A HIERARCHICAL BAYESIAN APPROACH FOR ANALYZING A POLYCHOTOMOUS RESPONSE FROM A CLUSTER SAMPLE

Michael E. Schuckers, Iowa State University, Hal S. Stern, Iowa State University
Michael E. Schuckers,Department of Statistics, Iowa State University Ames, Iowa 50011

Key Words: Dirichlet-Multinomial, finite population.

1. INTRODUCTION

Cluster sampling is a commonly used technique in finite population sampling, primarily due to its convenience. For example, choosing clusters that represent a physical location requires less effort and less time for data collection than a simple random sample that would require many more physical locations. The work in this paper is motivated by an analysis of a Slovenian cluster-design survey by Rubin, Stern and Vehovar(1995). For simplicity they ignored the clustering design and treated the sample as a simple random sample (SRS). The goal of this paper is to provide a Bayesian hierarchical model to analyze observations from cluster sampling design where the measured response is polychotomous.

A hierarchical Bayesian approach to finite population sampling with a Gaussian response was described by Scott and Smith(1969). Similar work has been done on responses that are transformable to normality by Malec and Sedransk(1985) and on binomial responses by Stasny(1989) and (1991). These authors considered Bayesian hierarchical models. A recent non-hierarchical approach is Ghosh and Meeden's (1997) inference from finite population sampling from the Bayesian bootstrap perspective. Frequentist analyses of data from cluster samples can be found in such classical works as Cochran(1977) and Kish(1965).

Section 2 of this paper describes a probability model for cluster samples with polychotomous responses. Section 3 considers how to make inferences from this model. In Section 4 we present an application of this model to an actual data set. Finally we present a few comments and some suggestions for future work in Section 5.

2. THE MODEL

In this section, we present a general model for analyzing polychotomous data from cluster sampling. Here, we will refer to the ultimate sampling unit as individuals since our motivating example is a public opinion survey. The model is, clearly, not limited to that situation. Suppose that we sample J clusters from a population of $J+K$ clusters and within each sampled cluster a simple random sample is performed and n_j of N_j individuals in the cluster are measured. Each of the n_j responses will be in one of I categories. We let Y_{ij} be the number of individuals with response i; $i = 1, 2, \ldots, I$, that come from cluster j. Also let $\underline{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{Ij})^T$ and $\underline{Y} = (\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_J)^T$. Also, let θ_{ij} be the probability that an individual in the jth cluster yields response i with $\underline{\theta}_j = (\theta_{1j}, \theta_{2j}, \ldots, \theta_{Ij})^T$ and $\underline{\Theta} = (\underline{\theta}_1, \underline{\theta}_2, \ldots, \underline{\theta}_J)^T$.

We propose the following Bayesian superpopulation model. The data $\underline{Y}_j$ are multinomial,

$$\underline{Y}_j \mid \underline{\theta}_j, n_j \sim \text{Multinomial}(n_j, \underline{\theta}_j).$$

The cluster level probability vectors are modeled as independent draws from a Dirichlet,

$$\underline{\theta}_j \mid \underline{\alpha} \sim \text{Dirichlet}(\underline{\alpha}),$$

where $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_I)^T$ represents the parameters describing the population of θ_{ij}'s with $E[\theta_{ij}] = \alpha_i / \Sigma \alpha_m$. Finally the prior distribution for $\underline{\alpha}$ is the improper distribution,

$$p(\underline{\alpha}) \propto \left(\sum_{i=1}^{I} \alpha_i \right)^{-\frac{2I+1}{2}} I_{\{\alpha_i > 0, \forall i\}}.$$

This choice of prior is uniform on the simplex form by the ratios of the α_i to $\Sigma \alpha_i$ and $\Sigma \alpha_i^{-1/2}$. For an example with the Beta-binomial distribution consult Gelman et al.(1995).

Here we are treating the clusters as independently drawn since they are chosen via SRS. We can express the joint distribution of $\underline{Y}$, $\underline{\Theta}$, and $\underline{\alpha}$, as follows:

$$p(\underline{\Theta}, \underline{\alpha}, \underline{Y}) = \left\{ \prod_{j=1}^{J} p(\underline{Y}_j \mid \underline{\theta}_j) p(\underline{\theta}_j \mid \underline{\alpha}) \right\} p(\underline{\alpha}).$$

In the Bayesian nomenclature, the distributions of the $\underline{Y}_j$'s are referred to as sampling distributions. The distributions of the $\underline{\theta}_j$'s and $\underline{\alpha}$ are known as the prior distributions and the hyperprior distribution, respectively. The hyperprior distribution on $\underline{\alpha}$ is called a "diffuse" or "vague" prior, meaning that the density places mass somewhat evenly throughout the entire parameter. The advantage of this type of prior or hyperprior is that it allows the data to shape the posterior and, consequently, to shape the inference to be made. Here, the hyperprior is clearly improper;

however, it yields a proper marginal posterior $p(\underline{\alpha}|\underline{Y})$. A sketch of a proof of this is given in the Appendix.

