BAYESIAN ESTIMATION OF A PROPORTION UNDER NONIGNORABLE NONRESPONSE

Balgobin Nandram
Department of Mathematical Sciences, Worcester Polytechnic Institute

Jai Won Choi

National Center for Health Statistics

Key words: Beta-Binomial model; Dirichlet process prior; Exchangeability; Griddy Gibbs sampler; Latent variable; Selection approach.

Abstract:

We use a Dirichlet process prior (DPP) to restrict the pooling of nonresponse binary data from small areas which may seem to be similar. Our objective is to estimate the proportion of individuals with a particular characteristic from each of a number of areas under nonignorable nonresponse. All hyperparameters have proper prior densities. The griddy Gibbs sampler is used to perform the computation. For illustration, we use data on victimization in ten domains from the National Crime Survey (NCS). We show empirically that there could be difference in inference between the two nonignorable models.

Key words: Beta-Binomial model; Dirichlet process prior; Exchangeability; Griddy Gibbs sampler; Latent variable; Selection approach.

1. Introduction

Recently there has been much activity in the analysis of survey nonresponse. Indeed, the response rates in many surveys have been decreasing internationally (De Heer 1999 and Groves and Couper 1998). For many of these surveys the responses are binary. To permit a flexibility in robustness to the prior specifications, we study a nonparametric hierarchical Bayes model that can be used to study nonignorable nonresponse for binary data from many areas.

Stasny (1991) used a hierarchical Bayesian model to study victimization in the National Crime Survey (NCS). She used the Bayesian selection approach which was developed primarily to study sample selection problems (e.g., Heckman 1976 and Olson 1980). However, the Stasny Bayes empirical Bayes approach assumes that the hyper-parameters are fixed but unknown, and these parameters are estimated using maximum likelihood methods. This approach has been extended in several directions. See Nandram and Choi (2002 a, b) and Nandram, Han and Choi (2002) for full Bayesian analyses.

In small area estimation, it is usual to assume that the parameters indexing the areas share an effect. That is, the parameters follow a common probability density function. This assumption can lead to overshrinkage.

We believe that there are groups of areas whose members are more similar than others. Furthermore, we also believe that the compositions of these groups are unknown to us. A natural way to deal with this situation is to use the Dirichlet process prior (henceforth, DPP). Ferguson (1973, 1974); Escobar (1994) and Escobar and West (1995) have used the DPP to perform nonparametric Bayesian analysis on normal data. Also Kong, Liu, and Wong (1994) have used nonparametric Bayesian analysis for binomial data.

We use a nonparametric Bayesian method to analyze nonignorable nonresponse binary data. We start with the model proposed by Stasny (1991) to model nonresponse data with several areas. But, we assume a DPP for these parameters. Thus, in our model with a DPP, the Stasny's nonignorable model is our baseline model. However, unlike Forster and Smith (1998) and Nandram and Choi (2002 a, b), we do not express uncertainty about ignorability in this paper.

A related literature is on what is now known as uncertain pooling used primarily for pooling experiments. The experiments are partitioned and there can be many partitions. The experiments in each partition set are believed to be similar and the partition sets are different. There is uncertainty about which partition is the correct one. This methodology works well for a small number of experiments, but for problems with many experiments (or areas) it may be infeasible. Malec and Sedransk (1992), Evans and Sedransk (2001), Mallick and Walker (1997) and Consoni and Veronese (1995) discussed Bayesian methodology for combining results from several normal or binomial experiments.

We consider a nonparametric hierarchical

Bayesian model with a DPP to study the proportion of individuals possessing a characteristic in the presence of nonignorable nonresponse when there is uncertainty about the hyper-parameters. For illustration, we use data from the National Crime Survey (NCS) which we describe briefly in Section 2. In Section 3 we describe the hierarchical Bayesian model and how to analyze this model using Markov chain Monte Carlo (MCMC) methods. In Section 4 we present some empirical results.

