Bayesian Small Area Analyses of the Unrelated Question Design with Multiple Sensitive Questions

by

Yuan Yu

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Statistics

by

April 8, 2019

APPROVED:

Professor Balgobin Nandram, Advisor
Department of Mathematical Sciences
Worcester Polytechnic Institute

Professor Jian Zou
Department of Mathematical Sciences
Worcester Polytechnic Institute

Professor Ewart Thomas
Department of Psychology
Stanford University

Professor Higgins Huong
Business School
Worcester Polytechnic Institute

Dr. Jai Won Choi
Statistical Consultant, Meho Inc., MD

Abstract

Elicitation of answers for sensitive questions is a delicate issue, and even questions of basic demographics (e.g., age, race, sex) can be offensive to some people. In sample surveys with sensitive questions, randomized response techniques have a huge advantage in estimating population quantities (e.g., proportion of people cheating on their tax returns) because they can reduce the bias caused by non-response or untruthful response (measurement error). Using hierarchical Bayesian models, we implement multiple sensitive questions into the simple unrelated question design for small areas (or clusters).

Most of the work on the unrelated question design rely on large sample sizes to get admissible estimates and there are limited discussions about applications on data from small areas. Bayesian methods work well because they allow pooling of data from desperate (limited data) areas and they can utilize prior information. In addition, few discussions have been made exploring the benefits of a combined design involving multiple items (e.g., two sensitive questions) under the Bayesian paradigm. Therefore, in our study, given binary response data from two or more sensitive questions from many small areas, we use a hierarchical Dirichlet-multinomial model to estimate the sensitive proportions. A blocked Gibbs sampler is used to sample the joint posterior density and the posterior distributions of finite population proportions can be obtained. We apply our method to college cheating data, obtained from our students at WPI with permission from IRB. We also use a simulation study to validate our method, and we investigate the effects on posterior inference of increasing the number of areas (clusters) and the correlation between the sensitive items.

When there is a large number of areas, our procedure is computationally intensive. Also, the Dirichlet distribution gives negative correlated probabilities and this is inflexible. Therefore, to make our procedure more useful, we propose a generalized mixed effects model which will set free the constraint of the Dirichlet parameters that must add up

to unity. Then based on the new parameter setting, we are able to either use a full Gibbs sampler or an integrated nested normal approximation to make posterior inference about the finite population proportions of students cheating in different courses. This alternative method allows for much faster computing and many more areas (courses). This model has much fewer parameters, and therefore, there are gains in precision when the finite population proportions are estimated. It also permits incorporating covariates, when available, in a straightforward manner.

Finally, we propose that our randomized response procedure can be used to provide masked public-used data, that is an important activity for many government agencies, where although other procedures are used, the randomized response procedure was never attempted for privacy protection of released data.

Acknowledgements

I would like to express my very great appreciation to my advisor, professor Balgobin Nandram, who not only provided this meaningful topic but also put so much effort to help me get through all the challenges of my dissertation. I am also inspired and encouraged to explore new knowledge during this period under his instruction.

I would also like to thank all of the committee members, professor Jian Zou, Dr. Jai Won Choi, professor Ewart Thomas and professor Joe Sedransk for their time and guidance. I am also grateful to professor Huong Higgins for assistance with the college cheating survey.

I wish to acknowledge all the great programming advice from my fellow colleagues. In addition, my special thanks are extended to all the professors who imparted valuable statistical knowledge to me, and their assistance since I came here, and to all the staff for their help and kindness.

Last but not least, I would like to express my deepest gratitude to my parents and family who always support me unconditionally.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we give an overview of the randomized response techniques together with some extensions of their Bayesian solutions, and some extensions for multiple sensitive items. In addition, we introduce the college cheating data that we will be using through out the dissertation. Finally a dissertation plan is presented.

1.1 Overview of the Randomized Response Technique

In survey sampling, a curious and problematic issue to consider is the collection of information about sensitive questions like habitual tax evasion, drunken driving, gambling, consuming drugs, etc. However, for these sensitive and stigmatizing items, a respondent has great tendency to refuse to answer, or answer untruthfully because of privacy protection. Accordingly, Warner (1965) gave a standard procedure for estimating the proportion of people bearing the sensitive character A, on adopting a suitable randomization device. In Warner's design, each individual is required to play a random game, like flipping a coin, with the probability of getting heads, denoted as p. With the whole procedure unobserved by the interviewer, the respondent chooses one of two questions to answer according to the result of heads or tails. That is, with probability p, the respondent will report the true response of the sensitive question A, and he/she will report

the answer of the opposite question A^C with the probability $1 - p$. This randomized response technique is also called the mirrored question design since the true response to the sensitive question is masked by an opposite one.

In this way the respondents should be more comfortable to answer the question because the investigator can never know which question the respondents are answering. When randomized response techniques are used, a respondent's individual answer is not of interest, rather inference for the population is wanted. Because the respondent does not provide a direct answer to the sensitive question, his/her identity is protected while the true answer to the sensitive question is elicited.

To illustrate this important randomized response technique, we consider for the following questions,

Question 1: I cheated on my income tax return last year, is it true? (A)

Question 2: I did not cheat on my income tax return last year, is it true? (A^C)

Circle your response. [Yes, No]

Let π_A represent the true probability of A in the population; p represent the probability of selecting A; $\lambda = p\pi_A + (1-p)(1-\pi_A)$ represent the proportion of the 'yeses'; y represents the total number of 'yeses' obtained from the sample of n respondents. Figure 1.1 shows Warner's mirrored question design.

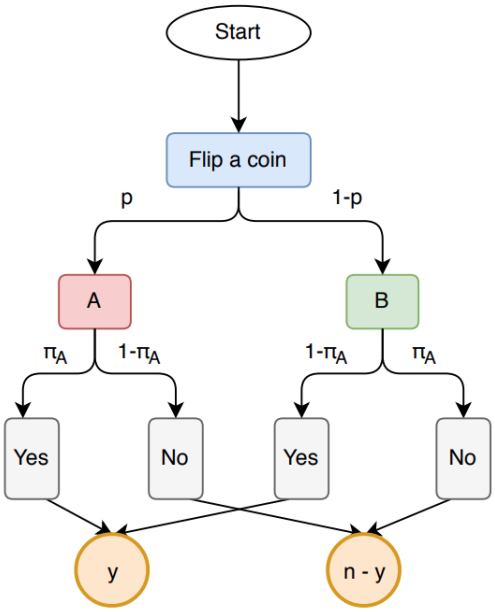


Figure 1.1: Warner's Design

Under the assumption that these 'yes' and 'no' reports are made truthfully and are random, we have

$$y \sim \text{Binomial}\{n, p\pi_A + (1-p)(1-\pi_A)\}.$$

Therefore, the maximum likelihood estimator of π_A is obtained by the solving

$$\hat{\lambda} = p\hat{\pi}_A + (1-p)(1-\hat{\pi}_A),$$

to get

$$\hat{\pi}_A = \frac{1}{2p-1}(\hat{\lambda} + p - 1), \; p \neq \frac{1}{2}, \quad \hat{\lambda} = \frac{y}{n}. \tag{1.1.1}$$

It is known that $\hat{\pi}_A$ is an unbiased estimator of π_A; see Warner (1965).

An important extension of the Warner's mirrored method is the unrelated question design by Greenberg, *et al.* (1969). Instead of the opposite question, an unrelated non-sensitive (innocuous) question is asked. In this design, the true probability of 'yes' of the unrelated characteristic is also unknown. For estimation, two samples are needed. For the individual from each sample, the following two questions are asked,

Question 1: Have you ever cheated in any WPI final exam? (A)

Question 2: Do you like living in Massachusetts? (U)

Circle your response. [Yes, No]

Let π_A be the true probability of 'yes' of A in the population and π_U to be the true probability of 'yes' U in the population. Because there are two unknown parameters, two samples are needed. In the j^{th} sample (j=1, 2), let p_j denote the probability of selecting the sensitive question and $\lambda_j = p_j\pi_A + (1-p_j)(1-\pi_A)$ denote the probability of 'yeses' collected from both questions; y_j is the total number of 'yeses' obtained out of n_j respondents. The unrelated question design is shown in Figure 1.2.

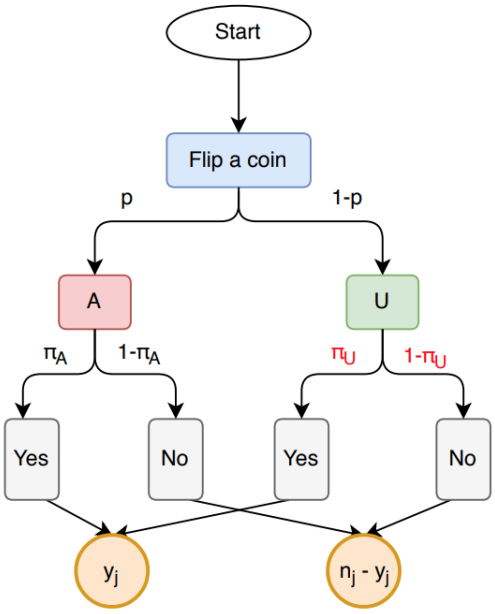


Figure 1.2: Unrelated Question Design

Then we have

$$y_j \overset{ind}{\sim} \text{Binomial}(n_j, p_j\pi_A + (1-p_j)(1-\pi_A)),\ j = 1, 2,$$

the conditions for the log likelihood to get a maximum are

$$\hat{\lambda}_1 = p_1\hat{\pi}_A + (1-p_1)\hat{\pi}_U,$$
$$\hat{\lambda}_2 = p_2\hat{\pi}_A + (1-p_2)\hat{\pi}_U.$$

By solving the set of equations above, one can get the maximum likelihood estimators,

$$\hat{\pi}_A = \frac{\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1)}{p_1 - p_2},\ \hat{\lambda_j} = \frac{y_j}{n_j},\ p_1 \neq p_2. \tag{1.1.2}$$

Here p_1 and p_2 cannot be too close, otherwise the estimation is highly likely to exceed 1. An optimal choice for the random mechanism probability is p_1 in $(0.1, 0.3)$ and $p_2 = 1-p_1$; see Greenberg *et al.* (1969).

A design that is closely related to the unrelated question design is the following. Because two samples are needed, we can run the unrelated question design on one sample, and in the other sample only the nonsensitive question is asked. This design is more efficient than the randomized designs applied to both samples. However, the problem with this design is that the 'nonsensitive' question may be also sensitive to some people. For example, the question "Were you born in Massachusetts?" may be sensitive to some respondents. This leads naturally to optional design in which a respondent is given the option to answer the 'sensitive' question if he/she is comfortable to do so (see Gupta, Gupta and Singh, 2002 and Gupta, Javid and Supriti, 2010 for optional designs for quantitative data).

Direct questioning exposes a respondent's privacy which leads to biased estimates. Any randomized response technique, which adds noise to the response, will be less efficient than a direct questioning design because of the response burden. One cannot compromise respondents' privacy, but one can compromise respondents' burden and efficiency. Besides, it has been argued that socially desirable answers and refusals are expected when sensitive questions are asked directly (e.g., see Tourangeau, Rips and Rasinski, 2000 and

Tourangeau and Yan 2007). Therefore, as supported by many psychologists, sensitive questions should not be asked directly.

We assume that respondents respond truthfully. It should be obvious that this assumption is more easily attained under indirect questioning than the direct one. In direct questioning, it is more likely that there will be nonresponse that may be nonignorable, and we need to develop nonignorable nonresponse models (Nandram and Choi 2002, 2010) to handle them. So on at least two fronts, indirect questioning is preferred.

There is also an important literature of various randomized response techniques. The forced response design, introduced by Boruch (1971), is an extension of the mirrored question design. Fox and Tracy (1986) developed a similar design based on this idea, but using the results coming from two Bernoulli trials instead. The disguised design (Kuk, 1990) is conducted the other way around, in terms of the way that the randomized response (noise) is added. The respondents need to report the results from two Bernoulli trials based on their answers to the sensitive questions. Blair, Imai, and Zhou (2015) gave an excellent review paper on these four methods. Other nonrandomised designs like crosswise and triangular designs (e.g. Tan *et al.*, 2009) can be viewed as extensions of the unrelated question design, which get rid of the random mechanism. See also Nandram and Yu (2018a) for a Bayesian extension.

For continuous response, Greenberg *et al.* (1971) and Eriksson (1973) extended the unrelated question model of Greenberg *et al.* (1969) to the case in which the response is quantitative. Pollock and Bek (1976) described the additive/multiplicative models, which involve the respondents adding/multiplying the answer to the sensitive question by a random number from a known distribution. More recently, optional designs for quantitative data has been discussed in Gupta, Gupta and Singh (2002) and Gupta, Javid and Supriti (2010). It is worth mentioning here that these optional designs for quantitative data have been extended to binary data (e.g., see Gupta *et al.*, 2013 for the unrelated question design).

1.2 Bayesian Randomized Response Technique (RRT)

There are some concerns for the traditional method of getting the maximum likelihood estimation of the proportion. First, direct estimates from solving the equation systems cannot guarantee a reasonable solution between (0,1), and also large sample sizes are needed to get an admissible estimate (e.g. Lee *et al.*, 2013b). In the unrelated question design, $\hat{\pi}_A$ and $\hat{\pi}_U$ can be highly correlated, reducing the correlation by increasing the sample size will be costly. The design-based estimator may be practically biased in small samples. More discussion about the individual Bayesian model is available in Nandram and Yu (2017).

There are not many works within the Bayesian paradigm of randomized response models. Nonetheless, attempts have been made on the Bayesian analysis of designs using randomized response techniques. For example, Winkler and Franklin (1979) gave an approximate Bayesian analysis of Warner's mirrored design, O'Hagan (1987) derived Bayesian linear estimators for the unrelated question design, and Oh (1994) used data augmentation to introduce latent variables to Gibbs sampling of the mirrored design, the unrelated question design and the two-stage designs with binary and polychotomous responses. van den Hout and Klugkist (2009) proposed Bayesian inference that takes into account assumptions with respect to non-compliance under simple random sampling. Also, Tian, Yuen, Tang and Tan (2009) proposed Bayesian approaches to non-randomized response models without using random mechanisms. Avetisyan and Fox (2012) used beta-binomial and multinomial-Dirichlet models for empirical Bayes analysis and to a less extent Bayesian analysis for a small sample from a single population. They considered multiple items with multiple categories, but this latter work is not within the small area context, although each person may be considered a small area, see details about the small area estimation by Rao and Molina (2015). Most recently, Song and Kim (2017) gave a Bayesian analysis of two unrelated questions with rare outcomes (i.e., Poisson modeling rather Binomial modeling). Bayesian methods, with useful prior information, deserve

more attention because it is easy to obtain proper estimates. In addition, hierarchical Bayesian models can be used to study data arising from sample surveys with randomized responses.

In Nandram and Yu (2017, 2018a), binary data are collected from a single small area using a version of the unrelated-question design, and the sensitive proportion is of interest. With a random mechanism of probability p_j for j^{th} group of size n_j, the Bayesian model is built for each area independently (see Oh, 1994),

$$y_j \mid \pi_1, \pi_2 \overset{ind}{\sim} \text{Binomial}\{n_j, p_j\pi_1 + (1-p_j)\pi_2\},$$

where $0 \leq y_j \leq n_j,\ j = 1, \ldots, g \geq 2$. Note that g does not have to be exactly two. With a flat conjugate prior

$$\pi_1 \overset{ind}{\sim} \text{Beta}(1,1) \quad \pi_2 \overset{ind}{\sim} \text{Beta}(1,1).$$

Then apply the same model repeatedly on ℓ areas will form the individual area model (IAM). Obviously, the IAM does not borrow information across the areas. In fact, we run a sensitivity analysis of the different choices of $p_1,\ p_2$ under different sample sizes, to illustrate the fact that IAM will be very sensitive to the choice of the random mechanism when sample sizes are small.

The execution of the randomized response technique requires a known random mechanism, which is p_1 and p_2. In the previous simulation section, for group j, we use p_j as the probability to answer the sensitive questions and $(1-p_j)$ to answer the nonsensitive questions. A lot of studies on the different choices of p have been carried out for the under the non-Bayesian model.

In Table 1.1, we provide a Bayesian sensitivity study for $\pi_1 = 0.25$ both on the different choices of p_1, p_2 and the group sample size ℓ within each group. The sensitivity analysis shows that when the sample size is reasonable large ($\ell = 300, 400$), the estimation is not sensitive to most of the choices of p_1 and p_2, except for the combination of (0.3,0.7), where the estimation is 0.177 and 0.220 respectively. However, if the sample size is not large enough (especially below 100), certain combination choices of p_1, p_2 will influence the estimation.

Table 1.1: Sensitivity analysis of π_1 with respect to choices of p_1, p_2 and ℓ.

		ℓ					
p_1	p_2	25	50	100	200	300	400
0.1	0.7	$.040_{.000}$	$.021_{.015}$	$.083_{.037}$	$.202_{.051}$	$.239_{.045}$	$.243_{.039}$
0.1	0.8	$.012_{.006}$	$.053_{.027}$	$.171_{.053}$	$.243_{.044}$	$.248_{.036}$	$.250_{.031}$
0.1	0.9	$.030_{.020}$	$.121_{.050}$	$.233_{.051}$	$.248_{.037}$	$.250_{.030}$	$.248_{.026}$
0.2	0.7	$.054_{.000}$	$.011_{.006}$	$.045_{.025}$	$.170_{.052}$	$.228_{.049}$	$.243_{.043}$
0.2	0.8	$.027_{.003}$	$.015_{.011}$	$.110_{.044}$	$.231_{.047}$	$.247_{.038}$	$.248_{.033}$
0.2	0.9	$.017_{.012}$	$.086_{.038}$	$.203_{.051}$	$.247_{.037}$	$.249_{.031}$	$.249_{.026}$
0.3	0.7	$.067_{.000}$	$.025_{.003}$	$.025_{.016}$	$.089_{.040}$	$.177_{.050}$	$.220_{.049}$
0.3	0.8	$.053_{.002}$	$.013_{.009}$	$.056_{.030}$	$.174_{.049}$	$.235_{.042}$	$.245_{.035}$
0.3	0.9	$.034_{.011}$	$.058_{.030}$	$.165_{.048}$	$.231_{.038}$	$.244_{.031}$	$.248_{.027}$

Table 1.2 provides a Bayesian sensitivity study for $\pi_2 = 0.75$ on the different choices of p_1, p_2 and the sample size ℓ within each group. The sensitivity analysis also shows that when the sample size is reasonable large ($\ell = 300, 400$), the estimation is not sensitive to the choices of p_1 and p_2. However, if the sample size is not large enough, certain combination choices of p_1, p_2 will overestimate π_2 more or less.

Table 1.2: Sensitivity analysis of π_2 with respect to choices of p_1, p_2 and ℓ.

		ℓ					
p_1	p_2	25	50	100	200	300	400
0.1	0.7	$.962_{.013}$	$.936_{.031}$	$.835_{.047}$	$.767_{.038}$	$.755_{.031}$	$.753_{.027}$
0.1	0.8	$.979_{.012}$	$.919_{.037}$	$.797_{.052}$	$.755_{.037}$	$.752_{.030}$	$.750_{.026}$
0.1	0.9	$.969_{.020}$	$.878_{.049}$	$.765_{.052}$	$.753_{.037}$	$.751_{.030}$	$.750_{.026}$
0.2	0.7	$.947_{.002}$	$.987_{.008}$	$.950_{.027}$	$.818_{.049}$	$.765_{.042}$	$.755_{.035}$
0.2	0.8	$.973_{.003}$	$.987_{.011}$	$.890_{.044}$	$.766_{.047}$	$.757_{.038}$	$.754_{.033}$
0.2	0.9	$.989_{.005}$	$.942_{.028}$	$.831_{.052}$	$.758_{.045}$	$.753_{.036}$	$.752_{.031}$
0.3	0.7	$.933_{.000}$	$.975_{.003}$	$.976_{.016}$	$.911_{.040}$	$.822_{.050}$	$.781_{.049}$
0.3	0.8	$.945_{.000}$	$.990_{.007}$	$.951_{.027}$	$.835_{.052}$	$.771_{.049}$	$.756_{.043}$
0.3	0.9	$.966_{.000}$	$.983_{.012}$	$.913_{.038}$	$.802_{.051}$	$.764_{.045}$	$.753_{.039}$

In Table 1.3, we keep π_1 fixed at 0.25 and modify π_2 from 0.1 to 0.9 to see the estimation effect, under different random mechanisms and sample sizes. We can make a similar conclusion that large sample will make the estimation for π_1 not that sensible to the values of π_2. For extreme values like 0.1 and 0.9, there is still sensitivity, for example under the random mechanism $(p_1, p_2) = (0.3, 0.7)$, if $\pi_2 = 0.1$ or 0.9, even the sample of size 400 large cannot guarantee a very close estimation of 0.25. But it appears that the sensitivity will wash out for even larger sample size.

Table 1.3: Sensitivity analysis of π_1 with respect to choices of π_2 and ℓ under different random mechanisms.

	ℓ					
π_2	25	50	100	200	300	400
a. $(p_1, p_2) = (0.2, 0.8)$						
0.1	$.285_{.077}$	$.298_{.061}$	$.315_{.049}$	$.303_{.040}$	$.278_{.033}$	$.264_{.029}$
0.25	$.287_{.056}$	$.238_{.046}$	$.245_{.059}$	$.250_{.043}$	$.247_{.034}$	$.251_{.030}$
0.5	$.143_{.012}$	$.078_{.034}$	$.182_{.055}$	$.243_{.045}$	$.249_{.037}$	$.249_{.032}$
0.75	$.027_{.003}$	$.015_{.011}$	$.110_{.044}$	$.231_{.047}$	$.247_{.038}$	$.248_{.033}$
0.9	$.005_{.003}$	$.021_{.014}$	$.083_{.035}$	$.174_{.041}$	$.201_{.035}$	$.216_{.032}$
b. $(p_1, p_2) = (0.3, 0.7)$						
0.1	$.335_{.082}$	$.283_{.059}$	$.294_{.044}$	$.323_{.038}$	$.324_{.035}$	$.313_{.033}$
0.25	$.358_{.044}$	$.258_{.042}$	$.248_{.034}$	$.255_{.058}$	$.248_{.049}$	$.248_{.042}$
0.5	$.212_{.003}$	$.141_{.009}$	$.075_{.032}$	$.165_{.055}$	$.220_{.053}$	$.243_{.046}$
0.75	$.067_{.000}$	$.025_{.003}$	$.025_{.016}$	$.089_{.040}$	$.177_{.050}$	$.220_{.049}$
0.9	$.022_{.002}$	$.011_{.004}$	$.034_{.017}$	$.088_{.034}$	$.147_{.038}$	$.168_{.035}$

When sample sizes are small, IAM is apparently not applicable, thus we come up with a small area model (SAM) instead in order to combine information across the areas. In Nandram and Yu (2018b), a hierarchical Bayesian model is used to capture the variation

in the observed binomial counts from the clusters within the small areas and to estimate the sensitive proportions for all areas. The SAM is

$$y_{ij} \mid \pi_{i1}, \pi_{i2} \overset{ind}{\sim} \text{Binomial}\{n_{ij}, p_{ij}\pi_{i1} + (1-p_{ij})\pi_{i2}\},\ i = 1, \ldots, \ell,\ j = 1, \ldots, g_i \geq 2, \quad (1.2.1)$$

$$\pi_{i1}|\mu_1, \tau \overset{ind}{\sim} \text{Beta}(\mu_1\tau, (1-\mu_1)\tau) \quad \pi_{i2}|\mu_2, \tau \overset{ind}{\sim} \text{Beta}(\mu_2\tau, (1-\mu_2)\tau),\ i = 1, \ldots, \ell,$$

$$\pi(\mu_1, \mu_2, \tau) = \frac{1}{(1+\tau)^2},\ 0 < \mu_1, \mu_2 < 1,\ \tau > 0.$$

The subscription $i = 1, \ldots, \ell$ is added to indicate the corresponding parameters from the i^{th} area. Using Bayes' theorem, we have the joint posterior density,

$$\pi(\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \mu_1, \mu_2, \tau_1, \tau_2 \mid \underset{\sim}{y}) \propto$$

$$\prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \{p_{ij}\pi_{i1} + (1-p_{ij})\pi_{i2}\}^{y_{ij}} \{p_{ij}(1-\pi_{i1}) + (1-p_{ij})(1-\pi_{i2})\}^{n_{ij}-y_{ij}} \right]$$

$$\times \frac{1}{(1+\tau_1)^2} \frac{1}{(1+\tau_2)^2} \prod_{i=1}^{\ell} \frac{\pi_{i1}^{\mu_1\tau_1-1}(1-\pi_{i1})^{(1-\mu_1)\tau_1-1}}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)} \frac{\pi_{i2}^{\mu_2\tau_2-1}(1-\pi_{i2})^{(1-\mu_2)\tau_2-1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}. \quad (1.2.2)$$

Latent variables z_{ij} and w_{ij} are introduced to deal with the difficulty involving the additional term in (1.2.1); consequently a blocked Gibbs sampler is constructed for the augmented joint posterior density,

$$\pi(\underset{\sim}{z}, \underset{\sim}{w}, \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \mu_1, \mu_2, \tau_1, \tau_2 \mid \underset{\sim}{y}) \propto$$

$$\prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \binom{y_{ij}}{z_{ij}} (p_{ij}\pi_{i1})^{z_{ij}} \{(1-p_{ij})\pi_{i2}\}^{y_{ij}-z_{ij}} \right.$$

$$\left. \times \binom{n_{ij}-y_{ij}}{\omega_{ij}} \{p_{ij}(1-\pi_{i1})\}^{w_{ij}} \{(1-p_{ij})(1-\pi_{i2})\}^{n_{ij}-y_{ij}-w_{ij}} \right]$$

$$\times \frac{1}{(1+\tau_1)^2} \frac{1}{(1+\tau_2)^2} \prod_{i=1}^{\ell} \frac{\pi_{i1}^{\mu_1\tau_1-1}(1-\pi_{i1})^{(1-\mu_1)\tau_1-1}}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)} \frac{\pi_{i2}^{\mu_2\tau_2-1}(1-\pi_{i2})^{(1-\mu_2)\tau_2-1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}.$$

The basic scheme is to draw the latent variables from the conditional posterior distribution

given all the other parameters. Then all the other parameters can be drawn as a whole block from the conditional joint posterior given the latent variables. Nandram and Yu (2018b) also demonstrated posterior propriety under noninformative priors for μ_1 and μ_2.

Theorem

The joint posterior density (1.2.2) is proper for any prior of the form $\pi(\mu_1, \mu_2, \tau_1, \tau_2) \propto \{\mu_1(1-\mu_1)\}^{-s_1}\{\mu_2(1-\mu_2)\}^{-s_2}\pi(\tau_1, \tau_2)$, where $0 \leq s_1, s_2 < \ell$ and $\pi(\tau_1, \tau_2)$ is proper.

Proof

Let $z_{i\cdot} = \sum_{j=1}^{g_i} z_{ij}$, $w_{i\cdot} = \sum_{j=1}^{g_i} w_{ij}$, $n_{i\cdot} = \sum_{j=1}^{g_i} n_{ij}$ and $y_{i\cdot} = \sum_{j=1}^{g_i} y_{ij}, i = 1, \ldots, \ell$. Then, integrating out π_{i1} and π_{i2}, we get

$$\pi(\underset{\sim}{z}, \underset{\sim}{w}, \mu_1, \mu_2, \tau_1, \tau_2 \mid \underset{\sim}{y}) \propto \{\mu_1(1-\mu_1)\}^{-s_1}\{\mu_2(1-\mu_2)\}^{-s_2}\frac{1}{(1+\tau_1)^2}\frac{1}{(1+\tau_2)^2}$$

$$\times \prod_{i=1}^{\ell}\prod_{j=1}^{g_i}\left[\frac{\binom{y_{ij}}{z_{ij}}\binom{n_{ij}-y_{ij}}{w_{ij}}}{\binom{n_{ij}}{z_{ij}+w_{ij}}}\binom{n_{ij}}{z_{ij}+w_{ij}}p_{ij}^{z_{ij}+w_{ij}}(1-p_{ij})^{n_{ij}-z_{ij}-w_{ij}}\right]$$

$$\times \prod_{i=1}^{\ell}\left\{\frac{B(z_{i\cdot}+\mu_1\tau_1, w_{i\cdot}+(1-\mu_1)\tau_1)}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)}\frac{B(y_{i\cdot}-z_{i\cdot}+\mu_2\tau_2, n_{i\cdot}-y_{i\cdot}-w_{i\cdot}+(1-\mu_2)\tau_2)}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}\right\}.$$

Under the double product, the first term is a hypergeometric probability and the second term is a binomial probability, and so these terms are bounded uniformly in z_{ij} and w_{ij}. Therefore,

$$\pi(\underset{\sim}{z}, \underset{\sim}{w}, \mu_1, \mu_2, \tau_1, \tau_2 \mid \underset{\sim}{y}) \leq \{\mu_1(1-\mu_1)\}^{-s_1}\{\mu_2(1-\mu_2)\}^{-s_2}\frac{1}{(1+\tau_1)^2}\frac{1}{(1+\tau_2)^2}$$

$$\times \prod_{i=1}^{\ell}\left\{\frac{B(z_{i\cdot}+\mu_1\tau_1, w_{i\cdot}+(1-\mu_1)\tau_1)}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)}\frac{B(y_{i\cdot}-z_{i\cdot}+\mu_2\tau_2, n_{i\cdot}-y_{i\cdot}-w_{i\cdot}+(1-\mu_2)\tau_2)}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}\right\}.$$

Assuming that $z_{i\cdot} \geq 1$ and $w_{i\cdot} \geq 1$, and $y_{i\cdot} > z_{i\cdot}$ and $n_{i\cdot} - y_{i\cdot} > w_{i\cdot}$, it is easy to show that

$$\pi(\underset{\sim}{z}, \underset{\sim}{w}, \mu_1, \mu_2, \tau_1, \tau_2 \mid \underset{\sim}{y}) \leq \frac{1}{(1+\tau_1)^2}\frac{1}{(1+\tau_2)^2}\{\mu_1(1-\mu_1)\}^{\ell-s_1}\{\mu_2(1-\mu_2)\}^{\ell-s_2}.$$

Therefore, any improper prior on μ_1 and μ_2 of the form, $\pi(\mu_1, \mu_2) \propto \{\mu_1(1-\mu_1)\}^{-s_1}\{\mu_2(1-\mu_2)\}^{-s_2}$, where $s_1, s_2 \leq \ell$, works, and because the $z_{i\cdot}$ and $w_{i\cdot}$ are bounded (finite ranges),

$$\pi(\tau_1, \tau_2 \mid \underset{\sim}{y}) \leq \frac{1}{(1+\tau_1)^2}\frac{1}{(1+\tau_2)^2}$$

is proper and the joint posterior density is proper.

Nandram and Yu (2018b) performed two examples, one on college cheating and the other, a simulation study, to show significant reductions in the posterior standard deviations of the sensitive proportions under the small-area model as compared to the corresponding individual-area model (Nandram and Yu 2017, 2018a). The simulation study also demonstrates that the estimates under the small-area model are closer to the truth than for the corresponding estimates under the individual-area model. We are also working on a new design that does not need specifications of the random mechanism p_{ij}. Our objective is to extend this model to multiple sensitive items.

1.3 RRT for Multiple Sensitive Items

In many instances in reality, multiple sensitive questions are provided all together in a survey, which leads to various work focusing on estimating the correlations or covariance matrix between different attributes, see Bellhouse (1995). Edgell *et al.* (1986) provided a further statistical efficiency study about the correlation. Kwan *et al.* (2010) used a method-of-moments approach to estimate the covariance matrix of sensitive quantitative attributes. A more recent paper of Chung *et al.* (2018) made causal inference among the sensitive attributes. They showed that, based on two loss functions of the covariance matrix estimators, their Bayesian RRT outperforms Kwan's (2010). However, there are limited discussions on how the correlation between the sensitive questions will influence the estimation accuracy, especially for the small area data.

Randomized response techniques (RRT) currently proposed for the estimation of multiple sensitive items are based on repeated applications of the randomized response pro-

cess (e.g., Barksdale, 1971; Clickner and Iglewicz, 1976), where the RRT is implemented separately for each question pair. Tamhane (1981) has reviewed some of these proposed repeated randomized response procedures and proposed a new technique called multiple randomized response trials technique, which is based on the repeated Warner's technique and Simmon's technique earlier proposed by Barksdale (1971). However, Tamhane (1981) finds that even though the use of repeated applications of RRT provide estimates of joint proportions, they will be costly and result in lower degree of cooperation even with only three or four repetitions of RRT. Truthfulness in responding may depend on which questions were answered on previous trials (see Barksdale, 1971).

