Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Bayesian pooling for analyzing categorical data from small areas

by Aejeong Jo, Balgobin Nandram and Dal Ho Kim

Release date: June 24, 2021

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2021
Vol. 47, No. 1, pp. 191-213
Statistics Canada, Catalogue No. 12-001-X

Bayesian pooling for analyzing categorical data from small areas

Aejeong Jo, Balgobin Nandram and Dal Ho Kim[1]

Abstract

Bayesian pooling strategies are used to solve precision problems related to statistical analyses of data from small areas. In such cases, the subpopulation samples are usually small, even though the population might not be. As an alternative, similar data can be pooled in order to reduce the number of parameters in the model. Many surveys consist of categorical data on each area, collected into a contingency table. We consider hierarchical Bayesian pooling models with a Dirichlet process prior for analyzing categorical data based on small areas. However, the prior used to pool such data frequently results in an overshrinkage problem. To mitigate for this problem, the parameters are separated into global and local effects. This study focuses on data pooling using a Dirichlet process prior. We compare the pooling models using bone mineral density (BMD) data taken from the Third National Health and Nutrition Examination Survey for the period 1988 to 1994 in the United States. Our analyses of the BMD data are performed using a Gibbs sampler and slice sampling to carry out the posterior computations.

Key Words: Categorical data; Dirichlet process; Nonparametric hierarchical Bayesian pooling; Slice sampling; Small area.

1. Introduction

Many surveys collect categorical data for individual areas, which are then stored in a contingency table. For example, in a typical obesity rate comparison survey, the researcher might classify the measured sample data by the degree of obesity. Then, the regional obesity rate is estimated using the number of samples assigned to each category. In such cases, we need to consider how the precision is affected by the sample size and the number of parameters in the model, particularly for estimations based on small areas (Rao and Molina, 2015). In general, the precision of a model decreases as the number of parameters increases, assuming the same sample data. To prevent this decrease in precision, the constructed model needs to be as simple as possible. That is, the number of parameters must be reduced in the model. However, the model loses the ability to reflect the detailed effects in each area. Another way to resolve the precision problem is to increase the sample size allowed per parameter. That is, we can employ pooling strategies when analyzing categorical data based on small areas.

Interest in pooling methods is growing among researchers. Malec and Sedransk (1992) developed a Bayesian procedure for estimating the mean of an experiment in a set of seemingly similar experiments. They constructed the prior distribution for a location parameter to reflect their assumptions. They identified subsets of parameters, with subscripts indicating the similarity between the subsets, in which there is uncertainty about the composition of the subsets. They specified the prior distribution for a parameter by conditioning on the same subscript in similar experiments. Their flexible prior distribution

1. Aejeong Jo, Researcher, Division for Healthcare Technology Assessment Research, National Evidence-based Healthcare Collaborating Agency, 173, Toegye-ro, Jung-gu, Seoul, 100-705, Korea. E-mail: joaejung@neca.re.kr; Balgobin Nandram, Professor, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, U.S.A. E-mail: balnan@wpi.edu; Dal Ho Kim, Professor, Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu, 702-701, Korea. E-mail: dalkim@knu.ac.kr.

allows the intensity and nature of the pooling to be influenced by the sample data. Later, Evans and Sedransk (1999) proposed a more flexible Bayesian model using covariates. In addition, Evans and Sedransk (2003) provided a fully Bayesian justification for the results of Malec and Sedransk (1992). These three works have since been extended based on the same key concept of specifying a model in which subscripts are used to indicate similar experiments (Consonni and Veronese, 1995 and DuMouchel and Harris, 1983). Also, Dunson (2009) suggested a generalization of the Dirichlet process (DP) proposed by Ferguson (1973) that allows for dependent local pooling and the borrowing of information. The goal is to borrow information in order to more efficiently estimate the individual functions. The proposed process for local pooling offers a simple, but flexible approach to specifying the local selection process. They suggest using slice sampling, proposed by Walker (2007), to carry out the posterior computations. This is a simple and efficient method that allows for posterior computations for an infinite-dimensional process that is similar to those of a finite-dimensional process. Here, we construct a pooling model using these basic concepts for data based on small areas. Recently, Nandram, Zhou and Kim (2019) proposed a pooled Bayes test of independence for sparse contingency tables. They constructed the model based on a Dirichlet-multinomial hierarchical Bayesian model, see also Nandram (1998) who constructed a prior using the Dirichlet distribution for pooling the data in the models. Of course, a DP is assumed for the parameters of interest, which is the cell probability parameters in our contingency table.

In this study, we use bone mineral density (BMD) data, taken from the Third National Health and Nutrition Examination Survey (NHANES III) for the six-year period from October 1988 to September 1994. BMD is the quantity of mineral in bone tissue, measured as the optical density per cm^2 of bone surface using medical imaging. BMD is used in clinical arenas as an indirect indicator of osteopenia, osteoporosis, fractures, and so on. BMD is statistically correlated with the probability of fractures, which are an important public health problem, especially in elderly women. Therefore, BMD data are important indicators used to identify osteoporotic patients who might benefit from early management to improve their bone strength.

NHANES III contains clinical data on 33,994 people who participated in the survey and is sampled for individual areas. Each person is categorized into three BMD levels: (1) normal, (2) osteopenia and (3) osteoporosis. Our study used Bayesian inference on categorical tables. See Agresti and Hitchcock (2005) and Leonard (1977) for inference on second multinomial tables. The original data were gathered from mobile examination centers across the United States. NHANES III, which is an important program of the National Center for Health Statistics (NCHS), examines the state of health and nutritional in the United States. The program started in the 1960s, and has conducted surveys on various health- and nutrition-related topics. As a result, NHANES provides surveys based on large samples in the United States. However, the NCHS is also interested in estimates for smaller geographical areas and study domains. When the sample size of a subpopulation is small, we need to consider an alternative estimator based on a pooling strategy in order to analyze the data.

As a result, we focus on predicting the finite population proportion of each area. The finite population proportion is estimated by inputting the sample data into the model to predict the unobserved nonsample part of the finite population, then obtaining the weighted sum of the observed sample data and the predicted nonsample obtain to the sample proportion. First, we estimate the cell probability parameters

from the sample data. During this process, the observed count by category in NHANES III is employed as the sample part of the finite population. Second, these parameters are used to predict values for the nonsample part. Finally, we get the finite population proportions by combining the sample data and prediction value of nonsample part.

The remainder of the paper proceeds as follows. In Section 2, we introduce the hierarchical Bayesian pooling strategies used to analyze categorical data from small areas. In Section 3, we present and discuss the results of our data analysis of the BMD survey data set. Section 4 concludes the paper. Appendix A and B include the computation process for hierarchical Bayesian pooling model.

2. Hierarchical Bayesian models

2.1 Parametric models

For a hierarchical Bayesian baseline model, we consider an $I \times K$ contingency table, where n_{ik} indicates the k^{th} response in the i^{th} area, for $i = 1, \ldots, I$, $k = 1, \ldots, K$. Let π_{ik} denote the corresponding proportion for each cell. Then, we assume that

$$\mathbf{n}_i \,|\, \boldsymbol{\pi}_i \overset{\text{iid}}{\sim} \text{Multinominal}(n_{i.}, \boldsymbol{\pi}_i), \tag{2.1}$$

where $\mathbf{n}_i = (n_{i1}, \ldots, n_{iK})$ for $i = 1, \ldots, I$, is a vector of responses, $n_{i.} = \sum_{k=1}^{K} n_{ik}$ is the sum of the responses in area i, and $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iK})$, $0 \le \pi_{ik} \le 1$, $\sum_{k=1}^{K} \pi_{ik} = 1$, is the proportion vector for each area. The model does not allow any pooling and is denoted as a baseline to compare our models. The parameters π_i are independent and do not share a common effect. That is, the areas are unrelated.

There are five categories of parametric pooling models, classified according to the priors of the proportion vectors in a multinomial distribution. First, the four prior distributions for parametric Bayesian inferences are given as follows:

1) No pooling, $\boldsymbol{\pi}_i \overset{\text{iid}}{\sim} \text{Dirichlet}(\mathbf{1})$;
2) Complete pooling, $\boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1})$ with $\boldsymbol{\pi}_1 = \ldots = \boldsymbol{\pi}_I = \boldsymbol{\pi}$;
3) Adaptive pooling, $\boldsymbol{\pi}_i \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau)$;
4) Restricted pooling, $\boldsymbol{\pi}_i \overset{\text{ind}}{\sim} \phi\ \text{Dirichlet}(\boldsymbol{\mu}\tau) + (1 - \phi)\,\text{Dirichlet}(\mathbf{1})$;

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$, $0 \le \mu_k \le 1$, $\sum_{k=1}^{K} \mu_k = 1$ and $\tau > 0$ are the hyperparameters of the Dirichlet distribution. We further assume that $\pi(\boldsymbol{\mu}, \tau) = (K-1)!/(1+\tau)^2$, a shrinkage prior. We note that Yin and Nandram (2020a, b) place a Dirichlet process on the sampling process to accomodate gaps, outliers and ties in survey data, see also Nandram and Yin (2016a, b) for additional discussion of the Dirichlet process. The $\phi \sim \text{uniform}(1/2, 1)$, $\phi > 1/2$ means that more weight is attached to adaptive pooling.

Model 1 is a no-pooling model that estimates the parameter without any data sharing from other areas. Model 2, on the other hand, is a complete pooling model that estimates the parameter while treating the different areas as one. When conducting parameter estimations on a small area with a small number of data using Model 1, the estimation may face the small area problem, as the parameter is estimated by relying on insufficient data. Although the complete pooling model alleviates this issue, it faces problems

of its own: Overshrinkage and individual areas cannot be discerned. Hence, this paper introduces various pooling approaches to find a model that delivers better estimates. In Model 3, the adaptive pooling model introduced by Nandram, Zhou and Kim (2019), all areas share the same hyperparameters; hence, they share their area data information as well. This is an indirect complete pooling method that preserves some area variation but, in general, assumes all areas to have identical traits; see also Nandram (1998). This creates variations in estimating the hyperparameters. Thus, we propose the restricted pooling model, which alters Model 3 by removing data-sharing information between distinct areas. In this new model, distinct areas use their local data to estimate parameters, and areas with similar traits share information through pooling based on the same hyperparameter, thereby improving the estimation. This model, however, assigns the same hyperparameter to areas with similar traits, which may lead to the overshrinkage problem when smoothing in the category occurs. To mitigate this issue, we propose Model 5, the global-local pooling model, see Dunson (2009). The global-local pooling model pools information in data among areas with similar traits but also preserves each variation in the category through the local effect model, thereby reducing the smoothing in the categorical effect. Indeed, Model 5 is flexible and robust.

