

# Constraint Bayesian Inference for Count Data from Small Areas

by

Xinyu Chen

A PhD Dissertation

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

---

December, 2021

APPROVED:

---

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

---

Professor Joseph Sedransk  
Joint Program in Survey Methodology  
University of Maryland

---

Professor Buddika Peiris  
Department of Mathematical Sciences  
Worcester Polytechnic Institute

---

Professor Fangfang Wang  
Department of Mathematical Sciences  
Worcester Polytechnic Institute

---

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

---

Dr. Myron Katzoff  
Laurel, MD

## Abstract

The small area analysis of survey data has received a lot of attention. Borrowing information from other areas can provide reliable and accurate estimates when the sample size of an area is small. In many applications, it is necessary to take into account possible order restrictions of the unknown parameters of interest, and it is reasonable to make such an assumption. With the order restriction assumption, pooling data can provide more accurate estimates, where parameters can increase up to a point (mode) and decrease thereafter. In this dissertation, we assume unimodal order restrictions on parameters of interest. First, we describe Bayesian hierarchical multinomial-Dirichlet models with order restrictions for count data from small areas. Second, we incorporate uncertainty into the model with unimodal order restrictions as an extension. Third, we describe the models with exchangeability–nonexchangeability (EXNEX) priors, which allow borrowing information across similar areas while avoiding too optimistic borrowing for extreme areas.

First, due to the natural characteristics of the data, making unimodal order restriction assumptions to parameter spaces is relevant. We present the models with order restrictions on different parameters of interest to explore how borrowing information under different order restriction assumptions works differently. In the simulation, we compare these models with order restrictions under different scenarios, where we assume three levels of heterogeneity between areas. In a small heterogeneity scenario, the model with stronger order restrictions on parameters borrow more information among areas and has smaller relative bias, posterior

standard deviation, and higher credible interval coverage than other models. We develop methods to generate posterior samples for the models with different order restrictions assumptions.

Second, in our application to body mass index (BMI) data from the NHANES III, we assume people may have a high chance to have overweight BMI level. We assume the same unimodal order restriction across all counties, where the mode is at the third position. But we notice the same unimodal order restriction for all areas may not hold. To have a more robust model, we incorporate uncertainty into the unimodal order restriction. We let the modal position for each area be a random variable and have mixture probabilities for the modal position, which means each area can have different order restrictions. We provide an approximation of log-pseudo marginal likelihood as a model diagnostic procedure. In the application to the BMI data and simulated data, we compare the performance of different models with or without order restrictions. We show that the performance of the model, incorporating uncertainty about order restrictions, is consistent and it can provide relatively accurate estimates of parameters in the application. We demonstrate how the model with order restrictions can borrow information among areas differently from the model without order restriction.

Third, when population means are clustered into two or more subgroups, shrinking all the means towards a common weighted average is inappropriate. A useful substitute for exchangeability in the above situation is partial exchangeability. We present exchangeability–nonexchangeability (EXNEX) models, which allow borrowing information across similar areas while avoiding too optimistic borrowing for extreme areas. We present a gridgy Gibbs sampler to draw samples from the joint posterior distribution of the binomial-Beta EXNEX model. In the simula-

tion, we illustrate the robustness of EXNEX models, which have small relative bias under different scenarios. Then we extend the approach to a multinomial-Dirichlet EXNEX model with order restrictions. In the application to BMI data, we compare the multinomial-Dirichlet EXNEX model with order restrictions and the multinomial-Dirichlet model with order restrictions. We show that the EXNEX model with order restrictions can borrow information across similar areas while avoiding borrowing from very different areas. So the EXNEX model with order restrictions is preferred in some cases.

Overall, borrowing information among areas is a key idea in small area estimation. The hierarchical structure of the models with order restrictions is easy to apply to small area estimation problems. The main issue we focus on here is to borrow information with the unimodal order restrictions on cell probabilities, which can borrow more information among areas than the model without order restrictions. As extensions of our approach, incorporating uncertainty about the order restrictions may solve the problem that the same unimodal order restrictions across areas may not hold. Partial exchangeability of parameters are recommended to allow borrowing across similar areas and avoid optimistic borrowing for very different areas. Our theoretical and methodological work can help provide accurate and efficient small area statistics for many national surveys.

## **Acknowledgement**

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. He also inspired and encouraged me to explore new knowledge during this period.

I would also like to extend my sincere thanks to my committee members, Professor Joe Sedransk, Professor Fangfang Wang, Professor Buddika Peiris, Dr. Jai Won Choi, and Dr. Myron Katzoff, for their time and effort in making my dissertation better. Each member has provided me invaluable comments and insight in improving my work.

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

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

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview of Small Area Estimation . . . . .	1
1.1.1	Model Based Approaches . . . . .	3
1.2	Order Restriction . . . . .	6
1.2.1	Order Restrictions for General Statistic Inference . . . . .	7
1.2.2	Order Restrictions for Small Area Estimation . . . . .	8
1.3	Partial Exchangeability . . . . .	14
1.4	National Health and Nutrition Examination Survey . . . . .	17
1.5	Gibbs Sampler and Convergence Diagnostics . . . . .	18
1.6	Model Diagnostics . . . . .	21
1.7	Contributions of this Research and Dissertation Organization . . . . .	22
<b>2</b>	<b>Bayesian Hierarchical Multinomial-Dirichlet Model</b>	<b>24</b>
2.1	The Model with Order Restrictions on $\mu$ Only . . . . .	25
2.2	The Model with Order Restrictions on $\theta$ Only . . . . .	27
2.3	The Model with Order Restrictions on $\theta$ and $\mu$ . . . . .	28
2.4	Simulation Under Six Scenarios . . . . .	30

<i>CONTENTS</i>	vii
2.5 Concluding Remarks . . . . .	33
2.6 Appendix . . . . .	34
2.6.1 Griddy Gibbs Sampler for $\mu, \tau$ in $M_\mu$ . . . . .	34
2.6.2 Griddy Gibbs Sampler for $\mu$ and $\tau$ in $M_\theta$ . . . . .	35
2.6.3 Griddy Gibbs Sampler for $\mu$ and $\tau$ in $M_2$ and $M_3$ . . . . .	36
2.6.4 Sampling from Conditional Posterior Distribution of $\theta$ in $M_\theta, M_2,$ and $M_3$ . . . . .	38
<b>3 Uncertainty about Order Restrictions</b>	<b>40</b>
3.1 The Model Incorporated Uncertainty . . . . .	40
3.2 Numerical Example: Body Mass Index . . . . .	43
3.2.1 Body Mass Index . . . . .	43
3.2.2 MCMC Convergence Diagnostics . . . . .	44
3.2.3 Model Comparison of $M_1, M_2, M_3,$ and $M_4$ . . . . .	46
3.3 Numerical Example: Smoothed BMI . . . . .	57
3.4 Concluding Remarks . . . . .	60
3.5 Appendix . . . . .	63
3.5.1 Sampling Method for Parameter $\theta, \mu,$ and $\tau$ in $M_4$ . . . . .	63
3.5.2 Bayesian Diagnostics for $M_2, M_3,$ and $M_4$ . . . . .	64
3.5.3 Posterior Summary of $\theta$ in $M_1, M_2, M_3,$ and $M_4$ . . . . .	66
<b>4 Partial Exchangeability</b>	<b>70</b>
4.1 Binomial-Beta EXNEX Model . . . . .	70
4.1.1 Simulation Under Three Scenarios . . . . .	73
4.2 Multinomial-Dirichlet EXNEX Model without Order Restriction . . . . .	76
4.3 Multinomial-Dirichlet EXNEX Model with Order Restrictions . . . . .	78

<i>CONTENTS</i>	viii
4.4 Numerical Example: BMI . . . . .	81
4.5 Concluding Remarks . . . . .	91
4.6 Appendix . . . . .	93
4.6.1 Griddy Gibbs Sampler for the Binomial-Beta EXNEX Model 93	
4.6.2 Griddy Gibbs Sampler for the EXNEX Model $M_1^{exnex}$ . .	94
4.6.3 Griddy Gibbs Sampler for the EXNEX Model with Order Restrictions $M_3^{exnex}$ . . . . .	95
4.6.4 Posterior Summary of Model $M_1, M_3, M_1^{exnex}$ , and $M_3^{exnex}$ for BMI . . . . .	98
<b>5 Concluding Remarks and Future Works</b>	<b>101</b>
5.1 Concluding Remarks . . . . .	101
5.2 Future works . . . . .	103



# List of Figures

1.1	Good trace plot . . . . .	20
3.1	Overall females BMI in five categories . . . . .	44
3.2	Traceplots of $\mu$ in $M_2$ and $M_3$ . . . . .	45
3.3	Posterior densities of $\theta$ for County 2 showing different order restrictions under different models. . . . .	48
3.4	Posterior densities of $\theta$ for County 3 showing different order restrictions under different models. . . . .	49
3.5	Posterior densities of $\theta$ for County 13 showing different order restrictions under different models. . . . .	50
3.6	Posterior densities of $\theta$ for County 35 showing different order restrictions under different models. . . . .	51
3.7	Boxplot of $\theta$ posterior samples: Part I . . . . .	53
3.8	Boxplot of $\theta$ posterior samples: Part II . . . . .	54
3.9	Standard deviation comparison between those models to show improvement. . . . .	55
3.10	CPOs for 35 counties under different models . . . . .	56

3.11	Simulation method to have the unimodal order restriction . . . . .	58
3.12	CPOs for 35 counties under different models (simulation) . . . . .	60
4.1	Traceplots of $\mu$ in $M_3^{exnex}$ . . . . .	82
4.2	Densities of $\mu$ in $M_3^{exnex}$ . . . . .	82
4.3	Densities of $\theta$ in $M_3$ and $M_3^{exnex}$ for County 2 . . . . .	87
4.4	Densities of $\theta$ in $M_3$ and $M_3^{exnex}$ for County 3 . . . . .	88
4.5	Densities of $\theta$ in $M_3$ and $M_3^{exnex}$ for County 13 . . . . .	89
4.6	Densities of $\theta$ in $M_3$ and $M_3^{exnex}$ for County 35 . . . . .	90
4.7	Densities of $\theta$ in $M_3$ and $M_3^{exnex}$ for County 21 . . . . .	90

# List of Tables

2.1	Simulation summary for four models under three scenarios . . . . .	31
3.1	Number of females at five levels of BMI . . . . .	43
3.2	Model notations in Chapter 3 . . . . .	44
3.3	Effective sample sizes of $\mu$ and $\tau$ . . . . .	45
3.4	P-values of Geweke tests for $\mu$ and $\tau$ . . . . .	45
3.5	Total posterior standard variance of $\theta$ . . . . .	52
3.6	LPMLs of model $M_1, M_2, M_3,$ and $M_4$ for BMI . . . . .	57
3.7	LPMLs of model $M_1, M_2, M_3,$ and $M_4$ for simulated data . . . . .	59
3.8	Posterior summary of $\theta$ part I: Counties 1-9 . . . . .	66
3.9	Posterior summary of $\theta$ part II: Counties 10-18 . . . . .	67
3.10	Posterior summary of $\theta$ part III: Counties 19-27 . . . . .	68
3.11	Posterior summary of $\theta$ part III: Counties 28-35 . . . . .	69
4.1	True success probabilities $\theta_i$ and area sizes for seven areas under three scenarios with different heterogeneity . . . . .	73
4.2	P-values of Geweke tests for $\mu, \rho,$ and $\phi$ in the Model $BBM$ and $BBM^{exnex}$ for one simulation . . . . .	74

4.3 Effective sample sizes of  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $BBM$  and  $BBM^{exnex}$  for one simulation . . . . . 74

4.4 Posterior standard deviation (PSD) of the success probabilities  $\theta_i$  for seven areas under three scenarios (one simulation) . . . . . 75

4.5 Average relative bias and RMSE of  $\theta_i$  in Model  $BBM$  and  $BBM^{exnex}$  under three scenarios (relatively different heterogeneity) . . . . . 75

4.6 Model notations in Chapter 4 . . . . . 81

4.7 P-values of Geweke tests for  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $M_1^{exnex}$  and  $M_3^{exnex}$  . . . . . 83

4.8 Effective sample sizes of  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $M_1^{exnex}$  and  $M_3^{exnex}$  . . . . . 83

4.9 Posterior summary of  $\theta$  in four models for County 3 . . . . . 84

4.10 Posterior summary of  $\theta$  in four models for County 27 . . . . . 85

4.11 Posterior summary of  $\theta$  in four models for County 9 . . . . . 86

4.12 Posterior summary of  $\theta$  in four models for County 21 . . . . . 86

4.13 EXNEX posterior summary part I: Counties 1-9 . . . . . 98

4.14 EXNEX posterior summary part II: Counties 10-22 . . . . . 99

4.15 EXNEX posterior summary part IV: Counties 23-35 . . . . . 100

# Chapter 1

## Introduction

Small area estimation is a statistical technique involving the estimation of parameters for small sub-populations that is included in a larger survey. Small area estimation plays a crucial role in sample surveys, which can borrow information across areas. In the context of small area estimation, hierarchical Bayes approaches are widely used. One important example is that the multinomial-Dirichlet model works well for count data. In many cases, such as the age compositions of the population of Atlantic cod, it is necessary to take into account the order restrictions of the unknown parameters of interest. Several studies have been developed to incorporate different order restriction assumptions into different models. The hierarchical models with order restrictions can improve the accuracy of estimates.

### 1.1 Overview of Small Area Estimation

Collecting and analyzing information play important roles in modern society. Data are regularly collected to satisfy the need for information about specified

sets of elements, called finite population. One of the most important methods is a sample survey, which is recognized as cost-effective and scientifically valid. As an example, it may be desired to know income and poverty in US for distributing funds and managing federal programs. In this situation, samples from school districts, counties and states would be obtained via national surveys. Then estimates of income and poverty would be calculated. A properly conducted sample survey will support inference from the sample that is valid about the population. It costs less than a complete enumeration, and it is usually less time consuming and can be more accurate with regard to the avoidance of non-sampling errors when, for example, the possibility of response error is a major concern.

Sample surveys provide estimates not only for the entire population of interest but also for a variety of subpopulation, which may be defined by geographic areas or socio-demographic groups (Rao and Molina 2015). Example of geographic areas include a state, county, school district, metropolitan area, and health service area. An area or subdomain is regarded as “small” if the area or subdomain contains too few ultimate population units to provide a sample of sufficient size and support direct estimates of adequate precision from the sample survey. For example, one aspect of a national survey is to provide proper estimates of poverty in states. The sample sizes of counties may not be large enough to support direct estimates for subgroups of the state populations, such as persons in each county. Due to cost considerations, it is often not possible to have a large enough overall sample size to support those estimates for all domains. It is also impossible to design another survey only for those small areas.

The use of survey data in developing reliable small area estimation (SAE), possibly in conjunction with the census and administrative data, has received lots of

attention. Demographers have long been using a variety of indirect methods for SAE of population and other characteristics of interest. In recent years, legislative acts by national governments have increasingly created a need for small area statistics and business demand for small area statistics from private sector has also accelerated the trend. Rao and Molina (2015) provided a very-detailed review on models and methods of SAE, but this is out of date now because the field is growing rapidly.

### **1.1.1 Model Based Approaches**

Model based approaches have several advantages. First, model diagnostics can be used to find proper models to satisfy different needs. Second, area-specific measures of precision can be associated with each small area estimate, unlike average over small areas. Third, linear mixed models and nonlinear models can be applied to complex data structures. Random effects models can achieve accurate small area inferences.

We can classify model based approaches into two broad types. (1) Area level models that relate the small area means to area-specific auxiliary variables; and (2) Unit level models that relate the unit values of the study variable to unit-specific auxiliary variables.

There are three essential model based approaches for small area estimation, which are empirical best linear unbiased prediction (EBLUP), parametric empirical Bayes (EB), and parametric hierarchical Bayes (HB). In the hierarchical Bayes (HB) approach, a subjective prior distribution  $f(\lambda)$  on the model parameters  $\lambda$  is

specified and the posterior distribution  $f(\boldsymbol{\mu}|\mathbf{y})$  can be obtained from

$$f(\boldsymbol{\mu}|\mathbf{y}) = \int f(\boldsymbol{\mu}, \lambda|\mathbf{y})d\lambda = \int f(\boldsymbol{\mu}|\mathbf{y}, \lambda)f(\lambda|\mathbf{y})d\lambda.$$

Note that  $f(\boldsymbol{\mu}|\mathbf{y}, \hat{\lambda})$  is for EB estimators (Rao and Molina 2015).

As a special case in the hierarchical Bayes approach, Bayesian hierarchical binomial-Beta models are effectively designed for binary data. Assume that given  $p_i$ , the number of success,  $Y_i, \dots, Y_k$ , are independent with

$$Y_i|p_i \sim \text{Binomial}(n_i, p_i), \text{ for } i = 1, 2, \dots, k,$$

$$p_1, \dots, p_k|\mu, \tau \sim \text{Beta}(\mu\tau, (1-\mu)\tau),$$

where  $p(p_i|\mu, \tau) = \frac{\Gamma(\tau)}{\Gamma(\mu\tau)\Gamma((1-\mu)\tau)}p_i^{\mu\tau-1}(1-p_i)^{(1-\mu)\tau-1}$ .

A vague non-informative prior can be used,

$$\pi(\mu, \tau) = \frac{1}{(1+\tau)^2}, \text{ where } 0 < \mu < 1, \tau > 0.$$

To perform the hierarchical Bayes analysis, Gibbs sampler can be used.

As an extension of the binomial-Beta model, the hierarchical Bayesian multinomial Dirichlet model is appropriate for count data. Nandram (1998) considered the problem of pooling data from several multinomial populations using a hierarchical Bayesian model. He specified a hierarchical multinomial-Dirichlet model for multinomial data, which allows for the uncertainty in the estimation of all the hyper-parameters. Letting  $n_{ij}$  be the cell counts,  $p_{ij}$  the corresponding cell probabilities,  $i = 1, \dots, \ell$ ,  $j = 1, \dots, K$ ,  $n_{i.} = \sum_{j=1}^K n_{ij}$ ,  $\mathbf{p}_i = (p_{i1}, \dots, p_{iK})'$ ,



$\sum_{j=1}^K p_{ij} = 1$ , given  $p_i$ ,  $n_i$  follow a multinomial distribution,

$$\mathbf{n}_i | \mathbf{p}_i \sim \text{Multinomial}(n_i, \mathbf{p}_i).$$

At the second stage, given hyper-parameters  $\boldsymbol{\mu}$  and  $\tau$ , the  $p_i$  follow a Dirichlet distribution,

$$\mathbf{p}_i | \boldsymbol{\mu}, \tau \stackrel{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau),$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$ ,  $0 < \mu_j < 1$ ,  $\tau > 0$ ,  $i = 1, \dots, I$ .

At the third stage, Nandram (1998) assumed that

$$\boldsymbol{\mu} \sim \text{Dirichlet}(\boldsymbol{\mu}_0\tau_0),$$

$$\tau \sim \text{Gamma}(\eta_0, \nu_0),$$

where  $\boldsymbol{\mu}_0, \tau_0, \eta_0, \nu_0$  are specified. Or at the third stage, it can be  $\pi(\boldsymbol{\mu}, \tau) = \frac{(K-1)!}{(1+\tau)^2}$  as a vague non-informative prior of  $\boldsymbol{\mu}$  and  $\tau$  for the multinomial-Dirichlet model without any order restriction (Nandram, Kim, and Zhou 2019).

To study the association between bone mineral density and body mass index (BMI) from several U.S. counties, Nandram, Kim, and Zhou (2019) provided a clear discussion of the general hierarchical multinomial-Dirichlet model and their methodology for small area estimation. In their application to the third National Health and Nutrition Examination Survey (NHANES III), 31 counties have been used. Pooling the data using small area estimation techniques is necessary since counts in some cells are very small. Specifically, let  $n_s = \sum_{j=1}^r \sum_{k=1}^c n_{sjk}$  be the cell counts,  $\mathbf{p}_s = (p_{s1}, p_{s2}, \dots, p_{sr})'$  and  $\mathbf{q}_s = (q_{s1}, q_{s2}, \dots, q_{sc})'$  are the marginal probabilities of bone mineral density and BMI in county  $s$ ,  $\text{vec}(p_s \mathbf{q}_s')$

be the corresponding cell probabilities for county  $s$ ,  $s = 1, 2, \dots, S$ , row  $j = 1, 2, \dots, r$ , column  $k = 1, 2, \dots, c$ . The model, which estimates the association between bone mineral density and BMI, is

$$\begin{aligned} \mathbf{n}_s | \mathbf{p}_s, \mathbf{q}_s &\sim \text{Multinomial}\{\mathbf{n}_{s\cdot}, \text{vec}(\mathbf{p}_s \mathbf{q}'_s)\}, s = 1, 2, \dots, S, \\ \mathbf{p}_s | \boldsymbol{\mu}_1, \tau_1 &\sim \text{Dirichlet}(\boldsymbol{\mu}_1 \tau_1), \\ \mathbf{q}_s | \boldsymbol{\mu}_2, \tau_2 &\sim \text{Dirichlet}(\boldsymbol{\mu}_2 \tau_2), \\ \pi(\boldsymbol{\mu}_1, \tau_1, \boldsymbol{\mu}_2, \tau_2) &= \frac{(r-1)! (c-1)!}{(1+\tau_1)^2 (1+\tau_2)^2}. \end{aligned}$$

Note the  $\pi(\boldsymbol{\mu}_1, \tau_1, \boldsymbol{\mu}_2, \tau_2)$  is proper.

Without any prior information, they take  $\boldsymbol{\mu}_1$ ,  $\boldsymbol{\mu}_2$ ,  $\tau_1$  and  $\tau_2$  to be independent. As an interpretation of hyper-parameters,  $\boldsymbol{\mu}$  are cell means and  $\tau$  is a prior sample size. This model features stratification and hyper-parameters to pool information from different strata.

We assume the same unimodal order restrictions on parameters into the Bayesian hierarchical multinomial-Dirichlet models. Our proposed models are different from their work, and can provide accurate estimates of parameters of interest when the order restriction assumptions hold.

## 1.2 Order Restriction

In this section, we briefly review some of the existing constrained inference on the parameters of interest.

### 1.2.1 Order Restrictions for General Statistic Inference

In many statistical problems, it is necessary to take into account the order restrictions of the unknown parameters of interest. Sometimes, it is reasonable to assume order restrictions on the parameters. Robertson, Wright, and Dykstra (1998) provided a comprehensive treatment of the topic of statistical inference under inequality constraints, where much of the theory is based on the principles of maximum likelihood estimation and likelihood ratio tests. Silvapulle and Sen (2004) discussed a broad range of inequality-constrained inference problems. They clearly illustrated concepts with practical examples from a variety of fields, focusing on sociology, econometrics, and biostatistics. Heck and Davis-Stober (2019) provided a comprehensive discussion about linear inequality constraints, such as the set of monotonic order constraints for binary choice probabilities on the parameters of multinomial distributions for psychological theories. Wu, Meyer and Opsomer (2016) combined domain estimation and the pooled adjacent violators algorithm to construct new design-weighted constrained estimators of wage for U.S. National Compensation Survey. The survey estimates mean wages for many job categories, and these mean wages are expected to be non-decreasing according to job levels. They assumed constrained estimators satisfied monotonicity which improved point estimation and sharpened confidence intervals. Their design-based approach is not our concern and their constraints are different from our unimodal order restrictions.

Nandram (2005) provided a Bayesian method to obtain the best subset of products and to assess the quality of the products. A priori information about which subset is the best is incorporated, and a stochastic ordering is modified to select the best subset of the products. But their constraints are different from our order

restrictions.

Nandram and Peiris (2018) developed a robust Bayesian analysis to study count data obtained from a nine-point hedonic scale at two time points. They robustified a standard binormal model by modeling skewness to obtain the receiver operating characteristic (ROC) curve and the area under the curve. They also incorporated a stochastic ordering to enforce ROC curve with robust measures. But their stochastic ordering on ROC curve is totally different from our unimodal order restrictions on parameters.

Li (2008) has a nice review about treatment effects models with order restrictions, since the treatment effect would be higher for a higher dose. The purpose of using estimators that take the order restrictions into consideration is to gain efficiency. He considered different statistical problems in which an order restriction on the unknown parameters is either natural or reasonable, and discussed methods of estimation and inference under the restrictions.

There are lots of discussion about how to make inferences with order restrictions. They show the importance of incorporating order restrictions into the model, which can have the considerable improvement in precision. But their approaches are not aimed for small area estimation.

### **1.2.2 Order Restrictions for Small Area Estimation**

In context of small area estimation, there is a trend considering the order parameters if one is interested in estimating jointly the best and worst areas's parameters, such as the largest multinomial probability and smallest multinomial probability. Dunson and Neelon (2003) proposed a Bayesian approach for inference on order-constrained parameters in generalized linear models. They focused

on the case where the data consist of normally distributed observations for subjects in  $k$  groups and there is prior knowledge of an ordering in the means. For  $j = 1, \dots, k$  group, let  $y_i$  denote an  $n_j \times 1$  vector observations of i.i.d.  $N(\mu_j, \sigma^2)$  random variables. Assuming a conjugate prior density,  $\mu_j | \sigma^2 \sim N(\mu_{0j}, \sigma^2 / k_{0j})$  and  $\sigma^2 \sim IG(a_0, b_0)$ , instead of choosing a prior distribution with support on the constrained space, which can result in major computational difficulties, they proposed to map draws from an unconstrained posterior density using an isotonic regression transformation. In particular, they focused inference on  $\mu^* = g(\mu)$ , where  $\mu$  is the unconstrained mean,  $\mu^*$  is the constrained mean, and  $g(\cdot)$  is the isotonic regression transformation. This approach allows flat regions over which increases in the level of a predictor have no effect. Bayes factors, which is a Bayesian diagnostic criterion, for assessing ordered trends can be computed based on the output from a Gibbs sampling algorithm.

Malinovsky and Rinott (2010) studied the prediction of ordered random effects in a basic unit level model for small area estimation. They considered a simplified Fay-Herriot model as in

$$y_{ij} = \mu + u_i + e_{ij}, j = 1, \dots, n, i = 1, \dots, m,$$

where  $y_{ij}$  is observed for area  $i = 1, \dots, m$ ,  $\mu$  is an unknown constant,  $u_i \sim N(0, \sigma_u^2)$  and  $e_{ij} \sim N(0, \sigma_e^2/n)$ .

Set  $\theta_i = \mu + u_i$  and  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ . The purpose is to predict the ordered random variables  $\boldsymbol{\theta} = \{\theta_{(1)} \leq \dots \leq \theta_{(m)}\}$  from the observed  $y$ . They suggested shrinkage-type predictors, such as  $\widehat{\theta}_{(i)} = \gamma y_{(i)} + (1 - \gamma)\bar{y}$ , where  $y_{(1)} \leq \dots \leq y_{(m)}$  denote the order statistics of  $y_1, \dots, y_m$ , with an appropriate amount of shrinkage

for the particular problem of ordered parameters, are better and close to the optimal predictors.

But those empirical Bayesian models with order restrictions are not appropriate for binary or count data. Their order restrictions are not the same as ours.

As a special case of Bayesian hierarchical models for small area estimation, the multinomial-Dirichlet model with order restrictions can be applied to the count data where the order restriction about parameters may exist.

Sedransk, Monahan and Chiu (1985) described a Bayesian method for estimation of finite population parameters in general population surveys. A multinomial distribution is used to model the variable of interest,  $Y$ , from a categorical table. A Dirichlet prior distribution is assigned to the cell probabilities,  $\mathbf{p} = (p_1, \dots, p_t)$ . They added order restrictions to the model to capture the unimodal smoothness relationships among cell probabilities  $(p_1, \dots, p_t)$ , such as

$$p_1 \leq \dots \leq p_k \geq p_{k+1} \geq \dots \geq p_t.$$

If the modal value,  $k$ , is known, then the prior distribution is assumed to be given by

$$f(\mathbf{p}) = C_k(\beta_1, \dots, \beta_t) \prod_{i=1}^t p_i^{\beta_i - 1},$$

where  $\mathbf{p} \in R_t^{(k)} = \{\mathbf{p} : p_1 \leq \dots \leq p_k \geq p_{k+1} \geq \dots \geq p_t\}$ . They showed how to use Monte Carlo integration to evaluate the posterior moments or probabilities. But their model is not designed for small area estimation. The lack of Bayesian diagnostics for their model suggests that there are a lot of improvement to be made.

Gelfand, Smith and Lee (1992) provided very-detailed Gibbs sampler struc-

tures for Bayesian analysis of constrained parameters. They suggested that a Dirichlet prior should be used for ordered multinomial parameters, which are  $p_1 \leq p_2 \leq \dots \leq p_t \geq \dots \geq p_K$ . A Bayesian model for their problem is given by defining  $Y_j = \#$  of observations in category  $j$ . So the likelihood is

$$Y|\mathbf{p} \sim \text{Mult}(n; p_1, \dots, p_K),$$

Given  $t$ , they specified a prior

$$f(\mathbf{p}|t) \propto c(\beta_1, \dots, \beta_K, t) \prod_{j=1}^J p_j^{\beta_j-1},$$

where  $c(\beta_1, \dots, \beta_K, t)$  is the normalizing constant, and  $R^K = \{(p_1, \dots, p_K) | p_1 \leq p_2 \leq \dots \leq p_t \geq \dots \geq p_K, 0 \leq p_j \leq 1, \sum_{j=1}^K p_j = 1\}$ .

Gelfand, Smith and Lee noted that the Gibbs sampler cannot be employed directly when  $t$  is unknown and prior  $Pr(t = j) = w_j, j = 1, \dots, K$ . But the marginal posterior for  $t$  can be calculated directly, taking the form

$$Pr(t = j|Y) = \frac{c(\beta_1, \dots, \beta_K, j)w_j/c(\beta_1 + Y_1, \dots, \beta_K + Y_K, j)}{\sum_{j=1}^K c(\beta_1, \dots, \beta_K, j)w_j/c(\beta_1 + Y_1, \dots, \beta_K + Y_K, j)}.$$

They showed Bayesian inference on order parameters can have higher precision. However their Dirichlet multinomial model with the ordered parameters does not consider stratification and hyper-parameters to borrow information from small areas.

For a stratified population, Nandram and Sedransk (1995) showed the precision of inference about  $\pi_{ij}$ , which is the proportion of firms in stratum  $i$  belonging to

the sales and receipts class  $j$ , can be dramatically increased by using Dirichlet multinomial model with appropriate order restrictions on  $\pi_{ij}$ . They considered uncertain modal positions to increase their model flexibility. Prior specification of parameter  $\alpha$  in Dirichlet distribution has to be proper and reasonable to the model inference. Their order restriction is more complicated than the previous model due to the stratification. They also consider the case where there is uncertainty about the vector of modal positions  $L$ , which can take  $g$  possible values,  $\ell_1, \ell_2, \dots, \ell_g$ . The position probabilities are given below,

$$Pr(L = \ell_s) = w_s, s = 1, 2, \dots, g, \text{ where } w_s \text{ are specified and } \sum_{s=1}^g w_s = 1.$$

They directly applied Monte Carlo integration to estimate the posterior  $w_s = Pr(L = \ell_s | \mathbf{n})$ . Adopting a Bayesian view, they showed that the posterior variances can be dramatically reduced by including order restrictions among  $\pi_{ij}$ , both within and between the strata. But their model does not have hyper-parameters to borrow information among strata.