Further more, Clickner and Iglewicz (1976), demonstrated that the extension to itemwise RRT procedures will increase the variance of the estimate of joint probability. Soeken and Macready (1986) proposed a setwise RRT, which uses a single randomization across sets of paired questions. They found that the estimated proportions of positive responses to sensitive items are no more negatively biased under setwise RRT than those obtained assuming itemwise RRT, though there was no further discussion about the correlation effect. Actually the setwise RRT and itemwise RRT are two approaches that we will refereed as the "combined model" and "separate question model" and compare under the Bayesian scheme later in Chapter 2. However, we use different notifications as combined model and separate question model instead through out the dissertation. Lee *et al.* (2013a), have created crossed Warner's design to get estimates of two sensitive proportions, which is more efficient than applying basic Warner's design twice on each one. Pal (2017) also suggest a bootstrap technique in dealing with complex randomized response surveys with two sensitive characters.

A natural extension of the unrelated question design to the multiple sensitive items is to ask all at the same time. In this dissertation, with this multi-item unrelated question design, an attempt is made to assess the correlation effect in estimating the marginal proportion (proportion of a single sensitive attribute) under the Bayesian paradigm.

1.4 College Cheating Data

Academic cheating is a serious problem in college. In the cheating proportion study, a survey with direct questions will obviously tend to underestimate the sensitive proportion. Some studies show that even when the conventional anonymous survey is conducted, the under reporting still exists. Scheers and Dayton (1978) compared the cheating proportion estimation from basic unrelated question design for single question with those from the anonymous survey, indicating that the anonymous questionnaire is an inadequate data collection device when a survey involves sensitive issue.

We will use the unrelated question design for multi-item sensitive questions to conduct a real survey to study the college cheating problem through which we can make inference about cheating proportions in the final exam and the proportions of the students with the ambition of getting a high GPA between 3.5 to 4.0. These are the two sensitive features that of interest, even though asking about students' ambition to get a high GPA may not be that sensitive compare to cheating, it could still be sensitive to some students. We are also interested in the proportion of having both sensitive features (e.g. 'yes' in both questions).

Therefore we proposed the following design for two sensitive questions. First, the sensitive question are set as "Have you ever cheated in any WPI final exam?" and "Are you very eager to get a GPA between 3.5 and 4.0?". To pair up with the sensitive questions, two unrelated questions are necessary so that the answers have the same dimension. In our design, we use "Do you like living in Massachusetts?" and "Do you like snow during winter?" as the nonsensitive questions.

The data are collected from students of 15 class sections at WPI, including both undergraduate and graduate courses, with some sections from the same course. For each section (area), the students are divided into two groups almost evenly. They are given a die to generate the random number between 1 to 6, the first group of students will answer the sensitive questions when they get 3, 4, 5, 6; otherwise they should answer the two

nonsensitive questions. For another group, they will answer the sensitive questions when 1, 2 come out, otherwise they go to the nonsensitive ones. In the last step, they need to provide their answers in one of the following types (No, No), (No, Yes), (Yes, No), (Yes, Yes). Since the survey still involves the sensitive topic, the data are collected under the approval of university Internal Review Board of WPI. The questionnaire is carefully designed under the help of professor Higgins, making the students feel comfortable to give out the honest responses. Two types of the questionnaires (Type I and Type II) with a different random mechanism are sent out to students evenly in each area, attached in the end of this section.

Table 1.4 provides the counts data from 15 class sections, each section having two sample groups. For example, in section 1, among the 14 students of first group, no one gives answer of (No, No), 10 students give (Yes, No), 2 students give (Yes, No) and the other 2 students give (Yes, Yes). Interest is on estimating various proportions of students cheating in their final exams and the proportions of the students who are eager to get high GPA between 3.5 to 4.0. Through out this dissertation, we will use this college cheating dataset for different models. However, the counts data are very sparse.

Questionnaire - Type I

Declaration: This is a nonprofit survey designed to supply real data for a PhD dissertation at WPI. It is an anonymous survey designed to protect your privacy. The researchers cannot know who submits answers to which questions. The researchers will use the data to advance research on "survey methods", not on "student cheating".

To begin with, please indicate your categories:

- ☐ Male.
- ☐ Female.
- ☐ Other.
- ☐ Prefer not to answer.
- ☐ I (or a family member) have worked on an undergraduate degree from WPI.
- ☐ I have an undergraduate degree elsewhere.

Step 1. Throw a die only once and make sure that the result is known to you, not the investigator.

Step 2. Based on the result above,

If you get **1** or **2**, please go to the questions below (**School Questions**) and keep your answers in mind. **Do not mark the answers in this box.** Please write your answers in Step 3.	If you get **3, 4, 5, 6**, please go to the questions below (**Life Questions**) and keep your answers in mind. **Do not mark the answers in this box.** Please write your answers in Step 3.
School Questions: 1. Have you ever cheated in any WPI final exam? Yes or No. 2. Are you very eager to get a GPA between 3.5 and 4.0? Yes or No.	**Life Questions:** 1. Do you like living in Massachusetts? Yes or No. 2. Do you like snow during winter? Yes or No.

(Note: The definition of cheating is based on the WPI academic dishonesty policy. DO NOT answer both school and life questions.)

Step 3. Now mark your answers by checking one of the boxes below. (For example, (No, Yes) means you answer "No" to the 1st question and "Yes" to the 2nd question.)

☐ (No, No) ☐ (No, Yes) ☐ (Yes, No) ☐ (Yes, Yes)

Questionnaire - Type II

Declaration: This is a nonprofit survey designed to supply real data for a PhD dissertation at WPI. It is an anonymous survey designed to protect your privacy. The researchers cannot know who submits answers to which questions. The researchers will use the data to advance research on "survey methods", not on "student cheating".

To begin with, please indicate your categories:

- ☐ Male.
- ☐ Female.
- ☐ Other.
- ☐ Prefer not to answer.
- ☐ I (or a family member) have worked on an undergraduate degree from WPI.
- ☐ I have an undergraduate degree elsewhere.

Step 1. Throw a die only once and make sure that the result is known to you, not the investigator.

Step 2. Based on the result above,

If you get **3, 4, 5, 6**, please go to the questions below (**School Questions**) and keep your answers in mind. **Do not mark the answers in this box.** Please write your answers in Step 3.	If you get **1** or **2**, please go to the questions below (**Life Questions**) and keep your answers in mind. **Do not mark the answers in this box.** Please write your answers in Step 3.
School Questions: 1. Have you ever cheated in any WPI final exam? Yes or No. 2. Are you very eager to get a GPA between 3.5 and 4.0? Yes or No.	**Life Questions:** 1. Do you like living in Massachusetts? Yes or No. 2. Do you like snow during winter? Yes or No.

(Note: The definition of cheating is based on the WPI academic dishonesty policy. DO NOT answer both school and life questions.)

Step 3. Now mark your answers by checking one of the boxes below. (For example, (No, Yes) means you answer "No" to the 1st question and "Yes" to the 2nd question.)

☐ (No, No) ☐ (No, Yes) ☐ (Yes, No) ☐ (Yes, Yes)

Table 1.4: Counts data collected from the questionnaire of 15 class sections from WPI.

Section	Group	(No, No)	(No, Yes)	(Yes, No)	(Yes, Yes)
1	1	0	10	2	2
	2	1	3	4	5
2	1	0	5	1	0
	2	1	2	1	2
3	1	3	5	0	5
	2	0	7	1	3
4	1	1	3	0	6
	2	1	3	1	4
5	1	0	4	0	0
	2	0	3	1	0
6	1	1	5	1	0
	2	2	2	1	2
7	1	0	6	0	2
	2	0	2	0	5
8	1	1	6	2	2
	2	1	2	2	5
9	1	2	6	0	1
	2	3	1	1	4
10	1	2	8	0	1
	2	0	4	4	5
11	1	1	0	1	4
	2	0	3	2	1
12	1	2	5	2	3
	2	4	2	1	4
13	1	0	4	2	3
	2	0	4	0	6
14	1	3	12	2	4
	2	3	8	1	9
15	1	4	8	2	8
	2	1	12	4	8

1.5 Dissertation Plan

This dissertation serves as a multi-question extension to the work of Nandram and Yu (2018b). We gain estimation strength by joining multiple sensitive questions and pooling

small areas. More areas involved is preferred and the correlation effect between the questions has also been explored. Another contribution is that we propose a parsimonious generalized mixed effects model to to reduce the variability and improve the computing efficiency. The plan of the rest of the dissertation is as follows.

In Chapter 2, a hierarchical Bayesian model for the unrelated question design of two sensitive questions and the computational methodology are presented. In addition, we present same data analysis on the college cheating data using our Bayesian model and provide a simulation study for three different models with different number of areas and correlations.

In Chapter 3, we study the generalized mixed effects model which let us cope with the intensive computing when large number of areas involved. Meanwhile it also allows us to use integrated nested normal approximation (INNA). Applying to the college cheating data, we also present the four-cell proportions estimated from the exact generalized mixed effects model and its normal approximate model. This leads to improved precision.

Finally, in Chapter 4, concluding remarks and the future work are presented. In addition, we state how our randomized response procedure can be used to provide masked data that have the same sensitive (non-sensitive) proportions that the original data would provide.

Chapter 2

Unrelated Question Design for Multi-sensitive Items

In this chapter, we discuss a Bayesian methodology to analyze the binary response data collected from the combined sensitive questions design for college cheating. Assume that we have more than one sensitive question from small areas. For each area i, we sample g_i (≥ 2) groups of people, letting them flip an unfair coin, or playing another random game chosen by the interviewer. Depending on different results (heads or tails), the respondents will answer different set of questions either sensitive or non-sensitive. Let p_{ij} denote the success probability of jth group (cluster) from ith area, so with the probability p_{ij} the respondents will get the chance to answer the combined sensitive questions $S_1\&S_2$; otherwise they should answer the non-sensitive questions $N_1\&N_2$. Then we collect the binary response data of four types: (No, No), (No, Yes), (Yes, No), (Yes, Yes), for example (No, Yes) refer to the response to the first and second questions. Let $\underset{\sim}{y}_{ij} = (y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4})$ denote the corresponding counts, $i = 1 \ldots \ell, j = 1 \ldots g_i$, with $\ell = 15$ and $g_i = 2$ in the college cheating data. Figure 2.1 shows the unrelated question design for multiple sensitive items.

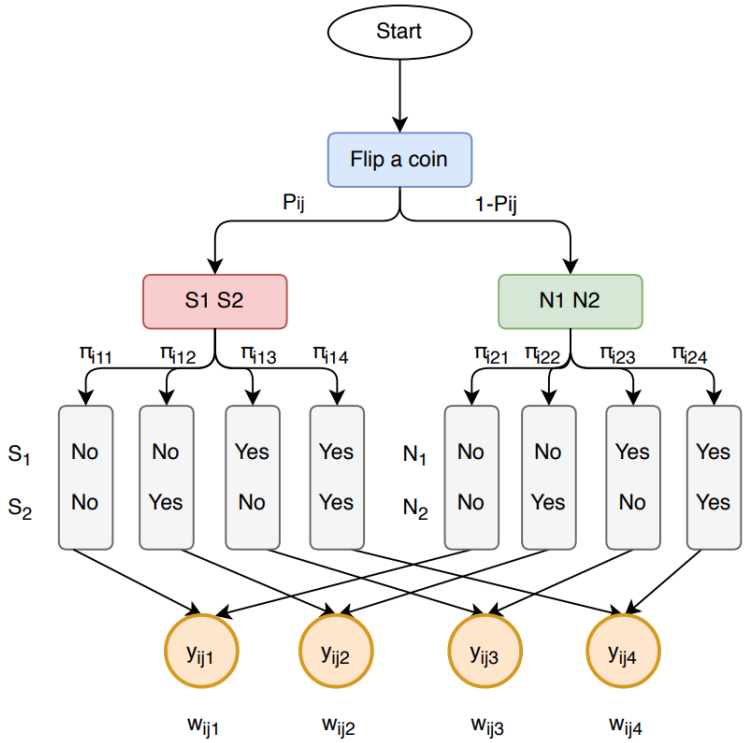


Figure 2.1: Unrelated Question Design

Since the interviewer does not know which branch the response comes from, the respondents will feel more comfortable to give the true response to the sensitive questions, which will lead to a more accurate estimation of the sensitive population proportion $\pi_{\underset{\sim}{i1}}$. This will lead to the estimation of the sensitive population proportions $\phi_{i11} = \pi_{i13} + \pi_{i14}$, the proportion of S_1; $\phi_{i12} = \pi_{i12} + \pi_{i14}$ the proportion of S_2, with $\underset{\sim}{\pi}_{i1} = (\pi_{i11}, \pi_{i12}, \pi_{i13}, \pi_{i14})$ representing the true proportions of four categories coming from sensitive questions (S_1&S_2). If applying the traditional method by solving equations (Greenberg, *et al.*, 1969), exactly two groups of samples are needed for each area. Then we may solve the system of eight equations

$$\frac{y_{ijk}}{n_{ij}} = p_{ij}\hat{\pi}_{i1k} + (1 - p_{ij})\hat{\pi}_{i2k}, \;\; j = 1, \;\; 2, k = 1, \ldots, 4, \tag{2.0.1}$$

subject to the constraints $\sum_{k=1}^{4} \hat{\pi}_{ijk} = 1, j = 1, 2.$ where $n_{ij} = \sum_{k=1}^{4} y_{ijk}$. But note that a proper solution is not guaranteed. Another concern would be, even though college

cheating data $g_i = 2$, in reality, more than two groups of samples maybe available ($g_i > 2$), thus the above method cannot be applied directly.

2.1 Hierarchical Bayesian Model

In order to estimate the population proportion for each area, instead of combining the equation system which may not guarantee a solution, we propose a three-stage Bayesian model with latent variables. It is natural to think of the count data from the four types of responses for each sample group follow a multinomial distribution with the four cell probabilities given above. Based on that, we developed the hierarchical Bayesian model to solve the problem.

2.1.1 Joint Posterior Density

Generally, the model consists of three stages. First,

$$\underset{\sim}{y}_{ij} \mid \underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \overset{ind}{\sim} \text{Multinomial}\{n_{ij}, \underset{\sim}{a}_{ij}\}, i = 1, \ldots, \ell, j = 1, \ldots, g_i,$$

where

$$\underset{\sim}{a}_{ij} = \Big(p_{ij}\pi_{i11}+(1-p_{ij})\pi_{i21}, p_{ij}\pi_{i12}+(1-p_{ij})\pi_{i22}, p_{ij}\pi_{i13}+(1-p_{ij})\pi_{i23}, p_{ij}\pi_{i14}+(1-p_{ij})\pi_{i24}\Big)$$

represents the four cell probabilities for each group, with each cell probability coming from two sources, sensitive and nonsensitive. Secondly, the parameters ($\underset{\sim}{\pi}_{i1}$, $\underset{\sim}{\pi}_{i2}$) represent the inherent probabilities within each area, and they follow independent Dirichlet distribution given the hyper parameters ($\underset{\sim}{\mu}_1$, $\underset{\sim}{\mu}_2$, τ),

$$\underset{\sim}{\pi}_{i1} \overset{ind}{\sim} \text{Dirichlet}\{\underset{\sim}{\mu}_1\tau\} \text{ and } \underset{\sim}{\pi}_{i2} \overset{ind}{\sim} \text{Dirichlet}\{\underset{\sim}{\mu}_2\tau\} \ .$$

Here $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ represent probabilities with all areas grouped together. Here τ can be treated as a prior sample size that plays a part in weighting the parameters of the Dirichlet distribution. For simplicity, we assume that there is no difference between $\underset{\sim}{\pi}_{i1}$ and $\underset{\sim}{\pi}_{i2}$ for sample size, we only use one τ instead of τ_1, τ_2. Notice that $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ and τ are independent, $\underset{\sim}{\mu}_1 \sim \text{Dirichlet}(1,1,1,1)$ and $\underset{\sim}{\mu}_2 \sim \text{Dirichlet}(1,1,1,1)$, so a priori,

$$\pi(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau) = \frac{1}{(1+\tau)^2} \frac{1}{[\text{D}(1,1,1,1)]^2} = \frac{36}{(1+\tau)^2} \ ,$$

here $\underset{\sim}{\mu}_1$, $\underset{\sim}{\mu}_2$ and τ are all independent.

The joint probability mass function of $\underset{\sim}{y}$ is

$$\pi(\underset{\sim}{y} \mid \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2) = \prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \frac{n_{ij}!}{y_{ij1}! y_{ij2}! y_{ij3}! y_{ij4}!} \prod_{k=1}^{4} (p_{ij}\pi_{i1k} + (1-p_{ij})\pi_{i2k})^{y_{ijk}} \right].$$

The summation term will cause difficulties in applying the Gibbs sampler, so we introduce the latent variables $\underset{\sim}{\omega}_{ij} = (\omega_{ij1}, \omega_{ij2}, \omega_{ij3}, (y_{ij4} - \omega_{ij4}))$ denoting those 4-cell counts elicited from the sensitive questions. For example, ω_{ij1} is the number of respondents from the sensitive item (No, No) among y_{ij1}, see Figure 2.1. So the augmented joint posterior density function is

$$\pi(\underset{\sim}{y}, \underset{\sim}{\omega} \mid \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2) = \prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \prod_{k=1}^{4} \binom{y_{ijk}}{\omega_{ijk}} (p_{ij}\pi_{i1k})^{\omega_{ijk}} ((1-p_{ij})\pi_{i2k})^{y_{ijk}-\omega_{ijk}} \right],$$

where $0 \leq \omega_{ijk} \leq y_{ijk}$.

By incorporating the priors from the other two stages, we get the joint posterior density

by Bayes' theorem,

$$
\begin{aligned}
&\pi(\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\omega}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
&\propto \pi(\underset{\sim}{y}, \underset{\sim}{\omega} \mid \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2)\pi(\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2 \mid \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau)\pi(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau) \\
&= \prod_{i=1}^{\ell}\left[\prod_{j=1}^{g_i}\left[\prod_{k=1}^{4}\binom{y_{ijk}}{\omega_{ijk}}(p_{ij}\pi_{i1k})^{\omega_{ijk}}((1-p_{ij})\pi_{i2k}))^{y_{ijk}-\omega_{ijk}}\right]\right] \\
&\times \prod_{i=1}^{\ell}\left(\prod_{s=1}^{2}\left(\frac{\pi_{is1}^{\mu_{s1}\tau-1}\pi_{is2}^{\mu_{s2}\tau-1}\pi_{is3}^{\mu_{s3}\tau-1}\pi_{is4}^{(1-\sum_{k=1}^{3}\mu_{sk})\tau-1}}{D(\mu_{s1}\tau, \mu_{s2}\tau, \mu_{s3}\tau, (1-\sum_{k=1}^{3}\mu_{sk})\tau)}\right)\right) \\
&\times \frac{36}{(1+\tau)^2}.
\end{aligned} \tag{2.1.1}
$$

2.1.2 Blocked Gibbs sampler

We intend to build a blocked Gibbs sampler to get draws of $\underset{\sim}{\pi}_1$ using the following sampling scheme that has two blocks,

$$\pi(\underset{\sim}{\omega} \mid \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{y}) \text{ and } \pi(\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{\omega}, \underset{\sim}{y}).$$

Based on the joint probability density function, we can run the blocked Gibbs sampler from the conditional distribution. The nice thing is that the latent variables have simple distributions, which are independent binomial distributions. Also, $\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2}$ follow Dirichlet distributions.

Step 1. $\underset{\sim}{\omega} \mid \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{y}$

The latent variable ω_{ijk} follow binomial distributions independently

$$\omega_{ijk} \mid \underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2}, \underset{\sim}{y} \overset{ind}{\sim} \text{Binomial}\{y_{ijk}, \frac{p_{ij}\pi_{i1k}}{p_{ij}\pi_{i1k} + (1-p_{ij})\pi_{i2k}}\},$$

$i = 1, \ldots, \ell,\ j = 1, \ldots, g_i, k = 1, 2, 3, 4$. Thus, given data and other parameters, we can draw ω_{ijk} easily from a simple form distribution.

By the multiplication rule,

$$\pi(\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{\omega}, \underset{\sim}{y}) = \pi(\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2 \mid \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{\omega}, \underset{\sim}{y})\pi(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{\omega}, \underset{\sim}{y}).$$

By integrating out $(\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2})$ from $\pi(\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{\omega}, \underset{\sim}{y})$, we can apply grid method based on the conditional joint pdf of $(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau)$ below.

Step 2. $\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{\omega}, \underset{\sim}{y}$

$$\pi(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{\omega}, \underset{\sim}{y}) \propto \prod_{i=1}^{\ell} \left(\frac{D(\omega_{i1\cdot} + \mu_{11}\tau, \omega_{i2\cdot} + \mu_{12}\tau, \omega_{i3\cdot} + \mu_{13}\tau, \omega_{i4\cdot} + (1 - \sum_{k=1}^{3} \mu_{1k})\tau)}{D(\mu_{11}\tau, \mu_{12}\tau, \mu_{13}\tau, (1 - \sum_{k=1}^{3} \mu_{1k})\tau)} \right.$$
$$\left. \times \frac{D(y_{i1\cdot} - \omega_{i1\cdot} + \mu_{21}\tau, y_{i2\cdot} - \omega_{i2\cdot} + \mu_{22}\tau, y_{i3\cdot} - \omega_{i3\cdot} + \mu_{23}\tau, y_{i4\cdot} - \omega_{i4\cdot} + (1 - \sum_{k=1}^{3} \mu_{2k})\tau)}{D(\mu_{21}\tau, \mu_{22}\tau, \mu_{23}\tau, (1 - \sum_{k=1}^{3} \mu_{2k})\tau)} \right)$$
$$\times \frac{36}{(1+\tau)^2}.$$

This is a complex form involving fractions of Dirichlet functions. Thus we could only draw $\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau$ by grid method. Here we actually use a variable substitution letting $\rho = 1/(1+\tau)^2$. So we can draw the new parameter ρ from the (0,1) range, and change it back later to get draws of τ.

Step 3. $\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2 \mid \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{\omega}, \underset{\sim}{y}$

$$\underset{\sim}{\pi}_{i1} \mid \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{\omega}, \underset{\sim}{y} \overset{ind}{\sim} \text{Dirichlet}(\omega_{i\cdot 1} + \mu_{11}\tau, \omega_{i\cdot 2} + \mu_{12}\tau, \omega_{i\cdot 3} + \mu_{13}\tau, \omega_{i\cdot 4} + (1 - \sum_{k=1}^{3} \mu_{1k})\tau),$$

$$\underset{\sim}{\pi}_{i2} \mid \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{\omega}, \underset{\sim}{y} \overset{ind}{\sim} \text{Dirichlet}(y_{i\cdot 1} - \omega_{i\cdot 1} + \mu_{21}\tau, y_{i\cdot 2} - \omega_{i\cdot 2} + \mu_{22}\tau, y_{i\cdot 3} - \omega_{i\cdot 3} + \mu_{23}\tau,$$
$$y_{i\cdot 4} - \omega_{i\cdot 4} + (1 - \sum_{k=1}^{3} \mu_{2k})\tau),$$

where $\omega_{i\cdot k} = \sum_{j=1}^{g_i} \omega_{ijk}, y_{i\cdot k} = \sum_{j=1}^{g_i} y_{ijk},\ k = 1, 2, 3, 4.$
Once we obtain draws from $\underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau$, we go back to the first step to update $\underset{\sim}{\omega}$ and continue with this Gibbs sampling scheme until it converges.

We can obtain Rao-Blackwellized estimators of $\underset{\sim}{\pi}_{i1}$ and $\underset{\sim}{\pi}_{i2}$, which provide smaller mean square error, because the joint distribution conditioned on data $\underset{\sim}{y}$ only can be

expressed as

$$\pi(\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \mid \underset{\sim}{y}) = \int_{\tau} \int_{\underset{\sim}{\mu}_2} \int_{\underset{\sim}{\mu}_1} \sum_{\underset{\sim}{\omega}_{i1}=0}^{\underset{\sim}{y}_{i1}} \cdots \sum_{\underset{\sim}{\omega}_{ig_i}=0}^{\underset{\sim}{y}_{ig_i}} \pi(\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \mid \underset{\sim}{\omega}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{y}) \pi(\underset{\sim}{\omega}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{y})$$

$$= \int_{\tau} \int_{\underset{\sim}{\mu}_2} \int_{\underset{\sim}{\mu}_1} \sum_{\underset{\sim}{\omega}_{i1}=0}^{\underset{\sim}{y}_{i1}} \cdots \sum_{\underset{\sim}{\omega}_{ig_i}=0}^{\underset{\sim}{y}_{ig_i}} \pi(\underset{\sim}{\pi}_{i1} \mid \underset{\sim}{\omega}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{y}) \pi(\underset{\sim}{\pi}_{i2} \mid \underset{\sim}{\omega}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau, \underset{\sim}{y}) \pi(\underset{\sim}{\omega}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \tau \mid \underset{\sim}{y}).$$

Let $\underset{\sim}{\omega}^{(h)} = (\omega_{ijk}^{(h)}), j = 1, \ldots, g_i, k = 1, \ldots, 4, h = 1, \ldots, M$, denote a random sample of size M from the posterior density, $\pi(\underset{\sim}{\omega} \mid \underset{\sim}{y})$, obtained from the Gibbs sampler. So that, the Rao-Blackwellized density estimator of $\pi(\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \mid \underset{\sim}{y})$ is

$$\widehat{\pi(\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \mid \underset{\sim}{y})} = \frac{1}{M} \sum_{h=1}^{M} \pi(\underset{\sim}{\pi}_{i1} \mid \underset{\sim}{\omega}^{(h)}, \underset{\sim}{\mu}_1{}^{(h)}, \underset{\sim}{\mu}_2{}^{(h)}, \tau^{(h)}, \underset{\sim}{y})$$
$$\pi(\underset{\sim}{\pi}_{i2} \mid \underset{\sim}{\omega}^{(h)}, \underset{\sim}{\mu}_1{}^{(h)}, \underset{\sim}{\mu}_2{}^{(h)}, \tau^{(h)}, \underset{\sim}{y}), \quad i = 1, \ldots, \ell.$$

Then we can get the Rao-Blackawellized estimator of the sensitive proportion $\phi_{i11} = \pi_{i13} + \pi_{i14}$ and $\phi_{i12} = \pi_{i12} + \pi_{i14}, i = 1, ..., \ell$ for each area.

Next, we are able to do the prediction in a finite population under simple random sampling. Assume that each sample unit is drawn from a finite population of size N, let X_s denote the total counts of 'yeses' from s^{th} sensitive question. Therefore,

$$X_s \mid \phi_{i1s} \overset{ind}{\sim} \text{Binomial}(N, \phi_{i1s}), s = 1, 2.$$

Then, the finite population proportion $P_s = X_s/N, s = 1, 2$, and inference about the P_s can be made in a straightforward manner under the Bayesian model by using the draws. For generality, we assume a sample of 0.1% from a finite population in the college cheating data.

2.2 Application on College Cheating Data

In this section, we will use this unrelated question design to analyze the college cheating data. Also we provide a comparison of results with those from separate question model and individual area model.

Table 2.1 and Table 2.2 show a comparison of the combined model and the separate question model in posterior means (PM), posterior standard deviations (PSD) and correlations (Cor) between ϕ_{11} and ϕ_{12}. They indicate the combined model has a smaller PSD and CV than the individual area model and the separate question model. For the combined model, we are not surprised finding that the cheating proportion in the final exam are all less then 0.224, since the majority of students will follow WPI's policy of academic honesty. Besides, the proportion of those who are very eager to get high GPA between 3.5 to 4.0 are almost about 0.8, indicating the importance of grades to most but not all of the college students. The separate question model has similar results, whereas the individual model could have aberrant estimates for class sections 3, 4, 11, 13, 15. Comparing the three different methods, the combined model always has a smaller posterior standard deviation, except for the ϕ_{11} in class section 9 with the separate question model giving a little bit smaller PSD of 0.094 than 0.095 from the combined model. Additionally, we can obtain the correlation estimation between cheating and their ambition to get a higher GPA, which seems to be negative consistently across 12 classe sections, with the other three sections have a very small positive correlation not more than 0.082. As expected, the separate question model gives a correlation close to 0, failing to catch any correlation between the cheating proportion and their ambition of the higher GPA. As for the individual area model, the correlation estimations are quite unstable, vibrating between large positive and negative correlations, and even cannot be calculated due to the sparsity of some individual sections.

Table 2.1: Comparison of the Combined model and the Separate question model using posterior means (PM), posterior standard deviations (PSD), posterior coefficient of variations (PCV)

	ϕ_{11}			ϕ_{12}			
	PM	PSD	PCV	PM	PSD	PCV	Cor
a. Combined model							
1	0.172	0.099	0.574	0.835	0.086	0.103	-0.309
2	0.163	0.102	0.624	0.820	0.094	0.115	-0.264
3	0.175	0.104	0.596	0.808	0.086	0.106	0.029
4	0.215	0.122	0.567	0.810	0.095	0.117	0.082
5	0.152	0.095	0.625	0.839	0.086	0.103	-0.289
6	0.161	0.097	0.601	0.791	0.101	0.128	-0.161
7	0.169	0.106	0.624	0.849	0.084	0.099	-0.105
8	0.187	0.108	0.576	0.792	0.096	0.121	-0.266
9	0.157	0.095	0.602	0.777	0.103	0.132	-0.028
10	0.150	0.092	0.614	0.821	0.086	0.105	-0.115
11	0.224	0.122	0.544	0.789	0.101	0.128	-0.058
12	0.193	0.110	0.572	0.757	0.107	0.141	0.019
13	0.202	0.114	0.562	0.828	0.095	0.115	-0.215
14	0.157	0.091	0.583	0.805	0.090	0.111	-0.112
15	0.198	0.106	0.537	0.800	0.085	0.106	-0.102
b. Separate question model							
1	0.139	0.111	0.796	0.849	0.103	0.122	-0.040
2	0.129	0.108	0.838	0.831	0.112	0.135	-0.005
3	0.149	0.120	0.807	0.834	0.106	0.127	0.048
4	0.206	0.155	0.752	0.856	0.098	0.114	0.047
5	0.119	0.104	0.876	0.850	0.105	0.124	0.011
6	0.120	0.104	0.872	0.806	0.126	0.156	0.019
7	0.138	0.112	0.813	0.883	0.082	0.093	0.052
8	0.157	0.123	0.785	0.815	0.115	0.142	0.034
9	0.112	0.094	0.842	0.821	0.118	0.144	-0.014
10	0.109	0.097	0.886	0.834	0.111	0.133	0.035
11	0.237	0.166	0.699	0.799	0.133	0.167	0.026
12	0.155	0.125	0.803	0.786	0.134	0.171	0.015
13	0.190	0.139	0.731	0.843	0.099	0.118	0.026
14	0.126	0.102	0.812	0.825	0.106	0.129	0.017
15	0.172	0.127	0.738	0.809	0.108	0.134	-0.044

Table 2.2: Comparison of the Combined model and the Individual area model using posterior means (PM), posterior standard deviations (PSD), posterior coefficient of variations (PCV)

	ϕ_{11}			ϕ_{12}			
	PM	PSD	PCV	PM	PSD	PCV	Cor
a. Combined model							
1	0.172	0.099	0.574	0.835	0.086	0.103	-0.309
2	0.163	0.102	0.624	0.820	0.094	0.115	-0.264
3	0.175	0.104	0.596	0.808	0.086	0.106	0.029
4	0.215	0.122	0.567	0.810	0.095	0.117	0.082
5	0.152	0.095	0.625	0.839	0.086	0.103	-0.289
6	0.161	0.097	0.601	0.791	0.101	0.128	-0.161
7	0.169	0.106	0.624	0.849	0.084	0.099	-0.105
8	0.187	0.108	0.576	0.792	0.096	0.121	-0.266
9	0.157	0.095	0.602	0.777	0.103	0.132	-0.028
10	0.150	0.092	0.614	0.821	0.086	0.105	-0.115
11	0.224	0.122	0.544	0.789	0.101	0.128	-0.058
12	0.193	0.110	0.572	0.757	0.107	0.141	0.019
13	0.202	0.114	0.562	0.828	0.095	0.115	-0.215
14	0.157	0.091	0.583	0.805	0.090	0.111	-0.112
15	0.198	0.106	0.537	0.800	0.085	0.106	-0.102
c. Individual are model							
1	0.240	0.178	0.741	0.853	0.144	0.169	-0.58
2	0.227	0.228	1.004	0.807	0.207	0.257	-0.619
3*	0.452	0.217	0.480	0.772	0.155	0.201	0.208
4*	0.633	0.238	0.376	0.873	0.140	0.160	0.324
5	0.097	0.202	2.083	0.903	0.202	0.223	0.999
6	0.201	0.198	0.982	0.712	0.229	0.321	-0.368
7*	0.216	0.226	1.045	1.000	0.000	0.000	NaN
8	0.362	0.214	0.592	0.715	0.194	0.271	-0.42
9	0.176	0.180	1.023	0.773	0.199	0.257	0.114
10	0.165	0.163	0.988	0.827	0.143	0.173	-0.286
11*	0.779	0.217	0.278	0.700	0.242	0.345	0.138
12	0.396	0.205	0.517	0.669	0.202	0.303	0.067
13*	0.505	0.276	0.546	0.827	0.151	0.182	-0.318
14	0.222	0.152	0.683	0.774	0.135	0.174	-0.139
15*	0.531	0.184	0.346	0.723	0.137	0.189	0.048

$*$ Aberrant areas, note the computational instability in class section 7.