The fifth prior distribution used for parametric Bayesian inferences is called global–local pooling. In this case, we use different notation for the proportion vector of each area. Let $\mathbf{p}_i$, for $i = 1, \ldots, I$, denote the corresponding cell proportion vector in the i^{th} area. We assume that

$$\mathbf{n}_i \,|\, \mathbf{p}_i \overset{\text{ind}}{\sim} \text{Multinomial}(n_{i.}, \mathbf{p}_i),\ i = 1, \ldots, I, \tag{2.2}$$

where $\mathbf{p}_i = \left(e^{\theta+\eta_{i1}} \Big/ \left(1 + \sum_{k=1}^{K-1} e^{\theta+\eta_{ik}}\right), \ldots, e^{\theta+\eta_{i(K-1)}} \Big/ \left(1 + \sum_{k=1}^{K-1} e^{\theta+\eta_{ik}}\right), 1 \Big/ \left(1 + \sum_{k=1}^{K-1} e^{\theta+\eta_{ik}}\right)\right)'$. Here, $\mathbf{p}_i$ is composed of two components, namely, θ and η_{ik}, and θ reflects the basic probability that brought all the areas together. The global-local effect is reflected in the component $\boldsymbol{\eta}_i = \left(\eta_{i1}, \ldots, \eta_{i(K-1)}\right)'$. Specifically,

$$\boldsymbol{\eta}_i \overset{\text{iid}}{\sim} I_{(z_i=0)} \prod_{k=1}^{K-1} \text{N}(0, \sigma^2) + I_{(z_i=1)} \prod_{k=1}^{K-1} \text{N}(0, \sigma_k^2), \tag{2.3}$$

where $\text{N}(0, \sigma^2)$ is the normal distribution of the global parameter σ^2, $\text{N}(0, \sigma_k^2)$ is the normal distribution of the local parameters σ_k^2 for each category, where each area is denoted by a different index. Then, $z_i, i = 1, \ldots, I,$ follows a Bernoulli distribution with a hyperparameter ϕ, which adjusts the proportion between the global and local effects. Thus, if we need to focus on the global effect, the prior for ϕ is set using the uniform distribution on $(1/2, 1)$. Specifically, we assume that

$$z_i \,|\, \phi \overset{\text{iid}}{\sim} \text{Bernoulli}(\phi),\ i = 1, \ldots, I,$$

$$\phi \sim \text{Uniform}\left(\frac{1}{2}, 1\right),$$

$$\pi(\theta) = \frac{1}{\pi(1 + \theta^2)}, \text{ a Cauchy prior,}$$

$$\pi(\sigma^2, \sigma_1^2, \ldots, \sigma_{K-1}^2) = \frac{1}{(1 + \sigma^2)^2} \prod_{k=1}^{K-1} \frac{1}{(1 + \sigma_k^2)^2},$$

where $-\infty < \eta_{ik} < \infty$, $k = 1, \ldots, K-1$, $\sigma^2 > 0$, and $\sigma_k^2 > 0$, $k = 1, \ldots, K-1$. We have used the Cauchy prior for θ and shrinkage priors for variance components. The shrinkage priors are similar to the half Cauchy prior and they are mathematically more convenient when we make transformations to $(0, 1)$.

The models proposed in this paper are based on the adaptive pooling model (Nandram, Zhou and Kim, 2019), which applies the principle of assigning the same subscripts of parameter in prior distribution to similar experiments (Malec and Sedransk, 1992) to categorical data. In particular, the no pooling model and complete pooling model represent two extreme cases of adaptive pooling, with parameters $\boldsymbol{\mu}$ and τ. On the other hand, the restricted pooling model has a pooling principle such as adaptive pooling model, but the model also reflects uncertainty through the weighting parameter ϕ. However, the same parameter in the prior distribution used to pool such data frequently results in an overshrinkage problem. To compensate for this problem, we propose the global-local pooling model. The effects of the parameters are separated into global and local effects in this model. As a result, we propose two new models, restricted pooling model and global-local pooling model, and we compare these models with existing ones.

2.2 Nonparametric models

We consider the DP prior for $\boldsymbol{\pi}_i$ of (2.1) in Section 2.1. The prior structure is as follows:

$$\pi_i \,|\, G \overset{\text{iid}}{\sim} G$$
$$G \sim \text{DP}(\alpha, G_0),$$

where G_0 is the base distribution and the positive real number, α, is the concentration parameter in the DP prior. The model is specified by the structure of the base distribution. We note here that Yin and Nandram (2020a, b) used a DP on the sampling process to accommodate gaps, outliers and ties in survey data. First, we define the model using two prior distributions, as follows:

6) Nonparametric adaptive pooling $\quad G_0 \equiv \text{Dirichlet}(\boldsymbol{\mu}\tau)$;

7) Nonparametric restricted pooling $\quad G_0 \equiv \phi\ \text{Dirichlet}(\boldsymbol{\mu}\tau) + (1-\phi)\,\text{Dirichlet}\left(\mathbf{1}_{1\times K}\right)$.

We assume $\pi(\boldsymbol{\mu}, \tau) = (K-1)!/(1+\tau)^2$, $\pi(\alpha) \propto 1/(1+\alpha)^2$, and $\phi \sim \text{uniform}(1/2, 1)$. In Models 6 and 7, we use a stict-breaking process for the DP prior (Sethuraman, 1994).

The last model is a nonparametric version of (2.2) in Section 2.1, used for global-local pooling. Here, $\boldsymbol{\eta}_i$ are used to construct the nonparametric Bayesian setting, as follows:

$$\boldsymbol{\eta}_i \overset{\text{iid}}{\sim} I_{(z_i=0)}G_0 + I_{(z_i=1)}\prod_{k=1}^{K-1} G_k$$

$$G_0 \sim \text{DP}(\alpha_0, \text{MVN}(\mathbf{0}, \sigma^2\mathbf{I})),\ G_k \sim \text{DP}(\alpha_1, N(0, \sigma_k^2)),$$

where $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{i(K-1)})'$, $z_i \sim \text{Bernouilli}(\phi)$, $\pi(\theta) \propto 1$, $\phi \sim \text{uniform}(1/2, 1)$, $\pi(\sigma^2, \sigma_1^2, \ldots, \sigma_{K-1}^2) = 1\Big/\left\{(1+\sigma^2)^2 \prod_{k=1}^{K-1}(1+\sigma_k^2)^2\right\}$, for $i = 1, \ldots, I$. This is model (A.1), nonparametric global-local pooling.

The distribution of η_i involves a mixture of global and local pooling areas. While global pooling is conducted according to the same principle as the aforementioned nonparametric models, the Dirichlet

process prior, where the normal distribution is the base distribution for each cell, is independently defined, thereby alleviating the overshrinkage problem that could arise owing to global pooling. Here, z_i, which represents the weight of the model, follows the Bernouilli distribution with ϕ greater than $1/2$ as its parameter, and ϕ follows the uniform distribution. Hence, the local pooling area is weighted so that it is defined as the form that can better alleviate the degree of shrinkage. In addition, to ensure simplicity of the model, the heavy tail and noninformative characteristic of θ distribution follow the improper prior which is identical to, but simpler than, the previous parametric model and the posterior distribution is presented properly.

3. Data analysis

In this section, we present the empirical results from comparing the performance of the five parametric pooling models described in Section 2 and the three nonparametric pooling versions described in Section 3. We use BMD data for the period 1988 to 1994, taken from the NHANES III, which collected data from mobile examination centers across the United States.

Our analysis is conducted using contingency tables with a cell count for three categories of BMD in 31 counties in the U.S. Here, BMD is categorized into one of three levels. The normal category is defined as those with a BMD value less than one standard deviation (SD) below the non-Hispanic white (NHW) adult mean $(0.82\,\mathrm{mg/cm^2} < \text{BMD})$. The osteopenia category is defined as a BMD value between 1 and 2.5 SD below the young NHW adult mean $(0.64\,\mathrm{mg/cm^2} < \text{BMD} \le 0.82\,\mathrm{mg/cm^2})$. Then, the osteoporosis category corresponds to a BMD of more than 2.5 SD below the young NHW adult mean $(\text{BMD} \le 0.64\,\mathrm{mg/cm^2})$.

We predict the finite population proportion for the BMD distribution in each area using the Bayesian pooling model. The survey covers roughly 0.02% of the population, and prediction as needed for the remainder 99.98%, an enormous job. Table 3.1 shows the sample data, which have a cell count for each categorized level in each area. We estimate the finite population proportion by predicting the nonsample part of the finite population from a multinomial distribution with parameter $\boldsymbol{\pi}_i$, $i = 1, \ldots, I\,(= 31)$ at each MCMC iteration. Specifically, let N_{ik} for $k = 1, 2, 3$ be the total BMD level in area i, where the value is unknown. We have the value (n_{ik}) for the sample part of the finite population. Then, we compute the finite population proportion (P_{ik}) for $i = 1, \ldots, I\,(I = 31)$ as follows:

$$P_{ik} = \frac{1}{N_i}\{n_{ik} + (N_{ik} - n_{ik})\},\ k = 1, 2, 3, \tag{3.1}$$

where $N_i = \sum_{k=1}^{3} N_{ik}$, $N_{ik} - n_{ik}$ is the nonsample part for each BMD level $k\,(k = 1, 2, 3)$ in area i, taken from the multinomial distribution with parameter $\hat{\boldsymbol{\pi}}_i$, estimated using the MCMC in each model. Then, the posterior mean and standard deviation of P_{ik} are obtained using the estimated empirical distribution of P_{ik}.

Table 3.1
BMD categorical data for 31 counties from NHANES III

Areas	BMD		
	Normal	Osteopenia	Osteoporosis
1	33	24	9
2	46	39	5
3	40	25	8
4	48	25	6
5	40	15	10
6	74	30	12
7	47	19	7
8	38	15	6
9	49	16	11
10	99	40	14
11	39	18	2
12	63	27	4
13	48	18	5
14	42	16	4
15	40	15	4
16	110	44	7
17	37	14	3
18	55	18	5
19	47	12	6
20	296	95	17
21	59	18	4
22	78	21	7
23	196	55	15
24	149	44	9
25	69	19	5
26	49	10	6
27	73	19	3
28	76	14	3
29	77	13	4
30	96	13	6
31	88	12	4

We use 1,000 iterations to "burn in" the MCMC samples, and take every 10^{th} value to obtain the 1,000 iterations. In addition, we use autocorrelation plots of the model to adjust the number of repetitions and thinning intervals. For example, in a nonparametric model with a relatively large number of parameters, we take every 20^{th} estimated value from 1,001 to 20,000. We set the initial value of proportion $\boldsymbol{\pi}_i$ for $i = 1, \ldots, 31$ based on the column proportion of the sample values in each area.

The groups are categorized according to the quartile values of the first column proportion. The tuning parameter ξ_j, $j = 1, \ldots, J$, is initially set to $\xi = 0.5$, and then is revised based on the performance of each model.