2. National Crime Survey

We used the data created by Stasny (1991), who took a random start at the record for the eighth household (ordered on the original longitudinal file) in the full data set and then every fifteenth record after that. The data are poststratified into domains according to three neighborhood characteristics: (i) urban (U) and rural (R), (ii) central city (C), other incorporated place (I), and unincorporated or not a place (N), and (iii) low poverty level (L) (9% or fewer of families below poverty level) and high poverty level (H) (10% or more of families below poverty level). Since it is practically impossible for a rural area to be a central city, as observed by Stasny (1991), this poststratification results in ten domains. Let

$$y_{ij} = \begin{cases} 1, & \text{if household } j \text{ in area } i \text{ is victimized} \\ 0, & \text{if household } j \text{ in area } i \text{ is not victimized} \end{cases}$$

and

$$r_{ij} = \begin{cases} 1, & \text{if household } j \text{ in area } i \text{ is a respondent} \\ 0, & \text{if household } j \text{ in area } i \text{ is not a respondent} \end{cases}$$

$i = 1, \ldots, \ell,\ \ j = 1, \ldots, n_i$. Essentially our models start with the y_{ij} and r_{ij}. We define $y_i = \sum_{j=1}^{r_i} y_{ij}$ and $r_i = \sum_{j=1}^{n_i} r_{ij}$.

Throughout, y_i is the number of successes (i.e., households with crimes in the NCS), r_i is the number of respondents and n_i is the number of households sampled in the i^{th} domain (or area), $i = 1, \ldots, \ell$ where $\ell = 10$ domains. The nonresponse rate in these domains ranges from 9.4% to 16.9%, and one reason for nonresponse is that a woman may be embarrassed to report a rape committed by an attacker.

Stasny (1991) suggested that nonresponse does not occur at random with respect to victimization status (see also Stasny 1990 and Saphire 1984). For the analysis of this data set, Nandram and Choi (2002 a) made two key contributions (a) discern whether nonresponse is ignorable or not, and (b) introduce a new model in which the degree of ignorability may vary from one area to another. Here, our contribution is to perform a nonparametric Bayesian analyis of these data by providing a prior that makes our procedure more robust. This can help to reduce overshrinkage, a nuisance in small area estimation.

3. Hierarchical Bayes Nonresponse Models

In this section, we describe the baseline nonignorable nonresponse model and the nonparametric Bayesian model.

3.1 Baseline Nonresponse Model

Our baseline model is a nonignorable nonresponse model, and is given by

$$y_{ij} \mid p_i \overset{iid}{\sim} \text{Bernoulli } (p_i),\ \ j = 1, \ldots, n_i,\ \ i = 1, \ldots, \ell,$$

$$r_{ij} \mid y_{ij} = s,\ \pi_{is} \overset{iid}{\sim} \text{Bernoulli } (\pi_{is}),\quad s = 0, 1. \quad (1)$$

$$p_i \mid \mu_{21}, \tau_{21} \overset{iid}{\sim} \text{Beta } (\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21}), \quad (2)$$

$$\pi_{is} \mid \mu_{2,s+1}, \tau_{2,s+1} \overset{iid}{\sim} \text{Beta } (A, B), \quad (3)$$

where $A = \mu_{2,s+1}\tau_{2,s+1}$ and $B = (1 - \mu_{2,s+1})\tau_{2,s+1}$, s = 0, 1.

Assumptions (2) and (3) express similarity among the states. This similarity helps when the weakly identified parameters like the π_{i1} and π_{i2} are estimated. This is not very robust because it may encourage too much pooling. Therefore, to restrict the pooling one may use a more robust prior specification.

We complete the prior specification by taking $\mu_k, k = 1, 2, 3$ and $\tau_k, k = 1, 2, 3$ to be independent. Specifically,

$$\mu_1 \overset{iid}{\sim} U(0, 1)$$

and

$$\mu_2 \mid \mu_3 \sim U(\mu_3, 1) \text{ and } \mu_3 \sim U(0, 1). \quad (4)$$

The assumption that $\mu_2 \geq \mu_3$ is used to avoid a possible computational difficulty. In the NCS this is a reassonable assumption because it is known that households that are victimized tend not to respond to the survey. In other situations, with expert opinion this equality can be reverse. For the τ_k, we take

$$\tau_k \overset{iid}{\sim} S(1),\ \ k = 1, 2, 3$$

The notation $X \overset{iid}{\sim} S(a)$ means that $p(x) = a/(a + x)^2,\ x \geq 0$ and $a \geq 0$ is the shrinkage prior density.