Nandram, Sedransk and Smith (1997) improved estimation of the age composition of the population of Atlantic cod with the help of order-restricted Bayesian estimation. Let  $p_{ij}$  denote the cell probability that a fish belongs to length stratum  $i$  and age class  $j$ . To simplify the analysis, the likelihood of  $\mathbf{p}$  is

$$\ell(\mathbf{p} | \mathbf{n}) \propto \prod_{i=1}^I \prod_{j=1}^J p_{ij}^{n_{ij}}.$$



They took independent Dirichlet distributions as prior; that is

$$f(\mathbf{p}|\boldsymbol{\alpha}) \propto \prod_{i=1}^I \prod_{j=1}^J p_{ij}^{\alpha_{ij}-1},$$

where  $\alpha_{ij} > 0$  is a fixed quantity, within stratum  $i$ ,  $p_{i1} \leq \dots \leq p_{ik_i} \geq \dots \geq p_{iu_i}$  for some  $k_i \in Z_i$ . In the set  $Z_i$ , the cells labelled  $(i, \ell_i), (i, \ell_i + 1), \dots, (i, u_i)$ , where  $\ell_i$  and  $u_i$  are integers such that  $1 \leq \ell_i \leq u_i \leq J$ . The  $p_{ij}$  are to be estimated for each  $j \in Z_i$ .

In their Atlantic cod study, let  $i = 1$  correspond to the stratum with the shortest fish and  $j = 1$  correspond to the youngest fish. It is expected that as  $i$  increases, the relative values of the  $\{p_{ij} : j \in Z_i\}$  will change. The order restrictions are not just within strata, but also among strata, such as

$$p_{i1} \leq \dots \leq p_{it} \geq \dots \geq p_{iK},$$

$$p_{j1} \leq \dots \leq p_{jt^*} \geq \dots \geq p_{jK} \text{ where } i < j \text{ and } t < t^*.$$

They also considered the case where there is uncertainty about the modal position  $L$ ,

$$Pr(L = \ell_s) = w_s, s = 1, 2, \dots, g.$$

They showed the joint posterior distribution of  $\pi$  and  $L$  is

$$f(\pi, L = \ell_s | n) = \frac{w_s C_{\ell_s}(\boldsymbol{\alpha}) \prod_{i=1}^I g_{n_i}(\pi_i)}{\sum_{s'=1}^g w_{s'} C_{\ell_{s'}}(\boldsymbol{\alpha}) / C_{\ell_{s'}}(\mathbf{n})}.$$

Their model with order restrictions is different from our proposed models, since their order restrictions are within strata and among strata.

Overall, making inference with order restrictions for small area estimation is feasible and necessary.

### 1.3 Partial Exchangeability

The idea to exploit the similarity and borrow information across strata has a long tradition. The methods are known to improve estimation accuracy over estimate obtained from stratification or complete pooling. Standard hierarchical models assume full exchangeability of parameters, which is usually expressed via random effects. The multinomial-Dirichlet model with order restrictions can improve estimation accuracy, since it can borrow more information across small areas than the model without order restriction. We found the same order restrictions may not be true for all small areas, which will make the performance of the model with order restriction worse.

Malec and Sedransk (1992) have pointed out the weakness of Bayesian models based solely on the exchangeability assumption, where we might have the same issue in the model for our application. To illustrate their approach, for  $i = 1, \dots, L, j = 1, \dots, n_i$ , the  $Y_{ij}$  are independent with  $Y_{ij} \sim N(\mu_i, \sigma_i^2)$ , where the  $\sigma_i^2$  are assumed to be known. They proposed a flexible prior distribution for parameter  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_L)$  to permit the amount and nature of the pooling to be influenced by the sample data. They use grouping method to have subsets of  $\boldsymbol{\mu}$  such that the  $\mu_i$  within each subset are similar, and there is uncertainty about the composition of such subsets of  $\boldsymbol{\mu}$ . But our approach to have partial exchangeability is different from their grouping method.

A useful substitute for exchangeability in the above situation is partial ex-

changeability, where the components within a subgroup are exchangeable, but the different subgroups are not. Ghosh and Kim (2002) considered a general Bayesian model which allows multiple grouping of parameters, where the components within a subgroup are exchangeable. In other word, they assumed the model with partially exchangeability. The general idea is then illustrated for the normal means estimation problem under priors which are scale mixture of normals. They discussed also implementation of the Bayes procedure via Markov chain Monte Carlo integration techniques. They illustrated the proposed methods with a numerical example. But their approach to have partially exchangeability is different from ours.

Kaizer, Koopmeiners and Hobbs (2017) introduced multisource exchangeability models, a general Bayesian approach for integrating multiple, potentially nonexchangeable, supplemental data sources into the analysis of a primary data source. They give us the inspiration to construct a model with partially exchangeability.

Neuenschwander et al. (2016) proposed models that allow each stratum-specific parameter to be exchangeable with parameters from other similar strata, or nonexchangeable with any of the other strata parameters. In other word, this approach can automatically borrow information from those similar areas, and borrow much less information from those very different areas.

These exchangeability-nonexchangeability (EXNEX) models are robust and can provide accurate estimates of parameters of interest, since they allow borrowing information across similar strata while avoiding too optimistic borrowing from extremely different strata. However, the EXNEX models proposed by them are designed for clinical trials. First, they use weakly informative or informative priors, which heavily rely on the previous study or historical data. Second, their models

for clinical trials are totally different from ours. Third, they do not have any order restriction assumption on parameters of interest.

Our approach to have partially exchangeability is inspired by their work. To simply illustrate a EXNEX model, a general hierarchical binomial-beta EXNEX model is given below. For area  $i = 1, \dots, \ell$ , the number of successes  $y_i$  in  $n_i$  trials are independent with

$$r_i | \theta_i \sim \text{binomial}(n_i, \theta_i),$$

where  $\theta_i$  are the corresponding success probabilities in the  $i$ th area. With mixture probabilities  $p_i$ , we assume fully exchangeability (EX), where the parameter vector  $\boldsymbol{\theta}$  is exchangeable and

$$\boldsymbol{\theta}_i | \alpha, \beta \sim \text{Beta}(\alpha, \beta),$$

and the prior of  $\alpha$  and  $\beta$  is  $\pi(\alpha, \beta) = 1$ .

With mixture probabilities  $1 - p_i$ , we assume nonexchangeability (NEX), where  $\boldsymbol{\theta}_i$  is nonexchangeable and

$$\boldsymbol{\theta}_i | \alpha_0, \beta_0 \sim \text{Beta}(1, 1).$$

We notice that the label switching problem arises when taking a Bayesian approach to parameter estimation and clustering using mixture models. In a Bayesian context the invariance of the likelihood under relabeling of the mixture components can lead to the posterior distribution of the parameters being highly symmetric and multimodal, making it difficult to summarize (Stephens 2000). We consider con-

strained parameter spaces to deal with label switching issue in the mixture models, which are our EXNEX models.

## **1.4 National Health and Nutrition Examination Survey**

The performance of our method is studied using the Third National Health and Nutrition Examination Survey, NHANES III. NHANES III is a stratified multistage probability design targeted to obtain a representative sample of the total civilian noninstitutionalized U.S. population age 2 months and older. The sample was selected from households across the United States during the period October 1988 through September 1994. Some individuals are selected with different probabilities. Nandram and Choi (2010) emphasized that NHANES III was designed to oversample the two largest minority groups in the U.S. population: black non-Hispanic and Hispanic-Americans. For confidentiality reasons, the final data set for this study uses only the 35 largest counties (from 14 states) with a population of at least 500,000 for selected age categories by sex (male, female) and race (white non-Hispanic, black non-Hispanic, Hispanic, other).

One of the variables in this survey is body mass index (BMI; body weight in kilograms divided by  $[(\text{height in meters})^2]$ ), which is currently used as a measure to diagnose overweight and obesity in children and adolescents. Specifically, 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 purpose of the thesis is to predict the percentile BMI for the finite population of female adults, post-stratified by counties. We use a selected subset of BMI data from NHANES III, where we only use the female adults older than 20

years from 35 counties with a population at least 500,000. The final female BMI data set for this study uses only 3149 samples, which is small relative to the finite population. So we use methods associated with small area estimation. Our method can be easily applied to other subsets of BMI data, such as the male BMI data set.

The original sensitive attributes BMI data are transformed to categorical data based on the criteria defined by the Centers for Disease Control (CDC), which are underweight, normal, overweight, obese I, and obese II. If BMI is less than 18.5, it falls within the underweight range. If BMI is from 18.5 to 25, it falls within the normal. If BMI is from 25.0 to 30, it falls within the overweight range. If BMI is from 30.0 to 35, it falls within the obese I range. If BMI is 35.0 or higher, it falls within the obese II range. We can look at small domains, but the cells of the multinomial tables will become sparse. We can eliminate some counties that become small or we can combine some counties. The hierarchical multinomial-Dirichlet model is ideal for contingency tables, but the framework of this dissertation does not cover any covariate. Due to the natural characteristics of BMI, unimodal order restrictions are reasonable to incorporate. Most people have higher chance to have normal or overweight BMI, and less likely to be underweight or obese.

## 1.5 Gibbs Sampler and Convergence Diagnostics

In the hierarchical Bayesian approaches, to evaluate the desired posterior density  $f(\boldsymbol{\mu}|\mathbf{y})$  or estimate a parameter of interest by its posterior mean,  $\hat{\Theta} = E[h(\boldsymbol{\mu}|\mathbf{y})]$ , high-dimensional integrations may be involved. Gibbs sampler, one of Markov chain Monte Carlo (MCMC) methods, is useful to overcome the computational difficulties from multi-dimensional integrations.

The point of Gibbs sampler is that given a multivariate distribution it is simpler to sample from a conditional distribution than to marginalize by integrating over a joint distribution. Hobert and Casella (1996) showed how to use Gibbs sampling in hierarchical linear mixed models even with improper priors. Gibbs sampling output cannot provide information that the posterior is improper. But the hierarchical binomial-Beta models and multinomial-Dirichlet models have been widely used. It is easy to see the propriety of their posterior.

There are three issues associated with Gibbs sampler to be considered. Gibbs sampler, generate Markov Chains which simulate the posterior distribution  $\pi(\cdot)$ . Early proportion of the chain that may not converge to the target posterior distribution should be dropped off, which is called burn-in. Thinning samples will reduce the final autocorrelation in the sample, and provide valid inference on the posterior.

Roy (2020) discussed most widely used MCMC convergence diagnostics tools. Even most of the proposed diagnostics have shortcomings, we decide to use a few criteria, such as Geweke's diagnostic, effective sample size, trace plots to acquire the converged posterior samples.

Geweke (1992) used spectral analysis to evaluate numerical accuracy formally and constructed diagnostics for convergence. Geweke's convergence diagnostic is a Z test of equality of means where autocorrelation in the samples is taken into account while calculating the standard error. Let  $\bar{g}_{n_1}$  and  $\bar{g}_{n_2}$  be the time averages based the first  $n_1$  and the last  $n_2$  observations,  $\widehat{S}_g(0)$  be the estimate of its standard error. Thus, Geweke's statistics is

$$Z_n = (\bar{g}_{n_1} - \bar{g}_{n_2}) / \sqrt{\widehat{S}_g(0)/n_1 + \widehat{S}_g(0)/n_2}$$

Geweke suggests using  $n_1 = 0.1n_{total}$  and  $n_2 = 0.5n_{total}$ .

The effective sample size (ESS) of MCMC measures the amount by which autocorrelation in samples increases uncertainty (standard errors) relative to an independent sample, given by

$$ESS = \frac{n}{1 + 2 \sum_{k=1}^{\infty} \rho(k)},$$

where  $n$  is the number of samples and  $\rho(k)$  is the correlation at lag  $k$ . If samples are independent, the effective sample size equals the actual sample size.

The most common graphical convergence diagnostic method is the trace plot. The trace plot is a time series plot that shows the realizations of the Markov chain at each iteration against the iteration numbers. This graphical method is used to visualize how the Markov chain is moving around the state space, that is, how well it is mixing. It is often said that a good trace plot should look like a hairy caterpillar, as in Figure 1.1.

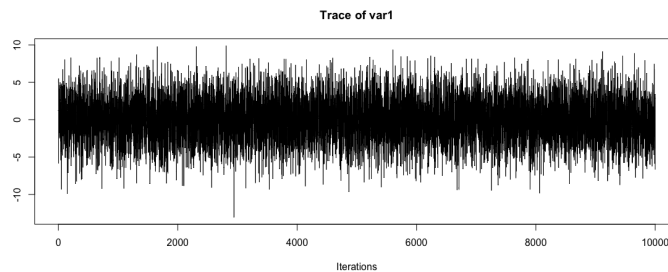


Figure 1.1: Good trace plot



## 1.6 Model Diagnostics

Gelfand, Dey and Chang (1992) used predictive distributions to address the issues of model adequacy and model selection. They proposed the conditional predictive ordinate for the model determination. The conditional predictive ordinate (CPO) is based on leave-one-out cross validation. Let  $\mathbf{n} = (n_1, \dots, n_K)$  is a set of data,  $n_{(i)}$  are the data omitting the  $i$ th observation. CPO estimates the probability of observing  $n_i$  in the future if after having already observed  $n_{(i)}$ , given as

$$CPO_i = f(n_i | n_{(i)}),$$

where  $f(\cdot | n_{(i)})$  is the predictive distribution of a new observation given  $n_{(i)}$ . The CPO is a probability, which lies in  $[0, 1]$ . Models with larger CPO values provide better fit to the observed data. Low CPO values suggest possible outliers, high-leverage and influential observations.

Compared with the marginal likelihood in Bayes factors, CPO can be estimated by taking the inverse of the posterior mean of the inverse density function value of  $n_i$ , which is a harmonic mean of the likelihood of  $n_i$ . Thus,  $\widehat{CPO}_i = [\frac{1}{S} \sum_{s=1}^S \frac{1}{f(n_i | \boldsymbol{\theta}^{(s)})}]^{-1}$ , where  $\boldsymbol{\theta}^{(s)}$  are posterior samples from  $f(\boldsymbol{\theta} | \mathbf{n})$ . This form may be computationally unstable though, and some care is needed. Generally, it is easier to estimate CPO than Bayes factors, since it is easier to have posterior samples than integrating the marginal likelihood for most cases.

The log-pseudo marginal likelihood  $LPML = \sum_{i=1}^K \log(CPO_i)$ , which is an estimator for the log marginal likelihood, can be used to quantify the support for a model over another. The 'best' model among competing models have the largest

LPML.

## **1.7 Contributions of this Research and Dissertation Organization**

We provide a brief introduction about small area estimation and Bayesian hierarchical approach with order restrictions. We introduce body mass index (BMI) data in NHANES III for later illustration. We have given a comprehensive overview in Chapter 1.

In Chapter 2, we discuss the Bayesian hierarchical model with order restrictions. The multinomial-Dirichlet model with three different order restriction assumptions are discussed as well. Then we show how to draw samples from the posterior distribution using MCMC method. To compare different order restrictions assumptions, we present a simulation study.

In Chapter 3, in order to accommodate the different degrees of heterogeneity between small areas, we describe a Bayesian hierarchical multinomial-Dirichlet model with uncertain order restrictions. In particular, we describe the model incorporated with the uncertainty about the modal position. We show how to draw samples from the proposed model using MCMC method with some approximation approaches because of the complexity of the model structure. Then we apply our model to the BMI data to provide accurate estimates of the BMI proportions for each county. We also compare the model incorporated uncertainty with the multinomial-Dirichlet model with order restrictions.

In Chapter 4, we discuss the partial exchangeability assumption in the Bayesian hierarchical model. As an introduction, we describe a simple binomial Beta model

with exchangeability-nonexchangeability (EXNEX) priors. Then we show how to draw samples from the joint posterior distribution. We present a simple simulation from different heterogeneity scenarios to show the robustness of the model with EXNEX priors. We also describe a multinomial-Dirichlet model with EXNEX priors and how to draw posterior samples from it.

In Chapter 5, we summarize our result that the model with order restrictions may be necessary for small area estimation. We present concluding remarks to show the complexity and difficulty of the model with order restrictions and provide some extensions of the multinomial-Dirichlet model with order restrictions. We discuss future research work to extend the application of the current method for small area estimation.

## Chapter 2

# Bayesian Hierarchical Multinomial-Dirichlet Model

We notice that there are benefits to add order restriction assumptions into the hierarchical models. We also notice that the potential of Bayesian hierarchical multinomial-Dirichlet model for small area estimation can be elaborated. Due to the complexity of the multinomial-Dirichlet model with order restrictions, the efficient Bayesian inference and Bayesian diagnostics should be discussed as well.

To be consistent, we use some notations from the general hierarchical multinomial-Dirichlet model for further reference. As a hierarchical Bayesian approach for small area estimation, let  $n_{ij}$  be the cell counts,  $\theta_{ij}$  be the corresponding cell probabilities, for areas  $i = 1, 2, \dots, \ell$ , categories  $j = 1, 2, \dots, K$ , and  $n_{i.} = \sum_{j=1}^K n_{ij}$ . The general hierarchical multinomial-Dirichlet model, in Nandram, Kim, and Zhou

(2019 ) and Nandram (1998), is

$$\begin{aligned} \mathbf{n}_i | \boldsymbol{\theta}_i &\overset{ind}{\sim} \text{Multinomial}(\mathbf{n}_{i\cdot}, \boldsymbol{\theta}_i), \\ \boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau &\overset{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), \\ \pi(\boldsymbol{\mu}, \tau) &= \frac{(K-1)!}{(1+\tau)^2}, \end{aligned}$$

where  $f(\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau) = \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1}$ ,  $D(\boldsymbol{\mu}\tau)$  is the normalization constant for  $\text{Dirichlet}(\boldsymbol{\mu}\tau)$ ,  $\mu_j > 0$ ,  $\sum_{j=1}^K \mu_j = 1$ ,  $\tau > 0$ .

Without any prior information, let  $\boldsymbol{\mu}$  and  $\tau$  be independent,  $E(\theta_{ij}) = \mu_j$ ,  $\sum_{j=1}^K \mu_j = 1$ . As an interpretation of hyper-parameters,  $\boldsymbol{\mu}$  are cell means and  $\tau$  is a prior sample size. This model features stratification and hyper-parameters to pool information from different strata together. In our BMI data application and simulation, we denote the general multinomial-Dirichlet model as  $M_1$ .

In the following section, we will provide three models with different order restriction assumptions.

## 2.1 The Model with Order Restrictions on $\mu$ Only

Let  $n_{ij}$  be the cell counts,  $\theta_{ij}$  be the corresponding cell probabilities, areas  $i = 1, 2, \dots, \ell$ , categories  $j = 1, 2, \dots, K$ , and  $n_{i\cdot} = \sum_{j=1}^K n_{ij}$ . The hierarchical multinomial-Dirichlet model is

$$\mathbf{n}_i | \boldsymbol{\theta}_i \overset{ind}{\sim} \text{Multinomial}(\mathbf{n}_{i\cdot}, \boldsymbol{\theta}_i), \quad i = 1, \dots, \ell.$$

At the second stage we assume

$$\theta_i | \boldsymbol{\mu}, \tau \stackrel{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), i = 1, \dots, \ell.$$

For hyperparameters  $\boldsymbol{\mu}$  and  $\tau$ , we consider non-informative prior

$$\pi(\boldsymbol{\mu}, \tau) = \frac{K(m-1)!(K-m)!}{(1+\tau)^2}, \quad \mu_j > 0, \quad \sum_{j=1}^K \mu_j = 1, \quad \boldsymbol{\mu} \in C_{\boldsymbol{\mu}},$$

where

$$C_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} : \mu_1 \leq \dots \leq \mu_m \geq \dots \geq \mu_K\}.$$

There may be an unimodal structure on the cell probabilities  $\boldsymbol{\theta}$ . Since  $E(\theta_{ij}) = \mu_j$ , it is reasonable to assume the expectation of cell probabilities have unimodal order restrictions. We assume an order restriction assumption on the hyperparameter  $\boldsymbol{\mu}$ ,  $\boldsymbol{\mu} \in C_{\boldsymbol{\mu}}$ , and assume the modal position  $m$  in  $C_{\boldsymbol{\mu}}$  is known,

In our BMI data application and simulation, we denote the multinomial-Dirichlet model with order restriction on  $\boldsymbol{\mu}$  as  $M_{\boldsymbol{\mu}}$ .

The joint posterior distribution of  $M_{\boldsymbol{\mu}}$  is

$$\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n} \propto \prod_{i=1}^{\ell} \left\{ \prod_{j=1}^K \theta_{ij}^{n_{ij}} \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} \right\} \frac{K(m-1)!(K-m)! I_{C_{\boldsymbol{\mu}}}}{(1+\tau)^2},$$

where  $D(\boldsymbol{\mu}\tau) = \frac{\prod_{j=1}^K \Gamma(\mu_j \tau)}{\Gamma(\sum_{j=1}^K \mu_j \tau)}$  is the normalization constant of Dirichlet distribution and  $I_{C_{\boldsymbol{\mu}}}$  is an indicator function that the constraint that  $\boldsymbol{\mu} \in C_{\boldsymbol{\mu}}$  is satisfied.

In Appendix 2.6.1, we use a gridy Gibbs sampler to draw samples from the joint posterior distribution for parameters and make statistical inference on those

posterior samples.

## 2.2 The Model with Order Restrictions on $\theta$ Only

Let  $n_{ij}$  be the cell counts,  $\theta_{ij}$  be the corresponding cell probabilities, areas  $i = 1, 2, \dots, \ell$ , categories  $j = 1, 2, \dots, K$ , and  $n_{i\cdot} = \sum_{j=1}^K n_{ij}$ . The hierarchical multinomial-Dirichlet model is

$$\mathbf{n}_i | \boldsymbol{\theta}_i \stackrel{ind}{\sim} \text{Multinomial}(\mathbf{n}_{i\cdot}, \boldsymbol{\theta}_i), \quad i = 1, \dots, \ell.$$

At the second stage we assume

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau \stackrel{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), \boldsymbol{\theta}_i \in C, \quad i = 1, \dots, \ell,$$

where  $C = \{\boldsymbol{\theta}_i : \theta_{i1} \leq \dots \leq \theta_{im} \geq \dots \geq \theta_{iK}, i = 1, \dots, \ell\}$ , and assume the modal position  $m$  in  $C$  is known.

For hyperparameters  $\boldsymbol{\mu}$  and  $\tau$ , we consider non-informative prior

$$\pi(\boldsymbol{\mu}, \tau) = \frac{(K-1)!}{(1+\tau)^2}, \quad \mu_j > 0, \quad \sum_{j=1}^K \mu_j = 1, .$$

In our BMI data application and simulation, we denote the multinomial-Dirichlet model with order restriction on  $\theta$  as  $M_\theta$ .

The joint posterior density is

$$\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n} \propto \prod_{i=1}^{\ell} \left\{ \prod_{j=1}^K \theta_{ij}^{n_{ij}} \frac{1}{D(\boldsymbol{\mu}\tau)C(\boldsymbol{\mu}\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} \right\} \frac{(K-1)!}{(1+\tau)^2},$$

where  $D(\boldsymbol{\mu}\tau) = \frac{\prod_{j=1}^K \Gamma(\mu_j\tau)}{\Gamma(\sum_{j=1}^K \mu_j\tau)}$  is the normalization constant of Dirichlet distribution,  $C(\boldsymbol{\mu}\tau) = \int_{\boldsymbol{\theta}_i \in C} \frac{\Gamma(\sum_{j=1}^K \mu_j\tau)}{\prod_{j=1}^K \Gamma(\mu_j\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j\tau-1} d\boldsymbol{\theta}_i$  is the normalization constant of the truncated Dirichlet distribution.

In Appendix 2.6.2, we use a gridy Gibbs sampler to draw samples from the joint posterior distribution for parameters  $\boldsymbol{\mu}$  and  $\tau$ . Then in Appendix 2.6.4, we show how to draw samples from the conditional posterior distribution of  $\boldsymbol{\theta}$  and make statistical inference on those posterior samples.

### 2.3 The Model with Order Restrictions on $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$

Chen and Nandram (2019) incorporated the order restriction into the Bayesian hierarchical Dirichlet multinomial model. Let  $n_{ij}$  be the cell counts,  $\theta_{ij}$  the corresponding cell probabilities,  $i = 1, 2, \dots, \ell$ ,  $j = 1, 2, \dots, K$ ,  $\mathbf{n}_{i\cdot} = \sum_{j=1}^K n_{ij}$  and denote the mode of  $\boldsymbol{\theta}_i$ s as  $\theta_{im}$ ,  $1 \leq m \leq K$ .

Specifically, we assume

$$\mathbf{n}_i | \boldsymbol{\theta}_i \stackrel{ind}{\sim} \text{Multinomial}(\mathbf{n}_{i\cdot}, \boldsymbol{\theta}_i), \quad i = 1, \dots, \ell.$$

At the second stage we assume

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau \stackrel{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), \quad \boldsymbol{\theta}_i \in C, \quad i = 1, \dots, \ell,$$

where  $C = \{\boldsymbol{\theta}_i : \theta_{i1} \leq \dots \leq \theta_{im} \geq \dots \geq \theta_{iK}, i = 1, \dots, \ell\}$ , and assume the modal position  $m$  in  $C$  is known,

$$\pi(\boldsymbol{\mu}, \tau) = \frac{K(m-1)!(K-m)!}{(1+\tau)^2}, \quad \mu_j > 0, \quad \sum_{j=1}^K \mu_j = 1, \quad \boldsymbol{\mu} \in C\boldsymbol{\mu},$$



where

$$C_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} : \mu_1 \leq \dots \leq \mu_m \geq \dots \geq \mu_K\}.$$

Since  $E(\theta_{ij}) = \mu_j$ ,  $\boldsymbol{\mu}$  should have the same order restriction as  $\boldsymbol{\theta}_i$ , which is  $\boldsymbol{\mu} \in C_{\boldsymbol{\mu}}$ . We assume the modal position  $m$  in  $C_{\boldsymbol{\mu}}$  is known.

A posteriori  $\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, \mathbf{n}_i \stackrel{ind}{\sim} \text{Dirichlet}(\mathbf{n}_i + \boldsymbol{\mu}\tau)$ ,  $\boldsymbol{\theta}_i \in C, i = 1, \dots, \ell$ , where

$$\begin{aligned} f(\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, \mathbf{n}) &= \frac{\frac{\Gamma[\sum_{j=1}^K (n_{ij} + \mu_j \tau)]}{\prod_{j=1}^K \Gamma(n_{ij} + \mu_j \tau)} \prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1}}{\int_{\boldsymbol{\theta}_i \in C} \frac{\Gamma[\sum_{j=1}^K (n_{ij} + \mu_j \tau)]}{\prod_{j=1}^K \Gamma(n_{ij} + \mu_j \tau)} \prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1} d\boldsymbol{\theta}_i} \\ &= \frac{\frac{\Gamma[\sum_{j=1}^K (n_{ij} + \mu_j \tau)]}{\prod_{j=1}^K \Gamma(n_{ij} + \mu_j \tau)} \prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1}}{C(\mathbf{n}_i + \boldsymbol{\mu}\tau)}. \end{aligned}$$

In our BMI data application, there are five cells and we use model  $M_2$  denote the model with order restrictions and its mode is at the second position, which is normal weight. Model  $M_3$  denote the model with order restrictions and its mode is at the third position, which is overweight weight. Here,  $M_2$  and  $M_3$  are the similar models but not the same.

The joint posterior distribution of  $M_2$  or  $M_3$  is

$$\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n} \propto \prod_{i=1}^{\ell} \left\{ \prod_{j=1}^K \theta_{ij}^{n_{ij}} \frac{1}{D(\boldsymbol{\mu}\tau)C(\boldsymbol{\mu}\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} \right\} \frac{(K(m-1)!(K-m)!I_{C_{\boldsymbol{\mu}}})}{(1+\tau)^2},$$

where  $D(\boldsymbol{\mu}\tau) = \frac{\prod_{j=1}^K \Gamma(\mu_j \tau)}{\Gamma(\sum_{j=1}^K \mu_j \tau)}$  is the normalization constant of Dirichlet distribution,  $C(\boldsymbol{\mu}\tau) = \int_{\boldsymbol{\theta}_i \in C} \frac{\Gamma(\sum_{j=1}^K \mu_j \tau)}{\prod_{j=1}^K \Gamma(\mu_j \tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} d\boldsymbol{\theta}_i$ , is the normalization constant of the truncated Dirichlet distribution,  $\boldsymbol{\theta} \in C, \boldsymbol{\mu} \in C_{\boldsymbol{\mu}}$  and  $I_{C_{\boldsymbol{\mu}}}$  is an indicator function that the constraint that  $\boldsymbol{\mu} \in C_{\boldsymbol{\mu}}$  is satisfied.

In Appendix 2.6.3, we use a gridy Gibbs sampler to draw samples from the joint posterior distribution for parameters  $\mu$ ,  $\tau$ . In Appendix 2.6.4, we draw samples of  $\theta$  from the conditional posterior distribution and make statistical inference on those posterior samples.

## 2.4 Simulation Under Six Scenarios

In this section, we run 100 simulations for each model under six scenarios to compare models' order restriction assumptions. We generate tables of true cell probabilities from a Dirichlet distribution with parameter  $\mu$  and  $\tau$ ,  $\mu$  is the expectation value of cell probabilities  $\theta_i$  in each area  $i$ . Small value of  $\tau$  in the Dirichlet means larger variances or covariances. In other word, small value  $\tau$  means larger heterogeneity between samples of Dirichlet distribution.

In the first three scenarios, we let  $\mu = (0.1, 0.2, 0.4, 0.2, 0.1)$ , where the unimodal order restriction may hold and the mode is at third position. In the last three scenarios, we let  $\mu = (0.2, 0.2, 0.2, 0.2, 0.2)$ , where the unimodal order restriction does not hold. We consider three categories of heterogeneity, which  $\tau = (50, 300, 1000)$  represents relatively large heterogeneity, moderate heterogeneity and small heterogeneity. For the size of each area, we randomly pick  $n_i$  from 60 to 150.

We use the following model comparison criteria,

- Relative bias, which is the difference between posterior mean and the true cell probabilities over the posterior mean, is  $\sum_{i=1}^{\ell} \sum_{j=1}^K \frac{|\widehat{\theta}_{ij} - \theta_{ij}^{true}|}{\theta_{ij}^{true}} / (\ell * K)$ , where  $\theta_{ij}^{true}$  are true cell probabilities and  $\widehat{\theta}_{ij}$  are the posterior means. Smaller is better.
- Root Mean Square Error (RMSE) is  $\sqrt{bias^2 + PSD^2}$ , where  $bias = \widehat{\theta}_{ij} - \theta_{ij}^{true}$

and PSD is the posterior standard deviation of samples. Smaller is better.