In Table 3.2, we report the posterior means (PM) and posterior standard deviations (PSD) of the finite population proportions for the eight models and some areas. The cases of Model 1 and Model 2 are the most extreme pooling structures. The PM of Model 1 has the results 0.511, 0.496, 0.901, 0.826, and 0.820 for the corresponding areas 1, 2, 28, 29, and 31, respectively, in the normal BMD, implying that the areas' fluctuations are greater than those in Model 2's PM (0.652, 0.654, 0.714, 0.716, 0.719). For Model 3, the fluctuations (0.644, 0.612, 0.793, 0.816, 0.798) show a trend similar to that in Model 1's PM for each area but are smoother than those in Model 1. This could be interpreted as the indirect pooling effect through the

hyper-parameter rather than the direct pooling effect in Model 2. For the restricted pooling, in areas that show similar characteristics, the estimated values are calculated through indirect pooling and hyper-parameters as in Model 3, and the estimated values are smoother than those in Model 1 (PM of areas 28, 29, 31: 0.827, 0.779, 0.864, respectively). However, in areas where similar characteristics are not shown, the estimated values are close to those in Model 1 because the parameter is estimated by solely relying on the information in its associated areas (PM of areas 1, 2: 0.523, 0.510, respectively). Model 5, which is proposed to alleviate the overshrinkage problem that could arise owing to information pooling, shares information in nearby areas and alleviates the excessive shrinkage by reflecting the local effect of each cell, thereby rendering the estimated values that are between those in Model 1 and Model 2 (PM of Model 5: 0.581, 0.511, 0.711, 0.836, 0.808). It should be noted that the nonparametric Bayesian model assigns the areas with indexes as the same group according to the characteristics of information through hyper-parameter d_i, $i = 1, \ldots, I\,(= 31)$, and the same group shares the parameter directly. For BMD data, the number of the group ranges from one to three, showing the highest frequency, and the PM is estimated as in Model 2. The characteristics of the estimated values of PM for each model are shown identically in osteopenia BMD and osteoporosis BMD as well.

The posterior means, standard deviations (SD) and posterior coefficients of variation (CV) of the finite population proportions for the eight models can also be seen in Figure 3.1-3.3.

In Figure 3.1, the variation of PM for eight models is the largest in the normal BMD, which takes up the largest proportion. Especially, we can see that PMs of the nonparametric model are similar to that of the complete pooling model. During data analysis, we found one to three group index. Through this, we were able to discover that the BMD distribution is quite similar across the areas in NHANES III. Hence, all areas share their information with others to estimate the same hyperparameter. Thus, the nonparametric and complete pooling models give similar estimates. Therefore, pooling can solve problems associated with small area estimates when areas share similar traits. Furthermore, Figure 3.2 shows that the performance of the nonparametric models are good through the fact that the SD of the nonparametric models with many parameters are similar or smaller than that of the parametric models.

In addition, we can see the CV of the models by BMD status in Figure 3.3. In the case of CV, osteoporosis BMD shows the greatest difference between models. Also it can be seen that the CV of the nonparametric versions is relatively low compared to the parametric version, which is not different for each BMD status. Furthermore, it is very meaningful that the nonparametric version had a smaller CV than the models of the parametric version, even though it had infinite parameter space.

To estimate the parameters, we use a Gibbs sampler. Whereas the parameters with restricted parameter spaces are sampled using the grid method, the other parameters are sampled using the Metropolis-Hastings algorithm. We tune to get acceptance rate 30-70%. In the actual analysis, the acceptance rate of the algorithm is 34-49%. We compare the two measures in terms of the performance of each model. First, we calculate the deviance information criterion (DIC), a typical Bayesian model choice criterion, to compare the hierarchical Bayesian models. The DIC was proposed by Spiegelhalter et al. (2002), where a lower DIC value indicates better performance. Second, we evaluate the performance of the eight models by

calculating the logarithmic conditional predictive ordinate (LCPO), which is a comparison method that uses cross validation. The average LCPO, proposed by Gneiting and Raftery (2007), is calculated as follows:

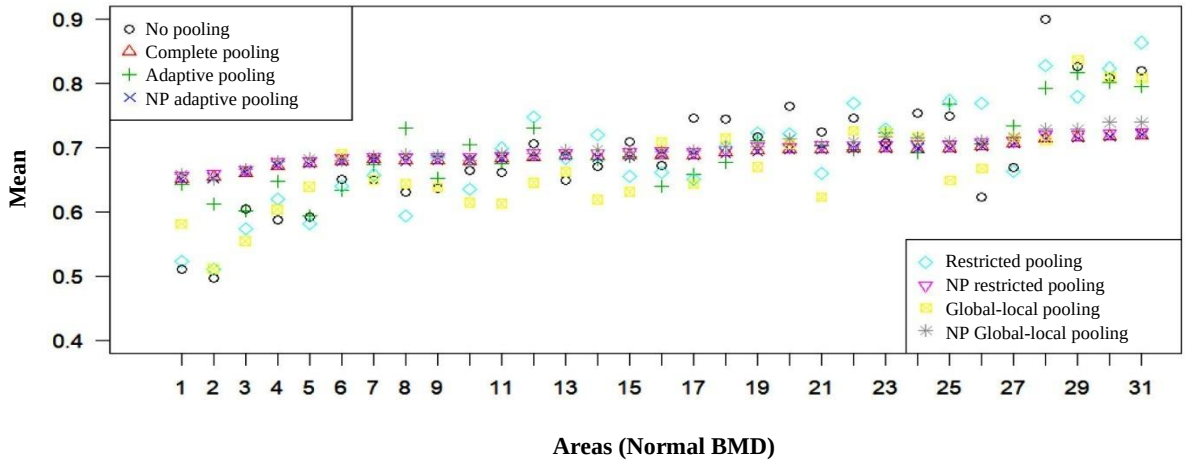
$$\overline{\text{LCPO}} = -\frac{1}{I}\sum_{i=1}^{I}\log\left(\hat{\text{CPO}}_i\right), \tag{3.2}$$

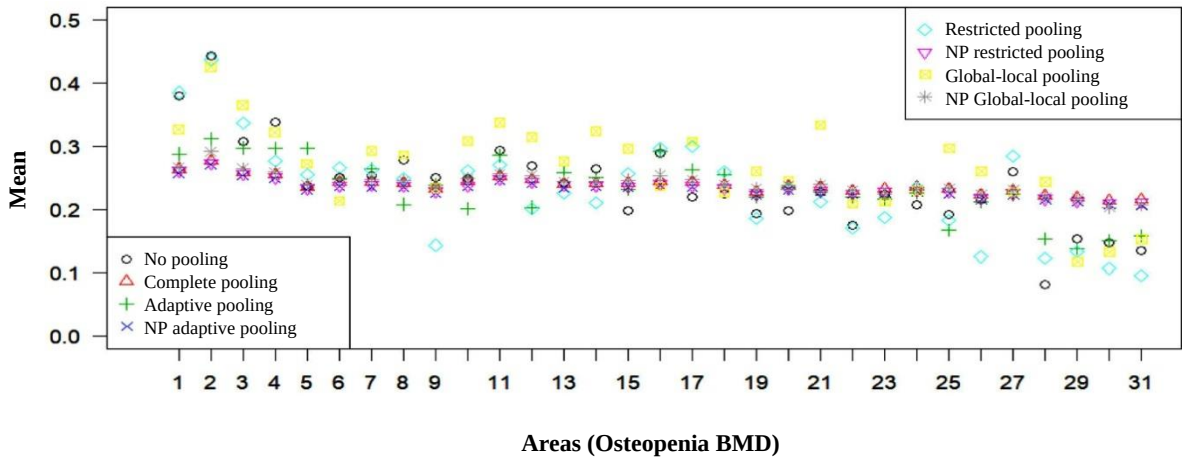
where $\hat{\text{CPO}}_i = \sum_{h=1}^{H} w_h P(Y = y|\boldsymbol{\Omega}^{(h)})$, $w_h = \sum_{h=1}^{H} f(Y = y|\boldsymbol{\Omega}^{(h)}) / f(Y = y|\boldsymbol{\Omega}^{(h)})$, for $i = 1, \ldots, I$, and $P(Y = y|\boldsymbol{\Omega}^{(h)})$ is the likelihood of a single observation of a given parameter $\boldsymbol{\Omega}^{(h)}$, and $h = 1, \ldots, H$ denotes the iterations from the MCMC result under the hierarchical Bayesian pooling model.

Table 3.2
Posterior summaries for finite population proportions of BMD data under the eight models by areas