3.2 Model with Dirichlet Process Prior

We maintain the structure in (1), but instead of the prior densities in (2) and (3), we use the DPP.

Letting $\boldsymbol{\theta_i}$ =$(p_i, \pi_{i0}, \pi_{i1})$, we assume that, given a cumulative distribution function, G say, that

$$\boldsymbol{\theta_i} \mid G \overset{iid}{\sim} G(\cdot).$$

To express the uncertainty about $G(\cdot)$, we assume that given α and $G_0(\cdot)$,

$$G(\cdot) \sim \text{Dirichlet}\left\{\alpha G_0(\cdot)\right\},$$

a Dirichlet procces defined by α, a positive real number, and $G_0(\cdot)$ the prior specification of $G(\cdot)$. In fact, $E(G(\theta)) = G_0(\cdot)$ for all θ and α is a precision parameter, determining the concentration of the prior distribution for $G(\cdot)$ around $G_0(\cdot)$. Here α is assumed unknown, and $G_0(\cdot)$ has a specified form with its parameters unknown, as in the baseline model.

A key feature of DPP is associated with the discretness of $G(\cdot)$ under the Dirichlet process assumption (Ferguson, 1973). In any sample, $\theta_i, i = 1, ..., \ell$, from $G(\cdot)$, there is a positive probability that some of these θ_i coincide. That is, there are k, $1 \le k \le \ell$, parameters that describe the ℓ areas. The structure is such that the posterior distribution will strongly support common values of individual parameters, θ_i and $\theta_i^{(\prime)}$ for data points (y_i, r_i) and $(y_i^{(\prime)}, r_i^{(\prime)}$ that are close. Thus, we can combine information locally in the sample space to estimate the local structure.

In our application, we take the prior $G_0(\cdot)$ for θ_i as

$$G_0(p_i, \pi_{i0}, \pi_{i1}) = G_{01}(p_i)G_{02}(\pi_{i0})G_{03}(\pi_{i1}), \qquad (5)$$

where the prior densities for p_i, π_{i0} and π_{i1}, which are specified as in the baseline, would be

$$p_i \mid \mu_{21}, \tau_{21} \overset{iid}{\sim} Beta\ (\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21}),$$

$$\pi_{is} \mid \mu_{2,s+2}, \tau_{2,s+2} \overset{iid}{\sim} \text{Beta}\ (A, B),$$

where $A = \mu_{2,s+2}\tau_{2,s+2}$ and $B = (1-\mu_{2,s+2})\tau_{2,s+2}$, s = 0,1; we complete the prior specification by taking $\mu_k, k = 1, 2, 3$ and $\tau_k, k = 1, 2, 3$ to be exactly same as in (4).

Let $\theta_{(i)} = (\theta_1, ..., \theta_{i\text{-}1}, \theta_{i+1}, ..., \theta_\ell)^{(\prime)}, \quad i = 1, ..., \ell$. That is, $\theta_{(i)}$ consists of all components except the i^{th} one. It is pertinent to describe the conditional posterior density of $\theta_i \mid \theta_{(i)}, r, y, \mu, \tau, \alpha$. First, we describe two important components of this distribution.

Under the baseline model, the likelihood function is

$$\boldsymbol{p(y_i, r_i \mid \theta_i) = \begin{pmatrix} n_i \\ r_i \end{pmatrix} \begin{pmatrix} r_i \\ y_i \end{pmatrix}}$$

$$\times(\pi_{i1}p_i)^{y_i}(\pi_{i0}(1-p_i))^{r_i - y_i}$$
$$\times((1-\pi_{i1})p_i + (1-\pi_{i0})(1-p_i))^{n_i - r_i},$$

$y_i = 0, ..., r_i$ and $r_i = y_i, y_{i+1}, ..., n_i$, with independence over $i,\ i = 1, ..., \ell$.

