BAYESIAN PREDICTIVE INFERENCE FOR A FINITE POPULATION PROPORTION: TWO STAGE CLUSTER SAMPLING

B. Nandram, WPI and J. Sedransk, NCHS
B. Nandram, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609

KEY WORDS: Beta distribution, exchangeable, mixture distributions

Abstract
Given binary data from a two stage cluster sample, we present a method to carry out Bayesian predictive inference for a finite population proportion. Our probabilistic specification should be useful for many surveys of this type, and yields simple analytical expressions for the prior and posterior mean and variance. Within cluster k, we assume that the Y_{ki} are a random sample from the Bernoulli distribution with probability θ_k. Conditional on β and τ, $\theta_1,\ldots,\theta_N$ are a random sample from a beta distribution. Finally, β has a discrete distribution with specified probabilities. We use data from the National Health Interview Survey to illustrate the methodology and to show how to choose values for the parameters in the prior distribution.

1. Introduction

Given binary data from a two stage cluster sample, one may wish to use Bayesian methods to make inference about a finite population proportion, P. To do so, we use a probabilistic specification that is an extension of one proposed by Albert and Gupta (1983; Section 3.1), and that should be useful in many situations where two stage cluster sampling is employed. This probabilistic specification may be viewed as a discrete analogue of the model for two stage cluster sampling with normal data described by Scott and Smith (1969).

It is assumed throughout that n clusters are sampled from the N clusters in the population. Denote by M_k and m_k the known number of units and sample size in cluster k $(0 \le m_k \le M_k)$. Letting Y_{ki} denote the Bernoulli random variable corresponding to the i^{th} unit in cluster k, it is assumed that $\{Y_{ki} : i = 1,\ldots,M_k\}$ are independent with

$$\Pr(Y_{ki} = 1 \mid \theta_k) = \theta_k. \qquad (1.1)$$

Letting $M_k' = \sum_{i=1}^{M_k} Y_{ki}$, $\Delta_k = M_k'/M_k$ and $\rho_k = M_k / \sum_{k=1}^{N} M_k$, we wish to make inference about

$$P = \sum_{k=1}^{N} M_k' / \sum_{k=1}^{N} M_k = \sum_{k=1}^{N} \rho_k \Delta_k. \qquad (1.2)$$

In small area estimation one may also make inference about the individual Δ_k.

For the prior distribution, given β and τ, we take $\theta_1,\ldots,\theta_N$ to be distributed independently with beta density function

$$p(\theta|\beta,\tau)=(B(\beta,\tau-\beta))^{-1}\theta^{\beta-1}(1-\theta)^{\tau-\beta-1} \qquad (1.3)$$

where $B(a,b) = \Gamma(a)\Gamma(b)\{\Gamma(a+b)\}^{-1}$.
For $\beta \in \{a_r : 0 < a_1 < a_2 < .. < a_R < \tau\}$,

$$\Pr(\beta = a_r) = \omega_r \qquad (1.4)$$

where $\sum_{r=1}^{R} \omega_r = 1$. It is assumed throughout that τ and the ω_r are fixed quantities; methods for assigning values are discussed in Section 3.

Albert and Gupta (1983), considering only the special case, $a_r = r$, $R = \tau - 1 > 0$ and $\omega_r = (\tau-1)^{-1}$, used (1.3) and (1.4) to model similarity among the probabilities in an $N \times 2$ contingency table. Lehoczky and Schervish (1987) considered empirical Bayes predictive inference for quantities such as P in (1.2). For a two stage sample they would use only (1.3) and estimate β and τ using the marginal distribution of the observed Y_{ki}. We share with Lehoczky and Schervish their concern about understated variability when one uses empirical Bayes methods. By careful choice of $\{a_r, \omega_r : r = 1,\ldots,R\}$, (1.3) and (1.4) provide a flexible, marginal (mixture) prior distribution for $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_N)$; i.e.,

$$p'(\underset{\sim}{\theta}|\tau) = \sum_{r=1}^{R} \omega_r\{B(a_r,\tau-a_r)\}^{-N} \prod_{k=1}^{N} \theta_k^{a_r-1}(1-\theta_k)^{\tau-a_r-1}, \qquad (1.5)$$

and is the basis for a fully Bayes analysis. The mixture of natural conjugate priors (1.5) is a very convenient class of priors; see Dalal and Hall (1983).