Figure 2.2 gives the comparison 95% HPD intervals of ϕ_{11}.

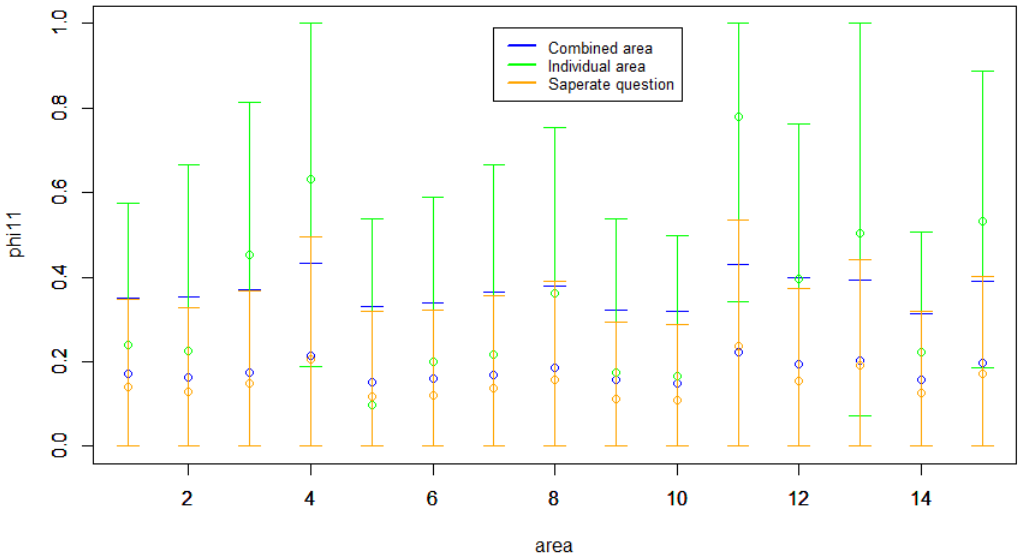


Figure 2.2: 95% HPD interval of ϕ_{11}

Figure 2.3 provides a direct comparison of the coefficient of variation (CV) for ϕ_{11}.

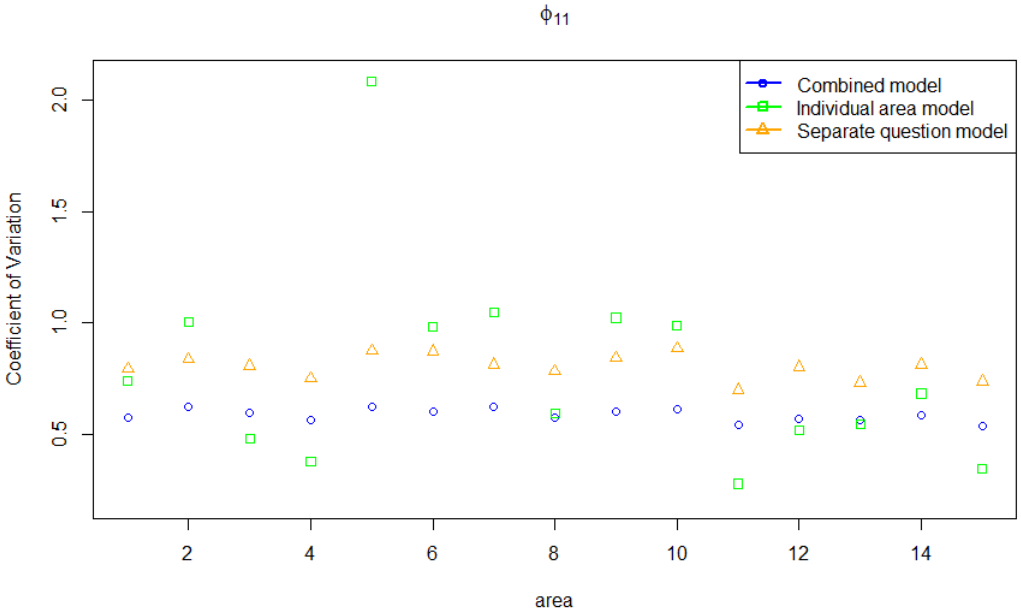


Figure 2.3: Coefficient of Variation (CV) of ϕ_{11}

Figure 2.4 gives the 95% HPD interval of ϕ_{12}.

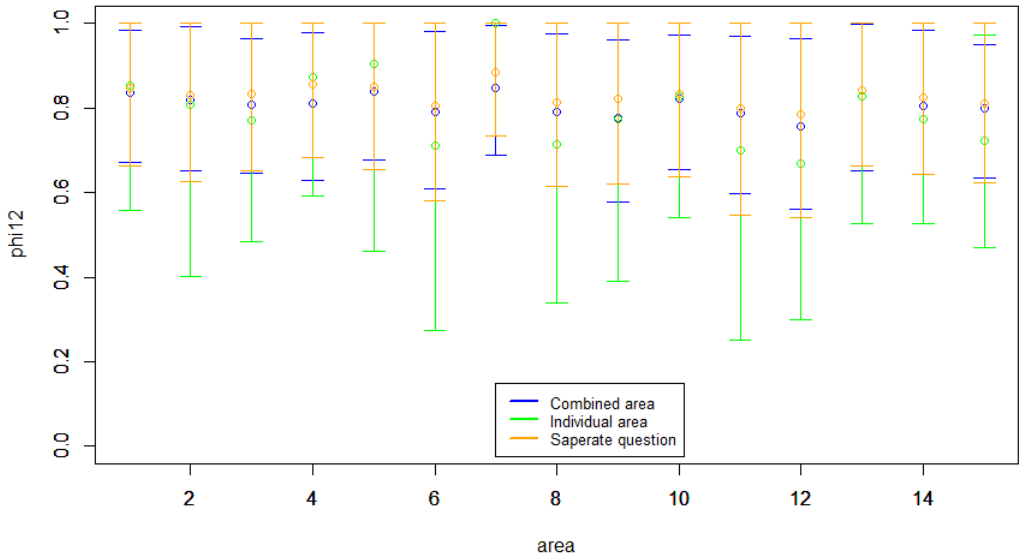


Figure 2.4: 95% HPD interval of ϕ_{12}

Figure 2.5 provides a direct comparison of the coefficient of variation for ϕ_{12}. The combined model also gives a smaller CV for both π_1 and π_2.

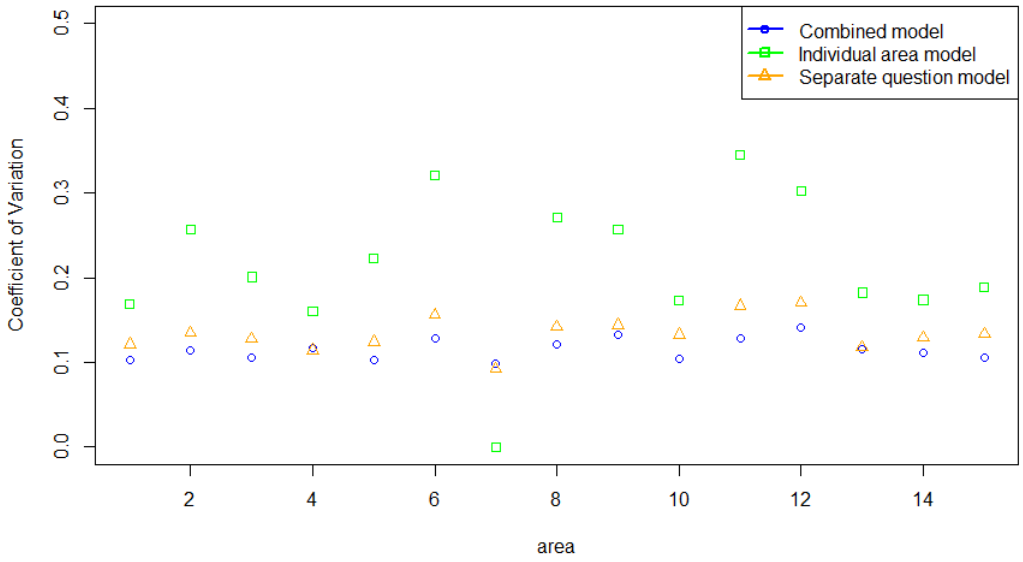


Figure 2.5: Coefficient of Variation (CV) of ϕ_{12}

2.3 Simulation study

In Section 2.3.1, we preform a simulation study to assess the performance of the combined-area model compared with individual-area model and separate question model. In Section 2.3.2, we adjust the parameter setting to increase the number of areas and the correlation within the sensitive and non-sensitive questions to see the possible gains in estimation accuracy.

2.3.1 Comparison of three models

In this section, we are going to test our combined model using the simulated data. Based on the 1000 simulated runs, we provide a comparison of our combined model with the separate question model and the individual area model. The combined model is discussed in Section 2.2; next we list the individual area model and separate question model for multi-item RRT here.

Individual Area Model for Multi-item Case

Because the areas are modelled individually, only a flat Dirichlet prior is used for $\underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2}$,

$$\underset{\sim}{y}_{ij} \mid \underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \overset{ind}{\sim} \text{Multinomial}\{n_{ij}, \underset{\sim}{a}_{ij}\}, i = 1, \ldots, \ell; j = 1, \ldots, g_i \geq 2,$$

where $\underset{\sim}{a}_{ij} = (p_{ij}\pi_{i11} + (1-p_{ij})\pi_{i21}, p_{ij}\pi_{i12} + (1-p_{ij})\pi_{i22}, p_{ij}\pi_{i13} + (1-p_{ij})\pi_{i23}, p_{ij}\pi_{i14} + (1-p_{ij})\pi_{i24})$, denoting the four cell probabilities for each group with size $n_{ij} = \sum_{k=1}^{4} y_{ijk}$.

$$\underset{\sim}{\pi}_{i1} \overset{ind}{\sim} \text{Dirichlet}(1,1,1,1) \quad \underset{\sim}{\pi}_{i2} \overset{ind}{\sim} \text{Dirichlet}(1,1,1,1), i = 1, \ldots, \ell.$$

Separate Question Model

The separate question model is the SAM we introduced in Section 1.2 being applied to

each single sensitive question,

$$y_{ij} \mid \pi_{i1}, \pi_{i2} \overset{ind}{\sim} \text{Binomial}\{n_{ij}, p_{ij}\pi_{i1} + (1 - p_{ij})\pi_{i2}\},$$

$i = 1, \ldots, \ell,\ j = 1, \ldots, g_i \geq 2,$

$$\pi_{i1}|\mu_1, \tau \overset{ind}{\sim} \text{Beta}(\mu_1\tau, (1 - \mu_1)\tau) \quad \pi_{i2}|\mu_2, \tau \overset{ind}{\sim} \text{Beta}(\mu_2\tau, (1 - \mu_2)\tau),$$

$i = 1, \ldots, \ell,$

$$\pi(\mu_1, \mu_2, \tau) = \frac{1}{(1 + \tau)^2}.$$

Assuming that the probability of answering 'yes' to the first and second sensitive questions are ϕ_{11} and ϕ_{12}; the probability of answering 'No' to the first and second nonsensitive questions are ϕ_{21} and ϕ_{22}. Then we simulated the correlated response data with correlation ρ_1= 0.5 and ρ_2= 0.5 with respect to the sensitive and nonsensitive questions correspondingly among 10 areas. In other words, with true value set as $\phi_{11} = 0.5, \phi_{12} = 0.5$, the four types of response (No, No), (No, Yes),(Yes, No),(Yes, Yes) for the sensitive questions are generated with the probability $\underset{\sim}{\pi}_{i1}$. After we construct $\underset{\sim}{\mu}_1 = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14})$ formulated as

$$\mu_{11} = (1-\phi_{11})(1-\phi_{12})+\gamma,\ \mu_{12} = (1-\phi_{11})\phi_{12}-\gamma,\ \mu_{13} = \phi_{11}(1-\phi_{12})-\gamma,\ \mu_{14} = \phi_{11}\phi_{12}+\gamma,$$

where $\gamma = \rho_1\sqrt{\phi_{11}(1 - \phi_{11})\phi_{12}(1 - \phi_{12})}$. Because we have selected ϕ_{11} and ϕ_{12} equal, we $0 < \rho < 1$; see Yu, Bhadra and Nandram, (2017). We can get $\underset{\sim}{\pi}_{i1} \overset{ind}{\sim} \text{Dirichlet}(\underset{\sim}{\mu}_1\tau)$ where the above equations give $\underset{\sim}{\mu}_1$ when we substitute $\phi_{11} = 0.5, \phi_{12} = 0.5$ and $\tau = 100$.

Again we can generate the response from correlated nonsensitive questions in the same way for $\underset{\sim}{\pi}_{i2}$, with the true value set as $\phi_{21} = 0.4, \phi_{22} = 0.4$.

Now we want to simulate the sampling process as follows. For each individual from i^{th} area and j^{th} group ($i = 1, .., \ell = 10, j = 1, .., g_i \geq 2$). At first, we generate the number of groups uniformly from 2 to 5. For each individual coming from i^{th} area and j^{th} group

with size n_{g_i}= (20, 25, 30, 35, 40), we choose a random mechanism with probability p_{ij}= (.25, .75, .2, .7, .3) to answer the sensitive questions and (1-p_{ij}) to answer the nonsensitive questions. Following the simulation process, we are able to collect the combined binary response data from both sensitive and nonsensitive questions, without knowing which exact question the respondent answer.

To find the finite population estimation of the probability of answering 'yes' to the sensitive questions, we can fit the three-stage Bayesian models to get the estimates of individual $\underset{\sim}{\pi}_{i1}$'s first, and then get the corresponding finite population estimation afterwards based on the probability relationship $\phi_{i11} = \pi_{i13} + \pi_{i14}$ and $\phi_{i12} = \pi_{i12} + \pi_{i14}$. We calculated the relative absolute bias, $RAB = (PM - T)/T$, and the posterior root mean squared error, $PRMSE = \sqrt{(PM - T)^2 + PSD^2}$, where T denotes the true proportions, ϕ_{11} or ϕ_{12} (known by simulation). To compare the combined model with the individual area model and separate question model, we present a 95% boxplot of the RAB and PRMSE of the 1000 simulations from 10 areas.

In Figure 2.6, we can observe that the combined model always has a smaller relative absolute bias than the separate question model and individual area model. In Figure 2.7, the combined model still outperforms the other two in the sense of the posterior mean square error across all the areas.

Figure 2.6: Boxplot of RAB for combined model, separate question model and individual area model of 10 areas for 1000 simulations under the combined model

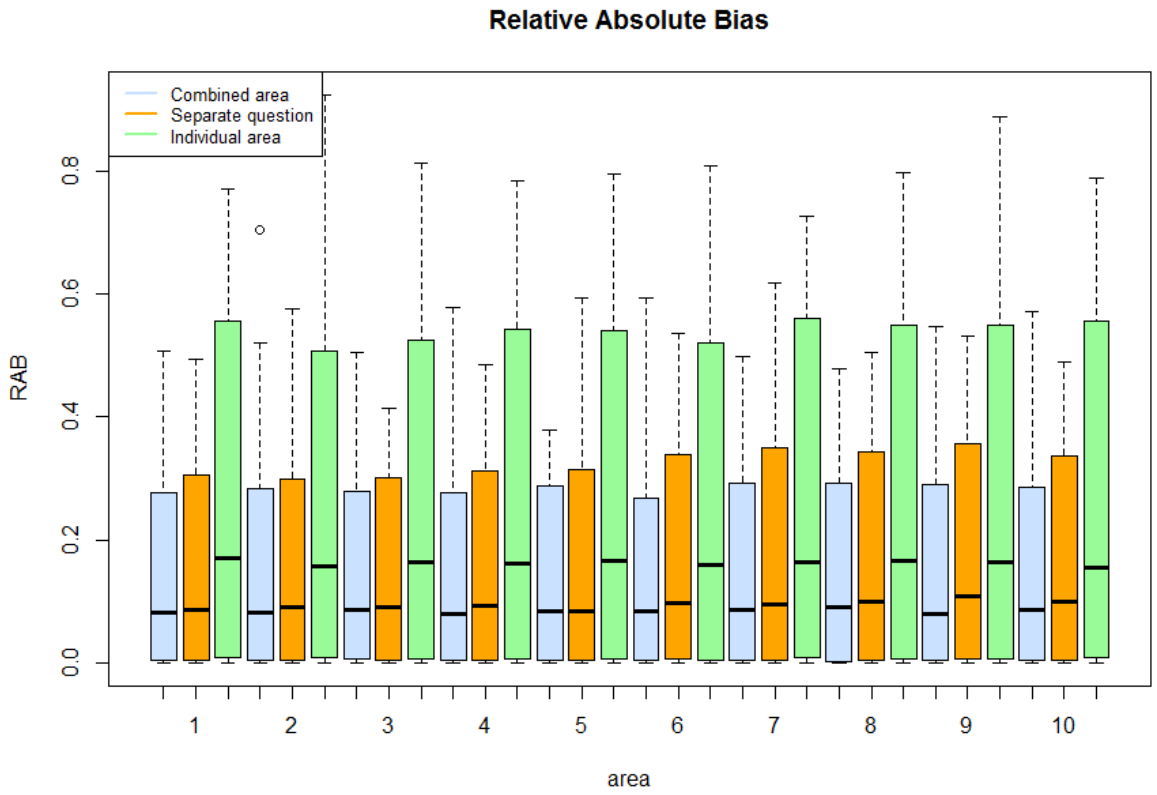
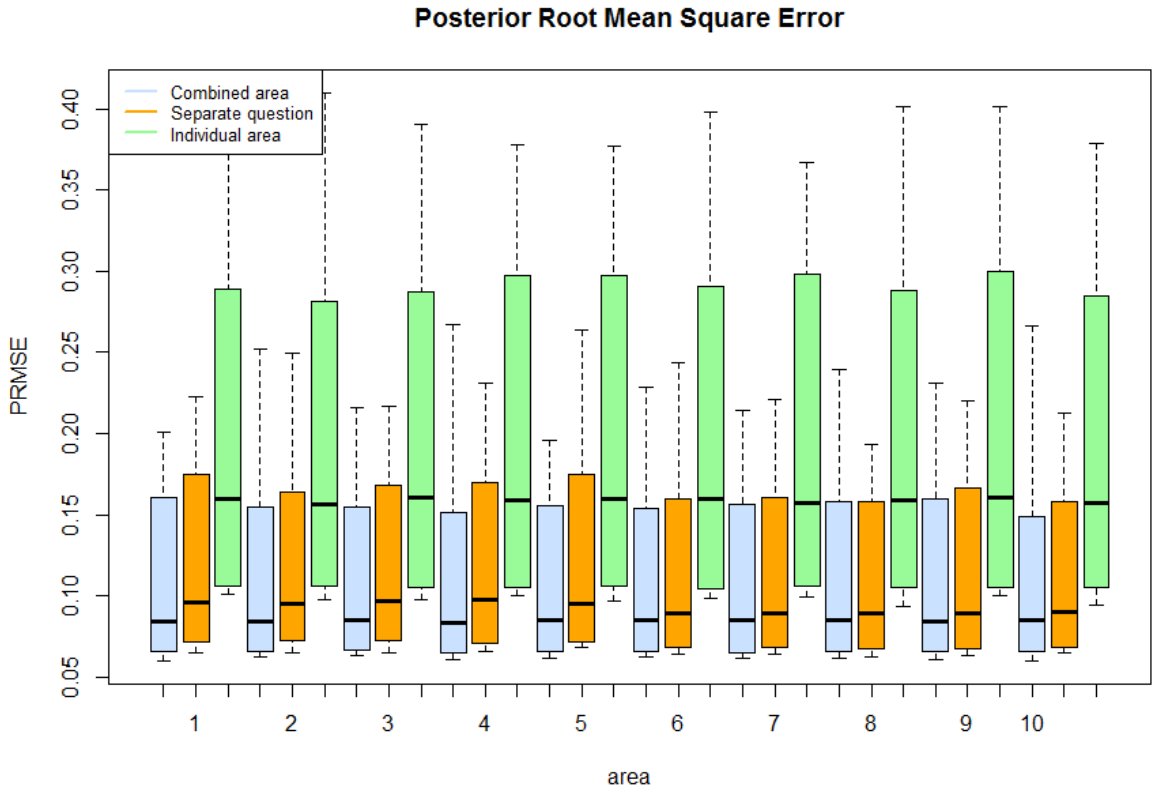


Figure 2.7: Boxplot of PRMSE for combined model, separate question model and individual area model of 10 areas for 1000 simulations under the combined model



2.3.2 Discussion of the effect of number of locations and correlation

In order to test the model with more areas and compare the effect of the number of locations, besides RAB and PRMSE, we also computed their average width (Wid) of the 95% HPD intervals and the coverage (C), which is the proportion of intervals containing the true value in the 1000 simulated runs. We simulated the correlated data with correlation $\rho = 0.5$ with respect to the sensitive and nonsensitive questions correspondingly among 10 areas.

In Table 3.5, we provided the simulation results from different area sizes ($\ell = 10, 25$). We observed that for both the combined Bayesian model (cb) and the separate question model (Sep), the average ralative absolute bias and the root mean squared error get smaller as the number of areas increase from 10 to 25 for both ϕ_{11} and ϕ_{12}; the average width of 95% HPD interval for ϕ_{11} is 0.21, shorter than 0.28. Even though we obtained a smaller coverage of 0.948 when the number of areas equal to 25, it is still very close to the expected 95%, given a shorter 95% HPD interval. However, the 95% HPD interval for the individual area model is much wider. Similar conclusions can be drawn for ϕ_{12}. In the case of $\ell = 10$, the FORTRAN codes running at 30 computers in parallel will take about 12 hours to finish 1000 simulations for combined model and about 6 hours for the separate question model. The individual area model will finish in 2.9 seconds. The computing time for 25 areas is 1.5 times longer.

Table 2.3: Relative absolute bias, posterior root mean squared error, coverage of 95% credible intervals and width of 95% credible interval averaged over the 1000 runs and different area sizes (ℓ=10, 25) for combined model (cb), separate question model (sep) and individual area model (ind) .

		$\hat{\phi}_{11}$				$\hat{\phi}_{12}$			
ℓ	*Model*	RAB	PRMSE	C	Wid	RAB	PRMSE	C	Wid
10	cb	0.099	0.092	0.976	0.280	0.100	0.092	0.975	0.280
	Sep	0.106	0.103	0.983	0.322	0.106	0.103	0.983	0.321
	Ind	0.192	0.170	0.945	0.494	0.193	0.170	0.945	0.495
25	cb	0.086	0.073	0.948	0.210	0.087	0.073	0.944	0.210
	Sep	0.087	0.080	0.974	0.244	0.088	0.080	0.971	0.245
	Ind	0.194	0.171	0.942	0.493	0.187	0.164	0.944	0.493

In Table 2.4, we showed the comparison results of different correlations when the number of areas is fixed to be 10. First, for the estimation of π_1 from the combined model (cb), we observed that the relative absolute bias (RAB) is 0.097 for the highly correlated data ($\rho = 0.9$), which is pretty close to 0.102 of the independent data ($\rho = 0$). The PRMSE is .087, which is also just a little bit smaller than 0.093 for the less correlated one. The changes according to the correlation under the combined model can also be seen in Figure 2.5 and Figure 2.6, where we show the boxplot of RAB and PRMSE of the combined model.

Table 2.4: Relative absolute bias, posterior root mean squared error, coverage of 95% credible intervals and width of 95% credible interval averaged over the 1000 runs and 10 areas under different levels of correlations $(\rho_1, \rho_2) = (0, 0),\ (0.5, 0.5),\ (0.9, 0.9)$.

		$\hat{\phi}_{11}$				$\hat{\phi}_{12}$			
ρ_1, ρ_2	*Model*	RAB	PRMSE	C	Wid	RAB	PRMSE	C	Wid
0, 0	cb	0.102	0.093	0.972	0.283	0.103	0.094	0.973	0.283
(0.153, 0.094)	Sep	0.109	0.104	0.981	0.322	0.110	0.104	0.980	0.322
	Ind	0.188	0.168	0.948	0.490	0.188	0.168	0.951	0.490
0.5, 0.5	cb	0.099	0.092	0.976	0.280	0.100	0.092	0.975	0.280
(0.525, 0.556)	Sep	0.106	0.103	0.983	0.322	0.106	0.103	0.983	0.321
	Ind	0.192	0.170	0.945	0.494	0.193	0.170	0.945	0.495
0.9, 0.9	cb	0.097	0.087	0.967	0.259	0.097	0.087	0.967	0.259
(0.948,0.892)	Sep	0.104	0.101	0.982	0.314	0.104	0.101	0.982	0.314
	Ind	0.203	0.174	0.934	0.496	0.204	0.157	0.934	0.496

NOTE: $(\hat{\rho}_1, \hat{\rho}_2) = (0.153, 0.094),\ (0.525, 0.556),\ (0.948, 0.892)$ are the actual correlations of the simulated data.

In Table 2.5, we find the correlation effect on bias, posterior mean, posterior standard

deviation and coefficient of variation also change little as the correlation varies for the combined model.

Table 2.5: B, PM, PSD, CV averaged over the 1000 runs and 10 areas under different levels of correlations $(\rho_1, \rho_2) = (0, 0),\ (0.5, 0.5),\ (0.9, 0.9)$.

		$\hat{\phi}_{11}$				$\hat{\phi}_{12}$			
ρ_1, ρ_2	*Model*	B	PM	PSD	CV	B	PM	PSD	CV
0, 0	cb	0.002	0.503	0.073	0.147	0.001	0.502	0.073	0.148
(0.153, 0.094)	sep	0.003	0.503	0.083	0.168	0.001	0.502	0.083	0.168
	ind	-0.007	0.494	0.128	0.279	-0.009	0.492	0.128	0.280
0.5, 0.5	cb	0.000	0.501	0.072	0.146	0.000	0.500	0.072	0.146
(0.525, 0.556)	sep	0.001	0.501	0.083	0.168	0.000	0.501	0.083	0.168
	ind	-0.003	0.497	0.129	0.281	-0.004	0.496	0.130	0.281
0.9, 0.9	cb	0.001	0.501	0.067	0.135	0.001	0.502	0.067	0.135
(0.948,0.892)	sep	0.001	0.502	0.081	0.164	0.001	0.502	0.081	0.164
	ind	-0.002	0.499	0.130	0.284	-0.001	0.499	0.130	0.284

Figure 2.8: Boxplot of RAB for 1000 simulations per area of different correlations under the combined model

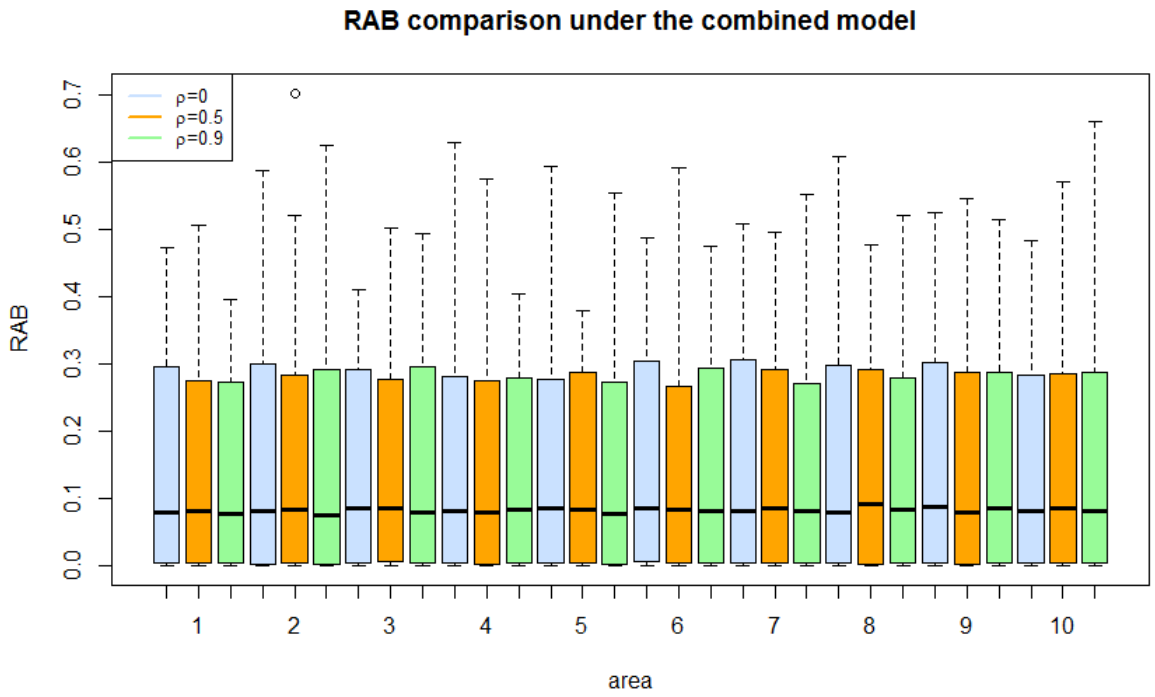
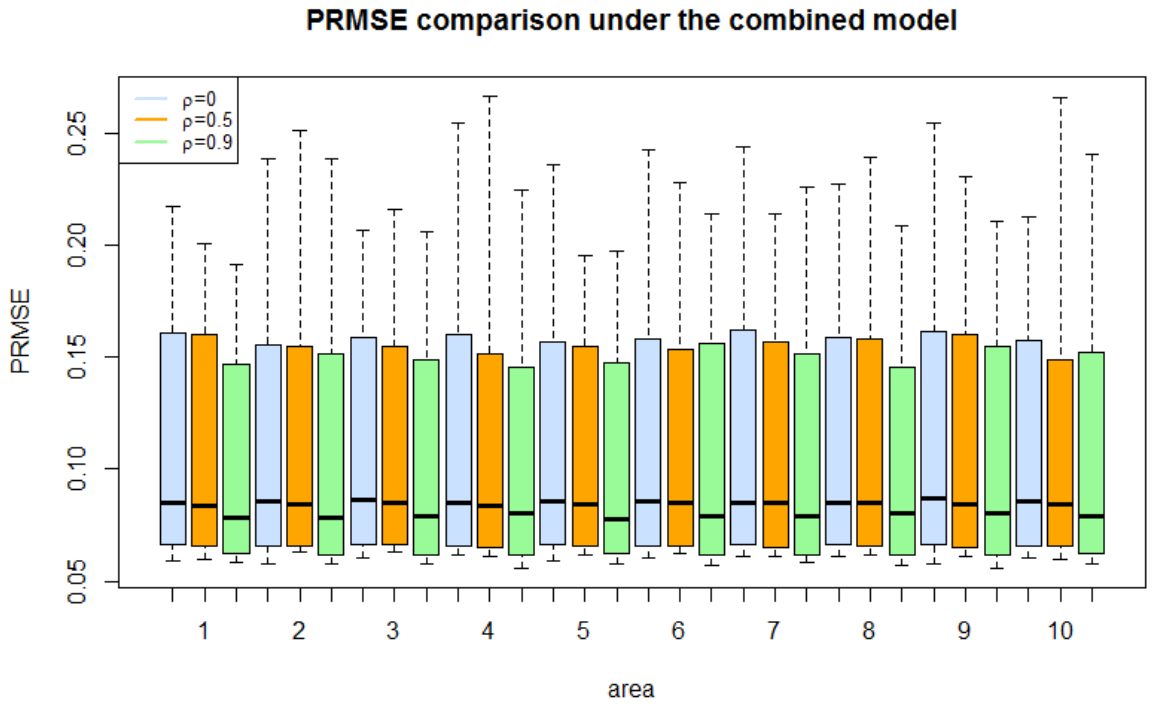


Figure 2.9: Boxplot of PRMSE for 1000 simulations per area of different correlations under the combined model



Since it is not intuitive that there are just small changes as the correlation increases (see Table 2.5, Table 2.6 Figure 2.8, Figure 2.9), to further study the correlation effect on the proportion estimation, we provide a study on the four-cell probability. Table 2.6 gives the bias (B) and posterior mean (PM) estimation of the four-cell probability of π_1 and π_2 generated under three different correlation levels (0,0.5, 0.75). We use correlation 0.75 instead of 0.9 to avoid the extreme situation that there is no counts of the off-diagonal cell (No&Yes, Yes&No), seeing that 0.75 is a comparative larger correlation which already makes the off-diagonal cell probability lower than 0.08.

