j}{\alpha} \frac{\alpha+n}{\alpha(\alpha+1)} n.$$

Finally, the mean vector and the covariance matrix are given by

$$E\{\boldsymbol{x}\} = \frac{n}{\alpha} a, \qquad \operatorname{Cov}\{\boldsymbol{x}\} = \left[\mathscr{A} - \frac{1}{\alpha} a' a\right] \frac{\alpha+n}{\alpha(\alpha+1)} n$$

where a' is the transpose of a.

The data vector $X = (x_1, \ldots, x_k, n_2)$, for the nonresponse data presented in Section 2, follows the DM model; that is, $X \sim \mathrm{DM}(n; \alpha_{11}, \ldots, \alpha_{k1}, \alpha_{\cdot 2})$. In this case

$$a = (\alpha_{11}, \ldots, \alpha_{k1}, \alpha_{\cdot 2}), \qquad \mathscr{A} = \begin{pmatrix} \alpha_{11} & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{21} & \cdots & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & & \alpha_{k1} & 0 \\ 0 & 0 & \cdots & 0 & \alpha_{\cdot 2} \end{pmatrix}. \tag{3.4}$$

The mean vector and the covariance matrix for $\hat{\boldsymbol{\Pi}}$, the Bayes estimator given by (2.16) are

$$\begin{aligned} E\{\hat{\boldsymbol{\Pi}}\} &= \frac{1}{\alpha+n} [(\alpha_{1\cdot}, \ldots, \alpha_{k\cdot}) + E\{X\}M], \\ \operatorname{Cov}\{\hat{\boldsymbol{\Pi}}\} &= \left(\frac{1}{\alpha+n}\right)^2 M' \operatorname{Cov}\{X\} M. \end{aligned} \tag{3.5}$$

Using (3.4), we have that

$$E\{\hat{\boldsymbol{\Pi}}\} = \frac{1}{\alpha} (\alpha_{1\cdot}, \ldots, \alpha_{k\cdot}),$$

$$M' \mathscr{A} M = \begin{pmatrix} \alpha_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \alpha_{k1} \end{pmatrix} + \frac{1}{\alpha_{\cdot 2}} (\alpha_{12}, \ldots, \alpha_{k2})' (\alpha_{12}, \ldots, \alpha_{k2}),$$

$$\frac{1}{\alpha} M' a' a M = \frac{1}{\alpha} (\alpha_{1\cdot}, \ldots, \alpha_{k\cdot})' (\alpha_{1\cdot}, \ldots, \alpha_{k\cdot}),$$

which imply

$$\begin{aligned} \operatorname{Var}\{\hat{\boldsymbol{\Pi}}_i\} &= \frac{n}{\alpha(\alpha+1)(\alpha+n)} \left[\alpha_{i1} + \frac{\alpha_{i2}^2}{\alpha_{\cdot 2}} - \frac{1}{\alpha} \alpha_{i\cdot}^2\right], \\ \operatorname{Cov}\{\hat{\boldsymbol{\Pi}}_i, \hat{\boldsymbol{\Pi}}_j\} &= \frac{n}{\alpha(\alpha+1)(\alpha+n)} \left[\frac{\alpha_{i2}\alpha_{j2}}{\alpha_{\cdot 2}} - \frac{1}{\alpha} \alpha_{i\cdot} \alpha_{j\cdot}\right] \end{aligned} \tag{3.6}$$

for $i, j = 1, \ldots, k$ and $i \neq j$.

In the particular case where $\alpha_1 = \cdots = \alpha_k = 1$, the probability function is $f(\boldsymbol{x}) =$

$\binom{n+k-1}{n}^{-1}$ and this corresponds to the Bose–Einstein statistic in Statistical Mechanics. See Feller (1968) for additional discussion.

4. The DM distribution: a natural family of priors for finite population studies

A sample of fixed size n is taken from a population of finite size N which is partitioned in $k \leq N$ categories. The category frequency counts are represented by $\theta_1, \ldots, \theta_k$ with $\sum_{i=1}^{k} \theta_i = N$. From the sample, an inference about $(\theta_1, \ldots, \theta_k)$ is required. Corresponding to each θ_i $(i=1, \ldots, k)$, x_i is the sample frequency count of the i-th category, where $\sum_{i=1}^{k} x_i = n$.

