

Article

A Bayesian analysis of small area probabilities under a constraint

by Balgobin Nandram and Hasanjan Sayit



December 2011

A Bayesian analysis of small area probabilities under a constraint

Balgobin Nandram and Hasanjan Sayit¹

Abstract

In many sample surveys there are items requesting binary response (*e.g.*, obese, not obese) from a number of small areas. Inference is required about the probability for a positive response (*e.g.*, obese) in each area, the probability being the same for all individuals in each area and different across areas. Because of the sparseness of the data within areas, direct estimators are not reliable, and there is a need to use data from other areas to improve inference for a specific area. Essentially, a priori the areas are assumed to be similar, and a hierarchical Bayesian model, the standard beta-binomial model, is a natural choice. The innovation is that a practitioner may have much-needed additional prior information about a linear combination of the probabilities. For example, a weighted average of the probabilities is a parameter, and information can be elicited about this parameter, thereby making the Bayesian paradigm appropriate. We have modified the standard beta-binomial model for small areas to incorporate the prior information on the linear combination of the probabilities, which we call a constraint. Thus, there are three cases. The practitioner (a) does not specify a constraint, (b) specifies a constraint and the parameter completely, and (c) specifies a constraint and information which can be used to construct a prior distribution for the parameter. The gridgy Gibbs sampler is used to fit the models. To illustrate our method, we use an example on obesity of children in the National Health and Nutrition Examination Survey in which the small areas are formed by crossing school (middle, high), ethnicity (white, black, Mexican) and gender (male, female). We use a simulation study to assess some of the statistical features of our method. We have shown that the gain in precision beyond (a) is in the order with (b) larger than (c).

Key Words: Accept-reject algorithm; Binomial distribution; Generalized beta distribution; Gridgy Gibbs sampler; Simulation.

1. Introduction

It is a standard practice to use models to “borrow strength” in small area estimation (Rao 2003). Owing to the sparseness of the data in each area, direct estimates for small areas are typically not reliable. Our procedure allows a practitioner to incorporate prior information about a linear combination of binomial probabilities, one for each area. This is a constraint that we include as a weighted average of the area probabilities in the standard beta-binomial model. The weighted average can be assumed known or unknown. In the case when this value is unknown, we consider the scenario when there is some information which can be elicited from an expert in the form of prior distribution. This is different from standard practice in design based survey sampling in which auxiliary information is incorporated as in ratio and regression estimators (Cochran 1977). When the value can be specified exactly, there will be an increase in precision because prior information is incorporated into the model.

The beta-binomial model has been studied extensively. For example, Nandram and Sedransk (1993), Nandram (1998) and Nandram and Choi (2002) show how to do Bayesian predictive inference of finite population proportions of the small areas for binomial and multinomial data. These models assume that the binomial probabilities share a common effect, thereby permitting adaptive pooling of the

data from small areas (or clusters). However, it is possible to improve on these models further by including additional information using covariates via generalized linear models (*e.g.*, see Ghosh, Natarajan, Stroud and Carlin 1998). It is worth noting that none of these works propose ways to incorporate prior information about linear combination of model parameters. Substantial gains in precision are expected when such prior information is incorporated in small area models; see Silvapulle and Sen (2006) for a book-length discussion of constrained statistical inference. It is also worth noting that Lazar, Meeden and Nelson (2008) showed how to include constraints in nonparametric Bayesian approach via a Polya urn scheme to predictive distribution of finite population parameters.

Our procedure is related to external benchmarking which occurs when a pre-specified estimator is obtained from external sources, such as a different survey, a census, or other administrative records. In benchmarking one wants the parts to add up to the whole. For example, when surveys are conducted over time, there are typically monthly surveys and annual surveys which are of much better quality than the monthly surveys. When the monthly surveys are estimated such that these estimates add up to the annual survey totals, there is a protection against model failure and therefore improved estimates (*i.e.*, reduced bias and possibly an increase in precision). These problems are prevalent in the government agencies especially in employment and sales;

1. Balgobin Nandram and Hasanjan Sayit, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280. E-mail: balnan@wpi.edu, hs7@wpi.edu.

see Hillmer and Trabelsi (1987) for an example on retail sales of hardware stores from the U.S. Census Bureau.

Prior information from external benchmarking will lead to improved precision but can produce severely biased estimators as well. This will depend on how different the current survey is from the prior ones. Nandram, Toto and Choi (2011) applied external benchmarking to estimate the finite population means of small areas. The constraint is the finite population mean for the entire population is a prespecified value which again can be obtained from a prior survey, census or administrative records. In our current work we are not incorporating information about a linear combination of the finite population values, but rather we are inputting information about a linear combination of the superpopulation parameters (in this case binomial probabilities).

We consider the problem in which binomial counts are obtained from similar small areas, and inference is required about the binomial probabilities. In the conclusion, we discuss how to extend our method to obtain the predictive distribution of finite population proportions. The standard beta-binomial model may be inadequate, and additional prior information must be incorporated. Our thesis is that there is an increase in precision over the standard beta-binomial small area model when prior information about the weighted average of the probabilities (*e.g.*, average of the probabilities) is incorporated. That is, we incorporate prior information about a linear combination of binomial probabilities (a weighted average). The weights can be proportional to population sizes, and under proportional allocation they can be proportional to the sample sizes themselves. The purpose of incorporating prior information about the binomial probabilities is to increase precision, and at the same time one needs to control the bias.

It is much easier for a survey practitioner to specify the value of the overall probability rather than the individual area probabilities. That is, the overall probability can be specified with relatively much less error than the individual probabilities. Of course, one can specify the overall probability using prior information (a prior survey, census or administrative records), and so the specification of the overall probability will depend on the quality of the prior information. Thus, the problem falls naturally within the Bayesian paradigm because we are incorporating prior information about a parameter via a distribution. Thus, there will be gains in precision because of the extra information. However, a practitioner can still proceed when there is no prior information. One can use the ratio of the total success and total sample size over areas to form a reasonable specification of the overall probability which is typically not of interest. This estimate will have much higher precision than the one for individual areas. There will still be gain in

precision, but clearly such gain is due to using the current data (double use) and the constraint.

One example of a survey in which reliable information can be obtained to perform the benchmarking is the National Health Interview Survey (NHIS) which is conducted annually by the National Center for Health Statistics to assess an aspect of Health of the U.S. population. This is a population-based survey and there are many health indicators of interest; one of these indicators is the number of doctor visits made in the past two weeks, and an informative quantity is the proportion of people who made at least one doctor visit last year (*e.g.*, Nandram and Choi 2002). These proportions are useful for small domains formed by crossing age, race and sex for a particular state last year. Because the estimates over a state change very slowly over the previous years, the overall estimate from the year immediately preceding last year can be used as a reliable benchmark for last year. If a reliable estimate cannot be obtained for the benchmark, one can construct an informative prior distribution for it. For example, one can use the method of moments to equate the sample mean and sample variance of the overall estimates for the past few years to the mean and variance of a beta distribution to get a beta prior distribution. In either case, our procedure can be applied.

The plan of this paper is as follows. In Section 2 we describe the methodology. Specifically, we describe the standard beta-binomial model, and we develop two additional models to incorporate the extra information using appropriate prior distributions. We also describe posterior inference and how to perform the nonstandard computations. In Section 3 we describe an illustrative example on obesity, and a simulation study to assess empirically the statistical properties of our models. Section 4 has concluding remarks. We also discuss how to do Bayesian predictive inference for finite population proportions. While we discuss binary data, we also show how one can extend our method to polychotomous data.

2. Methodology

We show how to incorporate the constraint into the beta-binomial model in two ways, thereby providing a set of alternative models. In Section 2.1 we describe the models and in Section 2.2 we describe posterior inference. We attempt to explain what the constraint does to the estimates of the probabilities using an approximation. In Section 2.3 we describe the computation, and we describe a new algorithm as well.

2.1 Models

We assume that binary data are available from ℓ small areas, and we assume that the probability that an individual

responds in the i^{th} area is π_i , $i = 1, \dots, \ell$. Let n_i be the number of individuals sampled from the i^{th} area, $i = 1, \dots, \ell$. Also let s_i denote the number of individuals with the characteristic and $f_i = n_i - s_i$ be the number of individuals without the characteristic in the i^{th} area, $i = 1, 2, \dots, \ell$. Then the standard beta-binomial hierarchical Bayesian model is

$$s_i | \pi_i \sim \text{Binomial}(n_i, \pi_i), \quad (1)$$

$$\pi_i | \mu, \tau \sim \text{Beta}\{\mu\tau, (1-\mu)\tau\}, \quad i = 1, \dots, \ell \quad (2)$$

and

$$p(\mu, \tau) = \frac{1}{(1+\tau)^2}, \quad 0 < \mu < 1, \tau \geq 0. \quad (3)$$

We use a shrinkage prior for τ because it is proper and noninformative, and there are no conjugate priors. Priors of the form $p(\tau) \propto 1/\tau$ are discouraged; see, for example, Gelman (2006). Other alternatives are half Cauchy densities and gamma densities (one would need to specify the hyperparameters). Henceforth, we will call the model specified by (1), (2) and (3) the unrestricted (UR) model or Model 1.

We next describe the restricted model, which is an extension of the unrestricted model. We obtain a simple linear combination of the binomial probabilities. Letting $\tilde{\pi}_i = s_i/n_i$ and

$$\omega_i = \frac{n_i}{\sum_{i=1}^{\ell} n_i}, \quad i = 1, \dots, \ell,$$

we have

$$\frac{\sum_{i=1}^{\ell} s_i}{\sum_{i=1}^{\ell} n_i} = \sum_{i=1}^{\ell} \omega_i \tilde{\pi}_i.$$

Thus, taking the π_i unknown, the linear combination is $\sum_{i=1}^{\ell} \omega_i \pi_i$.

Therefore, we need to make an adjustment in (2) to incorporate the restriction, $\sum_{i=1}^{\ell} \omega_i \pi_i = \theta$ conditional on θ . We do so by introducing the variable $\phi = \sum_{i=1}^{\ell} \omega_i \pi_i - \theta$; so that the restriction is equivalent to $\phi = 0$. Now one of the variables, π_i , $i = 1, \dots, \ell$, is redundant. It is worth noting that one can choose any one of π_1, \dots, π_{ℓ} , and without loss of generality and for ease of exposition, we choose π_{ℓ} . Thus, to incorporate the restriction, we transform π_{ℓ} to $\phi = \sum_{i=1}^{\ell} \omega_i \pi_i - \theta$, keeping $\pi_1, \dots, \pi_{\ell-1}$ untransformed, and we let $\boldsymbol{\pi}_{(\ell)} = (\pi_1, \dots, \pi_{\ell-1})'$.

