A Nonparametric Bayesian Analysis of Response Data with Gaps, Outliers and Ties

Jiani Yin[1] and Balgobin Nandram[2]
[1]*Takeda Pharmaceuticals*
[2]*Department of Mathematical Sciences*
Worcester Polytechnic Institute

Abstract

Typically survey data have responses with gaps, outliers and ties, which we call GOT data, and the distributions of the responses might be skewed. Our application is on body mass index (BMI) data, which have these features, and inference is required about the 85th and 95th finite population percentiles. Because the data are collected using a two-stage sample design, usually predictive inference is done using a two-level Bayesian model with normality at both levels (responses and random effects). This is the Scott-Smith (S-S) model and it might not be robust against these GOT features. We use a two-level nonparametric Bayesian model, called the Dirichlet process Gaussian (DPG) model, with several independent Dirichlet processes at the first stage and a normal distribution on the random effects to accommodate the GOT data. The DPG model is different from the more popular two-level Dirichlet process mixture (DPM) model that has a single Dirichlet process on the random effects and independent normal distributions at the first level. Clearly, this DPM model has a shortcoming for survey data because the first stage has a normal distribution. We use our application on BMI GOT data and a very limited simulation study to compare the three models (S-S, DPM, DPG), which show, with the appropriate data, that the DPG model is preferred.

Key words: Bayesian computation; DPM model; GOT data; Predictive inference; Robust model; Survey data.

---

1.   Introduction

When samples are selected from a finite population, the most commonly used method for making inferences in current statistical literature is design-based. This method is non-parametric and it requires large sample sizes for reliable inference. Model-based inference for finite populations has been proposed as an alternative to the design-based theory. Typically survey data have responses with gaps, outliers and ties, and the distributions of the responses might be skewed. Henceforth, we use the acronym, GOT, to describe these features of

Corresponding Author: Balgobin Nandram
Email: balnan@wpi.edu

our data, and here we focus on a nonparametric Bayesian analysis of GOT data. [For convenience, all acronyms, which are used in the paper, are presented in Table 1.] The United States' National Center of Health Statistics has been collecting health data since the 1950's and one of them, body mass index (BMI) data that we discuss, has these features. We consider three two-level models with the first level accommodating the responses and the second level accounting for heterogeneity of groups of data.

We assume that data are obtained from a two-stage sample survey, for example, a two-stage cluster sampling, stratified or post-stratified sampling that is often seen in small area problems. The sampled values are observed and the nonsampled values are to be predicted using the two-level models. To gain robustness, these models start with a simple idea that uses a random distribution (*e.g.*, a Dirichlet process) in the model instead of some parametric distributions. Assuming a specific parametric form is typically motivated by technical convenience rather than by genuine prior beliefs.

In many surveys, we want to estimate quantities not only for the population as a whole, but also for sub-populations (*e.g.*, to estimate the average income for every county in the United States in order to allocate funds for needed areas). Once a hierarchical model is specified, inferences can be drawn from available data for the population quantities at any level. From a Bayesian perspective, these estimators which can be regarded as posterior means often have better properties than area-specific direct estimators. This makes two-level, and more generally hierarchical Bayesian models, useful in the problem of small area estimation (SAE) (*e.g.*, Rao and Molina 2015). That is, the sample size for a given area or domain may be too small to provide reliable estimates for themselves and it may be needed to borrow information from neighboring areas, or from areas with similar characteristics. Typically, in this two-level model, the first level accommodates the response data and the second level is used to accommodate random effects or means (*i.e.*, the small areas).

BMI is a person's weight in kilograms divided by the square of her/his height in meters and it is used as a screening tool for overweight or obesity. A high BMI can be an indicator of high body fatness. If your BMI is less than 18.5, it falls within the underweight range. If your BMI is 18.5 to 24.9, it falls within the normal or healthy weight range. If your BMI is 25.0 to 29.9, it falls within the overweight range. If your BMI is 30.0 or higher, it falls within the obese range. A child's weight status is determined using an age- and sex-specific percentile for BMI rather than BMI categories used for adults. Overweight is defined as a BMI at or above the 85th percentile and below the 95th percentile for children and teens of the same age and sex. Obesity is defined as a BMI at or above the 95th percentile for children and teens of the same age and sex.

The Expert Committee on Clinical Guidelines for Overweight in Adolescent Prevention Services published criteria for overweight to be integrated into routine screening of adolescents. BMI should be used routinely to screen for overweight and obesity in children and adolescents. Several disorders have been linked to overweight in childhood. A potential increase in type 2 diabetes mellitus is related to the increased prevalence of overweight in children (Fagot-Campagna 2000), as are cardiovascular risk factors, high cholesterol levels, and abnormal glucose levels.

We provide a Bayesian analysis of BMI data from the third National Health and Nutrition Examination Survey (NHANES III), conducted during the period October 1988 through September 1994. Due to confidentiality reasons, the final data set for this study uses only the 35 largest counties with population sizes at least 500,000. The sample sizes are less than 0.02% of the population sizes; see also Flegal *et al.* (2005, 2007) for discussions of other aspects of the NHANES III data. Kuczmarski *et al.* (2002) developed 85th and 95th percentile growth curves for US boys and girls age 2–20 years. Youths with a BMI in at least the 95th percentile for age and sex, or at least 30 (World Health Organization Consultation of Obesity 2000) should be considered overweight and referred for in depth medical follow-up to explore underlying diagnoses. Adolescents with a BMI with at least the 85th percentile (25) but below the 95th percentile (30) should be considered at risk of overweight and referred for a second-level screen; these are different for adults. (See Himes and Dietz 1994 for a summary of this discussion.) Dietz (1998) discussed health consequences of obesity in youth and childhood predictors of adult disease. Currently, obesity is one of the most serious health problems facing the world.

Nandram and Choi (2005) obtained finite population mean for children and young adults under a nonignorable nonresponse model for small domains. But the 85th and the 95th percentiles are more important and informative for BMI data. So Nandram and Choi (2010), using BMI data from NHANES III, showed how to predict these finite population percentiles of BMI for some US counties, incorporating additional measures to minimize possible biases. These measures are the inclusion of survey weights into the nonignorable nonresponse model to reflect the higher probabilities of selection among black, non-Hispanics and Hispanic-Americans. Here, we perform a Bayesian analysis of BMI data from NHANES III to obtain the 85th and 95th finite population percentiles for adults older than twenty years. We do not incorporate survey weights or covariates (age, race, sex) into our analysis. A related objective might be to estimate the proportion of obese individuals using logistic regression (*e.g.*, Nandram, Chen, Fu and Manandhar 2018), but this is far from our main objective in this paper. Here, our main objective is to compare the performance of the three two-level Bayesian models for the analysis of these data.

In Figure 1, we have shown dot plots of the BMI data by county. We can see that there are many gaps, ties and outliers in all counties. A gap occurs because no value exists between two adjacent values. For example, because the BMI values are recorded to one decimal place, there are no values between 20.0 and 20.1 (gaps), there are several values at 24.0 (ties) and there are extreme values in the right tails of the dot plots (outliers). This is why we are troubled by GOT data. Other data, such as income, when they are elicited in surveys, are also GOT data.