- Credible interval coverage rate, which is the true cell probabilities locate in the credible intervals. Larger coverage rate is better.
- Credible interval length, and shorter length is better.

Table 2.1: Simulation summary for four models under three scenarios

Criteria	Models	Scenario					
		$\mu = (0.2, 0.2, 0.2, 0.2, 0.2)$			$\mu = (0.1, 0.2, 0.4, 0.2, 0.1)$		
		$\tau = 50$	$\tau = 500$	$\tau = 1000$	$\tau = 50$	$\tau = 500$	$\tau = 1000$
Relative Bias	$M_1$	0.129	0.078	0.073	0.123	0.107	0.106
	$M_\mu$	0.131	0.082	0.080	0.126	0.104	0.104
	$M_\theta$	0.223	0.090	0.084	0.134	0.099	0.089
	$M_3$	0.225	0.131	0.124	0.140	0.088	0.077
RMSE	$M_1$	0.043	0.032	0.031	0.034	0.033	0.033
	$M_\mu$	0.044	0.035	0.037	0.035	0.033	0.033
	$M_\theta$	0.038	0.032	0.028	0.038	0.031	0.026
	$M_3$	0.038	0.029	0.027	0.048	0.032	0.028
CI Coverage	$M_1$	0.962	0.987	0.992	0.960	0.974	0.978
	$M_\mu$	0.950	0.979	0.985	0.960	0.980	0.980
	$M_\theta$	0.588	0.781	0.845	0.951	0.981	0.985
	$M_3$	0.627	0.784	0.797	0.949	0.984	0.993
CI Length	$M_1$	0.111	0.104	0.101	0.103	0.102	0.101
	$M_\mu$	0.104	0.098	0.100	0.103	0.095	0.095
	$M_\theta$	0.085	0.072	0.079	0.110	0.095	0.096
	$M_3$	0.098	0.086	0.085	0.111	0.093	0.089

Notation:  $M_1$  is the multinomial-Dirichlet model without order restrictions,

$M_\mu$  is the multinomial-Dirichlet model with order restrictions on  $\mu$ ,

$M_\theta$  is the multinomial-Dirichlet model with order restrictions on  $\theta$ ,

$M_3$  is the multinomial-Dirichlet model with order restrictions on  $\mu$  and  $\theta$ .

When  $\mu = (0.2, 0.2, 0.2, 0.2, 0.2)$  and  $\tau = 50$ , the unimodal order restriction

may not hold and the heterogeneity between areas is large. The model with order restrictions,  $M_\theta$  and  $M_3$ , have large relative bias and small credible interval coverage. When  $\mu = (0.1, 0.2, 0.4, 0.2, 0.1)$  and  $\tau = 1000$ , the unimodal order restriction may hold and the heterogeneity between areas is small. The model with order restrictions,  $M_\theta$  and  $M_3$ , have very small relative bias and very high credible interval coverage.

The model with order restrictions on  $\mu$ ,  $M_\mu$ , has similar performance to the model without order restriction,  $M_1$ . One possible explanation is that the order restriction assumption in  $M_1$  is weak. It will not force the categorical cell probabilities  $\theta_i$  to have the unimodal order restrictions. The model with order restrictions on  $\mu$  may not borrow much more information among areas than the model without order restrictions. There is not significant difference between  $M_1$  and  $M_\mu$ .

The model with order restrictions on  $\mu$  and  $\theta$ ,  $M_3$ , has smaller relative bias than the model with order restrictions on  $\mu$ ,  $M_\mu$ , when the heterogeneity between areas is small and order restriction assumption may hold.  $M_3$  has larger credible interval coverage than  $M_\mu$  when order restriction assumption may hold. It is consistent with the conclusion that the order restriction on  $\mu$  is weak and borrowing information is not much different. We want to maximize the advantage of incorporating order restrictions into the model to have small relative bias and large credible interval coverage. The model with order restriction on  $\mu$  and  $\theta$ ,  $M_3$ , is preferred for the further study.

With increasing values of  $\tau$ , the unimodal order restriction holds across areas and the heterogeneity between areas is getting smaller. The model with order restrictions on parameters  $\mu$  and  $\theta$  has smaller relative bias, RMSE, and higher credible interval coverage. Overall, if the order restriction assumption may hold and

the heterogeneity across areas is relatively small, the model with order restrictions,  $M_3$ , is the best model we should use.

## 2.5 Concluding Remarks

From the outcome of the simulation, the multinomial-Dirichlet model with order restrictions on parameters  $\mu$  and  $\theta$ ,  $M_3$ , shows that the model can borrow more information than the hierarchical model without any order restriction. The model with order restriction on hyperparameter  $\mu$  does not have much difference than the model without any order restriction. If the order restriction assumption holds on the data we have, we should consider the model with order restrictions to have a better estimates on parameters. The model with order restrictions on  $\theta$  only will not make computation easier than  $M_3$ . If the order restriction assumption holds for all areas and heterogeneity is small,  $M_3$  is better than other models. If the order restriction assumption may not hold, the model with order restrictions will have higher relative bias than the model without order restriction. In practice, it is not easy to verify the unimodal order restrictions in the data and decide the correct model we should use. It is worth fitting the model with order restrictions in the data where the order restriction may hold.

## 2.6 Appendix

### 2.6.1 Griddy Gibbs Sampler for $\mu, \tau$ in $M_\mu$

In this section we discuss how to obtain samples from the joint posterior density for each parameter and make statistical inference on those posterior samples. The joint posterior distribution of  $M_\mu$  is

$$\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n} \propto \prod_{i=1}^I \left\{ \prod_{j=1}^K \theta_{ij}^{n_{ij}} \frac{1}{D(\boldsymbol{\mu}\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} \right\} \frac{K(m-1)!(K-m)! I_{C_\mu}}{(1+\tau)^2},$$

where  $D(\boldsymbol{\mu}\tau) = \frac{\prod_{j=1}^K \Gamma(\mu_j \tau)}{\Gamma(\sum_{j=1}^K \mu_j \tau)}$  is the normalization constant of Dirichlet distribution and  $I_{C_\mu}$  is an indicator function that the constraint that  $\mu \in C_\mu$  is satisfied.

Obviously, the conditional posterior distribution of  $\boldsymbol{\theta}_i$  is Dirichlet distribution with parameter  $\mathbf{n}_i + \boldsymbol{\mu}\tau$ ,

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, \mathbf{n} \stackrel{ind}{\sim} \text{Dirichlet}(\mathbf{n}_i + \boldsymbol{\mu}\tau).$$

After integrating out  $\boldsymbol{\theta}$ , the conditional posterior distribution of  $\boldsymbol{\mu}, \tau$  is

$$\pi(\boldsymbol{\mu}, \tau | \mathbf{n}) \propto \prod_{i=1}^I \left\{ \frac{D(\mathbf{n}_i + \boldsymbol{\mu}\tau)}{D(\boldsymbol{\mu}\tau)} \right\} \frac{K(m-1)!(K-m)! I_{C_\mu}}{(1+\tau)^2}.$$

We use a griddy Gibbs sampler to draw the posterior samples of  $\boldsymbol{\mu}, \tau$ . Specifically, we use 200 points in the uniformly spaced grid. Since  $\boldsymbol{\mu} \in C_\mu = \{\boldsymbol{\mu} = (\mu_1, \dots, \mu_K) : \mu_1 \leq \dots \leq \mu_m \geq \dots \geq \mu_K, 0 \leq \mu_j \leq 1\}$ , and for  $j$  from  $m-1$

to 1 the range of  $\mu_j$  is

$$0 < \mu_j \leq \min\left\{\mu_{j+1}, \frac{1 - \sum_{t=1, t \neq m, t \neq j}^K \mu_t}{2}\right\},$$

and for  $j$  from  $m + 1$  to  $K$ ,

$$0 < \mu_j \leq \min\left\{\mu_{j-1}, \frac{1 - \sum_{t=1, t \neq m, t \neq j}^K \mu_t}{2}\right\}.$$

## 2.6.2 Griddy Gibbs Sampler for $\mu$ and $\tau$ in $M_\theta$

In this section we discuss how to obtain samples from the joint posterior density for each parameter and make statistical inference on those posterior samples.

The joint posterior density is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n}) \propto \prod_{i=1}^I \left\{ \prod_{j=1}^K \theta_{ij}^{n_{ij}} \frac{1}{D(\boldsymbol{\mu}\tau)C(\boldsymbol{\mu}\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j\tau-1} \right\} \frac{(K-1)!}{(1+\tau)^2},$$

where  $D(\boldsymbol{\mu}\tau) = \frac{\prod_{j=1}^K \Gamma(\mu_j\tau)}{\Gamma(\sum_{j=1}^K \mu_j\tau)}$  is the normalization constant of Dirichlet distribution,

$$C(\boldsymbol{\mu}\tau) = \int_{\boldsymbol{\theta}_i \in C} \frac{\Gamma(\sum_{j=1}^K \mu_j\tau)}{\prod_{j=1}^K \Gamma(\mu_j\tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j\tau-1} d\boldsymbol{\theta}_i$$

is the normalization constant of the truncated Dirichlet distribution.

Obviously, the conditional posterior distribution of  $\boldsymbol{\theta}_i$  is Dirichlet distribution with parameter  $\mathbf{n}_i + \boldsymbol{\mu}\tau$  and  $\boldsymbol{\theta} \in C$ ,

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, \mathbf{n} \sim \text{Dirichlet}(\mathbf{n}_i + \boldsymbol{\mu}\tau), \quad \boldsymbol{\theta}_i \in C, \quad i = 1, \dots, I,$$

where  $C = \{\boldsymbol{\theta}_i : \theta_{i1} \leq \dots \leq \theta_{im} \geq \dots \geq \theta_{iK}, i = 1, \dots, I\}$ , and assume the

modal position  $m$  in  $C$  is known.

After integrating out  $\theta$ , the conditional posterior distribution of  $\mu, \tau$  is

$$\begin{aligned}\pi(\mu, \tau | \mathbf{n}) &\propto \prod_{i=1}^I \left\{ \frac{D(\mathbf{n}_i + \mu\tau)C(\mathbf{n}_i + \mu\tau)}{D(\mu\tau)C(\mu\tau)} \right\} \frac{(K-1)!}{(1+\tau)^2} \\ &\propto \prod_{i=1}^I \left\{ \frac{\int_{\theta_i \in C_i} \prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j\tau - 1} d\theta_i}{\int_{\theta_i \in C_i} \prod_{j=1}^K \theta_{ij}^{\mu_j\tau - 1} d\theta_i} \right\} \frac{(K-1)!}{(1+\tau)^2}.\end{aligned}$$

We use a griddy Gibbs sampling method to draw the posterior samples of  $\mu, \tau$ . Specifically, we use 200 points in the uniformly spaced grid in each interval. Due to the ratio of the normalization constants with order restrictions, we propose a Monte Carlo importance sampling method to deal with the ratio, shown as below.

Draw samples of  $\theta$ :  $\theta^{(q)} \sim \text{Dirichlet}(\gamma\bar{\mathbf{n}}_j)$ , where  $\theta \in C_i$  and  $\gamma$  is the importance ratio,  $\bar{\mathbf{n}}_j$  is the average over small areas for each category, the approximate joint posterior distribution of  $\mu$  and  $\tau$  is

$$\begin{aligned}\pi(\mu, \tau | \mathbf{n}) &\propto \prod_{i=1}^I \left\{ \frac{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij} - \gamma\bar{\mathbf{n}}_j + \mu_j\tau}}{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma\bar{\mathbf{n}}_j + \mu_j\tau}} \right\} \frac{(K-1)!}{(1+\tau)^2} \\ &\propto \prod_{i=1}^I \left\{ \sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij}} \frac{\prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma\bar{\mathbf{n}}_j + \mu_j\tau}}{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma\bar{\mathbf{n}}_j + \mu_j\tau}} \right\} \frac{(K-1)!}{(1+\tau)^2} \\ &\propto \prod_{i=1}^I \left\{ \sum_{q=1}^M w_q \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij}} \right\} \frac{(K-1)!}{(1+\tau)^2}, \text{ where } w_q = \frac{\prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma\bar{\mathbf{n}}_j + \mu_j\tau}}{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma\bar{\mathbf{n}}_j + \mu_j\tau}}.\end{aligned}$$

Since there is no order restriction about  $\mu$ , the range of  $\mu_j$  is  $0 \leq \mu_j \leq \sum_{t=1, t \neq j}^{K-1} \mu_t$  for  $j = 1, \dots, K-1$ .

### 2.6.3 Griddy Gibbs Sampler for $\mu$ and $\tau$ in $M_2$ and $M_3$

In  $M_2$  and  $M_3$ , we use a griddy Gibbs sampler, a Markov chain Monte Carlo (MCMC) algorithm, for  $\mu \in C_\mu$  and  $\tau$ . Specifically, we use 200 points in the



uniformly spaced grid in each interval. The joint posterior density is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n}) \propto \prod_{i=1}^I \left\{ \frac{\prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1} I_C I_{C_\mu}}{D(\boldsymbol{\mu}\tau)C(\boldsymbol{\mu}\tau)} \right\} \frac{I_{C_\mu}}{(1 + \tau)^2},$$

where

$$C(\boldsymbol{\mu}\tau) = \int_{\boldsymbol{\theta}_i \in C} \frac{\Gamma(\sum_{j=1}^K \mu_j \tau)}{\prod_{j=1}^K \Gamma(\mu_j \tau)} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} d\boldsymbol{\theta}_i,$$

and  $I_{C_\mu}$  is an indicator function that the constraint that  $\boldsymbol{\mu} \in C_\mu$  is satisfied.

There is no recognizable conditional distribution of  $\boldsymbol{\mu}$  and  $\tau$  to generate samples. So we use a grid method to draw  $\boldsymbol{\mu}$  and  $\tau$  from  $\pi(\boldsymbol{\mu}, \tau | \mathbf{n})$  after integrating with respect to  $\boldsymbol{\theta}$ , we get

$$\begin{aligned} \pi(\boldsymbol{\mu}, \tau | \mathbf{n}) &\propto \prod_{i=1}^I \left\{ \frac{D(\boldsymbol{\mu}\tau + \mathbf{n}_i)C(\boldsymbol{\mu}\tau + \mathbf{n}_i)}{D(\boldsymbol{\mu}\tau)C(\boldsymbol{\mu}\tau)} \right\} \frac{I_{C_\mu}}{(1 + \tau)^2} \\ &\propto \prod_{i=1}^I \left\{ \frac{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau + n_{ij} - 1} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} d\boldsymbol{\theta}_i} \right\} \frac{I_{C_\mu}}{(1 + \tau)^2}. \end{aligned}$$

Chen and Shao (1997) mentioned that importance sampling could be used to estimate the ratio,  $\frac{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau + n_{ij} - 1} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} d\boldsymbol{\theta}_i}$ . We consider Dirichlet( $r\bar{n}_j$ ) as our importance function for all counties, where  $r$  is an adjustable ratio and  $\bar{n}_j = \frac{\sum_{i=1}^I n_{ij}}{I}$ .

It combines information together. Since our importance function does not depend on the unknown parameters,  $\boldsymbol{\mu}$  and  $\tau$ , we can generate one set of numbers for all iterations. In our late application, it has been proved as an efficient way to generate posterior samples.

**Algorithm 1:**

1. Draw  $\tau$  from  $\pi(\tau|\boldsymbol{\mu}, \mathbf{n})$ ;
2. For  $j$  from  $m-1$  to  $1$ , draw  $\mu_j$  from  $\pi(\mu_j|\boldsymbol{\mu}^{(-j)}, \tau, \mathbf{n})$ , where
 
$$0 < \mu_j < \min\left\{\mu_{j+1}, \frac{1 - \sum_{t=1, t \neq m, t \neq j}^K \mu_t}{2}\right\};$$
3. For  $j$  from  $m+1$  to  $K$ , draw  $\mu_j$  from  $\pi(\mu_j|\boldsymbol{\mu}^{(-j)}, \tau, \mathbf{n})$ , where
 
$$0 < \mu_j < \min\left\{\mu_{j-1}, \frac{1 - \sum_{t=1, t \neq m, t \neq j}^K \mu_t}{2}\right\};$$
4. Get  $\mu_m = 1 - \sum_{j=1, j \neq m}^K \mu_j$ , repeat Step 1 to Step 4 to get converged MCMC samples,  $\boldsymbol{\mu}^{(-j)} = (\mu_1, \dots, \mu_{j-1}, \mu_{j+1}, \dots, \mu_K)$ .

Notice that the upper end of an interval can be small and 200 grids might be an overkill.

#### 2.6.4 Sampling from Conditional Posterior Distribution of $\boldsymbol{\theta}$ in $M_\theta$ , $M_2$ , and $M_3$

In models  $M_\theta$ ,  $M_2$ , and  $M_3$ , the posterior of  $\boldsymbol{\theta}$  has a recognizable distribution, which is the Dirichlet distribution with the order restriction. Instead of drawing samples directly from the Dirichlet distribution with the order restriction, Chen and Nandram (2019) present a direct sampling from truncated Gamma distributions, where Nadarajah and Kotz (2006) offered a method for truncated Gamma.

**Algorithm 2 :** To draw  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$ ,  $\boldsymbol{\theta} \in C$ ,

denote  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ ,

If  $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_m \geq \dots \geq \theta_K$ , the mode is  $\theta_m$ ,

$0 \leq \beta_1 \leq \beta_2 \leq \dots \leq \beta_m \geq \dots \geq \beta_K$ , the mode is  $\beta_m$ ,

1. Draw  $\beta_m \sim \text{Gamma}(\alpha_m, 1)$ , where  $0 \leq \beta_m < \infty$ ,
2. Draw  $\beta_{m-1} \sim \text{Truncated Gamma}(\alpha_{m-1}, 1)$ , where  $0 \leq \beta_{m-1} \leq \beta_m$ ,

...  $\beta_1 \sim \text{Truncated Gamma}(\alpha_1, 1)$ , where  $0 \leq \beta_1 \leq \beta_2$ ,

3. Draw  $\beta_{m+1} \sim \text{Truncated Gamma}(\alpha_{m+1}, 1)$ , where  $0 \leq \beta_{m+1} \leq \beta_m$ ,

...  $\beta_K \sim \text{Truncated Gamma}(\alpha_K, 1)$ , where  $0 \leq \beta_K \leq \beta_{K-1}$ .

Then,

$$\theta_1 = \frac{\beta_1}{\beta_1 + \dots + \beta_K}, \dots, \theta_{K-1} = \frac{\beta_{K-1}}{\beta_1 + \dots + \beta_K}, \theta_K = 1 - \sum_{i=1}^{K-1} \theta_i.$$

Therefore, to draw samples from the conditional posterior distribution of  $\theta$  in  $M_\theta$ ,  $M_2$ , and  $M_3$ , we draw samples of  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ , where  $0 \leq \beta_1 \leq \beta_2 \leq \dots \leq \beta_m \geq \dots \geq \beta_K$ . Each  $\beta_j$  follows a truncated Gamma distribution with parameter  $\alpha_j$  and 1. Then we can get the posterior samples of  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ .

## Chapter 3

# Uncertainty about Order Restrictions

The same unimodal order restriction for all areas may be too strong to be true for some cases. To increase flexibility, we add uncertainty to the unimodal order restriction. Each area will have similar unimodal patterns, but not the same. We use the same notations for model  $M_1$ ,  $M_2$ ,  $M_3$  in Chapter 2 for the following discussion.

### 3.1 The Model Incorporated Uncertainty

Chen and Nandram (2021) consider adding uncertainty to the model to increase the robustness and flexibility. Let  $L_{pos}$  be the mode position of cell probabilities. Thus our new hierarchical multinomial-Dirichlet model, denoted as  $M_4$ , is given as below,

$$\mathbf{n}_i | \boldsymbol{\theta}_i, L_{pos} = \ell \stackrel{ind}{\sim} \text{Multinomial}(\mathbf{n}_i, \boldsymbol{\theta}_i), \quad i = 1, \dots, I, \quad \ell = 1, \dots, K,$$

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, L_{pos} = \ell \stackrel{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), i = 1, \dots, I, \boldsymbol{\theta}_i \in C_\ell,$$

$$\pi(\boldsymbol{\mu}, \tau | L_{pos} = \ell) = \frac{K(m_\ell - 1)!(K - m_\ell)!}{(1 + \tau)^2}, \mu_j > 0, \sum_{j=1}^K \mu_j = 1, \boldsymbol{\mu} \in C_{\boldsymbol{\mu}_\ell},$$

where

$$C_\ell = \{\boldsymbol{\theta}_i : \theta_{i1} \leq \dots \leq \theta_{im_\ell} \geq \dots \geq \theta_{iK}\},$$

$$C_{\boldsymbol{\mu}_\ell} = \{\boldsymbol{\mu} : \mu_1 \leq \dots \leq \mu_{m_\ell} \geq \dots \geq \mu_K\},$$

and

$$P(L_{pos} = \ell) = w_\ell, \sum_{\ell=1}^K w_\ell = 1, \ell = 1, \dots, K.$$

Modes are the same for all areas but we are uncertain about where they are.

Then the joint posterior distribution is

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \tau | \mathbf{n}) &\propto \sum_{L_{pos}=1}^K w_{L_{pos}} \prod_{i=1}^I \left\{ \prod_{j=1}^K \theta_{ij}^{n_{ij}} \frac{\prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} I_{C_{L_{pos}}} I_{C_{\boldsymbol{\mu}_{L_{pos}}}}}{D(\boldsymbol{\mu}\tau) C(\boldsymbol{\mu}\tau)} \right\} \frac{1}{(1 + \tau)^2} \\ &\propto \sum_{L_{pos}=1}^K w_{L_{pos}} \prod_{i=1}^I \left\{ \frac{\prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1} I_{C_{L_{pos}}} I_{C_{\boldsymbol{\mu}_{L_{pos}}}}}{D(\boldsymbol{\mu}\tau) C(\boldsymbol{\mu}\tau)} \right\} \frac{1}{(1 + \tau)^2}, \end{aligned}$$

where  $I_{C_{L_{pos}}}$  and  $I_{C_{\boldsymbol{\mu}_{L_{pos}}}}$  are the indicator functions under that order restriction.

In Appendix 3.5.1, we present an approximation approach to estimate the posterior of  $P(L_{pos} = \ell)$  and obtain the posterior samples of parameters  $\boldsymbol{\theta}$ .

Chen and Nandram (2020a) and Chen and Nandram (2020c) presented a method to compute the conditional predictive ordinate (CPO) and the log-pseudo marginal likelihood (LPML) as Bayesian model selection criteria. We present CPO and LPML of  $M_2$ ,  $M_3$ , and  $M_4$  in Appendix 3.5.2.

CPO of  $M_2$  and  $M_3$ (similar to  $M_2$ ) are

$$\widehat{CPO}_{i(M_2)} = \left[ \frac{1}{M} \sum_{h=1}^M \frac{\prod_{j=1}^K n_{ij}!}{n_i!} \int_{\theta_i \in C} \frac{\prod_{j=1}^K \theta_{ij}^{\mu(h)\tau(h)-1}}{\prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu(h)\tau(h)-1}} \frac{\prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu(h)\tau(h)-1}}{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu(h)\tau(h)-1} d\theta_i} d\theta_i \right]^{-1},$$

where  $\frac{\prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu(h)\tau(h)-1}}{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu(h)\tau(h)-1} d\theta_i}$  is the density function of  $\theta_i$ , and  $\theta_i \in C$ .

Then we get the LPML of  $M_2$  and  $M_3$  as  $LPML_{M_2} = \sum_{i=1}^I \log(\widehat{CPO}_{iM_2})$ .

and  $LPML_{M_3} = \sum_{i=1}^I \log(\widehat{CPO}_{iM_3})$ .

CPO of  $M_4$  is

$$CPO_{i(M_4)} \approx \left[ \sum_{\ell=1}^K P(\widehat{L} = \ell | n) \frac{1}{\widehat{CPO}_{i(M_L|L=\ell)}} \right]^{-1},$$

where  $\widehat{CPO}_{i(M_L|L=\ell)}$  are known, such as  $\widehat{CPO}_{i(M_2)}$  and  $\widehat{CPO}_{i(M_3)}$ . Without extra computation, taking advantage of known CPOs from  $M_2$  and  $M_3$ , and the estimated  $P(L = \ell | n)$ , we can easily acquire the CPO of  $M_4$ .

Then we get the LPML as  $LPML_{M_4} = \sum_{i=1}^I \log(\widehat{CPO}_{iM_4})$ .

We notice the marginal likelihood of the model with order restrictions is

$$f(\mathbf{n} | M_2 \text{ or } M_3) = \int_{\mu, \tau} \prod_{i=1}^I \left\{ \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu_j\tau-1} d\theta_i}{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j\tau-1} d\theta_i} \right\} \frac{K(m-1)!(K-m)!}{(1+\tau)^2} d\mu d\tau.$$

We cannot utilize the posterior samples to compute Bayes factor directly. Hence CPOs as Bayesian diagnostic are preferred.

## 3.2 Numerical Example: Body Mass Index

### 3.2.1 Body Mass Index

In our application, we use a selected subset of the body mass index (BMI) data from NHANES III, where we use only the female BMI data from the 35 largest counties with a population at least 500,000. For counties with large population, we assume people randomly fall into five BMI categorical levels, which are underweight, normal, overweight, obese I, and obese II. Our goal is to estimate the proportions of the BMI levels. Table 3.1 gives an illustration of the female BMI data of a few counties, where it can be seen that the cell probability is largest for the normal range and other probabilities roughly tail off on both sides suggest the unimodal order restriction. Indeed, there are violations in some counties in the earliest and latest cells.

County ID	BMI_lvl1	BMI_lvl2	BMI_lvl3	BMI_lvl4	BMI_lvl5
1	3	40	37	13	4
2	1	36	38	15	1
3	3	20	49	13	5
⋮	⋮	⋮	⋮	⋮	⋮
35	1	41	41	9	0
Total	45	1201	1318	496	89

Table 3.1: Number of females at five levels of BMI

For each county, the BMI counts can be assumed to follow a multinomial distribution because each individual person can be assumed to exist independently. Figure 3.1 shows a histogram of all BMI values for females aggregated into a single large sample. It can be clearly seen that the unimodal order restriction holds. Because the data in the individual counties are generally sparse, it is difficult to

tell whether the unimodal order restriction holds. However, it is sensible to assume that the same unimodal restriction holds within all the counties. Therefore, we can use multinomial distributions to model the female BMI counts.

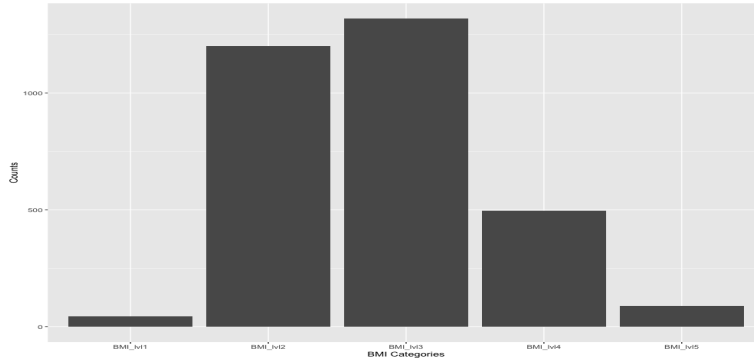


Figure 3.1: Overall females BMI in five categories

### 3.2.2 MCMC Convergence Diagnostics

To be convenient, Table 3.2 provides the model notations used in the following discussion.

Table 3.2: Model notations in Chapter 3

Notation	Model	Order Restrictions	Location
$M_1$	the multinomial-Dirichlet model	NA	Chapter 2
$M_2$ (Normal)	the multinomial-Dirichlet model with order restrictions	$\theta_1 \leq \theta_2 \geq \theta_3 \geq \theta_4 \geq \theta_5$	Section 2.3
$M_3$ (Overweight)	the multinomial-Dirichlet model with order restrictions	$\theta_1 \leq \theta_2 \leq \theta_3 \geq \theta_4 \geq \theta_5$	Section 2.3
$M_4$	the multinomial-Dirichlet EXNEX model with uncertain order restrictions	Uncertainty about the modal position	Section 3.1

For each model, we run 20,000 MCMC iterations, take 10,000 as a ‘burn in’ and use every 10th to obtain 1,000 converged posterior samples to maintain consistency. Figure 3.2 shows trace plots of posterior samples of  $\mu$ . The trace plots indicate that posterior samples mix well and MCMC is stationary. Table 3.3 gives



the effective sample sizes of the parameters  $\mu$ ,  $\tau$  for the model with the order restriction and the general model. The effective sample sizes are almost 1,000. Table 3.4 provides p-values of the Geweke test to check the convergence of the parameters (Geweke et. al. 1992). All p-values are large enough to not reject the null hypothesis that the MCMC is stationary. Then posterior samples can be used for the further inference.

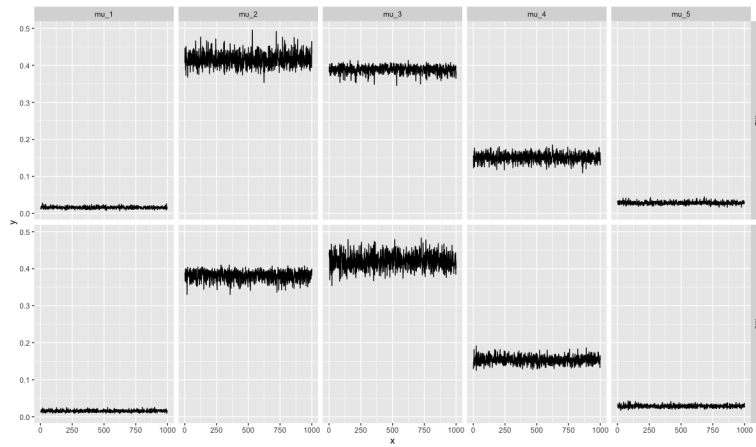


Figure 3.2: Traceplots of  $\mu$  in  $M_2$  and  $M_3$

	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\tau$
$M_1$	1000	1123.7	1000	1000	895.4	1000
$M_2$ (Mode at 2nd)	1000	1000	1000	1000	1150.2	1000
$M_3$ (Mode at 3rd)	1000	887	889	1000	1173.9	1000

Table 3.3: Effective sample sizes of  $\mu$  and  $\tau$

	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\tau$
$M_1$	0.623	0.558	0.899	0.767	0.959	0.514
$M_2$ (Mode at 2nd)	0.964	0.705	0.507	0.511	0.837	0.999
$M_3$ (Mode at 3rd)	0.817	0.559	0.580	0.557	0.812	0.516

Table 3.4: P-values of Geweke tests for  $\mu$  and  $\tau$

### 3.2.3 Model Comparison of $M_1$ , $M_2$ , $M_3$ , and $M_4$

With the approximate mixture probabilities, we mix posterior samples of  $M_2$  and  $M_3$  together to construct samples of  $M_4$ .