We restrict the choice of the prior distribution for $\boldsymbol{\theta}$ to the family of DM distributions. Given the sample $\boldsymbol{x} = (x_1, \ldots, x_k)$, we want to derive the posterior distribution of $\boldsymbol{\theta} - \boldsymbol{x} = (\theta_1 - x_1, \ldots, \theta_k - x_k)$, the composition of the unsampled part of the population. In order to reach this goal, we use only intuitive arguments since an algebraic analysis, besides being tedious (albeit easy), would bury the beauty of the argument.

Let $e_1 = (1, 0, \ldots, 0), \ldots, e_k = (0, \ldots, 0, 1)$ be the standard orthonormal basis for R^k. To each unit j $(j = 1, \ldots, N)$ of the population $\mathscr{P}$, we associate an incidence vector y_j which is equal to e_i if the category of j is c_i. More specifically, let $\mathscr{P} = \{1, \ldots, N\}$ be an enumeration of the population units. Associated with $\mathscr{P}$ are the incidence vectors $y_1, \ldots, y_N$ described above. The unknown vector is $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) = \sum_{j=1}^{N} y_j$. We are considering the case where the sample selection is noninformative. That is, the selection of the n units (sample) from $\mathscr{P}$ is based only on the labels $1, \ldots, N$, which are themselves uninformative about the incidence vectors $y_1, \ldots, y_N$.

A natural way to introduce the prior model $\boldsymbol{\theta} \sim \mathrm{DM}(N; \alpha_1, \ldots, \alpha_k)$ is to consider a random vector $\boldsymbol{p} = (p_1, \ldots, p_k) \sim D(\alpha_1, \ldots, \alpha_k)$ and to stipulate that for $j = 1, \ldots, N$

$$y_j \mid \boldsymbol{p} \sim M(1; \boldsymbol{p}) \quad \text{and} \quad y_1 \amalg \cdots \amalg y_N \mid \boldsymbol{p}.$$

In other words, given $\boldsymbol{p}$ the y_j's are i.i.d. with common distribution $M(1; \boldsymbol{p})$. Since $(y_1, \ldots, y_N)$ is an exchangable finite sequence, without loss of generality we can consider our sampled items as being the first n population items, say $\{1, 2, \ldots, n\}$. Now, the sample is represented by the vector $\boldsymbol{x} = \sum_{i=1}^{n} y_i$, and the unknown quantity of interest is the vector $\boldsymbol{\theta} - \boldsymbol{x} = \sum_{i=n+1}^{N} y_i$.

In terms of the pseudo parameter $\boldsymbol{p}$ we then have, a priori, the following:

$$\boldsymbol{p} \sim D(\alpha_1, \ldots, \alpha_k), \tag{4.1)(i}$$

$$\boldsymbol{x} \mid \boldsymbol{p} \sim \mathrm{M}(n; \boldsymbol{p}), \tag{4.1)(ii}$$

$$\boldsymbol{x} \sim \mathrm{DM}(n; \alpha_1, \ldots, \alpha_k), \tag{4.1)(iii}$$

$$(\boldsymbol{\theta} - \boldsymbol{x}) \mid \boldsymbol{p} \sim \mathrm{M}(N - n; \boldsymbol{p}), \tag{4.1)(iv}$$

$$(\boldsymbol{\theta} - \boldsymbol{x}) \sim \mathrm{DM}(N - n; \alpha_1, \ldots, \alpha_k), \tag{4.1)(v}$$

$$(\boldsymbol{\theta} - \boldsymbol{x}) \amalg \boldsymbol{x} \mid \boldsymbol{p}, \tag{4.1)(vi}$$

$$\theta \mid \boldsymbol{p} \sim \mathrm{M}(N; \boldsymbol{p}). \tag{4.1)(vii}$$

The result below is useful to our discussion.

Lemma 2. *If two independent random vectors X, and Y are such that* $X \sim \mathrm{M}(n_1; \boldsymbol{p})$ *and* $Y \sim \mathrm{M}(n_2; \boldsymbol{p})$*, then the conditional distribution of* $X \mid X+Y$ *is the Multivariate Hypergeometric with parameter* $(n_1+n_2, n_1, X+Y)$; *that is,*

$$X \mid X+Y \sim \mathrm{H}(n_1+n_2, n_1, X+Y).$$

(*Note that this distribution does not depend on the value of* $\boldsymbol{p}$.)