However, we often know very little about the specific parametric forms of the distributions, and it is also difficult to completely validate the parametric assumptions. The parametric Bayesian models based on distributional assumptions may be problematic because inferences are sensitive to such assumptions. It may be more appealing to use a nonparametric Bayesian approach because we are interested in extreme percentiles. For example, as stated already, for BMI data interest is usually on the 85th and 95th finite population percentiles. These are in the extreme right tails of the BMI distribution, and because the

BMI data have ties, outliers and gaps, it is dubious for normal distribution to represent them.

Here, we discuss the statistical modeling associated with the analysis of two-level survey data. Our intention is to propose nonparametric Bayesian alternatives using the Dirichlet process (DP). This permits robustification of inference by embedding parametric models in nonparametric models, thereby avoiding critical dependence on parametric assumptions and to allow for heterogeneity, gaps, outliers, ties, *etc.*

The existence of the DP was established by Ferguson (1973). It is a distribution over distributions; each draw from a DP itself is a distribution (*i.e.*, operating on functional spaces). The DP has gained a lot of attention recently. It has nice properties such as clustering and borrowing information which is attractive to SAE, and it can be used to address the nonparametric analysis of GOT data. The Dirichlet process mixture (DPM) model has normality on the responses (not appropriate for GOT data) and a DP on the random effects.

The more appropriate model, introduced by Yin and Nandram (2020), has several Dirichlet processes on the response data and a normal distribution for the random effects; therefore, we call it the DPG model. The Scott-Smith (Scott and Smith 1969, S-S) model has normality on both levels. The difference between the DPM model and the DPG model is that DPM model (normality on the responses) does not accommodate GOT data but its main strength is its clustering property among the small areas (*i.e.*, random effects). The DPM model is actually the opposite of the DPG model, and they are both different from the S-S model that has normality at both levels. In this paper, we compare the analysis of BMI data using the three models (S-S, DPM and DPG); our contribution is not theoretical nor methodological. Incorporating the survey weights into the DPM or the DPG needs new theory and methodology.

In Section 2, we briefly review the Scott-Smith (S-S) model and the DPM model. In Section 3, we discuss the DPG model. Section 3.1 describes the DPG model and its computation. Section 3.2 shows how to do prediction of the finite population quantities under the DPG model. In Section 4, to compare the three models, we present an analysis of BMI GOT data and in the appendix a limited simulation study. In Section 5, we present our conclusions.

2.   Scott-Smith Model and Dirichlet Process Mixture Model

In Section 2.1, we first review the S-S model, a two-level parametric model. It is used as a baseline model for the other two nonparametric models that we wish to discuss. In Section 2.2, we present a review of the Dirichlet process mixture (DPM) model. It is worth noting here that the S-S model is not robust against outliers in the data (*e.g.*, Gershunskaya and Lahiri 2018). This is also true for the DPM model. In addition, they do not make any adjustments for gaps and ties in the responses.

We assume that there are ℓ areas, and within the ith area there are N_i (known) individuals. A sample of n_i individuals is available from the ith area, and the remaining $N_i - n_i$

values are unknown. Inference is required for extreme percentiles (85th and 95th) of each area.

It is convenient to momentarily describe some notations. Let y_{ij} denote the value for the jth unit within the ith area, $i = 1, \ldots, \ell, j = 1, \ldots, N_i$. We assume that y_{ij}, $i = 1, \ldots, \ell, j = 1, \ldots, n_i$, are observed, and inference is required for the two finite population percentiles of the ith area. Let $y = (y_s, y_{ns})$, where $y_s = \{y_{ij}, i = 1, \ldots, \ell, j = 1, \ldots, n_i\}$ is the vector of observed values and $\underset{\sim}{y}_{ns} = \{y_{ij}, i = 1, \ldots, \ell, j = n_i + 1, \ldots, N_i\}$ vector of unobserved values.

2.1. Two-stage Scott-Smith (S-S) parametric model

We describe the Bayesian version of the model of Scott and Smith (1969). This S-S model was developed by Nandram, Toto and Choi (2011) for continuous data y_{ij}, where $i = 1, \ldots, \ell, j = 1, \ldots, N_i$. Letting $\delta^2 = \frac{\rho}{1-\rho}\sigma^2$, our two-level normal model (baseline parametric model) is then

$$y_{ij}|\mu_i \overset{ind}{\sim} N\left(\mu_i, \sigma^2\right), \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i, \tag{1}$$

$$\mu_i \overset{ind}{\sim} N\left(\theta, \frac{\rho}{1-\rho}\sigma^2\right), \tag{2}$$

$$\pi(\theta, \sigma^2, \rho) = \frac{1}{\pi(1+\theta^2)}\frac{1}{(1+\sigma^2)^2}, \quad -\infty < \theta < \infty, \quad \sigma^2 > 0, \quad 0 \le \rho \le 1.$$

Here, we consider a reparameterization of the S-S model, (1) and (2), together with proper non-informative priors that can allow computation of marginal likelihood and Bayes factors. We replace δ^2 by $\frac{\rho}{(1-\rho)}\sigma^2$ to gain some analytical and computational simplicity. Note that $\rho = \delta^2/(\delta^2 + \sigma^2)$ is a common intra-class correlation; see Toto and Nandram (2010), Nandram, Toto and Choi (2011) and Molina, Nandram and Rao (2014). We have used the Cauchy prior centered at 0 for θ; one can use a location-scale Cauchy prior distribution (*e.g.*, Gelman, Jakulin, Pittau and Su, 2008), but one would need to specify the location and scale parameters using the data (double using the data is forbidden in Bayesian statistics). The prior on σ^2 is a standard shrinkage prior (almost noninformative). The sampling importance resampling (SIR) algorithm is used to draw samples from the posterior distribution $\pi(\underset{\sim}{\mu}, \theta, \sigma^2, \rho|\underset{\sim}{y}_s)$; see Yin and Nandram (2020).

2.2. Dirichlet process mixture model

As pointed out by a reviewer, the Dirichlet process (DP) is well known and there is no need to review it. However, to set the stage, we need a brief description. First, we note that Binder (1982) was the first to introduce this model to survey sampling; more recently, see Nandram and Yin (2016 a,b). Let $(\Theta, \mathcal{B})$ be a measurable space, with G_0 a baseline measure (nonrandom) on the space; see Ferguson (1973) for a definition of the DP.

A Dirichlet process, $DP(\alpha, G_0)$, is defined as the distribution of a random probability measure G over $(\Theta, \mathcal{B})$ such that, for any finite measurable partition of the measur-

able space Θ, $\{A_i\}_{i=1}^n$, $\{G(A_1), \ldots, G(A_n)\} \sim \text{Dirichlet} \{\alpha G_0(A_1), \ldots, \alpha G_0(A_n)\}$. We write $G \sim \text{DP}(\alpha, G_0)$, if G is a random probability measure with a distribution given by the DP, where α is the concentration parameter. Sethuraman (1994) presented an enormously useful representation of DP in the form of a stick-breaking algorithm.