As the jacobian is $1/\omega_{\ell}$,

$$p(\boldsymbol{\pi}_{(\ell)}, \phi | \mu, \tau, \theta) = \frac{1}{\omega_{\ell}} \prod_{i=1}^{\ell-1} \frac{\pi_i^{\mu\tau-1} (1-\pi_i)^{(1-\mu)\tau-1}}{B\{\mu\tau, (1-\mu)\tau\}} \times \frac{\left[\frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}} \right]^{\mu\tau-1} \left[1 - \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}} \right]^{(1-\mu)\tau-1}}{B\{\mu\tau, (1-\mu)\tau\}}, \quad (4)$$

where

$$0 < \pi_i < 1, \quad i = 1, \dots, \ell,$$

$$0 < \mu < 1, \tau > 0, \phi + \theta - \omega_{\ell} \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \phi + \theta,$$

and

$$\pi_{\ell} = \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}}. \quad (5)$$

Note that the joint prior density of $(\boldsymbol{\pi}_{(\ell)}, \phi)$ in (4) is well defined. We wish to take $\phi = 0$ in (5) to incorporate the restriction, but when $\phi = 0$ the joint density of $\boldsymbol{\pi}_{(\ell)}$ is not well defined.

We assume μ, τ, θ are independent a priori with $p(\mu, \tau, \theta) = p_1(\mu, \tau) p_2(\theta)$, where

$$p_1(\mu, \tau) = \frac{1}{(1+\tau)^2}, \quad 0 < \mu < 1, \tau \geq 0$$

as in (3), and $p_2(\theta)$ is given by

$$\theta \sim \text{Beta}\{\mu_0\tau_0, (1-\mu_0)\tau_0\}. \quad (6)$$

For the restricted model we consider two scenarios. Letting $\tau_0 \rightarrow \infty$, θ becomes a point mass at μ_0 , and in this case $\theta = \mu_0$ is to be specified by a practitioner; we will call the adjusted model the fixed (FI) model or Model 2. We have a second scenario in which a practitioner specifies μ_0 and τ_0 but not θ ; we will call this adjusted model the informative (IN) model or Model 3. Thus, there are three models, including the unrestricted model. To provide a unified framework, we need all our priors to be proper. The exact value of θ is likely to be unknown in most applications, and this can lead to estimates which are not internally coherent.

It is worth noting that we have considered an additional model to help study the gain in precision of IN relative to FI. For comparison we want to impose a proper but noninformative prior on θ , so that $\theta \sim \text{Uniform}(0, 1)$ is not an unreasonable choice. Letting $\mu_0 = 1/2, \tau_0 = 2$, we get $\theta \sim \text{Uniform}(0, 1)$ with this prior, and we will call the adjusted model the uniform (UN) model or Model 4; of

course, we do not need to specify μ_0 and τ_0 . It is worth noting that the prior corresponding to $\tau \rightarrow \infty$ is improper as it corresponds to $\theta \sim \text{Beta}(0, 0)$. We do not consider this model further; however, although UN does not have a constraint, we will consider it briefly throughout.

2.2 Posterior inference

We consider making posterior inference about π_i , $i = 1, \dots, \ell$. Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)'$ and $\boldsymbol{\pi}_{(i)} = (\pi_1, \dots, \pi_{i-1}, \pi_{i+1}, \dots, \pi_\ell)'$ [e.g., $\boldsymbol{\pi}_{(\ell)} = (\pi_1, \dots, \pi_{\ell-1})'$ as defined above].

We use Bayes' theorem to find the joint posterior densities of all parameters. First, under the unrestricted model specified by (1), (2) and (3) the joint posterior density of $\boldsymbol{\pi}, \mu, \tau$ is

$$g(\boldsymbol{\pi}, \mu, \tau | s) \propto \prod_{i=1}^{\ell} \frac{\pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1-\mu)\tau - 1}}{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}} \times \prod_{i=1}^{\ell} \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \times \frac{1}{(1 + \tau)^2}, \tag{7}$$

$0 < \pi_i < 1, 0 < \mu < 1, \tau > 0, i = 1, \dots, \ell$.

Lemma 1 Under the unrestricted model the joint posterior density, $g(\boldsymbol{\pi}, \mu, \tau | s)$, is proper.

A proof of Lemma 1 is given in Appendix A.

Under the restricted model the joint posterior density of $\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta, \phi$ is

$$p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta, \phi | s) \propto \prod_{i=1}^{\ell-1} \frac{\pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1-\mu)\tau - 1}}{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}} \times \frac{\left[\frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{s_\ell + \mu\tau - 1} \left[1 - \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{f_\ell + (1-\mu)\tau - 1}}{B\{s_\ell + \mu\tau, f_\ell + (1-\mu)\tau\}} \times \prod_{i=1}^{\ell} \left[\frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \right] \times \theta^{\mu_0\tau_0 - 1} (1 - \theta)^{(1-\mu_0)\tau_0 - 1} \times \frac{1}{(1 + \tau)^2}, \tag{8}$$

$0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0, \phi + \theta - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \phi + \theta, 0 < \theta < 1$. Note that $\pi_\ell = (\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$.

We get the pertinent joint posterior density by incorporating the constraint ($\phi = 0$) into (8). That is, $p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta | s, \phi = 0) \propto p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta, \phi = 0 | s)$, where

$$p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta | s, \phi = 0) \propto \prod_{i=1}^{\ell-1} \frac{\pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1-\mu)\tau - 1}}{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}} \times \frac{\left[\frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{s_\ell + \mu\tau - 1} \left[1 - \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{f_\ell + (1-\mu)\tau - 1}}{B\{s_\ell + \mu\tau, f_\ell + (1 - \mu)\tau\}} \times \prod_{i=1}^{\ell} \left[\frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \right] \times \theta^{\mu_0\tau_0 - 1} (1 - \theta)^{(1-\mu_0)\tau_0 - 1} \times \frac{1}{(1 + \tau)^2}, \tag{9}$$

$0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0, \theta - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta, 0 < \theta < 1$. Note again that $\pi_\ell = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$. It is worth noting that the joint posterior density (9) incorporates the constraint, $\sum_{i=1}^{\ell} \omega_i \pi_i = \theta$, exactly because $\pi_\ell = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell, \theta - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta$. That is, the joint posterior density is not a function of π_ℓ , and posterior inference about π_ℓ follows from the identity, $\pi_\ell = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$. Thus, there is absolutely no difference between θ and $\sum_{i=1}^{\ell} \omega_i \pi_i$.

Theorem 1 Under the restricted model the joint posterior density, $p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta | s, \phi = 0)$, is proper.

A proof of Theorem 1 is given in Appendix A.

We note the difference between the densities for the unrestricted model in (7) and the restricted model in (9). Essentially, the term

$$\left(\frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right)^{s_\ell + \mu\tau - 1} \times \left(1 - \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right)^{f_\ell + (1-\mu)\tau - 1} \times \theta^{\mu_0\tau_0 - 1} (1 - \theta)^{(1-\mu_0)\tau_0 - 1}$$

in (9) replaces $\pi_\ell^{s_\ell + \mu\tau - 1} (1 - \pi_\ell)^{f_\ell + (1-\mu)\tau - 1}$ in (7). Note that in (9),

$$\pi_\ell = \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell}.$$

Let $a_i = s_i + \mu\tau, b_i = f_i + (1 - \mu)\tau, i = 1, \dots, \ell$. Also let

$$c_i = \frac{\theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j - \omega_\ell}{\omega_i}$$

and

$$d_i = \frac{\theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j}{\omega_i}, i = 1, \dots, \ell - 1.$$

Then,

$$p(\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0) \propto \pi_i^{a_i-1} (1 - \pi_i)^{b_i-1} (\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}, \quad (10)$$

$c_i < \pi_i < d_i, i = 1, \dots, \ell - 1$. Note that this density function consists of two terms $\pi_i^{a_i-1} (1 - \pi_i)^{b_i-1}$ and $(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}$; note the interchange between a_ℓ and b_ℓ in the second term. The first term is the conditional posterior density under the unrestricted model, and the second term is a generalized beta density [*i.e.*, a beta(b_ℓ, a_ℓ) distribution in the interval (c_i, d_i)]. Thus, the unrestricted beta density is adjusted by the generalized beta density. In the rest of the paper we denote by $\text{GenBeta}(a, b, c, d)$ the generalized beta random variable with density function,

$$p(x) = (x - c)^{a-1} (d - x)^{b-1} / \{(d - c)^{a+b-1} B(a, b)\},$$

$$c \leq x \leq d, a > 1, b > 1.$$

That is, $(X - c) / (d - c) \sim \text{Beta}(a, b)$ if and only if $X \sim \text{GenBeta}(a, b, c, d)$.

It is worth noting that we have ordered the areas in order of their counts (smallest to largest). This is convenient and advantageous both theoretically and computationally.