We provide posterior mean (PM), posterior standard deviation (PSD) and coefficient of variation (CV) of  $\theta$ s for all counties, which can be found in Appendix 3.5.3.

To compare model difference visually, we present the posterior densities plots about different counties in those models as Figure 3.3, Figure 3.4, Figure 3.5 and Figure 3.6. We use different colors to indicate five BMI levels and dashed lines for the posterior means. Due to different capability of borrowing information among areas, we can see different flatness of posterior density curves in the models. With different order restriction assumptions, those posterior density curves center at different places and may overlap differently. We mainly focus on density curves of normal BMI and overweight BMI, since the modal position might be second or third.

In Figure 3.3 has posterior density plots for County 2 applying different models. The number of observations with normal BMI level, which is 36, is close to the number of observations with overweight BMI level, which is 38. The unimodal order restriction may not hold in County 2. Maybe for this reason, there is a significant overlap between normal level and overweight level in the first plot after applying  $M_1$  to our BMI data. The second plot and the third plot show much less overlap in density curves, due to the strong order restriction assumption. The last plot, which is the density curve from  $M_4$ , is similar to the density curves in  $M_3$ . Based on the observations in County 2, the order restriction that the modal posi-

tion is at the third may be reasonable. The density curves in  $M_3$  and  $M_4$  may be appropriate for County 2.

In Figure 3.4 has posterior density plots for County 3 applying different models. Unlike in County 2, the density curves of  $\theta$  from model  $M_1$  in County 3 shows a very strong unimodality because we have 49 people in overweight BMI level which dominates this county. The second plot from  $M_2$ , which assumes that the mode is at normal BMI level, has a significant overlap. Its order restriction assumption that the modal position is at the second position may not hold in this county. The third plot from  $M_3$ , which assumes that the mode is at overweight BMI level, is similar as the density curve in  $M_1$ . The posterior mean of normal BMI level probability is higher than in  $M_1$ . This phenomenon can be considered as an evidence that  $M_3$  has a stronger borrowing ability than  $M_1$ . Overall, the modal position among 35 counties may be at the third.  $M_3$  can borrow more information among those counties than other models. Then the last plot, which is the density curve from  $M_4$ , has a little overlap. But the unimodal pattern is still in  $M_4$ .

In Figure 3.5, they are posterior density plots for County 13 applying different models. Only  $M_2$  with an assumption that the mode is at normal BMI level does not show a significant overlap. Since more people are at overweight BMI level, that assumption may be validate in County 13.

Figure 3.6 provides posterior density plots for County 35, which has almost same amount of people in normal and overweight BMI level.  $M_2$  and  $M_3$  with different unimodal assumptions have opposite conclusion about normal and overweight probabilities. In this county,  $M_1$  and  $M_4$  may be better models.

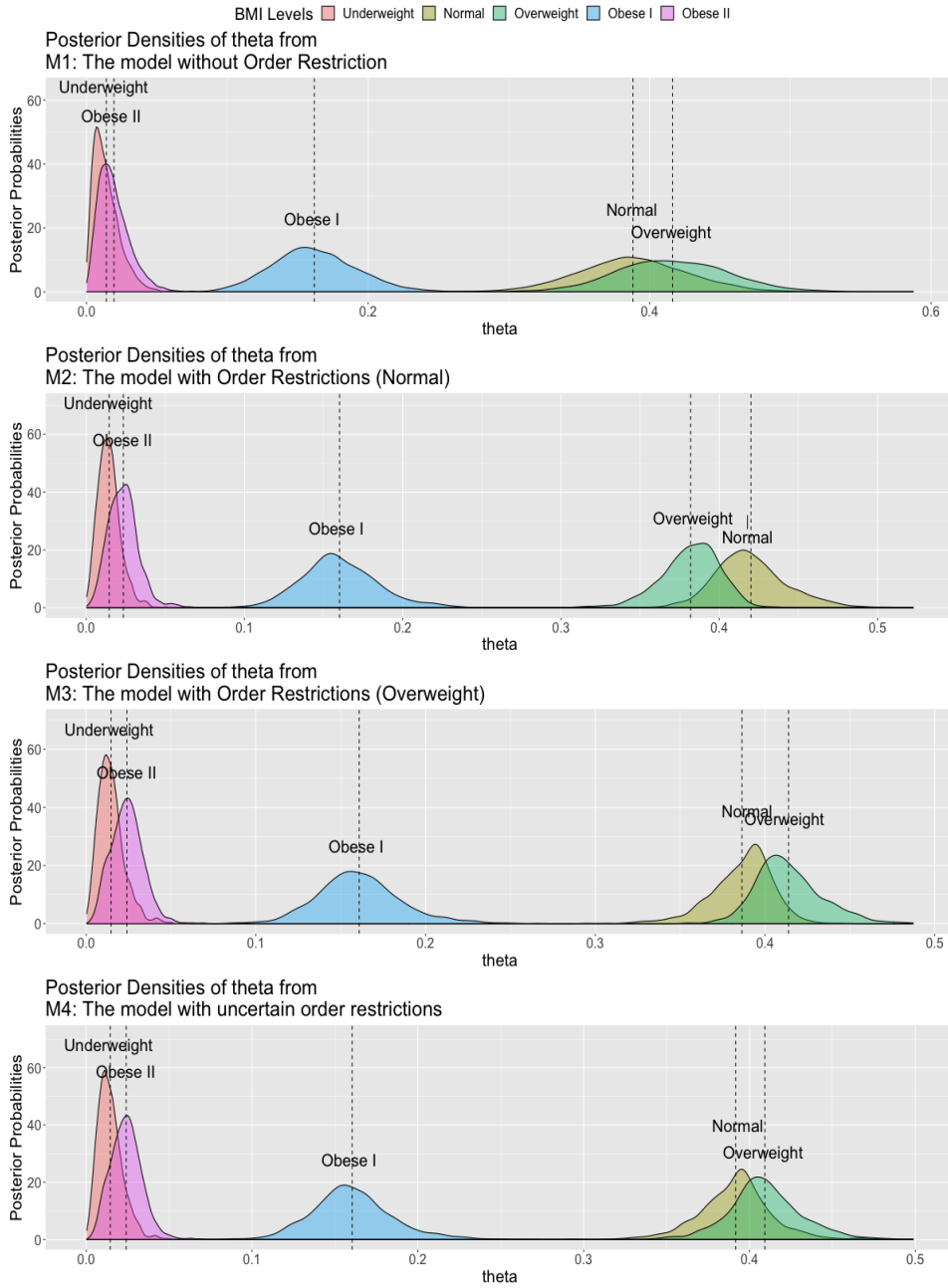


Figure 3.3: Posterior densities of  $\theta$  for County 2 showing different order restrictions under different models.

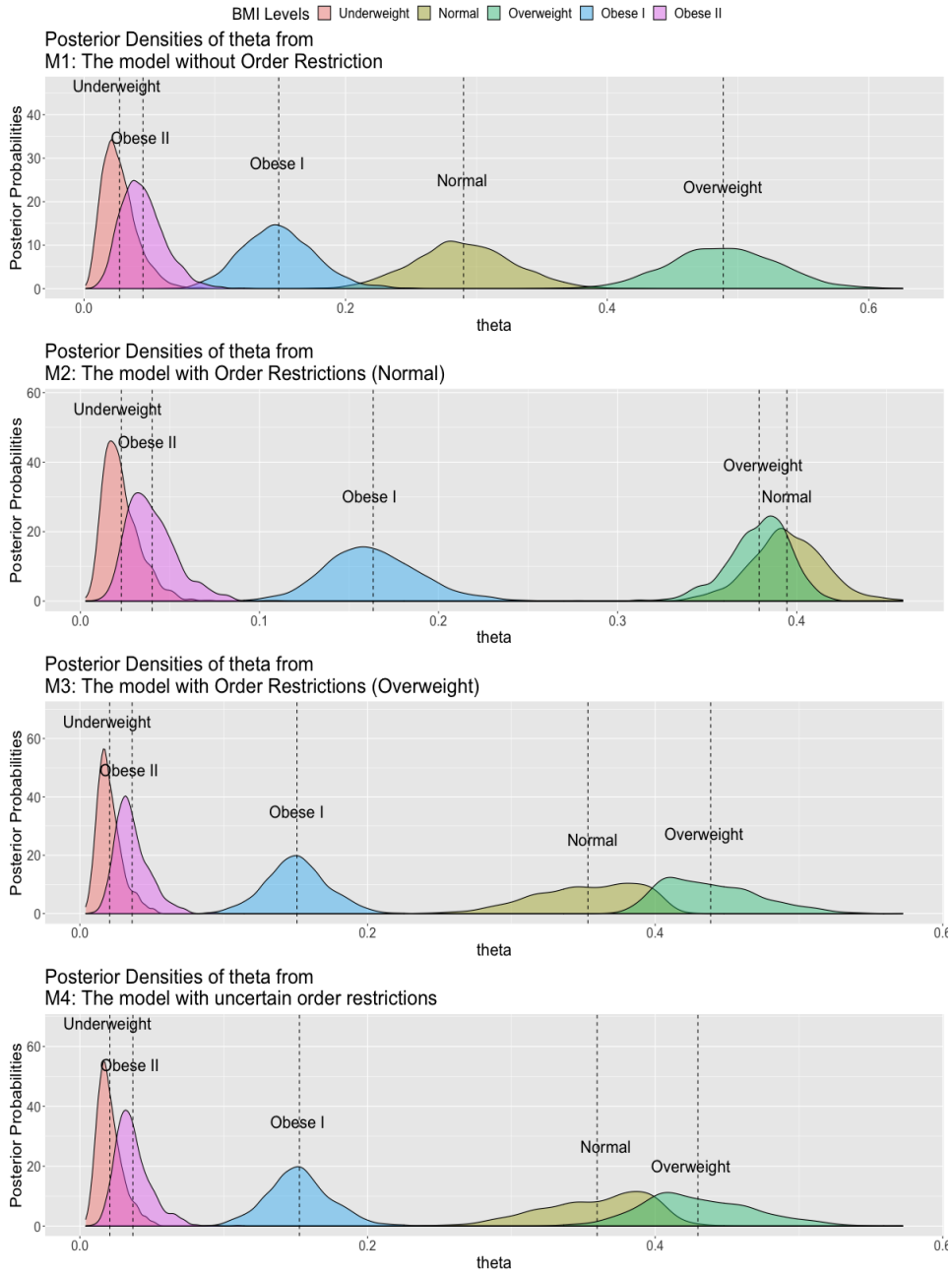


Figure 3.4: Posterior densities of  $\theta$  for County 3 showing different order restrictions under different models.

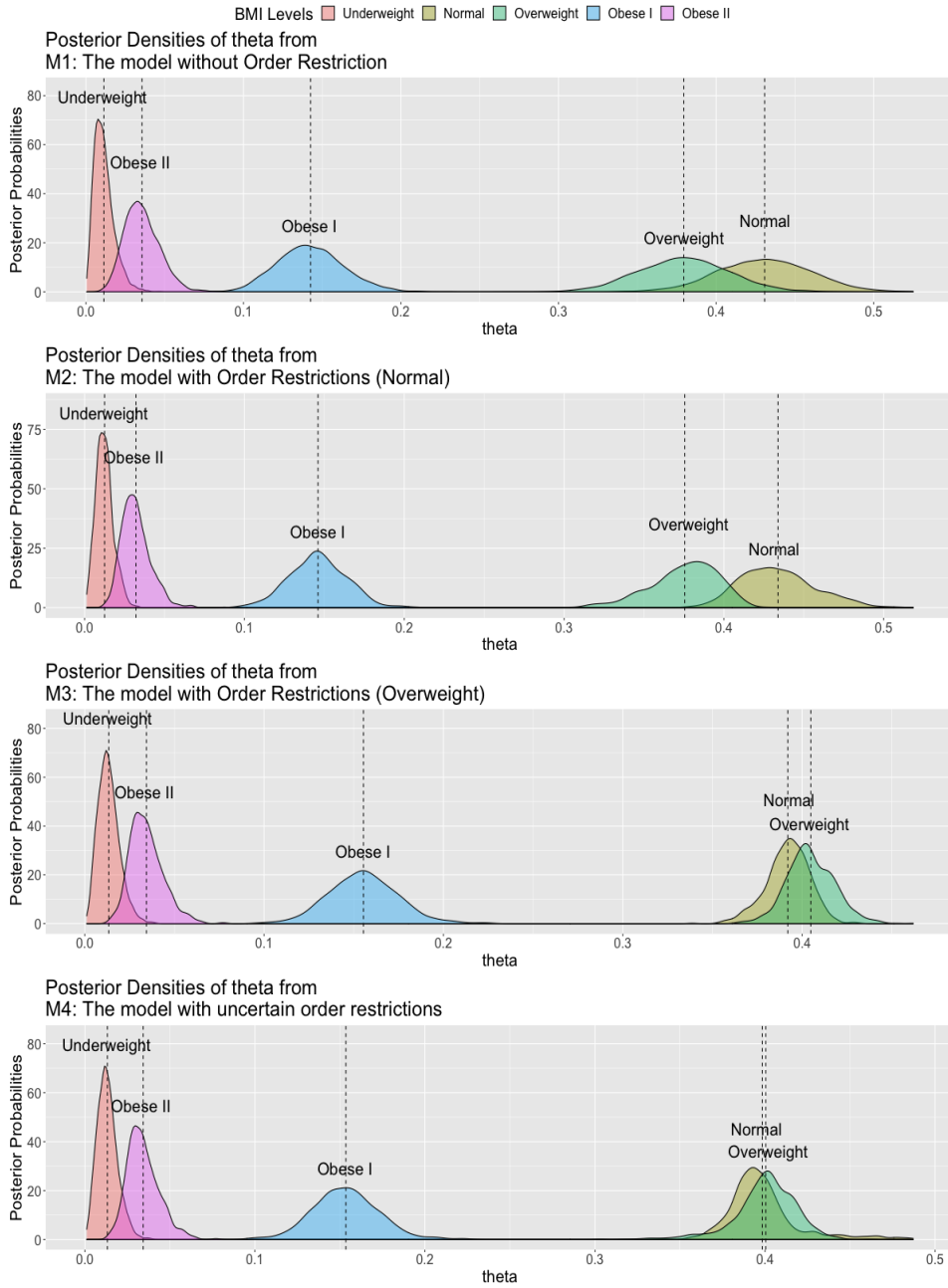


Figure 3.5: Posterior densities of  $\theta$  for County 13 showing different order restrictions under different models.

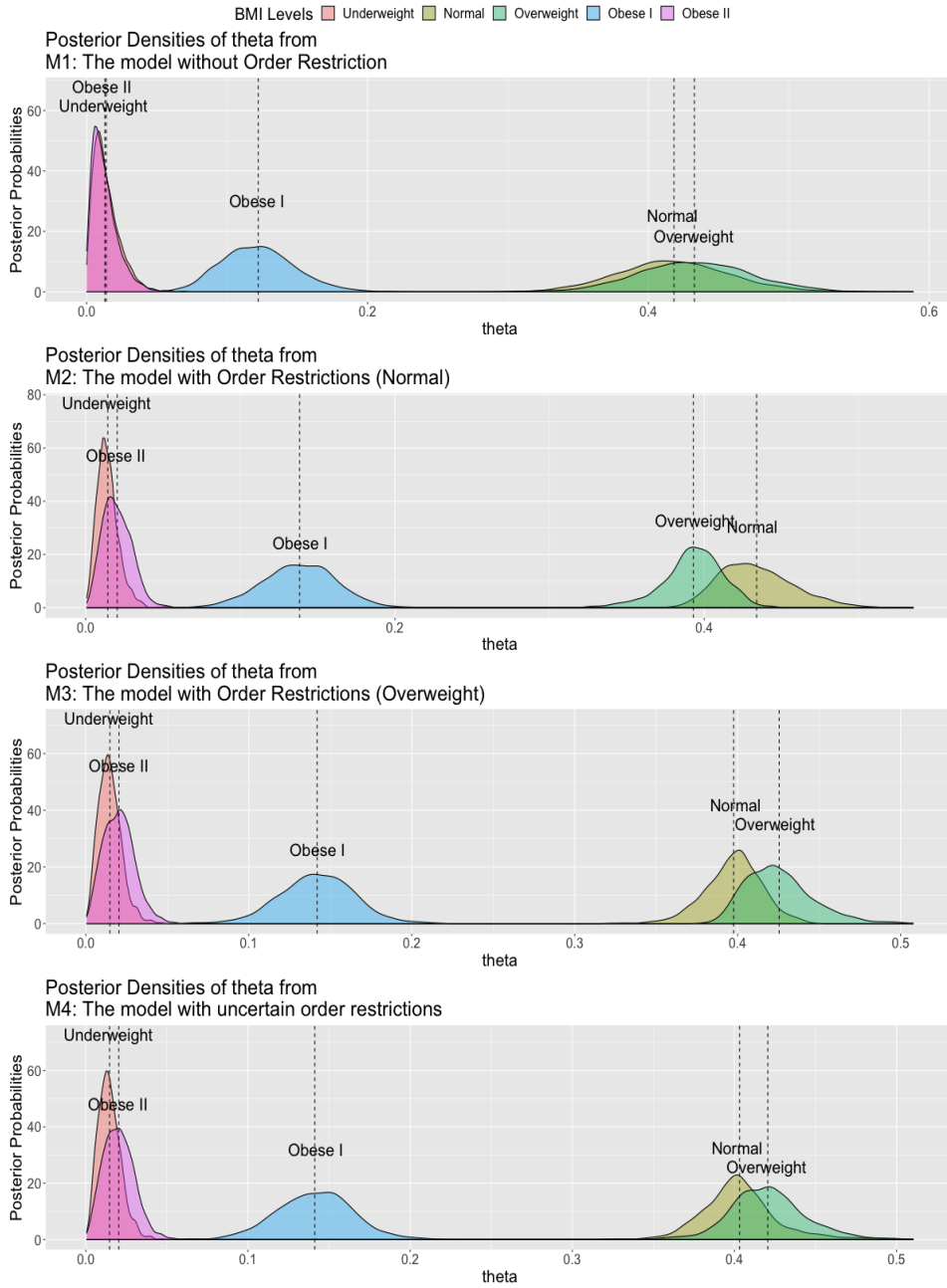


Figure 3.6: Posterior densities of  $\theta$  for County 35 showing different order restrictions under different models.

Overall, the model with order restrictions,  $M_2$  and  $M_3$ , can borrow more information among areas than the model without order restriction,  $M_1$ . The model with uncertain order restriction,  $M_4$ , borrow less information among areas than  $M_2$  or  $M_3$ . For this reason,  $M_2$  and  $M_3$  have sharper posterior density curves than  $M_1$ ,  $M_4$  has slightly flatter posterior density curves than  $M_2$  and  $M_3$ . For the same reason, as shown in Table 3.5,  $M_1$  has the largest total variance, which is the sum of posterior variance of all counties' cell probabilities.  $M_2$  and  $M_3$  have the smallest variance due to its strong unimodal order restriction assumption.  $M_4$ 's variance is between  $M_1$  and  $M_3$ (or  $M_2$ ) since  $M_4$  is a mixture of  $M_2$  and  $M_3$ .

$M_1$	$M_2$ (mode at normal)	$M_3$ (mode at overweight)	$M_4$
0.172	0.063	0.069	0.107

Table 3.5: Total posterior standard variance of  $\theta$

Figure 3.7 and Figure 3.8 are boxplots of  $\theta$ s' posterior samples. The first (Underweight) and last (Obese II) blocks show that different models do not have much difference in estimating the cell probabilities of underweight, normal, and obese I. In the box plots, short line segments from  $M_2$ ,  $M_3$ , and  $M_4$  and long line segments from  $M_1$  show that the models with order restrictions ( $M_2$ ,  $M_3$ ,  $M_4$ ) have smaller variances than the model without order restriction ( $M_1$ ). The models with order restrictions can borrow more information than the model without order restriction. The differences between each box of  $M_1$  are larger than the differences in  $M_2$ ,  $M_3$ , and  $M_4$ . In other word, the differences between posterior mean of each county in  $M_1$  are larger than other models'. It proves that the models with order restrictions borrow more information among areas than the model without order restriction.



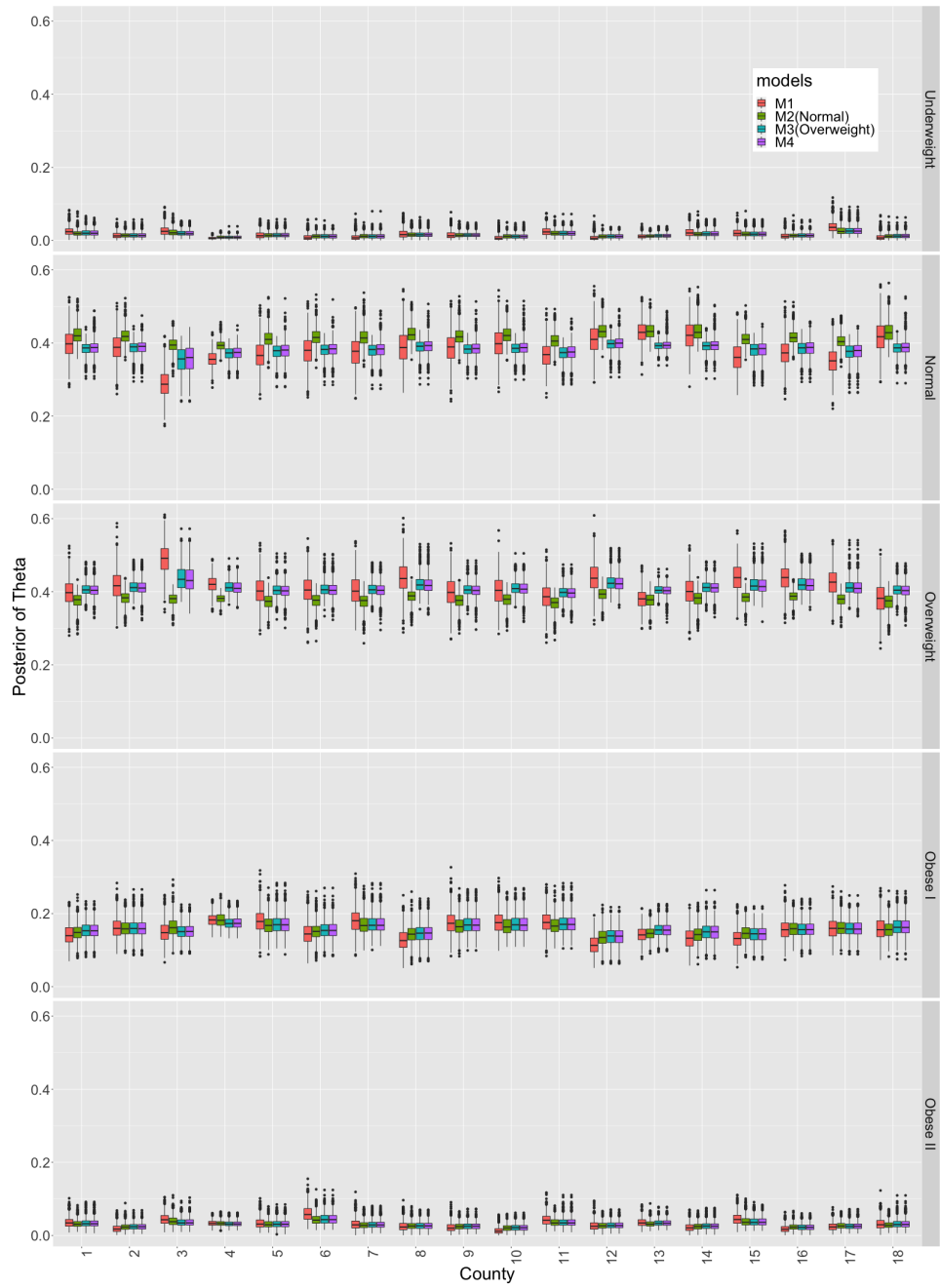


Figure 3.7: Boxplot of  $\theta$  posterior samples: Part I

This is a county-wise comparison for different BMI categories under different models.

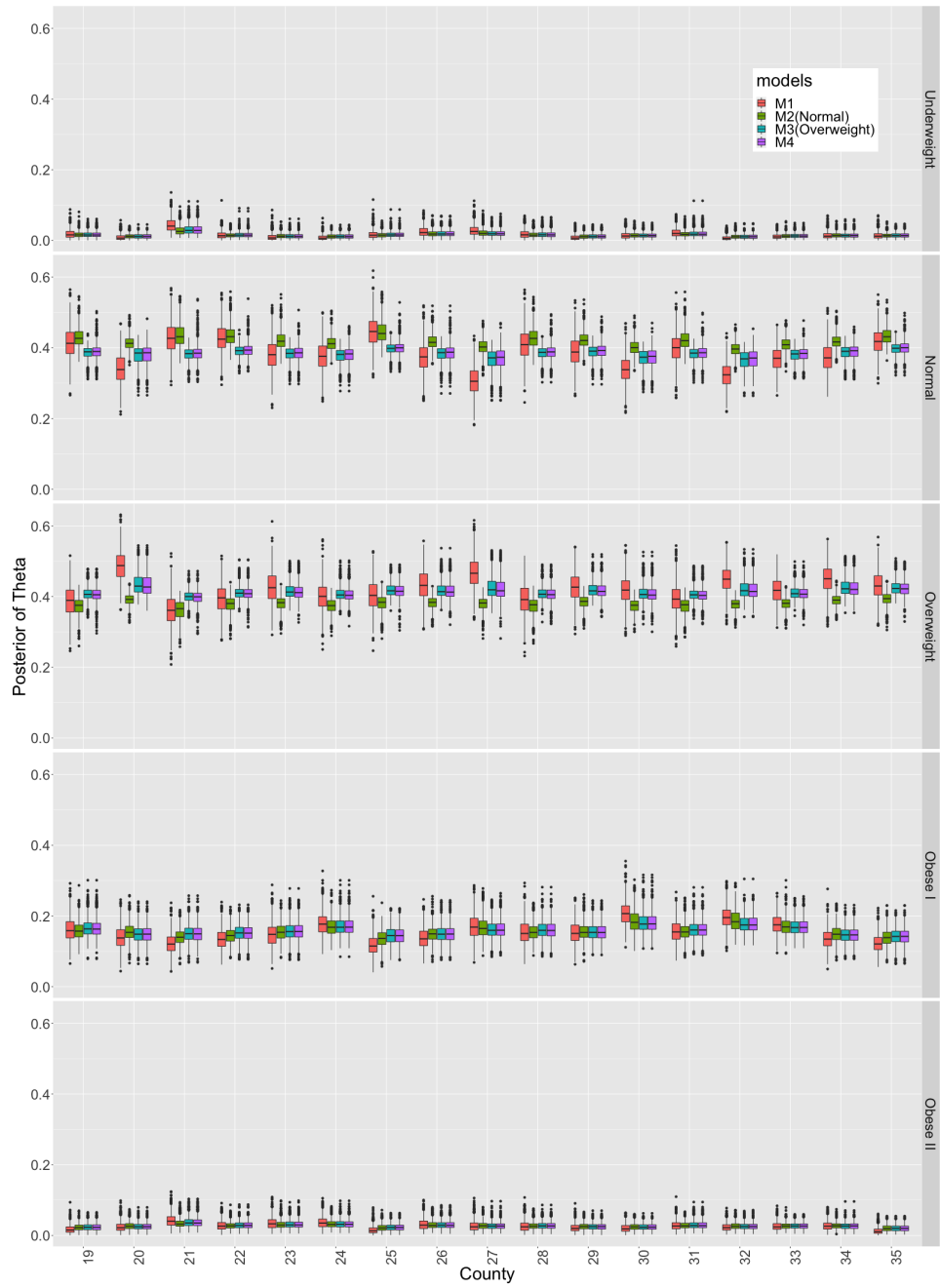


Figure 3.8: Boxplot of  $\theta$  posterior samples: Part II

This is a county-wise comparison for different BMI categories under different models.

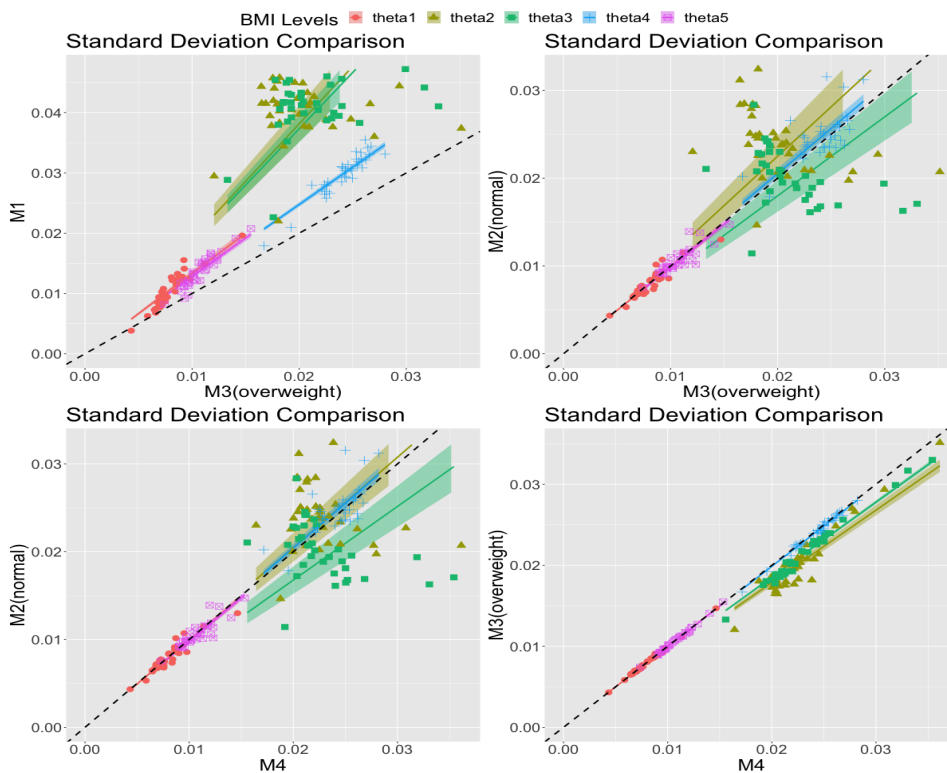


Figure 3.9: Standard deviation comparison between those models to show improvement.

In Figure 3.9, we have some regression lines to show the overall posterior standard deviation comparison among those models. The black dashed line is a reference line whose slope is one. The first plot shows a comparison between  $M_1$  and  $M_3$  (mode at overweight). All of regression lines are above the reference line, which means that  $M_3$  (mode at overweight) has smaller standard deviation. We gain higher precision on estimation of cell probabilities among 35 counties in  $M_3$ . The second plot shows a comparison between  $M_2$  (mode at normal) and  $M_3$  (mode at overweight). The regression lines about underweight, Obese I and Obese II are around the reference line. Only the regression line about overweight shows sig-

nificant difference. It means  $M_3$  (mode at overweight) is slightly better than  $M_2$  (mode at normal). In other word, the assumption that overweight BMI probability is the highest may be more reasonable. The last two plots in Figure 3.9 is a comparison between  $M_2$  (mode at normal) and  $M_4$ ,  $M_3$  (mode at overweight) and  $M_4$ .  $M_4$ 's performance is slightly worse than  $M_3$  and  $M_2$ .