Another implied representation of the DP is the generalized Polya urn scheme, which is obtained by integrating out the random measure, G. Now considering the predictive distribution for θ_{n+1} conditioned on $\theta_1, \ldots, \theta_n$ with G integrated out, we have

$$\theta_{n+1} | \theta_1, \ldots, \theta_n \sim \frac{\alpha}{\alpha + n} G_0(\theta_{n+1}) + \frac{1}{\alpha + n} \sum_{i=1}^n \Delta_{\theta_i}(\theta_{n+1}),$$

where $\Delta_a(x)$ is the cdf of a point mass at a. The sequence of predictive distributions for $\theta_1, \theta_2, \ldots$ is called the generalized Polya urn scheme (Blackwell and MacQueen 1973). Here, it is interesting that the probability measure G is discrete with probability one, but the k distinct values $\theta_1^*, \ldots, \theta_k^* \overset{iid}{\sim} G_0$, a continuous measure (*i.e.*, the θ_i are continuous, yet $\theta_i = \theta_j, i \neq j$). There is also a slightly more compressed form that we use for prediction.

In many applications, the almost sure discreteness of the DP measure may be inappropriate. As we noted, the most popular application of the DP is in clustering data using mixture models. There is a set of latent variables, $\{\mu_1, \ldots, \mu_\ell\}$, and as for finite populations, the model is

$$
\begin{aligned}
y_{ij} | \mu_i, \phi &\overset{ind}{\sim} h(y_{ij}; \mu_i, \phi), \quad j = 1, \ldots, N_i, i = 1, \ldots, \ell, \\
\mu_i | G &\sim G, \\
G &\sim \text{DP}(\alpha, G_0).
\end{aligned}
\tag{3}
$$

This model is referred to as a Dirichlet process mixture (DPM) model; see Lo (1984) where the DPM was introduced. There are numerous applications of the DPM but see Nandram and Choi (2004) and Polettini (2017) for applications on SAE. Each μ_i is a latent parameter modeling y_{ij}, while G is the unknown distribution over parameters modeled using a DP. It can be seen as a Dirichlet process mixture of $h(y_{ij}; \mu_i, \phi)$, where y_{ij}'s with the same value of μ_i belong to the same cluster. The DPM model removes the constraint from discrete measures. It is worth noting that the DPM model for the response data is usually normal, and so it will not fit the GOT data very well. The corresponding parametric baseline model with G_0 replacing the random probability measure G is,

$$
\begin{aligned}
y_{ij} | \mu_i, \phi &\overset{ind}{\sim} h(y_{ij}; \mu_i, \phi), \quad j = 1, \ldots, N_i, i = 1, \ldots, \ell, \\
\mu_i &\sim G_0.
\end{aligned}
$$

Kalli, Griffin and Walker (2011), who suggested slice-efficient samplers, gave an improved slice sampling scheme to fit the DPM model that we use in our work, and it is based on the stick-breaking construction (Sethuraman 1994) without truncation error. The idea is to introduce latent variables that permit sampling a finite number of variables at each iteration.

In our context, DPM model is

$$y_{ij}|\mu_i, \sigma^2 \overset{ind}{\sim} \text{Normal}(\mu_i, \sigma^2), \quad j = 1, \ldots, N_i, \tag{4}$$

$$\mu_i|G \sim G, \quad i = 1, \ldots, \ell,$$

$$G \mid \theta, \sigma^2, \gamma, \rho \sim \text{DP}\left\{\gamma, \text{Normal}\left(\theta, \frac{\rho}{1-\rho}\sigma^2\right)\right\}, \tag{5}$$

$$\pi(\theta, \sigma^2, \gamma, \rho) = \frac{1}{\pi(1+\theta^2)} \frac{1}{(1+\sigma^2)^2} \frac{1}{(1+\gamma)^2}, \tag{6}$$

where $-\infty < \theta < \infty$, $\sigma^2 > 0$, $\gamma > 0$, $0 \le \rho \le 1$, and γ is the concentration parameter. In this formulation the S-S model is a baseline model; the DPM model is centered on the S-S model and γ controls how close DPM model gets to the S-S model. Here, G is a random distribution function, discrete with probability one, with distribution $DP(\cdot, \cdot)$.

3. Dirichlet Process Gaussian (DPG) Model

Since there are gaps, outliers and ties (GOT) in survey data, it is reasonable to use a random distribution drawn from the DP for the sampling population. One drawback of the S-S model is over-shrinkage; the posterior mean of certain areas may be shrunk too much towards the overall mean. Using the DP allows borrowing information moderately within some of the areas instead of all. Moreover, since there are gaps, outliers and ties in the survey data, it is reasonable to use a random distribution drawn from the DP for the sampling population. Thus, it is important to use a nonparametric procedure.

3.1. Model and computation

We consider a nonparametric hierarchical Bayesian extension of the parametric baseline model with the uncertainty on the distribution of our sampling population. Using DPs in the first level and a parametric distribution as prior gives

$$y_{ij}|G_i \overset{ind}{\sim} G_i, \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i, \tag{7}$$

$$G_i|\mu_i \overset{ind}{\sim} \text{DP}\{\alpha_i, G_0(\mu_i)\},$$

$$\mu_i \overset{iid}{\sim} H_0(\cdot),$$

where $G_0(\mu_i)$ and $H_0(\cdot)$ can be any parametric distributions. When we have strong beliefs that the area means are from a normal distribution, we may choose to use the normal likelihood in the second level. In particular, we consider $G_0 = N(\mu_i, \sigma^2)$ and $H_0(\cdot) = N(\theta, \delta^2)$, where $\delta^2 = \frac{\rho}{1-\rho}\sigma^2$ in (7) to be consistent with the two-level normal model. A full Bayesian model can be obtained by adding prior distributions. For example, we can use proper non-informative priors,

$$\pi(\alpha_i) = \frac{1}{(\alpha_i + 1)^2}, \quad \alpha_i > 0, \quad i = 1, \ldots, \ell, \tag{8}$$

$$\pi(\theta, \sigma^2, \rho) = \frac{1}{\pi(1+\theta^2)} \frac{1}{(1+\sigma^2)^2}, \tag{9}$$

$$-\infty < \theta < \infty, 0 < \sigma^2 < \infty, 0 \le \rho \le 1,$$

with independence. Generally, it is not sensible to assume that the α_i are identically distributed because they can be very different. As apparent, we have been calling (7), (8) and (9) the DPG model.