In order to explain the gain in precision, we attempt to study (10) further by making two approximations. First, because the restriction under study is rather mild we do not expect c_i to be much different from 0 and d_i to be much different from 1. Under this assumption, we can approximate (10) by

$$p_a(\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0) \propto (\pi_i - c_i)^{a_i-1} (d_i - \pi_i)^{b_i-1} (\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1},$$

$$c_i < \pi_i < d_i.$$

Then, incorporating the normalization constant into $p_a(\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0)$, we have

$$p_a(\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0) = \frac{(\pi_i - c_i)^{a_i-1} (d_i - \pi_i)^{b_i-1} (\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}}{\int_{c_i}^{d_i} (\pi_i - c_i)^{a_i-1} (d_i - \pi_i)^{b_i-1} (\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1} d\pi_i}$$

$$= \frac{(\pi_i - c_i)^{a_i-1} (d_i - \pi_i)^{b_i-1}}{(d_i - c_i)^{a_i+b_i-1} B(a_i, b_i)}$$

$$\times \frac{(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}}{E[(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}]}, c_i < \pi_i < d_i, \quad (11)$$

where the expectation is taken over the generalized Beta distribution $\pi_i \sim \text{GenBeta}(a_i, b_i, c_i, d_i), i = 1, \dots, \ell - 1$. But under this latter density, $(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}$ is an unbiased estimator of $E[(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}]$. In addition, by construction a_ℓ and b_ℓ are relatively large and therefore $(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}$ and its variance are expected to be small. Then, our second approximation is

$$(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1} \approx E[(\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}]. \quad (12)$$

Therefore, combining (11) and (12), our final approximation of (10) is

$$\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0 \sim \text{GenBeta}(a_i, b_i, c_i, d_i). \quad (13)$$

It follows from (13) that

$$E_r(\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0) \approx c_i + (d_i - c_i) E_u(\pi_i | \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{s})$$

and

$$\text{Var}_r(\pi_i | \boldsymbol{\pi}_{(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{s}, \phi = 0) \approx (d_i - c_i)^2 \text{Var}_u(\pi_i | \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{s}), \quad (14)$$

where u refers to the unrestricted model and r restricted model. Note that when $c_i = 0$ and $d_i = 1$, we get $E_r(\pi_i | \cdot) = E_u(\pi_i | \cdot)$ and $\text{Var}_r(\pi_i | \cdot) = \text{Var}_u(\pi_i | \cdot)$. Generally though the estimates of π_i will be a bit different from one scenario to the other. It is also interesting that $\text{Var}_r(\pi_i | \cdot) \leq \text{Var}_u(\pi_i | \cdot)$ at least approximately. Thus, the restriction $\sum_{i=1}^{\ell} \omega_i \pi_i = \theta$ will reduce variability, when the π_i are estimated. This is true because the $\pi_i, i = 1, \dots, \ell$, belong to an $\ell - 1$ dimensional simplex in the ℓ dimensional hypercube while for the unrestricted model $\pi_i, i = 1, \dots, \ell$, belong to the ℓ dimensional hypercube. We expect the largest gain in precision when θ is completely specified, followed by the case when μ_0 is specified and $\tau_0 \gg 2$, and the least gain in precision when $\theta \sim \text{Uniform}(0, 1)$.

2.3 Computation

We show how to draw samples from the unrestricted and restricted models. For the unrestricted model we are able to draw random samples from (7) without using Markov chain Monte Carlo methods. However, for the restricted model we use the griddy Gibbs sampler (Ritter and Tanner 1992) to draw samples from (9).

2.3.1 Unrestricted model

We collapse over the π_i , draw samples from $p(\boldsymbol{\mu}, \boldsymbol{\tau} | \boldsymbol{s})$ using random draws from a bivariate grid, and finally obtain samples from the Rao-Blackwellized densities $\pi_i | \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{s}$.

Then,

$$\pi_i | \mu, \tau, s \sim \text{Beta}\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}, i = 1, \dots, \ell, \quad (15)$$

and integrating out π , we get

$$p(\mu, \tau | s) \propto \prod_{i=1}^{\ell} \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \times \frac{1}{(1 + \tau)^2},$$

$0 < \mu < 1, \tau > 0$. Letting $\delta = \tau / \tau + 1$, we have

$$p(\mu, \delta | s) \propto \left[\prod_{i=1}^{\ell} \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \right]_{\tau = \frac{\delta}{1 - \delta}}, \quad 0 < \mu, \delta < 1.$$

First we draw $\mu, \delta | s$ using a bivariate grid on $(0, 1)^2$ to obtain a sample of $M \approx 10,000$ values of $(\mu^{(h)}, \delta^{(h)})$, $h = 1, \dots, M$, $\tau^{(h)} = \delta^{(h)} / (1 - \delta^{(h)})$. Then we perform a data augmentation in (15) to obtain $\pi^{(h)}$, $h = 1, 2, \dots, M$, using a composition method. That is, we simply draw $\pi_i \sim \text{Beta}\{s_i + \mu^{(h)}\tau^{(h)}, f_i + (1 - \mu^{(h)})\tau^{(h)}\}$, $i = 1, \dots, \ell$, $h = 1, \dots, M$.

To perform the bivariate grid method for sampling from the posterior density of (μ, δ) , we divide the interval $(0, 1)$ into 100 sub-intervals; so there are 10,000 little squares in the original unit square. We obtain the heights of the posterior density (without the normalization constant) at the center of each of the 10,000 squares. Because these little squares have the same area, the heights of the bivariate density are proportional to the posterior probabilities that (μ, δ) fall in each of these squares. Thus, we have constructed a joint posterior probability mass function of (μ, δ) on very fine grids. It is easy to draw a sample from the discrete bivariate probability mass function by using the cumulative distribution method. This is actually a random draw of one of the 10,000 squares with probabilities proportional to the heights of the little squares. Then within the selected square we choose a point at random by drawing two uniform random variables (*i.e.*, uniform random jittering). Indeed, this is a very accurate random draw from the joint posterior density of (μ, δ) . We draw $M = 10,000$ samples from this approximation for posterior inference in a standard Monte Carlo procedure with independent samples, not a Markov chain. Because of the random jittering the numbers are different with probability one.

2.3.2 Restricted model

We show how to draw samples from the restricted model using the Gibbs sampler. The joint conditional posterior density of $\pi_1, \dots, \pi_{\ell-1}$ is

$$p(\pi_1, \dots, \pi_{\ell-1} | \mu, \tau, \theta, s, \phi = 0) \propto \prod_{i=1}^{\ell-1} \left\{ \pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1 - \mu)\tau - 1} \right\} \times \left(\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i \right)^{s_{\ell} + \mu\tau - 1} \left\{ \sum_{i=1}^{\ell-1} \omega_i \pi_i - \theta + \omega_{\ell} \right\}^{f_{\ell} + (1 - \mu)\tau - 1} \quad (16)$$

where

$$\theta < \omega_{\ell}, \theta - \omega_{\ell} < \sum_{i=1}^{\ell-1} \omega_i \pi_i < \theta, \pi_{\ell} = \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}}.$$

Thus, we would obtain samples of $\pi_1, \dots, \pi_{\ell-1}$ and we set

$$\pi_{\ell} = \frac{\left(\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i \right)}{\omega_{\ell}}$$

to complete the vector π_1, \dots, π_{ℓ} . That is, the constraint is obtained exactly. The conditional posterior density of θ is

$$p(\theta | \pi_{(\ell)}, \mu, \tau, s, \phi = 0) \propto \left\{ \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i \right\}^{s_{\ell} + \mu\tau - 1} \left\{ \omega_{\ell} + \sum_{i=1}^{\ell-1} \omega_i \pi_i - \theta \right\}^{f_{\ell} + (1 - \mu)\tau - 1} \times \theta^{\mu_0\tau_0 - 1} (1 - \theta)^{(1 - \mu_0)\tau_0 - 1}, \quad (17)$$

where

$$\sum_{i=1}^{\ell-1} \omega_i \pi_i < \theta < \omega_{\ell} + \sum_{i=1}^{\ell-1} \omega_i \pi_i.$$

The joint conditional posterior density of μ and τ is

$$p(\mu, \tau | \pi_{(\ell)}, \theta, s, \phi = 0) \propto \frac{q^{\mu\tau} r^{(1 - \mu)\tau}}{[B(\mu\tau, (1 - \mu)\tau)]^{\ell}} \times \frac{1}{(1 + \tau)^2}, \quad (18)$$

$0 < \mu < 1, \tau > 0, q = \prod_{i=1}^{\ell} \pi_i, r = \prod_{i=1}^{\ell} (1 - \pi_i)$.

To perform the Gibbs sampler, we need to draw samples from (16), (17) and (18), each in turn, until convergence. We draw μ, τ from $p(\mu, \tau | \pi_{(\ell)}, \theta, s)$ in a manner similar to drawing from $p(\mu, \tau | \pi_{(\ell)})$ in the unrestricted model. It is more difficult to draw sample from (16) and (17). However, we use essentially the same method to draw samples from the conditional posterior density of $\pi_i, i = 1, \dots, \ell - 1$, obtained from (16) and θ from (17) which are both proportional to the product of two density functions, one is a truncated beta density and the other a generalized beta density. We next develop some theory to draw a sample from such a density. For this purpose, we state and prove Lemma 2 and Theorem 2.

The density function of interest is

$$f(x) = Af_1(x)f_2(x), 0 \leq c < x < d \leq 1, \quad (19)$$

where

$$f_1(x) = \frac{x^{g-1}(1-x)^{h-1}}{\int_c^d x^{g-1}(1-x)^{h-1} dx}, c < x < d, g, h > 0, \quad (20)$$

$$f_2(x) = (x-c)^{a-1}(d-x)^{b-1} / \{(d-c)^{a+b-1} B(a, b)\}, \\ c < x < d, a, b > 1, \quad (21)$$

and, of course,

$$A = 1 / \int_c^d f_1(x)f_2(x) dx. \quad (22)$$

It is worth noting that we are not assuming $g, h > 1$. If this was the case, then $f_1(x)$ and $f_2(x)$ will be both log-concave, thereby making $f(x)$ logconcave, and in this case one can draw a sample from $f(x)$ using the adaptive rejection sampler (ARS, Gilks and Wild 1992). We are providing a specialized algorithm to draw a sample from $f(x)$ which is not logconcave. Even if $f_1(x)$ was logconcave (*i.e.*, $g, h > 1$) this specialized algorithm will still be better than the ARS because the ARS is a general purpose algorithm; see Robert and Casella (1999, page 59). Our algorithm requires less computation and does not need logconcavity; even if there is logconcavity the ARS can perform poorly in the tails of the density function.

Lemma 2 Consider the density functions $f_1(x)$ and $f_2(x)$ with $a, b > 1$.

(a) *Then*

$$\sup_{c < x < d} f_2(x) = \frac{\delta^{a-1}(1-\delta)^{b-1}}{(d-c)B(a, b)}, \delta = (a-1) / (a+b-2).$$

(b) *For any $g > 0, h > 0$ there exist two constants H_1 and H_2 such that*

$$0 < H_1 \leq A^{-1} \leq H_2 < \infty.$$

A proof of Lemma 2 is given in Appendix A.

Theorem 2 Let $F_{g,h}(\cdot)$ be the cdf of Beta(g, h) random variable and $F_{g,h}^{-1}(\cdot)$ be its inverse. Let

$$U, V \stackrel{\text{ind}}{\sim} \text{Uniform}(0, 1),$$

and let

$$X = F_{g,h}^{-1}\{UF_{g,h}(d) + (1-U)F_{g,h}(c)\}.$$

If for two real numbers $a, b > 1$,

$$V \leq \frac{1}{(d-c)^{a+b-2}} \left(\frac{X-c}{\delta} \right)^{a-1} \left(\frac{d-X}{1-\delta} \right)^{b-1},$$

where $\delta = (a-1) / (a+b-2)$, then X has the density $f(x) = Af_1(x)f_2(x)$.