The following conclusion based on (4.1) and Lemma 2 is important since it defines the likelihood function:

$$\boldsymbol{x} \mid (\theta, \boldsymbol{p}) \sim \boldsymbol{x} \mid \theta \sim \mathrm{H}(N, n, \theta) \tag{4.2}$$

From the Bayesian analysis of the multinomial case, we recall that

$$\boldsymbol{p} \mid \boldsymbol{x} \sim D(\alpha_1 + x_1, \ldots, \alpha_k + x_k).$$

On the other hand we notice that the conditional distribution of $(\theta - \boldsymbol{x}) \mid \boldsymbol{x}$ may be viewed as a composition of the distribution of $(\theta - \boldsymbol{x}) \mid \boldsymbol{p} \sim (\theta - \boldsymbol{x}) \mid (\boldsymbol{p}, \boldsymbol{x})$ (see (4.1)) by the distribution of $\boldsymbol{p} \mid \boldsymbol{x}$. Now, from the definition of the DM distribution, we have that $(\theta - \boldsymbol{x}) \mid \boldsymbol{x} \sim \mathrm{DM}(N-n; \alpha_1 + x_1, \ldots, \alpha_k + x_k)$. This is the main result of this section and may be summarized as:

Theorem 2. *For the finite population sampling situation described, if* $\theta \sim \mathrm{DM}(N; \alpha_1, \ldots, \alpha_k)$ *a priori, then* $(\theta - \boldsymbol{x}) \mid \boldsymbol{x} \sim \mathrm{DM}(N-n; \alpha_1 + x_1, \ldots, \alpha_k + x_k)$ *a posteriori.*

The next section is devoted to the nonresponse problem in finite populations.

5. Nonresponse: the Multivariate Hypergeometric model

The data for the nonresponse problem is presented in the $(k \times 2)$-tabular form as in (2.7). Instead of (1.4), the population parameters have the following representation:

	R	R'	
A_1	θ_{11}	θ_{12}	θ_1
A_2	θ_{21}	θ_{22}	θ_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_k	θ_{k1}	θ_{k2}	θ_k
	ψ	$N-\psi$	N

(5.1)

Now, the parameter of interest is $\theta=(\theta_1, \ldots, \theta_k)$, and the likelihood may be written as

$$L=\frac{\binom{\theta_{11}}{x_1}\cdots\binom{\theta_{k1}}{x_1}}{\binom{\psi}{n_1}}\;\frac{\binom{\psi}{n_1}\binom{N-\psi}{n_2}}{\binom{N}{n}}. \tag{5.2}$$

Suppose that, a priori, a DM distribution for $\theta=(\theta_{11}, \ldots, \theta_{k1}, \theta_{12}, \ldots, \theta_{k2})$ is considered; that is, a priori

$$\theta \sim \mathrm{DM}(N; \alpha_{11}, \ldots, \alpha_{k1}, \alpha_{12}, \ldots, \alpha_{k2}). \tag{5.3}$$

An additional notation is introduced:

$$\theta_1=(\theta_{11}, \ldots, \theta_{k1}) \quad \text{and} \quad \theta_2=(\theta_{12}, \ldots, \theta_{k2}).$$

Recall that $\theta=\theta_1+\theta_2$ is the parameter of interest, and that $X=(x_1, \ldots, x_k, n_2)$ is the data vector which may also be represented in the slightly abbreviated form $\boldsymbol{x}=(x_1, \ldots, x_k)$. Let

$$\alpha_{.j}=\sum_{i=1}^{k} \alpha_{ij} \quad (j=1,2), \qquad \alpha_{i.}=\alpha_{i1}+\alpha_{i2} \quad (i=1, \ldots, k),$$

and $\alpha=\sum\sum \alpha_{ij}$.

Writing the parameters in the extended form $(\psi, \theta_1, \theta_2)$, it is convenient to describe the prior (5.3) in the following equivalent form (see Theorem 1):

$$\psi \sim \mathrm{BB}(N; \alpha_{.1}, \alpha_{.2}), \tag{5.4)(i}$$

$$\begin{aligned}&\theta_1 \mid \psi \sim \mathrm{DM}(\psi; \alpha_{11}, \ldots, \alpha_{k1}),\\ &\theta_2 \mid \psi \sim \mathrm{DM}(N-\psi; \alpha_{12}, \ldots, \alpha_{k2}),\end{aligned} \tag{5.4)(ii}$$

and

$$\theta_1 \amalg \theta_2 \mid \psi \tag{5.4)(iii}$$

The theorem presented below is the main result of this section. It allows a simple derivation of the Bayes estimator.