3. POSTERIOR INFERENCE

From the Bayesian perspective, the posterior distribution $p(\underline{\Theta},\underline{\alpha}|Y)$ describes the uncertainty in the parameters after observing the data. In this case we can factor the posterior into two pieces:

$$p(\underline{\Theta},\underline{\alpha}|\underline{Y}) = p(\underline{\Theta}|\underline{\alpha},\underline{Y})\, p(\underline{\alpha}|\underline{Y}).$$

The first piece, the conditional posterior of $\underline{\Theta}|\underline{\alpha},\underline{Y}$, is simply a product of Dirichlet distributions with parameters $\underline{Y}_j + \underline{\alpha}$. The second piece, the marginal posterior of $\underline{\alpha}|\underline{Y}$, is the product of a Dirichlet-Multinomial(DM) distribution and the hyperprior distribution on $\underline{\alpha}$. The resulting posterior distribution is not recognizable. Consequently, we use a Markov Chain Monte Carlo (MCMC) procedure to obtain samples from the posterior.

Supposing that we have obtained M samples from the posterior distribution of $\underline{\alpha}$, we can use these samples to make our inferences. Then we can estimate any function, $\varphi(\underline{\Theta})$, of the parameters using samples from its posterior. For each of the M samples from the posterior, we calculate $\varphi(\underline{\Theta})$. This yields $\varphi^{(m)}$, realizations from the posterior distribution of φ. From these realizations of φ, any inferential measures can be calculated, e.g. posterior quantiles, posterior means.

It is important to point out that a variety of choices of φ are possible. For example,

$$\sum_{j=1}^{J+K} n_j z_j \left(\sum_{j=1}^{J+K} n_j \right)^{-1}$$

where $z_j \sim$ Dirichlet$(Y_j + a)$ for $j=1, \ldots, J$ and $z_j \sim$ Dirichlet(a) for $j=J+1, \ldots, J+K$, is the superpopulation estimate of the I cell proportions in the total population. For this paper, we are interested in understanding the relationship of the Bayesian hierarchical model to the traditional inferential methods. Therefore, we construct an estimand that allows us to recreate the traditional interval inference for cluster samples.

Specifically, we take φ to be the proportion of responses for J randomly chosen clusters that occur in a particular cell or cells.

Here, let $\varphi^{(m)} = \dfrac{1}{J} \sum_{n=1}^{J} \underline{\lambda}^T \underline{\tilde{\theta}}_n^{(m)}$, (1)

where $\underline{\tilde{\theta}}_n^{(m)}$ is a sample from a Dirichlet$(Y_j + \alpha^{(m)})$ with probability $1/(J+K)$ for each sample cluster, j, and is a sample from a Dirichlet$(\alpha^{(m)})$ with probability $K/(J+K)$. λ is a column vector of zeroes

and ones indicating which cells are of interest. This φ attempts to capture the sample to sample variability that characterizes the classical inferential methods.

4. AN APPLICATION: THE SLOVENIAN PLEBISCITE

On October 8, 1991 Slovenia became an independent republic. Ten months earlier a plebiscite was held, and the Slovenian citizens overwhelming voted for independence. In the month proceeding the plebiscite, the annual Slovenian Public Opinion (SPO) Survey was conducted and incorporated into the survey were the following questions:

1. Are you in favor of Slovenian independence?
2. Are you in favor of Slovenia's secession from Yugoslavia?
3. Will you attend the plebiscite?

The last question was especially relevant since an individual not attending the plebiscite would be counted as a no vote. The results recorded were YES, NO and Don't Know(DK) for each question. For the purposes of this paper, we will ignore the DK's, so that each question will have only two responses. The response of interest for this paper will be the $2^3=8$ possible responses to the three questions above. That is, we will consider the I=8 possible combinations of answers to the above questions as our response.

The SPO was carried out via a three-stage design. 139 of 1000 PSU's were sampled and then three secondary sampling units(SSU) were sampled; finally five individuals within each SSU were sampled. For this paper, we will ignore the SSU information and treat the data as if an SRS of size 15 was performed within each cluster. Like most large scale surveys, not all selected individuals responded. However, 2074 of the 2085 did respond (including some substitutes). Eliminating those individuals who chose DK as one of their responses to the three questions that concerned the plebiscite further reduces this number. Consequently, data from 1454 individuals were analyzed. This data set was originally analyzed ignoring the clustering in Rubin, Stern and Vehovar(1994). Their emphasis was on treating the DK responses. Additional information concerning the Slovenian plebiscite and the SPO can be found there.