Areas	Model	Normal BMD			Osteopenia BMD			Osteoporosis BMD		
		PM	PSD	95% CI	PM	PSD	95% CI	PM	PSD	95% CI
1	1	0.511	0.024	(0.467, 0.561)	0.380	0.024	(0.333, 0.427)	0.109	0.015	(0.082, 0.139)
	2	0.652	0.023	(0.603, 0.692)	0.264	0.021	(0.227, 0.306)	0.084	0.013	(0.061, 0.109)
	3	0.644	0.024	(0.594, 0.691)	0.288	0.023	(0.242, 0.338)	0.068	0.011	(0.048, 0.091)
	4	0.523	0.024	(0.476, 0.570)	0.385	0.024	(0.336, 0.432)	0.092	0.013	(0.070, 0.118)
	5	0.581	0.022	(0.536, 0.627)	0.327	0.022	(0.285, 0.370)	0.092	0.013	(0.067, 0.118)
	6	0.651	0.022	(0.606, 0.696)	0.258	0.020	(0.218, 0.297)	0.090	0.014	(0.067, 0.117)
	7	0.656	0.023	(0.612, 0.700)	0.260	0.021	(0.218, 0.300)	0.084	0.013	(0.061, 0.109)
	8	0.658	0.022	(0.615, 0.703)	0.267	0.021	(0.230, 0.311)	0.075	0.012	(0.052, 0.100)
2	1	0.496	0.021	(0.458, 0.534)	0.442	0.021	(0.402, 0.484)	0.061	0.010	(0.042, 0.080)
	2	0.654	0.021	(0.617, 0.693)	0.277	0.019	(0.242, 0.316)	0.068	0.010	(0.047, 0.087)
	3	0.612	0.020	(0.570, 0.652)	0.313	0.019	(0.273, 0.353)	0.075	0.012	(0.053, 0.098)
	4	0.510	0.022	(0.470, 0.551)	0.438	0.022	(0.396, 0.478)	0.052	0.009	(0.036, 0.071)
	5	0.511	0.023	(0.466, 0.554)	0.425	0.023	(0.381, 0.472)	0.064	0.011	(0.045, 0.085)
	6	0.653	0.020	(0.618, 0.693)	0.271	0.017	(0.237, 0.304)	0.075	0.012	(0.053, 0.096)
	7	0.659	0.020	(0.618, 0.702)	0.273	0.018	(0.239, 0.311)	0.068	0.011	(0.047, 0.089)
	8	0.651	0.021	(0.609, 0.689)	0.292	0.019	(0.254, 0.329)	0.058	0.011	(0.040, 0.079)
28	1	0.901	0.011	(0.881, 0.920)	0.081	0.010	(0.062, 0.099)	0.018	0.005	(0.010, 0.028)
	2	0.714	0.020	(0.677, 0.751)	0.223	0.017	(0.189, 0.256)	0.064	0.011	(0.043, 0.088)
	3	0.793	0.017	(0.757, 0.826)	0.154	0.015	(0.126, 0.184)	0.053	0.010	(0.034, 0.073)
	4	0.827	0.016	(0.796, 0.856)	0.123	0.013	(0.099, 0.148)	0.050	0.010	(0.032, 0.069)
	5	0.711	0.021	(0.672, 0.751)	0.244	0.020	(0.209, 0.283)	0.046	0.010	(0.028, 0.064)
	6	0.715	0.018	(0.680, 0.748)	0.216	0.016	(0.184, 0.249)	0.070	0.011	(0.049, 0.092)
	7	0.721	0.019	(0.683, 0.757)	0.217	0.018	(0.185, 0.251)	0.062	0.011	(0.043, 0.084)
	8	0.729	0.022	(0.686, 0.770)	0.220	0.020	(0.181, 0.261)	0.052	0.011	(0.034, 0.073)
29	1	0.826	0.015	(0.796, 0.855)	0.153	0.015	(0.124, 0.184)	0.020	0.005	(0.011, 0.030)
	2	0.716	0.019	(0.679, 0.755)	0.219	0.017	(0.187, 0.253)	0.065	0.010	(0.047, 0.085)
	3	0.816	0.016	(0.785, 0.847)	0.139	0.014	(0.113, 0.168)	0.045	0.008	(0.030, 0.060)
	4	0.779	0.017	(0.746, 0.815)	0.134	0.014	(0.105, 0.160)	0.087	0.012	(0.065, 0.111)
	5	0.836	0.016	(0.807, 0.866)	0.117	0.013	(0.095, 0.145)	0.047	0.009	(0.031, 0.067)
	6	0.715	0.020	(0.677, 0.755)	0.213	0.018	(0.179, 0.248)	0.072	0.012	(0.051, 0.096)
	7	0.721	0.019	(0.685, 0.756)	0.214	0.017	(0.183, 0.251)	0.066	0.011	(0.045, 0.085)
	8	0.729	0.021	(0.690, 0.768)	0.217	0.020	(0.179, 0.257)	0.053	0.011	(0.034, 0.075)
31	1	0.820	0.015	(0.792, 0.849)	0.136	0.014	(0.110, 0.161)	0.044	0.008	(0.030, 0.061)
	2	0.719	0.019	(0.680, 0.758)	0.216	0.017	(0.181, 0.253)	0.065	0.010	(0.046, 0.087)
	3	0.796	0.016	(0.765, 0.827)	0.159	0.015	(0.133, 0.188)	0.046	0.008	(0.031, 0.063)
	4	0.864	0.013	(0.838, 0.889)	0.095	0.011	(0.075, 0.119)	0.041	0.008	(0.027, 0.056)
	5	0.808	0.018	(0.773, 0.841)	0.153	0.017	(0.122, 0.188)	0.038	0.008	(0.024, 0.054)
	6	0.721	0.018	(0.685, 0.756)	0.207	0.017	(0.177, 0.239)	0.072	0.011	(0.052, 0.092)
	7	0.724	0.018	(0.688, 0.758)	0.211	0.017	(0.180, 0.243)	0.065	0.010	(0.045, 0.085)
	8	0.739	0.020	(0.698, 0.774)	0.208	0.019	(0.171, 0.247)	0.053	0.011	(0.033, 0.073)

Figure 3.1 The posterior means plot of the finite population proportion.





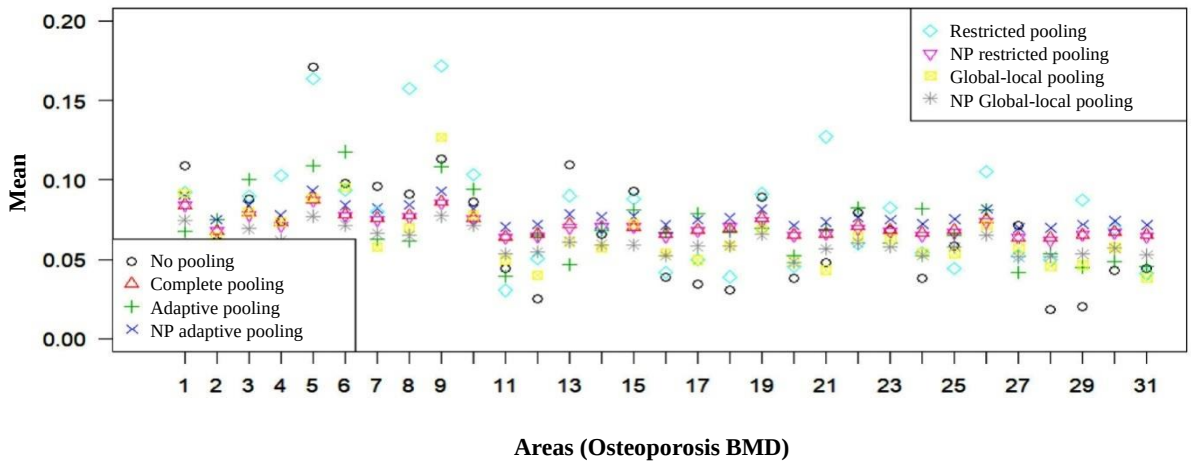
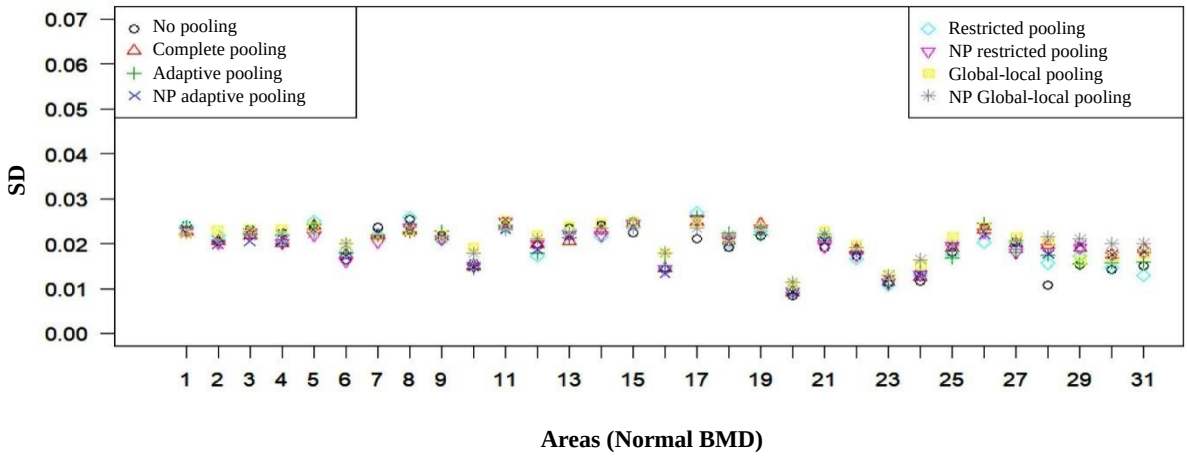
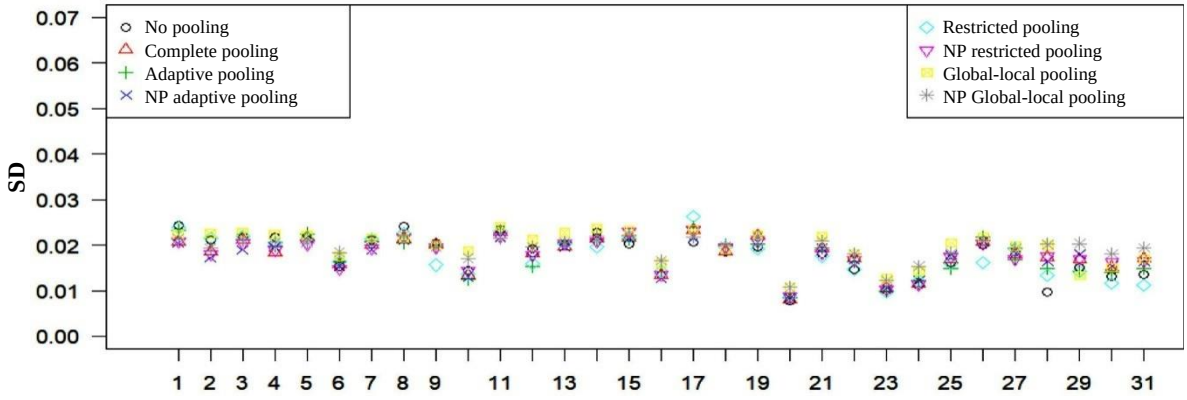


Figure 3.2 The posterior standard deviations plot of the finite population proportion.





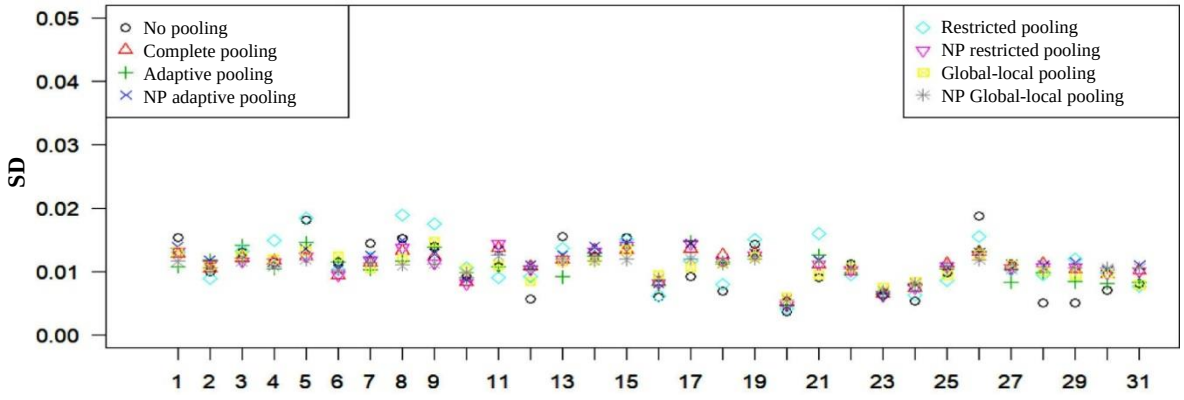
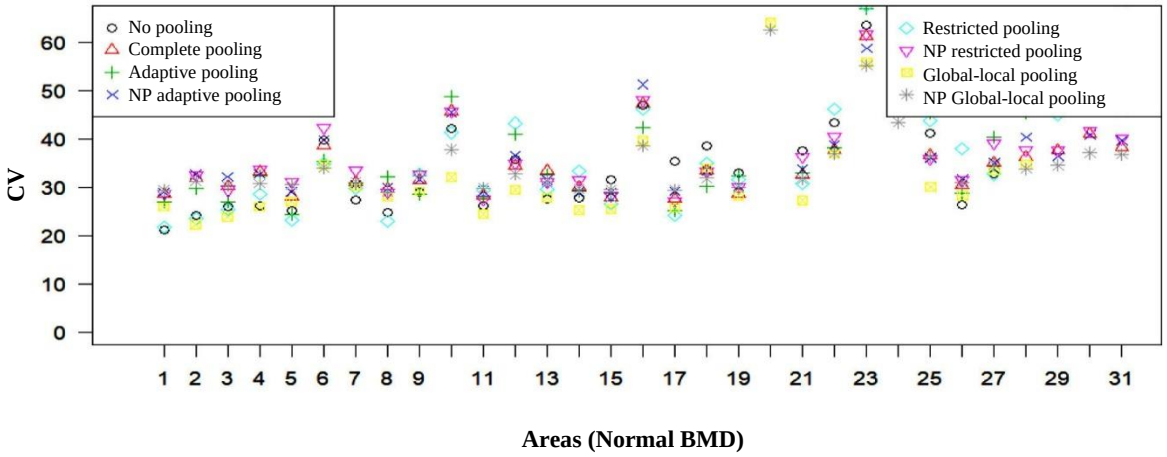
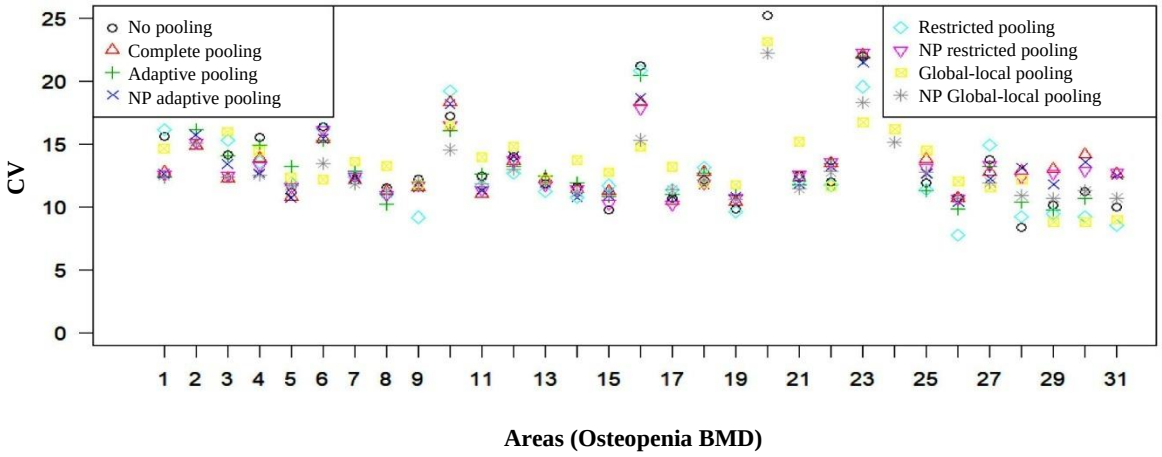