A proof of Theorem 2 is given in Appendix A.

Theorem 1 gives us the following algorithm for drawing samples from $f(\pi) \propto \pi^{g-1}(1-\pi)^{h-1}(\pi-c)^{a-1}(d-\pi)^{b-1}$, $c < \pi < d, g, h > 0, a, b > 1$.

Algorithm

(a) Draw $U \sim \text{Uniform}(0, 1)$ and set

$$\pi = F_{g,h}^{-1}\{UF_{g,h}(d) + (1-U)F_{g,h}(c)\}.$$

(b) Draw $V \sim \text{Uniform}(0, 1)$. If

$$V \leq \frac{1}{(d-c)^{a+b-2}} \left(\frac{\pi-c}{\delta} \right)^{a-1} \left(\frac{d-\pi}{1-\delta} \right)^{b-1},$$

accept π , otherwise go to (a).

Because the binomial sample sizes are arranged in increasing order, in any application it will be true that $a, b > 1$ and $g, h > 0$ (possibly greater than 1 as well). Thus, the algorithm will work. Indeed, in all our examples (one presented here) and simulation exercises the algorithm runs very quickly.

Now, we show how to draw $\pi_i, i = 1, \dots, \ell$, and θ . For π_i ,

$$p(\pi_i | \boldsymbol{\pi}_{(i,\ell)}, \theta, \mu, \tau, s, \phi = 0)$$

$$\propto \pi_i^{a_i-1} (1-\pi_i)^{b_i-1} (\pi_i - c_i)^{b_\ell-1} (d_i - \pi_i)^{a_\ell-1}, c_i < \pi_i < d_i,$$

where $\boldsymbol{\pi}_{(i,\ell)}$ is the vector containing the elements of $\boldsymbol{\pi}$ except for π_i and π_ℓ , and $a_i = s_i + \mu\tau, b_i = f_i + (1-\mu)\tau, i = 1, \dots, \ell$,

$$c_i = \left(\theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j - \omega_\ell \right) / \omega_i,$$

$$d_i = \left(\theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j \right) / \omega_i, i = 1, \dots, \ell - 1.$$

Apply the theorem to $p(\pi_i | \boldsymbol{\pi}_{(i)}, \theta, \mu, \tau, s), a_\ell > 1, b_\ell > 1, i = 1, \dots, \ell - 1$.

For θ , we have

$$p(\theta \mid \boldsymbol{\pi}, \mu, \tau, s, \phi = 0) \propto \theta^{\mu_0 \tau_0 - 1} (1 - \theta)^{(1 - \mu_0) \tau_0 - 1} (\theta - \tilde{c})^{a_\ell - 1} (\tilde{d} - \theta)^{b_\ell - 1}, \tilde{c} < \theta < \tilde{d},$$

where

$$\tilde{c} = \sum_{i=1}^{\ell-1} \omega_i \pi_i, \tilde{d} = \omega_\ell + \sum_{i=1}^{\ell-1} \omega_i \pi_i.$$

Again, apply the theorem, $a_\ell > 1, b_\ell > 1$.

When θ is fully specified (*i.e.*, θ is not random), we do not have to draw θ . However, when $\theta \sim \text{Uniform}(0, 1)$ a priori ($\mu_0 = 1/2, \tau_0 = 2$), we have a simplification. In this case,

$$\theta \mid \boldsymbol{\pi}_{(\ell)}, \mu, \tau, s, \phi = 0 \sim \text{GenBeta}(a_\ell, b_\ell, \tilde{c}, \tilde{d})$$

and $\theta = \tilde{c} + (\tilde{d} - \tilde{c})X$, where $X \sim \text{Beta}(a_\ell, b_\ell)$, has the required density.

For both the unrestricted and restricted models we use 10,000 iterates to make posterior inference about the binomial probabilities, π_i . Under the unrestricted model these are simply random draws and no monitoring is required. For the restricted model, running the griddy Gibbs sampler, we drew 11,000 iterates, used 1,000 as a “burn in” (a conservative number because convergence occur much earlier as evident in the trace plots) and we found negligible correlations among the iterates. Thus, we used 10,000 iterates to make inference about the binomial probabilities. For both the unrestricted and the three restricted models it takes only a few seconds on our 2×833 MHz alpha computer.

3. Numerical studies

In Section 3.1 we describe an illustrative example to show the main features of the restriction. In Section 3.2 we describe a simulation study to show frequentist properties of the Bayes estimators, and we show deeper insight into the differences among the four scenarios. Note again that when we performed the computations, it is convenient to order the domain sizes so that the largest domain comes last.

3.1 Illustrative example

We have used data in the third National Health and Nutrition Examination (NHANES) Survey to illustrate our method. We have studied body mass index for teenagers, and we have data on the sample obtained. The domains (small areas) are formed by crossing ethnicity (white, black, Mexican) and sex (male, female). We have separated out the teenagers with respect to whether they were in middle

school or high school at the time of the survey. Thus, there are 12 small domains. The data are presented in the first four columns of Table 1 by domain. Note that domains MWM, MBF, MWF and HBF are relatively sparse with 4, 2, 5, 5 obese teenagers respectively; for the twelve domains the sample consists of 959 with 130 obese teenagers (*i.e.*, the overall proportion of obese individuals is 0.136 approximately). In column 4 of Table 1 we have also presented the direct estimates by domains, and these estimates range from 0.069 to 0.228. The estimates for the smallest domains will be unreliable. Moreover, when the beta-binomial models are used, these estimates will regress to the overall sample mean of 0.136, creating a possible bias. Our method is expected to increase precision beyond the unrestricted model because the restricted model uses more information about the weighted sum. Clearly, predictors based on either the restricted model or the beta-binomial model are biased if the specified model is wrong.

We have taken $\mu_0 = 0.136$, the overall sample proportion, and $\tau_0 = 959$, the total sample size. Less optimistic choices can be used. For example, $\tau_0 = 100$, say; but this choice makes very little difference. However, it is worth noting that using the observed data to specify the prior distribution can artificially decrease the posterior variance. Typically a survey practitioner will have an appropriate specification from a prior survey or a census. One cannot specify values for μ_0 and τ_0 which are completely out of line and will create huge biases. Here τ_0 is a prior sample size and μ_0 is a prior mean of θ . This method permits a sensible value for θ ; we are essentially adding a degree of uncertainty about knowledge of the linear combination. Thus, these specifications are not unreasonable.

We have applied our method as described for the four scenarios. In the other columns of Table 1 we study the estimates of the small area probabilities. We present the posterior mean (PM), posterior standard deviation (PSD),

$$\text{RMSE} = \sqrt{(\hat{\pi} - \text{PM})^2 + \text{PSD}^2},$$

where $\hat{\pi}$ is the direct estimate, and the 95% highest posteriori density (HPD) interval (Int). As is expected, the PSDs are roughly in the increasing order: Model 2, Model 3, Model 4 and Model 1; in some cases the differences are important. The PMs for Models 1, 2 and 3 are mostly similar, but for Model 4 the PMs are mostly smaller than the other three models. There is much improvement of Models 2 and 3 over Model 1 at least in terms of precision. This gain becomes less important for Model 4, the model with the greatest uncertainty about θ .

Table 1
Comparison of the four models using posterior mean (PM), posterior standard deviation, root mean square error (RMSE), and 95% credible HPD intervals (Int) of π_i by domain (D) for the NHANES data

D	s	n	$\hat{\pi}$	PM	PSD	RMSE	Int	PM	PSD	RMSE	Int
				Model 1				Model 2			
1	4	47	0.085	0.114	0.033	0.044	(0.051, 0.179)	0.111	0.032	0.041	(0.049, 0.170)
2	2	29	0.069	0.112	0.037	0.057	(0.042, 0.183)	0.111	0.036	0.055	(0.041, 0.178)
3	10	44	0.227	0.175	0.044	0.068	(0.100, 0.264)	0.177	0.041	0.065	(0.108, 0.260)
4	5	62	0.081	0.107	0.030	0.040	(0.047, 0.159)	0.107	0.027	0.038	(0.054, 0.160)
5	10	74	0.135	0.134	0.030	0.030	(0.077, 0.194)	0.134	0.028	0.028	(0.080, 0.190)
6	12	69	0.174	0.158	0.036	0.039	(0.089, 0.227)	0.155	0.031	0.036	(0.095, 0.214)
7	8	79	0.101	0.116	0.028	0.031	(0.065, 0.173)	0.115	0.027	0.030	(0.065, 0.166)
8	5	62	0.081	0.107	0.030	0.040	(0.052, 0.169)	0.105	0.029	0.038	(0.042, 0.153)
9	28	123	0.228	0.196	0.036	0.048	(0.129, 0.262)	0.196	0.032	0.045	(0.131, 0.253)
10	10	111	0.090	0.106	0.026	0.030	(0.059, 0.155)	0.105	0.024	0.028	(0.061, 0.150)
11	16	122	0.131	0.132	0.026	0.026	(0.083, 0.183)	0.130	0.023	0.023	(0.090, 0.179)
12	20	137	0.146	0.144	0.026	0.026	(0.094, 0.194)	0.141	0.022	0.023	(0.100, 0.184)
				Model 3				Model 4			
1	4	47	0.085	0.111	0.033	0.042	(0.044, 0.169)	0.109	0.032	0.040	(0.050, 0.172)
2	2	29	0.069	0.111	0.037	0.056	(0.039, 0.179)	0.108	0.036	0.053	(0.037, 0.173)
3	10	44	0.227	0.175	0.043	0.068	(0.093, 0.260)	0.170	0.044	0.072	(0.091, 0.255)
4	5	62	0.081	0.106	0.029	0.038	(0.050, 0.160)	0.103	0.030	0.038	(0.048, 0.164)
5	10	74	0.135	0.134	0.029	0.029	(0.077, 0.189)	0.129	0.030	0.030	(0.067, 0.184)
6	12	79	0.174	0.156	0.034	0.038	(0.090, 0.217)	0.151	0.036	0.043	(0.087, 0.222)
7	8	69	0.101	0.118	0.028	0.033	(0.062, 0.171)	0.111	0.028	0.029	(0.061, 0.167)
8	5	62	0.081	0.107	0.030	0.040	(0.051, 0.165)	0.102	0.030	0.036	(0.050, 0.159)
9	28	123	0.228	0.195	0.034	0.047	(0.138, 0.265)	0.189	0.035	0.052	(0.123, 0.255)
10	10	111	0.090	0.107	0.024	0.029	(0.062, 0.156)	0.104	0.025	0.029	(0.051, 0.149)
11	16	122	0.131	0.132	0.024	0.024	(0.086, 0.179)	0.126	0.025	0.025	(0.083, 0.179)
12	20	137	0.146	0.143	0.024	0.024	(0.095, 0.191)	0.137	0.025	0.027	(0.091, 0.189)

Note: The four models are: Model 1 - no restriction; Model 2 - fixed θ ; Model 3 - informative prior for θ ; Model 4 - uniform prior for θ . Domains are formed by crossing school (middle school - M, high school - H), race (white - W, black - B, mexican american - M) and sex (male - M, female - F). Thus, the domains are: 1-MWM, 2-MBF, 3-MMM, 4-MWF, 5-MBM, 6-MMF, 7-HWM, 8-HBF, 9-HMM, 10-HWF, 11-HBM, 12-HMF (e.g., the first domain consists of middle school white boys). n is the number of teenagers and s the number of obese teenagers in each domain. Data are taken from the 35 largest counties in the US. An estimate of the overall probability is $130 / 959 \approx 0.136$, and for the first domain $\hat{p} = 4 / 47 = 0.085$; the numerical standard errors are all smaller than 0.001; $RMSE = \sqrt{(\hat{\pi} - PM)^2 + PSD^2}$.