Inference of the DPG model can be easily performed. We denote $(\mu, \theta, \sigma^2, \rho)$ as $\underset{\sim}{\psi}$ and $\underset{\sim}{\alpha} = \{\alpha_1, \ldots, \alpha_\ell\}$. The posterior density of α_i are independent with other parameters $\underset{\sim}{\psi}$ in the model, conditioning on only the distinct values. Let k_i denote the number of distinct values for each area in the observed data, $\underset{\sim}{k} = \{k_i, i = 1, \ldots, \ell\}$ be the vector of k_i, $y_{i1}^*, \ldots, y_{ik_i}^*$ be the k_i distinct sample values for each i and $\underset{\sim}{y}^* = \{y_{i1}^*, \ldots, y_{ik_i}^*, i = 1, \ldots, \ell\}$ be the vector of y_{ij}^*. Thus the joint posterior density is

$$\pi(\underset{\sim}{\alpha}, \underset{\sim}{\psi} \mid \underset{\sim}{k}, \underset{\sim}{y}^*) = \left[\prod_{i=1}^{\ell} \pi(\alpha_i \mid k_i)\right] \pi(\underset{\sim}{\psi} \mid \underset{\sim}{y}^*), \tag{10}$$

where $\pi(\alpha_i|k_i) \propto \pi(k_i \mid \alpha_i)\pi(\alpha_i)$. For the other parameters $\underset{\sim}{\psi}$, we have

$$y_{ij}^*|\mu_i \overset{ind}{\sim} N\left(\mu_i, \sigma^2\right), \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, k_i, \tag{11}$$

$$\mu_i \overset{iid}{\sim} N\left(\theta, \frac{\rho}{1-\rho}\sigma^2\right),$$

$$\pi(\theta, \sigma^2, \rho) = \frac{1}{\pi(1+\theta^2)}\frac{1}{(1+\sigma^2)^2}, -\infty < \theta < \infty, 0 < \sigma^2 < \infty, 0 \leq \rho \leq 1.$$

Therefore, the algorithm for the DPG model is

Step 1 : For each i $(i = 1, \ldots, \ell)$, draw α_i from $\pi(\alpha_i|k_i) \propto \alpha^{k_i}\frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+n_i)}\frac{1}{(\alpha_i+1)^2}$ (Antoniak 1974).

Step 2: Draw $\underset{\sim}{\psi}$ from the parametric model (11) which is easy to fit.

Finally, we highlight how the DPG model takes care of GOT responses; this is apparent in the sampling process. When we integrate out the random probability measure (Blackwell and MacQueen, 1973), we get

$$f(y_i \mid \mu_i, \sigma^2, \alpha_i) = \frac{1}{\sigma}\phi(\frac{y_{i1} - \mu_i}{\sigma})$$

$$\times \prod_{k=2}^{n_i} \left\{\frac{k-1}{\alpha_i + k - 1}\frac{\sum_{j=1}^{k-1}\delta_{y_{ij}}(y_{ik})}{k-1} + \frac{\alpha_i}{\alpha_i + k - 1}\frac{1}{\sigma}\phi(\frac{y_{ik} - \mu_i}{\sigma})\right\}, \tag{12}$$

where $\delta_a(b)$ means that b is a point mass at a; so ties are accommodated. Therefore, in each area we are mixing the distributions in (12) using normal mixing distributions in the DPG model. The DPM is different being a Dirichlet process mixture of normals. The DPM model actually produces ties among the random effects or area means (clustering), which is its major strength. But it does not model gaps, outliers, ties and skewness among the responses. By putting DPs on the responses in different areas, we are attempting to accommodate the GOT data.

3.2. Prediction for the finite population

We have a simple random sample of size n_i from a finite population of size N_i, $i = 1, \ldots, \ell$. Let $y_{i1}, \ldots, y_{in_i}$ denote the sampled values. We want to predict $y_{in_i+1}, \ldots, y_{iN_i}$, the nonsampled values, and obtain the posterior predictive distributions for the 85th and 95th finite population percentiles for each area. The sampling process is

$$y_{ij}|G_i \overset{ind}{\sim} G_i, \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i,$$
$$G_i|\mu_i \overset{ind}{\sim} DP\{\alpha_i, G_0(\mu_i)\}.$$

Predictive inference for the DPG model simply uses the generalized Polya urn scheme (Blackwell and MacQueen 1973) for each area, since all areas are independent (see Nandram and Yin 2016 a,b). Once we have obtained the nonsampled $y_{ij}, j = n_i+1, \ldots, N_i, i = 1, \ldots, \ell$, we can now calculate any finite population quantity of interest. For example, for BMI data, we are interested in the finite population 85th percentile (overweight individuals) and the 95th percentile (obese individuals). The N_i are assumed known, and they are obtained from the 1990 census. To obtain the percentiles, one simply sort all the data (sample values and predicted non-sample values) in increasing order. Then, for the 85th percentile, pick the value at $.85N_i$ (nearest integer) position, and for the 95th percentile, pick the value at $.95N_i$ (nearest integer) position. Also, it is more difficult to estimate the two percentiles because they are in the right tail of the posterior distributions.

Because the N_i are very large ($N_i = n_i/.0002$), it takes relatively more time to compute the percentiles than other finite population quantities. One needs to sort $y_{i1}, \ldots, y_{iN_i}$ at each iteration, and the observed values can take different positions in the sorting. Prediction is relatively easier in the S-S and DPM models because it is done under normality, whereas in the DPG model, it is done under the Polya urn scheme.

4. Application to Body Mass Index Data

We fit the three models (S-S, DPM and DPG) to the BMI data. Our objective is mainly to compare the three models. As we mentioned in previous sections, survey data tend to have gaps, outliers and ties. The BMI data set is an example because in practice, BMI is rounded to one decimal place which creates many ties, and therefore the BMI data are a prime example of GOT data. We present the dot plots for all thirty-five areas (see Figure 1). The observations are more concentrated and having ties within the range around 25. It is also clear that the data are clustered and present gaps. Especially outside the normal weight range, the data become sparse and present bigger gaps.

The Gibbs sampler is needed for only the DPM model; for the S-S model and DPG model, we use random samplers, and therefore no monitoring is required. For the DPM model, we ran $10,000$ iterations, used $5,000$ as a "burn in" and thin every 5th to obtain $1,000$ converged posterior samples. We have computed the p values of the Geweke test and the effective sample sizes for the parameters σ^2, θ, δ^2 and γ for the DPM model. The p values are respectively .48, .41, .46, .62 and the effective sample sizes are respectively

1000, 1000, 698, 1085, thereby showing that the chain is stationary and strongly mixing. Also, trace plots and auto-correlation plots indicate that the chains converge.

For model assessment, we computed the delete-one cross validation (CV) divergence measure (Wang *et al.* 2012). The CV, obtained by predicting y_{ij} when it is deleted to obtain $y_{(ij)}$, is

$$\text{CV} = \frac{1}{\sum_{i=1}^{\ell} n_i} \sum_{i}^{\ell} \sum_{j=1}^{n_i} |y_{ij} - \text{E}(y_{ij}|y_{(ij)})|,$$

$$\text{E}(y_{ij}|y_{(ij)}) = \text{E}_{\Omega|y_{(ij)}} \text{E}(y_{ij} \mid y_{(ij)}, \Omega) = \int \text{E}(y_{ij} \mid y_{(ij)}, \Omega) f(\Omega|y_{(ij)}) d\Omega.$$

where Ω is the set of all parameters. A Monte Carlo estimator of $\text{E}(y_{ij}|y_{(ij)})$ is

$$\widehat{\text{E}(y_{ij}|y_{(ij)})} = \frac{\sum_{h=1}^{M} \{f(y_{ij} \mid y_{(ij)}, \Omega^{(h)})\}^{-1} \text{E}(y_{ij} \mid y_{(ij)}, \Omega^{(h)})}{\sum_{h=1}^{M} \{f(y_{ij} \mid y_{(ij)}, \Omega^{(h)})\}^{-1}},$$

$j = 1, \ldots, n_i, i = 1, \ldots, \ell$. Note that this measure is essentially a weighted average of the $\text{E}(y_{ij} \mid y_{(ij)}, \Omega^{(h)})$ (*i.e.*, a prediction-based measure), it is not based directly on a likelihood function. For the S-S model, DPM model and DPG model the CVs are respectively 0.765, 0.766 and 0.772. So based on this measure, there is virtually no difference among these models.