Figure 3.3 The coefficients of variation plot of the finite population proportion.





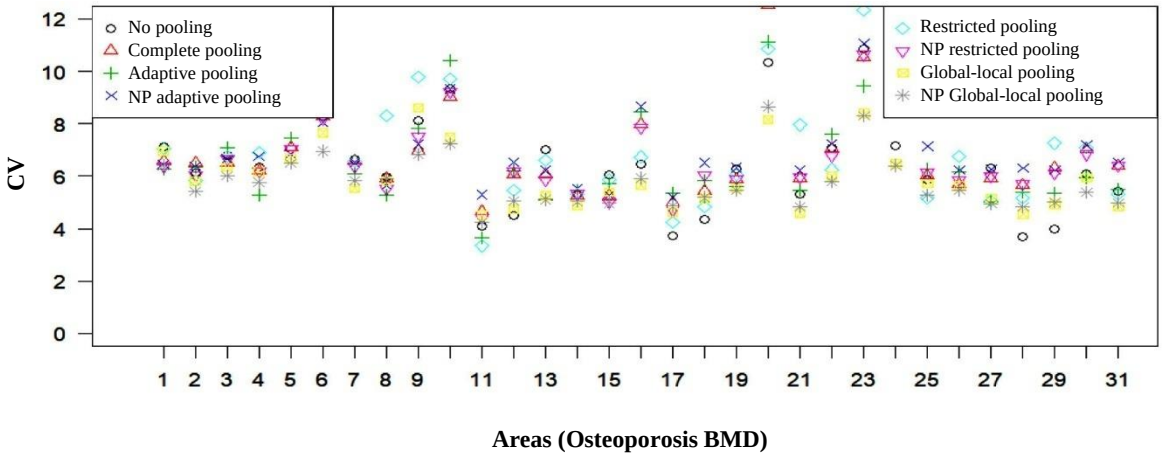


Table 3.3 shows the results of the two measures for each Bayesian pooling model. The model is considered to perform better as its estimated measures are smaller. The LCPO and DIC in parametric models are compared using the global-local pooling model, and LCPO and DIC have 12.707 and 4,984.56, respectively, implying the best performance. The restricted pooling model in Model 4 has an LCPO value of 12.886, showing the second-best performance after the global-local pooling model, but the DIC in the adaptive pooling model has a value of 4,998.90, which is lower than that in the restricted pooling model.

Table 3.3
Comparisons of $\overline{\text{LCPO}}$ and DIC (95% CI) under Models 1-8

Model		$\overline{\text{LCPO}}$	DIC
Parametric models	No pooling	13.167	4,999.19
	Complete pooling	13.151	5,011.45
	Adaptive pooling	13.411	4,998.90
	Restricted pooling	12.886	5,000.21
	Global-local pooling	12.707	4,984.56
Nonparametric models	Adaptive pooling	13.105	5,001.17
	Restricted pooling	12.837	4,983.88
	Global-local pooling	12.694	4,768.47

Another point to keep in mind is that the nonparametric version using the same pooling method shows similar values to those of the parametric method. In particular, although the nonparametric global-local pooling model has the greatest number of parameters to be estimated, its LCPO and DIC have values of 12.694 and 4,768.47, respectively, thereby indicating that it has the best performance among all the models. Additionally, in the restricted pooling model, the nonparametric model's LCPO and DIC have values of 12.837 and 4,983.88, respectively, showing better performance than the parametric model. These results are identical in the LCPO scale of the adaptive pooling (i.e., base) model (LCPO (parametric vs nonparametric) = (13.411 vs 13.105)). This means the performance of the nonparametric version is very good for our data, even though the parameter space has infinite dimensions.

Table 3.4 illustrates the calculated statistical values to estimate the shrinkage of the model. To estimate shrinkage, we calculated the average and standard deviation of the absolute difference from the no shrinkage model for each BMD category. Let PM_{ik}, $i = 1, \ldots, I$, $k = 1, 2, 3$ denote the posterior mean of the finite population proportion for each cell in area i. The average (ASE) and standard deviation (SDSE) of the shrinkage estimator are

$$\text{ASE}_k = \frac{1}{I}\sum_{i=1}^{I}\frac{|\text{PM}_{ik} - \text{PM}_{0ik}|}{\text{PM}_{0ik}},\ I = 31, \tag{3.3}$$

$$\text{SDSE}_k = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}\left(\frac{|\text{PM}_{ik} - \text{PM}_{0ik}|}{\text{PM}_{0ik}} - \text{ASE}_k\right)^2}, \tag{3.4}$$

where PM_{0ik} is the posterior mean for the finite population proportion of Model 1, a no poooling model.

Based on this calculation, we show that the ASE and SDSE in the normal cell are the smallest in the global-local pooling model. Drawing from the data analysis, with the BMD data applied in this study, the number of groups of slice sampling is 1 to 3, and we can confirm that the data characteristics are identical for most regions. In the global-local model, however, it can be shown that the problem of overshrinkage induced by pooling is solved by looking at the smallest shrinkage degree. In addition, osteopenia cells and ostoporosis cells could confirm the tendency of low shrinkage relative to other models, and the SDSE in the global-local pooling model can be shown to be small. Meanwhile, in the case of nonparametric models, the group index had the highest number of 1 because of slice sampling; therefore, it could be suspected that data dependence was excessive in the no pooling model.

Also, Geweke's test, autocorrelation plot and effective sample size (ESS) were applied for the diagnosis of the model, and they showed strongly mixing chains.

Table 3.4
A comparision of shrinkage in the eight models, see equations (3.3) and (3.4)

Model	Normal		Osteopenia		Osteoporosis	
	Mean	Std	Mean	Std	Mean	Std
No pooling(no shrinkage model)						
Complete pooling	0.087	0.072	0.217	0.320	0.522	0.622
Adaptive pooling	0.066	0.064	0.147	0.166	0.474	0.501
Restricted pooling	0.088	0.071	0.210	0.302	0.587	0.728
Global-local pooling	0.054	0.043	0.157	0.137	0.456	0.682
NP Adaptive pooling	0.088	0.074	0.209	0.306	0.524	0.609
NP Restricted pooling	0.065	0.049	0.249	0.354	0.335	0.362
NP Global-local pooling	0.085	0.072	0.206	0.311	0.427	0.430

NP: Nonparametric

4. Conclusion

In this study, we construct hierarchical parametric Bayesian pooling models and their nonparametric versions using the Ferguson (1973) Dirichlet process prior to pool the data. The pooling methodologies developed here are useful for analyzing survey data. We used the grid method to draw the parameters with nonstandard posterior densities and support that lies in a finite interval. However, we used the Metropolis-Hastings algorithm to draw the parameters with support in an infinite interval.

The Dirichlet process is assumed for the parameter of interest $\boldsymbol{\pi}_i$, for $i = 1, \ldots, I$, in our models. We apply the slice sampling algorithm for the specification of Dirichlet process prior, which is an extension of the widely used stick-breaking prior proposed by Ishwaran and James (2001). Five parametric models are modeled in a finite-dimensional parameter space, and three nonparametric versions have an infinite-dimensional parameter space. The eight hierarchical Bayesian models are also distinguished according to the type of effects in the model parameters. For the basic model (2.1), we can construct more effective and

efficient models that allow for a borrowing effect from neighboring areas in small-area estimations. However, exchangeable priors in a hierarchical Bayesian model may cause an overshrinkage problem. To compensate for this problem, the effect of a parameter is divided into two elements, as shown in the basic model (2.2), called the hierarchical Bayesian global-local pooling model. The model allows for grouping of similar experiments (area) and the borrowing of information in each area.

To compare the eight models using real data, we use the BMD data provided by the NHANES III. BMD is statistically correlated with the probability of fractures, which are an important public health problem, especially in elderly women. Therefore, BMD is an important indicator in diagnoses of osteoporosis, where patients might benefit from early management to improve their bone strength. For each sample, we assign an indicator based on three categories (normal, osteopenia, osteoporosis) before analyzing the data. The resulting hierarchical models with a pooling prior for BMD data outperformed the other models. To compare the models' performances, we calculated the DIC and the LCPO. Here, we found the best performance in the global–local pooling model. Although the nonparametric versions of the models have an infinite-dimensional parameter space, they showed similar values for the two comparison measures to those of the parametric pooling model with a finite-dimensional parameter space. Therefore, we should be careful in interpreting the results.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2016R1D1A1B03932261). This research was also supported by a grant from the Simons Foundation (#353953, Balgobin Nandram).