Since the number of victimizations for the nonrespondents is unknown, we denote it by the latent variable z_i, and the number of households with no victimizations among the nonrespondents is $n_i - r_i - z_i$. The z_i simplify the computations. Then, the augmented likelihood function is

$$p(y_i, r_i, z_i \mid \boldsymbol{\theta_i}) =$$

$$\begin{pmatrix} n_i \\ r_i \end{pmatrix} \begin{pmatrix} r_i \\ y_i \end{pmatrix} (\pi_{i1}p_i)^{y_i}(\pi_{i0}(1-p_i))^{r_i - y_i}$$
$$\times\{(1-\pi_{i1})p_i\}^{z_i}\{(1-\pi_{i0})(1-p_i)\}^{n_i - r_i - z_i},$$

$i,\ i = 1, ..., \ell$.

Then, marginalizing over θ_i, we have

$$A(y_i, r_i) = \begin{pmatrix} n_i \\ r_i \end{pmatrix} \begin{pmatrix} r_i \\ y_i \end{pmatrix} \sum_{z_i=0}^{n_i - r_i} \begin{pmatrix} n_i - r_i \\ z_i \end{pmatrix}$$
$$\times \frac{B(y_i + z_i + \mu_1\tau_1,\ r_i - y_i - z_i + (1-\mu_1)\tau_1)}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)}$$
$$\times \frac{B(r_i - y_i + \mu_2\tau_2,\ n_i - r_i - z_i + (1-\mu_2)\tau_2)}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}$$
$$\times \frac{B(y_i + \mu_3\tau_3,\ z_i + (1-\mu_3)\tau_3)}{B(\mu_3\tau_3, (1-\mu_3)\tau_3)},$$

$i = 1, ..., \ell$.

The second quantity is the posterior density of θ_i under the baseline model. It is easy to show that

$$f(p_i, \pi_{0i}, \pi_{1i} \mid \mathbf{y}, \mathbf{r}, \mu, \tau) \propto \sum_{z_i=0}^{n_i - r_i} \left\{ \begin{pmatrix} n_i - r_i \\ z_i \end{pmatrix} \right.$$
$$\times B(y_i + z_i + \mu_1\tau_1, r_i - y_i - z_i + (1-\mu_1)\tau_1)$$
$$\times B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)$$
$$\times B(y_i + \mu_3\tau_3, z_i + (1-\mu_3)\tau_3)$$
$$\times \frac{p_i^{y_i + z_i + \mu_1\tau_1 - 1}(1-p_i)^{r_i - y_i - z_i + (1-\mu_1)\tau_1 - 1}}{B(y_i + z_i + \mu_1\tau_1, r_i - y_i - z_i + (1-\mu_1)\tau_1)}$$
$$\times \frac{\pi_{i0}^{r_i - y_i + \mu_2\tau_2 - 1}(1-\pi_{i0})^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)}$$
$$\left. \times \frac{\pi_{i1}^{y_i + \mu_3\tau_3 - 1}(1-\pi_{i1})^{z_i + (1-\mu_3)\tau_3 - 1}}{B(y_i + \mu_3\tau_3, z_i + (1-\mu_3)\tau_3)} \right\},$$

$i = 1, ..., \ell$.

Then, using Theorem 1 of Escobar (1994) and expressing probability

$$Q_j = \frac{p(y_i, r_i \mid \theta_j)}{\alpha A(r_i, y_i) + \sum_{j=1, j\neq i}^{\ell} p(y_i, r_i \mid \theta_j)},$$

we have $\theta_i \mid \theta_{(i)}, \mathbf{r}, \mathbf{y}, \mu, \tau, \alpha =$

$$\begin{cases} \theta_j \quad \text{with} \quad Q_j, \quad i \neq j \\ \propto \pi_0(\theta_i \mid \mathbf{r}, \mathbf{y}, \mu, \tau, \alpha) \quad \text{with} \quad (1 - \sum_{j=1, j\neq i}^{\ell} Q_j). \end{cases}$$

Note that the original ℓ areas are replaced by at most ℓ areas (i.e., this conditional posterior density describes how the discreteness arises).