The primary measure of interest for this paper is the proportion of individuals in the population who will attend and who are in favor of independence. Following the procedure outlined in section 3, we took M=15000 samples from the posterior of $\underline{\alpha}$. From those samples we found that a

95% central posterior predictive interval for the proportion described in the previous paragraph runs from 0.9054 to 0.9355. Table 1 contains the results from the Bayesian analysis along with the results from the classical analysis following Cochran(1997). Both intervals are approximately the same length, 0.0301 versus 0.0308. There is a slight discrepancy in the center of each interval, but overall there is good agreement between the intervals. As expected for our choice of φ, the results from the Bayesian hierarchical analysis closely mirror those from this classical cluster analysis.

Table 1. Posterior estimates of the proportion who intend to attend the plebiscite and to vote for independence.

Percentiles	2.5^{th}	50th	97.5^{th}
Posterior	0.9054	0.9216	0.9355
Classical Cluster	0.9124	0.9278	0.9432

5. COMMENTS AND FUTURE WORK

We have developed a Bayesian hierarchical model for cluster samples with a polychotomous response. This probability model treats each cluster as a random sample from a multinomial distribution, with each cluster having its own probabilities. In turn, those probabilities are modeled as coming from a Dirichlet superpopulation. The parameters of this population are then given a "diffuse" improper distribution. One important aspect of this model is that inference can be made on cells for which no responses were recorded

Also, in Section 4, we discarded individuals whose response to any of the three questions of interest included a DK. A more preferable analysis would be to utilize modern missing-data methods to more accurately assess the affect of the DK's on inference of the kind presented in this paper. This was done in the analysis of Rubin, Stern and Vehovar(1995).

REFERENCES

Brier, S. (1980) "Analysis of contingency tables under cluster sampling". Biometrika, 67, 591-596.

Cochran, W. G. (1977) Sampling Techniques (3rd ed.), New York: John Wiley and Sons.

Ericson, W. A.(1969) "Subjective Bayesian Models in Sampling Finite Populations(with Discussion)". Journal of the Royal Statistical Society (B), 31, 195-233.

Gelman A. and D. Rubin(1992), "Inference from Interactive Simulation using Multiple Chains(with Discussion)". Statistical Science, 7, 457-511.

Gelman, A, J. Carlin, H. Stern and D. Rubin(1995), Bayesian Data Analysis New York: Chapman-Hall.

Gelman, A., G. Roberts and W. Gilks,(1995) "Efficient Metropolis Jumping Rules" In Bayesian Statistics V, ed. J. Bernardo et al. Oxford: Oxford University Press, 599-607.

Ghosh M. and G. Meeden(1997) Bayesian Methods in Finite Population Sampling New York :Chapman-Hall.

Kish, L.(1965) Sampling Techniques. New York: John Wiley and Sons.

Malec, D. and J. Sedransk(1985). "Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling". Journal of the American Statistical Association, 80, 897-902.

Rubin, D, H. Stern and V. Vehovar(1995), "Handling "Don't Know" Survey Responses: The case of the Slovenian Plebiscite". Journal of the American Statistical Association, 90, 822-828

Scott, A. and T. M. F. Smith.(1969) "Estimation in Multi-Stage Surveys". Journal of the American Statistical Association, 64, 830-840

Stasny, E.(1989) "Hierarchical Modeling for Non-response and Victimizations in the National Crime Survey". ASA Proceedings of the Survey Methodology Section, 446-451.

Stasny, E. (1991)"Hierarchical Models for the Probability of a Survey Classification and Nonresponse: An Example From the National Crime Survey", Journal of the American Statistical Association, 86, 1991, 296-303.

We need to show that $p(\alpha|y) \propto p(y|\alpha)p(\alpha)$ is integrable. That is, we must show that

$$\iint \ldots \int p(y|\alpha)p(\alpha)d\alpha_1 \ldots d\alpha_I < \infty$$

or

$$\iint \ldots \int \prod_{j=1}^{J}\left[\frac{\Gamma\left(\sum_{i=1}^{I}\alpha_i\right)}{\Gamma\left(n_j + \sum_{i=1}^{I}\alpha_i\right)}\prod_{i=1}^{I}\frac{\Gamma(y_{ij}+\alpha_i)}{\Gamma(\alpha_i)}\right]\left(\sum_{i=1}^{I}\alpha_i\right)^{-\left(\frac{2I+1}{2}\right)}d\alpha_1 \cdots d\alpha_I < \infty$$

There are two cases that need to be considered.