Appendix

A. Computations for parameteric models

Let $\mathbf{n} = (\mathbf{n}_1, \ldots, \mathbf{n}_I)'$ be the response matrix and $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_I)'$ be the proportion parameter matrix for (2.1). In an adaptive pooling model, let $\boldsymbol{\Omega}_i = (\boldsymbol{\pi}_i, \boldsymbol{\mu}, \tau)$, $i = 1, \ldots, I$. Here, no pooling and complete pooling are special cases of adaptive pooling, with parameters $\boldsymbol{\mu}$ and τ. The full conditional posterior density of the parameters for the given data is obtained in the usual way by combining the likelihood and the priors, as follows:

$$\begin{aligned}\pi(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_I \,|\, \mathbf{n}) &\propto \left\{\prod_{i=1}^{I} f(\mathbf{n} \,|\, \boldsymbol{\pi}_i)\,\pi(\boldsymbol{\pi}_i)\right\}\pi(\boldsymbol{\mu}, \tau) \\ &\propto \prod_{i=1}^{I}\left\{\frac{1}{D(\boldsymbol{\mu}\tau)}\prod_{k=1}^{K}\pi_{ik}^{n_{ik}+\mu_k\tau-1}\right\}\frac{(K-1)!}{(\tau+1)^2},\end{aligned}$$

where $D(\boldsymbol{\mu}\tau) = \prod_{k=1}^{K}\Gamma(\mu_k\tau)\Big/\Gamma\left(\sum_{k=1}^{K}\mu_k\tau\right)$.

To run a Gibbs sampler, we draw values as follows:

(a) Full conditional for $\boldsymbol{\pi}_i$, $i = 1, \ldots, I$: Draw $\boldsymbol{\pi}_i \,|\, \mathbf{n}_i, \boldsymbol{\mu}, \tau \sim \text{Dirichlet}((\mathbf{n}_i + \boldsymbol{\mu})\tau)$.

(b) Full conditional for $\boldsymbol{\mu}$: Draw

$$\pi(\boldsymbol{\mu} \mid \mathbf{n}, \boldsymbol{\pi}, \tau) \propto \prod_{i=1}^{I} \left\{ \frac{\Gamma\left(\sum_{k=1}^{K} \mu_k \tau\right)}{\prod_{k=1}^{K} \Gamma(\mu_k \tau)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1} \right\}. \tag{A.1}$$

Let $\boldsymbol{\mu}_{(k)}$, $k = 1, \ldots, K$ denote the vector of parameters other than the k^{th} component μ_k. Then, we obtain the conditional posterior density of μ_k, given $\boldsymbol{\mu}_{(k)}$, in each stage. Here, we need to estimate the $K - 1$ components of parameter $\boldsymbol{\mu}$ sequentially. Then, we can calculate the K^{th} component value of $\boldsymbol{\mu}$ using $\mu_K = 1 - \sum_{k=1}^{K-1} \mu_k$. Using the conditional posterior density (3.1), we can draw μ_k, $k = 1, \ldots, K - 1$ using the grid method, with support $1 - \sum_{k'=1,\, k' \neq k}^{K-1} \mu_{k'}$.

(c) Full conditional for τ: Draw

$$\pi(\tau \mid \mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\mu}) \propto \prod_{i=1}^{I} \left\{ \frac{\Gamma\left(\sum_{k=1}^{K} \mu_k \tau\right)}{\prod_{k=1}^{K} \Gamma(\mu_k \tau)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1} \right\} \frac{1}{(\tau + 1)^2}.$$

We can use the grid method for τ in this case as well. Because the grid method can be used for closed support, we transform τ to $\rho = 1/(1 + \tau)$, $0 < \rho < 1$. The absolute Jacobian is $1/\rho^2$. Then, the conditional posterior density of ρ can be expressed as follows:

$$\pi(\rho \mid \mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\mu}) \propto \prod_{i=1}^{I} \left\{ \frac{\Gamma\left(\sum_{k=1}^{K} \mu_k \frac{1-\rho}{\rho}\right)}{\prod_{k=1}^{K} \Gamma\left(\mu_k \frac{1-\rho}{\rho}\right)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \frac{1-\rho}{\rho} - 1} \right\}.$$

The full conditional posterior density in the case of restricted pooling is obtained from the likelihood and the priors, constructed from $\pi(\boldsymbol{\mu}, \tau)$, and from an additional prior for ϕ, $i = 1, \ldots, I$:

$$\pi(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_I \mid \mathbf{n}) = \left\{ \prod_{i=1}^{I} f(\mathbf{n} \mid \boldsymbol{\pi}_i)\, \pi(\boldsymbol{\pi}_i) \right\} \pi(\boldsymbol{\mu}, \tau)\, \pi(\phi),$$

where $\boldsymbol{\Omega}_i = (\boldsymbol{\pi}_i, \boldsymbol{\mu}, \tau, \phi)$, $i = 1, \ldots, I$.

For the Gibbs sampler, we consider the latent variables z_i, $i = 1, \ldots, I$ from a Bernoulli distribution with parameter ϕ. Here, the joint posterior density is given as follows:

$$\begin{aligned} \pi(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_I \mid \mathbf{n}) &\propto \left\{ \prod_{i=1}^{I} f(\mathbf{n} \mid \boldsymbol{\pi}_i)\, \pi(\boldsymbol{\pi}_i \mid z_i)\, \pi(z_i) \right\} \pi(\boldsymbol{\mu}, \tau)\, \pi(\phi) \\ &\propto \left\{ \prod_{i=1}^{I} \left[\phi \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau + n_{ik} - 1} \right]^{z_i} \left[(1 - \phi) \prod_{k=1}^{K} \pi_{ik}^{n_{ik}} \right]^{1 - z_i} \right\} \\ &\quad \times \frac{(K-1)!}{(\tau + 1)^2} I_{\left(\phi \in \left(\frac{1}{2}, 1\right)\right)}. \end{aligned}$$

Then, our Gibbs sampler is described as follows:

(a) Full conditional for ϕ: Draw

$$\pi(\phi \mid \mathbf{n}, \boldsymbol{\mu}, \tau, \mathbf{z}) \propto \phi^{\sum_{i=1}^{I} z_i} (1 - \phi)^{I - \sum_{i=1}^{I} z_i}, \frac{1}{2} \leq \phi \leq 1.$$

In other words, given $\boldsymbol{\mu}$, τ, $\mathbf{z}$, and the data, ϕ follows a truncated beta distribution with parameters $\sum_{i=1}^{I} z_i$ and $I - \sum_{i=1}^{I} z_i$, and the lower bound of ϕ is $1/2$.

(b) Full conditional for z_i, $i = 1, \ldots, I$: Draw $z_i \,|\, \mathbf{n}, \boldsymbol{\pi}_i, \boldsymbol{\mu}, \tau \sim \text{Bernoulli}(p_i)$, where

$$p_i = \left(\phi \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1} \Big/ D(\boldsymbol{\mu}\tau)\right) \Big/ \left(\phi \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1} \Big/ D(\boldsymbol{\mu}\tau) + (1-\phi)\, I_{\left(\sum_{k=1}^{K} \pi_{ik} = 1\right)} \Big/ K\right).$$

(c) Full conditional for $\boldsymbol{\pi}_i$, $i = 1, \ldots, I$: Draw

$$\pi(\boldsymbol{\pi}_i \,|\, \mathbf{n}, z_i, \boldsymbol{\mu}, \tau) \propto \left[\prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau + n_{ik} - 1}\right]^{z_i} \left[\prod_{k=1}^{K} \pi_{ik}^{1 + n_{ik} - 1}\right]^{1 - z_i}.$$

In the case of $z_i = 1$, we can generate the value of $\boldsymbol{\pi}_i$ from a Dirichlet distribution with parameters $\boldsymbol{\mu}$ and τ. In other cases, we can interpret that as the uncertainty of the modeling. That is, for $\boldsymbol{\pi}_i$ given $z_1 = 0$, $\mathbf{n}$ draws its value from a uniform Dirichlet distribution with a $K \times 1$ parameter vector, where each component has the value one.

(d) Full conditional for $\boldsymbol{\mu}$: Draw

$$\pi(\boldsymbol{\mu} \,|\, \mathbf{n}, \boldsymbol{\pi}, \tau) \propto \prod_{z_i = 1} \left\{\frac{\Gamma\left(\sum_{k=1}^{K} \mu_k \tau\right)}{\prod_{k=1}^{K} \Gamma(\mu_k \tau)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1}\right\}.$$

(e) Full conditional for τ: Draw

$$\pi(\tau \,|\, \mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\mu}) \propto \prod_{z_i = 1} \left\{\frac{\Gamma\left(\sum_{k=1}^{K} \mu_k \tau\right)}{\prod_{k=1}^{K} \Gamma(\mu_k \tau)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1}\right\} \frac{1}{(\tau + 1)^2}.$$

Of course, the generating process for parameters $\boldsymbol{\mu}$ and τ is similar to that of the parameter in adaptive pooling. In addition, the data used for $\boldsymbol{\mu}$ and τ are $z_i = 1$, $i = 1, \ldots, I$.

Otherwise, the parameter vector corresponding to area i in global-local pooling consists of θ, $\boldsymbol{\eta}$, $\mathbf{z}$, ϕ, σ^2, σ_1^2, …, and σ_{K-1}^2. Then, the full conditional posterior density for the given data in the model is as follows:

$$\begin{aligned}
\pi(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_I \,|\, \mathbf{n}) &\propto \left\{\prod_{i=1}^{I} f(\mathbf{n}_i \,|\, \boldsymbol{\eta}_i, \theta)\right\} \left\{\prod_{i=1}^{I} \pi(\boldsymbol{\eta}_i \,|\, z_i, \sigma^2, \sigma_1^2, \ldots, \sigma_{K-1}^2)\right\} \pi(\sigma^2) \\
&\quad \times \pi(\sigma_1^2, \ldots, \sigma_{K-1}^2)\, \pi(\theta)\, \pi(\mathbf{z} \,|\, \phi)\, \pi(\phi) \\
&= \left\{\prod_{i=1}^{I} \frac{n_i!}{\prod_{k=1}^{K} n_{ik}!} \left\{\prod_{k=1}^{K-1} \left(\frac{e^{\theta + \eta_{ik}}}{1 + \sum_{l=1}^{K-1} e^{\theta + \eta_{il}}}\right)^{n_{ik}}\right\} \left(\frac{1}{1 + \sum_{l=1}^{K-1} e^{\theta + \eta_{ik}}}\right)^{n_{iK}}\right\} \\
&\quad \times \left\{\prod_{i=1}^{I} \left[\prod_{k=1}^{K-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \eta_{ik}^2\right)\right]^{z_i} \left[\prod_{k=1}^{K-1} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{1}{2\sigma_k^2} \eta_{ik}^2\right)\right]^{1 - z_i}\right\} \\
&\quad \times \frac{1}{(1 + \sigma^2)^2} \left\{\prod_{k=1}^{K-1} \frac{1}{(1 + \sigma_k^2)^2}\right\} \frac{1}{\pi(1 + \theta^2)} \left\{\prod_{i=1}^{I} \phi^{z_i} (1 - \phi)^{1 - z_i}\right\} I_{(1/2 \le \phi \le 1)},
\end{aligned}$$

where $\boldsymbol{\Omega}_i = (\theta, \boldsymbol{\eta}_i, z_i, \phi, \sigma^2, \sigma_1^2, \ldots, \sigma_{K-1}^2)$, $i = 1, \ldots, I$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_I)'$ and $\mathbf{z} = (z_1, \ldots, z_I)'$.