We have studied other likelihood-based measures. However, when a parametric model is nested in a nonparametric alternative, any likelihood-based diagnostics (*e.g.*, deviance information criterion, Bayesian predictive p values, log-pseudo-marginal likelihood, Bayes factor) will be misleading because we are comparing infinite dimensional distributions. One possible explanation of this fact is that the DP generates discrete distributions with probability one. This phenomenon can arise, more generally, in different contexts (*e.g.*, using the DP in goodness of fit testing). Carota and Parmiginani (1996) and Petrone and Raftery (1997) pointed out that the discreteness of the DP can have a large effect on inferences of posterior distributions and Bayes factors, when the data are partially exchangeable with an unknown partition.

We perform the predictive inference of the 85th and 95th finite population percentiles for each area using the three models (S-S, DPM and DPG). We also use a Bayesian bootstrap (*e.g.*, Yin and Nandram 2020) to do prediction in each county individually without borrowing across counties as a comparison (*i.e.*, the assumption of similarity across counties is not used). Note that for the county level, all sample sizes are roughly 100, about .02% of the population sizes. We have compared the DPG model to the S-S model, the DPM model and Bayesian bootstrap.

In Tables 2 and 3, we present summary statistics, posterior mean (PM) and posterior standard deviation (PSD), of the 85th and 95th finite population percentiles for each county of BMI data under the three models (S-S, DPM and DPG) and Bayesian bootstrap respectively. Again, the bootstrap does not allow for pooling.

First, consider the 85th percentile in Table 2. The PMs are roughly the same with those for the DPG model slightly higher, thereby showing how the DPG model takes care

of the data in the right tail of the data distribution. The PSDs under bootstrap are always larger than those under the three models, sometimes as much as two times. This shows that pooling of information across counties is helpful. However, the PSDs under the three models are roughly in increasing order: S-S model, DPM model and DPG model; the PSDs under the DPG model are always the largest, again showing how the DPG model takes care of the data in the right tail of the data distribution.

Second, we note that there are similar results in Table 3, but the differences are sharper. For example, the PSDs under the bootstrap are much too large and they are much larger under the DPG model than under the S-S model and DPM model.

In short, Tables 2 and 3 are very informative. It is not true that because the PSDs under the DPG model are larger than those under the S-S and DPM model, the DPG is worse. On the contrary, it is true that the DPG model has higher PSDs because it takes care of the GOT features of the BMI data. The S-S model and DPM model understate the PSDs because they do not take care of the gaps, outliers and ties in the data; the DPG model has a stochastic mechanism that accounts for the gaps, outliers and ties.

In Figures 2 and 3, we present plots of the posterior means with credible bands versus direct estimates for BMI data. The predictive inferences of the population percentiles are similar under the S-S model and DPM model. However, the DPG model tends to have higher predictions of the population percentiles with similar credible bands as compared with the other models. We suspect that S-S model and DPM model might underestimate the 85th and 95th finite population percentile when the GOT data are right skewed. In both Figures 2 and 3, we notice that the points under the DPG model are closer to the 45-degree straight line than those under the S-S model and DPM model. This effect is more intense in Figure 3 than in Figure 2, where the points under the S-S model and the DPM model appear to be more horizontally spread out. This is an important point because it shows that there is too much pooling in the S-S model and the DPM model. In particular, it shows that because the DPG model takes care of the GOT features of the BMI data, it is able to represent the pooling effect the best, and without the restrictive parametric assumptions in the sampling process, the DPG model tends to provide less biased estimation.

Finally, in Appendix A, we have performed a limited simulation study. We have generated data like the BMI data using the DPG model and then we fit all three models (S-S, DPM, DPG) to it. All we need from the simulation study is to describe when the DPG model performs better than the S-S model and the DPM model. We have shown that when posterior inference is performed for the 85th and 95th finite population percentiles, the DPG model performs much better than the S-S model and the DPM models, as required; see Appendix A.

Since BMI data are right skewed with gaps, outliers in the right tails and ties, the estimations given by parametric models may be incorrect. Thus based on a belief that the parametric model is too restrictive, we prefer the analysis based on the nonparametric DPG model.

5.   Concluding Remarks

Bayesian nonparametric methods are motivated by the desire to avoid overly restrictive parametric assumptions. For GOT data (gaps, outliers and ties), we believe that our DPG model will play an important role for future analysis of "continuous" survey data. The S-S model and the DPM model have normality assumption on the response data, and hence they do not address the problem of GOT data. Based on the analyses, we have positive view of the DPG model.

Our main conclusion is that when data come from the DPG model, it will do much better than the S-S model and the DPM model. The DPM model, while nonparametric (good), it has the DP on the wrong level (random effects) for GOT responses. The S-S model has normality on both levels. Normality on the responses is not appropriate for GOT data. Hence, neither of these two models can accommodate GOT responses. Based on this point, we believe that among the three models (S-S, DPM, DPG), it is the DPG model that is most appropriate to represent the BMI GOT data. The DPG model is more important when interest is on the 85th and 95th finite population percentiles because there are larger gaps in the right tails of the data distributions (*i.e.*, outliers). However, the DPM model is attractive for an important reason; it provides clustering among the small areas but not the data. This clustering among the small areas can be accommodated in the DPG model; see Nandram and Yin (2019).

It is important to include survey weights in the DPG model. This can be done using a standardized composite likelihood. Covariates can also be incorporated into the DPG model. However, our main contribution in this paper has been to demonstrate the superiority of the DPG model when it is fitted to the BMI GOT data.

Acknowledgments

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, **2(6)**, 1152-1174.

Binder, D. A. (1982). Non-parametric Bayesian models for samples from finite populations. *Journal of the Royal Statistical Society*, Series B, **44(3)**, 388-393.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *The Annals of Statistics*, **1(2)**, 353-355.

Carota, C. and Parmigiani, G. (1996). On Bayes factor for nonparametric alternatives. *Bayesian Statistics*, **5**, 507-511.

Dietz, W. H. (1998). Health consequences of obesity in youth: Childhood predictors of adult disease. *Pediatrics*, **101**, 518–525.

Fagot-Campagna, A. (2000). Emergence of type 2 diabetes mellitus in children: Epidemiological evidence. *Journal of Pediatric Endocrinology Metabolism*, **13**, 1395–1405.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1(2)**, 209-230.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association*, **293**, 1861–1867.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. (2007). Cause-specific excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association*, **298**, 2028–2037.

Gelman, A., Jakulin, A., Pittau, M. P. and Su, Y-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, **2(4)**, 1360-1383.