3.3 Computations for the Model with Dirichlet Process Prior

We use the griddy Gibbs sampler to obtain samples from the joint posterior density of all the parameters. See Tanner (1993) for a more elaborate pedagogy on the Griddy Gibbs Sampler. We can draw from $\pi_0(\boldsymbol{\theta_{(i)}} \mid \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\tau})$ as follows. We note that $\pi_0(\boldsymbol{\theta_{(i)}}, z_i \mid \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\tau}) = \pi_0(\boldsymbol{\theta_{(i)}} \mid z_i, \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\tau})$ $p(z_i \mid \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\tau})$.

Note that, given z_i, y_i, r_i, μ, τ, the parameters p_i, π_{10}, and π_{i1} are independent with

$$p_i \mid z_i, y_i, r_i \mu_1, \tau_1 \overset{ind}{\sim}$$
$$\text{Beta}(z_i + y_i + \mu_1\tau_1, r_i - z_i - y_i + (1-\mu_1)\tau_1),$$
$$\pi_{i0} \mid z_i, y_i, r_i, \mu_2, \tau_2 \overset{ind}{\sim}$$
$$\text{Beta }(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2),$$
$$\pi_{i1} \mid z_i, y_i, r_i, \mu_3, \tau_3 \overset{ind}{\sim}$$
$$\text{Beta}(y_i + \mu_3\tau_3, z_i + (1-\mu_3)\tau_3).$$

The posterior distribution of z_i is as follows. Letting

$$p(Z_i = z_i \mid \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\tau}) \propto \frac{\omega_{z_i}}{\sum_{z_i=0}^{n_i - r_i} \omega_{z_i}}$$

where

$$\begin{aligned} \omega_{z_i} &= \binom{n_i - r_i}{z_i} \\ &\times \quad B(z_i + y_i + \mu_1\tau_1,\ r_i - z_i - y_i + (1-\mu_1)\tau_1) \\ &\times \quad B(r_i - y_i + \mu_2\tau_2,\ n_i - r_i - z_i + (1-\mu_2)\tau_2) \\ &\times \quad B(y_i + \mu_3\tau_3, z_i + (1-\mu_3)\tau_3), \end{aligned}$$

$z_i = 0, \ldots, n_i - r_i$ and $i = 1, \ldots, \ell$. Thus, we draw z_i from $p(Z_i = z_i \mid \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\tau})$, and, with this z_i, we draw p_i, π_{i0}, and π_{i1}, independently.

The conditional posterior density of α is only related to k. Escobar and West (1995) show how to get samples from the conditional posterior density. We use grids to obtain samples from the conditional posterior density of (μ, τ), given $\boldsymbol{\theta}$. Let $\boldsymbol{\theta_1^*}$,..., $\boldsymbol{\theta_k^*}$ be the k distinct values, where $\boldsymbol{\theta_i^*}$=$(p_i^*, \pi_0^*, \pi_1^*)$. Then,

$$p(\mu, \tau \mid \theta^{(*)}, k) \propto \prod_{i=1}^{k} \left\{ \frac{p_i^{*(\mu_1\tau_1 - 1)}(1-p_i^*)^{(1-\mu_1)\tau_1 - 1}}{B(\mu_1\tau_1, \quad (1-\mu_1)\tau_1)} \right.$$
$$\times \frac{\pi_{i0}^{*\,(\mu_2\tau_2 - 1)}(1-\pi_{i0}^*)^{(1-\mu_2)\tau_2 - 1}}{B(\mu_2\tau_2, \quad (1-\mu_2)\tau_2)}$$
$$\left. \times \frac{\pi_{i1}^{*\,(\mu_3\tau_3 - 1)}(1-\pi_{i1}^*)^{(1-\mu_3)\tau_3 - 1}}{B(\mu_3\tau_3, \quad (1-\mu_3)\tau_3)} \right\}$$
$$\times \prod_{r=1}^{3} \frac{1}{(1+\tau_r)^2}.$$