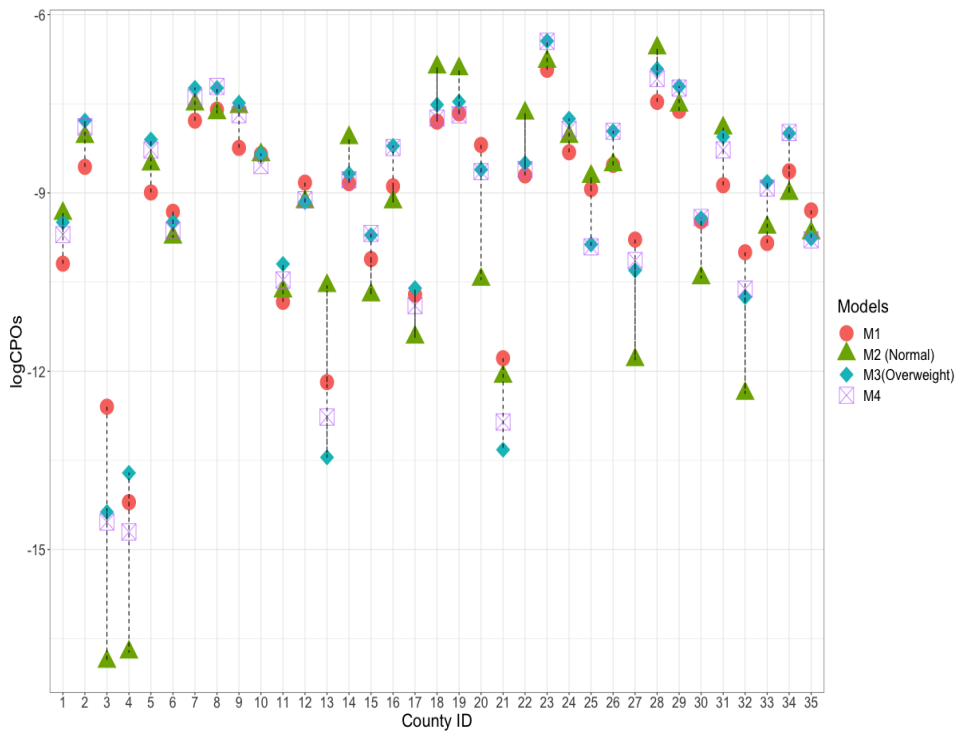


Figure 3.10: CPOs for 35 counties under different models

Note: Lower CPO suggests possible outliers, high-leverage and influential observations.

In Figure 3.10, we use different symbols to represent the CPOs in each model for all 35 counties. For each county, we can see different CPOs because of different model performance in each county. Small CPO values suggest possible outliers, high-leverage and influential observations. In our case, small CPO may suggest

improper order restriction assumption in the model for the county or the county is very different from the other counties. In BMI data, County 4 has the largest observations, which shows lowest CPO value among others. Due to the borrowing feature from the models, County 3 has a low CPO which may be affected by County 4. Another possible explanation for small CPOs in County 3 is that the people with overweight BMI dominates other categories, which may cause that County 3 is very different from the other counties. For most counties, the model with order restriction which assumes the mode is at overweight position can have large CPO, compared with other models. As a summary, in Table 3.6,  $M_3$  (mode at overweight) has the largest LMPL, which should be the ‘best’ model for our BMI data.

$M_1$	$M_2$ (mode at normal)	$M_3$ (mode at overweight)	$M_4$
-326.76	-331.76	-319.11	-323.17

Table 3.6: LPMLs of model  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  for BMI

Comparison of the four models using LPML.

### 3.3 Numerical Example: Smoothed BMI

To have a better comparison between those models, Chen and Nandram (2020b) construct a simulated data transformed from BMI using the idea of Pool-Adjacent-Violators Algorithm (PAVA) to have strong order restrictions as  $\theta_1 \leq \dots \leq \theta_m \geq \dots \geq \theta_K$  (Mair et al. 2009). It is a simple iterative algorithm for solving the quadratic problem.

Generally, given a sequence of  $n$  data points  $y_1, \dots, y_n$ , we start with  $y_1$  on the left. We move to the right until we encounter the first violation  $y_i > y_{i+1}$ . Then

we replace this pair by their average, and back-average to the left as needed, to get monotonicity. We continue this process to the right, until finally we reach  $y_n$ . We can have a reconstructed data set to fit our order restrictions better. Fitting models to the simulated data, we can discover the advantage of hierarchical multinomial-Dirichlet model with order restrictions easily.

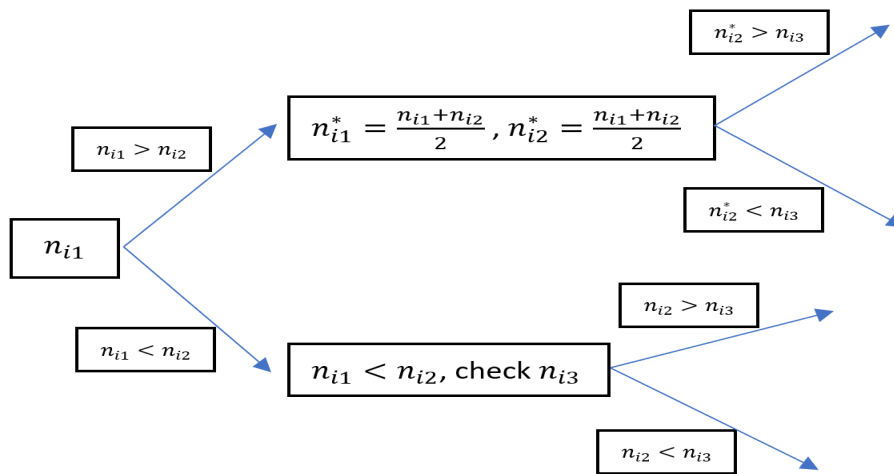


Figure 3.11: Simulation method to have the unimodal order restriction

Here, for each county, we start from BMI level 1 to the mode using PAVA to create an increasing sequence. Then from the mode to BMI level 5, we apply PAVA to create a decreasing sequence. To make sure that each BMI level has an integer number, we take the nearest integer that is larger than the mode to replace the mode, and take the nearest integer that is smaller than  $n_{ij}$  (except the mode) to replace those non-modes. Now our assembled BMI data have strong order restrictions. But we also notice that our current approach cannot be used for a general case to create an unimodal structure. It works for BMI data when the numbers of level 2 and level

3 are significantly larger than others. Now we have a simulated BMI data which mode is at the third position (overweight).

$M_1$	$M_2$ (mode at normal)	$M_3$ (mode at overweight)	$M_4$
-319.83	-330.73	-310.39	-311.26

Table 3.7: LPMLs of model  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  for simulated data  
Comparison of the four models using LPML.

Since the mode is at the third position, the LPML of  $M_3$  is significantly larger than others, which is -310.39. The LPML of  $M_4$  is -311.26, due to the robustness of  $M_4$ . The LPML of  $M_2$  is the smallest, which is -330.73. The LPML of  $M_1$  is -319.83. The LMPLs show that the model with order restrictions can have the best performance if the unimodal assumption is correct. Model  $M_4$ , which incorporates uncertainty about order, has a similar performance as Model  $M_3$ . In Figure 3.12,  $M_3$  and  $M_4$  have consistently large CPO values for 35 counties among those models.  $M_2$  have lowest CPO values at County 3 and 4, which suggests possible outliers, high-leverage and influential observations. For most of counties,  $M_3$  has the largest CPOs and  $M_2$  has the smallest CPOs because of the order restriction assumption may be correct in  $M_3$  but not in  $M_2$ .

In the simulated BMI data, CPO and LMPL are proved to be able to select more adequate models. Model  $M_4$  is robust and consistent for most cases.

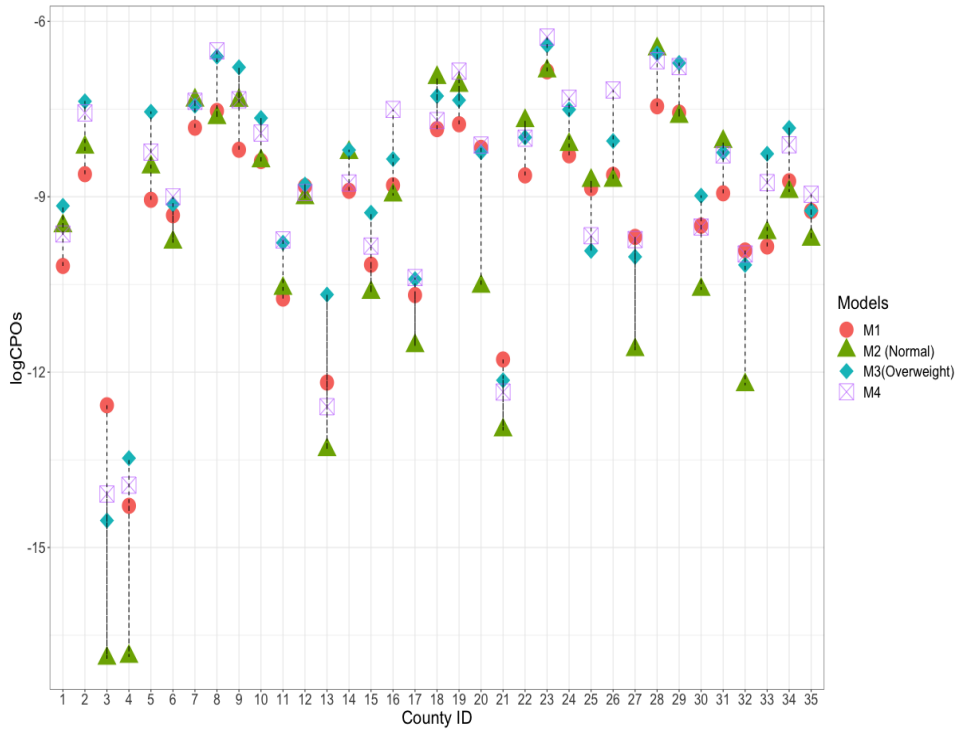


Figure 3.12: CPOs for 35 counties under different models (simulation)

Note: Lower CPO suggests possible outliers, high-leverage and influential observations.

### 3.4 Concluding Remarks

In the numerical example and simulated example, the multinomial-Dirichlet model with order restrictions,  $M_3$ , is the best model which has the largest LPML among those models.  $M_3$  has small posterior standard deviations of parameters of interest due to borrowing information among areas. It borrows more information across areas than the model without order restriction,  $M_1$ , and the model with uncertain order restriction,  $M_4$ . It is consistent with the discussion in Chapter 2.

In Figure 3.10, some counties in  $M_3$  have the smallest CPO among four mod-



els, such as County 13 and County 21. In those counties, the unimodal order restriction that the modal position is at third position may not hold. In Figure 3.12,  $M_3$  does not have small CPO for all areas among those models in the simulated example, which forces the modal position to be at third. CPO can indicate that the unimodal order restriction assumption may not hold for some areas. Overall, LPML and CPO are reasonable Bayesian model selection criteria.

As shown in Figure 3.3 and Figure 3.4, the same unimodal order restriction for all counties may not hold. Some counties have more people with normal BMI than people with overweight BMI, such as County 21. Nandram and Sedransk (1995) and Nandram, Sedransk and Smith (1997) presented a good discussion about unimodal order restriction in a stratified population. They pointed out the potential problem that the same unimodal order restriction may not hold for all areas. They incorporated the uncertainty about the proportion of firms and fish belonging to each of several classes when there are unimodal order relations among the proportions. Their work proved that the model with uncertain order restrictions can be used for cases where the same unimodal order restriction may not hold for all areas. However, their work cannot be used for small area estimation and their model cannot borrow information across areas. Our problem is much more difficult than theirs when we incorporate a similar uncertainty about modal positions into the model.

The multinomial-Dirichlet model with uncertain order restrictions,  $M_4$ , is an extension of  $M_2$  and  $M_3$ , which can be used for the case that the same unimodal order restriction may not hold for all areas. In the numerical example and the simulated example,  $M_4$  has the second-largest LPML. It shows that the model with uncertain order restrictions,  $M_4$ , can provide consistent and accurate inference

about parameters of interest. For many cases, we may not know the modal position of all areas and the unimodal order restrictions may not hold. For this reason, incorporating uncertainty about the modal position into the multinomial-Dirichlet model is necessary. Since we do not have to assume the known modal position,  $M_4$  is more flexible than  $M_2$  and  $M_3$ . We also notice that because of its complexity, it is hard to draw posterior samples directly from its joint posterior density and compute its marginal likelihood for model selection criteria, such as the Bayes factor. We show a method to estimate the posterior probabilities of the modal position, which is  $P(\widehat{L_{pos}} = \ell | \mathbf{n})$ . We also show how to obtain the posterior samples of parameters of interest by mixing posterior samples from  $M_2$  and  $M_3$  together.

## 3.5 Appendix

### 3.5.1 Sampling Method for Parameter $\theta$ , $\mu$ , and $\tau$ in $M_4$

To generate samples of  $\theta$ ,  $\mu$  and  $\tau$ , we have to deal with the uncertainty indicator  $L_{pos}$ . In  $M_4$ , variable  $L_{pos}$  has prior  $P(L_{pos} = \ell) = w_\ell$  and posterior

$$P(L_{pos} = \ell | \mathbf{n}) = \frac{w_{L_{pos}} \int \boldsymbol{\theta}, \boldsymbol{\mu}, \tau \prod_{i=1}^I \left\{ \frac{\prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1} I_{C_{L_{pos}}} I_C \boldsymbol{\mu}_{L_{pos}}}{D(\boldsymbol{\mu} \tau) C(\boldsymbol{\mu} \tau)} \right\} \frac{1}{(1+\tau)^2} d\boldsymbol{\theta} d\boldsymbol{\mu} d\tau}{\sum_{L_{pos}=1}^K w_{L_{pos}} \int \boldsymbol{\theta}, \boldsymbol{\mu}, \tau \prod_{i=1}^I \left\{ \frac{\prod_{j=1}^K \theta_{ij}^{n_{ij} + \mu_j \tau - 1} I_{C_{L_{pos}}} I_C \boldsymbol{\mu}_{L_{pos}}}{D(\boldsymbol{\mu} \tau) C(\boldsymbol{\mu} \tau)} \right\} \frac{1}{(1+\tau)^2} d\boldsymbol{\theta} d\boldsymbol{\mu} d\tau}.$$

Chen and Nandram (2020) notice the order restrictions will significantly increase the computational difficulty, especially for the marginal likelihood. We can use the posterior samples from the model without order restriction to obtain that probability. However, it is difficult for the hierarchical multinomial-Dirichlet model with order restrictions to obtain the probability correctly. We suggest the following approximation method,

#### Computing the Approximation of Mixture Probabilities:

1. Apply  $M_1$  model to the entire data set and acquire posterior samples of  $\theta$  and  $\mu$ .
2. For each iteration of  $\mu$ , check the unimodal order restriction and denote its mode position as  $m_\ell = 1, \dots, K$ .
3. At the same iteration, check  $\theta$ 's unimodal order restrictions for each  $i = 1 \dots, \ell$ , where  $\theta$  is a  $k \times \ell$  matrix and count how many of them have the unimodal order restriction where the mode is the same as  $\mu$ 's mode.
4. For each mode position  $m_\ell$ , add counts of each cases together and generate the ratios.

For example, in our application BMI, 6.595% of posterior samples of  $\theta$  and

$\mu$  has mode at the second position, 93.405% of posterior samples of  $\theta$  and  $\mu$  has mode at the third position. Then we can have  $P(\widehat{L_{pos}} = 2|\mathbf{n}) \approx 0.066$  and  $P(\widehat{L_{pos}} = 3|\mathbf{n}) \approx 0.934$  as probabilities to mix samples from  $M_2$  (mode at 2nd) and samples from  $M_3$  (mode at 3rd) together.

### 3.5.2 Bayesian Diagnostics for $M_2$ , $M_3$ , and $M_4$

We present CPO and LPML of  $M_2$ ,  $M_3$ , and  $M_4$  as Bayesian model selection criteria in the section.

CPO of  $M_2$  and CPO of  $M_3$  are very similar to each other. CPO of  $M_2$  are

$$\begin{aligned}\widehat{CPO}_{i(M_2)} &= \left[ \frac{1}{M} \sum_{h=1}^M \frac{\prod_{j=1}^K n_{ij}!}{n_i!} \frac{D(\boldsymbol{\mu}^{(h)})\tau^{(h)}C(\boldsymbol{\mu}^{(h)})\tau^{(h)}}{D(\mathbf{n}_i + \boldsymbol{\mu}^{(h)})\tau^{(h)}C(\mathbf{n}_i + \boldsymbol{\mu}^{(h)})\tau^{(h)}} \right]^{-1} \\ &= \left[ \frac{1}{M} \sum_{h=1}^M \frac{\prod_{j=1}^K n_{ij}!}{n_i!} \frac{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu^{(h)}\tau^{(h)}-1} d\boldsymbol{\theta}_i}{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu^{(h)}\tau^{(h)}-1} d\boldsymbol{\theta}_i} \right]^{-1} \\ &= \left[ \frac{1}{M} \sum_{h=1}^M \frac{\prod_{j=1}^K n_{ij}!}{n_i!} \int_{\theta_i \in C} \frac{\prod_{j=1}^K \theta_{ij}^{\mu^{(h)}\tau^{(h)}-1}}{\prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu^{(h)}\tau^{(h)}-1}} \frac{\prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu^{(h)}\tau^{(h)}-1}}{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu^{(h)}\tau^{(h)}-1} d\boldsymbol{\theta}_i} d\boldsymbol{\theta}_i \right]^{-1},\end{aligned}$$

where  $\frac{\prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu^{(h)}\tau^{(h)}-1}}{\int_{\theta_i \in C} \prod_{j=1}^K \theta_{ij}^{n_{ij}+\mu^{(h)}\tau^{(h)}-1} d\boldsymbol{\theta}_i}$  is the density function of  $\boldsymbol{\theta}_i$ , and  $\boldsymbol{\theta}_i \in C$ .

We notice  $\boldsymbol{\mu}^{(h)}$  and  $\tau^{(h)}$ ,  $h = 1, \dots, M$  are the posterior samples from section 2.6.3. For each pair of  $\boldsymbol{\mu}^{(h)}$  and  $\tau^{(h)}$ , we can draw  $\boldsymbol{\theta}_i$  from  $\text{Dirichlet}(\mathbf{n}_i + \boldsymbol{\mu}^{(h)})\tau^{(h)}$ .

$$\widehat{CPO}_{i(M_2)} = \left[ \frac{1}{M} \sum_{h=1}^M \frac{\prod_{j=1}^K n_{ij}!}{n_i!} \left( \frac{1}{M'} \sum_{h'=1}^{M'} \prod_{j=1}^K \theta_{ij}^{(h')-n_{ij}} \right) \right]^{-1},$$

where

for  $h' = 1, \dots, M'$ ,  $M'\boldsymbol{\theta}_i^{(h')} \sim \text{Dirichlet}(\mathbf{n}_i + \boldsymbol{\mu}^{(h)})\tau^{(h)}$  with order restriction.

Then we get the LPML as  $\widehat{LPML} = \sum_{i=1}^I \log(\widehat{CPO}_i)$ . As we can see from the following equation, it is not easy to compute  $CPO_i$  or  $\widehat{CPO}_i$  of  $M_4$ .

$$CPO_i = f(n_i|n_{(i)}) = \sum_{\ell=1}^K [P(L = \ell|n_{(i)})f(n_i|n_{(i),L=\ell})].$$

While we know  $f(n_i|n_{(i),L=\ell})$  can be computed like  $CPO$ s of  $M_2$ , it is hard to get the posterior probability of  $P(L = \ell|n_{(i)})$ . Here,

$$\begin{aligned} CPO_{i(M_4)} &= f(n_i|n_{(i)}) = \left( \frac{f(n_{(i)})}{f(n)} \right)^{-1} \\ &= \left[ \frac{\sum_{\ell=1}^K P(L = \ell) \int \int f(n_{(i)}|\boldsymbol{\mu}, \tau, L = \ell) f(\boldsymbol{\mu}, \tau|L = \ell) d\boldsymbol{\mu} d\tau}{f(n)} \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K P(L = \ell) \int \int \frac{f(n_{(i)}|\boldsymbol{\mu}, \tau, L = \ell) f(\boldsymbol{\mu}, \tau|L = \ell)}{f(n)} d\boldsymbol{\mu} d\tau \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K P(L = \ell) \int \int \frac{f(n_i|\boldsymbol{\mu}, \tau, L = \ell) f(n_{(i)}|\boldsymbol{\mu}, \tau, L = \ell) f(\boldsymbol{\mu}, \tau|L = \ell)}{f(n_i|\boldsymbol{\mu}, \tau, L = \ell) f(n)} d\boldsymbol{\mu} d\tau \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K P(L = \ell) \int \int \frac{f(n|\boldsymbol{\mu}, \tau, L = \ell) f(\boldsymbol{\mu}, \tau|L = \ell)}{f(n_i|\boldsymbol{\mu}, \tau, L = \ell) f(n)} d\boldsymbol{\mu} d\tau \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K P(L = \ell) \int \int \frac{f(n|L = \ell)}{f(n_i|\boldsymbol{\mu}, \tau, L = \ell) f(n)} \frac{f(n|\boldsymbol{\mu}, \tau, L = \ell) f(\boldsymbol{\mu}, \tau|L = \ell)}{f(n|L = \ell)} d\boldsymbol{\mu} d\tau \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K P(L = \ell) \frac{f(n|L = \ell)}{f(n)} \int \int \frac{1}{f(n_i|\boldsymbol{\mu}, \tau, L = \ell)} \frac{f(n|\boldsymbol{\mu}, \tau, L = \ell) f(\boldsymbol{\mu}, \tau|L = \ell)}{f(n|L = \ell)} d\boldsymbol{\mu} d\tau \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K \frac{P(L = \ell) f(n|L = \ell)}{\sum_{\ell=1}^K P(L = \ell) f(n|L = \ell)} \int \int \frac{1}{f(n_i|\boldsymbol{\mu}, \tau, L = \ell)} f(\boldsymbol{\mu}, \tau|n, L = \ell) d\boldsymbol{\mu} d\tau \right]^{-1} \\ &= \left[ \sum_{\ell=1}^K P(L = \ell|n) \int \int \frac{1}{f(n_i|\boldsymbol{\mu}, \tau, L = \ell)} f(\boldsymbol{\mu}, \tau|n, L = \ell) d\boldsymbol{\mu} d\tau \right]^{-1}, \end{aligned}$$

then  $\widehat{CPO}_i \approx \left[ \sum_{\ell=1}^K P(\widehat{L} = \ell|n) \frac{1}{\widehat{CPO}_{i(M_L|L=\ell)}} \right]^{-1}$ , where  $\widehat{CPO}_{i(M_L|L=\ell)}$  are known, such as  $\widehat{CPO}_{i(M_2)}$  from  $M_2$  and  $\widehat{CPO}_{i(M_3)}$  from  $M_3$ . Without extra computation, taking advantage of known CPOs from  $M_2$  and  $M_3$ , and the estimated  $P(L = \ell|n)$  from the previous section, we can easily acquire the CPO of  $M_4$ .

### 3.5.3 Posterior Summary of $\theta$ in $M_1, M_2, M_3,$ and $M_4$

Table 3.8: Posterior summary of  $\theta$  part I: Counties 1-9

County ID	Model	Underweight			Normal			Overweight			Obese I			Obese II		
		PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV
1	$M_1$	0.026	0.013	0.501	0.399	0.040	0.101	0.394	0.040	0.102	0.143	0.029	0.206	0.039	0.016	0.408
	$M_2$	0.021	0.009	0.425	0.421	0.023	0.056	0.376	0.021	0.056	0.148	0.023	0.153	0.033	0.010	0.316
	$M_3$	0.021	0.009	0.431	0.376	0.019	0.051	0.418	0.023	0.055	0.152	0.023	0.153	0.033	0.011	0.323
	$M_4$	0.021	0.009	0.431	0.393	0.030	0.076	0.404	0.030	0.075	0.150	0.023	0.156	0.033	0.010	0.315
2	$M_1$	0.014	0.010	0.704	0.390	0.040	0.102	0.417	0.041	0.098	0.160	0.030	0.189	0.019	0.011	0.580
	$M_2$	0.015	0.007	0.490	0.422	0.024	0.056	0.381	0.019	0.049	0.159	0.024	0.152	0.023	0.009	0.386
	$M_3$	0.015	0.007	0.494	0.375	0.020	0.055	0.426	0.025	0.059	0.161	0.023	0.143	0.023	0.010	0.405
	$M_4$	0.015	0.007	0.476	0.391	0.031	0.079	0.409	0.031	0.077	0.161	0.024	0.147	0.024	0.010	0.405
3	$M_1$	0.028	0.014	0.489	0.282	0.039	0.137	0.495	0.042	0.085	0.149	0.029	0.192	0.047	0.017	0.368
	$M_2$	0.024	0.011	0.459	0.393	0.021	0.054	0.378	0.018	0.047	0.166	0.028	0.167	0.040	0.015	0.368
	$M_3$	0.021	0.009	0.440	0.334	0.035	0.106	0.458	0.036	0.079	0.151	0.022	0.146	0.037	0.012	0.320
	$M_4$	0.022	0.010	0.452	0.354	0.042	0.118	0.429	0.050	0.117	0.156	0.026	0.163	0.038	0.013	0.342
4	$M_1$	0.007	0.004	0.543	0.356	0.022	0.062	0.421	0.022	0.053	0.183	0.018	0.096	0.034	0.009	0.252
	$M_2$	0.009	0.004	0.461	0.394	0.014	0.035	0.381	0.011	0.029	0.182	0.020	0.112	0.034	0.008	0.224
	$M_3$	0.009	0.004	0.451	0.363	0.018	0.050	0.422	0.019	0.046	0.174	0.017	0.098	0.032	0.007	0.220
	$M_4$	0.009	0.004	0.456	0.374	0.023	0.061	0.407	0.026	0.063	0.177	0.018	0.104	0.032	0.007	0.221
5	$M_1$	0.016	0.011	0.708	0.370	0.042	0.112	0.400	0.042	0.104	0.180	0.033	0.181	0.035	0.016	0.453
	$M_2$	0.015	0.008	0.515	0.413	0.024	0.057	0.372	0.021	0.057	0.168	0.027	0.158	0.032	0.012	0.360
	$M_3$	0.015	0.007	0.490	0.366	0.023	0.063	0.419	0.027	0.063	0.169	0.026	0.152	0.032	0.011	0.341
	$M_4$	0.015	0.008	0.493	0.382	0.032	0.084	0.402	0.033	0.083	0.169	0.026	0.154	0.032	0.011	0.356
6	$M_1$	0.009	0.009	0.943	0.380	0.045	0.118	0.402	0.044	0.108	0.147	0.032	0.217	0.063	0.021	0.339
	$M_2$	0.012	0.007	0.586	0.417	0.025	0.059	0.375	0.020	0.054	0.151	0.024	0.160	0.046	0.017	0.362
	$M_3$	0.012	0.007	0.569	0.371	0.023	0.061	0.423	0.026	0.061	0.151	0.023	0.150	0.043	0.015	0.355
	$M_4$	0.012	0.007	0.590	0.387	0.032	0.083	0.406	0.034	0.083	0.151	0.024	0.158	0.044	0.016	0.370
7	$M_1$	0.009	0.009	0.943	0.376	0.044	0.117	0.400	0.045	0.113	0.183	0.035	0.191	0.032	0.016	0.502
	$M_2$	0.012	0.007	0.575	0.416	0.025	0.059	0.374	0.022	0.058	0.169	0.028	0.163	0.030	0.012	0.389
	$M_3$	0.013	0.007	0.578	0.367	0.023	0.062	0.422	0.027	0.065	0.169	0.025	0.150	0.030	0.011	0.359
	$M_4$	0.012	0.007	0.590	0.384	0.033	0.087	0.405	0.034	0.084	0.169	0.027	0.156	0.030	0.011	0.372
8	$M_1$	0.019	0.014	0.726	0.387	0.048	0.123	0.443	0.050	0.112	0.126	0.033	0.265	0.025	0.015	0.597
	$M_2$	0.017	0.009	0.520	0.426	0.025	0.058	0.386	0.020	0.051	0.143	0.024	0.170	0.027	0.011	0.406
	$M_3$	0.016	0.008	0.488	0.376	0.023	0.061	0.437	0.029	0.066	0.144	0.023	0.160	0.027	0.010	0.387
	$M_4$	0.017	0.009	0.520	0.394	0.035	0.088	0.418	0.035	0.083	0.144	0.023	0.162	0.027	0.011	0.401
9	$M_1$	0.016	0.011	0.686	0.391	0.045	0.116	0.398	0.044	0.110	0.174	0.035	0.203	0.021	0.012	0.584
	$M_2$	0.015	0.008	0.504	0.421	0.027	0.064	0.373	0.021	0.058	0.165	0.025	0.152	0.026	0.010	0.389
	$M_3$	0.016	0.008	0.492	0.372	0.021	0.056	0.420	0.025	0.059	0.167	0.025	0.149	0.025	0.010	0.389
	$M_4$	0.015	0.008	0.496	0.390	0.033	0.084	0.403	0.033	0.081	0.166	0.025	0.148	0.026	0.010	0.383

Note: Posterior Mean (PM), Posterior Standard Deviation (PSD), Coefficient of Variation (CV)