Gershunskaya, J. and Lahiri, P. (2018). Robust empirical best small area finite population mean estimation using a mixture model. *Calcutta Statistical Association Bulletin*, **69(2)**, 183-204.

Himes, J. H., and Dietz, W. H. (1994). Guidelines for overweight in adolescent preventive services: Recommendations from an expert committee. *American Journal of Clinical Nutrition*, **59**, 307–316.

Kalli, M. and Griffin, J. E. and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21(1)**, 83-105.

Kuczmarski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., Wei, R., Curtin, L. R., Roche, A. L., and Johnson, C. L. (2002). 2000 CDC growth charts for the united states: Methods and development. *Vital and Health Statistics*, **11(246)**, 1–190.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, **12(1)**, 351-357.

Molina, I., Nandram, B. and Rao, J. N. K., (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics*, **8(2)**, 852-885.

Nandram, B. and Choi, J. W. (2004). Nonparametric Bayesian analysis of a proportion for a small area under nonignorable nonresponse. *Journal of Nonparametric Statistics*, **16(6)**, 821-839.

Nandram, B., Chen, L., Fu, S. and Manandhar, B. (2018). Bayesian Logistic Regression for Small Areas with Numerous Households. *Statistics and Application*, **16(1)**, 171-205.

Nandram, B. and Choi, J. W. (2010). Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *Journal of the American Statistical Association*, **105(489)**, 120-135.

Nandram, B. and Choi, J. W. (2005). Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the NHANES data. *Survey Methodology*, **31(1)**, 73-84.

Nandram, B. and Toto, M. C. S. and Choi, J. W. (2011). A Bayesian benchmarking of the Scott-Smith model for small areas. *Journal of Statistical Computation and Simulation*, **81(11)**, 1593-1608.

Nandram, B. and Yin, J. (2016a). Bayesian predictive inference under a Dirichlet process with sensitivity to the normal baseline. *Statistical Methodology*, **28**, 1-17.

Nandram, B. and Yin, J. (2016b). A nonparametric Bayesian prediction interval for a finite population mean. *Journal of Statistical Computation and Simulation*, **86(16)**, 3141-3157.

Nandram, B. and Yin, J. (2019). Hierarchical Bayesian Models for small areas with Dirichlet processes. *JSM Proceedings, 2594-2613, Survey Research Methods Section. Alexandria, VA: American Statistical Association.*

Polettini, S. (2017). A generalized semiparametric Bayesian Fay-Herriot model for small area estimation shrinking both means and variances. *Bayesian Analysis*, **12(3)**, 729-752.

Petrone, S. and Raftery, A. E. (1997). A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability. *Statistics and Probability Letters*, **36(1)**, 69-83.

Rao, J. N. K. Molina, I. (2015). *Small Area Estimation*, John Wiley and Sons, NY.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639-650.

Scott, A. and Smith, T. M. F. (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association*, **64(327)**,830-840.

Toto, M. C. S. and Nandram, B. (2010). A Bayesian predictive inference for small area means incorporating covariates and sampling weights. *Journal of Statistical Planning and Inference*, **140**, 2963-2979.

Wang, J. C., Holan, S. H., Nandram, B., Barboza, W., Toto, C. and Anderson, E. (2012). A Bayesian approach to estimating agricultural yield based on multiple repeated surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, **17(1)**, 84-106.

WHO Consultation on Obesity (2000).Obesity: Preventing and managing the global epidemic. *Technical Report 894*, World Health Organization, Geneva, Switzerland.

Yin, J. and Nandram, B. (2020). A Bayesian small area model with Dirichlet processes on responses. *Statistics in Transition, New Series*, **21(3)**, 1–19.

APPENDIX A: Simulation Study Showing the Strength of the DPG Model

We conduct a limited simulation study. We have simulated a single data set from the DPG model and we fitted all three models (S-S, DPM, DPG) to this data set.

We choose $\ell = 50$ and the sample sizes, n_i, for 50 areas. The sample sizes are 35 for each of the first 10 areas, 50 for each of the second 10 areas, 100 for each of the third 10 areas, 200 for each of the fourth 10 areas and 500 for each of the last 10 areas. Then, the population sizes are selected as $N_i = 100n_i, i = 1, \ldots, \ell$. These are comparable to the BMI data. For convenience, to simulate the data set, we have taken $\theta = 0.0$, $\sigma^2 = 0.01$, $\delta^2 = 0.04$, thereby making $\rho = 0.8$. For the concentration parameters of the Dirichlet processes, we have selected $\alpha_i \overset{ind}{\sim} 0.5 + \text{Beta}(5, 5), i = 1, \ldots, \ell$. These choices allow us to have data similar to the BMI data with some flexibility to get gaps, outliers and ties when data are simulated from the DPG model.

We use absolute bias (AB) and posterior root mean squared error (PRMSE) to compare the models. We know the true values of the finite population quantities, denoted by T. Then, $AB = \mid PM - T \mid$ and $PRMSE = \sqrt{(PM - T)^2 + PSD^2}$. We compute these quantities for each of the fifty counties for the 85th and 95th finite population percentiles, and respectively we average them over the fifty counties. We present AB and PRMSE in Table 4 for the case in which data are generated from the DPG model (*i.e.*, for GOT data). It is nice that AB and $PRMSE$ are manyfold smaller under the DPG model than the other two models (S-S, DPM). Therefore, it is clear that the DPG model performs much better than the S-S model and the DPM model, when inference is made about the 85th and 95th finite population percentiles. This is exactly what we want to happen.

Table 1: Acronyms used in the Presentation

Acronym	Meaning
GOT	gaps, outliers and ties
BMI	body mass index
SAE	small area estimation
S-S	Scott-Smith
DP	Diriclet process
DPM	Dirichlet process mixture
DPG	Dirichlet process Gaussian
AB	absolute bias
PRMSE	posterior root mean squared error
PM	posterior mean
PSD	posterior standard deviation
CV	cross validation

Table 2: Comparison of posterior mean (PM) and posterior standard deviation (PSD) of the finite population 85th percentile for each county of body mass index (BMI) data by te three models (S-S, DPM, DPG) and Bayesian bootstrap