For example, letting $a = \prod_{i=1}^{k} p_i^*$ and $b = \prod_{i=1}^{k}(1-p_i^*)$, for (μ_1, τ_1) the joint conditional posterior density is

$$p(\mu_1, \tau_1 \mid \boldsymbol{\theta^*}, k) \propto \frac{1}{(1+\tau_1)^2} \frac{a^{\mu_1\tau_1 - 1} b^{(1-\mu_1)\tau_1 - 1}}{B(\mu_1\tau_1,\ (1-\mu_1)\tau_1)}.$$

Then, for μ_1

$$p(\mu_1 \mid \tau_1, \boldsymbol{\theta^*}, k) \propto \frac{a^{(\mu_1\tau_1 - 1)} b^{(1-\mu_1)\tau_1 - 1}}{B(\mu_1\tau_1,\ (1-\mu_1)\tau_1)}.$$

For τ_1, we make the transformation $\tau_1 = \nu_1/(\mu_1 - \nu_1)$, with $0 < \nu_1 < \mu_1$, to have

$$p(\nu_1 \mid \mu_1, \boldsymbol{\theta^*}, k) \propto \frac{a^{\mu_1 \frac{\nu_1}{\mu_1 - \nu_1} - 1} b^{(1-\mu_1)\frac{\nu_1}{\mu_1 - \nu_1} - 1}}{B(\mu_1 \frac{\nu_1}{\mu_1 - \nu_1},\ (1-\mu_1)\frac{\nu_1}{\mu_1 - \nu_1})},$$

where $0 < \nu_1 < \mu_1$.

The conditional posterior density of $p(\alpha \mid k)$ is obtained as follows. Using results in Antoniak (1974), Escobar and West (1995) presented the probability mass function

$$p(k \mid \alpha) = s_n(k)\alpha^k \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)}, \quad k = 1, ..., n$$

where $s_n(k)$, not involving α, are the absolute values of the Sterling numbers of the first kind.

Letting D_n represent a configuration of the data into k groups, Escobar and West (1995) argued that

$$p(\alpha \mid k, \theta, D_n) = p(\alpha \mid k)$$

where clearly $p(\alpha \mid k) \propto p(\alpha)p(k \mid \alpha)$ and $p(\alpha)$ is the prior density for α. Escobar and West (1995) took $\alpha \sim G(a, b)$ where they specified $a = 2$ and $b = 4$ for the astronomy data studied by Roeder (1990).

Finally, introducing the latent variable γ, where

$$\gamma \mid \alpha, k \sim \text{Beta}(\alpha + 1, n),$$

they showed that

$$\alpha \mid \gamma, k \sim \lambda_{\gamma,k} G\{a + k, b - \log(\gamma)\} +$$
$$(1 - \lambda_{\gamma,k}) G\{a + k - 1, b - \log(\gamma)\}$$

where

$$\lambda_{\gamma,k} = \frac{a+k-1}{a+k-1+n(b-\log(\gamma))}.$$

From above, we have

$$p(\alpha \mid k) \propto p(\alpha)\alpha^k \frac{\Gamma(\alpha)}{\Gamma(\alpha+\ell)}, \quad 0 < \alpha < \infty.$$

We use the shrinkage prior

$$p(\alpha) \propto \mathrm{a}/(\mathrm{a}+\alpha)^2, \quad 0 < \alpha < \infty$$

where a is to be chosen. Then, the conditional posterior density for α is

$$p(\alpha \mid k) \propto \frac{1}{(\mathrm{a}+\alpha)^2}\alpha^k \frac{\Gamma(\alpha)}{\Gamma(\alpha+\ell)}, \quad 0 < \alpha < \infty.$$

Transforming α to $\rho = \alpha/(\alpha+1), \quad 0 < \rho < 1$, the conditional posterior density for α is

$$p(\alpha \mid k) \propto (1-\rho)^{-2}\{\rho/(1-\rho)\}^k\{\mathrm{a}+\rho/(1-\rho)\}^{-2}$$

$$\times \frac{\Gamma(\rho/(1-\rho))}{\Gamma(\ell+\rho/(1-\rho))}, \quad 0 < \rho < 1.$$

For both μ_1 and ν_1, our procedure is the same. Bounded intervals improve the grid method. We stratify the range into a large number of grids (e.g., 100) to approximate the probability density function by a probability mass function.