There are other Bayesian methods used to model similarity among the probabilities θ_k. Leonard (1972) used the logit transformation on the θ_k and assumed that the logits are a random sample from a normal distribution. Novick, Lewis and Jackson (1973) used arc–sine transformations on the θ_k and adjusted sample proportions, but again assumed normality of these transformed proportions. The essence of these approximations is the possibility of applying the general Bayesian normal linear model theory of Lindley and Smith (1972). The method proposed here uses a flexible prior distribution, is easily implemented and avoids the uncertain approximations in the Leonard and Novick, et. al. approaches.

Using (1.5) we show in Section 2 that the exact posterior moments of P have sensible forms and are easily calculated. Using data from the National Health Interview Survey (NHIS) we show in Section 3 how to choose τ in (1.3) and $\{(a_r,\omega_r) : r = 1,\ldots,R\}$ in (1.4), and illustrate predictive inference. Section 4 summarizes our approach and suggests extensions.

In larger surveys one might apply this methodology by first assigning all second stage units to a set of mutually exclusive and exhaustive domains. For each domain one would then allocate the first stage units in the population to a set of mutually exclusive and exhaustive strata in a way that there is exchangeability among the "first stage" parameters within each stratum. One would then apply the probabilistic specification in (1.1), (1.3) and (1.4) within each domain x stratum, and obtain national (or sub-national) estimates by summing the estimated totals obtained for each domain x stratum. Let Y denote the NHIS binary variable where $Y = 1$ if, and only if, the person has seen a doctor at least once during the past year. Then the domains may be defined by characteristics such as an individual's age, sex and race. If counties are the first stage units, the strata may be defined by socio-economic variables measured at the county level such as per capita income, education level, and number of doctors per 1,000 population.

2. Inference

2.1 Prior Moments of P

Letting $\underset{\sim}{M}' = (M_1',\ldots,M_N')'$ and $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_N)'$, it is clear from (1.1) that given $\underset{\sim}{\theta}$, the M_k' have independent binomial distributions:

$$p(M_k'|\theta_k) = \begin{bmatrix} M_k \\ M_k' \end{bmatrix} \theta_k^{M_k'}(1-\theta_k)^{M_k-M_k'} \qquad (2.1)$$

$k = 1,2,\ldots,N$. Using (1.5),

$$p(\underset{\sim}{M}'|\tau) = \sum_{r=1}^{R} \omega_r \prod_{k=1}^{N} \times \left\{ \begin{bmatrix} M_k+\tau-1 \\ M_k \end{bmatrix}^{-1} \begin{bmatrix} M_k'+a_r-1 \\ M_k' \end{bmatrix} \begin{bmatrix} M_k-M_k'+\tau-a_r-1 \\ M_k-M_k' \end{bmatrix} \right\}; \qquad (2.2)$$

where for any nonnegative real numbers $g, h, g \le h$, $\begin{bmatrix} h \\ g \end{bmatrix}$ denotes $\Gamma(h+1)\{\Gamma(g+1)\Gamma(h-g+1)\}^{-1}$. If the a_r are integers, the marginal prior distribution of $\underset{\sim}{M}'$ in (2.2) is a weighted average of negative hypergeometric distributions.

Let

$$E(\beta/\tau|\tau) = \eta_1 = \sum_{r=1}^{R} \omega_r a_r \tau^{-1},\ \mathrm{var}(\beta/\tau|\tau) = \eta_2 \text{ and}$$
$$E\{\beta(\tau-\beta)/\tau(\tau+1)|\tau\} = \eta_3.$$

With $\Delta_k = M_k'/M_k$, it is easily seen that

$$E(\Delta_k|\tau) = \eta_1 \qquad (2.3)$$

and

$$\mathrm{cov}(\Delta_k,\Delta_\ell|\tau) = \begin{cases} \eta_2 + \left[\tau^{-1} + M_k^{-1}\right]\eta_3 ; & k = \ell \\ \eta_2 ; & k \ne \ell \end{cases}$$

Now, $\mathrm{corr}(\Delta_k, \Delta_\ell|\tau) =$

$$\{1+(\tau^{-1}+M_k^{-1})(\eta_3/\eta_2)\}^{-1/2} \times \{1+(\tau^{-1}+M_\ell^{-1})(\eta_3/\eta_2)\}^{-1/2} \qquad (2.3a)$$

a monotone increasing function of τ.