We also study very briefly the nuisance parameter θ . We note that the weighted average of the direct estimators of the small areas is 0.136 (more accurately 0.135599). When θ is held fixed at 0.135599, the weighted average of the posterior means is 0.136. When θ has the informative prior, the weighted average of the posterior means is 0.136, and for θ the PM is 0.136, the PSD is 0.008, and a 95% HPD interval for θ is (0.122, 0.152). When θ has the uniform prior, the weighted average of the posterior means is 0.132, and for θ the PM is 0.131, the PSD is 0.011, and a 95% HPD interval for θ is (0.110, 0.151). This shows the deficiencies of the uniform prior which we use only for comparison. It is worth noting that $\mu_1, \dots, \mu_{\ell-1}$ and θ are computed first. Then μ_ℓ is obtained by subtraction. This is done at each iterate of the Gibbs sampler. Then, the posterior summaries for $\sum_{i=1}^{\ell} \omega_i \pi_i$ and θ are computed. So there will be very minor discrepancies which are due to rounding.

Finally, we have selected the four smallest domains to compare the posterior densities of the probabilities. We have used the Parzen-Rosenblatt kernel density estimator to estimate the posterior densities; see Silverman (1986) for details. Figure 1 compares the estimated posterior densities for the four models. It is interesting that as the domain sizes increase, the four models get closer together. Also, for all cases the tails of the distributions in each panel are very similar; the differences in these distributions though lie in the modal intervals (i.e., interval containing the mode), and their heights. As expected, the posterior density corresponding to the unrestricted model is the shortest, simply because it has more variability. Model 4 has posterior density shifted to the left and is slightly bimodal for the smallest domain. Thus, inference about the modes of these distributions will be different. But inference involving the tails will not be so different; except for Model 4, 95% credible intervals will be similar.

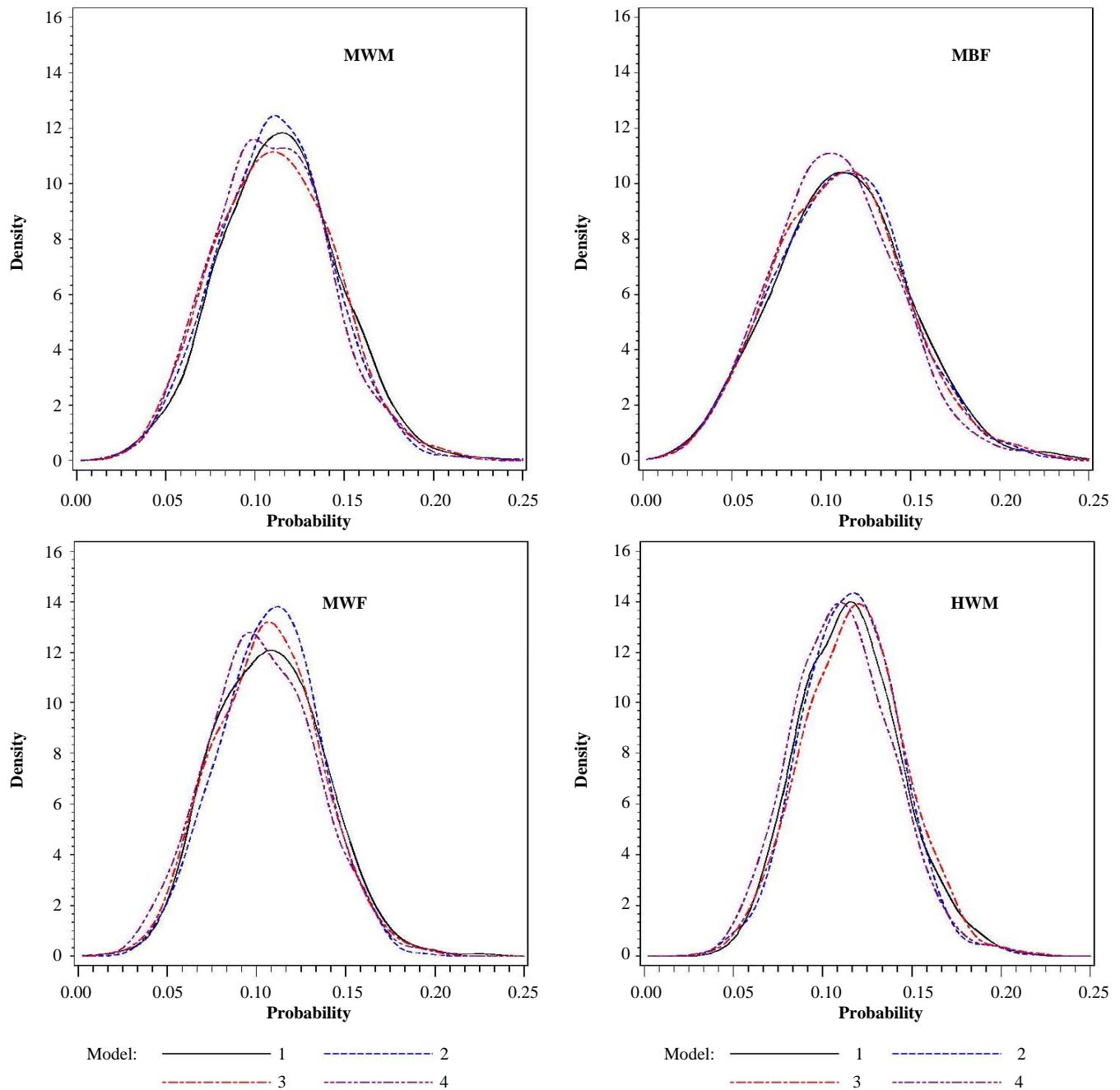


Figure 1 Plots of the estimated posterior densities of $\pi_1, \pi_2, \pi_4,$ and π_7 for the four models and NHANES data

3.2 Simulation study

We use a simulation study to assess the statistical properties of our method. We want to see if the gain in precision persists and to see how the estimators of the probabilities are shifted. We also study the frequentist properties of the estimators of the probabilities. In the description of the simulation it is convenient to use the abbreviated names of the models which are UR (Model 1, no restriction), FI (Model 2, fixed θ), IN (Model 3,

informative prior for θ) and UN (Model 4, uniform prior for θ).

We set $\theta_0 = 0.15, \mu_0 = \theta_0$ and $\tau_0 = 100$. We have selected three values of $\ell = 12, 24, 36, 12$ being the number of areas in the NHANES data. We drew the sample sizes from a uniform density in $(25, 150)$, again to reflect the NHANES data. First, we generated

$$\pi_i \stackrel{iid}{\sim} \text{Beta} \{ \mu_0 \tau_0, (1 - \mu_0) \tau_0 \}, i = 1, \dots, \ell.$$

To do this latter task, we drew sets of $\ell \pi_i$ until $\theta_0 - w_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta_0$; set $\pi_\ell = (\theta_0 - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / w_\ell$. Then, we generated

$$s_i \stackrel{\text{ind}}{\sim} \text{Binomial}(n_i, \pi_i).$$

We have generated 1,000 data sets in this manner for each of $\ell = 12, 24, 36$. Then, we fit the four models (one unrestricted and three restricted models). The process is very fast (*i.e.*, for samples sizes of 12, 24, 30 there were respectively 22, 90, 153 rejects in the 1,000 samples). We fit each data set using random samples for the unrestricted model and the gridy Gibbs sampler for the restricted models. We fit the 1,000 data sets in a couple of hours on our on our 2×833 MHz alpha computer.

For these 1,000 simulations we study PM, the coverage (C), the bias (B), PSD, RMSE and width (W) of the 95% credible intervals. For each domain we compute the bias $PM - \pi$, then we average these values over all domains and simulation runs, and this quantity we now call B . Associated with B we also computed AB , the average of $|PM - \pi|$. Similarly, we have computed

$$\text{RMSE} = \sqrt{(\text{PM} - \pi)^2 + \text{PSD}^2}$$

for each domain and each simulation run and we average these over all domains and simulation runs. Note that the true probabilities, π_i , are known by design. We obtain the coverage (C) by computing the proportion of all intervals containing the true value of π_i over all domains and simulation runs. We also obtain the average of the widths of the 95% credible intervals. Numerical standard errors are obtained for all quantities.

In Table 2 we study the estimates of the small area probabilities. It is convenient to use the shorter names of the four models for our discussion. For IN the PMs are close to the nominal value of 0.15, but for UN the PMs are smaller than the nominal value particularly for UN at $\ell = 12$. We observe that the coverage for all the models UR, FI and UN are always larger than the nominal value of 95%, but for model IN these coverages are smaller than the nominal value of 95%. A similar difference exists for the bias; while the bias is small for all models, models UR, FI (the specified value of θ is 0.15) and UN have negative biases but IN has positive bias. Except for $\ell = 36$ IN has the largest AB. The PSDs are mostly similar and the RMSEs share the same features; there are some differences at $\ell = 12$. The four models get similar as ℓ increases; when ℓ is large there appears to be no need for our method. However, again the gain in precision appears to be in the increasing order FI, IN, UN and UR.