We have used the griddy Gibbs sampler to fit both the baseline and the DPP models. The baseline model was fit using the Metropolis-Hastings sampler in our previous work. We drew 11,000 iterates, threw out the first 1000, and took every tenth. This is very conservative because convergence is very rapid.

4. Numerical Results

We have used the NCS data to consider inference about p in Table 1, in which we have presented the posterior means (PM), the posterior standard deviations (PSD), the numerical standard errors (NSE) and the 95% credible intervals (CI) for the DPP model. The PMs are very similar for the two models except for RIH, RNL, and RNH. As is expected, the PSDs are similar larger for the DPP model. The NSEs are larger for the first seven domains than the last three which are somewhat smaller. Consequently, the 95% credible intervals for the first three domains are wider than the last three.

We have considered sensitivity to inference for the specification of a in the prior density for α. We have looked at five choices: $a = 0.001, 0.01, 1.00, 100, 1000$. Table 2 shows some sensitivity to inference about the p_i, but this is reasonably small.

As a summary, there are differences between the DPP and the baseline models. The posterior means for p, π_0 and π_1 are very similar (not shown because of space limit), but their posterior standard deviations under the DPP model are generally larger. Inference for p is not very sensitive to the choice of a.

Table 1: Comparison of the posterior means (PM), posterior standard deviations (PSD), numerical standard errors (NSE), and 95% credible intervals (CI) for p from the baseline and Dirichlet process prior (DPP) models

Domain	PM	PSD	NSE	CI
(a) Baseline Model				
UCL	0.269	0.036	0.017	(0.200, 0.329)
UCH	0.262	0.038	0.016	(0.190, 0.328)
UIL	0.273	0.037	0.018	(0.198, 0.332)
UIH	0.254	0.034	0.014	(0.187, 0.313)
UNL	0.295	0.046	0.019	(0.205, 0.374)
UNH	0.291	0.056	0.020	(0.186, 0.408)
RIL	0.269	0.058	0.019	(0.164, 0.389)
RIH	0.178	0.049	0.020	(0.087, 0.274)
RNL	0.168	0.034	0.013	(0.105, 0.234)
RNH	0.213	0.034	0.015	(0.150, 0.275)
(b) DPP Model				
UCL	0.274	0.045	0.041	(0.196, 0.327)
UCH	0.274	0.045	0.041	(0.196, 0.327)
UIL	0.275	0.045	0.041	(0.196, 0.329)
UIH	0.272	0.045	0.040	(0.186, 0.327)
UNL	0.280	0.049	0.041	(0.197, 0.365)
UNH	0.276	0.048	0.042	(0.196, 0.342)
RIL	0.270	0.049	0.039	(0.185, 0.331)
RIH	0.186	0.049	0.022	(0.084, 0.265)
RNL	0.183	0.042	0.020	(0.096, 0.248)
RNH	0.223	0.051	0.030	(0.126, 0.312)

NOTE: $p = Pr(y = 1 \mid p)$ where $y = 1$ for a victimized household and 0 otherwise.

Bibliography

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Nonparametric Problems", The Annals of Statistics, **2**, 1152-1174.

Consoni, G. and Veronese, P. (1995), "A Bayesian Method for Combining the Results from Several Binomial Experiments", Journal of the American Statistical Association, **90**, 935-944.