Theorem 3. *The posterior distribution derived from the Bayes operation, when* (5.2) *is the likelihood and* (5.4) *defines the prior, is given by the following set of conditions:*

$$(\psi-n_1) \mid X \sim (\psi-n_1) \mid n_1 \sim \mathrm{BB}(N-n; \alpha_{.1}+n_1, \alpha_{.2}+n_2), \tag{5.5)(i)'}$$

$$\begin{aligned}&(\theta_1-\boldsymbol{x}) \mid (\psi, X) \sim \mathrm{DM}(\psi-n_1; \alpha_{11}+x_1, \ldots, \alpha_{k1}+x_k),\\ &\theta_2 \mid (\psi, X) \sim \theta_2 \mid \psi,\end{aligned} \tag{5.5)(ii)'}$$

and

$$\theta_1 \amalg \theta_2 \mid (\psi, X) \tag{5.5)(iii)'}$$

Proof. The second condition of (ii)′ follows from the fact that the likelihood does not depend on θ_2 when ψ is known. This fact together with the prior condition (iii), implies (iii)′.

To prove (i)′ and the first condition of (ii)′ we consider (as in Section 2) the invisible nonresponse sample frequency counts, say $\boldsymbol{y}=(y_1, \dots, y_k)$. If we had full response, the data would have been represented by $(\boldsymbol{x}, \boldsymbol{y})$. From Theorems 1 and 2 we have that

$$\psi - n_1 \mid (\boldsymbol{x}, \boldsymbol{y}) \sim \mathrm{BB}(N-n; \alpha_{.1}+n_1, \alpha_{.2}+n_2), \tag{a}$$

$$\theta_1 - \boldsymbol{x} \mid (\psi, \boldsymbol{x}, \boldsymbol{y}) \sim \mathrm{DM}(\psi - n_1; \alpha_{11}+x_1, \dots, \alpha_{k1}+x_2). \tag{b}$$

From (a) and (b) it follows that $(\psi - n_1) \mid (\boldsymbol{x}, \boldsymbol{y}) \sim (\psi - n_1) \mid n_1$, and that $(\theta_1 - \boldsymbol{x}) \mid (\psi, \boldsymbol{x}, \boldsymbol{y}) \sim (\theta_1 - \boldsymbol{x}) \mid (\psi, X)$ which imply (i)′ and the first condition of (ii)′ respectively. □

Note that we showed above that $\psi \perp\!\!\!\perp X \mid n_1$; that is, n_1 is partially Bayes sufficient to predict ψ. See Basu (1977) for a more complete discussion of this concept.

As in the multinomial case, the posterior (5.5) does not define a distribution in the same class as the prior was chosen from; that is, (5.5) does not define a DM distribution. It is easy to check, however, that

$$\begin{aligned}&(\theta_{11}-x_1, \dots, \theta_{k1}-x_k, N-\psi-n_2)\\ &\quad \sim \mathrm{DM}(N-n; \alpha_{11}+x_1, \dots, \alpha_{k1}+x_k, \alpha_{.1}+n_2).\end{aligned}$$

A more complete class might be considered by taking in (5.4) a $\beta \geq \alpha_{.2}$ for $\alpha_{.2}$ in (i).

From the posterior (5.5) we obtain the following results:

$$E\{\psi - n_1 \mid X\} = (N-n)\frac{\alpha_{.1}+n_1}{\alpha+n},$$

$$E\{N-\psi \mid X\} = n_2 + (N-n)\frac{\alpha_{.2}+n_2}{\alpha+n},$$

$$E\{\theta_{i1} - x_i \mid (\psi, X)\} = (\psi - n_1)\frac{\alpha_{i1}+x_i}{\alpha_{.1}+n_1},$$

$$E\{\theta_{i2} \mid (\psi, X)\} = E\{\theta_{i2} \mid \psi\} = (N-\psi)\frac{\alpha_{i2}}{\alpha_{.2}}.$$

Using now the properties of conditional expectation we have the Bayes estimators

$$\begin{aligned}\hat{\theta}_i &= E\{\theta_i \mid X\} = E\{\theta_{i1}+\theta_{i2} \mid X\} = E\{\theta_{i1} \mid X\} + E\{\theta_{i2} \mid X\}\\ &= x_i + \frac{\alpha_{i1}+x_i}{\alpha_{.1}+n_1}E\{\psi - n_1 \mid X\} + \frac{\alpha_{i2}}{\alpha_{.2}}E\{N-\psi \mid X\}\\ &= \frac{\alpha+N}{\alpha+n}\left(x_i + n_2\frac{\alpha_{i2}}{\alpha_{.2}}\right) + (N-n)\frac{\alpha_{i.}}{\alpha+n}.\end{aligned}$$