In most applications the exact value of θ is unknown. Therefore, the PSDs of the π_i , under the situation where θ is assumed known, are likely to underestimate the true PSDs. So we study the deviations of the PSDs of IN and UN from those of FI, and we compute the ratios, $R_1 = \text{PSD}_{\text{IN}} / \text{PSD}_{\text{FI}}$ and $R_2 = \text{PSD}_{\text{UN}} / \text{PSD}_{\text{FI}}$. In Table 3 we present the five-number summaries of these ratios by sample size. Most of the ratios are around 1 (*i.e.*, inter-quartile range) with some tendency for them to be larger than 1. (Note that the maxima at $\ell =$ and $\ell = 24$ are outliers possibly due to bad simulated samples.) Thus, overall the PSDs under IN and UN are not much larger under FI.

Table 2
Simulation: Comparison of the four models using coverage (C), bias and average absolute bias (B and AB), posterior standard deviation (PSD), root mean squared error (RMSE) and width of the 95% credible intervals (W) of π_i

ℓ	Model	C	B	AB	PSD	RMSE	W
12	UR	0.960 _{0.0018}	-0.002 _{0.0003}	0.0231 _{0.00016}	0.033 _{0.0001}	0.043 _{0.0001}	0.125 _{0.0003}
	FI	0.961 _{0.0018}	-0.000 _{0.0003}	0.0219 _{0.00020}	0.031 _{0.0001}	0.040 _{0.0001}	0.118 _{0.0003}
	IN	0.946 _{0.0021}	0.005 _{0.0003}	0.0275 _{0.00066}	0.032 _{0.0001}	0.043 _{0.0001}	0.122 _{0.0002}
	UN	0.956 _{0.0019}	-0.000 _{0.0003}	0.0261 _{0.00019}	0.032 _{0.0001}	0.042 _{0.0001}	0.122 _{0.0003}
24	UR	0.957 _{0.0013}	-0.001 _{0.0002}	0.0229 _{0.00012}	0.031 _{0.0000}	0.041 _{0.0001}	0.119 _{0.0002}
	FI	0.957 _{0.0013}	-0.000 _{0.0002}	0.0224 _{0.00013}	0.030 _{0.0000}	0.040 _{0.0001}	0.116 _{0.0002}
	IN	0.943 _{0.0015}	0.006 _{0.0002}	0.0252 _{0.00058}	0.030 _{0.0000}	0.041 _{0.0001}	0.116 _{0.0001}
	UN	0.952 _{0.0014}	-0.000 _{0.0002}	0.0236 _{0.00012}	0.031 _{0.0002}	0.041 _{0.0002}	0.118 _{0.0005}
36	UR	0.960 _{0.0010}	-0.001 _{0.0001}	0.0224 _{0.00009}	0.030 _{0.0000}	0.040 _{0.0001}	0.117 _{0.0001}
	FI	0.961 _{0.0010}	-0.000 _{0.0001}	0.0218 _{0.00009}	0.030 _{0.0000}	0.039 _{0.0001}	0.115 _{0.0001}
	IN	0.948 _{0.0012}	0.005 _{0.0002}	0.0224 _{0.00009}	0.030 _{0.0000}	0.040 _{0.0001}	0.114 _{0.0001}
	UN	0.957 _{0.0011}	-0.000 _{0.0001}	0.0228 _{0.00010}	0.030 _{0.0000}	0.040 _{0.0001}	0.116 _{0.0001}

Note: The four models are: Model 1 - no restriction (UR); Model 2 - fixed θ (FI); Model 3 - informative prior for θ (IN); Model 4 - uniform prior for θ (UN). $\text{RMSE} = \sqrt{(\pi - \text{PM})^2 + \text{PSD}^2}$. The notation a_b means a is an estimate and b is the standard error.

Table 3
Simulation: A study of the posterior standard deviation (PSD) of the π_i using five number summaries of the ratios, R_1 and R_2 , by sample size

ℓ	Ratio	Min	Q_1	Med	Q_3	Max
12	R_1	0.673	0.972	1.032	1.091	5.329
	R_2	0.022	0.984	1.034	1.086	85.370
24	R_1	0.019	0.965	1.005	1.047	16.017
	R_2	0.024	0.979	1.014	1.049	486.960
36	R_1	0.690	0.962	0.998	1.034	1.236
	R_2	0.837	0.979	1.011	1.044	1.243

Note: $R_1 = \text{PSD}_{\text{IN}} / \text{PSD}_{\text{FI}}$ and $R_2 = \text{PSD}_{\text{UN}} / \text{PSD}_{\text{FI}}$. The five summaries are minimum (min), first quartile (Q_1), median (med), third quartile (Q_3) and maximum (max).

In Table 4 we study the estimate of θ for the two pertinent models IN and UN. For both models the coverage probabilities are smaller than the nominal value, and the coverage for UN is smaller than the interval for IN. Bias is small for both models, positive for IN and negative for UN.

Table 4
Simulation: Comparison of the informative (IN) and the uniform (UN) models using posterior mean (PM), coverage (C), bias and average absolute bias (B and AB), posterior standard deviation (PSD), root mean squared error (RMSE) and width of the 95% credible intervals (W) of π_i

ℓ	Model	PM	C	B	AB	PSD	RMSE	W
12	IN	0.149 _{0.0012}	0.853 _{0.0112}	0.000 _{0.0003}	0.00152 _{0.00081}	0.008 _{0.0000}	0.012 _{0.0002}	0.030 _{0.0001}
	UN	0.138 _{0.0005}	0.881 _{0.0102}	-0.012 _{0.0004}	0.00038 _{0.00003}	0.011 _{0.0001}	0.016 _{0.0002}	0.042 _{0.0002}
24	IN	0.153 _{0.0015}	0.833 _{0.0118}	0.003 _{0.0015}	0.00212 _{0.00103}	0.007 _{0.0006}	0.012 _{0.0015}	0.024 _{0.0015}
	UN	0.145 _{0.0029}	0.842 _{0.0115}	-0.005 _{0.0003}	0.00012 _{0.00006}	0.008 _{0.0001}	0.012 _{0.0002}	0.030 _{0.0002}
36	IN	0.150 _{0.0002}	0.828 _{0.0119}	0.000 _{0.0002}	0.00004 _{0.00000}	0.004 _{0.0000}	0.007 _{0.0001}	0.017 _{0.0001}
	UN	0.145 _{0.0003}	0.794 _{0.0128}	-0.005 _{0.0002}	0.00009 _{0.00000}	0.006 _{0.0000}	0.010 _{0.0001}	0.024 _{0.0001}

Note: The two models considered are: Model 3 – informative prior for θ and model 4 - uniform prior for θ . $\text{RMSE} = \sqrt{(\theta_0 - \text{PM})^2 + \text{PSD}^2}$. The notation a_b means a is an estimate and b is the standard error.

Table 5
Simulation: Comparison of the four models using posterior standard deviation and root mean square error (RMSE) of π_i by domain (D)

D	Unrestricted		Fixed		Informative		Uniform	
	PSD	RMSE	PSD	RMSE	PSD	RMSE	PSD	RMSE
1	0.048 _{0.0003}	0.057 _{0.0004}	0.046 _{0.0003}	0.054 _{0.0004}	0.045 _{0.0002}	0.056 _{0.0005}	0.047 _{0.0004}	0.056 _{0.0005}
2	0.046 _{0.0003}	0.055 _{0.0004}	0.044 _{0.0003}	0.053 _{0.0004}	0.044 _{0.0002}	0.054 _{0.0005}	0.045 _{0.0004}	0.054 _{0.0005}
3	0.044 _{0.0002}	0.053 _{0.0004}	0.042 _{0.0002}	0.050 _{0.0004}	0.042 _{0.0002}	0.052 _{0.0005}	0.043 _{0.0003}	0.051 _{0.0004}
4	0.042 _{0.0002}	0.050 _{0.0004}	0.040 _{0.0002}	0.047 _{0.0004}	0.040 _{0.0002}	0.050 _{0.0004}	0.041 _{0.0002}	0.049 _{0.0004}
5	0.041 _{0.0002}	0.049 _{0.0004}	0.038 _{0.0002}	0.046 _{0.0004}	0.039 _{0.0002}	0.048 _{0.0004}	0.039 _{0.0003}	0.048 _{0.0005}
6	0.040 _{0.0002}	0.048 _{0.0004}	0.037 _{0.0002}	0.045 _{0.0004}	0.037 _{0.0002}	0.048 _{0.0004}	0.038 _{0.0003}	0.047 _{0.0005}
7	0.038 _{0.0002}	0.046 _{0.0004}	0.035 _{0.0002}	0.043 _{0.0003}	0.036 _{0.0002}	0.046 _{0.0004}	0.037 _{0.0003}	0.045 _{0.0004}
8	0.037 _{0.0002}	0.045 _{0.0003}	0.034 _{0.0002}	0.041 _{0.0003}	0.036 _{0.0002}	0.046 _{0.0004}	0.036 _{0.0003}	0.044 _{0.0004}
9	0.036 _{0.0002}	0.044 _{0.0003}	0.033 _{0.0002}	0.040 _{0.0004}	0.034 _{0.0001}	0.044 _{0.0004}	0.035 _{0.0003}	0.042 _{0.0004}
10	0.035 _{0.0002}	0.043 _{0.0003}	0.032 _{0.0002}	0.039 _{0.0003}	0.034 _{0.0001}	0.044 _{0.0004}	0.034 _{0.0003}	0.042 _{0.0004}
11	0.034 _{0.0001}	0.042 _{0.0003}	0.031 _{0.0002}	0.038 _{0.0003}	0.033 _{0.0001}	0.042 _{0.0004}	0.033 _{0.0003}	0.041 _{0.0004}
12	0.035 _{0.0002}	0.047 _{0.0005}	0.031 _{0.0002}	0.042 _{0.0004}	0.034 _{0.0003}	0.047 _{0.0006}	0.034 _{0.0007}	0.046 _{0.0008}

Note: The four models are: Model 1 - no restriction; Model 2 - fixed θ ; Model 3 - informative prior for θ ; Model 4 - uniform prior for θ . $\text{RMSE} = \sqrt{(\pi_i - \text{PM})^2 + \text{PSD}^2}$. The notation a_b means a is an estimate and b is the standard error. Here 12 domains are used and the original simulated sample sizes are divided by 2.

Except for $\ell = 36$ IN has by far the larger AB. The PSDs and RMSEs are generally smaller for IN, and the widths of the 95% credible intervals are significantly smaller for IN. It appears that it is difficult to estimate θ under UN, but IN appears to be somewhat better.