Table 2.6: B and PM comparison of π_1 and π_2 under different correlations

Model	ρ	$\hat{\pi}_1$				$\hat{\pi}_2$			
B									
cb	0	0.000	-0.001	0.001	0.000	-0.002	0.000	0.000	0.002
	0.5	-0.003	0.002	0.002	-0.002	-0.005	0.004	0.003	-0.001
	0.75	-0.010	0.010	0.010	-0.010	-0.013	0.011	0.011	-0.009
ind	0	0.006	-0.001	0.001	-0.007	-0.103	0.010	0.010	0.083
	0.5	-0.004	0.005	0.005	-0.006	-0.109	0.010	0.010	0.089
	0.75	-0.015	0.015	0.015	-0.016	-0.118	0.018	0.018	0.082
PM									
cb	0	0.251	0.249	0.250	0.250	0.359	0.240	0.240	0.162
	0.5	0.372	0.127	0.127	0.374	0.475	0.124	0.122	0.279
	0.75	0.428	0.072	0.072	0.428	0.528	0.071	0.071	0.330
ind	0	0.257	0.250	0.250	0.243	0.257	0.250	0.250	0.243
	0.5	0.371	0.130	0.130	0.370	0.371	0.130	0.130	0.370
	0.75	0.422	0.078	0.078	0.422	0.422	0.078	0.078	0.422

Table 2.7 and Table 2.8 give the comparison for posterior standard deviation (PSD), posterior critical value (PCV), relative absolute bias (RAB) and posterior root mean square error (PRMSE). As the correlation increases, there tends to be more counts in the diagonal of the contingency table, which have the counts for (No, No) and (Yes, Yes). While for the off-diagonal, the counts of (No, Yes) and (Yes, No) are getting smaller. The changes of the contingency table result in a smaller bias on the diagonal cell. The PSD for the off-diagonal cell is decreasing because it rarely has counts there; the PSD for the diagonal shows an increasing trend first and then it decreases as the correlation change from 0.5 to 0.75. The changes of the PRMSE are consistent with PSD since the bias changes a little. In comparison, we also provide the estimation results for individual model.

The second step of the block Gibbs sampler in drawing $\underset{\sim}{\mu_1}, \underset{\sim}{\mu_2}, \tau$ mainly takes time since the conditional posterior distribution is a complicated function when building the grids. Moreover, the Dirichlet probability parameter $\underset{\sim}{\mu_1}$ and $\underset{\sim}{\mu_2}$ have self constraints which keep changing the range for the grid method, slowing down the convergence. Another concern about the Multinomial Dirichlet model is that the Dirichlet distribution only models negative correlated probabilities which is not very flexible. We address these issues in Chapter 3.

Table 2.7: PSD and PCV comparison of π_1 and π_2 under different correlations

Model	ρ	$\hat{\pi}_1$				$\hat{\pi}_2$			
PSD									
cb	0	0.062	0.061	0.061	0.060	0.062	0.056	0.056	0.049
	0.5	0.068	0.046	0.046	0.067	0.064	0.042	0.042	0.059
	0.75	0.067	0.035	0.035	0.066	0.062	0.032	0.032	0.059
ind	0	0.113	0.108	0.108	0.102	0.113	0.108	0.108	0.102
	0.5	0.124	0.078	0.078	0.121	0.124	0.078	0.078	0.121
	0.75	0.128	0.059	0.058	0.125	0.128	0.059	0.058	0.125
CV									
cb	0	0.256	0.251	0.251	0.245	0.175	0.237	0.237	0.315
	0.5	0.187	0.382	0.380	0.182	0.137	0.355	0.358	0.214
	0.75	0.159	0.503	0.506	0.157	0.119	0.466	0.465	0.181
ind	0	0.515	0.507	0.508	0.495	0.515	0.507	0.508	0.495
	0.5	0.379	0.731	0.729	0.370	0.379	0.731	0.729	0.370
	0.75	0.331	0.899	0.902	0.344	0.331	0.899	0.902	0.344

Table 2.8: RAB and PRMSE comparison of π_1 and π_2 under different correlations

Model	ρ	$\hat{\pi}_1$				$\hat{\pi}_2$			
RAB									
cb	0	0.176	0.174	0.177	0.169	0.110	0.148	0.150	0.198
	0.5	0.129	0.278	0.273	0.125	0.085	0.228	0.230	0.132
	0.75	0.109	0.497	0.499	0.109	0.075	0.389	0.395	0.113
ind	0	0.335	0.326	0.327	0.307	0.359	0.382	0.382	0.700
	0.5	0.252	0.432	0.423	0.248	0.287	0.504	0.507	0.464
	0.75	0.249	0.585	0.591	0.249	0.281	0.642	0.647	0.402
PRMSE									
cb	0	0.080	0.078	0.078	0.076	0.077	0.069	0.069	0.061
	0.5	0.088	0.059	0.059	0.086	0.080	0.052	0.052	0.073
	0.75	0.087	0.044	0.044	0.086	0.078	0.039	0.039	0.074
ind	0	0.148	0.142	0.142	0.135	0.184	0.149	0.149	0.156
	0.5	0.165	0.099	0.098	0.161	0.199	0.103	0.103	0.185
	0.75	0.188	0.076	0.076	0.187	0.222	0.078	0.078	0.205

Chapter 3

Generalized Mixed Effects Model and Approximation

In this chapter, we build a generalized mixed effects model in which the parameters take values on the real line. Recall the underlying distribution for the counts is

$$\underset{\sim}{y}_{ij} \mid \underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2} \overset{ind}{\sim} \text{Multinomial}\{n_{ij}, \underset{\sim}{a}_{ij}\},\ i = 1, \ldots, \ell,\ j = 1, \ldots, g_i \geq 2,$$

where

$$\underset{\sim}{a}_{ij} = (p_{ij}\pi_{i11}+(1-p_{ij})\pi_{i21}, p_{ij}\pi_{i12}+(1-p_{ij})\pi_{i22}, p_{ij}\pi_{i13}+(1-p_{ij})\pi_{i23}, p_{ij}\pi_{i14}+(1-p_{ij})\pi_{i24})$$

and

$$\pi_{i1k} = \frac{\exp(\theta_{1k} + \nu_{1i})}{1 + \sum_{t=1}^{3} \exp(\theta_{1t} + \nu_{1i})}, k = 1, 2, 3;\ \pi_{i14} = \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{1t} + \nu_{1i})},$$
$$\pi_{i2k} = \frac{\exp(\theta_{2k} + \nu_{2i})}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}, k = 1, 2, 3;\ \pi_{i24} = \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})},$$
$$i = 1, \ldots, \ell.$$

The reparameterization actually permits a one-to-one map through $\log(\frac{\pi_{i1k}}{1-\pi_{i1k}}) = \theta_{1k} + \nu_{1i}, k = 1, 2, 3$, where we see that the logit of π_{i1k} is determined by the cell effect θ_{1k} and the area effect ν_{1i} together. Consequently, we can modify the combined model to a generalized mixed effects model which allow us to sample the parameters from an approximate normal distribution.

The generalized mixed effects model will allow for an integrated nested normal approximation (INNA) and improve the computation efficiency in two folds. First, they form a fast but approximate computing approach. Second, we use the approximations to feed into the exact method. In the end, we also apply it to the college cheating data to compare the two models.

3.1 Generalized Mixed Effects Model

The model after the reparameterization is

$$\underset{\sim}{y}_{ij} \mid \underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \nu_{1i}, \nu_{2i} \overset{ind}{\sim} \text{Multinomial}\{n_{ij}, \underset{\sim}{a}_{ij}(\underset{\sim}{\theta}, \underset{\sim}{\nu})\}, i = 1, \ldots, \ell,\ j = 1, \ldots, g_i,$$

with $$\underset{\sim}{a}_{ij}(\underset{\sim}{\theta}, \underset{\sim}{\nu}) = (p_{ij}\pi_{i11} + (1-p_{ij})\pi_{i21}, p_{ij}\pi_{i12} + (1-p_{ij})\pi_{i22},$$
$$p_{ij}\pi_{i13} + (1-p_{ij})\pi_{i23}, p_{ij}\pi_{i14} + (1-p_{ij})\pi_{i24}),$$

where

$$\pi_{i1k} = \frac{\exp(\theta_{1k} + \nu_{1i})}{1 + \sum_{t=1}^{3} \exp(\theta_{1t} + \nu_{1i})}, k = 1, 2, 3;\ \pi_{i14} = \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{1t} + \nu_{1i})},$$
$$\pi_{i2k} = \frac{\exp(\theta_{2k} + \nu_{2i})}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}, k = 1, 2, 3;\ \pi_{i24} = \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})},$$
$$i = 1, \ldots, \ell. \tag{3.1.1}$$

This parameter setting allows for $\sum_{k=1}^{4} \pi_{i1k} = 1$, with the cell probability $\underset{\sim}{\pi}_{i1}$ is only determined by a main cell effect $\underset{\sim}{\theta}_1 = \{\theta_{1k}\}$ $(k = 1, 2, 3)$ and an area effect $\underset{\sim}{\nu}_1 = \{\nu_{1i}\}$ $(i = 1, \ldots, \ell)$. Notice that θ_{1k} and ν_{1i} can take any values under the scheme. Similarly, $\underset{\sim}{\theta}_2$

and $\underset{\sim}{\nu}_2$ are the re-parameterized parameters for $\underset{\sim}{\pi}_{i2}$. Here we consider the situation of flat priors on $\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2$ momentarily; later we will put informative prior in them. Let

$$\underset{\sim}{\tau}_1 = \begin{pmatrix} \underset{\sim}{\nu}_1 \\ \underset{\sim}{\theta}_1 \end{pmatrix}, \quad \underset{\sim}{\tau}_2 = \begin{pmatrix} \underset{\sim}{\nu}_2 \\ \underset{\sim}{\theta}_2 \end{pmatrix},$$

then, conditional on $\underset{\sim}{\omega}$, we have the following joint posterior density of $\underset{\sim}{\tau}_1, \underset{\sim}{\tau}_2$,

$$\begin{aligned}
&\pi(\underset{\sim}{\tau}_1, \underset{\sim}{\tau}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) = \\
&\pi(\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2, \underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \propto \\
&\prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \prod_{k=1}^{3} \Big(\frac{p_{ij} \exp(\theta_{1k} + \nu_{1i})}{1 + \sum_{s=1}^{3} \exp(\theta_{1s} + \nu_{1i})} \Big)^{\omega_{ijk}} \Big(\frac{p_{ij}}{1 + \sum_{s=1}^{3} \exp(\theta_{1s} + \nu_{1i})} \Big)^{\omega_{ij4}} \right] \\
&\prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \prod_{k=1}^{3} \Big(\frac{(1 - p_{ij}) \exp(\theta_{2k} + \nu_{2i})}{1 + \sum_{s=1}^{3} \exp(\theta_{2s} + \nu_{2i})} \Big)^{y_{ijk} - \omega_{ijk}} \Big(\frac{1 - p_{ij}}{1 + \sum_{s=1}^{3} \exp(\theta_{2s} + \nu_{2i})} \Big)^{y_{ij4} - \omega_{ij4}} \right] \\
&= h_1(\underset{\sim}{\tau}_1) h_2(\underset{\sim}{\tau}_2),
\end{aligned}$$

where

$$h_1(\underset{\sim}{\tau}_1) = \prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \prod_{k=1}^{3} \Big(\frac{p_{ij} \exp(\theta_{1k} + \nu_{1i})}{1 + \sum_{s=1}^{3} \exp(\theta_{1s} + \nu_{1i})} \Big)^{\omega_{ijk}} \Big(\frac{p_{ij}}{1 + \sum_{s=1}^{3} \exp(\theta_{1s} + \nu_{1i})} \Big)^{\omega_{ij4}} \right]$$

and

$$h_2(\underset{\sim}{\tau}_2) = \prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \prod_{k=1}^{3} \Big(\frac{(1 - p_{ij}) \exp(\theta_{2k} + \nu_{2i})}{1 + \sum_{s=1}^{3} \exp(\theta_{2s} + \nu_{2i})} \Big)^{y_{ijk} - \omega_{ijk}} \Big(\frac{1 - p_{ij}}{1 + \sum_{s=1}^{3} \exp(\theta_{2s} + \nu_{2i})} \Big)^{y_{ij4} - \omega_{ij4}} \right]$$

We can separate out this likelihood by $h_1(\underset{\sim}{\tau}_1) h_2(\underset{\sim}{\tau}_2)$ since given $\underset{\sim}{y}, \underset{\sim}{\omega}$, the parameter set $\underset{\sim}{\theta}_1, \underset{\sim}{\nu}_1$ and $\underset{\sim}{\theta}_2, \underset{\sim}{\nu}_2$ are independent.

We need to approximate this likelihood to a multivariate normal density, then putting conjugate priors to $\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2, \underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2$, so that we have the simple forms for posterior densities for $\underset{\sim}{\nu}$ and $\underset{\sim}{\theta}$. We exemplify this procedure by approximating $h_1(\underset{\sim}{\tau}_1)$ and $h_2(\underset{\sim}{\tau}_2)$ separately by considering a generic function $h(\underset{\sim}{\tau})$.

3.2 Approximation

Lemma Let $h(\underset{\sim}{\tau})$ be a unimodal density function with a vector parameter $\underset{\sim}{\tau}$. Then approximately $\underset{\sim}{\tau}$ has a multivariate normal distribution

$$\underset{\sim}{\tau} \sim \text{Normal}(\underset{\sim}{\tau}^* - H^{-1}\underset{\sim}{g}, -H^{-1}),$$

where $\underset{\sim}{g}$ is the gradient vector of $f(\underset{\sim}{\tau}) = \log h(\underset{\sim}{\tau})$ evaluated at some point $\underset{\sim}{\tau}^*$ near the mode and H is the Hessian matrix evaluated at $\underset{\sim}{\tau}^*$. Since the multivariate Taylor expansion of $f(\underset{\sim}{\tau})$ at $\underset{\sim}{\tau}^*$ is

$$f(\underset{\sim}{\tau}) \approx f(\underset{\sim}{\tau}^*) + (\underset{\sim}{\tau} - \underset{\sim}{\tau}^*)'\underset{\sim}{g} + \tfrac{1}{2}(\underset{\sim}{\tau} - \underset{\sim}{\tau}^*)'H(\underset{\sim}{\tau} - \underset{\sim}{\tau}^*);$$

so that

$$h(\underset{\sim}{\tau}) \approx \exp\left(f(\underset{\sim}{\tau}^*) + (\underset{\sim}{\tau} - \underset{\sim}{\tau}^*)'\underset{\sim}{g} + \tfrac{1}{2}(\underset{\sim}{\tau} - \underset{\sim}{\tau}^*)'H(\underset{\sim}{\tau} - \underset{\sim}{\tau}^*)\right)$$

has a kernel of multivariate normal distribution (e.g., Nandram, Chen, Fu and Manandhar, 2018). It is costly to fit the exact methods because generally there are numerous parameters (e.g. INLA). For our study, we set $\underset{\sim}{\tau}^*$ as the quasi-modes, which are calculated from the EM algorithm.

Approximation Theorem

For the unimodal density,

$$h_1(\underset{\sim}{\tau}_1) = \prod_{i=1}^{\ell}\left[\prod_{j=1}^{g_i}\prod_{k=1}^{3}\Big(\frac{p_{ij}\exp(\theta_{1k}+\nu_{1i})}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}\Big)^{\omega_{ijk}}\Big(\frac{p_{ij}}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}\Big)^{\omega_{ij4}}\right],$$

by approximation Lemma, $\underset{\sim}{\tau}_1$ approximately has a multivariate normal distribution

$$\begin{pmatrix}\underset{\sim}{\nu}_1\\ \underset{\sim}{\theta}_1\end{pmatrix} = \underset{\sim}{\tau}_1 \sim \text{Normal}(\underset{\sim}{\tau}_1^* - H_1^{-1}\underset{\sim}{g}_1, -H_1^{-1}), \quad \underset{\sim}{\tau}_1^* = \begin{pmatrix}\underset{\sim}{\nu}_1^*\\ \underset{\sim}{\theta}_1^*\end{pmatrix},$$

where $\underset{\sim}{\tau}_1^*$ is the quasi-mode, $\underset{\sim}{g}_1$ and H_1 are the gradient vector and the Hessian matrix of $\log h_1(\underset{\sim}{\tau}_1)$ evaluated at $\underset{\sim}{\tau}_1^*$.

The same approximation Lemma applies for $\underset{\sim}{\tau}_2$, with the density function

$$h_2(\underset{\sim}{\tau}_2) = \prod_{i=1}^{\ell}\left[\prod_{j=1}^{g_i}\prod_{k=1}^{3}\Big(\frac{p_{ij}\exp(\theta_{2k}+\nu_{2i})}{1+\sum_{t=1}^{3}\exp(\theta_{2t}+\nu_{2i})}\Big)^{y_{ijk}-\omega_{ijk}}\Big(\frac{p_{ij}}{1+\sum_{t=1}^{3}\exp(\theta_{2t}+\nu_{2i})}\Big)^{y_{ij4}-\omega_{ij4}}\right],$$

to get the normal approximation for $\underset{\sim}{\nu}_2$,

$$\begin{pmatrix}\underset{\sim}{\nu}_2\\ \underset{\sim}{\theta}_2\end{pmatrix} = \underset{\sim}{\tau}_2 \sim \text{Normal}(\underset{\sim}{\tau}_2^* - H_2^{-1}\underset{\sim}{g}_2, -H_2^{-1}), \quad \underset{\sim}{\tau}_2^* = \begin{pmatrix}\underset{\sim}{\nu}_2^*\\ \underset{\sim}{\theta}_2^*\end{pmatrix},$$

where $\underset{\sim}{\tau}_2^*$ is the quasi-mode, $\underset{\sim}{g}_2$ and H_2 are the gradient vector and the Hessian matrix of $\log h_2(\underset{\sim}{\tau}_2)$ evaluated at $\underset{\sim}{\tau}_2^*$.

Construction of Quasi-Modes

Next, we describe how to find quasi-mode $\underset{\sim}{\tau}_1^*$ and $\underset{\sim}{\tau}_2^*$. To find the quasi-modes, we consider

$$\prod_{i=1}^{\ell}\Bigg\{\prod_{j=1}^{g_i}\Bigg[\prod_{k=1}^{3}(p_{ij}\pi_{i1k})^{\omega_{ijk}}(p_{ij}(1-\sum_{k=1}^{3}\pi_{i1k}))^{\omega_{ij4}} \prod_{k=1}^{3}((1-p_{ij})\pi_{i2k})^{y_{ijk}-\omega_{ijk}}((1-p_{ij})(1-\sum_{k=1}^{3}\pi_{i2k}))^{\omega_{ij4}-\omega_{ij4}}\Bigg]\Bigg\},$$

where

$$\pi_{i1k} = \frac{\exp(\theta_{1k}+\nu_{1i})}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}, k=1,2,3;\ \pi_{i14} = 1-\sum_{k=1}^{3}\pi_{i1k},$$

$$\pi_{i2k} = \frac{\exp(\theta_{2k}+\nu_{2i})}{1+\sum_{t=1}^{3}\exp(\theta_{2t}+\nu_{2i})}, k=1,2,3;\ \pi_{i24} = 1-\sum_{k=1}^{3}\pi_{i2k},$$

and

$$\log(\frac{\pi_{i1k}}{1-\sum_{t=1}^{3}\pi_{i1t}}) = \theta_{1k}+\nu_{1i}, i=1,\ldots,\ell, k=1,2,3,$$

$$\log(\frac{\pi_{i2k}}{1-\sum_{t=1}^{3}\pi_{i2t}}) = \theta_{2k}+\nu_{2i}, i=1,\ldots,\ell, k=1,2,3.$$

We perform the EM algorithm to obtain $\hat{\pi}_{i1k}, k=1,2,3,4$ separately in each area to

obtain

$$\widehat{\theta_{1k}+\nu_{1i}} = \log(\frac{\hat{\pi}_{i1k}}{1-\sum_{t=1}^{3}\hat{\pi}_{i1t}}), \quad \widehat{\theta_{2k}+\nu_{2i}} = \log(\frac{\hat{\pi}_{i2k}}{1-\sum_{t=1}^{3}\hat{\pi}_{i2t}}), \ k=1,2,3.$$

Then, we repeat the EM algorithm for all areas combined into a single one (setting $\nu_{1i}=0, \nu_{2i}=0$), to get

$$\hat{\theta}_{1k} = \log(\frac{\hat{\pi}_{1k}}{1-\sum_{t=1}^{3}\hat{\pi}_{1t}}), \ \hat{\theta}_{2k} = \log(\frac{\hat{\pi}_{2k}}{1-\sum_{t=1}^{3}\hat{\pi}_{2t}}), k=1,2,3.$$

Then

$$\hat{\nu}_{1i} = \sum_{k=1}^{3}(\widehat{\theta_{1k}+\nu_{1i}} - \hat{\theta}_{1k})/3, \ \hat{\nu}_{2i} = \sum_{k=1}^{3}(\widehat{\theta_{2k}+\nu_{2i}} - \hat{\theta}_{2k})/3, \ k=1,2,3.$$

We need to check that $-H_1^{-1}$ is positive definite at $\underset{\sim}{\tau}_1^*$ and $-H_2^{-1}$ is positive definite at $\underset{\sim}{\tau}_2^*$, where $\underset{\sim}{\tau}_1^* = (\hat{\nu}_{1i}, i=1\ldots,\ell, \ \hat{\theta}_{1k}, k=1,2,3)$, $\underset{\sim}{\tau}_2^* = (\hat{\nu}_{2i}, i=1\ldots,\ell, \ \hat{\theta}_{2k}, k=1,2,3)$ and $\underset{\sim}{\tau}_1^*$, $\underset{\sim}{\tau}_2^*$ are the solutions just obtained. If these negative Hessian matrices are not positive definite, we jitter $\underset{\sim}{\tau}_1^*$ and $\underset{\sim}{\tau}_2^*$ until they become positive definite. More details of obtaining the quasi-modes are provided in the Appendix A.

Integrated Nested Normal Approximation

From the derivation above, the approximate joint posterior density of $\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1 \mid \underset{\sim}{\omega}$ is

$$\underset{\sim}{\tau}_1 = \begin{pmatrix}\underset{\sim}{\nu}_1 \\ \underset{\sim}{\theta}_1\end{pmatrix} \mid \underset{\sim}{\omega} \sim \text{Normal}\Big\{\begin{pmatrix}\underset{\sim}{\mu}_{\nu_1} \\ \underset{\sim}{\mu}_{\theta_1}\end{pmatrix}, -H_1^{-1}\Big\}.$$

Here $\underset{\sim}{\mu}_{\nu_1}$ and $\underset{\sim}{\mu}_{\theta_1}$ are calculated with $\underset{\sim}{g}_1$, H_1 and $\underset{\sim}{\tau}_1^*$, where $\underset{\sim}{\tau}_1^*$ is the quasi-mode calculated from EM algorithm.

Similar approximation can be made for $\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$ and we get

$$\underset{\sim}{\tau}_2 = \begin{pmatrix}\underset{\sim}{\nu}_2 \\ \underset{\sim}{\theta}_2\end{pmatrix} \mid \underset{\sim}{y}, \underset{\sim}{\omega} \sim \text{Normal}\Big\{\begin{pmatrix}\underset{\sim}{\mu}_{\nu_2} \\ \underset{\sim}{\mu}_{\theta_2}\end{pmatrix}, -H_2^{-1}\Big\}.$$

Now we need to specify $\underset{\sim}{g}$ and H evaluated at $\underset{\sim}{\tau}_1^*$, $\underset{\sim}{\tau}_2^*$. Consider the log likelihood

function

$$\Delta = f(\underset{\sim}{\tau}) = \log h(\underset{\sim}{\tau}) = \sum_{i=1}^{\ell}\Big\{\sum_{j=1}^{g_i}\Big[\sum_{k=1}^{3}\omega_{ijk}(\theta_{1k}+\nu_{1i}) - \sum_{k=1}^{4}\omega_{ijk}\log(1+\sum_{t=1}^{3}e^{\theta_{1t}+\nu_{1i}})\Big\}\Big] + \sum_{k=1}^{3}\frac{1}{2}\lambda_{ai}\theta_{1k}^2 + \sum_{i=1}^{\ell}\frac{1}{2}\lambda_{bi}\nu_{1i}^2,$$

with λ_{ak} $(k = 1 \ldots 3)$ representing the coefficient of the regularization term $\frac{1}{2}\lambda_{ak}\theta_{1k}^2$ for $\underset{\sim}{\theta}_1$ and λ_{bi} $(i = 1 \ldots \ell)$ representing the coefficient of the regularization term $\frac{1}{2}\lambda_{bi}\nu_{1i}^2$ for $\underset{\sim}{\nu}_1$. This term is introduced to get rid of the multicollinearity in solving the inverse Hessian matrix. Just like in ridge regression which proceeds by adding a small value to the diagonal elements of the correlation matrix, here the added term $\frac{1}{2}\lambda_{bi}\nu_{1i}^2$ (or $\frac{1}{2}\lambda_{ak}\theta_{1k}^2$) actually result in a λ_{bi} (λ_{ak}) more variance for the variance of ν_{1i} (θ_{1k}), where λ_{bi} (λ_{ak}) takes a very small value. In the actual calculation, a λ_{bi} (λ_{ak}) is chosen to be a small value proportional to the variance, we use 0.001 here. These regularization terms work like the priors of the ν_{1i} (θ_{1k}).

Obtain the Approximate Normal Distribution

Consider $\underset{\sim}{\tau}_1$ first, once we got $\underset{\sim}{\tau}_1^*$, $\underset{\sim}{g}$ and H evaluated at the approximate posterior mode $\underset{\sim}{\tau}_1 = \underset{\sim}{\tau}_1^*$ can also be obtained as

$$\underset{\sim}{g}_1 = \Big(\frac{\partial\Delta}{\partial\nu_{11}} \;\cdots\; \frac{\partial\Delta}{\partial\nu_{1\ell}} \;\frac{\partial\Delta}{\partial\underset{\sim}{\theta}_1}\Big)^T \Big|_{\underset{\sim}{\nu}_1=\underset{\sim}{\nu}_1^*,\ \underset{\sim}{\theta}_1=\underset{\sim}{\theta}_1^*},$$

$$H_1 = \begin{pmatrix} \frac{\partial^2\Delta}{\partial\nu_{11}^2} & \cdots & 0 & \frac{\partial^2\Delta}{\partial\nu_{11}\partial\underset{\sim}{\theta}_1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{\partial^2\Delta}{\partial\nu_{1\ell}^2} & \frac{\partial^2\Delta}{\partial\nu_{1\ell}\partial\underset{\sim}{\theta}_1} \\ \frac{\partial^2\Delta}{\partial\nu_{11}\partial\underset{\sim}{\theta}_1} & \ddots & \frac{\partial^2\Delta}{\partial\nu_{1\ell}\partial\underset{\sim}{\theta}_1} & \frac{\partial^2\Delta}{\partial\underset{\sim}{\theta}_1^2} \end{pmatrix} \Big|_{\underset{\sim}{\nu}_1=\underset{\sim}{\nu}_1^*,\ \underset{\sim}{\theta}_1=\underset{\sim}{\theta}_1^*}.$$

The partial derivatives can be expressed in terms of latent variables ω_{ijk} as

$$\frac{\partial \Delta}{\partial \theta_{1k}} = \sum_{i=1}^{\ell}\sum_{j=1}^{g_i}\omega_{ijk} - \sum_{i=1}^{\ell}\sum_{j=1}^{g_i}\sum_{s=1}^{4}\omega_{ijs}\frac{\exp(\theta_{1k}+\nu_{1i})}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}, \qquad k=1,2,3;$$

$$\frac{\partial \Delta}{\partial \nu_{1i}} = -\sum_{j=1}^{g_i}\omega_{ij4} + \sum_{j=1}^{g_i}\sum_{s=1}^{4}\omega_{ijs}\frac{1}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}, \qquad i=1\ldots\ell;$$

$$\frac{\partial^2 \Delta}{\partial \theta_{1k}^2} = -\sum_{i=1}^{\ell}\sum_{j=1}^{g_i}\Big[\sum_{s=1}^{4}\omega_{ijs}\frac{\exp(\theta_{1k}+\nu_{1i})(1+\sum_{t\neq k}^{3}\exp(\theta_{1t}+\nu_{1i}))}{(1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i}))^2}\Big]$$

$$\frac{\partial^2 \Delta}{\partial \theta_{1k}\partial \theta_{1h}} = \sum_{i=1}^{\ell}\sum_{j=1}^{g_i}\Big[\sum_{s=1}^{4}\omega_{ijs}\frac{\exp(\theta_{1k}+\nu_{1i})\exp(\theta_{1h}+\nu_{1i})}{(1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i}))^2}\Big], \qquad 1\leq k\neq h\leq 3;$$

$$\frac{\partial^2 \Delta}{\partial \nu_{1i}^2} = -\sum_{j=1}^{g_i}\Big[\sum_{s=1}^{4}\omega_{ijs}\frac{\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}{(1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i}))^2}\Big],$$

and

$$\frac{\partial^2 \Delta}{\partial \nu_{1i}\partial \theta_{1k}} = -\sum_{j=1}^{g_i}\Big[\sum_{s=1}^{4}\omega_{ijs}\frac{\exp(\theta_{1k}+\nu_{1i})}{(1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i}))^2}\Big].$$

We use notation $\underset{\sim}{g}_1 = \begin{pmatrix}\underset{\sim}{g}_{11}\\ \underset{\sim}{g}_{12}\end{pmatrix}$ and $H_1 = -\begin{pmatrix}D_1 & C_1'\\ C_1 & B_1\end{pmatrix}$ for computational convenience, where

$$\underset{\sim}{g}_{11} = \Big(\frac{\partial \Delta}{\partial \nu_{11}} \;\cdots\; \frac{\partial \Delta}{\partial \nu_{1\ell}}\Big)^T, \quad \underset{\sim}{g}_{12} = \frac{\partial \Delta}{\partial \underset{\sim}{\theta}_1},$$

$$B_1 = -\frac{\partial^2 \Delta}{\partial \underset{\sim}{\theta}_1^2} + \Lambda_a,\; C_1 = -\Big(\frac{\partial^2 \Delta}{\partial \nu_{11}\partial \underset{\sim}{\theta}_1} \;\cdots\; \frac{\partial^2 \Delta}{\partial \nu_{1\ell}\partial \underset{\sim}{\theta}_1}\Big),\; D_1 = -\begin{pmatrix}\frac{\partial^2 \Delta}{\partial \nu_{11}^2} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{\partial^2 \Delta}{\partial \nu_{1\ell}^2}\end{pmatrix} + \Lambda_b,$$

where $\Lambda_a = \mathrm{diag}(\lambda_{a1}, \lambda_{a2}, \lambda_{a3})$ and $\Lambda_b = diag(\lambda_{b1}, \ldots, \lambda_{b\ell})$. Here B_1 is a 3×3 non-positive definite square matrix; D_1 is a ℓ^{th} order diagonal matrix which is also non-positive definite. The variance-covariance matrix $-H_1^{-1}$ can then be constructed from the block matrices above, since D_1 is a nonsingular matrix and the Schur complement $B_1 - C_1 D_1^{-1} C_1'$ of D_1

is invertible,

$$-H_1^{-1} = \begin{pmatrix} D_1 & C_1' \\ C_1 & B_1 \end{pmatrix}^{-1} = \begin{pmatrix} E_1 & F_1' \\ F_1 & G_1 \end{pmatrix},$$

where

$$\begin{aligned} E_1 &= D_1^{-1} + D_1^{-1}C_1'(B_1 - C_1D_1^{-1}C_1')^{-1}C_1D_1^{-1}, \\ F_1 &= -(B_1 - C_1D_1^{-1}C_1')^{-1}C_1D_1^{-1}, \\ G_1 &= (B_1 - C_1D_1^{-1}C_1')^{-1}. \end{aligned}$$

According to the Lemma, the estimated mean of the multivariate normal distribution is

$$\tau_1^* - H_1^{-1}\underset{\sim}{g}_1 = \begin{pmatrix} \underset{\sim}{\nu}_1^* \\ \underset{\sim}{\theta}_1^* \end{pmatrix} + \begin{pmatrix} E_1 & F_1' \\ F_1 & G_1 \end{pmatrix} \begin{pmatrix} \underset{\sim}{g}_{11} \\ \underset{\sim}{g}_{12} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\nu}_1^* + E_1\underset{\sim}{g}_{11} + F_1'\underset{\sim}{g}_{12} \\ \underset{\sim}{\theta}_1^* + F_1\underset{\sim}{g}_{11} + G_1\underset{\sim}{g}_{12} \end{pmatrix}.$$