Finally, from (2.3), $E(P|\tau) = \eta_1$ and

$$\mathrm{var}(P|\tau) = \eta_2 \sum_{k=1}^{N} \sum_{\ell=1}^{N} \rho_k\rho_\ell + \eta_3 \sum_{k=1}^{N} \rho_k^2\left[\tau^{-1} + M_k^{-1}\right]$$

where $\rho_k = M_k / \sum_{\ell=1}^{N} M_\ell$, $k = 1,2,\ldots,N$.

2.2 Posterior Moments of P

Let m_k' denote the number of sampled units in cluster k having the desired characteristic, $\underset{\sim}{m}' = (m_1',\ldots,m_n')'$, and $\underset{\sim}{m} = (m_1,\ldots,m_n)'$, the vector of sample sizes. Using (1.1), (1.3) and (1.4) the posterior distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{m}$, $\underset{\sim}{m}'$ and τ is given by

$$p(\underset{\sim}{\theta}|\underset{\sim}{m},\underset{\sim}{m}',\tau) =$$
$$\sum_{r=1}^{R} \omega_r^* \prod_{k=1}^{N} \{B(a_r+I_k m_k', \tau-a_r+I_k(m_k-m_k'))\}^{-1} \times \theta_k^{a_r+I_k m_k'-1}(1-\theta_k)^{\tau-a_r+I_k(m_k-m_k')-1} \qquad (2.4)$$

where

$$\omega_r^* = \Pr(\beta=a_r|\underset{\sim}{m},\underset{\sim}{m}',\tau) = \omega_r \prod_{k\epsilon s} \begin{bmatrix} \tau-2 \\ a_r-1 \end{bmatrix} \begin{bmatrix} m_k + \tau-2 \\ m_k'+a_r-1 \end{bmatrix}^{-1} \times \left\{ \sum_{r=1}^{R} \omega_r \prod_{k\epsilon s} \begin{bmatrix} \tau-2 \\ a_r-1 \end{bmatrix} \begin{bmatrix} m_k+\tau-2 \\ m_k'+a_r-1 \end{bmatrix}^{-1} \right\}^{-1}, \qquad (2.5)$$

$r = 1,2,\ldots,R$, and $I_k = 1$ if cluster k is in the sample ($k\epsilon s$) and $I_k = 0$, otherwise.

Given $\underset{\sim}{m}$, $\underset{\sim}{m}'$ and $\underset{\sim}{\theta}$, $M_k' - I_k m_k' (k=1,\ldots,N)$ are independent binomial random variables (see (1.1)). Using (2.4),

$$p(\{M_k'-I_k m_k' : k = 1,\ldots,N\} \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) =$$

$$\sum_{r=1}^{R} \omega_r^* \prod_{k=1}^{N} \left\{ \binom{M_k + \tau - 1}{M_k - I_k m_k}^{-1} \binom{M_k' + a_r - 1}{M_k' - I_k m_k'} \times \binom{M_k - M_k' + \tau - a_r - 1}{M_k - M_k' - I_k(m_k - m_k')} \right\}. \qquad (2.6)$$

As expected, (2.6) lies in the same class as (2.2).

Using (2.5),

$$E(\beta/\tau \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) = \hat{\eta}_1 = \sum_{r=1}^{R} \omega_r^* a_r \tau^{-1},$$

$$\mathrm{Var}\left[\beta/\tau \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau\right] = \hat{\eta}_2 = \sum_{r=1}^{R} \omega_r^* (a_r \tau^{-1} - \hat{\eta}_1)^2,$$

and

$$E\{\beta(\tau-\beta)/\tau(\tau+1) \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau\} =$$

$$\hat{\eta}_3 = (\tau+1)^{-1} \sum_{r=1}^{R} \omega_r^* a_r (1 - a_r \tau^{-1}). \qquad (2.7)$$