Table 3.9: Posterior summary of  $\theta$  part II: Counties 10-18

County ID	Model	Underweight			Normal			Overweight			Obese I			Obese II		
		PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV
10	$M_1$	0.008	0.007	0.940	0.396	0.041	0.103	0.403	0.042	0.104	0.180	0.033	0.184	0.013	0.010	0.760
	$M_2$	0.011	0.007	0.574	0.423	0.024	0.057	0.377	0.022	0.058	0.167	0.025	0.151	0.021	0.010	0.453
	$M_3$	0.012	0.007	0.573	0.376	0.021	0.055	0.422	0.024	0.057	0.169	0.025	0.146	0.021	0.009	0.438
	$M_4$	0.012	0.007	0.579	0.393	0.033	0.083	0.406	0.032	0.079	0.168	0.025	0.146	0.021	0.009	0.447
11	$M_1$	0.026	0.013	0.515	0.365	0.037	0.102	0.385	0.038	0.098	0.181	0.030	0.167	0.044	0.016	0.366
	$M_2$	0.021	0.009	0.420	0.407	0.024	0.058	0.367	0.021	0.057	0.169	0.025	0.148	0.036	0.012	0.323
	$M_3$	0.021	0.009	0.435	0.363	0.022	0.062	0.411	0.026	0.064	0.169	0.024	0.144	0.037	0.012	0.326
	$M_4$	0.021	0.009	0.440	0.379	0.031	0.081	0.395	0.031	0.078	0.169	0.024	0.140	0.036	0.012	0.322
12	$M_1$	0.008	0.007	0.937	0.415	0.041	0.099	0.439	0.042	0.095	0.113	0.027	0.235	0.026	0.013	0.507
	$M_2$	0.012	0.007	0.581	0.434	0.024	0.055	0.392	0.020	0.050	0.135	0.023	0.171	0.028	0.010	0.360
	$M_3$	0.012	0.007	0.557	0.386	0.022	0.056	0.438	0.026	0.059	0.137	0.024	0.173	0.027	0.010	0.355
	$M_4$	0.012	0.007	0.583	0.403	0.033	0.082	0.422	0.033	0.078	0.135	0.024	0.176	0.028	0.010	0.357
13	$M_1$	0.012	0.007	0.563	0.432	0.030	0.070	0.378	0.029	0.076	0.142	0.021	0.146	0.036	0.012	0.323
	$M_2$	0.013	0.006	0.426	0.434	0.023	0.053	0.375	0.020	0.053	0.146	0.018	0.123	0.033	0.009	0.272
	$M_3$	0.013	0.006	0.423	0.388	0.014	0.037	0.413	0.017	0.042	0.152	0.019	0.122	0.034	0.009	0.277
	$M_4$	0.013	0.006	0.426	0.405	0.028	0.069	0.399	0.025	0.063	0.150	0.019	0.124	0.033	0.009	0.273
14	$M_1$	0.024	0.013	0.545	0.425	0.045	0.106	0.399	0.044	0.110	0.131	0.030	0.228	0.022	0.012	0.567
	$M_2$	0.019	0.009	0.465	0.434	0.027	0.062	0.378	0.023	0.059	0.144	0.023	0.162	0.025	0.010	0.380
	$M_3$	0.019	0.009	0.463	0.383	0.021	0.055	0.426	0.024	0.057	0.147	0.024	0.162	0.026	0.010	0.389
	$M_4$	0.019	0.009	0.465	0.400	0.033	0.082	0.409	0.032	0.078	0.146	0.024	0.162	0.025	0.010	0.378
15	$M_1$	0.022	0.012	0.532	0.357	0.041	0.114	0.444	0.041	0.093	0.131	0.028	0.214	0.047	0.018	0.384
	$M_2$	0.018	0.008	0.438	0.412	0.021	0.050	0.384	0.017	0.045	0.148	0.025	0.166	0.039	0.013	0.334
	$M_3$	0.018	0.008	0.462	0.368	0.025	0.068	0.433	0.028	0.064	0.145	0.023	0.155	0.037	0.012	0.325
	$M_4$	0.018	0.008	0.448	0.383	0.032	0.083	0.416	0.035	0.083	0.146	0.024	0.167	0.037	0.012	0.327
16	$M_1$	0.013	0.009	0.695	0.372	0.037	0.100	0.439	0.041	0.092	0.158	0.029	0.183	0.018	0.010	0.584
	$M_2$	0.015	0.007	0.482	0.416	0.020	0.048	0.386	0.017	0.044	0.160	0.024	0.150	0.023	0.009	0.406
	$M_3$	0.014	0.007	0.480	0.371	0.023	0.062	0.436	0.028	0.063	0.157	0.021	0.135	0.023	0.009	0.383
	$M_4$	0.014	0.007	0.481	0.386	0.031	0.080	0.418	0.035	0.083	0.158	0.023	0.147	0.023	0.009	0.381
17	$M_1$	0.039	0.016	0.405	0.351	0.039	0.111	0.426	0.041	0.095	0.161	0.030	0.187	0.024	0.012	0.507
	$M_2$	0.028	0.012	0.418	0.406	0.021	0.051	0.378	0.017	0.045	0.161	0.025	0.153	0.027	0.010	0.362
	$M_3$	0.026	0.011	0.420	0.362	0.024	0.066	0.428	0.028	0.064	0.157	0.021	0.132	0.027	0.009	0.351
	$M_4$	0.027	0.012	0.425	0.377	0.030	0.080	0.410	0.034	0.083	0.159	0.023	0.142	0.027	0.010	0.365
18	$M_1$	0.009	0.009	0.964	0.420	0.045	0.108	0.376	0.043	0.114	0.164	0.036	0.220	0.032	0.017	0.519
	$M_2$	0.012	0.007	0.581	0.430	0.028	0.065	0.370	0.024	0.066	0.158	0.026	0.163	0.030	0.011	0.373
	$M_3$	0.013	0.007	0.552	0.378	0.019	0.051	0.417	0.024	0.056	0.162	0.025	0.153	0.031	0.011	0.362
	$M_4$	0.013	0.007	0.568	0.396	0.034	0.086	0.400	0.033	0.082	0.161	0.025	0.159	0.031	0.011	0.366

Note: Posterior Mean (PM), Posterior Standard Deviation (PSD), Coefficient of Variation (CV)

Table 3.10: Posterior summary of  $\theta$  part III: Counties 19-27

County ID	Model	Underweight			Normal			Overweight			Obese I			Obese II		
		PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV
19	$M_1$	0.019	0.013	0.693	0.416	0.048	0.116	0.384	0.047	0.123	0.164	0.035	0.214	0.016	0.012	0.767
	$M_2$	0.016	0.008	0.507	0.431	0.030	0.070	0.372	0.025	0.066	0.157	0.026	0.162	0.023	0.010	0.430
	$M_3$	0.017	0.009	0.532	0.378	0.020	0.053	0.420	0.025	0.059	0.162	0.025	0.158	0.024	0.010	0.407
	$M_4$	0.017	0.009	0.533	0.397	0.036	0.091	0.402	0.034	0.085	0.161	0.027	0.166	0.024	0.010	0.422
20	$M_1$	0.009	0.009	0.935	0.335	0.044	0.132	0.494	0.047	0.095	0.139	0.031	0.225	0.023	0.013	0.564
	$M_2$	0.013	0.008	0.610	0.413	0.020	0.048	0.390	0.017	0.043	0.157	0.027	0.171	0.027	0.011	0.406
	$M_3$	0.012	0.007	0.551	0.359	0.029	0.082	0.454	0.035	0.077	0.149	0.023	0.156	0.026	0.010	0.380
	$M_4$	0.012	0.007	0.599	0.378	0.037	0.098	0.432	0.043	0.100	0.152	0.025	0.166	0.026	0.010	0.396
21	$M_1$	0.048	0.021	0.431	0.431	0.050	0.116	0.353	0.051	0.145	0.123	0.033	0.269	0.046	0.021	0.453
	$M_2$	0.029	0.012	0.432	0.436	0.032	0.074	0.363	0.029	0.079	0.138	0.025	0.179	0.035	0.013	0.363
	$M_3$	0.029	0.014	0.485	0.377	0.020	0.052	0.412	0.024	0.058	0.146	0.025	0.174	0.036	0.013	0.364
	$M_4$	0.029	0.014	0.459	0.398	0.038	0.096	0.394	0.035	0.090	0.143	0.026	0.180	0.036	0.013	0.372
22	$M_1$	0.016	0.010	0.660	0.431	0.044	0.102	0.391	0.043	0.109	0.134	0.030	0.226	0.029	0.015	0.512
	$M_2$	0.015	0.008	0.500	0.434	0.027	0.062	0.378	0.023	0.060	0.145	0.024	0.163	0.028	0.010	0.369
	$M_3$	0.015	0.008	0.500	0.384	0.019	0.050	0.423	0.023	0.055	0.149	0.023	0.151	0.029	0.011	0.362
	$M_4$	0.015	0.008	0.508	0.402	0.034	0.083	0.407	0.032	0.078	0.147	0.024	0.160	0.029	0.011	0.376
23	$M_1$	0.011	0.011	0.979	0.379	0.048	0.126	0.426	0.048	0.112	0.149	0.034	0.230	0.035	0.018	0.516
	$M_2$	0.013	0.007	0.560	0.422	0.025	0.060	0.379	0.021	0.055	0.155	0.026	0.171	0.031	0.011	0.352
	$M_3$	0.013	0.007	0.568	0.371	0.024	0.064	0.431	0.029	0.068	0.154	0.025	0.162	0.032	0.012	0.378
	$M_4$	0.013	0.007	0.570	0.388	0.035	0.089	0.413	0.037	0.089	0.155	0.026	0.171	0.032	0.012	0.365
24	$M_1$	0.008	0.008	1.005	0.375	0.044	0.116	0.397	0.043	0.107	0.182	0.034	0.189	0.038	0.017	0.445
	$M_2$	0.012	0.007	0.596	0.414	0.024	0.058	0.373	0.021	0.055	0.167	0.027	0.160	0.033	0.011	0.339
	$M_3$	0.012	0.007	0.551	0.368	0.023	0.062	0.418	0.026	0.061	0.169	0.025	0.145	0.033	0.011	0.339
	$M_4$	0.012	0.007	0.581	0.385	0.033	0.085	0.403	0.032	0.079	0.168	0.026	0.153	0.032	0.011	0.343
25	$M_1$	0.018	0.012	0.676	0.449	0.047	0.103	0.402	0.045	0.112	0.117	0.029	0.248	0.015	0.011	0.751
	$M_2$	0.016	0.008	0.483	0.444	0.030	0.068	0.383	0.023	0.060	0.135	0.025	0.185	0.022	0.010	0.435
	$M_3$	0.016	0.008	0.512	0.390	0.020	0.050	0.428	0.024	0.055	0.143	0.025	0.177	0.023	0.010	0.422
	$M_4$	0.016	0.008	0.510	0.411	0.036	0.087	0.412	0.033	0.080	0.139	0.026	0.188	0.023	0.009	0.421
26	$M_1$	0.027	0.016	0.595	0.373	0.045	0.120	0.432	0.046	0.107	0.136	0.032	0.232	0.032	0.016	0.514
	$M_2$	0.021	0.010	0.483	0.417	0.023	0.056	0.383	0.019	0.050	0.148	0.026	0.173	0.031	0.012	0.378
	$M_3$	0.020	0.009	0.477	0.370	0.025	0.066	0.433	0.029	0.066	0.148	0.024	0.161	0.029	0.010	0.357
	$M_4$	0.020	0.009	0.463	0.387	0.034	0.087	0.415	0.035	0.084	0.148	0.025	0.168	0.030	0.011	0.365
27	$M_1$	0.030	0.018	0.582	0.302	0.045	0.148	0.473	0.049	0.103	0.170	0.037	0.219	0.026	0.016	0.600
	$M_2$	0.022	0.011	0.492	0.401	0.023	0.056	0.378	0.019	0.050	0.171	0.030	0.176	0.028	0.011	0.377
	$M_3$	0.020	0.009	0.463	0.346	0.034	0.099	0.446	0.037	0.082	0.160	0.024	0.150	0.027	0.011	0.386
	$M_4$	0.021	0.010	0.479	0.366	0.041	0.112	0.423	0.046	0.109	0.163	0.027	0.163	0.028	0.011	0.391

Note: Posterior Mean (PM), Posterior Standard Deviation (PSD), Coefficient of Variation (CV)



Table 3.11: Posterior summary of  $\theta$  part III: Counties 28-35

County ID	Model	Underweight			Normal			Overweight			Obese I			Obese II		
		PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV
28	$M_1$	0.019	0.013	0.687	0.410	0.047	0.115	0.389	0.048	0.122	0.156	0.035	0.221	0.025	0.015	0.594
	$M_2$	0.017	0.008	0.494	0.429	0.028	0.066	0.374	0.025	0.066	0.154	0.026	0.168	0.027	0.010	0.389
	$M_3$	0.017	0.008	0.504	0.377	0.022	0.058	0.421	0.025	0.059	0.159	0.027	0.167	0.027	0.010	0.373
	$M_4$	0.017	0.009	0.508	0.395	0.034	0.087	0.404	0.035	0.086	0.157	0.026	0.168	0.027	0.011	0.394
29	$M_1$	0.009	0.008	0.980	0.391	0.042	0.107	0.429	0.041	0.096	0.150	0.032	0.211	0.022	0.013	0.575
	$M_2$	0.012	0.007	0.621	0.424	0.023	0.055	0.384	0.020	0.051	0.155	0.024	0.156	0.025	0.010	0.394
	$M_3$	0.012	0.007	0.566	0.376	0.023	0.060	0.433	0.027	0.062	0.154	0.023	0.147	0.025	0.009	0.370
	$M_4$	0.012	0.007	0.591	0.393	0.033	0.083	0.416	0.033	0.081	0.155	0.023	0.149	0.025	0.009	0.372
30	$M_1$	0.015	0.010	0.702	0.338	0.041	0.121	0.420	0.044	0.104	0.207	0.034	0.166	0.020	0.012	0.590
	$M_2$	0.016	0.007	0.471	0.401	0.022	0.055	0.373	0.019	0.052	0.186	0.032	0.171	0.025	0.010	0.380
	$M_3$	0.015	0.007	0.466	0.355	0.027	0.075	0.427	0.028	0.066	0.179	0.028	0.155	0.024	0.009	0.386
	$M_4$	0.015	0.007	0.468	0.371	0.033	0.090	0.407	0.037	0.090	0.183	0.030	0.165	0.025	0.009	0.386
31	$M_1$	0.023	0.013	0.578	0.399	0.043	0.107	0.391	0.043	0.110	0.158	0.031	0.199	0.030	0.015	0.491
	$M_2$	0.019	0.009	0.462	0.423	0.026	0.062	0.373	0.022	0.060	0.156	0.025	0.161	0.029	0.011	0.374
	$M_3$	0.019	0.009	0.478	0.373	0.022	0.058	0.420	0.025	0.060	0.160	0.025	0.155	0.028	0.010	0.351
	$M_4$	0.019	0.009	0.472	0.391	0.033	0.083	0.403	0.033	0.082	0.159	0.025	0.158	0.029	0.010	0.355
32	$M_1$	0.007	0.007	0.941	0.319	0.037	0.116	0.450	0.039	0.086	0.200	0.032	0.159	0.024	0.012	0.511
	$M_2$	0.012	0.007	0.569	0.397	0.020	0.051	0.378	0.016	0.042	0.186	0.031	0.164	0.027	0.010	0.370
	$M_3$	0.011	0.006	0.576	0.348	0.029	0.084	0.439	0.030	0.068	0.177	0.026	0.144	0.026	0.009	0.345
	$M_4$	0.011	0.006	0.579	0.365	0.036	0.097	0.417	0.039	0.094	0.181	0.029	0.159	0.026	0.009	0.352
33	$M_1$	0.011	0.007	0.662	0.367	0.037	0.101	0.419	0.035	0.084	0.177	0.029	0.164	0.026	0.012	0.458
	$M_2$	0.014	0.007	0.510	0.411	0.020	0.049	0.381	0.017	0.044	0.168	0.024	0.140	0.027	0.009	0.331
	$M_3$	0.013	0.006	0.502	0.370	0.021	0.058	0.424	0.024	0.056	0.167	0.022	0.133	0.027	0.009	0.346
	$M_4$	0.013	0.007	0.519	0.384	0.029	0.076	0.408	0.031	0.076	0.169	0.023	0.135	0.027	0.009	0.352
34	$M_1$	0.015	0.010	0.695	0.373	0.041	0.110	0.452	0.042	0.092	0.134	0.030	0.222	0.026	0.013	0.503
	$M_2$	0.015	0.008	0.496	0.420	0.021	0.051	0.389	0.017	0.044	0.148	0.023	0.158	0.028	0.011	0.390
	$M_3$	0.015	0.007	0.485	0.372	0.024	0.065	0.443	0.029	0.065	0.144	0.022	0.153	0.027	0.010	0.363
	$M_4$	0.015	0.007	0.495	0.388	0.033	0.086	0.424	0.036	0.085	0.145	0.023	0.157	0.028	0.011	0.381
35	$M_1$	0.014	0.010	0.705	0.419	0.040	0.095	0.435	0.040	0.092	0.121	0.028	0.228	0.012	0.010	0.790
	$M_2$	0.015	0.007	0.488	0.436	0.024	0.055	0.392	0.020	0.050	0.138	0.022	0.162	0.020	0.009	0.447
	$M_3$	0.014	0.007	0.474	0.388	0.021	0.055	0.437	0.026	0.059	0.140	0.023	0.166	0.020	0.009	0.433
	$M_4$	0.015	0.007	0.486	0.406	0.032	0.080	0.421	0.033	0.077	0.139	0.023	0.167	0.020	0.009	0.439

Note: Posterior Mean (PM), Posterior Standard Deviation (PSD), Coefficient of Variation (CV)

## Chapter 4

# Partial Exchangeability

To avoid borrowing information from the extreme areas, we propose the model with partially exchangeability, which is the exchangeability-nonexchangeability (EXNEX) model. EXNEX models allow borrowing information across similar areas while avoiding too optimistic borrowing from very different areas. We propose a simple binomial-Beta EXNEX model to illustrate the partial exchangeability in a simulation study. Then we present a multinomial-Dirichlet EXNEX model with order restrictions and a numerical example.

### 4.1 Binomial-Beta EXNEX Model

It is difficult and time-consuming to run simulations for the multinomial-Dirichlet EXNEX model with order restrictions under several scenarios. To simply demonstrate the Bayesian hierarchical model with the partial exchangeability, we first present a binomial-Beta EXNEX model without any order restriction to show that the EXNEX model can borrow information among areas and avoid borrowing from

very different areas.

Let  $y_i$  denote the number of successes in  $n_i$  independent trials in the  $i$ th area,  $i = 1, \dots, \ell$  and  $\theta_i$  denote the probability of individual success in the  $i$ th area. For example,  $y_i$  might be the number of people in normal BMI category, and  $\theta_i$  might be the corresponding normal BMI probability in our numerical example.

We assume fully exchangeability in the binomial-Beta model. Denoted as *BBM*, the binomial-Beta exchangeability (EX) model is

$$\begin{aligned} y_i | \theta_i &\sim \text{Binomial}(n_i, \theta_i), i = \dots, \ell, \\ \theta_i | \mu, \rho &\sim \text{Beta}\left\{\mu\left(\frac{1-\rho}{\rho}\right), (1-\mu)\left(\frac{1-\rho}{\rho}\right)\right\}, \\ \pi(\mu, \rho) &\propto 1, 0 < \mu, \rho < 1. \end{aligned}$$

where hyper-parameters  $\mu$  and  $\rho = 1/(1 + \tau)$  have a noninformative prior.

We assume partial exchangeability in the binomial-Beta model, which is the exchangeability-nonexchangeability (EXNEX) model. We can consider that each area  $i$  has its own mixture probability  $p_i$  of borrowing information. The distribution of  $\theta_i$  is a mixture of two Beta distributions. One component,  $\text{Beta}\left\{\mu\left(\frac{1-\rho}{\rho}\right), (1-\mu)\left(\frac{1-\rho}{\rho}\right)\right\}$ , permits borrowing information from the other areas where we assume fully exchangeability. The other component is the noninformative  $\text{Beta}(1, 1)$ , permits no borrowing information from the other areas where we assume nonexchangeability. The binomial-Beta EX model, *BBM*, is a special case of *BBM<sup>exnex</sup>* where the mixture probability is 1. Denoted as *BBM<sup>exnex</sup>*, the proposed binomial-Beta EXNEX model is

$$y_i | \theta_i \sim \text{Binomial}(n_i, \theta_i), i = \dots, \ell,$$

$$\theta_i | \mu, \rho, p_i \sim p_i \text{Beta}(\mu(\frac{1-\rho}{\rho}), (1-\mu)(\frac{1-\rho}{\rho})) + (1-p_i) \text{Beta}(1, 1),$$

$$p_i | \phi \sim \text{Beta}(\phi\tau_0, (1-\phi)\tau_0),$$

where  $\tau_0$  is specified as 1, because it is weakly identified,

$$\pi(\mu, \rho, \phi) \propto 1, \text{ where } 1/2 < \phi < 1, 0 < \mu, \rho < 1.$$

However, label switching problem in the Bayesian mixture models is a well-known problem, which is caused by symmetry in the likelihood of the model parameters (Stephens 2000). In particular, the invariance of the likelihood under relabeling of the mixture components can lead to the posterior distribution of the parameters being highly symmetric and multimodal, making it difficult to identify the posterior distribution of parameters. We want to impose constraints on the mixture probabilities  $p_i$ ,  $1/2 < p_i < 1$ , to have  $p_i > 1 - p_i$ . Because we want to use the exchangeable component for the most of the time. Constraints on  $p_i$  may cause other problems, such as computation difficulty. We assume  $1/2 < \phi < 1$  to deal with label switching in the mixture model, where  $\phi$  is the expectation of  $p_i$ , and it will not increase the computation difficulty.

The joint posterior distribution of  $BBM^{exnex}$  is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \rho, \mathbf{p}, \phi | \mathbf{y}) \propto \prod_{i=1}^{\ell} \left\{ \binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i-y_i} \{ p_i \text{Beta}(\mu(\frac{1-\rho}{\rho}), (1-\mu)(\frac{1-\rho}{\rho})) + (1-p_i) \text{Beta}(1, 1) \}^{\frac{p_i^{\phi\tau_0-1} (1-p_i)^{(1-\phi)\tau_0-1}}{B(\phi\tau_0, (1-\phi)\tau_0)}} \right\},$$

where  $B(\phi\tau_0, (1-\phi)\tau_0)$  is the normalization constant of the Beta distribution.

In Appendix 4.6.1, we show how to use a gridy Gibbs sampler to draw samples of  $\mu$ ,  $\rho$ , and  $\phi$  from the binomial-Beta ENXEX model,  $BBM^{exnex}$ . As in Appendix 4.6.1, we can draw the posterior samples of the mixture probability  $p_i$

easily.

Given the mixture probability  $p_i$ , we draw posterior samples from the conditional posterior distribution of  $\theta$ , which is a mixture of Beta distributions,

$$\theta_i | p_i, \mu, \rho, \mathbf{y} \sim p_i \text{Beta}\left\{y_i + \mu\left(\frac{1-\rho}{\rho}\right), n_i - y_i + (1-\mu)\left(\frac{1-\rho}{\rho}\right)\right\} + (1-p_i) \text{Beta}\{y_i + 1, n_i - y_i + 1\}.$$

#### 4.1.1 Simulation Under Three Scenarios

To show the robustness of the model with partial exchangeability, which is the EXNEX model, denoted as  $BBM^{exnex}$ , we consider three simulation scenarios with different heterogeneity. The heterogeneity between areas is related to the success probabilities  $\theta_i$  and the numbers of observations in each area. In the simulation, we assume the true success probabilities and the range of area size are known in 7 areas, given in Table 4.1.

Scenario	Area Sizes	$\theta$	Area1	Area2	Area3	Area4	Area5	Area6	Area7	Heterogeneity
1	5~15	$\theta_i^{true}$	0.2	0.2	0.2	0.2	0.2	0.2	0.2	'Small'
2	5~25	$\theta_i^{true}$	0.2	0.2	0.2	0.2	0.2	0.6	0.6	'Moderate'
3	5~40	$\theta_i^{true}$	0.1	0.1	0.4	0.4	0.4	0.8	0.8	'Large'

Table 4.1: True success probabilities  $\theta_i$  and area sizes for seven areas under three scenarios with different heterogeneity

Under Scenario 1, all areas have the same event probabilities and the area size is between 5 and 15, which means the heterogeneity between areas is relatively very small. Under Scenario 2, the last two areas have larger event probabilities than the first five areas, and the each area size is between 5 to 25, which means the heterogeneity between areas might be moderate. Under Scenario 3, the first

two areas have smaller event probabilities than the middle areas, the last two areas have larger event probabilities than the middle areas. We assume the each area size is between 5 to 40. The difference between each area under Scenario 3 is significantly larger than the difference under Scenario 2. Under Scenario 3, it may have relatively large heterogeneity. We expect that  $BBM^{exnex}$  model might be better than  $BBM$  model under Scenario 3.

We run 50 simulations for  $BBM^{exnex}$  and  $BBM$  under three scenarios. For each simulation, we run 15,000 MCMC iterations, take 5,000 as a ‘burn in ’ and use every 10th to obtain 1,000 converged posterior samples to maintain consistency. We provide convergence diagnostics for one of simulations, such as Geweke test, effective sample size. Table 4.2 provides p-values of the Geweke test, which are large to show the MCMC is stationary. Table 4.3 provides the effective sample sizes of posterior samples. Our Gibbs sampler method works well in the simulation.

	$\mu$	$\rho$	$\phi$
$BBM$	0.541	0.994	
$BBM^{exnex}$	0.413	0.642	0.099

Table 4.2: P-values of Geweke tests for  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $BBM$  and  $BBM^{exnex}$  for one simulation

	$\mu$	$\rho$	$\phi$
$BBM$	950	1000	
$BBM^{exnex}$	1000	1054	1000

Table 4.3: Effective sample sizes of  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $BBM$  and  $BBM^{exnex}$  for one simulation

Scenario	Model	PSD of $\theta_1$	PSD of $\theta_2$	PSD of $\theta_3$	PSD of $\theta_4$	PSD of $\theta_5$	PSD of $\theta_6$	PSD of $\theta_7$
1	<i>BBM</i>	0.083	0.089	0.080	0.071	0.089	0.080	0.101
	<i>BBM<sup>exnex</sup></i>	0.092	0.098	0.089	0.077	0.100	0.089	0.121
2	<i>BBM</i>	0.063	0.061	0.060	0.071	0.061	0.085	0.072
	<i>BBM<sup>exnex</sup></i>	0.069	0.069	0.067	0.084	0.067	0.110	0.087
3	<i>BBM</i>	0.064	0.065	0.072	0.053	0.108	0.109	0.070
	<i>BBM<sup>exnex</sup></i>	0.065	0.071	0.075	0.056	0.119	0.127	0.139

Table 4.4: Posterior standard deviation (PSD) of the success probabilities  $\theta_i$  for seven areas under three scenarios (one simulation)

Table 4.4 provides the posterior standard deviation of the success probabilities  $\theta_i$  for each area in models, *BBM* and *BBM<sup>exnex</sup>*, under three scenarios. Overall, the binomial-Beta EXNEX model, *BBM<sup>exnex</sup>*, has larger posterior standard deviation of the success probabilities  $\theta_i$  than the binomial-Beta model, *BBM*. The possible explanation is that the mixture structure in the EXNEX model, *BBM<sup>exnex</sup>*, causes the model to borrow less information among areas than the model *BBM*.

Scenario	Heterogeneity	Model	Relative Bias	RMSE
1	'Small'	<i>BBM</i>	0.202	0.097
		<i>BBM<sup>exnex</sup></i>	0.223	0.103
2	'Moderate'	<i>BBM</i>	0.215	0.124
		<i>BBM<sup>exnex</sup></i>	0.212	0.125
3	'Large'	<i>BBM</i>	0.346	0.168
		<i>BBM<sup>exnex</sup></i>	0.305	0.160

Table 4.5: Average relative bias and RMSE of  $\theta_i$  in Model *BBM* and *BBM<sup>exnex</sup>* under three scenarios (relatively different heterogeneity)

In Table 4.5, we compare the relative bias and RMSE of  $\theta$  in each model under three scenarios. The relative bias are  $\sum_{i=1}^7 \frac{|\theta_i^{true} - \hat{\theta}_i|}{\theta_i^{true}} / 7$ , where  $\theta_i^{true}$  from Table 4.1

and  $\hat{\theta}_i$  are the posterior means. RMSE is  $\sqrt{bias^2 + PSD^2}$ , where  $bias = \theta_i^{true} - \hat{\theta}_i$  and PSD is the posterior standard deviation of samples.

Under Scenario 1 with the very small difference between areas, all models have relatively small absolute bias. But *BBM* model has the smaller relative bias than *BBM<sup>exnex</sup>* model. However, under Scenario 3, *BBM<sup>exnex</sup>* model has the smaller relative bias than *BBM*.

One explanation is that *BBM* model can borrow much more information than *BBM<sup>exnex</sup>* model, which is helpful under Scenario 1. *BBM<sup>exnex</sup>* can borrow information from similar areas and avoid too optimistic borrowing information from very different areas, which shows smaller relative bias under Scenario 3. For this reason, *BBM<sup>exnex</sup>* model is more robust than *BBM* model under large heterogeneity scenarios. Since the heterogeneity between areas might be hard to assess in practice, the Bayesian hierarchical EXNEX model is recommended when areas might be very different from each other.

## 4.2 Multinomial-Dirichlet EXNEX Model without Order Restriction

EXNEX models allow for a small number of partial exchangeability structures, which is that different mixture probabilities  $p_i$  for each area  $i$  represent different ratio of fully exchangeability and non-exchangeability. The mixture probabilities are related to the percentage of information borrowed from other areas, which will be illustrated in a numerical example. This is a useful and more robust method to avoid pooling from very different areas and borrow from similar areas.

To have EXNEX structures in the multinomial-Dirichlet model without order



restriction, we have a following model, denoted as  $M_1^{exnex}$ ,

$$\mathbf{n}_i | \boldsymbol{\theta}_i \stackrel{ind}{\sim} \text{Multinomial}(\mathbf{n}_{i\cdot}, \boldsymbol{\theta}_i), \quad i = 1, \dots, \ell.$$

At the second stage, the distribution of  $\boldsymbol{\theta}_i$  is a mixture of two Dirichlet distributions,

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, p_i \stackrel{ind}{\sim} p_i \text{Dirichlet}(\boldsymbol{\mu}\tau) + (1 - p_i) \text{Dirichlet}(1, 1, \dots, 1), i = 1, \dots, \ell,$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$ ,  $\tau$  are hyper-parameters,

$$\pi(\boldsymbol{\mu}, \tau) = \frac{(K-1)!}{(1+\tau)^2}, \quad \tau, \mu_j > 0, \quad \sum_{j=1}^K \mu_j = 1.$$

One component in the distribution of  $\boldsymbol{\theta}_i$  is  $\text{Dirichlet}(\boldsymbol{\mu}\tau)$ , which permits borrowing information from the other areas where we assume fully exchangeability (EX). The other component is the noninformative  $\text{Dirichlet}(1, 1, \dots, 1)$ , which permits no borrowing information among areas where we assume non-exchangeability (NEX). In Chapter 2, the model  $M_1$  is a special case of  $M_1^{exnex}$ , where the mixture probabilities  $p_i$  is 1.