Letting

$$\begin{aligned} \underset{\sim}{\mu}_{\nu_1} &= \underset{\sim}{\nu}_1^* + E_1\underset{\sim}{g}_{11} + F_1'\underset{\sim}{g}_{12}, \\ \underset{\sim}{\mu}_{\theta_1} &= \underset{\sim}{\theta}_1^* + F_1\underset{\sim}{g}_{11} + G_1\underset{\sim}{g}_{12}, \end{aligned}$$

then the joint posterior density of $\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$ is

$$\begin{pmatrix} \underset{\sim}{\nu}_1 \\ \underset{\sim}{\theta}_1 \end{pmatrix} \mid \underset{\sim}{y} \sim \text{Normal}\Big\{ \begin{pmatrix} \underset{\sim}{\mu}_{\nu_1} \\ \underset{\sim}{\mu}_{\theta_1} \end{pmatrix}, -H_1^{-1} \Big\}.$$

By a property of the multivariate normal distribution, the conditional posterior density of $\underset{\sim}{\nu}_1 \mid \underset{\sim}{\theta}_1, \underset{\sim}{y}, \underset{\sim}{\omega}$ and $\underset{\sim}{\theta}_1 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$ are

$$\begin{aligned} \underset{\sim}{\nu}_1 \mid \underset{\sim}{\theta}_1, \underset{\sim}{y}, \underset{\sim}{\omega} &\sim N(\underset{\sim}{\mu}_{\nu_1} - D_1^{-1}C_1'(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1}), D_1^{-1}), \\ \underset{\sim}{\theta}_1 \mid \underset{\sim}{y}, \underset{\sim}{\omega} &\sim N(\underset{\sim}{\mu}_{\theta_1}, G_1). \end{aligned}$$

Similarly for $\underset{\sim}{\tau}_2$, $\underset{\sim}{g}$ and H evaluated at the approximate posterior mode $\underset{\sim}{\tau}_2 = \underset{\sim}{\tau}_2^*$ can also be obtained as

$$\underset{\sim}{g}_2 = \Big(\frac{\partial \Delta}{\partial \nu_{21}} \ \cdots \ \frac{\partial \Delta}{\partial \nu_{2\ell}} \ \frac{\partial \Delta}{\partial \underset{\sim}{\theta}_2}\Big)^T \Big|_{\underset{\sim}{\nu}_2 = \underset{\sim}{\nu}_2^*,\ \underset{\sim}{\theta}_2 = \underset{\sim}{\theta}_2^*},$$

$$H_2 = \begin{pmatrix} \frac{\partial^2 \Delta}{\partial \nu_{21}^2} & \cdots & 0 & \frac{\partial^2 \Delta}{\partial \nu_{21} \partial \underset{\sim}{\theta}_2} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{\partial^2 \Delta}{\partial \nu_{2\ell}^2} & \frac{\partial^2 \Delta}{\partial \nu_{2\ell} \partial \underset{\sim}{\theta}_2} \\ \frac{\partial^2 \Delta}{\partial \nu_{21} \partial \underset{\sim}{\theta}_2} & \ddots & \frac{\partial^2 \Delta}{\partial \nu_{2\ell} \partial \underset{\sim}{\theta}_2} & \frac{\partial^2 \Delta}{\partial \underset{\sim}{\theta}_2^2} \end{pmatrix} \Big|_{\underset{\sim}{\nu}_2 = \underset{\sim}{\nu}_2^*,\ \underset{\sim}{\theta}_2 = \underset{\sim}{\theta}_2^*}.$$

The partial derivatives can be expressed in terms of latent variables ω_{ijk} as

$$\frac{\partial \Delta}{\partial \theta_{2k}} = \sum_{i=1}^{\ell} \sum_{j=1}^{g_i} (y_{ijk} - \omega_{ijk}) - \sum_{i=1}^{\ell} \sum_{j=1}^{g_i} \sum_{s=1}^{4} (y_{ijs} - \omega_{ijs}) \frac{\exp(\theta_{2k} + \nu_{2i})}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}, \qquad k = 1, 2, 3;$$

$$\frac{\partial \Delta}{\partial \nu_{2i}} = -\sum_{j=1}^{g_i} (y_{ij4} - \omega_{ij4}) + \sum_{j=1}^{g_i} \sum_{s=1}^{4} (y_{ijs} - \omega_{ijs}) \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}, \qquad i = 1 \ldots \ell;$$

$$\frac{\partial^2 \Delta}{\partial \theta_{2k}^2} = -\sum_{i=1}^{\ell} \sum_{j=1}^{g_i} \Big[\sum_{s=1}^{4} (y_{ijs} - \omega_{ijs}) \frac{\exp(\theta_{2k} + \nu_{2i})(1 + \sum_{t \neq k}^{3} \exp(\theta_{2t} + \nu_{2i}))}{(1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i}))^2} \Big]$$

$$\frac{\partial^2 \Delta}{\partial \theta_{2k} \partial \theta_{2h}} = \sum_{i=1}^{\ell} \sum_{j=1}^{g_i} \Big[\sum_{s=1}^{4} (y_{ijs} - \omega_{ijs}) \frac{\exp(\theta_{2k} + \nu_{2i}) \exp(\theta_{2h} + \nu_{2i})}{(1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i}))^2} \Big], \qquad 1 \leq k \neq h \leq 3;$$

$$\frac{\partial^2 \Delta}{\partial \nu_{2i}^2} = -\sum_{j=1}^{g_i} \Big[\sum_{s=1}^{4} (y_{ijs} - \omega_{ijs}) \frac{\sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}{(1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i}))^2} \Big],$$

and

$$\frac{\partial^2 \Delta}{\partial \nu_{2i} \partial \theta_{2k}} = -\sum_{j=1}^{g_i} \Big[\sum_{s=1}^{4} (y_{ijs} - \omega_{ijs}) \frac{\exp(\theta_{2k} + \nu_{2i})}{(1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i}))^2} \Big].$$

We use notation $\underset{\sim}{g}_1 = \begin{pmatrix} \underset{\sim}{g}_{21} \\ \underset{\sim}{g}_{22} \end{pmatrix}$ and $H_2 = -\begin{pmatrix} D_2 & C_2' \\ C_2 & B_2 \end{pmatrix}$ for computational convenience, where

$$\underset{\sim}{g}_{21} = \Big(\frac{\partial \Delta}{\partial \nu_{21}} \ \cdots \ \frac{\partial \Delta}{\partial \nu_{2\ell}}\Big)^T, \quad \underset{\sim}{g}_{22} = \frac{\partial \Delta}{\partial \underset{\sim}{\theta}_2},$$

$$B_2 = -\frac{\partial^2 \Delta}{\partial \underset{\sim}{\theta}_2^2} + \Lambda_c,\ C_2 = -\Big(\frac{\partial^2 \Delta}{\partial \nu_{21} \partial \underset{\sim}{\theta}_2} \ \cdots \ \frac{\partial^2 \Delta}{\partial \nu_{2\ell} \partial \underset{\sim}{\theta}_2}\Big),\ D_2 = -\begin{pmatrix} \frac{\partial^2 \Delta}{\partial \nu_{21}^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial^2 \Delta}{\partial \nu_{2\ell}^2} \end{pmatrix} + \Lambda_d \ ,$$

where $\Lambda_c = diag(\lambda_{c1}, \lambda_{c2}, \lambda_{c3})$ and $\Lambda_d = diag(\lambda_{d1}, \ldots, \lambda_{d\ell})$. Here B_2 is a 3×3 non-positive definite square matrix; D_2 is a ℓth order diagonal matrix which is also non-positive definite. The variance-covariance matrix $-H_2^{-1}$ can then be constructed in the same way,

$$-H_2^{-1} = \begin{pmatrix} D_2 & C_2' \\ C_2 & B_2 \end{pmatrix}^{-1} = \begin{pmatrix} E_2 & F_2' \\ F_2 & G_2 \end{pmatrix},$$

where

$$\begin{aligned} E_2 &= D_2^{-1} + D_2^{-1} C_2' (B_2 - C_2 D_2^{-1} C_2')^{-1} C_2 D_2^{-1}, \\ F_2 &= -(B_2 - C_2 D_2^{-1} C_2')^{-1} C_2 D_2^{-1}, \\ G_2 &= (B_2 - C_2 D_2^{-1} C_2')^{-1}. \end{aligned}$$

The mean estimation for $\underset{\sim}{\tau}_2$ is

$$\underset{\sim}{\tau}_2^* - H_2^{-1} \underset{\sim}{g}_2 = \begin{pmatrix} \underset{\sim}{\nu}_2^* \\ \underset{\sim}{\theta}_2^* \end{pmatrix} + \begin{pmatrix} E_2 & F_2' \\ F_2 & G_2 \end{pmatrix} \begin{pmatrix} \underset{\sim}{g}_{21} \\ \underset{\sim}{g}_{22} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\nu}_2^* + E_2 \underset{\sim}{g}_{21} + F_2' \underset{\sim}{g}_{22} \\ \underset{\sim}{\theta}_2^* + F_2 \underset{\sim}{g}_{21} + G_2 \underset{\sim}{g}_{22} \end{pmatrix}.$$

Let

$$\begin{aligned} \underset{\sim}{\mu}_{\nu_2} &= \underset{\sim}{\nu}_2^* + E_2 \underset{\sim}{g}_{21} + F_2' \underset{\sim}{g}_{22}, \\ \underset{\sim}{\mu}_{\theta_2} &= \underset{\sim}{\theta}_2^* + F_2 \underset{\sim}{g}_{21} + G_2 \underset{\sim}{g}_{22}, \end{aligned}$$

then the joint posterior density of $\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$ is

$$\begin{pmatrix} \underset{\sim}{\nu}_2 \\ \underset{\sim}{\theta}_2 \end{pmatrix} \mid \underset{\sim}{y} \sim \text{Normal}\Big\{ \begin{pmatrix} \underset{\sim}{\mu}_{\nu_2} \\ \underset{\sim}{\mu}_{\theta_2} \end{pmatrix}, -H_2^{-1} \Big\}.$$

Therefore, the conditional posterior density of $\underset{\sim}{\nu}_2 \mid \underset{\sim}{\theta}_2, \underset{\sim}{y}, \underset{\sim}{\omega}$ and $\underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$ are

$$\begin{aligned} \underset{\sim}{\nu}_2 \mid \underset{\sim}{\theta}_2, \underset{\sim}{y}, \underset{\sim}{\omega} &\sim N(\underset{\sim}{\mu}_{\nu_2} - D_2^{-1} C_2' (\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2}), D_2^{-1}), \\ \underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega} &\sim N(\underset{\sim}{\mu}_{\theta_2}, G_2). \end{aligned}$$

So far, we have a closed form of the approximate normal distribution of $\underset{\sim}{\nu}_1 \mid \underset{\sim}{\theta}_1, \underset{\sim}{y}, \underset{\sim}{\omega}$ and $\underset{\sim}{\theta}_1 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$; $\underset{\sim}{\nu}_2 \mid \underset{\sim}{\theta}_2, \underset{\sim}{y}, \underset{\sim}{\omega}$ and $\underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}$, based on which we build the integrated nested normal approximation Model (INNA).

3.3 The Integrated Nested Normal Approximation Model

Now we plan to add priors for $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$ upon the approximate conditional densities. We will keep the prior $\pi(\underset{\sim}{\theta}_1) = 1$, $\pi(\underset{\sim}{\theta}_2) = 1$ meanwhile.

3.3.1 Case of Independent $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$

First, we consider the situation when $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$ are independent, with priors

$$\underset{\sim}{\nu}_1 \mid \delta_1^2 \sim N(\underset{\sim}{0}, \delta_1^2 I), \quad \underset{\sim}{\nu}_2 \mid \delta_2^2 \sim N(\underset{\sim}{0}, \delta_2^2 I)$$

and together with the joint prior

$$\pi(\delta_1^2, \delta_2^2) \propto \frac{1}{(1+\delta_1^2)^2} \frac{1}{(1+\delta_2^2)^2}.$$

then we have the following approximation model for $(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1)$ and $(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2)$ separately, and the conditional distribution of all the stages are normal distributions, Finally, by the multiplication rule, we can write out the posterior joint distribution of all the parameters as

$$\begin{aligned}
\pi(\underset{\sim}{\nu}, \underset{\sim}{\theta}, \delta_1^2, \delta_2^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \propto & \pi(\underset{\sim}{\nu} \mid \underset{\sim}{\theta}, \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\nu} \mid \delta_1^2, \delta_2^2)\pi(\delta_1^2, \delta_2^2) \\
\propto & \pi(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
\propto & \pi(\underset{\sim}{\nu}_1 \mid \underset{\sim}{\theta}_1, \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\theta}_1 \mid \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\nu}_1 \mid \delta_1^2) \\
\times & \pi(\underset{\sim}{\nu}_2 \mid \underset{\sim}{\theta}_2, \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\nu}_2 \mid \delta_2^2) \\
\times & \pi(\delta_1^2, \delta_2^2).
\end{aligned}$$

Observing that $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$ are independent and also the prior for (δ_1^2, δ_2^2) are separable so that we can deal with the parameter set $(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2)$ and $(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2)$ independently given the data $\underset{\sim}{y}$ and latent variable $\underset{\sim}{\omega}$. Hence we study the posterior joint distribution for $(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2)$ first

$$\begin{aligned}
& \pi(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
& \propto \pi(\underset{\sim}{\nu}_1 \mid \underset{\sim}{\theta}_1, \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\theta}_1 \mid \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\nu}_1 \mid \delta_1^2)\pi(\delta_1^2) \\
& \propto \frac{1}{|D_1^{-1})|^{1/2}} e^{-\frac{1}{2}[\underset{\sim}{\nu}_1 - (\underset{\sim}{\mu}_{\nu_1} - D_1^{-1}C_1'(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1}))]' D_1 [\underset{\sim}{\nu}_1 - (\underset{\sim}{\mu}_{\nu_1} - D_1^{-1}C_1'(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1}))]} \\
& \times \frac{1}{|G_1|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})' G_1^{-1} (\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})} \times \frac{1}{|\delta_1^2 I|^{1/2}} e^{-\frac{1}{2}\underset{\sim}{\nu}_1'(\delta_1^2 I)^{-1}\underset{\sim}{\nu}_1} \times \frac{1}{(1+\delta_1^2)^2} ,
\end{aligned}$$

and consequently get the posterior conditional distributions for each parameter. Start with $\underset{\sim}{\nu}_1$, we have

$$\underset{\sim}{\nu}_1 \mid \delta_1^2, \underset{\sim}{\theta}_1, \underset{\sim}{y}, \underset{\sim}{\omega} \sim N\{(D_1 + \frac{1}{\delta_1^2} I)^{-1}(D_1 \underset{\sim}{\mu}_{\nu_1} - C_1'(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})), (D_1 + \frac{1}{\delta_1^2} I)^{-1}\}.$$

Since $\underset{\sim}{\nu}_1$ has a multivariate normal distribution, we can integrate out $\underset{\sim}{\nu}_1$ from the joint posterior density $\pi(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega})$ and get the joint posterior density of $\underset{\sim}{\theta}_1$ and δ_1^2 as

$$\begin{aligned}
&\pi(\underset{\sim}{\theta}_1, \delta_1^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
&\propto e^{-\frac{1}{2}\{[\underset{\sim}{\mu}_{\nu_1} - D_1^{-1}C_1'(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})]'(D_1+\delta_1^2 I)^{-1}[\underset{\sim}{\mu}_{\nu_1} - D_1^{-1}C_1'(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})] + (\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})' G_1^{-1}(\underset{\sim}{\theta}_1 - \underset{\sim}{\mu}_{\theta_1})\}} \\
&\times \frac{|D_1|^{1/2}}{|D_1 + \frac{1}{\delta_1^2} I|^{1/2} |\delta_1^2 I|^{1/2} |G_1|^{1/2}} \frac{1}{(1+\delta_1^2)^2} \ .
\end{aligned}$$

So that,

$$\begin{aligned}
&\underset{\sim}{\theta}_1 \mid \delta_1^2, \underset{\sim}{y}, \underset{\sim}{\omega} \sim \\
&N\{(C_1 D_1^{-1}(D_1^{-1} + \delta_1^2 I)^{-1} D_1^{-1} C_1' + G_1^{-1})^{-1}[(\underset{\sim}{\mu}_{\nu_1} + \underset{\sim}{\mu}_{\theta_1}' C_1 D_1^{-1})(D_1^{-1} + \delta_1^2 I)^{-1} D_1^{-1} C_1' + \underset{\sim}{\mu}_{\theta_1}' G_1^{-1}] \\
&\quad , (C_1 D_1^{-1}(D_1^{-1} + \delta_1^2 I)^{-1} D_1^{-1} C_1' + G_1^{-1})^{-1}\} \ .
\end{aligned}$$

Now by integrating out $\underset{\sim}{\theta}_1$

$$\begin{aligned}
&\delta_1^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega} \propto \\
&\frac{|D_1|^{1/2}}{|D_1 + \frac{1}{\delta_1^2} I|^{1/2} |\delta_1^2 I|^{1/2} |G_1|^{1/2}} \frac{1}{(1+\delta_1^2)^2} |C_1 D_1^{-1}(D_1^{-1} + \delta_1^2 I)^{-1} D_1^{-1} C_1' + G_1^{-1}|^{1/2}.
\end{aligned}$$

Then we can draw δ_1^2 using grid method. A Similar approach applies for the parameter set $(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2)$ with data $\underset{\sim}{y} - \underset{\sim}{\omega}$ instead of $\underset{\sim}{\omega}$ in obtaining the second order derivatives. Based on the joint posterior distribution for $(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2)$

$$\begin{aligned}
&\pi(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
&\propto \pi(\underset{\sim}{\nu}_2 \mid \underset{\sim}{\theta}_2, \underset{\sim}{y}, \underset{\sim}{\omega}) \pi(\underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \pi(\underset{\sim}{\nu}_2 \mid \delta_2^2) \pi(\delta_2^2) \\
&\propto \frac{1}{|D_2^{-1})|^{1/2}} e^{-\frac{1}{2}[\underset{\sim}{\nu}_2 - (\underset{\sim}{\mu}_{\nu_2} - D_2^{-1}C_2'(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2}))]' D_2 [\underset{\sim}{\nu}_2 - (\underset{\sim}{\mu}_{\nu_2} - D_2^{-1}C_2'(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2}))]} \\
&\times \frac{1}{|G_2|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})' G_2^{-1}(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})} \times \frac{1}{|\delta_2^2 I|^{1/2}} e^{-\frac{1}{2}\underset{\sim}{\nu}_2' (\delta_2^2 I)^{-1} \underset{\sim}{\nu}_2} \times \frac{1}{(1+\delta_2^2)^2} \ .
\end{aligned}$$

So that,

$$\underset{\sim}{\nu}_2 \mid \delta_2^2, \underset{\sim}{\theta}_2, \underset{\sim}{y}, \underset{\sim}{\omega} \sim N\{(D_2 + \frac{1}{\delta_2^2} I)^{-1}(D_2 \underset{\sim}{\mu}_{\nu_2} - C_2'(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})), (D_2 + \frac{1}{\delta_2^2} I)^{-1}\}.$$

Similarly, we can integrate out $\underset{\sim}{\nu}_2$ from the joint posterior density $\pi(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega})$ and get the joint posterior density of $\underset{\sim}{\theta}_2$ and δ_2^2 as

$$\begin{aligned}
&\pi(\underset{\sim}{\theta}_2, \delta_2^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
&\propto e^{-\frac{1}{2}\{[\underset{\sim}{\mu}_{\nu_2} - D_2^{-1}C_2'(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})]'(D_2 + \delta_2^2 I)^{-1}[\underset{\sim}{\mu}_{\nu_2} - D_2^{-1}C_2'(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})] + (\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})' G_2^{-1}(\underset{\sim}{\theta}_2 - \underset{\sim}{\mu}_{\theta_2})\}} \\
&\times \frac{|D_2|^{1/2}}{|D_2 + \frac{1}{\delta_2^2} I|^{1/2} |\delta_2^2 I|^{1/2} |G_2|^{1/2}} \frac{1}{(1 + \delta_2^2)^2} \ .
\end{aligned}$$

So that,

$$\begin{aligned}
&\underset{\sim}{\theta}_2 \mid \delta_2^2, \underset{\sim}{y}, \underset{\sim}{\omega} \sim \\
&N\{(C_2 D_2^{-1}(D_2^{-1} + \delta_2^2 I)^{-1} D_2^{-1} C_2' + G_2^{-1})^{-1}[(\underset{\sim}{\mu}_{\nu_2} + \underset{\sim}{\mu}_{\theta_2}' C_2 D_2^{-1})(D_2^{-1} + \delta_2^2 I)^{-1} D_2^{-1} C_2' + \underset{\sim}{\mu}_{\theta_2}' G_2^{-1}] \\
&\quad , (C_2 D_2^{-1}(D_2^{-1} + \delta_2^2 I)^{-1} D_2^{-1} C_2' + G_2^{-1})^{-1}\}.
\end{aligned}$$

Now by integrating out $\underset{\sim}{\theta}_2$

$$\begin{aligned}
&\delta_2^2 \mid \underset{\sim}{y}, \underset{\sim}{\omega} \propto \\
&\frac{|D_2|^{1/2}}{|D_2 + \frac{1}{\delta_2^2} I|^{1/2} |\delta_2^2 I|^{1/2} |G_2|^{1/2}} \frac{1}{(1 + \delta_2^2)^2} |C_2 D_2^{-1}(D_2^{-1} + \delta_2^2 I)^{-1} D_2^{-1} C_2' + G_2^{-1}|^{1/2}.
\end{aligned}$$

Then we can draw δ_2^2 using grid method.

3.3.2 Case of Correlated $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$

Now we assume a general case that $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$ are correlated with the correlation denoted as ρ, which is

$$\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2 \mid \delta_1^2, \delta_2^2, \rho \sim \text{Normal}\Big\{\underset{\sim}{0}, M\Big\}$$

where $M = \begin{pmatrix} \delta_1^2 I_{\ell\times\ell} & \rho\delta_1\delta_2 I_{\ell\times\ell} \\ \rho\delta_1\delta_2 I_{\ell\times\ell} & \delta_2^2 I_{\ell\times\ell} \end{pmatrix}$, representing the covariate matrix with correlation ρ and I is a $\ell \times \ell$ identity matrix, together with the joint prior of δ_1^2, δ_2^2

$$\pi(\delta_1^2, \delta_2^2, \rho) \propto \frac{1}{(1+\delta_1^2)^2}\frac{1}{(1+\delta_2^2)^2}.$$

With the combined parameter vector $\underset{\sim}{\nu} = \begin{pmatrix} \underset{\sim}{\nu}_1 \\ \underset{\sim}{\nu}_2 \end{pmatrix}$ and $\underset{\sim}{\theta} = \begin{pmatrix} \underset{\sim}{\theta}_1 \\ \underset{\sim}{\theta}_2 \end{pmatrix}$, the joint probability density function

$$\begin{aligned}
&\pi(\underset{\sim}{\nu}, \underset{\sim}{\theta}, \delta_1^2, \delta_2^2, \rho \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
\propto&\pi(\underset{\sim}{\nu} \mid \underset{\sim}{\theta}, \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \underset{\sim}{\omega})\pi(\underset{\sim}{\nu} \mid \delta_1^2, \delta_2^2, \rho)\pi(\delta_1^2, \delta_2^2) \\
\propto&\frac{1}{|D^{-1})|^{1/2}} e^{-\frac{1}{2}[\underset{\sim}{\nu}-(\underset{\sim}{\mu}_\nu - D^{-1}C'(\underset{\sim}{\theta}-\underset{\sim}{\mu}_\theta))]' D[\underset{\sim}{\nu}-(\underset{\sim}{\mu}_\nu - D^{-1}C'(\underset{\sim}{\theta}-\underset{\sim}{\mu}_\theta))]} \\
\times&\frac{1}{|G|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{\theta}-\underset{\sim}{\mu}_\theta)'G^{-1}(\underset{\sim}{\theta}-\underset{\sim}{\mu}_\theta)} \times \frac{1}{|M|^{1/2}} e^{-\frac{1}{2}\underset{\sim}{\nu}' M^{-1}\underset{\sim}{\nu}} \times \frac{1}{(1+\delta_1^2)^2}\frac{1}{(1+\delta_2^2)^2}
\end{aligned}$$

where $D = \begin{pmatrix} D_1 & \mathbf{0} \\ \mathbf{0} & D_2 \end{pmatrix}$, $C = \begin{pmatrix} C_1 & \mathbf{0} \\ \mathbf{0} & C_2 \end{pmatrix}$, $G = \begin{pmatrix} G_1 & \mathbf{0} \\ \mathbf{0} & G_2 \end{pmatrix}$, and D_1, C_1, G_1 are the matrix corresponding to the latent variable $\underset{\sim}{\omega}$ in obtaining the Hessian matrix of $\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1$; D_2, C_2, G_2 are those corresponding to $\underset{\sim}{y} - \underset{\sim}{\omega}$ in obtaining the Hessian matrix of $\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2$,

with $\underset{\sim}{\mu}_\nu = \begin{pmatrix} \underset{\sim}{\mu}_{\nu_1} \\ \underset{\sim}{\mu}_{\nu_2} \end{pmatrix}$ and $\underset{\sim}{\mu}_\theta = \begin{pmatrix} \underset{\sim}{\mu}_{\theta_1} \\ \underset{\sim}{\mu}_{\theta_2} \end{pmatrix}$. Then,

$$\underset{\sim}{\nu} \mid \underset{\sim}{\theta}, \delta_1^2, \delta_2^2, \rho, \underset{\sim}{y}, \underset{\sim}{\omega} \sim N\{(D + M^{-1})^{-1}(D\underset{\sim}{\mu}_\nu - C'(\underset{\sim}{\theta} - \underset{\sim}{\mu}_\theta)), (D + M^{-1})^{-1}\}.$$

Similarly $\underset{\sim}{\nu}$ has a multivariate normal distribution, we can integrate out $\underset{\sim}{\nu}$ from the joint posterior density $\pi(\underset{\sim}{\nu}, \underset{\sim}{\theta}, \delta_1^2, \delta_2^2, \rho \mid \underset{\sim}{y}, \underset{\sim}{\omega})$ and get the joint posterior density of $\underset{\sim}{\theta}$ and $\delta_1^2, \delta_2^2, \rho$ as

$$\begin{aligned} &\pi(\underset{\sim}{\theta}, \delta_1^2, \delta_2^2, \rho \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\ &\propto e^{-\frac{1}{2}\{[\underset{\sim}{\mu}_\nu - D^{-1}C'(\underset{\sim}{\theta} - \underset{\sim}{\mu}_\theta)]'(D+M)^{-1}[\underset{\sim}{\mu}_\nu - D^{-1}C'(\underset{\sim}{\theta} - \underset{\sim}{\mu}_\theta)] + (\underset{\sim}{\theta} - \underset{\sim}{\mu}_\theta)'G^{-1}(\underset{\sim}{\theta} - \underset{\sim}{\mu}_\theta)\}} \\ &\times \frac{|D|^{1/2}}{|D + M^{-1}|^{1/2}|M|^{1/2}|G|^{1/2}} \frac{1}{(1+\delta_1^2)^2} \frac{1}{(1+\delta_2^2)^2} \ . \end{aligned}$$

So that,

$$\begin{aligned} &\underset{\sim}{\theta} \mid \delta_1^2, \delta_2^2, \rho, \underset{\sim}{y}, \underset{\sim}{\omega} \sim \\ &N\{(CD^{-1}(D^{-1} + M)^{-1}D^{-1}C' + G^{-1})^{-1}[(\underset{\sim}{\mu}_\nu + \underset{\sim}{\mu}'_\theta CD^{-1})(D^{-1} + M)^{-1}D^{-1}C' + \underset{\sim}{\mu}'_\theta G^{-1}] \\ &\quad , (CD^{-1}(D^{-1} + M)^{-1}D^{-1}C' + G^{-1})^{-1}\} \ . \end{aligned}$$

Now by integrating out $\underset{\sim}{\theta}$,

$$\begin{aligned} &\delta_1^2, \delta_2^2, \rho \mid \underset{\sim}{y}, \underset{\sim}{\omega} \propto \\ &\frac{|D|^{1/2}}{|D + M^{-1}|^{1/2}|M|^{1/2}|G|^{1/2}} \frac{1}{(1+\delta_1^2)^2} \frac{1}{(1+\delta_2^2)^2} |CD^{-1}(D^{-1} + M)^{-1}D^{-1}C' + G^{-1}|^{1/2} \ . \end{aligned}$$

Then we can draw $\delta_1^2, \delta_2^2, \rho$ using a Gibbs sampler with grid method. When $\rho = 0$, it can be shown that the correlated model degenerates to the model of the independent $\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2$ case.

3.4 Complete Generalized Mixed Effects Model Without Approximations

We now use a proper Cauchy prior on $\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2$,

$$\pi(\theta_{sk}) = \frac{1}{\pi\sigma^*_{sk}[1 + (\frac{\theta_{sk}-\theta^*_{sk}}{\sigma^*_{sk}})^2]}, \quad s = 1, 2, \ k = 1, 2, 3,$$

where θ^*_{sk} and σ^*_{sk} are the posterior mean and standard deviation of θ_{sk}, calculated from the Gibbs sampler from the approximation model with correlated $\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2$. The Cauchy prior is necessary here which helps to avoid the extremely long range dependence in the Gibbs sampler. Back to the posterior density function of the generalized mixed effects model, under the correlated prior of $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$, which is

$$\begin{aligned}
&\pi(\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2, \underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \delta_1^2, \delta_2^2, \rho \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \\
&\propto \pi(\underset{\sim}{y}, \underset{\sim}{\omega} \mid \underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2, \underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2)\pi(\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2 \mid \delta_1^2, \delta_2^2, \rho)\pi(\delta_1^2, \delta_2^2)\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2) \\
&= \prod_{i=1}^{\ell}\Bigg\{\prod_{j=1}^{g_i}\Bigg[\prod_{k=1}^{3}\Big(\frac{p_{ij}\exp(\theta_{1k}+\nu_{1i})}{1+\sum_{s=1}^{3}\exp(\theta_{1s}+\nu_{1i})}\Big)^{\omega_{ijk}}\Big(\frac{p_{ij}}{1+\sum_{s=1}^{3}\exp(\theta_{1s}+\nu_{1i})}\Big)^{\omega_{ij4}} \\
&\quad \prod_{k=1}^{3}\Big(\frac{(1-p_{ij})\exp(\theta_{2k}+\nu_{2i})}{1+\sum_{s=1}^{3}\exp(\theta_{2s}+\nu_{2i})}\Big)^{y_{ijk}-\omega_{ijk}}\Big(\frac{1-p_{ij}}{1+\sum_{s=1}^{3}\exp(\theta_{2s}+\nu_{2i})}\Big)^{y_{ij4}-\omega_{ij4}}\Bigg] \\
&\quad \times \frac{1}{2\pi\delta_1\delta_2\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2(1-\rho^2)}\Big[\frac{\nu_{1i}^2}{\delta_1^2}-\frac{2\rho\nu_{1i}\nu_{2i}}{\delta_1\delta_2}+\frac{\nu_{2i}^2}{\delta_2^2}\Big]\Big)\Bigg\}\times\frac{1}{(1+\delta_1^2)^2}\frac{1}{(1+\delta_2^2)^2} \\
&\quad \times\prod_{s=1}^{2}\prod_{k=1}^{3}\frac{1}{\pi\sigma^*_{sk}[1+(\frac{\theta_{sk}-\theta^*_{sk}}{\sigma^*_{sk}})^2]} \\
&\propto \prod_{i=1}^{\ell}\Bigg\{\prod_{j=1}^{g_i}\Bigg[\frac{\exp\sum_{k=1}^{3}(\theta_{1k}+\nu_{1i})\omega_{ijk}}{[1+\sum_{s=1}^{3}\exp(\theta_{1s}+\nu_{1i})]^{\sum_{k=1}^{4}\omega_{ijk}}}\frac{\exp\sum_{k=1}^{3}(\theta_{2k}+\nu_{2i})(y_{ijk}-\omega_{ijk})}{[1+\sum_{s=1}^{3}\exp(\theta_{2s}+\nu_{2i})]^{\sum_{k=1}^{4}y_{ijk}-\omega_{ijk}}}\Bigg] \\
&\quad \times \frac{1}{2\pi\delta_1\delta_2\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2(1-\rho^2)}\Big[\frac{\nu_{1i}^2}{\delta_1^2}-\frac{2\rho\nu_{1i}\nu_{2i}}{\delta_1\delta_2}+\frac{\nu_{2i}^2}{\delta_2^2}\Big]\Big)\Bigg\}\times\frac{1}{(1+\delta_1^2)^2}\frac{1}{(1+\delta_2^2)^2} \\
&\quad \times\prod_{s=1}^{2}\prod_{k=1}^{3}\frac{1}{\pi\sigma^*_{sk}[1+(\frac{\theta_{sk}-\theta^*_{sk}}{\sigma^*_{sk}})^2]}.
\end{aligned}$$

Thus based on the joint posterior distribution above, we construct the conditional dis-

tribution of each parameter and execute a complete Gibbs sampler among the full parameter space. We denote the samples drawn from section 3.3.2 as $\{\underset{\sim}{\theta}_1^{(h)}, \underset{\sim}{\theta}_2^{(h)}, \delta_1^{2(h)}, \delta_2^{2(h)}, \rho^{(h)}\}, h = 1, \ldots, M$, with avg() and std() representing the process of taking the sample mean and the sample standard deviation.