Note that the $\hat{\eta}_i$ are updates of the η_i defined above (2.3). Defining

$$\lambda_k = \begin{cases} \{1+(\tau/M_k)\}\{1+(\tau/m_k)\}^{-1} ; & k \in s \\ 0 ; & k \notin s \end{cases},$$

it can be shown that

$$E(\Delta_k \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) = \lambda_k \left[\frac{m_k'}{m_k}\right] + (1 - \lambda_k)\hat{\eta}_1 \qquad (2.8)$$

and

$$\mathrm{cov}(\Delta_k, \Delta_\ell \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) =$$

$$\begin{cases} (1-\lambda_k)^2\, \hat{\eta}_2 + (1-\lambda_k)(\tau^{-1} + M_k^{-1})\hat{\nu}_k^2 ; & k = \ell \\ (1-\lambda_k)(1-\lambda_\ell)\, \hat{\eta}_2 ; & k \neq \ell \end{cases} \qquad (2.9)$$

where for $k \in s$

$$\hat{\nu}_k^2 =$$

$$\sum_{r=1}^{R} \omega_r^* (a_r + m_k')(\tau + m_k - (a_r + m_k'))/(\tau + m_k)(\tau + m_k + 1);$$

and for $k \notin s$ $\hat{\nu}_k^2 = \hat{\eta}_3$.

In (2.8) $E(\Delta_k \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau)$ has a sensible form: it is a weighted average of the sample proportion from cluster k and $\hat{\eta}_1$, the posterior expected value of η_1. (Recall from (2.3) that $E(\Delta_k \mid \tau) = \eta_1$.) Since $\hat{\eta}_1$ is a function of $\underset{\sim}{m}'$, there is a "borrowing of strength". However, if $m_k = M_k$, $\lambda_k = 1$ and no borrowing is required since all information is obtained about Δ_k. Conversely, if $m_k = 0$, $\lambda_k = 0$ and $E(\Delta_k \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) = \hat{\eta}_1$ with cluster k contributing nothing to $\hat{\eta}_1$.

In (2.8), since λ_k is a monotone decreasing function of τ, there is more borrowing as τ increases. This is to be expected since large τ reflects a belief that the proportions of individuals possessing the characteristic of interest are similar over the clusters; by (2.3a) $\mathrm{corr}(\Delta_k, \Delta_\ell \mid \tau)$ is a monotone increasing function of τ. In particular, $\lim_{\tau\to\infty} \lambda_k = m_k/M_k$, the sampling fraction in cluster k. Thus, if $\lim_{\tau\to\infty} \hat{\eta}_1$ exists,

$\lim_{\tau\to\infty} E(\Delta_k \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) = M_k^{-1}\{m_k' + (M_k - m_k)\lim_{\tau\to\infty} \hat{\eta}_1\}$. If the ω_r are all equal and $a_r = r$,

$$\lim_{\tau\to\infty} \hat{\eta}_1 = \left\{\sum_{k \in s} m_k' + 1\right\}\left\{\sum_{k \in s} m_k + 2\right\}^{-1};$$

see Albert and Gupta (1983).

Note that (2.8) and (2.9) have the same forms as formulas (2) and (3) in Scott and Smith (1969).

Defining $\varphi = 1 - \sum_{k \in s} \rho_k \lambda_k$, the posterior mean of P is

$$E(P \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) = \varphi\hat{\eta}_1 + (1 - \varphi)\hat{P} \qquad (2.10)$$

where $\hat{P} = \left\{\sum_{k \in s} \rho_k \lambda_k\right\}^{-1} \sum_{k \in s} \rho_k \lambda_k (m_k'/m_k)$. Thus the posterior mean of P is a weighted average of $\hat{\eta}_1$, the posterior expected value of the prior mean of P, and $\hat{P}$, a weighted average of the sample proportions within each cluster.