	Bootstrap		S-S		DPM		DPG	
	PM	PSD	PM	PSD	PM	PSD	PM	PSD
1	32.14	0.50	32.48	0.35	32.50	0.39	32.46	0.47
2	34.76	1.24	32.93	0.45	32.95	0.43	34.08	0.82
3	30.76	0.78	32.05	0.39	32.00	0.44	31.94	0.63
4	31.57	1.07	31.97	0.43	31.93	0.42	32.48	0.61
5	30.51	0.90	31.75	0.47	31.75	0.45	31.87	0.72
6	33.82	1.22	33.42	0.44	33.35	0.44	33.55	0.67
7	31.59	0.85	32.58	0.36	32.58	0.39	32.45	0.72
8	32.25	0.67	32.46	0.36	32.48	0.42	32.70	0.53
9	32.81	1.18	33.03	0.41	33.01	0.42	33.15	0.75
10	34.01	0.74	33.07	0.39	33.08	0.36	33.73	0.48
11	32.75	0.54	32.78	0.26	32.79	0.27	32.90	0.49
12	30.26	0.80	31.67	0.42	31.67	0.38	31.45	0.53
13	31.91	0.88	32.34	0.36	32.32	0.43	32.64	0.57
14	32.37	0.38	32.80	0.19	32.82	0.20	32.50	0.37
15	33.39	0.50	32.84	0.40	32.85	0.41	33.39	0.42
16	32.21	0.75	32.72	0.37	32.71	0.40	32.73	0.62
17	30.88	0.83	31.95	0.40	31.91	0.42	32.07	0.72
18	31.18	0.80	32.29	0.39	32.28	0.49	32.21	0.85
19	32.03	0.97	32.09	0.38	32.08	0.42	32.77	0.56
20	32.71	0.96	32.50	0.39	32.52	0.42	33.08	0.61
21	33.08	0.98	32.57	0.40	32.58	0.44	33.28	0.56
22	32.06	0.72	32.65	0.36	32.68	0.37	32.57	0.57
23	31.18	0.77	31.85	0.42	31.81	0.42	32.19	0.70
24	32.66	0.66	32.64	0.37	32.68	0.40	32.96	0.52
25	31.63	0.98	32.37	0.39	32.39	0.42	32.47	0.73
26	32.02	0.96	32.34	0.40	32.30	0.45	32.77	0.57
27	31.56	0.44	32.34	0.31	32.36	0.39	32.16	0.50
28	33.51	1.51	32.87	0.39	32.89	0.40	33.33	0.80
29	31.53	0.97	32.30	0.45	32.31	0.49	32.57	0.80
30	30.62	0.94	31.89	0.43	31.83	0.45	32.13	0.71
31	32.36	0.57	33.02	0.38	32.99	0.38	32.72	0.49
32	33.24	0.89	32.96	0.37	32.96	0.37	33.31	0.62
33	30.54	0.51	32.03	0.37	32.01	0.42	31.61	0.57
34	32.48	0.49	32.78	0.31	32.82	0.31	32.71	0.45
35	31.78	1.04	32.40	0.35	32.41	0.42	32.54	0.75

Table 3: Comparison of posterior mean (PM) and posterior standard deviation (PSD) of the finite population 95th percentile for each county of body mass index (BMI) data by the three models (S-S, DPM, DPG) and Bayesian bootstrap

	Bootstrap		S-S		DPM		DPG	
	PM	PSD	PM	PSD	PM	PSD	PM	PSD
1	35.52	1.27	35.79	0.42	35.81	0.45	36.21	0.88
2	40.88	2.32	36.45	0.46	36.47	0.45	38.83	1.54
3	34.90	2.58	35.36	0.47	35.32	0.51	36.16	1.43
4	35.59	1.12	35.31	0.45	35.27	0.45	36.26	0.85
5	35.82	1.61	35.19	0.51	35.19	0.50	36.53	0.92
6	39.32	1.58	37.00	0.44	36.94	0.44	38.45	0.74
7	35.93	1.12	35.95	0.40	35.94	0.44	36.50	0.69
8	37.32	1.49	35.90	0.43	35.92	0.48	37.26	0.86
9	38.76	1.54	36.55	0.45	36.53	0.46	38.02	0.84
10	39.82	1.64	36.48	0.41	36.48	0.41	38.32	1.14
11	37.49	0.94	36.19	0.28	36.21	0.29	37.36	0.71
12	35.84	1.50	35.17	0.47	35.18	0.44	36.46	0.89
13	36.13	1.20	35.68	0.40	35.66	0.45	36.65	0.93
14	36.90	0.80	36.16	0.22	36.19	0.23	36.96	0.69
15	36.04	1.47	36.00	0.48	36.03	0.49	36.64	0.89
16	36.44	1.40	36.08	0.41	36.08	0.44	36.79	0.93
17	34.70	0.99	35.27	0.44	35.23	0.45	35.77	0.83
18	35.57	0.81	35.68	0.38	35.65	0.46	36.16	0.78
19	34.88	0.88	35.31	0.40	35.30	0.44	35.85	0.78
20	37.08	1.89	35.82	0.42	35.84	0.46	37.11	1.14
21	35.75	1.03	35.75	0.44	35.77	0.47	36.30	0.84
22	35.56	1.08	35.94	0.43	35.98	0.42	36.12	0.89
23	36.46	1.46	35.29	0.45	35.24	0.46	36.84	0.92
24	37.80	2.17	36.02	0.44	36.06	0.45	37.40	1.33
25	37.29	2.60	35.76	0.43	35.77	0.46	37.23	1.47
26	36.18	1.92	35.67	0.52	35.62	0.55	36.90	1.10
27	36.09	1.30	35.75	0.38	35.77	0.44	36.51	0.78
28	40.33	1.37	36.50	0.44	36.53	0.46	38.84	0.96
29	35.71	1.10	35.66	0.52	35.67	0.52	36.43	0.78
30	34.57	1.11	35.20	0.48	35.15	0.49	35.87	0.83
31	35.43	1.06	36.28	0.39	36.26	0.39	36.01	0.68
32	39.12	1.40	36.43	0.41	36.43	0.40	38.24	1.00
33	34.10	0.83	35.31	0.42	35.30	0.46	35.32	0.88
34	35.98	1.02	36.09	0.36	36.12	0.36	36.36	0.85
35	37.83	1.13	35.92	0.38	35.92	0.44	37.57	0.92

Table 4: Comparison of absolute bias (AB) and posterior root mean squared error (PRMSE) of the finite population 85th percentile and 95th percentile for the simulated data from DPG model averaged over areas

	S-S Model		DPM Model		DPG Model	
	AB	PRMSE	AB	PRMSE	AB	PRMSE
85th percentile	379.0	385.9	384.1	394.6	18.05	40.0
95th percentile	550.2	556.3	555.3	563.7	35.5	101.0

NOTE: Data are generated from the DPG model, and all three models (S-S, DPM, DPG) are fitted to the data. The numbers in the table must be multiplied by 10^{-4}.