The hyper-prior of mixture probabilities  $p_i$  is

$$p_i | \phi, \tau_0 \sim \text{Beta}(\phi\tau_0, (1-\phi)\tau_0), \text{ where } \tau_0 \text{ is specified as 1,}$$

$$\phi \sim \text{Beta}(1, 1) \quad 1/2 < \phi < 1.$$

We use a constrained parameter space of  $\phi$  to deal with label switching problem in the mixture model, where we assume  $1/2 < \phi < 1$ . The joint posterior

distribution is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \tau, \mathbf{p}, \phi | \mathbf{n}) \propto \prod_{i=1}^{\ell} \left\{ \frac{n_{i\cdot}!}{\prod_{j=1}^K n_{ij}!} \prod_{j=1}^K \theta_{ij}^{n_{ij}} [p_i \text{Dirichlet}(\boldsymbol{\mu}\tau) \frac{(K-1)!}{(1+\tau)^2} + (1-p_i) \text{Dirichlet}(1)] \frac{p_i^{\phi\tau_0-1} (1-p_i)^{(1-\phi)\tau_0-1}}{B(\phi\tau_0, (1-\phi)\tau_0)} \right\}.$$

In Appendix 4.6.2, we show how to use griddy Gibbs sampler to draw  $\mu$ ,  $\rho$  and  $\phi$  from the multinomial-Dirichlet EXNEX model. We draw the posterior samples of the mixture probabilities  $p_i$  from the conditional posterior distribution. After having posterior samples of parameters  $\boldsymbol{\mu}$ ,  $\tau = 1/\rho - 1$ ,  $\phi$ ,  $p_i$ , we can draw the posterior of  $\boldsymbol{\theta}_i$  easily where

$$\boldsymbol{\theta}_i | p_i, \boldsymbol{\mu}, \tau, \mathbf{n} \sim p_i \text{Dirichlet}(\mathbf{n}_i + \boldsymbol{\mu}\tau) + (1-p_i) \text{Dirichlet}(\mathbf{n}_i + 1).$$

### 4.3 Multinomial-Dirichlet EXNEX Model with Order Restrictions

We present a multinomial-Dirichlet EXNEX model with order restrictions, denoted as  $M_3^{exnex}$ ,

$$\mathbf{n}_i | \boldsymbol{\theta}_i \stackrel{ind}{\sim} \text{Multinomial}(\mathbf{n}_{i\cdot}, \boldsymbol{\theta}_i), \quad i = 1, \dots, \ell.$$

At the second stage, the distribution of  $\boldsymbol{\theta}_i$  is a mixture of two Dirichlet distri-

butions,

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, p_i \stackrel{ind}{\sim} p_i \text{Dirichlet}(\boldsymbol{\mu}\tau) I_{\boldsymbol{\theta}_i \in C} + (1 - p_i) \text{Dirichlet}(1, 1, \dots, 1),$$

where  $I_{\boldsymbol{\theta}_i \in C}$  is an indicator that  $\boldsymbol{\theta}_i \in C$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$ ,  $\tau$  are hyper-parameters,

$$\pi(\boldsymbol{\mu}, \tau) = \frac{K(m-1)!(K-m)!}{(1+\tau)^2}, \quad \mu_j > 0, \quad \sum_{j=1}^K \mu_j = 1, \boldsymbol{\mu} \in C_\mu,$$

$$\text{and } C = \{\boldsymbol{\theta}_{ij} : \theta_{i1} \leq \theta_{i2} \leq \dots \leq \theta_{im} \geq \theta_{i(m+1)} \geq \dots \geq \theta_{iK}\},$$

$$C_\mu = \{\mu_j : \mu_1 \leq \mu_2 \leq \dots \leq \mu_m \geq \mu_{(m+1)} \geq \dots \geq \mu_K\}.$$

We assume the modal positions  $m$  in  $C$  and  $C_\mu$  are known.

One component in the distribution of  $\boldsymbol{\theta}_i$  is  $\text{Dirichlet}(\boldsymbol{\mu}\tau)$ , which permits borrowing information from the other areas, where we assume fully exchangeability (EX) and unimodal order restrictions on  $\boldsymbol{\theta}_i$ . The other component is the noninformative  $\text{Dirichlet}(1, 1, \dots, 1)$ , which permits no borrowing information among areas, where we assume non-exchangeability (NEX) and no order restriction. In Chapter 2, the model  $M_3$  is a special case of  $M_3^{exnex}$  where the mixture probabilities  $p_i$  is 1.

The hyper-prior of mixture probabilities  $p_i$  is

$$p_i | \phi, \tau_0 \sim \text{Beta}(\phi\tau_0, (1-\phi)\tau_0), \text{ where } \tau_0 \text{ is specified as 1,}$$

$$\phi \sim \text{Beta}(1, 1) \quad 1/2 < \phi < 1.$$

We assume  $1/2 < \phi < 1$  to deal with label switching in the mixture model.

The joint posterior distribution is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \tau, \boldsymbol{p}, \phi | \boldsymbol{n}) \propto \prod_{i=1}^{\ell} \left\{ \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \prod_{j=1}^K \theta_{ij}^{n_{ij}} \left[ p_i \frac{\text{Dirichlet}(\boldsymbol{\mu}\tau)}{\int_{\theta_i \in C} \text{Dirichlet}(\boldsymbol{\mu}\tau) d\theta_i} \frac{K(m-1)!(K-m)! I_{C\mu}}{(1+\tau)^2} + (1-p_i) \text{Dirichlet}(1) \right] \frac{p_i^{\phi\tau_0-1} (1-p_i)^{(1-\phi)\tau_0-1}}{B(\phi\tau_0, (1-\phi)\tau_0)} \right\}.$$

When there may be no unimodal order restriction in some small areas, those areas may have smaller mixture probabilities  $p_i$  than other areas. For example, in the following numerical example, County 21 may be considered a very-different area since the order restriction assumption that the mode is at the third position may not hold. We expect the mixture probability  $p_{21}$  of County 21 to be small. Then the EXNEX model with order restrictions,  $M_3^{exnex}$ , will borrow less information for County 21 from other counties.

In Appendix 4.6.3, we show how to use griddy Gibbs sampler to draw  $\mu$ ,  $\rho$  and  $\phi$  from the multinomial-Dirichlet EXNEX model. We draw the posterior samples of the mixture probability  $p_i$  from the conditional posterior distribution. After having posterior samples of parameters  $\boldsymbol{\mu}$ ,  $\rho$ ,  $\phi$ ,  $p_i$ , we can draw the posterior of  $\boldsymbol{\theta}_i$  easily, where

$$\boldsymbol{\theta}_i | \boldsymbol{\mu}, \tau, p_i, \boldsymbol{n} \sim p_i \text{Dirichlet}(\boldsymbol{n}_i + \boldsymbol{\mu}\tau) I_{\theta_i \in C} + (1-p_i) \text{Dirichlet}(\boldsymbol{n}_i + 1),$$

where  $I_{\theta_i \in C}$  is an indicator function that  $\theta_i \in C$ .

## 4.4 Numerical Example: BMI

We use BMI data from NHANES III to illustrate the multinomial-Dirichlet EXNEX model, denoted as  $M_1^{exnex}$  and the EXNEX model with order restrictions, denoted as  $M_3^{exnex}$ . Our primary purposes are to show that the EXNEX model can provide good estimates and borrow information from the similar areas while avoiding too optimistic borrowing from extreme areas. We fit  $M_1$ ,  $M_3$ ,  $M_1^{exnex}$ ,  $M_3^{exnex}$  in the numerical example. To be convenient, Table 4.6 provides the model notations used in the following discussion.

Table 4.6: Model notations in Chapter 4

Notation	Model	Order Restrictions	Location
$M_1$	the multinomial-Dirichlet model	NA	Chapter 2
$M_3$	the multinomial-Dirichlet model with order restrictions	$\theta_1 < \theta_2 < \theta_3 > \theta_4 > \theta_5$	Section 2.3
$M_1^{exnex}$	the multinomial-Dirichlet EXNEX model	NA	Section 4.2
$M_3^{exnex}$	the multinomial-Dirichlet EXNEX model with order restrictions	$\theta_1 < \theta_2 < \theta_3 > \theta_4 > \theta_5$	Section 4.3

For each model, we run 10,000 MCMC iterations, take 5,000 as a ‘burn in’ and use every 5th to obtain 1000 converged posterior samples. In Figure 4.1, the trace plots look like a random scatter around a mean value. There is no distinct jumps occur in the traces of  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_5)$  as the MCMC scheme moves. In Figure 4.2, they are unimodal distribution densities. We do not have the label switching problem. We have converged posterior samples and our MCMC is stationary. Table 4.7 provides p-values of the Geweke diagnostic to check the convergence of the parameters (Geweke 1992). All p-values are large to not reject the null hypothesis

that the MCMC is stationary. Table 4.8 provides the effective sample size of posterior samples. Our samples are independent since the effective sample size equals the actual sample size. Our Gibbs sampling method works well and the posterior samples can be used for the further inference.

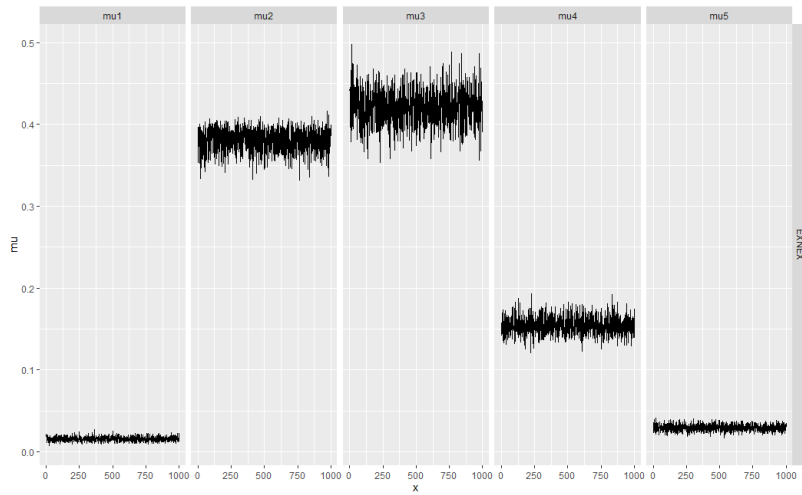


Figure 4.1: Traceplots of  $\mu$  in  $M_3^{exnex}$

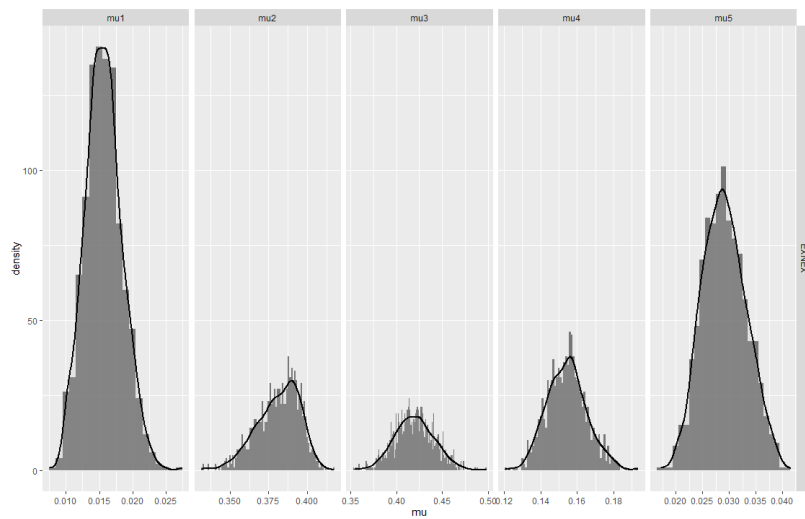


Figure 4.2: Densities of  $\mu$  in  $M_3^{exnex}$

	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\rho$	$\phi$
$M_1^{exnex}$	0.241	0.994	0.077	0.047	0.279	0.058	0.921
$M_3^{exnex}$	0.241	0.150	0.992	0.094	0.273	0.642	0.099

Table 4.7: P-values of Geweke tests for  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $M_1^{exnex}$  and  $M_3^{exnex}$

	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\rho$	$\phi$
$M_1^{exnex}$	950	1000	1060	866	1000	1025	1000
$M_3^{exnex}$	1000	1054	1057	1000	1000	1000	1000

Table 4.8: Effective sample sizes of  $\mu$ ,  $\rho$ , and  $\phi$  in the Model  $M_1^{exnex}$  and  $M_3^{exnex}$

In Appendix 4.6.4, we provide the posterior mean (PM) and posterior standard deviation (PSD), Posterior Coefficient of Variation (CV) of each category probability  $\theta$  after fitting four models. We also provide the posterior mean of mixture probabilities  $p_i$  in  $M_1^{exnex}$  and  $M_3^{exnex}$ . For mixture probabilities  $p_i$ , CVs are misleading for proportions.

Overall, the multinomial-Dirichlet EXNEX models,  $M_1^{exnex}$  and  $M_3^{exnex}$ , have relatively larger PSD than the multinomial-Dirichlet models,  $M_1$  and  $M_3$ , due to the partially borrowing information from other areas. The average mixture probabilities  $p_i$  are about 0.9 in  $M_1^{exnex}$  and  $M_3^{exnex}$ , which indicates that  $M_1^{exnex}$  and  $M_3^{exnex}$  borrow less information than  $M_1$  and  $M_3$ . For some extreme areas, such as County 3, County 21, County 27, the mixture probabilities  $p_i$  are smaller than the average and they borrow much less information from the other areas. Incorporating order restrictions,  $M_3$  has smaller PSD than  $M_1$ , and  $M_3^{exnex}$  has smaller PSD than  $M_1^{exnex}$ . But for some extreme areas such as County 21, the model with order restrictions may not have smaller PSD than the model without order restrictions. Incorporating order restrictions may cause those areas to be very different

from other areas, the mixture probabilities  $p_i$  may be smaller, and borrow less information from other areas.

To see the feature of EXNEX model, which is that it can borrow from similar areas and avoid borrowing from extreme areas, we discuss the posterior summary for four representative areas, such as County 3, County 9, County 21 and County 27 here. Areas where the unimodal order restrictions may not hold can be considered as extreme areas, such as County 21. Areas where the unimodal order restrictions are too good to be true can be considered as extreme areas, such as County 3 and County 27. In those extreme areas, the mixture probabilities  $p_i$  will be relatively small, which borrow less information from other areas.

Table 4.9: Posterior summary of  $\theta$  in four models for County 3

Model	$p_i$	$\theta$									
		Underweight		Normal		Overweight		Obese I		Obese II	
		PM1	PSD1	PM2	PSD2	PM3	PSD3	PM4	PSD4	PM5	PSD5
$M_1$		0.026	0.013	0.305	0.049	0.477	0.045	0.147	0.027	0.045	0.022
$M_3$		0.024	0.011	0.308	0.034	0.479	0.037	0.148	0.027	0.042	0.015
$M_1^{exnex}$	0.920	0.041	0.020	0.233	0.042	0.516	0.051	0.162	0.036	0.048	0.021
$M_3^{exnex}$	0.794	0.024	0.015	0.311	0.056	0.474	0.048	0.152	0.025	0.040	0.018

Note: Mixture Probabilities  $p_i$ , Posterior Mean (PM), Posterior Standard Deviation (PSD)

In Table 4.9, the mixture probability  $p_i$  in  $M_1^{exnex}$  is 0.92. However, the mixture probability  $p_i$  in  $M_3^{exnex}$  is 0.794, which is smaller than 0.92 even the order restriction may hold in County 3. The cell counts of County 3 are (3, 20, 49, 13, 5). One possible explanation is that the cell counts of County 3 are much smaller than the cell counts of County 4, which are (2, 145, 174, 77, 14). The cell counts of County 2 are (1, 36, 38, 15, 1). We notice that the order restriction assumption may hold in County 2, County 3, and County 4. But County 3 might be too good



to be true, since 20 observations in normal BMI level and 49 observations in overweight BMI level, which will make County 3 as a very-different area than others. We also notice the posterior means in  $M_1$  and  $M_3$  are very close. One possible explanation is that the unimodal order restrictions that the modal position is at the third may hold in County 3. But the posterior standard deviation in  $M_3$  is smaller than in  $M_1$ , due to borrowing information from other areas.

Table 4.10: Posterior summary of  $\theta$  in four models for County 27

Model	$p_i$	$\theta$									
		Underweight		Normal		Overweight		Obese I		Obese II	
		PM1	PSD1	PM2	PSD2	PM3	PSD3	PM4	PSD4	PM5	PSD5
$M_1$		0.027	0.016	0.322	0.056	0.459	0.049	0.162	0.033	0.030	0.024
$M_3$		0.024	0.013	0.328	0.037	0.460	0.042	0.162	0.031	0.027	0.014
$M_1^{exnex}$	0.934	0.054	0.030	0.230	0.052	0.491	0.062	0.209	0.051	0.016	0.016
$M_3^{exnex}$	0.869	0.025	0.020	0.327	0.061	0.455	0.047	0.166	0.031	0.028	0.014

Note: Mixture Probabilities  $p_i$ , Posterior Mean (PM), Posterior Standard Deviation (PSD)

From Table 4.10, the mixture probability  $p_i$  in  $M_1^{exnex}$  is 0.934, where the count data in County 27 is  $(2, 10, 26, 9, 1)$ . However the mixture probability  $p_i$  in  $M_3^{exnex}$  is 0.869, which is smaller than 0.934. County 27 has the similar problem as County 3, which is that the number of overweight dominates the other categories. The unimodal order restrictions in County 27 may be too good to be true, which makes County 27 as a very-different area than others. Similar to County 3, we also notice the posterior means in  $M_1$  and  $M_3$  are very close. But the posterior standard deviation in  $M_3$  is smaller than in  $M_1$ , due to borrowing information from other areas.

Table 4.11: Posterior summary of  $\theta$  in four models for County 9

Model	$p_i$	$\theta$									
		Underweight		Normal		Overweight		Obese I		Obese II	
		PM1	PSD1	PM2	PSD2	PM3	PSD3	PM4	PSD4	PM5	PSD5
$M_1$		0.017	0.011	0.394	0.04	0.394	0.042	0.171	0.031	0.024	0.019
$M_3$		0.016	0.010	0.368	0.024	0.417	0.029	0.175	0.032	0.023	0.012
$M_1^{exnex}$	0.945	0.026	0.017	0.385	0.053	0.373	0.053	0.203	0.044	0.012	0.012
$M_3^{exnex}$	0.914	0.015	0.009	0.373	0.026	0.416	0.034	0.172	0.029	0.025	0.011

Note: Mixture Probabilities  $p_i$ , Posterior Mean (PM), Posterior Standard Deviation (PSD)

From Table 4.11, the mixture probability  $p_i$  in  $M_1^{exnex}$  is 0.945 and the mixture probability  $p_i$  in  $M_3^{exnex}$  is 0.914, which are not much different, where the count data of County 9 is (1, 29, 28, 14, 1). Even the order restriction assumption may not hold, the number of observations in normal BMI level is very close to the number of observations in overweight BMI level. County 9 may not be considered as a very-different area and the mixture proportion  $p_i$  is 0.914, which is close to other areas'.

Table 4.12: Posterior summary of  $\theta$  in four models for County 21

Model	$p_i$	$\theta$									
		Underweight		Normal		Overweight		Obese I		Obese II	
		PM1	PSD1	PM2	PSD2	PM3	PSD3	PM4	PSD4	PM5	PSD5
$M_1$		0.044	0.024	0.426	0.048	0.361	0.053	0.123	0.032	0.046	0.032
$M_3$		0.041	0.017	0.373	0.023	0.405	0.027	0.140	0.033	0.042	0.017
$M_1^{exnex}$	0.878	0.087	0.036	0.449	0.063	0.292	0.055	0.124	0.042	0.049	0.027
$M_3^{exnex}$	0.578	0.055	0.042	0.412	0.064	0.356	0.080	0.125	0.043	0.052	0.033

Note: Mixture Probabilities  $p_i$ , Posterior Mean (PM), Posterior Standard Deviation (PSD)

From Table 4.12, in  $M_1^{exnex}$ , the mixture probability  $p_i$  is 0.878, which is smaller than others. In  $M_3^{exnex}$ , the mixture probability  $p_i$  is 0.578, which is

extremely smaller than others. The count data in County 21 is  $(1, 36, 29, 9, 2)$ , which does not match the third modal position assumption and the order restriction assumption may not hold. It suggested that the EXNEX model can avoid optimistic borrowing from the very different areas. The posterior mean of  $\theta$  are  $(0.055, 0.412, 0.356, 0.125, 0.052)$ , and the mode is at the second position. In the EXNEX model with order restriction,  $M_3^{exnex}$ , County 21 should be considered as a very-different area compared with other counties, since 36 observations in normal BMI level and 29 observations in overweight BMI level.

In Figure 4.3, Figure 4.4, Figure 4.5, and Figure 4.6, we provide the posterior densities of  $\theta$  in  $M_3$  and  $M_3^{exnex}$  for counties 2, 3, 13, and 35, correspondingly. The difference between  $M_3^{exnex}$  and  $M_3$  may be not significant since the mixture proportions of EXNEX in  $M_3^{exnex}$  are around 0.91.

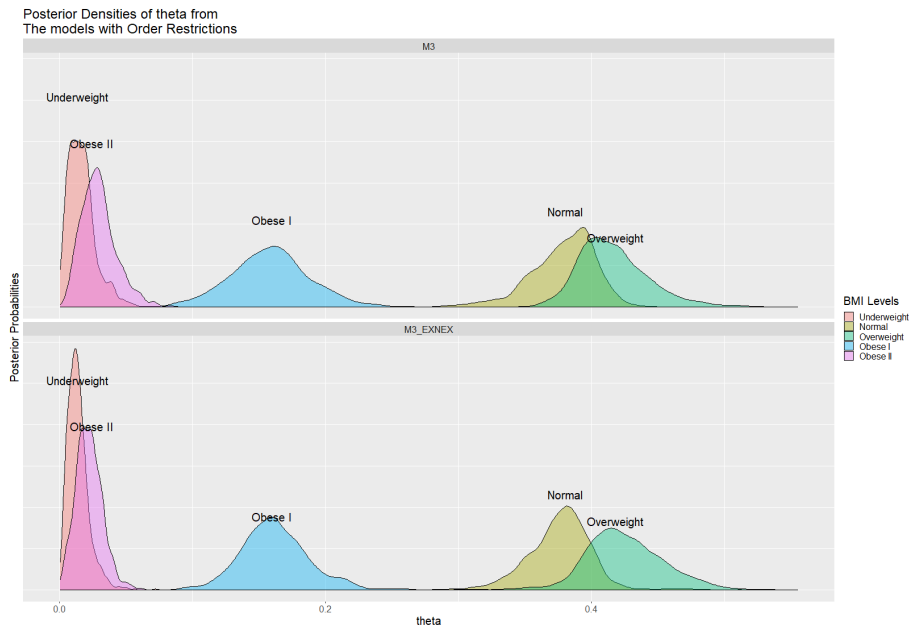


Figure 4.3: Densities of  $\theta$  in  $M_3$  and  $M_3^{exnex}$  for County 2

Figure 4.3 provides the posterior densities of  $\theta$  for County 2. The density curves in  $M_3^{exnex}$  are slightly flatter than in  $M_3$ . But the difference of density curves between  $M_3$  and  $M_3^{exnex}$  is not significant. The model  $M_3^{exnex}$  borrows almost the same amount of information from other counties as the model  $M_3$  does.

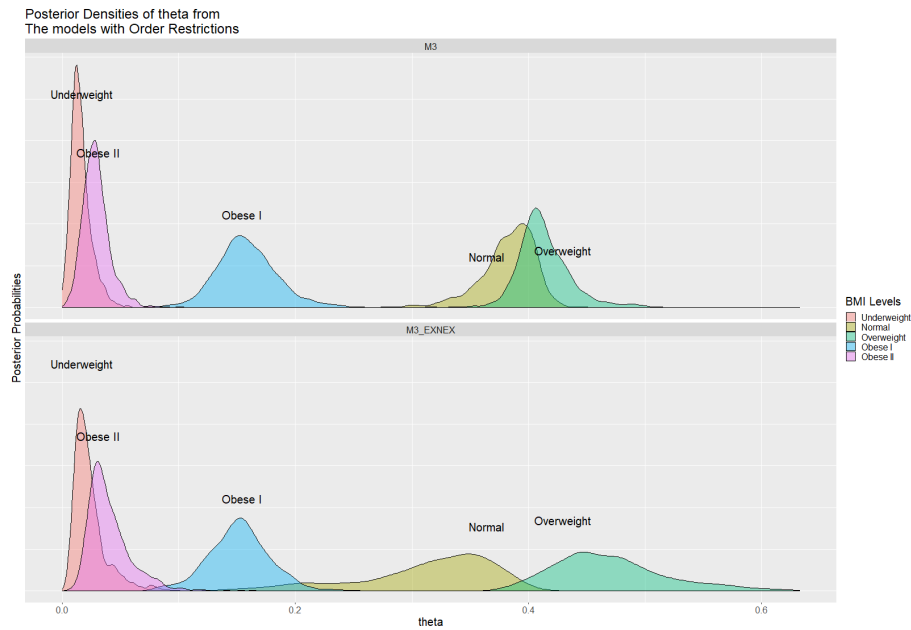


Figure 4.4: Densities of  $\theta$  in  $M_3$  and  $M_3^{exnex}$  for County 3

Figure 4.4 provides the posterior densities of  $\theta$  for County 3. The density curves in  $M_3^{exnex}$  are flatter than in  $M_3$ . That is because the mixture probability is 0.794 and the model  $M_3^{exnex}$  borrows much less information than the model  $M_3$  does. As discussed in the previous, one possible explanation is that a significant difference between County 3 and other counties, such as County 2 and County 4, which causes a small mixture probability.

In Figure 4.5 and Figure 4.6, the posterior density curves in  $M_3$  and  $M_3^{exnex}$  for County 13 and County 35 do not have significant difference. With order restric-

tions, County 13 and County 35 are similar with other counties where the order restriction assumption may hold. The mixture probabilities  $p_{13}$  and  $p_{35}$  are around 0.91, which is close to 1. For this reason, the performances of  $M_3$  and  $M_3^{exnex}$  are similar.

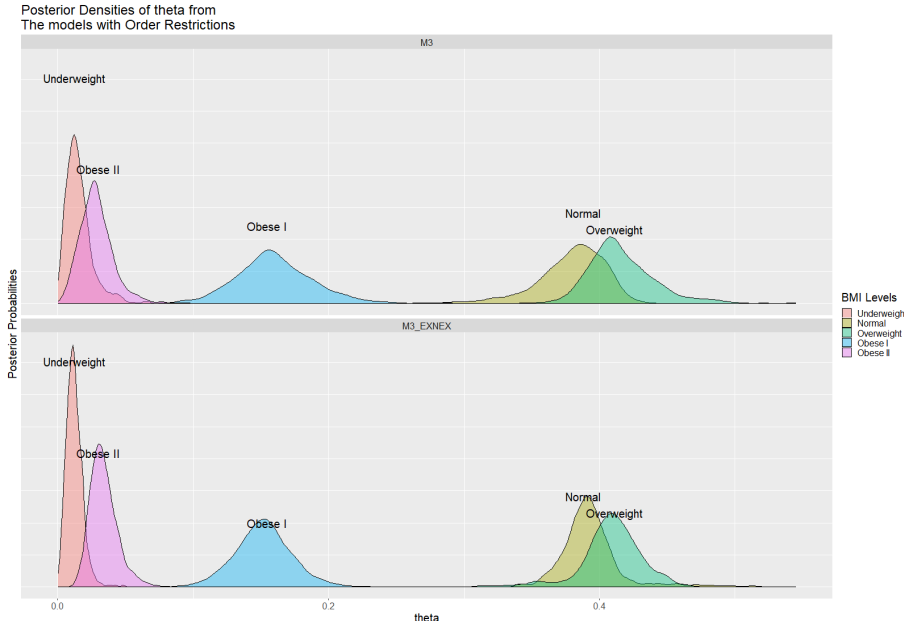


Figure 4.5: Densities of  $\theta$  in  $M_3$  and  $M_3^{exnex}$  for County 13

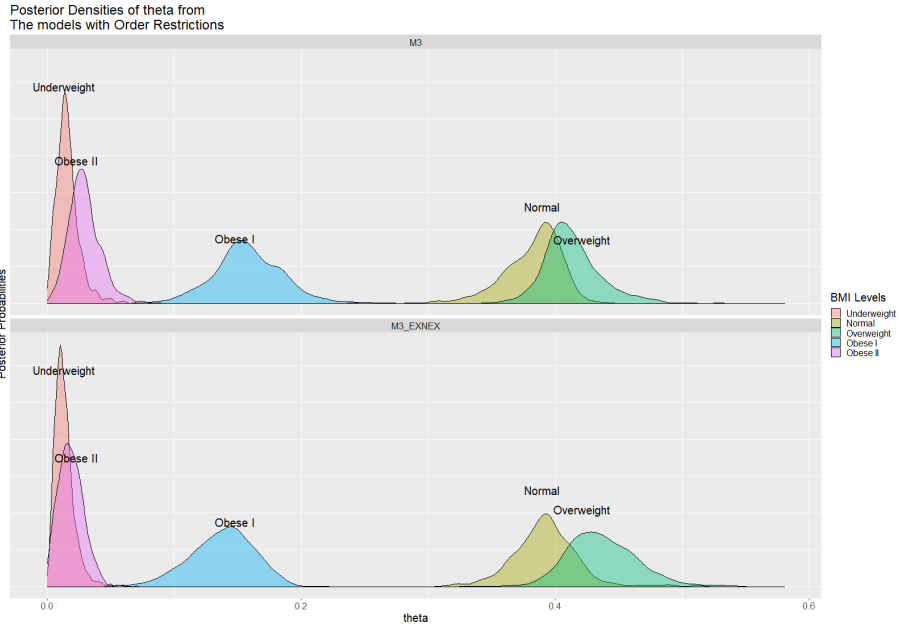


Figure 4.6: Densities of  $\theta$  in  $M_3$  and  $M_3^{exnex}$  for County 35

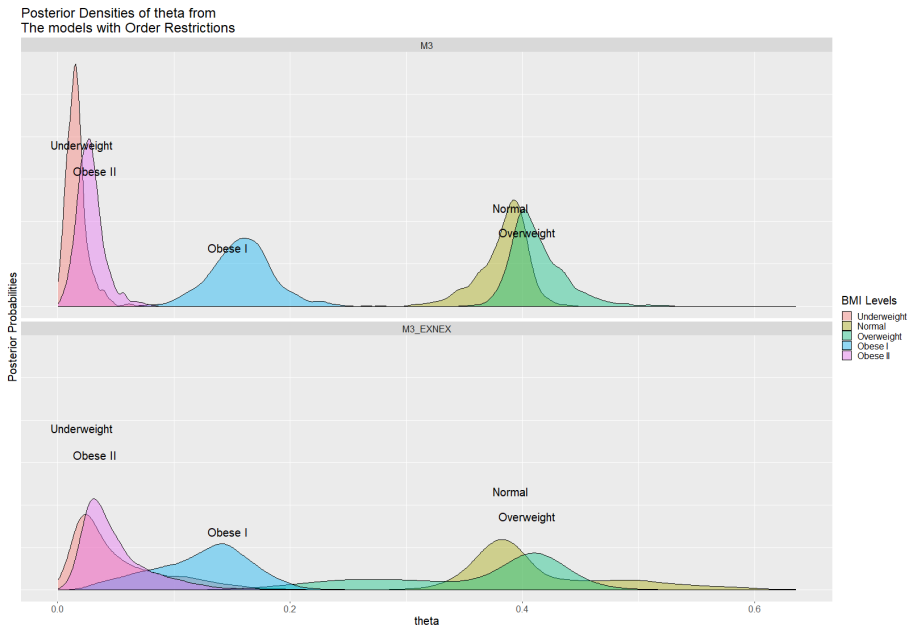


Figure 4.7: Densities of  $\theta$  in  $M_3$  and  $M_3^{exnex}$  for County 21

Additional to those, Figure 4.7 provide the posterior density curves of  $\theta_{21}$  in  $M_3$  and  $M_3^{exnex}$ . Large overlaps between normal BMI level and overweight BMI level in both models suggest that the unimodal order restriction assumption may not hold for County 21. We should avoid borrowing too much information from other counties. Therefore the EXNEX model with order restrictions,  $M_3^{exnex}$ , is a better fit for this situation than model  $M_3$ .