Finally, the posterior variance of P can be shown to be given by

$$\mathrm{var}(P \mid \underset{\sim}{m}, \underset{\sim}{m}', \tau) = \hat{\eta}_2 \sum_{k=1}^{N} \sum_{\ell=1}^{N} \rho_k \rho_\ell (1-\lambda_\ell)(1-\lambda_k)$$

$$+ \sum_{k=1}^{N} \rho_k^2 (1-\lambda_k)^2 \left[\tau^{-1} + M_k^{-1}\right] \hat{\nu}_k^2. \qquad (2.11)$$

3. Application to National Health Interview Survey Data

To illustrate the methodology described in Section 2, we consider the NHIS binary variable where $Y = 1$ if, and only if, the individual has seen a doctor at least once during the past year. The first stage sampling units are counties while the second stage units are individuals. As described in Section 1, we assign each county to a stratum and each individual to a domain, and assume that (1.1), (1.3) and (1.4) hold within each domain x stratum. National or sub–national estimates are obtained by summing the estimated totals from each domain x stratum. For this illustration, we make inference about P, the proportion of individuals who have made at least one doctor visit, for females 20 or older and males 70 or older, and one stratum consisting of counties identified as a cluster by a cluster analysis using the three variables PCPOV, MDPOP and PCINC. Here, PCPOV is the percent of the population below the poverty level, MDPOP is the number of physicians per 1,000 population and PCINC is per capita income. We simplify the analysis by treating as our population the sampled individuals in the 74 counties in the 1987 NHIS who are members of the domain x stratum just described. Our sample has 20 (of the 74) counties, and 1172 individuals, selected as a proportional sample of individuals from each of the sampled counties. The second and third columns of Table 1 have the values of m_k and $\hat{\Delta}_k = m_k'/m_k$ for the 20 counties.

There are very small changes over time in the values of the Δ_k. Thus, we used estimates of the Δ_k from the 1985/1986 NHIS for the 74 counties to assign values to the ω_r in (1.4). Defining $b_r = a_r/\tau$, we classified the 74 estimated Δ_k's into nine intervals with midpoints $b_1 = 0.63$, $b_2 = 0.67$, $b_3 = 0.73$, $b_4 = 0.76$, $b_5 = 0.79$, $b_6 = 0.81$, $b_7 = 0.84$, $b_8 = 0.87$ and $b_9 = 0.89$: ω_r is the proportion of counties in the 1985/1986 NHIS with estimated Δ_k in the interval with midpoint b_r. We took a_r as the rounded value of $b_r\tau$. Before selecting a single value for τ, we considered a range of values for τ. Approximating λ_k by $\lambda_k^* = \{1 + (\tau/m_k)\}^{-1}$, we took $0.10 \le \lambda_k^* \le 0.80$. Replacing m_k by its median value, 36, we have $9 \le \tau \le 324$. Rounding, we considered τ = 10, 20, 30,50, 100, 200 and 400. It is clear from the results in Table 2 that the prior and posterior expected values of P are affected minimally by the choice of τ. While the value of τ has a greater effect on the corresponding standard errors (SE's), the effect is small over large ranges of τ. Since we believe that the past data used to determine the ω_r are almost as good as the current data, we take $\tau = 30$ which is slightly smaller than the median of the observed m_k. To provide a contrast we also consider in the ensuing analysis a much larger value of τ i.e. $\tau = 200$.

We first consider the data based prior distribution for β/τ described in the preceding paragraph, and presented in the first two columns of Table 3a. Since there are small changes over time, it is not surprising that there are only small differences between the prior and posterior expected values of P, the proportion of individuals having at least one doctor visit (Table 2a). Changing the value of τ has little effect on these expected values. The posterior standard errors of P are small, ranging from 23% to 37% of the value of the prior standard error of P(Table 2a). We obtained results similar to those in Table 2 by doubling and halving the sample sizes (m_k) in Table 1. From Table 3a, it is clear that our prior distribution of β/τ is quite diffuse. However, even for a relatively small value of τ (e.g., $\tau = 30$) the posterior distribution of β/τ has substantially less variability.

Table 1. Sample size, sample proportion and posterior means of population proportion for each of twenty counties; τ = 30, 200.

			$E(\Delta_k \mid \underline{m}, \underline{m}', \tau)$	
County	m_k	$\hat{\Delta}_k$	τ=30	τ=200
1	119	.8319	.8259	.8176
2	63	.9048	.8716	.8320
3	33	.8485	.8263	.8146
4	97	.8557	.8430	.8243
5	31	.8065	.8042	.8087
6	21	.8095	.8050	.8091
7	78	.8974	.8709	.8338
8	30	.7667	.7843	.8035
9	48	.7708	.7828	.8016
10	126	.8095	.8081	.8092
11	17	.8824	.8310	.8148
12	79	.8228	.8170	.8129
13	58	.8966	.8643	.8287
14	19	.8947	.8379	.8165
15	151	.8411	.8346	.8228
16	31	.7097	.7550	.7957
17	17	.5294	.7033	.7871
18	36	.6667	.7281	.7873
19	98	.6327	.6723	.7510
20	20	.7000	.7611	.7991

NOTE: $\hat{\Delta}_k = m_k'/m_k$ and $E(\Delta_k \mid \underline{m}, \underline{m}', \tau)$ is given by (2.8).