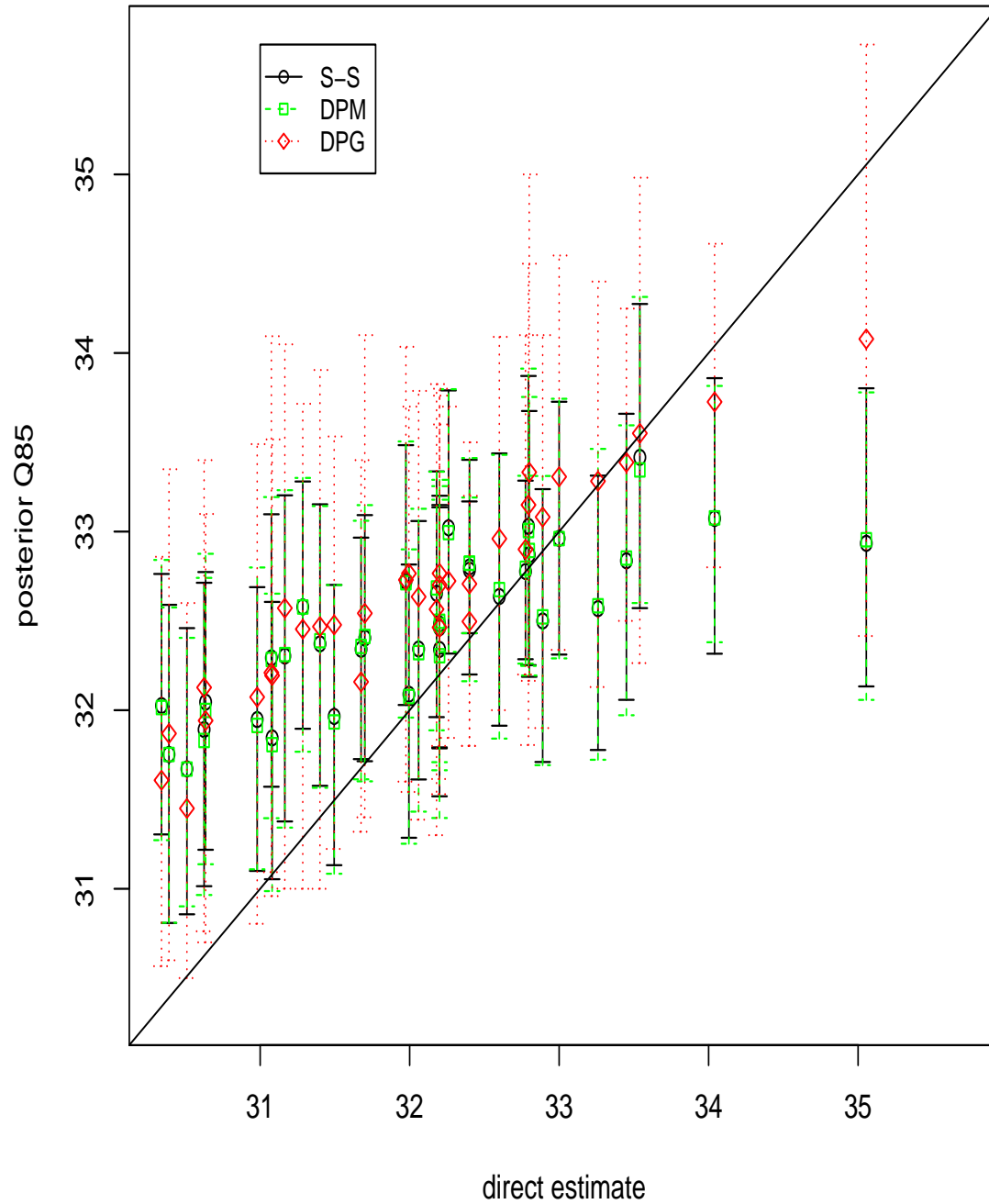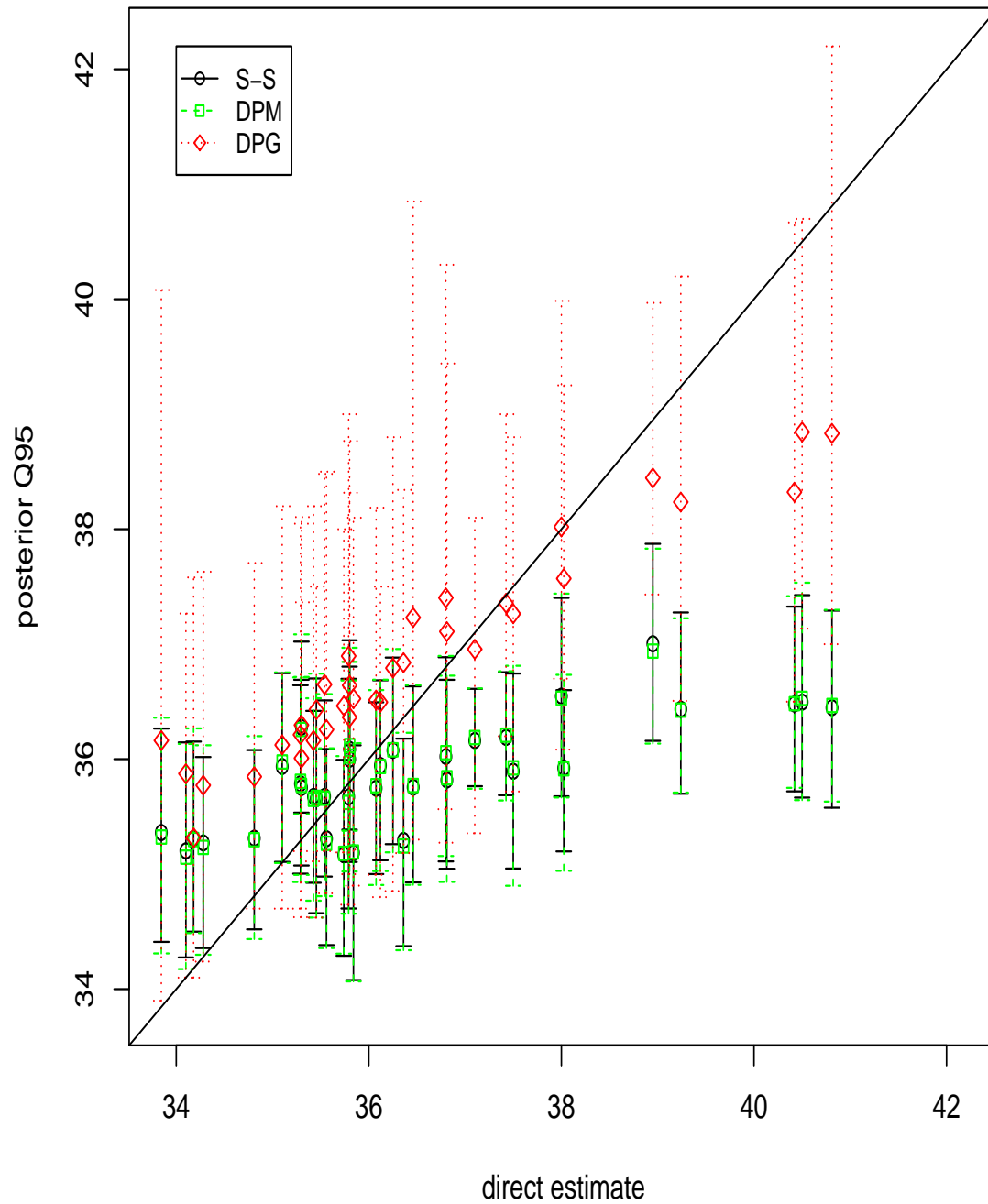Figure 1: Dot plots of body mass index (BMI) for thirty-five counties

Figure 2: Comparison for body mass index (BMI) data (posterior means with credible bands versus direct estimates): the predictive inference of the finite population 85th percentile for each county under the three models (S-S, DPM, DPG)

Figure 3: Comparison for body mass index (BMI) data (posterior means with credible bands versus direct estimates): the predictive inference of the finite population 95th percentile for each county under the three models (S-S, DPM, DPG)