In Table 5 we present more detailed result (*i.e.*, by domain) for the case when the number of domains is 12. To show further gains in precision, we have reduced the sample size to half as much [*i.e.*, we drew the sample sizes uniformly in the interval (12, 75)]. We present the posterior standard deviation and the posterior root mean square error, averaged over the simulation runs. Again the standard errors are presented. We note that all the probability contents (not presented) are at least the nominal value of 95%. The numerical standard errors are small in all cases. The PSDs and RMSEs are in the right order. Note that because the sample sizes are arranged in order from smallest to largest, there is a decrease in the PSDs and RMSEs as the domain numbers go up.

We study the posterior density of π_1 for $\ell = 12$, and we compare the four models. Again we use the Parzen-Rosenblatt density estimator. In Figure 2 we present the estimated posterior densities (Parzen-Rosenblatt) averaged over the 1,000 runs for $\ell = 12$. We obtain the same results as for the BMI data. Again the tails are similar. FI is the tallest density and UN is the shortest. UN is slightly shifted to the left of IN. In Figure 3 we present a systematic sample of 10 densities from the 1,000 simulation runs by model. We can see large variation among the 10 estimated posterior densities. Again we can see that FI is tallest; UR, FI and UN show similar variation with IN slightly taller. Thus, it is important to take the average for comparison as in Figure 2.

4. Concluding remarks

We have extended the beta-binomial model of small area estimation to accommodate a prior specification of a weighted average of the area probabilities. We have used the Bayesian approach which is particularly attractive for problems with awkward likelihood functions as in our application with the constraint of the weighted average of the beta-binomial model. We viewed the constraint as prior knowledge which can be precise or less informative. The griddy Gibbs sampler is used to fit the models, thereby avoiding the more sophisticated Metropolis-Hastings sampler. We have developed a theory which permits sampling from a density function which is proportional to the product of a truncated beta-binomial density and a generalized beta density. We have found that overall our complete algorithm forming the griddy Gibbs sampler runs efficiently and fast.

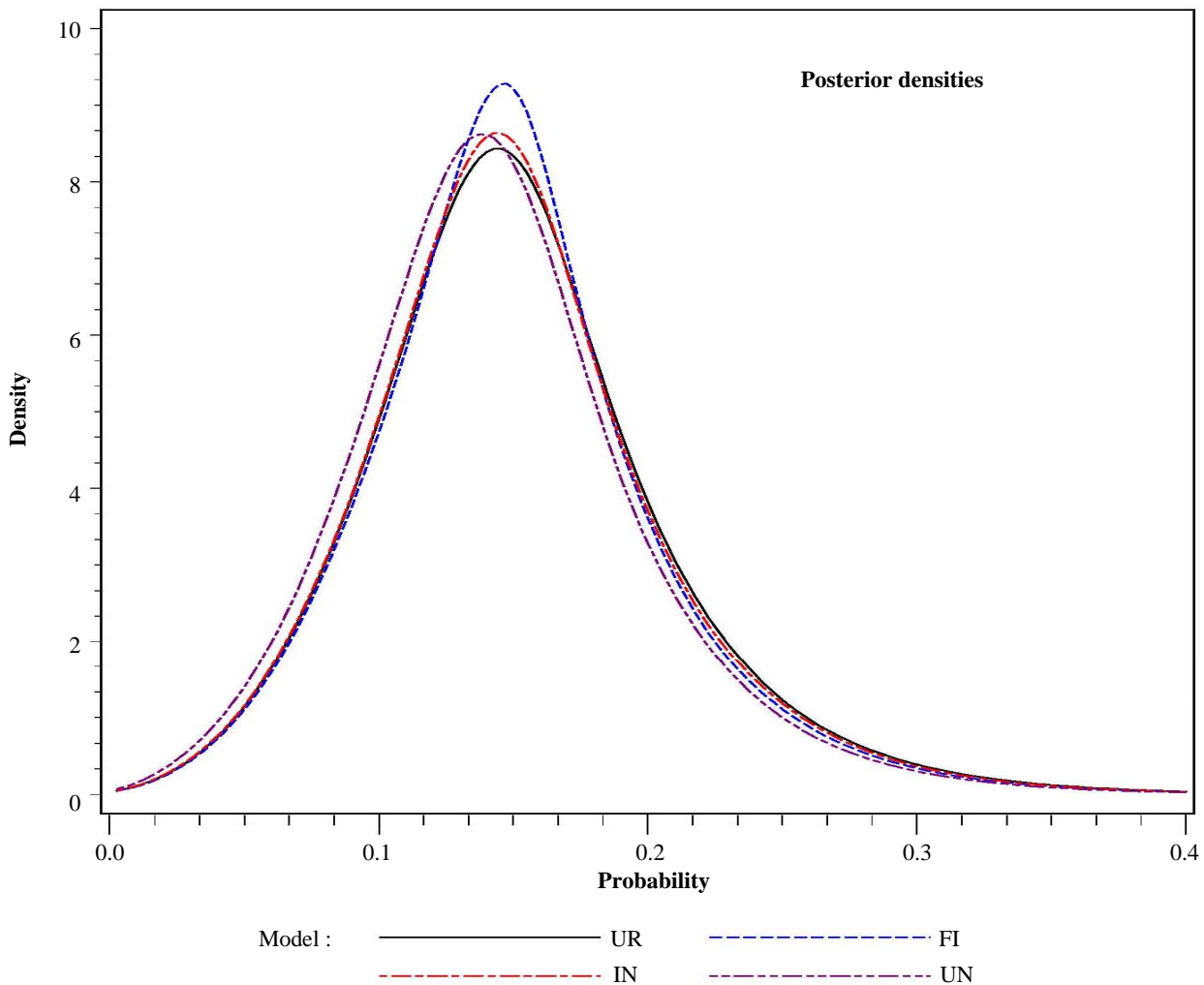


Figure 2 Plots of the estimated posterior densities of π_1 by model when there are 12 domains

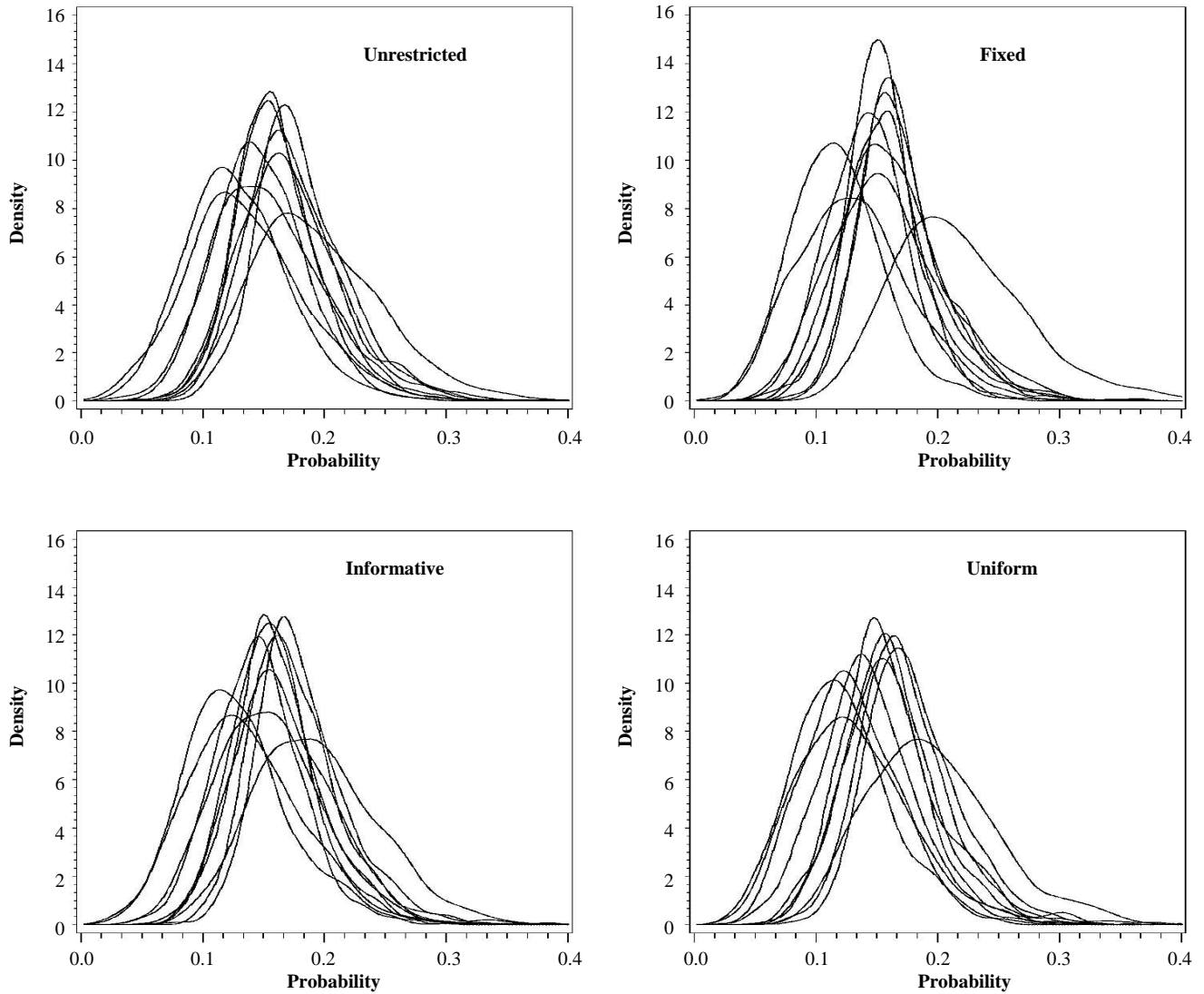


Figure 3 Plots of the estimated posterior densities of π_1 for a systematic sample of size 10 from the 1,000 runs by model when there are 12 domains

We have shown that there could be gains in precision when extra information is incorporated into the beta-binomial model. We have considered three scenarios in which a survey practitioner (a) can not specify any constraint (standard beta-binomial model for small areas), (b) can specify a constraint and the parameter completely, and (c) can specify a constraint and information which can be used to construct a prior distribution for the parameter. Our example on obesity of children in the National Health and Nutrition Examination Survey and simulation study showed that the gain in precision beyond (a) is in an order with (b) larger than (c). As the exact algebraic arguments are difficult, we obtained an analytical approximation which shows that indeed there could be gain in precision of (b) over (a). For comparison we have considered a fourth scenario in which θ has vague information, and as expected, it turned to be rather uninteresting and inefficient.