Step 1 Start from some initial values of $\delta_1^{2(0)}, \delta_2^{2(0)}, \rho^{(0)}, \underset{\sim}{\nu}_1^{(0)}, \underset{\sim}{\nu}_2^{(0)}, \underset{\sim}{\omega}^{(0)}$;

Step 2 Draw $\underset{\sim}{\theta}_1^{(0)}$ from $\underset{\sim}{\theta}_1 \mid \underset{\sim}{\nu}_1^{(0)}, \underset{\sim}{\omega}^{(0)}, \underset{\sim}{y}$ using grid method, using the intervals

$$(avg(\underset{\sim}{\theta}_1^{(h)}) - 10 \cdot std(\underset{\sim}{\theta}_1^{(h)}), avg(\underset{\sim}{\theta}_1^{(h)}) + 10 \cdot std(\underset{\sim}{\theta}_1^{(h)}));$$

draw $\underset{\sim}{\theta}_2^{(0)}$ from $\underset{\sim}{\theta}_2 \mid \underset{\sim}{\nu}_2^{(0)}, \underset{\sim}{\omega}^{(0)}, \underset{\sim}{y}$ using grid method, using the intervals

$$(avg(\underset{\sim}{\theta}_2^{(h)}) - 10 \cdot std(\underset{\sim}{\theta}_2^{(h)}), avg(\underset{\sim}{\theta}_2^{(h)}) + 10 \cdot std(\underset{\sim}{\theta}_2^{(h)})).$$

Step 3 Draw $\tau_1^{(1)} = 1/(1+\delta_1^{2(1)}), \tau_2^{(1)} = 1/(1+\delta_2^{2(1)}), \rho^{(1)}$ from $\tau_1, \tau_2, \rho \mid \underset{\sim}{\theta}_1^{(0)}, \underset{\sim}{\theta}_2^{(0)}, \underset{\sim}{\nu}_1^{(0)}, \underset{\sim}{\nu}_2^{(0)}, \underset{\sim}{\omega}^{(0)}, \underset{\sim}{y}$ using the grid method with the intervals (0,1), (0,1), (-1,1) correspondingly.

Step 4 Draw $\underset{\sim}{\omega}^{(1)}$ from binomial distribution given $\underset{\sim}{\theta}_1^{(0)}, \underset{\sim}{\theta}_2^{(0)}, \underset{\sim}{\nu}_1^{(0)}, \underset{\sim}{\nu}_2^{(0)}, \underset{\sim}{y}$,

$$\omega_{ijk} \mid \pi_{i1k}, \pi_{i2k}, y_{ijk} \overset{ind}{\sim} \text{Binomial}\{y_{ijk}, \frac{p_{ij}\pi_{i1k}}{p_{ij}\pi_{i1k} + (1-p_{ij})\pi_{i2k}}\},$$
$$i = 1, \ldots, \ell,\ j = 1, \ldots, g_i, k = 1, 2, 3, 4$$

with

$$\pi_{i1k} = \frac{\exp(\theta_{1k} + \nu_{1i})}{1 + \sum_{t=1}^{3} \exp(\theta_{1t} + \nu_{1i})}, k = 1, 2, 3;\ \pi_{i14} = \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{1t} + \nu_{1i})},$$
$$\pi_{i2k} = \frac{\exp(\theta_{2k} + \nu_{2i})}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}, k = 1, 2, 3;\ \pi_{i24} = \frac{1}{1 + \sum_{t=1}^{3} \exp(\theta_{2t} + \nu_{2i})}.$$

Step 5 We begin with a simplification step for the prior of $\underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2$. Given the multivariate

normal distribution

$$\begin{pmatrix} \underset{\sim}{\nu}_1 \\ \underset{\sim}{\nu}_2 \end{pmatrix} \sim \text{Normal}\left\{ \begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix}, \begin{pmatrix} \delta_1^2 I_{\ell\times\ell} & \rho\delta_1\delta_2 I_{\ell\times\ell} \\ \rho\delta_1\delta_2 I_{\ell\times\ell} & \delta_2^2 I_{\ell\times\ell} \end{pmatrix} \right\},$$

by the property of the multivariate normal distribution, the conditional distribution of the multivariate normal still follows normal distribution, so we draw ν_{1i} and ν_{2i} simultaneously as

$$\begin{aligned} \nu_{1i} \sim & \text{Normal}\{0, \delta_1^2\} \\ \nu_{2i} \mid \nu_{1i} \sim & \text{Normal}\{\frac{\rho\delta_2}{\delta_1}\nu_{1i}, \delta_2\sqrt{1-\rho^2}\}, i = 1, \ldots, \ell. \end{aligned}$$

so that we can draw ν_{1i} independently from ν_{2i}. The conditional posterior distribution of ν_{1i} is

$$\prod_{j=1}^{g_i} \left[\frac{\exp\sum_{k=1}^{3}(\theta_{1k}+\nu_{1i})\omega_{ijk}}{[1+\sum_{s=1}^{3}\exp(\theta_{1s}+\nu_{1i})]^{\sum_{k=1}^{4}\omega_{ijk}}} \right] \left[-\frac{1}{\sqrt{2\pi\delta_1^2}}\exp(-\frac{\nu_{1i}^2}{2\delta_1^2}) \right],$$

then we can draw $\underset{\sim}{\nu}_1^{(1)}$ from the conditional distribution $\underset{\sim}{\nu}_1 \mid \delta_1^{2(1)}, \underset{\sim}{\theta}_1^{(1)}, \underset{\sim}{\omega}^{(1)}, \underset{\sim}{y}$ using grid method with interval

$$(\underset{\sim}{\nu}_1^* - 10 \cdot avg(\delta_1^{(h)}), \underset{\sim}{\nu}_1^* + 10 \cdot avg(\delta_1^{(h)})),$$

where $\underset{\sim}{\nu}_1^*$ is the quasi-mode, and $\delta_1^{(h)}$ is the square root of $\delta_1^{2(h)}$.

Next $\underset{\sim}{\nu}_2 = \{\nu_{2i}, i = 1, \ldots, \ell\}$ is drawn from the conditional posterior distribution

$$\prod_{j=1}^{g_i} \left[\frac{\exp\sum_{k=1}^{3}(\theta_{2k}+\nu_{2i})(y_{ijk}-\omega_{ijk})}{[1+\sum_{s=1}^{3}\exp(\theta_{2s}+\nu_{2i})]^{\sum_{k=1}^{4}y_{ijk}-\omega_{ijk}}} \right] \left[-\frac{1}{\sqrt{2\pi(1-\rho^2)\delta_2^2}}\exp(-\frac{(\nu_{2i}-\frac{\rho\delta_2}{\delta_1}\nu_{1i})^2}{2\delta_1^2(1-\rho^2)}) \right],$$

given $\underset{\sim}{\nu}_1^{(1)}, \underset{\sim}{\theta}_2^{(1)}, \delta_2^{2(1)}, \rho^{(1)}, \underset{\sim}{\omega}^{(1)}, \underset{\sim}{y}$, with grid interval for $\nu_{2i}^{(1)}$ constructed as

$$\left(\frac{\rho^{(1)}\delta_2^{(1)}}{\delta_1^{(1)}}\nu_{1i}^{(1)} - 10 \cdot \delta_2^{(1)}\sqrt{1-(\rho^{(1)})^2},\ \frac{\rho^{(1)}\delta_2^{(1)}}{\delta_1^{(1)}}\nu_{1i}^{(1)} + 10 \cdot \delta_2^{(1)}\sqrt{1-(\rho^{(1)})^2} \right).$$

Once we get the $\underset{\sim}{\nu}_2^{(1)}$, we complete one circle of the full Gibbs sampler and get back to the first step to continue.

3.5 Application to College Cheating and Comparisons

In this section, we provide a detailed 4-cell estimation results on the college cheating data.

In summary, the full Gibbs sampler on the generalized mixed effects model provides a similar posterior mean with the combined Bayesian model. Since the generalized mixed effects model have less parameters, the posterior standard deviation is smaller for the first three cells. However, the posterior mean are similar.

Also, all the approximation methods based on the generalized mixed effects model, with the different degree of the flexibility involved, will provide a reasonable estimation for the finite population proportion estimation for most areas, except for area 3, 4, 11 and 15, the estimation seems to be larger than Bayesian combined model and the generalized mixed effects model.

Here we first consider a goodness-of-fit statistic called conditional predictive ordinate (CPO). Larger values of CPO indicates better fit, see Geisser and Eddy (1979). Then for each area group (ij), a Monte Carlo approximation of CPO_{ij} is

$$\widehat{\text{CPO}}_{ij} = \left\{ \frac{1}{M}\sum_{h=1}^{M} \frac{1}{f(n_{ij} \mid \underset{\sim}{\pi}_{ij}^{(h)})} \right\}^{-1},\ j = 1, \ldots, n_j = 2;\ i = 1, \ldots, \ell,$$

where

$$\pi_{ijk}^{(h)} = p_{ijk}\pi_{i1k}^{(h)} + (1 - p_{ij})\pi_{i2k}^{(h)}, k = 1, \ldots, 4;\ h = 1, \ldots, M,$$

$\pi_{i1k}^{(h)}$ and $\pi_{i2k}^{(h)}$, $k = 1, \ldots, 4;\ h = 1, \ldots, M$, are the samples from the block Gibbs sampler in Chapter 2, with $f(n_{ij} \mid \underset{\sim}{\pi}_{ij})$ denoting the multinomial likelihood function. Correspondingly, from the re-parameterization equation (3.1.1), $\underset{\sim}{\pi}_{ij}^{(h)}$ for the generalized mixed effects model can be obtained by transforming the samples of $\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\nu}_1, \underset{\sim}{\nu}_2$,

$$\pi_{ijk}^{(h)} = p_{ijk}\pi_{i1k}(\underset{\sim}{\theta}_1^{(h)}, \underset{\sim}{\nu}_1^{(h)}) + (1 - p_{ij})\pi_{i2k}(\underset{\sim}{\theta}_2^{(h)}, \underset{\sim}{\nu}_2^{(h)}),\ h = 1, \ldots, M.$$

In fact, $\widehat{\text{CPO}}_{ij}$ the harmonic mean of the likelihoods $f(n_{ij} \mid \underset{\sim}{\pi}_{ij}^{(h)})$, $h = 1, \ldots, M$. A summary statistic of the CPO_{ij}, log pseudo marginal likelihood (LPML) is given by

$$\text{LPML} = \sum_{i}^{\ell} \sum_{j}^{n_i} \log(\widehat{\text{CPO}}_{ij}).$$

Also, larger value indicates a better fitting.

With respect to the cheating data, 18 out of the 30 CPOs are larger for the generalized mixed effects model than the Bayesian combined model in Chapter 2. LPML for the generalized mixed effects model is -137.6, which is larger than -140.9, indicating that the generalized mixed effects model fits the college cheating data better.

For the 325 observations from 15 areas, the multinomial-Dirichlet model in Chapter 2 takes about 20 minutes while the approximation methods M1, M2.A, M2.B will all take less than 1 minute; the full Gibbs sampler on the exact generalized mixed effects model M2.C also takes about 6 minutes even though the process will experience a long run to reach the convergence.

When the number of areas increased to 105 by simply concatenating 7 college cheating data samples into a single one, the computation took more than 1 hour in Chapter 2 while the generalized mixed effects model took less than 3 minutes. The computing times are calculated based on the FORTRAN codes, and R program took more than an

hour, even after improving the efficiency by Rcpp package. In this regard, it will be a huge difference with more areas involved.

Table 3.1 shows the result from the approximation model with independent $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$ (M1). We keep the latent variable fixed at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^*$, which comes from the EM algorithm in obtaining the quasi-modes. $\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2$ and $\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2$ can be drawn independently. After drawing $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\theta}_1$ are drawn from the posterior conditional distribution which are both normal, and then δ_1^2 can be drawn from the grid method. Finally we get 1000 set of samples of $(\underset{\sim}{\nu}_1, \underset{\sim}{\theta}_1, \delta_1^2)$ and $(\underset{\sim}{\nu}_2, \underset{\sim}{\theta}_2, \delta_2^2)$, where we can get our finite population proportion estimation. Also, these samples are saved to fed into the full Gibbs sampler in the complete generalized mixed effects model.

Table 3.1: Finite population proportion estimation for $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_1$ of the college cheating data using the approximation method with independent $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ (M1) in compare with the combined model.

	PM				PSD			
Area	π_{11}	π_{12}	π_{13}	π_{14}	π_{11}	π_{12}	π_{13}	π_{14}
a. Combined model								
1	0.099	0.729	0.065	0.106	0.063	0.107	0.067	0.082
2	0.120	0.717	0.060	0.103	0.078	0.119	0.063	0.084
3	0.149	0.677	0.044	0.131	0.077	0.114	0.045	0.098
4	0.142	0.643	0.049	0.166	0.083	0.128	0.053	0.119
5	0.109	0.739	0.052	0.100	0.070	0.114	0.055	0.081
6	0.149	0.690	0.060	0.102	0.087	0.116	0.062	0.080
7	0.105	0.725	0.046	0.124	0.067	0.117	0.051	0.100
8	0.137	0.676	0.072	0.115	0.076	0.120	0.071	0.090
9	0.174	0.670	0.050	0.107	0.097	0.119	0.052	0.083
10	0.130	0.720	0.049	0.101	0.074	0.106	0.053	0.082
11	0.139	0.637	0.072	0.152	0.084	0.124	0.074	0.111
12	0.175	0.632	0.068	0.125	0.099	0.120	0.064	0.096
13	0.104	0.694	0.068	0.135	0.071	0.125	0.066	0.101
14	0.145	0.698	0.050	0.107	0.079	0.110	0.050	0.080
15	0.143	0.659	0.057	0.141	0.070	0.116	0.055	0.099
b. Approximation method M1								
1	0.129	0.727	0.087	0.057	0.028	0.056	0.028	0.057
2	0.125	0.705	0.084	0.087	0.030	0.085	0.029	0.099
3	0.071	0.402	0.047	0.479	0.024	0.101	0.020	0.129
4	0.057	0.319	0.038	0.586	0.023	0.109	0.019	0.139
5	0.122	0.685	0.082	0.111	0.032	0.104	0.029	0.127
6	0.126	0.707	0.084	0.083	0.031	0.085	0.029	0.100
7	0.127	0.712	0.085	0.076	0.029	0.068	0.029	0.077
8	0.126	0.711	0.085	0.079	0.029	0.084	0.028	0.099
9	0.126	0.711	0.084	0.079	0.029	0.081	0.029	0.095
10	0.128	0.722	0.086	0.064	0.029	0.069	0.028	0.077
11	0.050	0.278	0.033	0.639	0.028	0.138	0.020	0.177
12	0.114	0.640	0.076	0.170	0.028	0.085	0.025	0.100
13	0.112	0.629	0.075	0.185	0.028	0.087	0.026	0.107
14	0.130	0.736	0.088	0.046	0.027	0.056	0.028	0.055
15	0.079	0.444	0.053	0.424	0.021	0.078	0.019	0.098

Figure 3.1: The posterior density plot of $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, δ_1^2, δ_2^2 and ρ

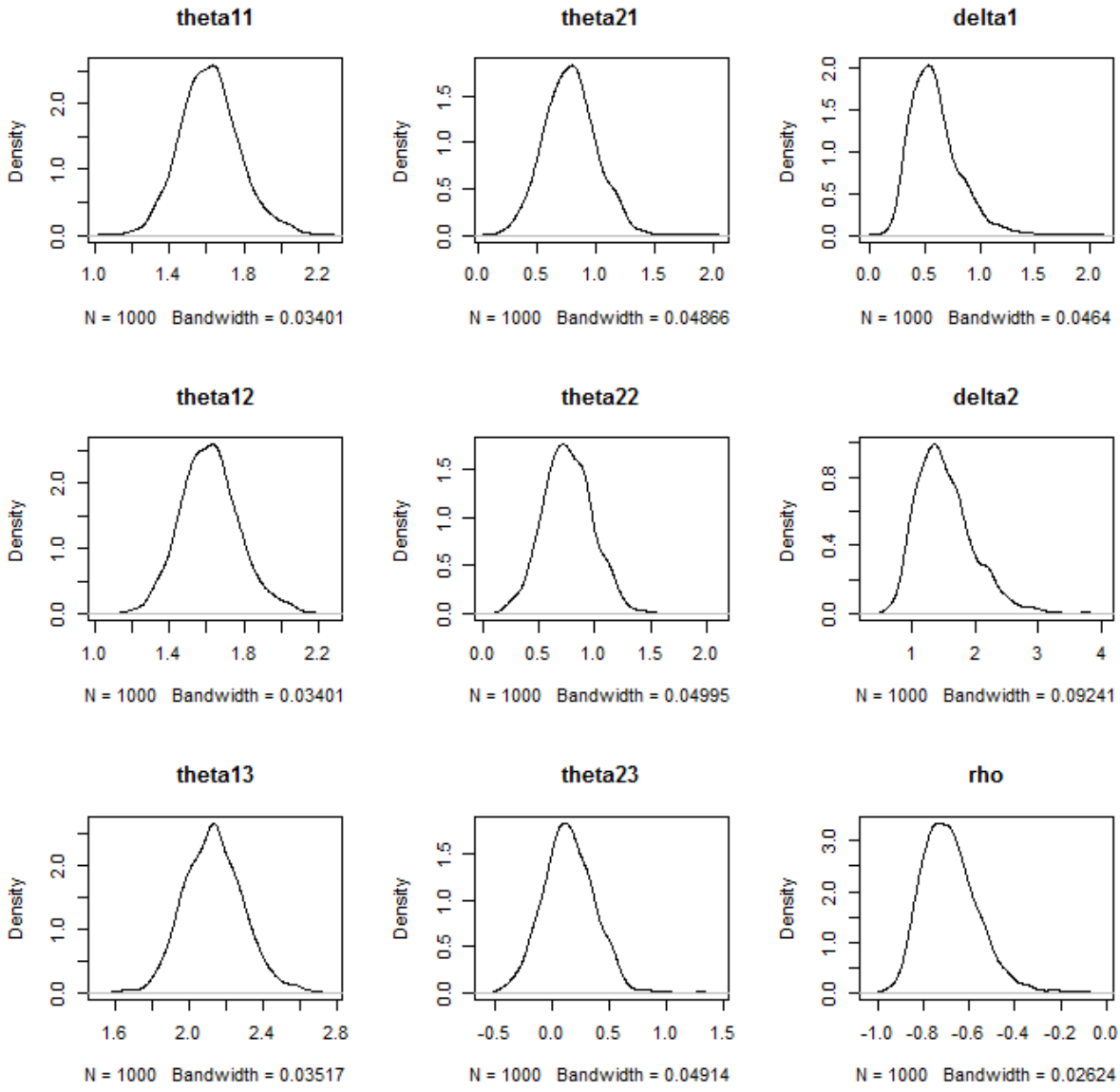


Table 3.2 shows the case when $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$ are correlated (M2.A). With the latent variable $\underset{\sim}{\omega}$ still fixed at $\underset{\sim}{\omega}^*$, those 1000 samples of (δ_1^2, δ_1^2) are fed into the model to draw the correlation parameter ρ, and in the end we get 1000 sets of samples of $(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \delta_1^2, \delta_2^2, \rho)$ to get the proportion estimation. Figure 3.1 shows the posterior density plot of those samples.

Table 3.2: Finite population proportion estimation for $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ of the college cheating data using the approximation method with correlated $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ (M2.A) in compare with the combined model.

	PM				PSD			
Area	π_{11}	π_{12}	π_{13}	π_{14}	π_{11}	π_{12}	π_{13}	π_{14}
a. Combined model								
1	0.099	0.729	0.065	0.106	0.063	0.107	0.067	0.082
2	0.120	0.717	0.060	0.103	0.078	0.119	0.063	0.084
3	0.149	0.677	0.044	0.131	0.077	0.114	0.045	0.098
4	0.142	0.643	0.049	0.166	0.083	0.128	0.053	0.119
5	0.109	0.739	0.052	0.100	0.070	0.114	0.055	0.081
6	0.149	0.690	0.060	0.102	0.087	0.116	0.062	0.080
7	0.105	0.725	0.046	0.124	0.067	0.117	0.051	0.100
8	0.137	0.676	0.072	0.115	0.076	0.120	0.071	0.090
9	0.174	0.670	0.050	0.107	0.097	0.119	0.052	0.083
10	0.130	0.720	0.049	0.101	0.074	0.106	0.053	0.082
11	0.139	0.637	0.072	0.152	0.084	0.124	0.074	0.111
12	0.175	0.632	0.068	0.125	0.099	0.120	0.064	0.096
13	0.104	0.694	0.068	0.135	0.071	0.125	0.066	0.101
14	0.145	0.698	0.050	0.107	0.079	0.110	0.050	0.080
15	0.143	0.659	0.057	0.141	0.070	0.116	0.055	0.099
b. Approximation method M2.A								
1	0.129	0.729	0.088	0.055	0.026	0.052	0.029	0.051
2	0.122	0.690	0.083	0.105	0.030	0.099	0.030	0.119
3	0.066	0.371	0.045	0.518	0.022	0.100	0.020	0.128
4	0.058	0.330	0.040	0.572	0.023	0.110	0.020	0.142
5	0.118	0.671	0.081	0.130	0.032	0.109	0.030	0.135
6	0.122	0.690	0.084	0.105	0.029	0.095	0.029	0.115
7	0.130	0.734	0.089	0.047	0.027	0.057	0.030	0.057
8	0.129	0.730	0.088	0.053	0.027	0.061	0.029	0.064
9	0.127	0.718	0.087	0.069	0.028	0.068	0.028	0.074
10	0.128	0.725	0.088	0.060	0.026	0.060	0.028	0.061
11	0.047	0.265	0.032	0.656	0.026	0.136	0.020	0.175
12	0.114	0.643	0.078	0.166	0.026	0.077	0.026	0.089
13	0.116	0.656	0.080	0.148	0.026	0.079	0.028	0.092
14	0.131	0.742	0.090	0.037	0.026	0.046	0.029	0.038
15	0.078	0.441	0.054	0.428	0.019	0.079	0.020	0.099

Table 3.3 (M2.B) shows the estimation result when the latent variable $\underset{\sim}{\omega}$ are drawn from the binomial distributions fixed at the quasi-mode of $\underset{\sim}{\nu}^*$. All the approximation methods so far will provide a reasonable finite population proportion estimation compared to the combined model in Chapter 2.

Table 3.3: Finite population proportion estimation for $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ of the college cheating data using the third approximation method with correlated $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ and flexible $\boldsymbol{\omega}$ (M2.B) in compare with the combined model.

	PM				PSD			
Area	π_{11}	π_{12}	π_{13}	π_{14}	π_{11}	π_{12}	π_{13}	π_{14}
a. Combined model								
1	0.099	0.729	0.065	0.106	0.063	0.107	0.067	0.082
2	0.120	0.717	0.060	0.103	0.078	0.119	0.063	0.084
3	0.149	0.677	0.044	0.131	0.077	0.114	0.045	0.098
4	0.142	0.643	0.049	0.166	0.083	0.128	0.053	0.119
5	0.109	0.739	0.052	0.100	0.070	0.114	0.055	0.081
6	0.149	0.690	0.060	0.102	0.087	0.116	0.062	0.080
7	0.105	0.725	0.046	0.124	0.067	0.117	0.051	0.100
8	0.137	0.676	0.072	0.115	0.076	0.120	0.071	0.090
9	0.174	0.670	0.050	0.107	0.097	0.119	0.052	0.083
10	0.130	0.720	0.049	0.101	0.074	0.106	0.053	0.082
11	0.139	0.637	0.072	0.152	0.084	0.124	0.074	0.111
12	0.175	0.632	0.068	0.125	0.099	0.120	0.064	0.096
13	0.104	0.694	0.068	0.135	0.071	0.125	0.066	0.101
14	0.145	0.698	0.050	0.107	0.079	0.110	0.050	0.080
15	0.143	0.659	0.057	0.141	0.070	0.116	0.055	0.099
b. Approximation method M2.B								
1	0.122	0.693	0.084	0.101	0.029	0.094	0.029	0.113
2	0.118	0.668	0.081	0.133	0.032	0.125	0.030	0.155
3	0.079	0.450	0.054	0.416	0.026	0.122	0.023	0.155
4	0.064	0.364	0.044	0.527	0.027	0.133	0.023	0.172
5	0.118	0.671	0.082	0.130	0.031	0.116	0.030	0.143
6	0.116	0.657	0.080	0.148	0.032	0.127	0.030	0.159
7	0.124	0.701	0.085	0.091	0.030	0.101	0.031	0.123
8	0.123	0.697	0.085	0.095	0.030	0.103	0.030	0.126
9	0.122	0.691	0.084	0.103	0.030	0.104	0.030	0.127
10	0.123	0.698	0.085	0.094	0.029	0.092	0.029	0.111
11	0.049	0.279	0.034	0.637	0.029	0.152	0.023	0.196
12	0.113	0.635	0.077	0.175	0.029	0.100	0.028	0.125
13	0.115	0.648	0.079	0.158	0.029	0.098	0.028	0.122
14	0.126	0.714	0.087	0.074	0.028	0.087	0.029	0.102
15	0.088	0.500	0.061	0.351	0.025	0.108	0.024	0.136

In the end, Table 3.4 provides the estimation result from the full Gibbs sampler for the exact posterior function model of the generalized mixed effects model assisted with the intervals provided by the approximation in M2.A. The finite population mean estimations for $\underset{\sim}{\pi}_1$ from both models are similar, the standard deviation estimation for π_{11}, π_{12} and π_{13} using the generalized mixed effects model are smaller than the combined model in Chapter 2. Meanwhile the standard deviation estimation for π_{14} tend to be larger since $\underset{\sim}{\pi}_1$ are added up to a fixed value 1. Table 3.5 also gives the corresponding 95% HPD interval estimation, illustrated by Figure 3.2 to Figure 3.5. In comparison with the combined model, the generalized mixed effects model gives shorter estimation intervals for π_{11}, π_{12} and π_{13}. Especially for π_{12}, the generalized mixed effects model provide informative smaller interval estimations for section 4, 11 and 13, which are consistent with the counts data we collected. As a result in Figure 3.5, the intervals are larger for π_{14} at those sections. For other sections, even though the generalized mixed effects model are slightly wider, they are still quite close with the intervals estimation given by the combined model. We are able to draw the similar conclusion for interval estimations for ϕ_{11} and ϕ_{12}.

We run 25,000 iterations burning the first 5,000 and gap every 20th to get a converged sample of 1,000 with the Geweke test and the effective sample size indicating the convergence. The computing time is 22.52 seconds compared with the 20 minutes for the Bayesian combined model. However, without taking advantage of the approximate intervals and the Cauchy prior, the full Gibbs sampler will need 210,000 iterates, with 10000 burn-in and taking every 200th sample to get a satisfactory convergence diagnostics.

In fact, the generalized mixed effects model is more parsimonious with much fewer parameters, which result in less variability. Meanwhile, it allows more flexible correlation than the multinomial-Dirichlet model. Besides, it is very convenient for adding covariates over the exponential part. Most important advantage is that we can achieve the fast computing through the generalized mixed effects model and its approximation.

Table 3.4: Finite population proportion estimation for $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ of the college cheating data using a full Gibbs sampler generalized mixed effects model (M2.C) in compare with the combined model.

	PM				PSD			
Area	π_{11}	π_{12}	π_{13}	π_{14}	π_{11}	π_{12}	π_{13}	π_{14}
a. Combined model								
1	0.099	0.729	0.065	0.106	0.063	0.107	0.067	0.082
2	0.120	0.717	0.060	0.103	0.078	0.119	0.063	0.084
3	0.149	0.677	0.044	0.131	0.077	0.114	0.045	0.098
4	0.142	0.643	0.049	0.166	0.083	0.128	0.053	0.119
5	0.109	0.739	0.052	0.100	0.070	0.114	0.055	0.081
6	0.149	0.690	0.060	0.102	0.087	0.116	0.062	0.080
7	0.105	0.725	0.046	0.124	0.067	0.117	0.051	0.100
8	0.137	0.676	0.072	0.115	0.076	0.120	0.071	0.090
9	0.174	0.670	0.050	0.107	0.097	0.119	0.052	0.083
10	0.130	0.720	0.049	0.101	0.074	0.106	0.053	0.082
11	0.139	0.637	0.072	0.152	0.084	0.124	0.074	0.111
12	0.175	0.632	0.068	0.125	0.099	0.120	0.064	0.096
13	0.104	0.694	0.068	0.135	0.071	0.125	0.066	0.101
14	0.145	0.698	0.050	0.107	0.079	0.110	0.050	0.080
15	0.143	0.659	0.057	0.141	0.070	0.116	0.055	0.099
b. Generalized mixed effects model (M2.C)								
1	0.128	0.676	0.076	0.120	0.034	0.075	0.025	0.082
2	0.129	0.680	0.077	0.114	0.033	0.077	0.025	0.082
3	0.119	0.626	0.071	0.185	0.033	0.096	0.024	0.113
4	0.094	0.495	0.056	0.356	0.034	0.130	0.023	0.164
5	0.130	0.685	0.077	0.108	0.034	0.079	0.025	0.085
6	0.129	0.683	0.077	0.111	0.034	0.078	0.025	0.085
7	0.119	0.632	0.071	0.177	0.033	0.102	0.025	0.120
8	0.123	0.650	0.073	0.154	0.034	0.089	0.025	0.103
9	0.126	0.667	0.075	0.132	0.034	0.082	0.025	0.091
10	0.129	0.681	0.077	0.113	0.034	0.074	0.025	0.078
11	0.101	0.532	0.060	0.307	0.036	0.139	0.025	0.175
12	0.122	0.640	0.072	0.166	0.034	0.088	0.025	0.105
13	0.112	0.596	0.067	0.225	0.033	0.117	0.024	0.139
14	0.128	0.674	0.076	0.122	0.033	0.074	0.025	0.079
15	0.113	0.598	0.067	0.222	0.032	0.097	0.024	0.115

Table 3.5: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model.