1. $\alpha_i \to 0 \ \forall i$
2. $\alpha_i \to \infty \ \forall i$

Case 1: Small alpha's

Note : For $\varepsilon > 0$,

$$\Gamma(1+\varepsilon) = \varepsilon\Gamma(\varepsilon) \Rightarrow \Gamma(\varepsilon) = \frac{\Gamma(1+\varepsilon)}{\varepsilon} \approx \frac{1}{\varepsilon}$$

For suitably small ε,

$$\int_0^{\varepsilon} \cdots \int_0^{\varepsilon} p(y|\underline{\alpha})p(\underline{\alpha})d\underline{\alpha} =$$

$$\int_0^{\varepsilon}\cdots\int_0^{\varepsilon}\prod_{j=1}^{J}\left[\frac{\Gamma\left(\sum_{i=1}^{I}\alpha_i\right)}{\Gamma(n_j+\kappa)_j}\prod_{i=1}^{I}\frac{\Gamma(y_{ij}+\alpha_i)}{\Gamma(\alpha_i)}\right]\left(\sum_{=1}^{I}\alpha_i\right)^{-\left(\frac{2I+1}{2}\right)}d\underline{\alpha}$$

$$\cong k(y,n)\int_0^{\varepsilon}\ldots\int_0^{\varepsilon}\alpha_1^{L_1}\alpha_2^{L_2}\cdots\alpha_I^{L_I}\left(\sum_{i=1}^{I}\alpha_i\right)^{-J}\left(\sum_{i=1}^{I}\alpha_i\right)^{-\left(\frac{2I+1}{2}\right)}d\underline{\alpha}$$

where $L_i = \sum_{j=1}^{J}I(Y_{ij} > 0)$

Consider the case where I=3 for simplicity:

$$\Psi = \int\limits_{0}^{\varepsilon}\int\limits_{0}^{\varepsilon}\int\limits_{0}^{\varepsilon} \alpha_3^{L_3} \alpha_2^{L_2} \alpha_1^{L_1} \left(\sum_{i=1}^{3}\alpha_i\right)^{-J}\left(\sum_{i=1}^{3}\alpha_i\right)^{-\left(\frac{7}{2}\right)} d\underline{\alpha}$$

Integration by parts yields

$$\Psi = \int\limits_{0}^{\varepsilon}\int\limits_{0}^{\varepsilon} \alpha_3^{L_3} \alpha_2^{L_2} \left[g_1(\varepsilon + \alpha_2 + \alpha_3) + (\alpha_2 + \alpha_3)^{-\left(\frac{7}{2}\right)-J+(L_1+1)} \right] d\alpha_2 d\alpha_3$$

Now we can ignore g_1 since it will integrate to a finite quantity. Repeat integration by parts, until

$$\int\limits_{0}^{\varepsilon} \alpha_3^{L_3} \left[g_2(\varepsilon + \alpha_3) + \alpha_3^{-\left(\frac{7}{2}\right)-J+L_1+L_2+2} \right] d\alpha_3$$

so $\Psi < \infty$ as long as $L_1 + L_2 + L_3 > J + 1/2$.

In the general I dimensional case, the posterior is proper as long as $\sum\limits_{i=1}^{I} L_i > J + 1/2$. That is, in order for the posterior to be proper there must be at least one cluster that has observed values in at least two cells.

Case 2: Large alpha's

Choose large M, such that each n_j is small relative to M.

- Replace Gamma functions with products.
- Integrate by parts.
- Repeat this for each alpha.

$$\int_{M}^{\infty}\cdots\int_{M}^{\infty} \prod_{j=1}^{J}\left[\frac{\Gamma\left(\sum\limits_{i=1}^{I}\alpha_i\right)}{\Gamma\left(n_j + \kappa\right)} \prod_{i=1}^{I} \frac{\Gamma\left(y_{ij}+\alpha_i\right)}{\Gamma(\alpha_i)} \right]\left(\sum_{i=1}^{I}\alpha_i\right)^{-\left(\frac{2I+1}{2}\right)} d\underline{\alpha}$$

$$Z = \int_{M}^{\infty}\cdots\int_{M}^{\infty} \prod_{j=1}^{J}\left[\frac{1}{\prod\limits_{k=0}^{n_j-1}(k+\kappa)} \prod_{i=1}^{I}\prod_{k=0}^{Y_{ij}-1}(k+\alpha_i) \right]\left(\sum_{i=1}^{I}\alpha_i\right)^{-\left(\frac{2I+1}{2}\right)} d\underline{\alpha}$$

$$Z = \int_{M}^{\infty}\left(\alpha_I^{Y_{Ij}}\right)\left[h_{I-1}(M + \alpha_I) + \left(\alpha_I\right)^{-\left(\frac{2I+1}{2}\right)-\sum\limits_{j=1}^{J}n_j + \sum\limits_{i=1}^{I-1}\sum\limits_{j=1}^{J}Y_{1j}+(I-1)} \right] d\alpha_I$$

$$Z = \int_{M}^{\infty}\alpha_I^{-\frac{3}{2}} d\alpha_I < \infty$$

Consequently, the posterior is proper.