It is straight forward to make Bayesian predictive inference about the finite population mean of each small area. Let $P_i = T_i / N_i$ denote the finite population proportion for the i^{th} area, where $T_i = \sum_{j=1}^{N_i} y_{ij}$, y_{ij} are the binary responses, and N_i , the number of individuals in the i^{th} area, is assumed known. Now $T_i = t_i^{(s)} + t_i^{(ns)}$, where $t_i^{(s)}$ and $t_i^{(ns)}$ are respectively the sample total and the nonsample total. Now under any of the models $t_i^{(ns)} | \pi_i \sim \text{Binomial}(n_i, \pi_i)$ and $p(t_i^{(ns)} | \mathbf{y}_s) = \int p(t_i^{(ns)} | \pi_i) p(\pi | \mathbf{y}_s) d\pi_i$, where $\mathbf{y}_s = (y_1, \dots, y_\ell)'$. Thus, it is easy to obtain the empirical posterior density of P_i using a sampling-based method. Nandram and Sedransk (1993) obtained some analytical features of P_i when τ is known, but not with the constraint; see also Nandram (1998).

We mention a generalization of our restricted beta-binomial hierarchical Bayesian model to the Dirichlet-multinomial model (e.g., Nandram 1998). Let \mathbf{y}_i be c -vector of

cell counts (*i.e.*, number of people possessing one of c traits), and let n_i denote the sample sizes within the i^{th} area, $i = 1, \dots, \ell$. We assume

$$y_i | \pi_i \sim \text{Multinomial}(n_i, \pi_i), \pi_i | \mu, \tau, \theta \stackrel{\text{iid}}{\sim} \text{Dirichlet}(\mu\tau)$$

with $\sum_{i=1}^{\ell} w_i \pi_i = \theta$. Finally $\theta \sim \text{Dirichlet}(\mu_0\tau_0)$, where μ_0 and τ_0 are to be specified, and independently $p(\mu, \tau) = (k-1)!/(1+\tau)^2, 0 < \mu_k < 1, k = 1, \dots, c, \sum_{k=1}^c \mu_k = 1$. With k constraints this problem is much more complex, but we plan to work on it. Other extensions to nonignorable non-response (Nandram and Choi 2002) and two-way categorical tables are possible.

Acknowledgements

The authors are grateful to the Associate Editor and the two referees who helped enormously to improve the quality of the presentation.

Appendix A

Proofs of lemmas 1, 2 and theorems 1, 2

Proof of lemma 1

This is a special case of a general result. Using the multiplication rule and because the prior is proper, it is clear that the joint density of π, μ, τ, s “integrates” to one. Therefore, the joint posterior density of π, μ, τ given s is proper.

Proof of theorem 1

Let $\mathcal{T} = \{(\pi, \mu, \tau, \theta): 0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0, \theta - \omega_{\ell} \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta, 0 < \theta < 1, \pi_{\ell} = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_{\ell}\}$ and $\mathcal{T}^* = \{(\pi, \mu, \tau): 0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0\}$; note that $\mathcal{T} \subset \mathcal{T}^*$.

Let $\tilde{g}(\pi, \mu, \tau | s)$ denote the right-hand side of the unrestricted posterior density in (7) and $\tilde{p}(\pi_{(\ell)}, \mu, \tau, \theta | s, \phi = 0)$ denote the right-hand side of the restricted posterior density in (9). Noting that $\pi_{\ell} = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_{\ell}$, we observe that

$$\tilde{p}(\pi_{(\ell)}, \mu, \tau, \theta | s, \phi = 0) = \tilde{g}(\pi, \mu, \tau | s) \times \theta^{\mu_0\tau_0-1} (1-\theta)^{(1-\mu_0)\tau_0-1}, (\pi, \mu, \tau, \theta) \in \mathcal{T}.$$

Because $\theta^{\mu_0\tau_0-1} (1-\theta)^{(1-\mu_0)\tau_0-1}$ is proportional to the density function of beta random variable, we have

$$\int_{\mathcal{T}} \tilde{p}(\pi_{(\ell)}, \mu, \tau, \theta | s, \phi = 0) d\pi d\mu d\tau d\theta = A \int_{\mathcal{T}^*} \tilde{g}(\pi, \mu, \tau | s) d\pi d\mu d\tau,$$

where $A = B\{\mu_0\tau_0, (1-\mu_0)\tau_0\}$ is the beta function. By lemma 1, $\int_{\mathcal{T}^*} \tilde{g}(\pi, \mu, \tau | s) d\pi d\mu d\tau < \infty$. Thus, $p(\pi_{(\ell)}, \mu, \tau, \theta | s, \phi = 0)$ is proper.

Proof of Lemma 2 (a)

This can be proved in two ways. The second derivative of $\log\{f_2(x)\}$ is negative in (c, d) , and so the first derivative, when set to zero, provides a unique mode which is $\delta d + (1-\delta)c$. Alternatively, because $(X-c)/(d-c) \sim \text{Beta}(a, b)$ with $a, b > 1$, there is a unique mode for $(X-c)/(d-c)$, and this translates to $\delta d + (1-\delta)c$; note that $\delta d + (1-\delta)c$ is a point in (c, d) . Thus, substituting $\delta d + (1-\delta)c$ into $f_2(x)$, we have

$$\sup_{c < x < d} f_2(x) = \delta^{a-1} (1-\delta)^{b-1} / (d-c) B(a, b).$$

Proof of Lemma 2 (b)

Because $a, b > 1, x \geq x-c$ and $1-x \geq d-x$, it is true that

$$A^{-1} \geq D^{-1} \int_c^d (x-c)^{a+g-2} (d-x)^{b+h-2} dx,$$

where $D = (d-c)^{a+b-1} B(a, b) B(g, h) \{F_{g,h}(d) - F_{g,h}(c)\}$ and $F_{g,h}(x)$ is the cdf of a standard beta random variable in $(0, 1)$. Note that because $c < d$ (strictly) and $F_{g,h}(x)$ is monotone increasing in $(0, 1)$, $F_{g,h}(d) - F_{g,h}(c) > 0$ (strictly). By comparison with the generalized beta density [*i.e.*, $\text{Beta}(a+g-1, b+h-1, c, d)$], the integral is $(d-c)^{a+b+g+h-3} B(a+g-1, b+h-1)$. Thus,

$$A^{-1} \geq \frac{(d-c)^{g+h-2} B(a+g-1, b+h-1)}{B(a, b) B(g, h) \{F_{g,h}(d) - F_{g,h}(c)\}} = H_1 > 0.$$

Also, we have

$$A^{-1} \leq \int_c^d f_1(x) \sup_{c < x < d} f_2(x) dx.$$

Then by Lemma 2 (a),

$$A^{-1} \leq \frac{\delta^{a-1} (1-\delta)^{b-1}}{(d-c) B(a, b)} \int_c^d f_1(x) dx = \frac{\delta^{a-1} (1-\delta)^{b-1}}{(d-c) B(a, b)} = H_2 < \infty.$$

Proof of theorem 2

To show the claim, we calculate the cdf $F_X(\cdot)$ of the random variable X defined in the Theorem. We have

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P[F_{g,h}^{-1}\{UF_{g,h}(d) + (1-U)F_{g,h}(c)\} \leq x] \\ &= P[UF_{g,h}(d) + (1-U)F_{g,h}(c) \leq F_{g,h}(x)] \\ &= P[U\{F_{g,h}(d) - F_{g,h}(c)\} \leq F_{g,h}(x) - F_{g,h}(c)] \\ &= P\left[U \leq \frac{F_{g,h}(x) - F_{g,h}(c)}{F_{g,h}(d) - F_{g,h}(c)}\right]. \end{aligned}$$

Now, since $U \sim \text{Uniform}(0, 1)$, from the above expression for $F_X(\cdot)$, we have $F_X(x) = 1$ if $x \geq d$ and $F_X(x) = 0$ if $x \leq c$. When $c \leq x \leq d$, we have

$$F_X(x) = \frac{F_{g,h}(x) - F_{g,h}(c)}{F_{g,h}(d) - F_{g,h}(c)}.$$

This shows that X has the truncated beta density $f_1(x)$ in (20).

Now, looking to use the accept-reject algorithm, consider

$$\frac{f(x)}{f_1(x)} = Af_2(x).$$

By Lemma 2, we have

$$\sup_{c < \pi < d} \left\{ \frac{f(x)}{f_1(x)} \right\} = A \sup_{c < \pi < d} f_2(x) = A \frac{\delta^{a-1}(1-\delta)^{b-1}}{(d-c)B(a, b)} < \infty.$$

Thus, by the accept-reject algorithm, if

$$V \leq \frac{1}{(d-c)^{a+b-2}} \left(\frac{X-c}{\delta} \right)^{a-1} \left(\frac{d-X}{1-\delta} \right)^{b-1},$$

then X has the density $f(x)$ in (19).

References

- Cochran, W.G. (1977). *Sampling Techniques*, third edition. New York: John Wiley & Sons, Inc.
- Gilks, W.R., and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society, Series C*, 41, 337-348.
- Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.
- Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998). Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93, 273-282.
- Hillmer, S.C., and Trabelsi, A. (1987). Benchmarking of economic time series. *Journal of the American Statistical Association*, 82, 1064-1071.
- Lazar, R., Meeden, G. and Nelson, D. (2008). A noninformative Bayesian approach to finite population sampling using auxiliary variables. *Survey Methodology*, 34, 51-64.
- Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, 61, 97-126.
- Nandram, B., and Choi, J.W. (2002). 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). A Bayesian analysis of a proportion under non-ignorable non-response. *Statistics in Medicine*, 21, 9, 1189-1212.
- Nandram, B., and Sedransk, J. (1993). Bayesian predictive inference for a finite population proportion: Two-stage cluster sampling. *Journal of the Royal Statistical Society, Series B*, 55, 399-408.
- Nandram, B., 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* (in press, preprint).
- Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons, Inc.
- Ritter, C., and Tanner, M.A. (1992). The gibbs sampler and the griddy gibbs sampler. *Journal of the American Statistical Association*, 87, 861-868.
- Robert, C.P., and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- Silvapulle, M.J., and Sen, P.K. (2006). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*. New York: John Wiley & Sons, Inc.
- Silverman, B.W. (1986). *Density Estimation*. London: Chapman and Hall.