In some applications, estimates of the individual Δ_k are needed. It is clear from Table 1 that use of $E(\Delta_k \mid \underline{m}, \underline{m}', \tau)$ rather than $\hat{\Delta}_k$ will be beneficial if there are ostensibly outlying values such as the one for county 17. In other cases the small, but non–negligible shrinkage (e.g., for counties 14, 18, 20 with $\tau = 30$) seems sensible in view of our prior knowledge of homogeneity of the counties within each stratum.

We have also considered the effect on inference about P of assigning a prior distribution for β/τ that is badly misspecified. (Compare the misspecified prior distribution in the first two columns of Table 3b with the data based prior distribution in Table 3a.) For each value of τ, the posterior distribution of β/τ, ω_r^*, corresponding to the misspecified prior distribution (Table 3b) is much closer to ω_r^* corresponding to the data–based prior distribution (Table 3a) than the

Table 2. Prior (single prime) and posterior (double prime) moments of P, the proportion of individuals having at least one doctor visit.

τ	E′(P)	SE′(P)	E″(P)	SE″(P)
		a. Data based prior distribution		
10	.8070	.0655	.8012	.0243
20	.8092	.0552	.7996	.0226
30	.8085	.0542	.8032	.0188
100	.8117	.0545	.8071	.0138
200	.8086	.0533	.8093	.0122
		b. Misspecified prior distribution		
10	.7195	.0753	.7916	.0364
20	.7134	.0716	.7814	.0292
30	.7154	.0677	.7929	.0220
100	.7159	.0665	.7981	.0159
200	.7149	.0642	.8026	.0141
		c. Uniform prior distribution		
10	.7778	.0947	.7998	.0274
20	.7778	.0839	.7964	.0247
30	.7778	.0844	.8006	.0192
100	.7789	.0849	.8039	.0145
200	.7761	.0836	.8066	.0129

misspecified prior distribution of β/τ, ω_r (Table 3b), is to the data–based prior distribution (Table 3a). [For example, with $\tau = 30$ compare (a) the third columns of Tables 3a and 3b with (b) the second columns of Table 3a and 3b.].

For each value of τ and for the misspecified prior distribution, the prior expected value of P is about 0.71 and the posterior expected value of P is about 0.79 (Table 2b). Thus, even with a prior distribution for β/τ that is badly misspecified, the posterior distribution of P has approximately the correct location. Moreover, note in both Table 2a and 2b that the effect on E″(P) of changing τ is negligible.

Finally, in the absence of significant prior information, one might assign to β/τ the uniform prior distribution given in the first two columns of Table 3c. Comparing the data based and uniform prior distributions, the values of the prior expected value of P, E′(P), are similar for each τ (Tables 2a, 2c). Consequently, it is not surprising that for each τ the values of the posterior expected value of P, E″(P), are close for these two prior distributions. Moreover, as expected, SE′(P) is larger for the uniform than for the misspecified prior distribution while SE″(P) is smaller for the uniform than for the misspecified prior distribution (Tables 2b, 2c). The posterior SE's for the data based and uniform prior distributions (Tables 2a, 2c) are surprisingly close. Finally, for each τ, the posterior distribution of β/τ corresponding to the uniform prior distribution is closer to the one corresponding to the data based prior than to the misspecified prior (Table 3).

Table 3. Prior and posterior distributions of β/τ: $\Pr(\beta/\tau = b_r)$; $\tau = 30, 200$.