As in (2.15), the Bayes estimator of the parameter of interest $\theta=\theta_1+\theta_2$ is given by:

$$\hat{\theta}=E\{\theta \mid X\}=\frac{\alpha+N}{\alpha+n}XM+\frac{N-n}{\alpha+n}(\alpha_{1.},\ldots,\alpha_{k.}).$$

Using the results (3.4), and (3.5) we finally have:

$$E\{\hat{\theta}\}=\frac{N}{\alpha}(\alpha_{1.},\ldots,\alpha_{k.}),\qquad \operatorname{Cov}\{\hat{\theta}\}=\left(\frac{\alpha+N}{\alpha+n}\right)^2 M'\operatorname{Cov}\{X\}M,$$

which implies that

$$\operatorname{Cov}\{\hat{\theta}_i,\hat{\theta}_j\}=\frac{n(\alpha+N)^2}{(\alpha+n)(\alpha+1)\alpha}\left(\delta_{ij}\alpha_{i1}+\frac{\alpha_{i2}\alpha_{j2}}{\alpha_{.2}}-\frac{\alpha_{i.}\alpha_{j.}}{\alpha}\right), \tag{5.6}$$

where δ_{ij} is the Kronecker delta, and $\operatorname{Var}\{\theta_i\}=\operatorname{Cov}\{\theta_i,\theta_i\}$.

6. Final remarks

(i) There are many follow-up techniques used to obtain response among some of the n_2 units that have not responded initially. For example, from the n_2 nonrespondents in our sample, we select a subsample of size $n_2'\leq n_2$ and offer an incentive to those who now would respond. In that way, information about Q_2 in (2.14) or about $\theta_2 \mid \psi$ in (5.4) might be improved. See Kaufman and King (1973), and Singh and Sedransk (1978) for a more specific discussion on this two stage sampling.

(ii) Although we have restricted ourselves to the nonresponse problem, it should be understood that our method applies equally well to the general problem of categorical data with missing entries. Consider, for instance, the categorical data where all but the first n cell entry data are missing. By using Lemma 1 for the multinomial case or Theorem 1 for the hypergeometric case, we would, analogously to (2.6) or (5.4), obtain the posterior distribution for the cell parameters.

(iii) One word about the relevance of the variance of Bayes estimators as presented in (3.6) and (5.6). Note that we are not talking about conditional variances (with the parameter fixed) but the variance of the marginal distribution of the estimator. Consider the $k=2$ multinomial case for instance. It is clear that

$$\operatorname{Var}\{\hat{\Pi}_1\}=\operatorname{Var}\{\Pi_1\}-E\{\operatorname{Var}\{\Pi_1 \mid X\}\};$$

that is, the variance of $\hat{\Pi}_1$ may be regarded as the expected amount of uncertainty removed, when uncertainty (De Groot (1962)) about the parameter is measured by its variance. Thus, the variance of the Bayes estimator is a kind of a measure of the amount of information in the experiment. The larger the variance of $\hat{\Pi}_1$ is, the better off we are!

The variance of the Bayes estimator may be used (see Appendix) to study the amount of information lost when the nonresponse portion of the sample is neglected as in many classical procedures.

(iv) We notice that, there are many important applications of the DM class of distributions. For example, in Quality Control [see A. Hald (1960,78)] and in Statistical Prediction [see J. Aitchison & I.R. Dunsmore (1975)].

(v) A special case of the DM distribution is considered in I.J. Good (1965, p. 36). Another characterization of this distribution in terms of mixtures of Negative-Binomial distributions can be found in D. Basu & C.A. de B. Pereira (1980).

(vi) Finally, we would like to notice that the Bayes estimators discussed in this paper are not consistent in the classical sense. That is, $E\{(\hat{\Pi}_i-\Pi_i)^2 \mid \Pi\}$ does not converge to zero as $n\to\infty$.

Appendix

The usual non-Bayesian methods for analyzing data with nonresponse do not use the nonresponse portion of the sample. The likelihood in this case is defined by the conditional probability of $\boldsymbol{x}=(x_1,\dots,x_k)$, the response vector, given $n_1=\sum_{i=1}^{k}x_i$. For instance, in the multinomial case this 'conditional' likelihood is $L_c=\prod_{i=1}^{k}q_{i1}^{x_i}$. It is intuitive that, by considering this reduction, one is not using the full information (about the parameter of interest) contained in the data. In order to clarify this point we define a reasonable measure of information and compute, in a particular case, its values for both the original and the conditional model.