Section	Model	π_{11}	π_{12}	π_{13}	π_{14}
1	BC	(0, 0.219)	(0.543, 0.961)	(0, 0.196)	(0, 0.262)
	GME	(0.059, 0.189)	(0.536, 0.820)	(0.033, 0.130)	(0.008, 0.285)
2	BC	(0, 0.260)	(0.469, 0.919)	(0, 0.187)	(0, 0.260)
	GME	(0.060, 0.188)	(0.526, 0.822)	(0.030, 0.126)	(0.010, 0.286)
3	BC	(0, 0.293)	(0.461, 0.895)	(0, 0.132)	(0, 0.317)
	GME	(0.051, 0.177)	(0.429, 0.774)	(0.028, 0.121)	(0.013, 0.397)
4	BC	(0, 0.295)	(0.403, 0.894)	(0, 0.153)	(0, 0.385)
	GME	(0.032, 0.158)	(0.217, 0.713)	(0.016, 0.099)	(0.080, 0.693)
5	BC	(0, 0.237)	(0.502, 0.920)	(0, 0.167)	(0, 0.259)
	GME	(0.066, 0.196)	(0.535, 0.839)	(0.033, 0.131)	(0.009, 0.298)
6	BC	(0, 0.306)	(0.466, 0.924)	(0, 0.185)	(0, 0.252)
	GME	(0.063, 0.195)	(0.517, 0.816)	(0.032, 0.129)	(0.007, 0.287)
7	BC	(0, 0.235)	(0.516, 0.945)	(0, 0.149)	(0, 0.306)
	GME	(0.059, 0.184)	(0.422, 0.812)	(0.025, 0.123)	(0.009, 0.424)
8	BC	(0, 0.273)	(0.439, 0.906)	(0, 0.212)	(0, 0.279)
	GME	(0.062, 0.192)	(0.474, 0.817)	(0.027, 0.125)	(0.007, 0.354)
9	BC	(0, 0.344)	(0.405, 0.871)	(0, 0.153)	(0, 0.262)
	GME	(0.060, 0.190)	(0.497, 0.816)	(0.032, 0.128)	(0.009, 0.320)
10	BC	(0, 0.262)	(0.515, 0.910)	(0, 0.161)	(0, 0.251)
	GME	(0.063, 0.193)	(0.544, 0.814)	(0.031, 0.128)	(0.007, 0.258)
11	BC	(0, 0.285)	(0.385, 0.864)	(0, 0.220)	(0, 0.360)
	GME	(0.038, 0.173)	(0.231, 0.751)	(0.017, 0.109)	(0.029, 0.657)
12	BC	(0, 0.350)	(0.407, 0.870)	(0, 0.191)	(0, 0.313)
	GME	(0.058, 0.190)	(0.459, 0.796)	(0.028, 0.121)	(0.006, 0.371)
13	BC	(0, 0.245)	(0.465, 0.944)	(0, 0.197)	(0, 0.322)
	GME	(0.053, 0.185)	(0.364, 0.801)	(0.022, 0.114)	(0.015, 0.499)
14	BC	(0, 0.288)	(0.466, 0.892)	(0, 0.156)	(0, 0.254)
	GME	(0.062, 0.191)	(0.529, 0.810)	(0.029, 0.125)	(0.014, 0.290)
15	BC	(0, 0.264)	(0.426, 0.865)	(0, 0.162)	(0, 0.322)
	GME	(0.052, 0.173)	(0.396, 0.768)	(0.023, 0.115)	(0.022, 0.441)

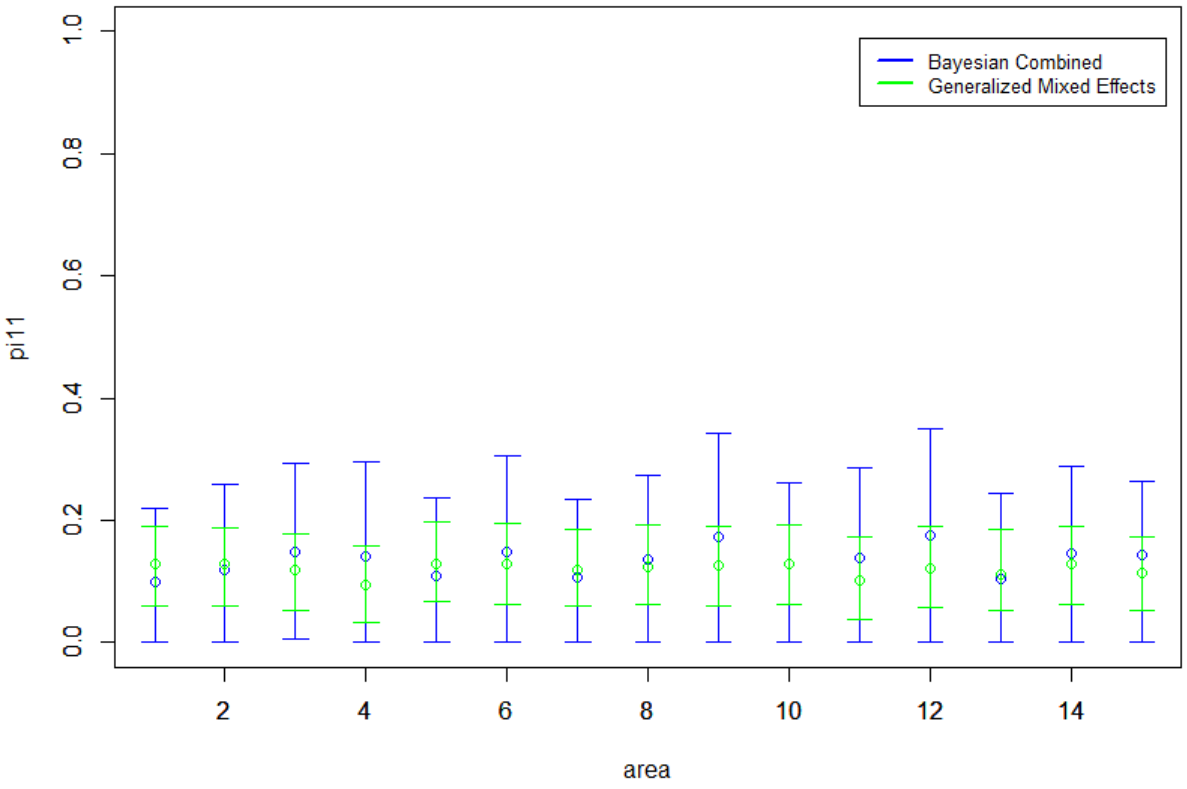


Figure 3.2: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model of π_{11}.

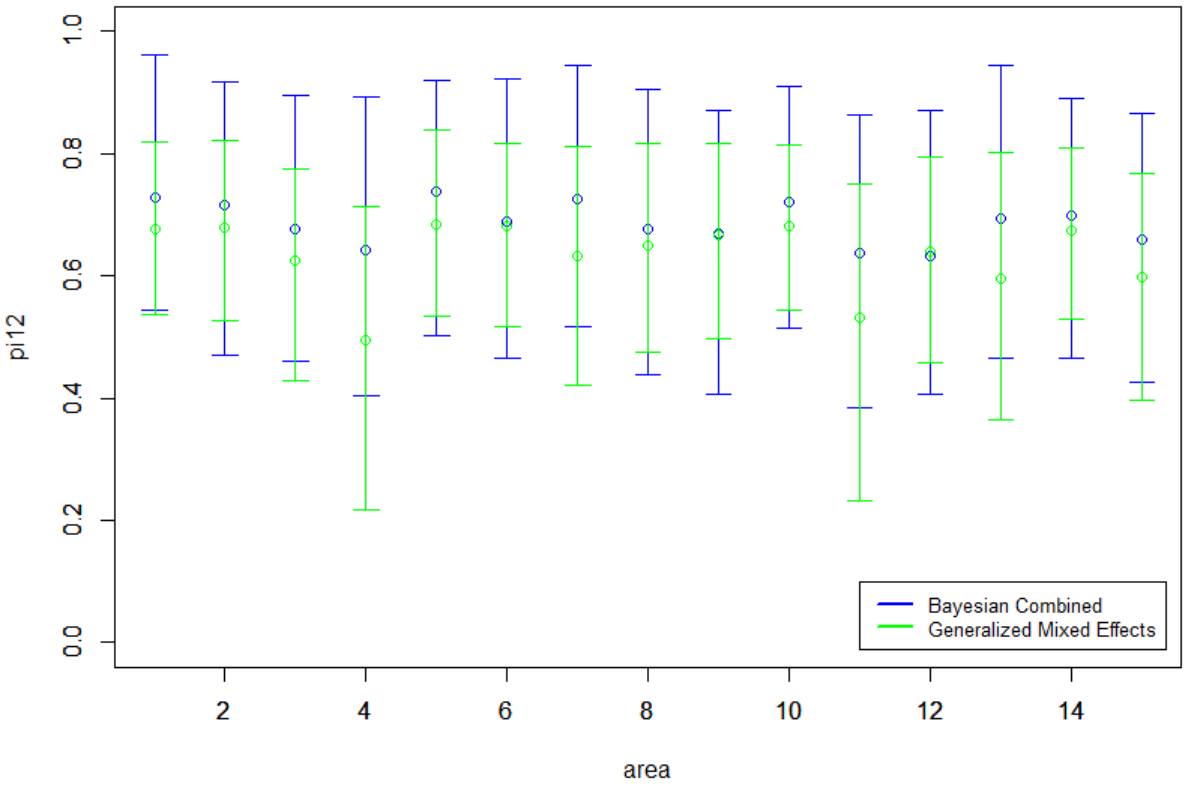


Figure 3.3: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model of π_{12}.

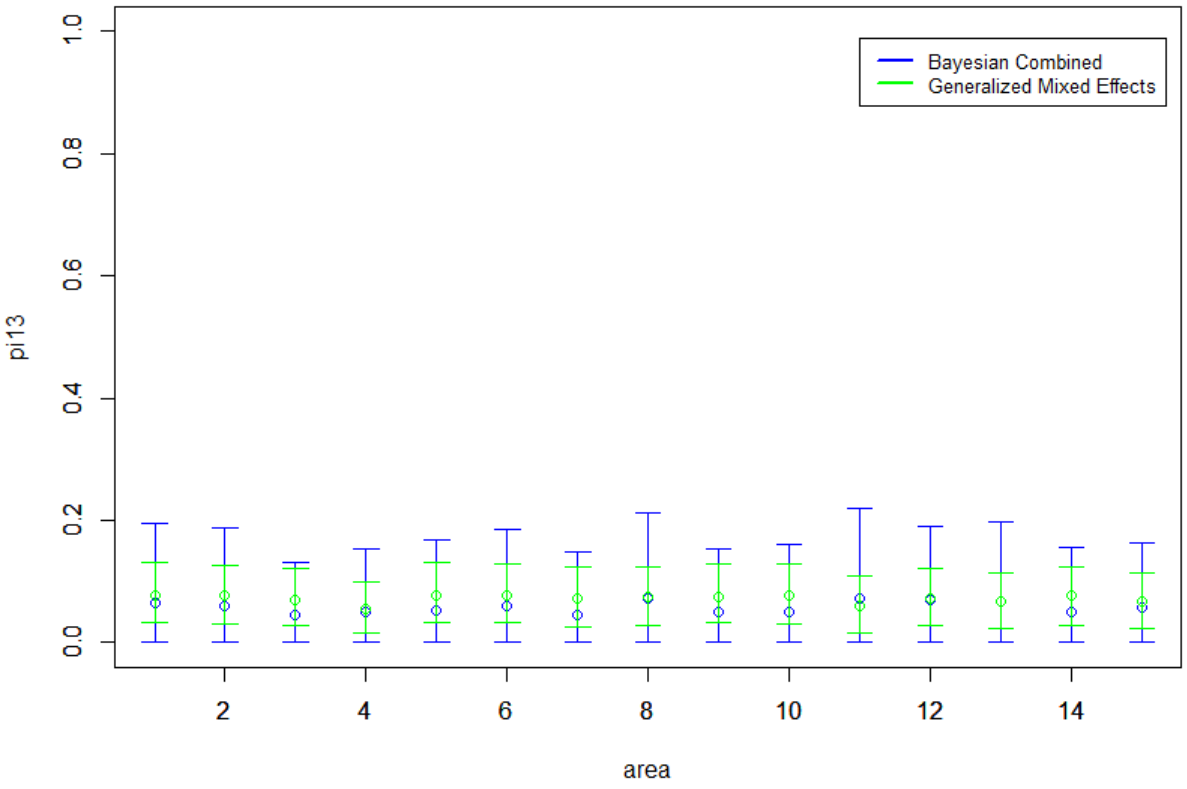


Figure 3.4: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model of π_{13}.

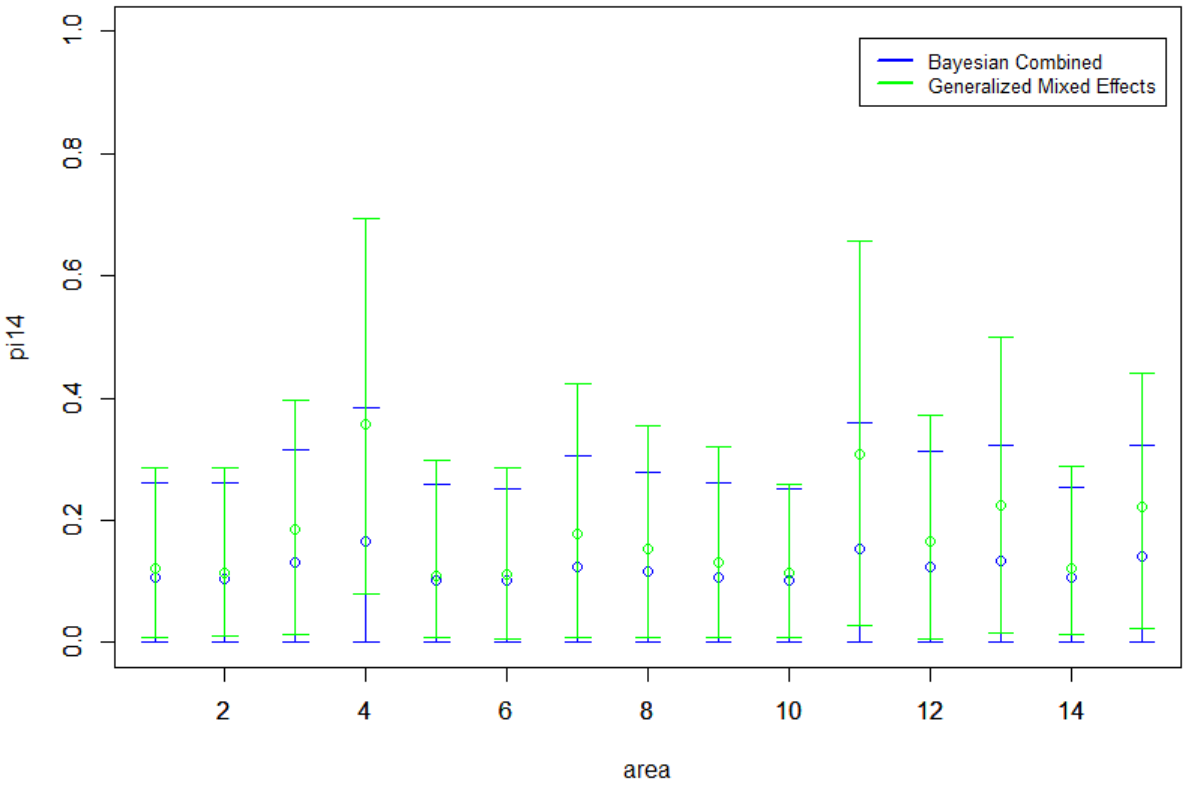


Figure 3.5: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model of π_{14}.

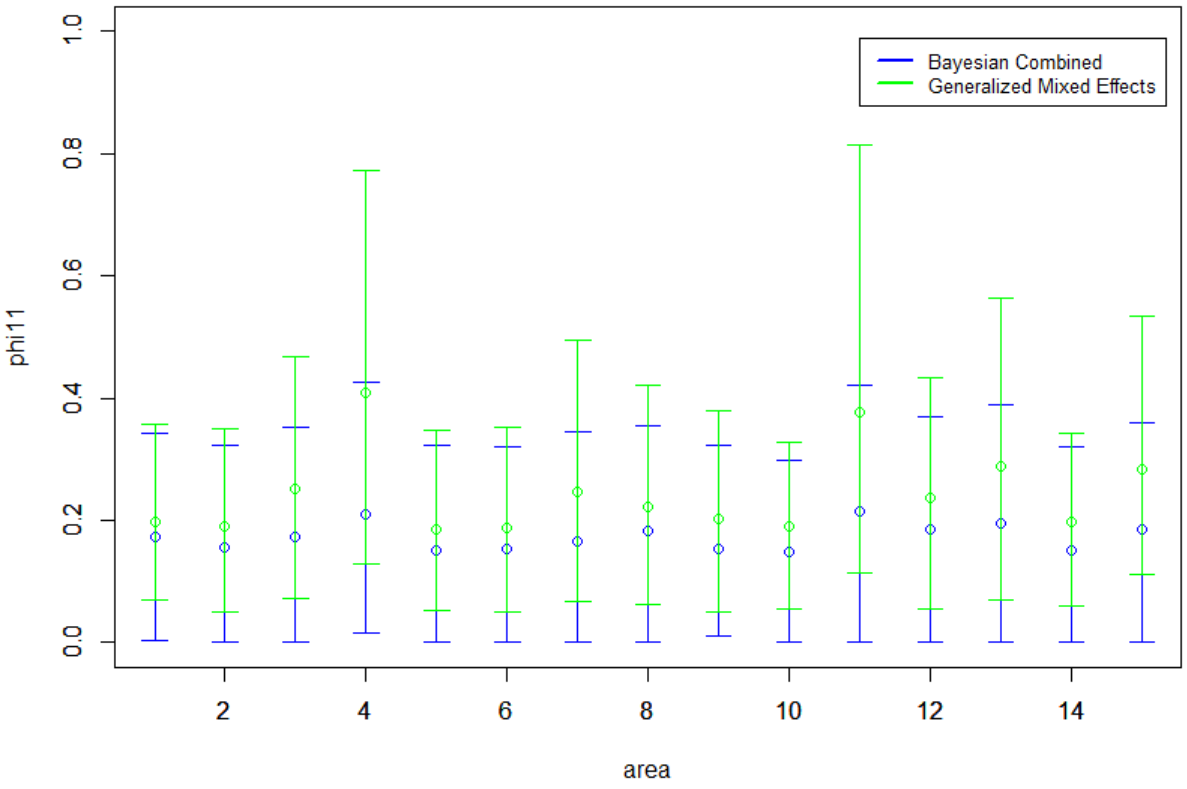


Figure 3.6: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model of ϕ_{11}.

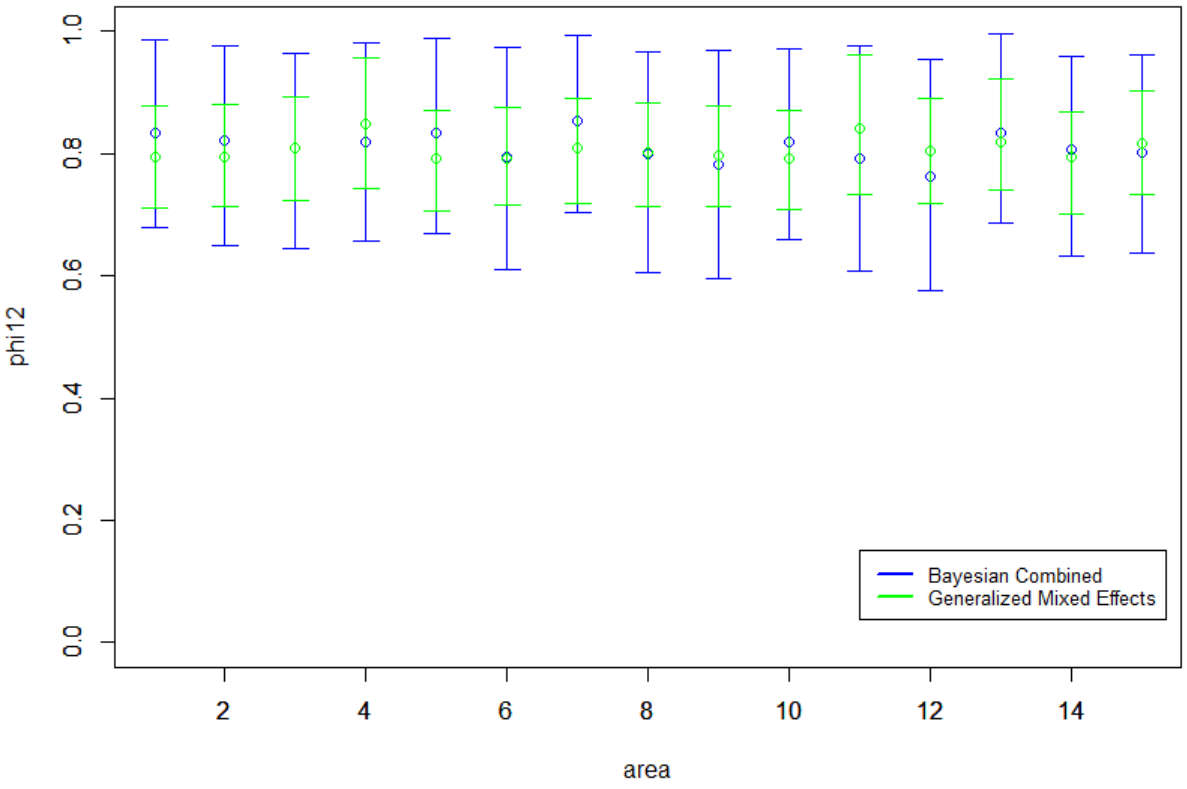


Figure 3.7: Comparison of 95% HPD intervals of the Bayesian combined model and the full Gibbs sampling of the generalized mixed effects model of ϕ_{12}.

Appendix

3.A Quasi-modes

We intend to get the quasi-modes of $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\nu}_1$ through the EM algorithm. We consider the likelihood function,

$$\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\nu}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\nu}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \propto \prod_{i=1}^{\ell} \left[\prod_{j=1}^{g_i} \prod_{k=1}^{4} (p_{ij}\pi_{i1k})^{\omega_{ijk}} ((1-p_{ij})\pi_{i2k}))^{y_{ijk}-\omega_{ijk}} \right],$$

where

$$\pi_{i1k} = \frac{\exp(\theta_{1k}+\nu_{1i})}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})}, k=1,2,3;\ \pi_{i14} = \frac{1}{1+\sum_{t=1}^{3}\exp(\theta_{1t}+\nu_{1i})},$$

$$\pi_{i2k} = \frac{\exp(\theta_{2k}+\nu_{2i})}{1+\sum_{t=1}^{3}\exp(\theta_{2t}+\nu_{2i})}, k=1,2,3;\ \pi_{i24} = \frac{1}{1+\sum_{t=1}^{3}\exp(\theta_{2t}+\nu_{2i})}.$$

It is worth noting that the reparameterization of $\underset{\sim}{\pi}_{i1}$ and $\underset{\sim}{\pi}_{i2}$ allows for a one-to-one mapping, so for the simplicity of computation we only need to apply the EM algorithm to $\underset{\sim}{\pi}_{i1}$ and $\underset{\sim}{\pi}_{i2}$ and transform back to $\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2$ later. Here we use $(\theta_{1k}+\nu_{1i}), k=1,2,3, for\ i=1,\ldots,\ell$, to represent mean effect combining all the areas θ_{1k}, incorporating with the area effect ν_{1i}.

i) Get the MLE of global effect $\widehat{\theta_{1k}+\nu_{1i}}, \widehat{\theta_{2k}+\nu_{2i}},\ i=1,\ldots,\ell$.

Start with initial value of $\underset{\sim}{\pi}_{i1}^{(0)}$ and $\underset{\sim}{\pi}_{i2}^{(0)}$, we can draw the latent variables ω_{ijk} independently

from Binomial distribution

$$\omega_{ijk} \mid \underset{\sim}{\pi}_{i1}, \underset{\sim}{\pi}_{i2}, \underset{\sim}{y} \overset{ind}{\sim} \text{Binomial}\{y_{ijk}, \frac{p_{ij}\pi_{i1k}}{p_{ij}\pi_{i1k} + (1-p_{ij})\pi_{i2k}}\}, i = 1, \ldots, l; j = 1, \ldots, g_i; k = 1, 2, 3, 4.$$

Step 1. Update $\underset{\sim}{\omega}$,

$$\omega_{ijk}^{(0)} = y_{ijk} \ \frac{p_{ij}\pi_{i1k}}{p_{ij}\pi_{i1k} + (1-p_{ij})\pi_{i2k}}.$$

Since

$$\underset{\sim}{\pi}_{i1} \mid \underset{\sim}{\omega} \overset{ind}{\sim} \text{Dirichlet}(\omega_{i1\cdot}, \omega_{i2\cdot}, \omega_{i3\cdot}, \omega_{i4\cdot}),$$

$$\underset{\sim}{\pi}_{i2} \mid \underset{\sim}{\omega}, \underset{\sim}{y} \overset{ind}{\sim} \text{Dirichlet}(y_{i1\cdot} - \omega_{i1\cdot}, y_{i2\cdot} - \omega_{i2\cdot}, y_{i3\cdot} - \omega_{i3\cdot}, y_{i4\cdot} - \omega_{i4\cdot}),$$

where $\omega_{i\cdot k} = \sum_{j=1}^{g_i} \omega_{ijk}, y_{i\cdot k} = \sum_{j=1}^{g_i} y_{ijk}, \ k = 1, 2, 3, 4.$

Step 2. Update $\underset{\sim}{\pi}_{i1}$ and $\underset{\sim}{\pi}_{i2}$,

$$\pi_{i1k}^{(1)} = \frac{\omega_{i\cdot k}^{(0)}}{\sum_{k=1}^{4} \omega_{i\cdot k}^{(0)}} \quad \text{and} \quad \pi_{i2k}^{(1)} = \frac{y_{i\cdot k} - \omega_{i\cdot k}^{(0)}}{\sum_{k=1}^{4} y_{i\cdot k} - \omega_{i\cdot k}^{(0)}}.$$

Step 3. Back to step 1 to get $\omega_{ijk}^{(1)}$ with $(\pi_{i1k}^{(1)}, \pi_{i2k}^{(1)})$ and continue with step 2 to get $\pi_{i1k}^{(2)}$ and $\pi_{i2k}^{(2)}$. Keep updating the parameters until convergence to $\hat{\pi}_{i1k}$ and $\hat{\pi}_{i2k}$, $i = 1 \ldots \ell, \ k = 1, 2, 3.$

Step 4. Transform back to get $\widehat{\theta_{1k} + \nu_{1i}}, \widehat{\theta_{2k} + \nu_{2i}}, \ i = 1, \ldots, \ell.$

$$\widehat{\theta_{1k} + \nu_{1i}} = \log \frac{\hat{\pi}_{i1k}}{1 - \sum_{t=1}^{3} \hat{\pi}_{i1t}}, \qquad \widehat{\theta_{2k} + \nu_{2i}} = \log \frac{\hat{\pi}_{i2k}}{1 - \sum_{t=1}^{3} \hat{\pi}_{i2t}}.$$

ii) Get the MLEs of global effects θ_{1k}^* and θ_{2k}^*.

For the mean effect, there is no area difference so we combine all the samples with the same random mechanism p_{ij} into 5 large groups. As a result, the likelihood function with only the global effect is

$$\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2 \mid \underset{\sim}{y}, \underset{\sim}{\omega}) \propto \prod_{j=1}^{5} \prod_{k=1}^{4} (p_j \pi_{1k})^{\omega_{\cdot jk}} ((1-p_j)\pi_{2k}))^{y_{\cdot jk} - \omega_{\cdot jk}}$$

where $\omega_{\cdot jk} = \sum_{i=1}^{\ell} \omega_{ijk}, y_{\cdot jk} = \sum_{i=1}^{\ell} y_{ijk},\ k = 1, 2, 3, 4.$

With the likelihood function above, we apply the EM algorithm in the same way. Start with initial value of $\underset{\sim}{\pi}_1^{(0)}$ and $\underset{\sim}{\pi}_2^{(0)}$, we can draw the latent variables $\omega_{\cdot jk}$ independently from Binomial distribution

$$\omega_{\cdot jk} \mid \underset{\sim}{\pi}_1, \underset{\sim}{\pi}_2, \underset{\sim}{y} \overset{ind}{\sim} \text{Binomial}\{y_{\cdot jk}, \frac{p_j \pi_{1k}}{p_j \pi_{1k} + (1 - p_j)\pi_{2k}}\}, j = 1, \ldots, 5; k = 1, 2, 3, 4.$$

Step 1. Update $\underset{\sim}{\omega} = (\omega_{\cdot jk}),\ j = 1, \ldots, 5; k = 1, \ldots, 4,$

$$\omega_{\cdot jk}^{(0)} = y_{\cdot jk}\ \frac{p_j \pi_{1k}}{p_j \pi_{1k} + (1 - p_j)\pi_{2k}}.$$

Since

$$\underset{\sim}{\pi}_1 \mid \underset{\sim}{\omega} \overset{ind}{\sim} \text{Dirichlet}(\omega_{\cdot 1\cdot}, \omega_{\cdot 2\cdot}, \omega_{\cdot 3\cdot}, \omega_{\cdot 4\cdot}),$$

$$\underset{\sim}{\pi}_2 \mid \underset{\sim}{\omega}, \underset{\sim}{y} \overset{ind}{\sim} \text{Dirichlet}(y_{\cdot 1\cdot} - \omega_{\cdot 1\cdot}, y_{\cdot 2\cdot} - \omega_{\cdot 2\cdot}, y_{\cdot 3\cdot} - \omega_{\cdot 3\cdot}, y_{\cdot 4\cdot} - \omega_{\cdot 4\cdot}),$$

where $\omega_{\cdot k\cdot} = \sum_{j=1}^{5} \omega_{\cdot jk}, y_{\cdot k\cdot} = \sum_{j=1}^{5} y_{\cdot jk},\ k = 1, 2, 3, 4.$

Step 2. Update $\underset{\sim}{\pi}_1$ and $\underset{\sim}{\pi}_2$,

$$\pi_{1k}^{(1)} = \frac{\omega_{\cdot k\cdot}^{(0)}}{\sum_{k=1}^{4} \omega_{\cdot k\cdot}^{(0)}} \quad \text{and} \quad \pi_{2k}^{(1)} = \frac{y_{i\cdot k} - \omega_{\cdot k\cdot}^{(0)}}{\sum_{k=1}^{4} y_{\cdot k\cdot} - \omega_{\cdot k\cdot}^{(0)}}.$$

Step 3. Back to step 1 to get $\omega_{\cdot jk}^{(1)}$ with $(\pi_{1k}^{(1)}, \pi_{2k}^{(1)})$ and continue with step 2 to get $\pi_{1k}^{(2)}$and $\pi_{2k}^{(2)}$. Keep updating the parameters until converge to π_{1k}^{*} and π_{2k}^{*}, $k = 1, 2, 3$.

Step 4. Transform back to get $\underset{\sim}{\theta}_1^*, \underset{\sim}{\theta}_2^*$.

$$\theta_{1k}^{*} = \log \frac{\hat{\pi}_{1k}}{1 - \sum_{t=1}^{3} \hat{\pi}_{1t}}, \qquad \theta_{2k}^{*} = \log \frac{\hat{\pi}_{2k}}{1 - \sum_{t=1}^{3} \hat{\pi}_{2t}}.$$

iii) Get $\underset{\sim}{\nu}_1^*, \underset{\sim}{\nu}_2^*$.

In consideration of the relationship between the global effect and the area effect, for each area, we can get the area effect ν_{1i}^{*} (ν_{2i}^{*}), $i = 1, \ldots, \ell$ by subtracting the mean effect θ_{1k}^{*}

(θ^*_{2k}) from the total effect $\widehat{\theta_{1k} + \nu_{1i}}$ $(\widehat{\theta_{2k} + \nu_{2i}})$,

$$\nu^*_{1i} = \sum_{k=1}^{3} (\widehat{\theta_{1k} + \nu_{1i}} - \theta^*_{1k})/3, \qquad \nu^*_{2i} = \sum_{k=1}^{3} (\widehat{\theta_{2k} + \nu_{2i}} - \theta^*_{2k})/3.$$

Chapter 4

Concluding Remarks and Future Work

4.1 Concluding Remarks

We provided a Bayesian method to estimate the finite population proportions for sensitive quantities through the unrelated question design when there are more than one sensitive question.

An application on the college cheating data also show that the combined model outperforms the individual area model and the separate question model in the terms of posterior standard deviation, coefficient of variation and correlation. Furthermore, the simulation study shows that for data from small areas, the Bayesian combined model (cb) gives a more accurate estimation, in terms of relative absolute bias and posterior root mean square error, compared to the individual area model (ind) and separate question model. In addition, we can gain strength by increasing the number of areas.

Although it seems that the variations in correlation rarely make any difference, the combining effect is significant. It could be the case that even if we are only interested in one sensitive question, we can include other sensitive questions into the design to get a better estimate, under the same degree of corporation of the respondents. Of course, the

same number of unrelated questions should be constructed since the masked responses should have the same dimension. Even though there might be a concern of extra cost by asking more questions, the availability of online survey tools will make the collection of data easier, but this might lead to nonprobalibity samples.

The generalized mixed effects model will provide a consistent finite population proportion estimation to the Bayesian combined model, with smaller posterior standard deviation for some cells, and fitted better. All kinds of INNA approach based on the generalized mixed effects model can also provide a similar estimation results for most of the areas compared to the combined model with at least 10 times faster computing process.

We have used proper prior for both models. This allows proper posterior densities. Improper posterior densities will make the MCMC suffer from slowly mixing, so that a large burn-in and huge thinning are needed. But proper prior do not guarantee a proper posterior.

Theorem Given $\pi^*(\theta) = g^*(\theta)\pi(\theta)$, the posterior $\pi^*(\theta)$ is proper if and only if $g^*(\theta) \leq A < \infty$ provided the prior $\pi(\theta)$ is proper.

Proof

Suppose $g^*(\theta) \leq A < \infty$, $\int \pi(\theta)g(\theta)d\theta \leq A \int \pi(\theta)d\theta = A < \infty$;

Suppose $\pi^*(\theta)$ is unbounded, then there exists c such that $\pi^*(\theta) \geq c$, so that $\int \pi(\theta)g(\theta)d\theta \geq c \int \pi(\theta)d\theta = c$. Thus when $c \to \infty$, the posterior density is improper.

In general, the essence of the study is borrowing information across small areas and multiple questions to make a better inference. In addition, we also proposed a generalized mixed effects model which is more flexible while allowing improved precision and fast computing.

4.2 Data Masking

We discuss how to provide public-used data from confidential data. This is a practical issue which is of concern for many government agencies, referred as statistical disclosure limitation.