		ω_r^*	
b_r	ω_r	τ=30	τ=200
a. Data based prior distribution			
.63	.0141	.0000	.0000
.67	.0282	.0000	.0000
.73	.0845	.0020	.0000
.76	.0986	.0730	.0034
.79	.1127	.2757	.1673
.81	.2113	.5169	.7282
.84	.2394	.1317	.1011
.87	.1690	.0007	.0000
.89	.0423	.0000	.0000
b. Misspecified prior distribution			
.63	.1220	.0003	.0000
.67	.3659	.0009	.0000
.73	.2439	.0196	.0000
.76	.1220	.3126	.0248
.79	.0488	.4131	.4235
.81	.0244	.2066	.4915
.84	.0244	.0465	.0602
.87	.0244	.0004	.0000
.89	.0244	.0001	.0000
c. Uniform prior distribution			
.63	.1111	.0001	.0000
.67	.1111	.0001	.0000
.73	.1111	.0037	.0000
.76	.1111	.1192	.0064
.79	.1111	.3938	.2756
.81	.1111	.3938	.6396
.84	.1111	.0885	.0783
.87	.1111	.0007	.0000
.89	.1111	.0001	.0000

NOTE: $b_r = a_r/\tau$, ω_r is given by (1.4) and ω_r^* by (2.5).

4. Summary and Extensions

The probabilistic specification in (1.1), (1.3) and (1.4) provides the basis for Bayesian predictive inference for a finite population proportion when there is binary data from a two stage cluster sample. Formulas (1.3) and (1.4) provide a flexible prior distribution, and permit a fully Bayes analysis. This obviates the need to use empirical Bayes methods which are often unsatisfactory because of understated variability. While our methodology is directed to small–to–moderate sized surveys, we suggest in Section 1 how one may try to extend these procedures for use in larger surveys.

One may extend the probabilistic specification in (1.1), (1.3) and (1.4) to accommodate some three–stage sample surveys. For example, let Y_{jki} denote the Bernoulli random variable with

$$\Pr(Y_{jki} = 1 | \theta_{jk}) = \theta_{jk},$$

and the $Y_{jki}(i = 1,\ldots,M_{jk},\ k = 1,\ldots,K_j,\ j = 1,\ldots,J)$ are assumed to be independent. Given β_j and τ, we take $\theta_{j1},\ldots,\theta_{jK_j}$ to be distributed independently with density given by (1.3).

For $\beta_j \in \{a_r : 0 < a_1 < \ldots < a_R < \tau\}$,

$$\Pr(\beta_j = a_r | \{\omega_{rj}\}) = \omega_{rj}$$

where $\sum_{r=1}^{R} \omega_{rj} = 1$, and the β_j are assumed to be independent. Finally, $\underset{\sim}{\omega}_1,\ldots,\underset{\sim}{\omega}_J$ are assumed to be a random sample from the Dirichlet distribution with parameter $\underset{\sim}{\alpha}$. Here $\underset{\sim}{\omega}_j = (\omega_{1j},\ldots,\omega_{Rj})$.

While one cannot obtain analytical expressions for the posterior moments of the finite population proportion, P, sampling–based methods such as the Gibbs sampler (see, for example, Gelfand and Smith (1990)) can be used to carry out Bayesian predictive inference for P.

References

Albert, J.H. and Gupta, A.K. (1983). Estimation in Contingency Tables Using Prior Information. Jour. Roy. Stat. Soc. B45, 60–69.

Dalal, S.R. and Hall, W.J. (1983). Approximating Priors by Mixtures of Natural Conjugate Priors. Jour. Roy. Stat. Soc., B45, 278–286.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling–Based Approaches to Calculating Marginal Densities. Jour. Amer. Stat. Assoc., 85, 398–409.

Lehoczky, J.P. and Schervish, M.J. (1987). Hierarchical Modelling and Multi–level Analysis Applied to the National Crime Survey. Paper presented at the Workshop on the National Crime Survey, July 6–17, 1987.

Leonard, T. (1972). Bayesian Methods for Binomial Data. Biometrika 59, 581–589.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes Estimates for the Linear Model (with Discussion). Jour. Roy. Stat. Soc. B34, 1–41.

Novick, M.R., Lewis, C. and Jackson, P.H. (1973). The Estimation of Proportions in m Groups. Psychometrika 38, 19–46.

Scott, A. and Smith, T.M.F. (1969). Estimation in Multi–Stage Surveys. Jour. Amer. Stat. Assoc. 64, 830–840.