Then, the Gibbs sampler is as follows:

(a) Full conditional for ϕ: Draw $\phi \,|\, \text{others} \sim \text{truncated Beta}\left(\sum_{i=1}^{I} z_i,\ I - \sum_{i=1}^{I} z_i,\ \frac{1}{2},\ 1\right)$.

(b) Full conditional for z_i, $i = 1, \ldots, I$: Draw $z_i \,|\, \text{others} \sim \text{Bernoulli}(p_i)$, with $p_i = \phi \prod_{k=1}^{K-1} \left(\exp(-\eta_{ik}^2 / 2\sigma^2)/\sqrt{2\pi\sigma^2}\right) \Big/ \left\{\phi \prod_{k=1}^{K-1} \exp(-\eta_{ik}^2/2\sigma^2)\Big/\sqrt{2\pi\sigma^2} + (1-\phi) \prod_{k=1}^{K-1} \left(\exp(-\eta_{ik}^2/2\sigma_k^2)/\sqrt{2\pi\sigma_k^2}\right)\right\}$.

(c) Full conditional for θ: Draw

$$\pi(\theta \,|\, \text{others}) \ \propto\ \left[\prod_{i=1}^{I}\prod_{k=1}^{K-1}\left(\frac{e^{\theta+\eta_{ik}}}{1+\sum_{l=1}^{K-1} e^{\theta+\eta_{il}}}\right)^{n_{ik}}\left(\frac{1}{1+\sum_{l=1}^{K-1} e^{\theta+\eta_{il}}}\right)^{n_{iK}}\right]\frac{1}{1+\theta^2}.$$

(d) Full conditional for η_{ik}, $i = 1, \ldots, I$, $k = 1, \ldots, (K-1)$: Draw

$$\pi(\eta_{ik} \,|\, \text{others}) \ \propto\ I_{(z_i=1)}\left(\frac{1}{1+\sum_{l=1}^{K-1} e^{\theta+\eta_{il}}}\right)^{\sum_{k=1}^{K} n_{ik}} \exp\left\{\sum_{k=1}^{K-1}\eta_{ik}n_{ik} - \frac{1}{2}\sum_{k=1}^{K-1}\frac{1}{\sigma^2}\eta_{ik}^2\right\}$$
$$+\ I_{(z_i=0)}\left(\frac{1}{1+\sum_{l=1}^{K-1} e^{\theta+\eta_{il}}}\right)^{\sum_{k=1}^{K} n_{ik}} \exp\left\{\sum_{k=1}^{K-1}\eta_{ik}n_{ik} - \frac{1}{2}\sum_{k=1}^{K-1}\frac{1}{\sigma_k^2}\eta_{ik}^2\right\}.$$

(e) Full conditional for σ^2: Draw

$$\pi(\sigma^2 \,|\, \text{others}) \ \propto\ \frac{1}{\sigma^{(K-1)\sum_{i=1}^{I} I(z_i=1)}\left(1+\sigma^2\right)^2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{z_i=1}\sum_{k=1}^{K-1}\eta_{ik}^2\right\}.$$

(f) Full conditional for σ_k^2 $k = 1, \ldots, (K-1)$: Draw

$$\pi(\sigma^2 \,|\, \text{others}) \ \propto\ \frac{1}{\sigma_k^{\sum_{i=1}^{I} I_{(z_i=0)}}\left(1+\sigma_k^2\right)^2} \exp\left\{-\frac{1}{2\sigma_k^2}\sum_{z_i=0}\eta_{ik}^2\right\}.$$

In this model, we suggest using the Metropolis-Hastings algorithm, which is the most commonly used Markov Chain Monte Carlo (MCMC) algorithm used to estimate the value of the location parameter, $(\theta, \boldsymbol{\eta})$. Of course, $(\sigma^2, \sigma_1^2, \ldots, \sigma_{(K-1)}^2)$ is drawn from the above full conditionals, and the Gibbs sampler is performed using the grid method.

B. Computations for nonparameteric models

In order to pool the parameters in nonparametric Bayesian models, we apply the slice sampling mehtod introduced by the Kalli, Griffin and Walker (2011). They proposed an efficient version of the slice sampler for Dirichlet process mixture models constructed by Walker (2007). Suppose that the observations y_i, $i = 1, \ldots, I$ are generated in the Dirichlet process mixture model with parameter θ. That is,

$$y_i \,|\, \mathrm{G} \overset{\text{iid}}{\sim} \mathrm{G}$$
$$\mathrm{G} \sim \mathrm{DP}(\alpha, \mathrm{G}_0)$$

Here, we write $\mathrm{G} \sim \mathrm{DP}(\alpha, \mathrm{G}_0)$ to denote that G follows a Dirichlet process with parameter $\alpha > 0$. Then G has a stick-breaking representation (Sethuraman, 1994) given by

$$P(y_i \,|\, G, \alpha) = \sum_{j=1}^{\infty} w_j f(y_i \,|\, \theta_j)$$

where $\theta_1, \theta_2, \theta_3, \ldots$ are independent and identically distributed (iid) form P_0 and

$$w_1 = \nu_1, \qquad w_j = \nu_j \prod_{l<j} (1 - \nu_l)$$

with the ν_j being iid from $\mathrm{Beta}(1, \alpha)$, see also Antoniak (1974).

The Slice sampler algorithm proposed by Walker (2007) introduces a latent variable $u \in (0, 1)$, $d_i = 1, 2, \ldots$ to perform sampling on the joint distribution. First, the latent variable u_i has a joint density as follow

$$P(y_i, u_i \,|\, G, \alpha) = \sum_{j=1}^{\infty} I_{(u_i < w_j)} f(y_i \,|\, \theta_j).$$

Later, they introduced a latent variable d_i representing the group assignment of the observation i. At this time, the joint density of (y_i, u_i, d_i) is as follow

$$P(y_i, u_i, d_i \,|\, G, \alpha) = \sum_{j=1}^{\infty} I_{(u_i < w_{d_i = j})} f(y_i \,|\, \theta_j).$$

Then we need to sample the parameter θ, α, and ν including latent variables u and d at each iteration of a Gibbs sampler. Kalli et al. (2011) introduces how to perform slice sampling for Dirichlet process mixture models by processing u and ν as blocks in the basic algorithm described by Walker (2007). The algorithm is as follow.

1. $\pi(\theta_j \,|\ldots) \propto G_0(\theta_j) \prod_{d_i = j} f(y_i \,|\, \theta_j)$,
2. $\pi(\nu_j) \propto \mathrm{Beta}(a_j, b_j)$, where $a_j = 1 + \sum_{i=1}^{I} I_{(d_i = j)}$ and $b_j = \alpha + \sum_{i=1}^{I} I_{(d_i > j)}$,
3. $\pi(u_i \,|\ldots) \propto I_{(0 < u_i < \xi_{d_i})}$, where $\xi_j = (1 - \kappa)\kappa^{j-1}$, and κ is a constant,
4. $P(d_i = k \,|\ldots) \propto I_{(k: \xi_k > u_i)} w_k / \xi_k f(y_i \,|\, \theta_k)$,
5. $\pi(\alpha \,|\ldots) \propto \alpha^J \prod_{j=1}^{J} (1 - \nu_j)^{\alpha - 1} \pi(\alpha)$.

For this paper, we need to sample the following variables at each iteration of a Gibbs sampler:

$$\{(\boldsymbol{\pi}_j, \nu_j),\ j = 1, 2, \ldots; (d_i, u_i),\ i = 1, \ldots, I\}.$$

In general, κ is equal to 0.5. However, we use κ as a tuning parameter for the hierarchical Bayesian model. Then, the full posterior density for the nonparametric adaptive pooling of he given data is given as follows:

$$P(\mathbf{n}, \mathbf{d}, \mathbf{u} \mid \boldsymbol{\nu}, \boldsymbol{\pi}) \left\{ \prod_{i=1}^{I} \pi(\boldsymbol{\pi}_i \mid \boldsymbol{\mu}, \tau) \right\} \pi(\boldsymbol{\mu}, \tau)\, \pi(\boldsymbol{\nu} \mid \alpha)\, \pi(\alpha)$$

$$\propto \left\{ \prod_{i=1}^{I} I_{(u_i < \xi_{d_i})} \frac{w_{d_i}}{\xi_{d_i}} n_{i\cdot}! \prod_{k=1}^{K} \frac{\pi_{d_i k}^{n_{ik}}}{n_{ik}!} \right\} \left\{ \prod_{i=1}^{I} \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{k=1}^{K} \pi_{ik}^{\mu_k \tau - 1} \right\}$$

$$\times \left\{ \frac{(K-1)!}{(1+\tau)^2} \right\} \left\{ \prod_{j=1}^{J} \frac{1}{B(1, \alpha)} (1 - \nu_j)^{\alpha - 1} \right\} \frac{1}{(1+\alpha)^2},$$

where $\mathbf{d} = (d_1, \ldots, d_I)$, $\mathbf{u} = (u_1, \ldots, u_I)$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_J)$, and the hyperparameters are mutually independent. Then, our Gibbs sampler can be performed in two steps.

The first step is the pooling of the data:

(a) Draw for u_i from Uniform$(0, \xi_{d_i})$.

(b) Draw for d_i from $P(d_i = j \mid \text{others}) = I_{(u_i < \xi_j)} w_j \Big/ \xi_j \prod_{k=1}^{K} \pi_{jk}^{n_{ik}} \Big/ n_{ik}!$. Next, we can generate the value of each parameter from the following conditional density:

(c) Draw $\boldsymbol{\pi}_j$, $j = 1, \ldots, J$, from Dirichlet$\left(\boldsymbol{\mu}\tau + \sum_{d_i = j} \mathbf{n}_i - 1\right)$.

(d) Draw ν_j, $j = 1, \ldots, J$, from Beta$\left(1 + \sum_{i=1}^{I} I_{(d_i = j)}, \alpha + \sum_{i=1}^{I} I_{(d_i = j)}\right)$.