De Heer, W. (1999), "International response

Table 2: Comparison of 95% credible intervals of p for various choices of a

dom	a=.001,	a=.01,	a=1,	a=100,	a=1,000
UCL	(.19,.33),	(.21,.33),	(.20,.33),	(.19,.33),	(.19,.33)
UCH	(.19,.33),	(.21,.33),	(.20,.33),	(.19,.33),	(.18,.33)
UIL	(.19,.33),	(.21,.33),	(.20,.33),	(.20,.33),	(.20,.33)
UIH	(.19,.33),	(.21,.33),	(.19,.33),	(.19,.33),	(.18,.32)
UNL	(.19,.33),	(.21,.34),	(.20,.37),	(.20,.38),	(.20,.37)
UNH	(.19,.33),	(.21,.33),	(.20,.34),	(.19,.40),	(.19,.40)
RIL	(.17,.33),	(.19,.33),	(.19,.33),	(.18,.38),	(.16,.39)
RIH	(.09,.30),	(.12,.30),	(.08,.27),	(.09,.28),	(.08,.27)
RNL	(.10,.27),	(.12,.26),	(.10,.25),	(.10,.24),	(.10,.23)
RNH	(.12,.31),	(.13,.32),	(.13,.31),	(.13,.30),	(.14,.29)

NOTE: $p = Pr(y = 1 \mid p)$ where $y = 1$ for a victimized household and 0 otherwise

trends: Results of an international survey", Journal of Official Statistics, **15**, 129-142.

Escobar, M. D. (1994), "Estimating normal means with Dirichlet process prior", Journal of the American Statistical Association,**89**, 268-277.

Escobar, M. D. and West, M. (1995), "Bayesian density estimation and inference using mixtures", Journal of the American Statistical Association, **90**, 577-588.

Evans, R. and Sedransk, J. (2001), "Combining Data from Experiments that may be similar", Biometrika, **88**, 643-656.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems", The Annals of Statistics, **1**, 209-230.

Ferguson, T. S. (1974), "Prior distributions on space of probability measures", The Annals of Statistics, **2**, 615-629.

Forster, J. J. and Smith, P. W. F. (1998), "Model-based inference for categorical survey data subject to non-ignorable non-response", Journal of the Royal Statistical Society, Series B, **60**, 57-70.

Groves, R. M. and Couper, M. P. (1998), Nonresponse in Household Interview Surveys, New York: Wiley.

Heckman, J. (1976), "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models", Annals of Economic and Social Measurement, **5**, 475-492.

Kong, A., Liu, J. S., and Wong, W. H. (1994), "Sequential imputations and Bayesian missing data problems", Journal of the American Statistical Association, **89**, 278-288.

Malec, D. and Sedransk, J. (1992), "Bayesian Methodology for Combining the Results from Different Experiments when the Specifications for Pooling are Uncertain", Biometrika, **79**, 593-601.

Mallick, B. K. and Walker, S. G. (1997), "Combining Information from Several Experiments with Nonparametric Priors", Biometrika, **84**, 697-706.

Nandram, B. and Choi, J. W. (2002 a), "Hierarchical Bayesian Nonresponse Models for Binary Data from small areas with Uncertainty about Ignorability", Journal of the American Statistical Association, **97**, 381-388.

Nandram, B. and Choi, J. W. (2002 b). "A Bayesian Analysis of a Proportion under Nonignorable Nonresponse", Statistics in Medicine, **21**, 1189-1212.

Nandram, B., G. Han. and Choi, J.W. (2002, December), "A Hierarchical Bayesian Nonignorable Nonresponse Model for Multinomial Data From Small Areas", Survey Methodology.

Olson, R. L. (1980), "A least squares correction for selectivity bias", Econometrica **48**, 1815-1820.

Roeder, K. (1990), "Density Estimation With Confidence Sets Exemplified by Superclusters and Voids in Galaxies", Journal of the American Statistical Association, **85**, 617-624.

Stasny, E. A. (1991), "Hierarchical models for the probabilities of a survey classification and nonresponse: An example from the National Crime Survey", Journal of the American Statistical Association, **86**, 296-303.

Tanner, M. A. (1993), Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions, 2nd Edition, New York: Springer-Verlag.