The confidential data values are replaced with the predicted ones from a statistical model, to create a synthetic dataset (e.g. Rubin, 1993; Reiter 2003, 2005; Fienberg and Jin, 2009). Hu, Reiter and Wang (2018) recently present a Dirichlet process mixture model for nested categorical data, and further for use of generating masked public-used files for household data especially for confidential variables. The key point of data masking is that the masked data can be used to draw a similar conclusion as if the original data were analyzed. Here we propose our randomized response procedure to mimic a masking procedure, and compare the estimation results with those from Bayesian logistic regression model.

4.2.1 Data Description

We mimic the body mass index (BMI) and bone mineral density (BMD) data from the third National Health and Nutrition Examination Survey (NHANES III) to provide an example. Note the design is not implemented in reality, in fact we have their responses to the sensitive questions, and also the nonsensitive ones. However we can obtain the binary response assuming that every individual gives his/her response following the design. As a result, only part of the responses are taken, which constitute the masked data. Thus, this example can provide a masking procedure using our Bayesian unrelated question model.

Obesity is one of today's leading public health problems and it increases the risk of morbidity due to diseases such as diabetes and hypertension. The survey is a program of studies run by CDC (Center of Disease Control and Prevention) to assess the health and nutritional status of adults and children in the United States, and it was conducted during the period October 1988 through September 1994. This survey contains BMI and

BMD data together with covariates of age, race and sex, where BMI is measured by an individual's weight and height and BMD is measured using Dual X-ray Absoptionetry (DEXA).

The final data set for this study uses only 6557 samples of the 35 largest counties with a population at least 500,000. Due to confidentiality reasons, the original sensitive attributes BMI and BMD are transformed to categorical data based on the criteria defined by the World Health Organization (WHO). So that there are 4 levels of BMI (=1,2,3,4) and 3 levels of BMD (=1,2,3), where BMI = $(3, 4)$ represents the BMI value greater than 25 which can be considered as overweight; and BMD = $(2, 3)$ means the BMD value smaller than 0.82 which indicates osteopenia or osteoporosis. In our case, our interest is the proportion of people from the targeted population who are overweight or have osteoporosis, and both can be considered as sensitive. Though mechanisms by which body weight influences bone mineral density are still unknown, many studies have shown a strong association between bone mineral density and body mass index for large populations (e.g., Nandram, Kim and Zhou, 2019). Hence we believe that the overweight and the osteoporosis are two correlated attributes. Of course, we can apply the Bayesian RRT twice for each attribute, yet we are more interested in whether we can benefit from utilizing the correlated data.

As compared to our survey design, the two sensitive questions are set to be 'Are you overweight?' and 'Do you have osteoporosis?'. In order to simulate our design, the other two unrelated non-sensitive questions need to be constructed from the covariates. Thus race and sex are selected to be the non-sensitive question as 'Are you white? ' and 'Are you male?', which are unrelated with the sensitive ones, assuming that there is no sex effect on the BMD. The combined response for the sensitive questions could only be four types (No, No), (No, Yes), (Yes, No), (Yes, Yes), same for the unrelated questions.

In the masking procedure, all the samples from each county are divided to 2-5 groups. For each individual from i^{th} county and j^{th} group, we set $p_{ij} = (.25, .75, .2, .7, .3)$ according to the group size. Then the sensitive responses are selected with probability p_{ij}, as the

i	j		p_{ij}				$1-p_{ij}$			
			BMI & BMD				white & male			
			(0,0)	(0,1)	(1,0)	(1,1)	(0,0)	(0,1)	(1,0)	(1,1)
1st County	Group 1	Individual 1			1			1		
		Individual 2	1						1	
		Individual 3				1		1		
		⋮								
		Individual 54	1							1
	Group 2	Individual 1		1			1			
		Individual 2			1			1		
		Individual 3		1					1	
		⋮								
		Individual 54			1		1			
2nd County	Group 1	⋮								
	Group 2									
	Group 3									
⋮	⋮	⋮								
35th County	Group 1	⋮								
	Group 2									
	Group 3									

Figure 4.1: Masking procedure for the NHANES data

nonsensitive attributes are selected with $1-p_{ij}$. In other words, either the response from the sensitive attributes or those from the nonsensitive attributes can be selected for each individual. Figure 4.1 illustrates the masking procedure. As a result, the counts data that we gathered using the randomized procedure can be treated like the data masked for the original data. Afterwards, we pass in the data into our Bayesian combined model to get the inference of the finite population proportions.

4.2.2 Bayessian Logistic Regression Estimation

The intuition of utilizing the mimicking data is to evaluate if the Bayesian hierarchical model designed for multiple sensitive questions provides the reliable proportion estimation. The "true" sensitive proportion that we intend to compare with is obtained from logistic regression exact method on the full sample, involving more information about age, race and sex. We consider the following Bayesian logistic model of the binary response

with covariates age, race and sex. The binary variable y_{ij} is an overweight indicator.

$$
\begin{aligned}
y_{ij} \mid \nu_i, \underset{\sim}{\beta}_{(0)} &\overset{ind}{\sim} \text{Bernoulli}\{\frac{e^{\underset{\sim}{x}'_{ij}\underset{\sim}{\beta}_{(0)}+\nu_i}}{1+e^{\underset{\sim}{x}'_{ij}\underset{\sim}{\beta}_{(0)}+\nu_i}}\},\\
\nu_i \mid \beta_0, \delta^2 &\overset{iid}{\sim} \text{Normal}(\beta_0, \delta^2),\\
\pi(\underset{\sim}{\beta}, \delta^2) &\propto \frac{1}{(1+\delta^2)^2}, \qquad \delta^2 > 0, i = 1, \ldots, \ell, j = 1, \ldots, n_i.
\end{aligned}
$$

From the joint posterior density of the parameters $(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{y})$

$$
\begin{aligned}
\pi(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{y}) &\propto \pi(\underset{\sim}{y} \mid \underset{\sim}{\nu}, \underset{\sim}{\beta}_{(0)})\pi(\underset{\sim}{\nu} \mid \beta_{(0)}, \delta^2)\pi(\underset{\sim}{\beta}, \delta^2)\\
&\propto \prod_{i=1}^{\ell}\Big\{\Big[\prod_{j=1}^{n_i}\frac{e^{(\underset{\sim}{x}'_{ij}\underset{\sim}{\beta}_{(0)}+\nu_i)y_{ij}}}{1+e^{\underset{\sim}{x}'_{ij}\underset{\sim}{\beta}_{(0)}+\nu_i}}\Big]\Big[\frac{1}{\sqrt{2\pi\delta^2}}e^{-\frac{(\nu_i-\beta_0)^2}{2\delta^2}}\Big]\Big\}\\
&\times \frac{1}{(1+\delta^2)^2}\ .
\end{aligned}
$$

Next we can use a blocked Gibbs sampler based on the conditional distribution $\pi(\underset{\sim}{\nu} \mid \underset{\sim}{\beta}, \delta^2, \underset{\sim}{y})$ and $\pi(\underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{\nu}, \underset{\sim}{y})$.

To assist the Gibbs sampler in the logistic regression model, we use Bayesian Fay-Herriot model to get the grid range constructed by the posterior draws of $(\underset{\sim}{\beta}, \delta^2)$. Briefly, the Bayesian Fay-Herriot model is

$$
\begin{aligned}
\hat{\theta}_i &\sim \text{N}(\theta_i, \hat{\sigma}_i^2),\\
\theta_i &\sim \text{N}(\underset{\sim}{x}'_i\underset{\sim}{\beta}, \delta^2),\\
\pi(\beta, \delta^2) &= \frac{1}{(1+\delta^2)^2}
\end{aligned}
$$

where $i = 1 \ldots \ell,\ j = 1 \ldots n_i$, $\hat{\theta}_i = \log(\frac{y_i+1/2}{n_i-y_i+1/2})$, $\hat{\sigma}_i^2 = \frac{(n_i+1)(n_i+2)}{n_i(y_i+1)(n_i-y_i+1)}$, and $\underset{\sim}{x}'_i$ are covariates. Then the following posterior densities can be obtained (see Nandram, Erciulescu and Cruze, 2019)

$$
\theta_i \mid \underset{\sim}{\beta}, \delta^2, \hat{\theta}_i \overset{ind}{\sim} \text{Normal}\{\lambda_i\hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}'_i\underset{\sim}{\beta}, (1-\lambda_i)\delta^2\}, i = 1, \ldots, \ell,
$$

$$\underset{\sim}{\beta} \mid \delta^2, \hat{\underset{\sim}{\theta}} \sim \text{Normal}(\hat{\underset{\sim}{\beta}}, \hat{\Sigma}),$$

$$\pi(\delta^2 \mid \hat{\underset{\sim}{\theta}}) \sim Q(\delta^2)\frac{1}{(1+\delta^2)^2}$$

where

$$\lambda_i = \frac{\delta^2}{\hat{\sigma}_i^2 + \delta^2}, i = 1, \ldots, \ell,$$

$$\hat{\underset{\sim}{\beta}} = \hat{\Sigma}\sum_{i=1}^{\ell} \frac{\hat{\theta}_i \underset{\sim}{x}_i}{\hat{\sigma}_i^2 + \delta^2}, \quad \hat{\Sigma}^{-1} = \sum_{i=1}^{\ell} \frac{\underset{\sim}{x}_i \underset{\sim}{x}_i'}{\hat{\sigma}_i^2 + \delta^2},$$

$$Q(\delta^2) = |\hat{\Sigma}|^{1/2} \prod_{i=1}^{\ell} \frac{1}{(\hat{\sigma}_i^2 + \delta^2)^{1/2}} \exp\Big\{ -\frac{1}{2}\sum_{i=1}^{\ell} \frac{1}{\hat{\sigma}_i^2 + \delta^2}(\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})\Big\}.$$

Then we can draw 1000 samples of $\underset{\sim}{\beta}$ from $\underset{\sim}{\beta} \mid \delta^2, \hat{\underset{\sim}{\theta}}$ which is normal and δ^2 from $\pi(\delta^2 \mid \hat{\underset{\sim}{\theta}})$ using grid method. The grid ranges are constructed as

$$(avg(\underset{\sim}{\beta}^{(h)}) - 10 \cdot std(\underset{\sim}{\beta}^{(h)}), avg(\underset{\sim}{\beta}^{(h)}) + 10 \cdot std(\underset{\sim}{\beta}^{(h)})),$$

$$(avg(\delta^{(2h)}) - 10 \cdot std(\delta^{(2h)}), avg(\delta^{(2h)}) + 10 \cdot std(\delta^{(2h)})),\ h = 1, \ldots, M.$$

Let $z_i = \frac{v_i - \beta_0}{\delta}$ with the standard normal distribution, the joint posterior density of $\pi(\underset{\sim}{z}, \underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{y})$ becomes

$$\pi(\underset{\sim}{z}, \underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{y}) \propto \prod_{i=1}^{\ell} \Big\{ \Big[\prod_{j=1}^{n_i} \frac{e^{(x_{ij}'\underset{\sim}{\beta} + \delta z_i) y_{ij}}}{1 + e^{x_{ij}'\underset{\sim}{\beta} + \delta z_i}} \Big] \Big[\frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{z_i^2}{2}} \Big] \Big\} \times \frac{1}{(1+\delta^2)^2}\ .$$

Then we run a block Gibbs sampler between $\pi(\underset{\sim}{z} \mid \underset{\sim}{\beta}, \delta^2, \underset{\sim}{y})$ and $\pi(\underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{z}, \underset{\sim}{y})$ using the grid method to draw $z_i,\ i = 1, \ldots, \ell$, within a range of $(-5, 5)$ and draw $(\underset{\sim}{\beta}, \delta^2)$ from the ranges given by the Fay-Herriort model. We use 1,2000 iterates here with burn-in the first 2000 and take every 10th getting 1000 converged samples of $(\underset{\sim}{\beta}^{(h)}, \delta^{2(h)})$ and $\{z_i^{(h)}\},\ i = 1, \ldots, \ell,\ h = 1, \ldots, M$. Correspondingly, the samples of $\{\nu_i^{(h)}\},\ i = 1, \ldots, \ell$ can be obtained from $\nu_i^{(h)} = \delta^{(h)} z_i^{(h)} + \beta_0^{(h)}, h = 1, \ldots, M$. With all the coefficient estimated, we

are able to make inference about the binary response y_{ij} and get the proportion estimation further.

4.2.3 Comparisons

In Table 4.1 we compare the overweight proportion estimates of 35 areas from the logistic regression exact method and those from the Bayesian hierarchical model. Let y_i and n_i denote respectively, the number of 'yeses' and sample size within each area for each question. The direct estimates are calculated directly from dividing n_i by y_i.

We treat estimates from the logistic regression as a close value to the true proportion since it is obtained based on all the samples from each area utilizing all the covariates available. Figure 4.2 provides the 95% HPD intervals of the overweight proportion ϕ_{11} from Bayesian combined model. We find that the logistic regression estimates are all inside the 95% HPD interval. These results demonstrate that we are able to get a masked data through the randomized response technique while maintaining the overweight proportion within certain range.

Figure 4.2: 95% HPD interval of the overweight proportion ϕ_{11} from Bayesian combined model for 35 counties.

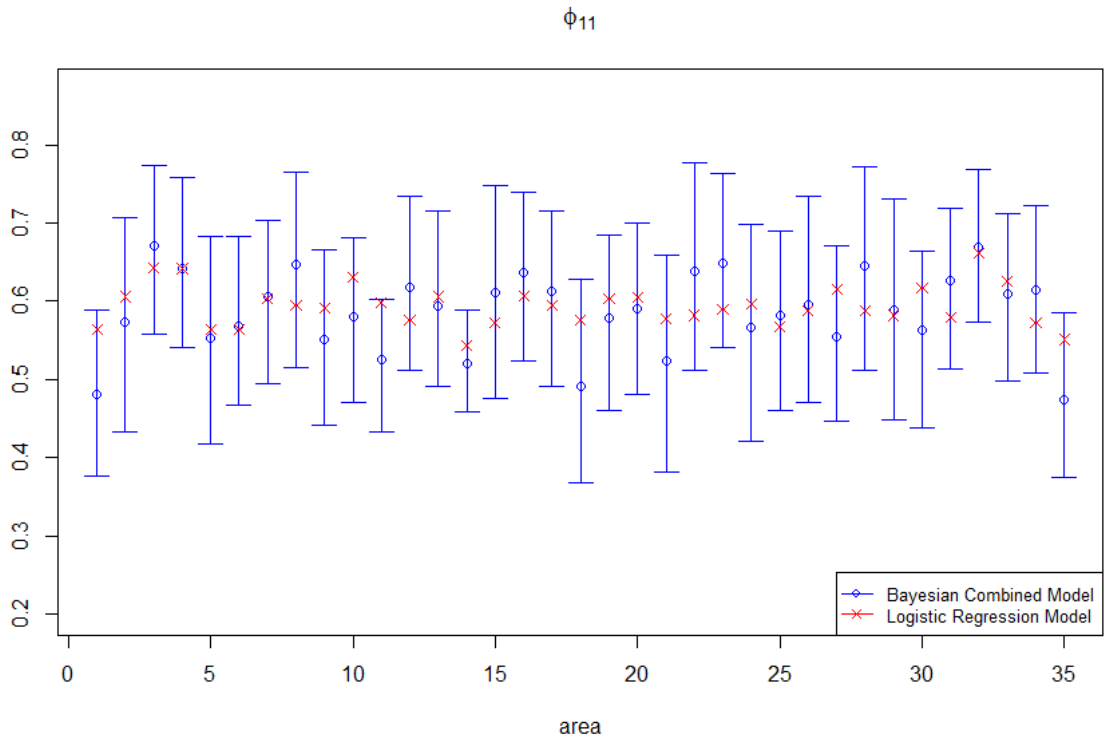


Table 4.1: Comparison of the logistic regression estimates (L) and the Bayesian estimates (B) of the overweight proportion ϕ_{11} (BMI) for 35 counties.

Area [1] -[9]	L	0.564	0.606	0.643	0.642	0.564	0.564	0.604	0.595	0.592
	B	0.481	0.576	0.672	0.635	0.506	0.569	0.607	0.646	0.555
Area [10] -[18]	L	0.631	0.598	0.576	0.606	0.544	0.573	0.607	0.595	0.576
	B	0.578	0.524	0.620	0.595	0.518	0.617	0.636	0.606	0.489
Area [19] -[27]	L	0.603	0.605	0.578	0.582	0.590	0.597	0.568	0.588	0.616
	B	0.576	0.593	0.521	0.640	0.648	0.568	0.580	0.599	0.551
Area [18] -[35]	L	0.588	0.582	0.617	0.580	0.662	0.625	0.573	0.551	
	B	0.650	0.592	0.564	0.624	0.670	0.611	0.618	0.476	

Table 4.2 and Figure 4.3 show the corresponding result for the osteoporosis proportion ϕ_{12}. Even though 3 areas out of 35 areas fail to fall between the 95% HPD interval, the finite population proportion estimation from the masked data are still close to those estimates from the Bayesian logistic regression model.

Figure 4.3: 95% HPD interval of the osteoporosis proportion ϕ_{12} from Bayesian combined model for 35 counties.

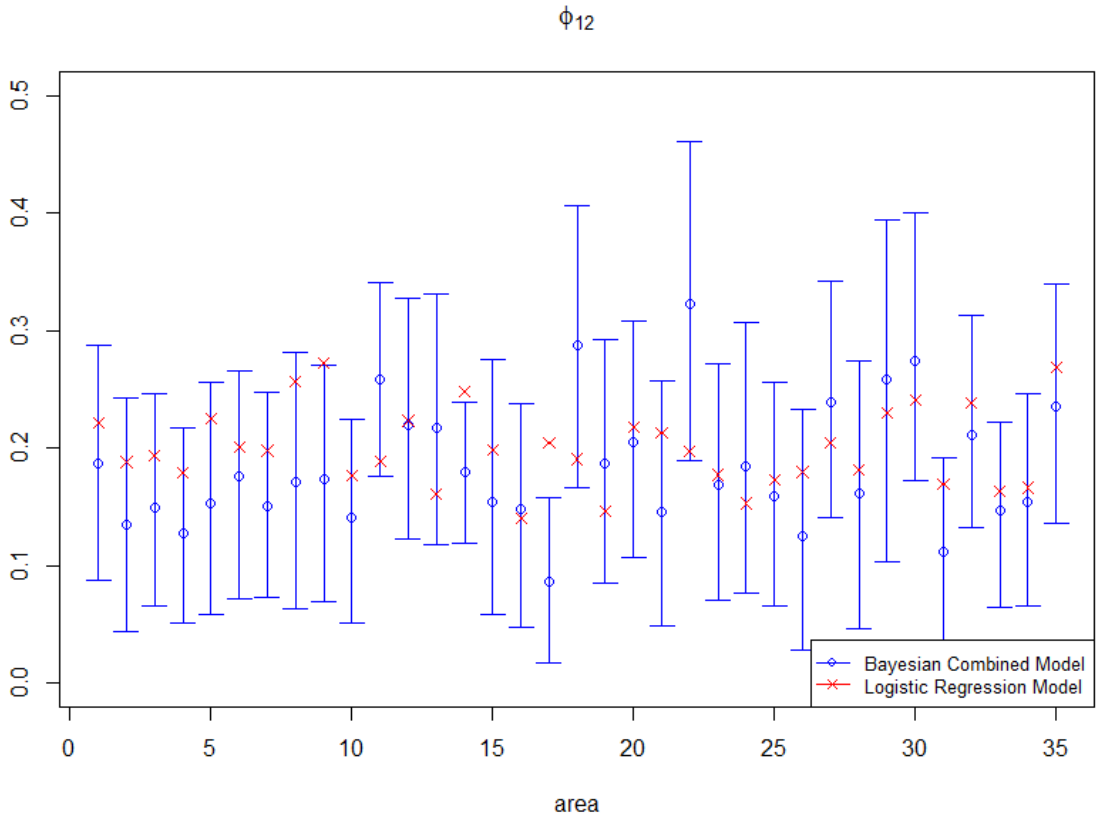


Table 4.2: Comparison of the logistic regression estimates (L) and the Bayesian estimates (B) of the osteoporosis proportion ϕ_{12} (BMD) for 35 counties.

Area [1] -[9]	L	0.222	0.188	0.194	0.179	0.225	0.201	0.198	0.257	0.272
	B	0.187	0.134	0.149	0.127	0.153	0.175	0.150	0.171	0.173
Area [10] -[18]	L	0.176	0.189	0.223	0.161	0.248	0.199	0.140	0.204	0.191
	B	0.140	0.259	0.220	0.217	0.180	0.154	0.148	0.087	0.287
Area [19] -[27]	L	0.146	0.217	0.213	0.197	0.177	0.153	0.173	0.180	0.204
	B	0.187	0.205	0.145	0.323	0.168	0.185	0.158	0.126	0.239
Area [18] -[35]	L	0.181	0.230	0.240	0.169	0.238	0.163	0.166	0.268	
	B	0.161	0.258	0.274	0.111	0.211	0.147	0.154	0.236	

Further more, we can generate the response for individual from each area. The response variable $y_{ij}, j = 1, \ldots, N_i$, indicates which group the individual falls in, the following types, "has neither overweight nor osteoporosis issues", "has osteoporosis but is not overwieght ", "overweight but does not have osteoporosis", "have both issues".

$$y_{ij} \overset{ind}{\sim} \text{Multinomial}\{1, \hat{\underset{\sim}{\pi}}_i\}, i = 1, \ldots, \ell,$$

where N_i is the sample size, $\hat{\underset{\sim}{\pi}}_i$ is the estimated finite population proportion estimation given by Table 4.3. Consequently, we are able to obtain a synthetic data set from any given sample size for each area in practice.

Table 4.3: Finite population proportion estimation for the four-cell probability of the NHANES III data using a Bayesian combined model.

	PM			
Area	π_{11}	π_{12}	π_{13}	π_{14}
1	0.382	0.139	0.435	0.044
2	0.333	0.091	0.536	0.041
3	0.226	0.103	0.627	0.044
4	0.268	0.091	0.608	0.033
5	0.363	0.129	0.425	0.082
6	0.329	0.103	0.495	0.073
7	0.306	0.083	0.543	0.068
8	0.244	0.110	0.585	0.061
9	0.335	0.114	0.494	0.058
10	0.310	0.111	0.548	0.031
11	0.328	0.148	0.416	0.108
12	0.248	0.132	0.536	0.083
13	0.289	0.120	0.494	0.097
14	0.346	0.135	0.477	0.042
15	0.291	0.094	0.556	0.060
16	0.302	0.065	0.550	0.083
17	0.331	0.060	0.586	0.023
18	0.291	0.221	0.426	0.063
19	0.323	0.103	0.490	0.085
20	0.283	0.129	0.514	0.074
21	0.387	0.095	0.474	0.044
22	0.191	0.174	0.490	0.146
23	0.275	0.078	0.560	0.087
24	0.322	0.113	0.492	0.073
25	0.304	0.115	0.537	0.043
26	0.331	0.075	0.541	0.054
27	0.271	0.179	0.491	0.059
28	0.233	0.116	0.606	0.045
29	0.212	0.197	0.530	0.062
30	0.293	0.142	0.435	0.130
31	0.304	0.071	0.586	0.039
32	0.245	0.086	0.545	0.123
33	0.337	0.053	0.517	0.093
34	0.267	0.112	0.578	0.042
35	0.393	0.132	0.374	0.101

4.3 Future Work

We discuss four possible extensions. These are about incorporating covariates in our combined model, survey weights, numerous items instead of two and polychotomous (more than two options) responses.

First, covariates are very useful, not only in constructing the non-sensitive questions like what we did in the survey example, but also can be incorporated into the probability parameters in a way similar to logistic regression. Furthermore, based on the proportion estimation results of each area given by either the Dirichlet-Multinomial model in Chapter 2 or the INNA model in Chapter 3, a masked data set with the same size can be generated through modeling with covariates.

Second, many complex surveys have survey weights. These can be included in our model using a normalized composite likelihood. This will help to reduce selection bias.

Third, the design can be generalized to the multiple-item case straightforwardly. However, the computation will be more expensive. As the number of items get larger, the model will be more complicated and even the blocked Gibbs sampler would experience a slowly mixing effect, which means more iterates are needed to get the converged draws in the end.

Fourth, it can also be extended to the polychotomous outcomes. Then the latent variables will follow multinomial distributions. A similar idea as in the combined model might be needed to avoid slow mixing in a Gibbs sampler. Of course, computation time will increase and we will need a way to minimize computational cost. This is also true in the third extension.

However, further effort still needed to explore whether the correlations will affect the estimation strength other than changing the contingency table of the counts.

Bibliography

[1] Blair, G., Imai, K. and Zhou, Y-Y. (2015), "Design and Analysis of the Randomized Response Technique," *Journal of the American Statistical Association, Review*, 110, 1304-1319.

[2] Bellhouse, D. R. (1995), "Estimation of Correlation in Randomized Response," *Survey Methodology*, 21, 13-19.

[3] Barksdale, W. B. (1971), "New Randomized Response Techniques for Control of Non-sampling Errors in Surveys," *unpublished Ph.D. thesis, University of North Carolina, Chapel Hill, Dept. of Biostatistics.*

[4] Chung, Ray S. W., Chu, Amanda M. Y. and So, Mike K. P. (1995), "Bayesian Randomized Response Technique with Multiple Sensitive Attributes: The Case of Information Systems Resource Misuse," *The Annals of Applied Statistics*, 12, 1969-1992.

[5] Clickner, R.P. and Iglewicz, B. (1980), "Warner's Randomized Response Technique: The Two Sensitive Questions Case," *South African Statist.*, 14, 77- 86.

[6] Edgell, S. E., Himmelfarb, S., and Cira, D. J. (1986), "Statistical Efficiency of Using Two Quantitative Randomized Response Techniques to Estimate Correlation," *Survey Methodology*, 100, 251-256.

[7] Eriksson, S. A. (1973), "A New Model for Randomized Response," *International Statistical Review*, 41, 101-113.

[8] Fox, J. A. and Tracy, P. E. (1986), *Randomized Response: A Method for Sensitive Surveys*, Sage: London.

[9] Fienberg, S. E. and Jin, J. (2009), "Statistical disclosure limitation for data access," *In Encyclopedia of Database Systems*, 27832789, Springer.

[10] Greenberg, B. G., Abul-Ela, A.-L. A., Simmons, W. R. and Horvitz, D. G. (1969), "The Unrelated Question Randomized Response Model: Theoretical Framework," *Journal of the American Statistical Association*, 64, 520-539.

[11] Greenberg, B. G., Kuebler, R. R., Abernathy, J. R. and Horvitz, D. G. (1971) "Application of the Randomized Response Technique in Obtaining Quantitative Data," *Journal of the American Statistical Association*, 66, 243-250.

[12] Gupta, S., Gupta, B. and Singh, S. (2002), "Estimation of Sensitivity Level of Personal Interview Survey Questions," *Journal of Statistical Planning and Inference*, 100, 239-247.

[13] Gupta, S., Javid, S. and Supriti, S. (2010), "Mean and Sensitivity Estimation in Optional Randomized Response Models," *Journal of Statistical Planning and Inference*, 140, 2870-2874.

[14] Geisser, S. and W. Eddy. (1979), "A Predictive Approach to Model Selection." *Journal of the American Statistical Association*, 74, 153160.

[15] Hu, J., Reiter, J. P., and Wang, Q. (2015), "Dirichlet Process Mixture Models for Modeling and Generating Synthetic Versions of Nested Categorical Data," *Journal of Statistical Planning and Inference*, 13, 183200.

[16] Kwak, S. G., Nandram, B. (2004) and Kim D. H. (2018), "Bayesian Inference on Contingency Tables with Uncertainty about Independence for Small Areas," *Journal of Applied Statistics*, 45, 2145-2163.

[17] Kwan, S. S. K., So, M. K. P. and Tam, K. Y. (2010), "Research Note - Applying the Randomized Response Technique to Elicit Truthful Responses to Sensitive Questions in IS research: The Case of Software Piracy Behavior," *Information Systems Research*, 21, 941-959.

[18] Lee, C-S., Sedory, S. A. and Singh, S. (2013a), "Estimating at Least Seven Measures of Qualitative Variables from a Single Sample using Randomized Response Technique," *Statistics & Probability Letters*, 83, 399-409.

[19] Lee, C-S., Sedory, S. A. and Singh, S. (2013b), "Simulated Minimum Sample Size Requirements in Various Randomized Response Models," *Communications in Statistics - Simulation and Computation*, 42, 771-789.

[20] Moors, J. (1971), "Optimization of the Unrelated Question Randomized Response Model," *Journal of the American Statistical Association*, 66, 627-629.

[21] 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.

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

[23] Nandram, B. and Choi, J. W. (2010), "A Bayesian Analysis of Body Mass Index Data from Small Domains under nonignorable nonresponse and selection," *Journal of the American Statistical Association*, 105, 120-135.

[24] Nandram, B., Kim, D. H. and Zhou, J. (2018), "A Pooled Bayes Test of Independence for Sparse Contingency Tables from Small Areas," *Journal of Statistical Computation and Simulation* (partially accepted).

[25] Nandram, B. and Yu, Y. (2017), "Bayesian Analysis of Sparse Counts Under the Unrelated Question Design," *JSM Proceeding, Survey Research Methodology Section*, American Statistical Association, Alexandria, VA, pp. 1162-1175.

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

[27] Nandram, B. and Yu, Y. (2018a), "Bayesian Analysis of Sparse Counts Under the Unrelated Question Design," (submitted).

[28] Nandram, B. and Yu, Y. (2018b), "Bayesian Analysis of a Sensitive Proportion for a Small Area," *International Statistical Review*, 0, 1-17, doi:10.1111/insr.12286.

[29] Nandram, B., Erciulescu, A.L. and Cruze N.B. (2019), "Bayesian Benchmarking of the Fay-Herriot Model Using Random Deletion," *Survey Methodology* (in press)

[30] Oh, M. (1994), "Bayesian Analysis of Randomized Response Models: A Gibbs Sampling Approach," *Journal of the Korean Statistical Society*, 23, 463-482.

[31] O'Hagan, A. (1987), "Bayes Linear Estimators for Randomized Response Models," *Journal of the American Statistical Association*, 82, 580-585.

[32] Pal, S. (2017), "Bootstrap technique for Randomized Response Surveys," *Statistics and Applications*, 15, 79-92.

[33] Reiter, J. P. (2003), "Inference for partially synthetic, public use microdata sets. *Survey Methodology*, 29, 181189.

[34] Reiter, J. P. (2005), "Releasing multiply-imputed, synthetic public use microdata: An illustration and empirical study. *Journal of the Royal Statistical Society, Series A*, 168, 185205.

[35] Rubin, D. B. (1993), "Discussion: Statistical disclosure limitation. *Journal of Official Statistics*, 9, 462468.

[36] Rao, J.N.K. and Molina, I. (2015), "Small Area Estimation," *Wiley Series in Survey Methodology*

[37] Rue, H., Martino, S. and Chopin, N. (2009), "Approximate Bayesian Inference for Latent Gaussian Models by using Integrated Nested Laplace Approximations," *Journal of the Royal Statistical Society*, Series B (Statistical Methodology) 71 (2), 319-392.

[38] Soeken, K. L. and Macready, G. B. (1986), "Application of Setwise Randomized Response Procedures for Surveying Multiple Sensitive Attributes," *Psychological Bulletin*, 99(2), 289-295.

[39] Song, J. J. and Kim, J-M. (2017), "Bayesian Estimation of Rare Sensitive Attribute," *Communications in Statistics - Simulation and Computation*, doi: 10.1080/03610918.2015.1109655.

[40] Tamhane, A. C. (1981), "Randomized Response Techniques for Multiple Sensitive Attributes," *Journal of the American Statistical Association*, 76, 916-923.

[41] Tian, G.-L., Yuen, K. C., Tang, M.-L. and Tan, M. T. (2009), "Bayesian Non-randomized Response Models for Surveys with Sensitive Questions," *Statistics and Its Interface*, 2, 13-25.

[42] Tourangeau, R., Rips, L. J. and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge University Press: Cambridge, England.

[43] Tourangeau, R. and Yan, T. (2007), "Sensitive Questions in Surveys," *Psychological Bulletin*, 133, 859-883.

[44] Warner, S. L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60, 63-69.

[45] Winkler, R. L., and Franklin, L. A. (1979), "Warner's Randomized Response Model: A Bayesian Approach," *Journal of the American Statistical Association*, 74, 207-214.

[46] Yu, Y., Bhadra, D., Nandram, B. (2017), "Tests of Independence for a Two-by-two Contingency Table with Random Margins," *International Journal of Statistics and Probability*, 6, 2, 106.

[47] Yu, Y. and Nandram, B. (2018), "Bayesian Analysis of Unrelated Question Design for Correlated Sensitive Questions from Small Areas," *JSM Proceeding., Section of Bayesian Statistical Science*, American Statistical Association, Alexandria, VA, pp, 2838-2848.