(e) Draw $\boldsymbol{\mu}$ from

$$\pi(\boldsymbol{\mu} \mid \text{others}) \propto \prod_{j=1}^{J} \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{k=1}^{K} \pi_{jk}^{\mu_k \tau - 1}.$$

(f) Draw τ from

$$\pi(\tau \mid \text{others}) \propto \left\{ \prod_{j=1}^{J} \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{k=1}^{K} \pi_{jk}^{\mu_k \tau - 1} \right\} \frac{(K-1)!}{(\tau+1)^2}.$$

For our Gibbs sampler, we need to transform τ to $\rho = 1/(1+\tau)$, $0 < \rho < 1$ because we need to use the grid method for τ, which is taken from the noninformative prior with variable support equal to $(0, \infty)$.

$$\pi(\rho \mid \text{others}) \propto \left\{ \prod_{j=1}^{J} \frac{1}{D\left(\boldsymbol{\mu} \frac{1-\rho}{\rho}\right)} \prod_{k=1}^{K} \pi_{jk}^{\mu_k \frac{1-\rho}{\rho} - 1} \right\} (K-1)!.$$

(g) Draw α from

$$\pi(\alpha \mid \text{others}) \propto \left\{ \prod_{j=1}^{J} (1 - \nu_j)^{\alpha - 1} \right\} \frac{\alpha^J}{(1+\alpha)^2}.$$

The parameter α is also taken from the noninformative prior with variable support equal to $(0, \infty)$, as in the case of τ. Therefore, we need to transform α to δ, with Jacobian $1/\delta^2$. Then, the conditional density for δ is given as follows:

$$\pi(\delta \mid \text{others}) \propto \left\{ \prod_{j=1}^{J} (1 - \nu_j)^{\frac{1-\delta}{\delta} - 1} \right\} \left(\frac{1-\delta}{\delta} \right)^J.$$

The nonparametric version for the restricted pooling has $\boldsymbol{\pi}_j = (\pi_{j1}, \ldots, \pi_{jK})$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$, $0 \le \mu_k \le 1$, $\sum_{k=1}^{K} \mu_k = 1$, $\mathbf{d} = (d_1, \ldots, d_I)$, and $d_i = j$ for $j = 1, 2, \ldots, J$. Then, we compose the full posterior density for the given data using equation, as follows:

$$P(\mathbf{n},\, \mathbf{d},\, \mathbf{u}\,|\,\boldsymbol{\nu},\, \boldsymbol{\pi})\left\{\prod_{j=1}^{J}\pi\left(\boldsymbol{\pi}_j\,|\boldsymbol{\mu},\, \tau,\, \phi\right)\right\}\pi(\boldsymbol{\mu},\, \tau)\;\pi(\phi)\;\pi(\boldsymbol{\nu}\,|\alpha)\;\pi(\alpha).$$

In our Gibbs sampler, we consider the latent variables z_j, for $j = 1, \ldots, J$, as the parametric version for restricted pooling, where the subscripts for the parameter are equal to the parameter $\boldsymbol{\pi}$. At this time, the pooling step for the data is the same as above, and generating the parameter is as follows:

(a) Draw ϕ from truncated Beta$\left(\sum_{j=1}^{J} z_j,\; J - \sum_{j=1}^{J} z_j,\; \phi \in \left(\frac{1}{2}, 1\right)\right)$.

(b) Draw z_j from Bernoulli$\left(p_j\right)$,

where $p_j = \left(\phi\prod_{k=1}^{K}\pi_{jk}^{\mu_k\tau-1}\Big/D(\boldsymbol{\mu}\tau)\right)\Big/\left\{\phi\prod_{k=1}^{K}\pi_{jk}^{\mu_k\tau-1}\Big/D(\boldsymbol{\mu}\tau) + (1-\phi)\,I_{\left(\sum_{k=1}^{K}\pi_{jk}=1\right)}\Big/K-1\right\}$.

(c) Draw $\boldsymbol{\pi}_j$ from Dirichlet$\left(I_{(z_j=1)}\boldsymbol{\mu}\tau + I_{(z_j=0)}\;\mathbf{1} + \sum_{d_i=j}\mathbf{n}_i\right)$.

(d) Draw ν_j from Beta$\left(1 + \sum_{i=1}^{I} I_{(d_i=j)},\; \alpha + \sum_{i=1}^{I} I_{(d_i>j)}\right)$.

(e) Draw $\boldsymbol{\mu}$ from

$$\pi(\boldsymbol{\mu}\,|\,\text{others}) \;\propto\; \prod_{j=1}^{J}\frac{1}{D(\boldsymbol{\mu}\tau)}\prod_{k=1}^{K}\pi_{jk}^{\mu_k\tau-1}.$$

(f) Draw τ from

$$\pi(\tau\,|\,\text{others}) \;\propto\; \left[\prod_{j=1}^{J}\frac{1}{D(\boldsymbol{\mu}\tau)}\prod_{k=1}^{K}\pi_{jk}^{*\mu_k\tau-1}\right]\frac{(K-1)!}{(\tau+1)^2}.$$

(g) Draw α from

$$\pi(\alpha\,|\,\text{others}) \;\propto\; \left\{\prod_{j=1}^{m}(1-\nu_j)^{\alpha-1}\right\}\frac{\alpha^J}{(1+\alpha)^2}.$$

Lastly, the full posterior density is calculated using the joint posterior density and the priors for their parameters in the nonparametric Bayesian global-local pooling model. Here, the data pooling algorithm is the same as above inference. Then, we estimate each parameter as follows:

(a) Draw ϕ from truncated Beta$\left(\sum_{j=1}^{j} z_j,\; I - \sum_{j=1}^{J} z_j,\; \frac{1}{2},\; 1\right)$.

(b) Draw z_j from Bernoulli$\left(p_j\right)$, where $p_j = \phi\prod_{k=1}^{K-1}\left(\exp\left(-\eta_{jk}^2/2\sigma^2\right)\Big/\sqrt{2\pi\sigma^2}\right)\Big/\left\{\phi\prod_{k=1}^{K-1}\left(\exp(-\eta_{jk}^2/2\sigma^2)\Big/\sqrt{2\pi\sigma^2}\right) + (1-\phi)\prod_{k=1}^{K-1}\left(\exp\left(-\eta_{jk}^2/2\sigma_k^2\right)\Big/\sqrt{2\pi\sigma_k^2}\right)\right\}$.

(c) Draw θ from

$$\pi(\theta\,|\,\text{others}) \;\propto\; \left[\prod_{i=1}^{I}\prod_{k=1}^{K-1}\left(\frac{e^{\theta+\eta_{d_ik}}}{1+\sum_{l=1}^{K-1}e^{\theta+\eta_{d_il}}}\right)^{n_{ik}}\left(\frac{1}{1+\sum_{l=1}^{K-1}e^{\theta+\eta_{d_il}}}\right)^{n_{iK}}\right].$$

(d) Draw for η_{jk}, $k = 1, \ldots, (K-1)$:

$$\pi\left(\eta_{jk}\,|\,\text{others}\right) \;\propto\; I_{(z_j=1)}\left(\frac{1}{1+\sum_{l=1}^{K-1}e^{\theta+\eta_{jl}}}\right)^{\sum_{d_i=j}\sum_{k=1}^{K}n_{ik}}\exp\left\{\sum_{d_i=j}\sum_{k=1}^{K-1}\eta_{jk}n_{ik} - \frac{1}{2\sigma^2}\sum_{k=1}^{K-1}\eta_{jk}^2\right\}$$

$$+\; I_{(z_i=0)}\left(\frac{1}{1+\sum_{l=1}^{K-1}e^{\theta+\eta_{jl}}}\right)^{\sum_{d_i=j}\sum_{k=1}^{K}n_{ik}}\exp\left\{\sum_{d_i=j}\sum_{k=1}^{K-1}\eta_{jk}n_{ik} - \frac{1}{2}\sum_{k=1}^{K-1}\frac{1}{\sigma_k^2}\eta_{jk}^2\right\}.$$

(e) Draw ν_j from $\text{Beta}\left(1 + \sum_{i=1}^{I} I_{(d_i = j)},\ I_{(z_j=1)}\alpha + I_{(z_j=0)}\alpha_0 + \sum_{i=1}^{I} I_{(d_i > j)}\right)$.

(f) Draw σ^2 from

$$\pi(\sigma^2 \mid \text{others}) \propto \frac{1}{\sigma^{K-1\sum_{j=1}^{J} I_{(z_j=1)}} (1+\sigma^2)^2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{z_j=1}\sum_{k=1}^{K-1}\eta_{jk}^2\right\}.$$

(g) Draw σ_k^2 $k = 1, \ldots, (K-1)$ from

$$\pi(\sigma_k^2 \mid \text{others}) \propto \frac{1}{\sigma_k^{\sum_{j=1}^{J} I_{(z_j=0)}} (1+\sigma_k^2)^2} \exp\left\{-\frac{1}{2\sigma_k^2}\sum_{z_j=0}\eta_{jk}^2\right\}.$$

(h) Draw α from

$$\pi(\alpha \mid \text{others}) \propto \frac{\alpha^{\sum_{j=1}^{J} I_{(z_j=1)}}}{(1+\alpha)^2}\left(\prod_{z_j=1}(1-\nu_j)\right)^{\alpha-1}.$$

(i) Draw α_0 from

$$\pi(\alpha_0 \mid \text{others}) \propto \frac{\alpha_0^{\sum_{j=1}^{J} I_{(z_j=0)}}}{(1+\alpha_0)^2}\left(\prod_{z_j=0}(1-\nu_j)\right)^{\alpha_0-1}.$$

References

Agresti, A., and Hitchcock, D.B. (2005). Bayesian inference for categorical data analysis. *Statistical Methods and Applications*, 14, 297-330.

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

Consonni, G., and Veronese, P. (1995). A Bayesian method for combining results from several binomial experiments. *Journal of the American Statistical Association*, 90, 935-944.

DuMouchel, W.H., and Harris, J.E. (1983). Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association*, 78, 293-308.

Dunson, D.B. (2009). Nonparametric Bayes local partition models for random effects. *Biometrika*, 96, 249-262.

Evans, R., and Sedransk, J. (1999). Methodology for pooling subpopulation regressions when sample sizes are small and there is uncertainty about which subpopulations are similar. *Statistica Sinica*, 9, 345-359.

Evans, R., and Sedransk, J. (2003). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain: II. *Journal of Statiatical Planning and Inference*, 111, 95-100.

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

Gneiting, T., and Reftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359-378.

Ishwaran, H., and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96, 161-173.

Kalli, M., Griffin, J.E. and Walker, S.G. (2011). Slice sampling mixture models. *Statistics and Computing*, 21, 93-105.

Leonard, T. (1977). Bayes simultaneous estimation for several multinomial distributions. *Communications in Statistics: Theory and Methods*, 6, 619-630.

Malec, D., and Sedransk, J. (1992). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain. *Biometrika*, 79, 593-601.

Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, 61, 97-112.

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, 3141-3157.

Nandram, B., Zhou, J. and Kim, D.H. (2019). A pooled Bayes test of independence for sparse contingency tables from small area. *Journal of Statistical Computation and Simulation*, 89(5), 899-926.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation*, Second Edition. New York: John Wiley & Sons, Inc.

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

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Linde, A.V.D. (2002). Bayesian measures of model complexity and fit. *Journal of royal statistical society, Series B*, 64(4), 538-639.

Walker, S.G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics-Simulation and Computation*, 36, 45-54.

Yin, J., and Nandram, B. (2020a). A nonparametric Bayesian analysis of response data with gaps, outlier and ties. *Statistics and Applications*, 18(2), 121-141.

Yin, J., and Nandram, B. (2020b). A Bayesian small area model with Dirichlet processes on the responses. *Statistics in Transition*, 21(1), 1-17.