Consider the Multinomial model for the case of two categories ($k=2$). Let the prior be the uniform distribution; that is, $\alpha_{11}=\alpha_{21}=\alpha_{12}=\alpha_{22}=1$. By using the variance (see Section 6, (iii)) as the uncertainty function, we define the measure of information as

$$I(\text{data})=(\operatorname{Var}\{\Pi_1\})^{-1}\operatorname{Var}\{E\{\Pi_1 \mid \text{data}\}\}.$$

Considering the original likelihood, the information measure is given by

$$I(X)=I=\frac{n}{2(4+n)}$$

since

$$\operatorname{Var}[\Pi_1\}=(20)^{-1} \text{ and } \operatorname{Var}[\hat{\Pi}_1]=\frac{\text{n}}{40(4+\text{n})}.$$

The posterior distribution under the conditional model (and same prior) is defined by the following conditions:

$$q'\sim B(2,2),\qquad q'_{11}\sim B(1+x_1,1+x_2),$$

$$q'_{12}\sim B(1,1),\qquad q'\amalg q'_{11}\amalg q'_{12}.$$

The Bayesian estimator in this case is given by

$$\tfrac{1}{2}\frac{1+x_1}{2+n_1}+\tfrac{1}{4},$$

and the respective measure of information is

$$I(\boldsymbol{x}\mid n_1)=I_{\mathrm{c}}=5\ \mathrm{Var}\left\{\frac{1+x_1}{2+n_1}\right\}.$$

Relative to the uniform prior, the distribution of (x_1,x_2,n_2) is $\mathrm{DM}(n;1,1,2)$. Considering the particular case of $n=4$ we obtain the following results:

$$I=\tfrac{1}{4},\qquad I_{\mathrm{c}}=\tfrac{51}{280},\qquad \frac{I-I_{\mathrm{c}}}{I}=0.27.$$

Here we might say that if an inference about Π_1 is required, then 27% of the information is expected to be lost (relatively) when the nonresponse portion is neglected.

Note that it is possible to have an analogous analysis for the Hypergeometric model. However, in addition to the value of n, we would have to fix a value for N, the population size. Here, the conditional model is given by

$$L_{\mathrm{c}}=\binom{\psi}{n_1}^{-1}\prod_{i=1}^{k}\binom{\theta_i}{x_i}.$$

For particular choices of N, the relative loss of information would appear to be more extreme.

Acknowledgements

We would like to thank Professor Shelley Zacks and the referee for their helpful suggestions.

References

Aitchison, J. and I.R. Dunsmore (1975). *Statistical Prediction Analysis.* Cambridge University Press, London.

Basu, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.* 72, 355–66.

Basu, D. (1979). A discussion on survey theory. *Symposium on Incomplete Data*, August 10–11, Washington, DC.

Basu, D. and C.A. de B. Pereira (1980). A Note on Blackwell Sufficiency and a Skibinsky Characterization of Distributions. *Sankhyā*, to appear.

Crook, J.F. and I.J. Good (1980). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables, Part II. *Annals of Statistics* 8, 1198–1218; Correction, Vol. 9.

Degroot, M.H. (1962). Uncertainty, information, and sequential experiments. *Ann. Math. Statistics* 33, 404–19.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications.* Vol. I, 3rd edn. Wiley, New York.

Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

Good, I.J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press, New York/Cambridge, MA.

Good, I.J. (1967). A Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. Ser. B*, 29, 399–431.

Good, I.J. (1967). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Annals of Statistics* 4, 1159–1189.

Hald, A. (1960). The compound hypergeometric distribution and a system of single sampling inspection plan based on prior distributions and costs. *Technometrics 2*, 275–340. Discussion, 370–372.

Hald, A. (1978). *Statistical Theory of Sampling Inspection by Attributes.* University of Copenhagen, Denmark.

Kaufman, G.M. and B. King (1973). A Bayesian Analysis of Nonresponse in Dichotomous Processes. *J. Amer. Statist. Soc. 68*, 670–678.

Pereira, C.A. de B. (1980). Bayesian Solution to Some Classical Problems of Statistics. PhD. Thesis, The Florida State University.

Singh, B. and I. Sedransk (1978). A two-phase sample design for estimating the finite population mean when there is nonresponse. In: N.K. Namboodiri, Ed., *Survey Sampling and Measurement.* Academic Press, New York.