## 4.5 Concluding Remarks

In summary, extending the full exchangeability (EX) model to mixtures that allow for exchangeability and nonexchangeability looks promising. From the simulation of binomial-Beta EXNEX model,  $BBM^{exnex}$ , the model with partial exchangeability will have smaller relative bias and are more robust under different scenarios than the fully exchangeability model, such as  $BBM$ . We present the multinomial-Dirichlet EXNEX model with order restrictions,  $M_3^{exnex}$ . We notice the computational difficulties under order restriction assumption.

In the numerical example, the EXNEX model with order restrictions,  $M_3^{exnex}$ , can borrow from similar areas and avoid borrowing too much from extreme areas, since it can have different mixture probabilities  $p_i$  for each area. Extreme areas could be the areas where the unimodal order restriction may not hold, such as County 21, or the areas where the unimodal structure may be too good to be true, such as County 3. Different mixture probabilities  $p_i$  are related to different portions of fully exchangeability and nonexchangeability, which means different portions of information borrowed from other areas. The mixture probabilities  $p_i$  of EXNEX models for  $i$ th area are related to the difference between that area and

others. Large difference causes small mixture probability  $p_i$ , and causes less information borrowed from other areas.

However the simulation of the EXNEX model with order restrictions should be provided in the future to show the robustness of the EXNEX model under different scenarios. Careful considerations for the specification of prior distributions and mixture weights are important, which may affect the partial exchangeability assumption in the model (Neuenschwander et al. 2016). Due to the complexity of the EXNEX model, we have much more parameters in the EXNEX model for the same amount of data than the EX model. So over-fitting should be concerned.



## 4.6 Appendix

### 4.6.1 Griddy Gibbs Sampler for the Binomial-Beta EXNEX Model

We show how to use a griddy Gibbs sampler to draw samples of  $\mu$ ,  $\rho$ , and  $\phi$  from the binomial-Beta EXNEX model,  $BBM^{exnex}$ . Specifically, we use 200 points in the uniformly spaced grid. The joint posterior distribution of  $BBM^{exnex}$  is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \rho, \mathbf{p}, \phi | \mathbf{y}) \propto \prod_{i=1}^{\ell} \left\{ \binom{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n_i - y_i} \left\{ p_i \text{Beta}\left(\mu \frac{1 - \rho}{\rho}, (1 - \mu) \frac{1 - \rho}{\rho}\right) + (1 - p_i) \text{Beta}(1, 1) \right\} \frac{p_i^{\phi \tau_0 - 1} (1 - p_i)^{(1 - \phi) \tau_0 - 1}}{B(\phi \tau_0, (1 - \phi) \tau_0)} \right\},$$

where  $B(\phi \tau_0, (1 - \phi) \tau_0)$  is the normalization constant of the Beta distribution.

Since  $E(p_i A + (1 - p_i) B) = A \phi + B(1 - \phi)$ , we can integrate out  $p_i$  and  $\theta_i$ ,

$$\begin{aligned} \pi(\boldsymbol{\mu}, \rho, \phi | \mathbf{y}) &\propto \prod_{i=1}^{\ell} \left\{ \int_{p_i, \theta_i} \binom{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n_i - y_i} \left\{ p_i \text{Beta}\left(\mu \frac{1 - \rho}{\rho}, (1 - \mu) \frac{1 - \rho}{\rho}\right) + (1 - p_i) \text{Beta}(1, 1) \right\} \frac{p_i^{\phi \tau_0 - 1} (1 - p_i)^{(1 - \phi) \tau_0 - 1}}{B(\phi \tau_0, (1 - \phi) \tau_0)} dp_i d\theta_i \right\} \\ &= \prod_{i=1}^{\ell} \left\{ \int_{p_i} \binom{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n_i - y_i} \left\{ \phi \text{Beta}\left(\mu \frac{1 - \rho}{\rho}, (1 - \mu) \frac{1 - \rho}{\rho}\right) + (1 - \phi) \text{Beta}(1, 1) \right\} dp_i \right\} \\ &= \prod_{i=1}^{\ell} \left\{ \phi \binom{n_i}{y_i} \frac{B(y_i + \mu(1/\rho - 1), (n_i - y_i) + (1 - \mu)(1/\rho - 1))}{B(\mu(1/\rho - 1), (1 - \mu)(1/\rho - 1))} + (1 - \phi) \binom{n_i}{y_i} \frac{B(y_i + 1, (n_i - y_i) + 1)}{B(1, 1)} \right\}. \end{aligned}$$

Then we use a griddy Gibbs sampler to draw posterior samples of  $\mu$ ,  $\rho$ , and  $\phi$ .

Next, with the help of a latent variable  $z_i$ , we can draw samples of  $p_i$  from the following part,

$$\pi(p_i, z_i | \mu, \rho, \phi, \mathbf{y}) \propto [p_i \binom{n_i}{y_i} \frac{B\{y_i + \mu(1/\rho - 1), (n_i - y_i) + (1 - \mu)(1/\rho - 1)\}}{B\{\mu(1/\rho - 1), (1 - \mu)(1/\rho - 1)\}}]^{z_i} [(1 - p_i) \frac{1}{n_i + 1}]^{1 - z_i} p_i^{\phi \tau_0 - 1} (1 - p_i)^{(1 - \phi) \tau_0 - 1},$$

then draw  $z_i$  and  $p_i$

$$z_i | p_i, \mu, \rho, \mathbf{y} \sim \text{Bernoulli} \left\{ \frac{p_i \binom{n_i}{y_i} \frac{B[y_i + \mu(1/\rho - 1), (n_i - y_i) + (1 - \mu)(1/\rho - 1)]}{B[\mu(1/\rho - 1), (1 - \mu)(1/\rho - 1)]}}{p_i \binom{n_i}{y_i} \frac{B[y_i + \mu(1/\rho - 1), (n_i - y_i) + (1 - \mu)(1/\rho - 1)]}{B[\mu(1/\rho - 1), (1 - \mu)(1/\rho - 1)]} + (1 - p_i) \frac{1}{n_i + 1}} \right\},$$

$$p_i | z_i, \phi \sim \text{Beta}\{z_i + \phi\tau_0, (1 - z_i) + (1 - \phi)\tau_0\}.$$

#### 4.6.2 Griddy Gibbs Sampler for the EXNEX Model $M_1^{exnex}$

We show how to use griddy Gibbs sampler to draw sample from the multinomial-Dirichlet EXNEX model. After integrating out  $p_i$  and  $\theta_i$ , the joint posterior of  $\mu, \tau, \phi$  in  $M_1^{exnex}$  is

$$\begin{aligned} \pi(\boldsymbol{\mu}, \tau, \phi | \mathbf{n}) &\propto \prod_{i=1}^{\ell} \left\{ \int_{p_i, \boldsymbol{\theta}_i} \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \prod_{j=1}^K \theta_i^{n_{ij}} [p_i \text{Dirichlet}(\boldsymbol{\mu}\tau) \frac{(K-1)!}{(1+\tau)^2} + \right. \\ &\quad \left. (1-p_i) \text{Dirichlet}(1)] \frac{p_i^{\phi\tau_0-1} (1-p_i)^{(1-\phi)\tau_0-1}}{B(\phi\tau_0, (1-\phi)\tau_0)} dp_i d\boldsymbol{\theta}_i \right\} \\ &= \prod_{i=1}^{\ell} \left\{ \phi \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + \boldsymbol{\mu}\tau)}{D(\boldsymbol{\mu}\tau)} \frac{(K-1)!}{(1+\tau)^2} + (1-\phi) \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + 1)}{D(1)} \right\}, \\ \pi(\boldsymbol{\mu}, \tau, \phi | \mathbf{n}) &\propto \prod_{i=1}^{\ell} \left\{ \phi \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\sum_j n_{ij} + \tau)}}{\frac{\prod_{j=1}^K \Gamma(\mu_j \tau)}{\Gamma(\tau)}} \frac{(K-1)!}{(1+\tau)^2} + (1-\phi) \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij} + 1)}{\Gamma(\sum_j n_{ij} + K)}}{\frac{1}{\Gamma(K)}} \right\}. \end{aligned}$$

Using the transformation  $\rho = \frac{1}{\tau+1}$  ( $\tau = 1/\rho - 1$ ),

$$\begin{aligned} \pi(\boldsymbol{\mu}, \tau, \phi | \mathbf{n}) &\propto \prod_{i=1}^{\ell} \left\{ \phi \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij} + \mu_j(1/\rho - 1))}{\Gamma(\sum_j n_{ij} + (1/\rho - 1))}}{\frac{\prod_{j=1}^K \Gamma(\mu_j(1/\rho - 1))}{\Gamma((1/\rho - 1))}} (K-1)! + (1-\phi) \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij} + 1)}{\Gamma(\sum_j n_{ij} + K)}}{\frac{1}{\Gamma(K)}} \right\} \\ &\propto \prod_{i=1}^{\ell} \left\{ \phi \frac{\Gamma(n_i + 1) \Gamma(K)}{\prod_{j=1}^K \Gamma(n_{ij} + 1)} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij} + \mu_j(1/\rho - 1))}{\Gamma(\sum_j n_{ij} + (1/\rho - 1))}}{\frac{\prod_{j=1}^K \Gamma(\mu_j(1/\rho - 1))}{\Gamma((1/\rho - 1))}} + (1-\phi) \frac{\Gamma(n_i + 1) \Gamma(K)}{\Gamma(n_i + K)} \right\}. \end{aligned}$$

We can use griddy Gibbs sampling to draw  $\mu, \tau, \phi$ , and  $0 \leq \mu_j \leq \sum_{t=1, t \neq j}^{K-1} \mu_t$ ,  $j = 1, \dots, K-1$ .

Then we show how to draw the mixture probabilities  $p_i$  with the help of a latent variable  $z_i$ ,

$$\begin{aligned} \pi(p_i, z_i | \boldsymbol{\mu}, \rho, \phi, \mathbf{n}) \propto & \{p_i \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + \boldsymbol{\mu}(1/\rho - 1))}{D(\boldsymbol{\mu}(1/\rho - 1))} \Gamma(K)\}^{z_i} \{(1 - p_i) \frac{\Gamma(n_i + 1)\Gamma(K)}{\Gamma(n_i + K)}\}^{1 - z_i} \\ & \times p_i^{\phi\tau_0 - 1} (1 - p_i)^{(1 - \phi)\tau_0 - 1}, \end{aligned}$$

then we draw sample from the conditional distributions,

$$z_i | p_i, \boldsymbol{\mu}, \rho, \mathbf{n} \sim \text{Bernoulli}\left(\frac{p_i \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + \boldsymbol{\mu}(1/\rho - 1))}{D(\boldsymbol{\mu}(1/\rho - 1))} \Gamma(K)}{p_i \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + \boldsymbol{\mu}(1/\rho - 1))}{D(\boldsymbol{\mu}(1/\rho - 1))} \Gamma(K) + (1 - p_i) \frac{\Gamma(n_i + 1)\Gamma(K)}{\Gamma(n_i + K)}}\right),$$

$$p_i | z_i, \phi, \tau_0 = 1 \sim \text{Beta}(z_i + \phi\tau_0, (1 - z_i) + (1 - \phi)\tau_0).$$

### 4.6.3 Griddy Gibbs Sampler for the EXNEX Model with Order Restrictions $M_3^{exnex}$

We show how to use a griddy Gibbs sampler to draw sample from the multinomial-Dirichlet model with ENXEX prior. Specifically, we use 200 points in the uniformly spaced grid. After integrating out  $p_i$  and  $\theta_i$ , the joint posterior of  $\boldsymbol{\mu}, \tau, \phi$  in  $M_3^{exnex}$  is

$$\begin{aligned} \pi(\boldsymbol{\mu}, \tau, \phi | \mathbf{n}) \propto & \prod_{i=1}^{\ell} \left\{ \int_{p_i, \boldsymbol{\theta}_i} \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \prod_{j=1}^K \theta_{ij}^{n_{ij}} [p_i \int_{\theta_i \in C} \frac{\text{Dirichlet}(\boldsymbol{\mu}\tau)}{\text{Dirichlet}(\boldsymbol{\mu}\tau)} d\boldsymbol{\theta}_i \frac{K(m-1)!(K-m)! I_{C_{\boldsymbol{\mu}}} +}{(1+\tau)^2} \right. \\ & \left. (1 - p_i) \text{Dirichlet}(1)] \frac{p_i^{\phi\tau_0 - 1} (1 - p_i)^{(1 - \phi)\tau_0 - 1}}{B(\phi\tau_0, (1 - \phi)\tau_0)} dp_i d\boldsymbol{\theta}_i \right\} \\ = & \prod_{i=1}^{\ell} \left\{ \phi \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + \boldsymbol{\mu}\tau) C(\mathbf{n}_i + \boldsymbol{\mu}\tau)}{D(\boldsymbol{\mu}\tau) C(\boldsymbol{\mu}\tau)} \frac{K(m-1)!(K-m)! I_{C_{\boldsymbol{\mu}}} +}{(1+\tau)^2} \right. \\ & \left. (1 - \phi) \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{D(\mathbf{n}_i + 1)}{D(1)} \right\}, \end{aligned}$$

where  $D(\boldsymbol{\mu}\tau)$  and  $D(\mathbf{n}_i + \boldsymbol{\mu}\tau)$  are the normalization constant of Dirichlet distributions, and  $C(\boldsymbol{\mu}\tau)$  and  $C(\mathbf{n}_i + \boldsymbol{\mu}\tau)$  are the adjusted constant of Dirichlet distributions with order restrictions.

$$\pi(\boldsymbol{\mu}, \tau, \phi | \mathbf{n}) \propto \prod_{i=1}^{\ell} \left\{ \phi \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau + n_{ij} - 1} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \tau - 1} d\boldsymbol{\theta}_i} \frac{K(m-1)!(K-m)! I_{C_\mu}}{(1+\tau)^2} + (1-\phi) \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij}+1)}{\Gamma(\sum_j n_{ij}+K)}}{\frac{1}{\Gamma(K)}} \right\}.$$

After transforming  $\tau$  into  $\rho$  ( $\tau = \frac{1-\rho}{\rho}$ ), use Monte Carlo integration to deal the ratio of two integrals, draw samples of  $\boldsymbol{\theta}$ :  $\boldsymbol{\theta}^{(q)} \sim \text{Dirichlet}(\gamma \bar{\mathbf{n}}_j)$ , where  $\boldsymbol{\theta} \in C$  and  $\gamma$  is the importance ratio,  $\bar{\mathbf{n}}_j$  is the average over small areas for each category, the approximated joint posterior distribution of  $\boldsymbol{\mu}$  and  $\rho$  is

$$\begin{aligned} \pi(\boldsymbol{\mu}, \rho, \phi | \mathbf{n}) &\propto \prod_{i=1}^{\ell} \left\{ \phi \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \frac{1-\rho}{\rho} + n_{ij} - 1} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i \in C} \prod_{j=1}^K \theta_{ij}^{\mu_j \frac{1-\rho}{\rho} - 1} d\boldsymbol{\theta}_i} K(m-1)!(K-m)! + (1-\phi) \frac{n_i!}{\prod_{j=1}^K n_{ij}!} \frac{\frac{\prod_{j=1}^K \Gamma(n_{ij}+1)}{\Gamma(\sum_j n_{ij}+K)}}{\frac{1}{\Gamma(K)}} \right\} \\ &\propto \prod_{i=1}^{\ell} \left\{ \phi \frac{\Gamma(n_i+1)K(m-1)!(K-m)!}{\prod_{j=1}^K \Gamma(n_{ij}+1)} \frac{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij} - \gamma \bar{\mathbf{n}}_j + \mu_j \frac{1-\rho}{\rho}}}{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma \bar{\mathbf{n}}_j + \mu_j \frac{1-\rho}{\rho}}} + (1-\phi) \frac{\Gamma(n_i+1)\Gamma(K)}{\Gamma(n_i+K)} \right\} \\ &\propto \prod_{i=1}^{\ell} \left\{ \phi \frac{\Gamma(n_i+1)K(m-1)!(K-m)!}{\prod_{j=1}^K \Gamma(n_{ij}+1)} \left\{ \sum_{q=1}^M w_q \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij}} \right\} + (1-\phi) \frac{\Gamma(n_i+1)\Gamma(K)}{\Gamma(n_i+K)} \right\}, \end{aligned}$$

$$\text{where } w_q = \frac{\prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma \bar{\mathbf{n}}_j + \mu_j \frac{1-\rho}{\rho}}}{\sum_{q=1}^M \prod_{j=1}^K [\theta_j^{(q)}]^{-\gamma \bar{\mathbf{n}}_j + \mu_j \frac{1-\rho}{\rho}}}.$$

We can use gridy Gibbs sampling to draw  $\mu, \rho, \phi$ , and for  $j$  from  $m - 1$  to 1 the range of  $\mu_j$  is

$$0 < \mu_j \leq \min\left\{\mu_{j+1}, \frac{1 - \sum_{t=1, t \neq m, t \neq j}^K \mu_t}{2}\right\},$$

and for  $j$  from  $m + 1$  to  $K$ ,

$$0 < \mu_j \leq \min\left\{\mu_{j-1}, \frac{1 - \sum_{t=1, t \neq m, t \neq j}^K \mu_t}{2}\right\}.$$

Then we draw the mixture probabilities  $p_i$  with the help of a latent variable  $z_i$ ,

$$z_i | p_i, \boldsymbol{\mu}, \rho, \boldsymbol{n} \sim \text{Bernoulli}\left(\frac{p_i \frac{\Gamma(n_i + 1) K (m-1)! (K-m)!}{\prod_{j=1}^K \Gamma(n_{ij} + 1)} \left\{ \sum_{q=1}^M w_q \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij}} \right\}}{p_i \frac{\Gamma(n_i + 1) K (m-1)! (K-m)!}{\prod_{j=1}^K \Gamma(n_{ij} + 1)} \left\{ \sum_{q=1}^M w_q \prod_{j=1}^K [\theta_j^{(q)}]^{n_{ij}} \right\} + (1 - p_i) \frac{\Gamma(n_i + 1) \Gamma(K)}{\Gamma(n_i + K)}}\right)$$

$$p_i | z_i, \phi, \tau_0 = 1 \sim \text{Beta}(z_i + \phi \tau_0, (1 - z_i) + (1 - \phi) \tau_0).$$

#### 4.6.4 Posterior Summary of Model $M_1$ , $M_3$ , $M_1^{exnex}$ , and $M_3^{exnex}$ for BMI

Table 4.13: EXNEX posterior summary part I: Counties 1-9

County	Models	$p_i$	$\theta$														
			PM1	PSD1	CV1	PM2	PSD2	CV2	PM3	PSD3	CV3	PM4	PSD4	CV4	PM5	PSD5	CV5
1	$M_1$		0.025	0.012	0.480	0.402	0.038	0.095	0.394	0.038	0.096	0.142	0.025	0.176	0.038	0.021	0.553
	$M_3$		0.024	0.012	0.500	0.375	0.023	0.061	0.418	0.026	0.062	0.147	0.028	0.190	0.036	0.014	0.389
	$M_1^{exnex}$	0.946	0.041	0.020	0.488	0.398	0.047	0.118	0.372	0.045	0.121	0.152	0.034	0.224	0.037	0.018	0.486
	$M_3^{exnex}$	0.912	0.021	0.011	0.524	0.380	0.023	0.061	0.415	0.031	0.075	0.151	0.025	0.166	0.033	0.013	0.394
2	$M_1$		0.015	0.009	0.600	0.394	0.038	0.096	0.413	0.039	0.094	0.157	0.026	0.166	0.021	0.017	0.810
	$M_3$		0.014	0.009	0.643	0.374	0.024	0.064	0.428	0.027	0.063	0.162	0.028	0.173	0.022	0.012	0.545
	$M_1^{exnex}$	0.942	0.021	0.014	0.667	0.385	0.047	0.122	0.404	0.048	0.119	0.179	0.038	0.212	0.012	0.010	0.833
	$M_3^{exnex}$	0.914	0.014	0.008	0.571	0.377	0.027	0.072	0.425	0.031	0.073	0.162	0.026	0.160	0.023	0.010	0.435
3	$M_1$		0.026	0.013	0.500	0.305	0.049	0.161	0.477	0.045	0.094	0.147	0.027	0.184	0.045	0.022	0.489
	$M_3$		0.024	0.011	0.458	0.308	0.034	0.110	0.479	0.037	0.077	0.148	0.027	0.182	0.042	0.015	0.357
	$M_1^{exnex}$	0.920	0.041	0.020	0.488	0.233	0.042	0.180	0.516	0.051	0.099	0.162	0.036	0.222	0.048	0.021	0.438
	$M_3^{exnex}$	0.794	0.024	0.015	0.625	0.311	0.056	0.180	0.474	0.048	0.101	0.152	0.025	0.164	0.040	0.018	0.450
4	$M_1$		0.008	0.004	0.500	0.359	0.022	0.061	0.421	0.021	0.050	0.179	0.018	0.101	0.033	0.010	0.303
	$M_3$		0.007	0.004	0.571	0.357	0.020	0.056	0.422	0.021	0.050	0.180	0.018	0.100	0.034	0.008	0.235
	$M_1^{exnex}$	0.941	0.008	0.004	0.500	0.350	0.023	0.066	0.419	0.023	0.055	0.190	0.018	0.095	0.033	0.008	0.242
	$M_3^{exnex}$	0.913	0.008	0.004	0.500	0.361	0.020	0.055	0.424	0.020	0.047	0.175	0.018	0.103	0.032	0.008	0.250
5	$M_1$		0.015	0.010	0.667	0.376	0.041	0.109	0.398	0.041	0.103	0.174	0.031	0.178	0.036	0.022	0.611
	$M_3$		0.015	0.009	0.600	0.359	0.025	0.070	0.414	0.030	0.072	0.178	0.030	0.169	0.034	0.014	0.412
	$M_1^{exnex}$	0.939	0.024	0.016	0.667	0.357	0.048	0.134	0.377	0.048	0.127	0.208	0.041	0.197	0.033	0.019	0.576
	$M_3^{exnex}$	0.913	0.014	0.008	0.571	0.365	0.027	0.074	0.417	0.031	0.074	0.173	0.029	0.168	0.031	0.012	0.387
6	$M_1$		0.010	0.009	0.900	0.387	0.044	0.114	0.401	0.043	0.107	0.143	0.031	0.217	0.059	0.031	0.525
	$M_3$		0.010	0.008	0.800	0.363	0.028	0.077	0.424	0.031	0.073	0.149	0.030	0.201	0.055	0.018	0.327
	$M_1^{exnex}$	0.942	0.016	0.014	0.875	0.364	0.054	0.148	0.382	0.055	0.144	0.161	0.041	0.255	0.077	0.029	0.377
	$M_3^{exnex}$	0.902	0.011	0.007	0.636	0.371	0.027	0.073	0.417	0.035	0.084	0.153	0.028	0.183	0.049	0.024	0.490
7	$M_1$		0.011	0.009	0.818	0.381	0.042	0.110	0.400	0.042	0.105	0.176	0.033	0.188	0.033	0.023	0.697
	$M_3$		0.010	0.008	0.800	0.361	0.027	0.075	0.420	0.031	0.074	0.178	0.033	0.185	0.031	0.014	0.452
	$M_1^{exnex}$	0.942	0.018	0.015	0.833	0.360	0.054	0.150	0.373	0.057	0.153	0.221	0.050	0.226	0.027	0.020	0.741
	$M_3^{exnex}$	0.914	0.011	0.007	0.636	0.366	0.029	0.079	0.418	0.032	0.077	0.175	0.030	0.171	0.030	0.013	0.433
8	$M_1$		0.019	0.013	0.684	0.390	0.045	0.115	0.435	0.045	0.103	0.127	0.033	0.260	0.028	0.023	0.821
	$M_3$		0.017	0.011	0.647	0.375	0.029	0.077	0.446	0.034	0.076	0.136	0.030	0.221	0.025	0.013	0.520
	$M_1^{exnex}$	0.944	0.034	0.022	0.647	0.377	0.059	0.156	0.440	0.060	0.136	0.133	0.040	0.301	0.016	0.016	1.000
	$M_3^{exnex}$	0.914	0.016	0.012	0.750	0.377	0.029	0.077	0.439	0.035	0.080	0.142	0.029	0.204	0.026	0.012	0.462
9	$M_1$		0.017	0.011	0.647	0.394	0.040	0.102	0.394	0.042	0.107	0.171	0.031	0.181	0.024	0.019	0.792
	$M_3$		0.016	0.010	0.625	0.368	0.024	0.065	0.417	0.029	0.070	0.175	0.032	0.183	0.023	0.012	0.522
	$M_1^{exnex}$	0.945	0.026	0.017	0.654	0.385	0.053	0.138	0.373	0.053	0.142	0.203	0.044	0.217	0.012	0.012	1.000
	$M_3^{exnex}$	0.914	0.015	0.009	0.600	0.373	0.026	0.070	0.416	0.034	0.082	0.172	0.029	0.169	0.025	0.011	0.440

Note: Mixture Probabilities  $p_i$ , Posterior Mean (PM), Posterior Standard Deviation (PSD), Coefficient of Variation (CV)







## **Chapter 5**

# **Concluding Remarks and Future Works**

In this dissertation, we study several Bayesian hierarchical multinomial-Dirichlet models with order restrictions for small area estimation. We provided very-detailed Bayesian analysis for each model. We have demonstrated our application to body mass index data from NHANES III.

### **5.1 Concluding Remarks**

First, in Chapter 2, we developed the Bayesian hierarchical multinomial-Dirichlet model with order restrictions to analyze multinomial data. The unimodal order restrictions are necessary when the parameters of interest may have the unimodal structure. We use gridy Gibbs sampler for each model. We show how the model with order restrictions can borrow information differently. A simulation study is presented to compare the different order restriction assumptions, which impact the

strength of borrowing information. The model with order restrictions on parameters  $\mu$  and  $\theta$  can have small standard variance when the order restriction holds and the heterogeneity of small areas is small. In practice, the same unimodal order restriction for all small areas may not hold. Hence, the performance of model with order restrictions is not good as expected.

Second, in Chapter 3, we proposed a multinomial-Dirichlet model, where we incorporated the order restriction uncertainty to allow different unimodal order restriction in each small area. We discussed an approximation method to compute the mixture probabilities of different order restrictions and avoid the computation difficulty. We proposed a Bayesian diagnostics method, LPML, to compare different models. In the application to the BMI data, the model with order restriction that the mode is at the third position, has the largest LPML, which suggests that the model is preferable when fitting the NHANES BMI data. The model, which incorporates the uncertainty, is slightly less-preferable but robust under different scenarios. It is recommended to use the model, which incorporates the uncertainty, when the modal position of unimodal order restriction is not known.

Third, in Chapter 4, we proposed an exchangeability-nonexchangeability model to allow partial exchangeability on parameters. We discussed a Gibbs sampler method for the EXNEX model. The model can borrow information from the similar areas while avoiding borrowing from the very different areas. For example, in the numerical example, after fitting the multinomial-Dirichlet EXNEX model with order restrictions,  $M_3^{exnex}$ , the posterior mean of mixture probability  $p_{21}$  is 0.578, which means that it borrows much less information in County 21 than other counties. This is an alternative approach when the same unimodal order restriction may not hold for all areas, which is a preferred approach when the heterogeneity across

small areas is large.

In summary, the multinomial-Dirichlet model with order restrictions may be appropriate for many applications. As in the application to BMI data, the model with order restrictions can have smaller posterior standard deviation and higher estimation accuracy than the model without order restrictions. The model with uncertain order restrictions in Chapter 3 and the EXNEX model with order restrictions in Chapter 4 are great extensions of our current approach. This dissertation provide novel contributions in making inference with order restrictions for small area estimation. But we notice the computational difficulty may exist in the multinomial-Dirichlet model with order restrictions. Even assuming partial exchangeability is a great alternative approach for the hierarchical models for small area estimations, the carefully considerations of prior specification should be addressed (Neuenschwander 2016).

## **5.2 Future works**

In the dissertation, we use a Bayesian hierarchical model with order restrictions for small area estimation of categorical data in Chapter 2. In Chapters 3 and 4, we develop two types of models to handle the situation where the same unimodal order restrictions may not hold for all areas. Exploring different order restriction assumptions can extend the usage of models with order restrictions. We focus on parametric statistical models in this dissertation. We can have nonparameteric Bayesian analysis, see Quintana (1998), to make the procedure more robust. Yin and Nandram (2020a) propose a two-stage non-parametric Bayesian model with several independent Dirichlet processes at the first stage that represents the data, to

take care of the gaps, outliers and ties in the body mass index data, see also Yin and Nandram (2019), Yin and Nandram (2020b). All these models are Dirichlet process mixture (DPM) models.

The Dirichlet process mixture model in Nandram and Yin (2016b) is motivated by Nandram and Choi (2004). Nandram and Yin (2016a) showed how to add a Dirichlet process prior to do Bayesian predictive inference. Their work is motivated by Binder (1982). We can use the idea of DPM for the model with order restrictions to relax the parametric assumption. Specifically,

$$\begin{aligned} \mathbf{n}_i | \boldsymbol{\theta}_i &\sim \text{Multinomial}(\boldsymbol{\theta}_i), \\ \boldsymbol{\theta}_i | G &\sim G, i = 1, \dots, \ell, \\ G | \alpha, \boldsymbol{\mu}, \tau &\sim \text{DP}\{\alpha, \text{Dirichlet}(\boldsymbol{\mu}\tau)\}, \boldsymbol{\theta}_i \in C, \\ \pi(\boldsymbol{\mu}, \tau, \alpha) &\propto \frac{I_{C_\mu}}{(1 + \tau)^2} \frac{1}{(1 + \alpha)^2}, \boldsymbol{\mu} \in C_\mu, \end{aligned}$$

where  $\alpha$  is a concentration parameter,  $C = \{\boldsymbol{\theta}_i : \theta_{i1} \leq \dots \leq \theta_{im} \geq \dots \geq \theta_{iK}, i = 1, \dots, \ell\}$ ,  $C_\mu = \{\boldsymbol{\mu} : \mu_1 \leq \dots \leq \mu_m \geq \dots \geq \mu_K\}$ , and assume the modal positions  $m$  in  $C$  and  $C_\mu$  are known.

The final BMI data set for this study uses only the 35 largest counties with a population of at least 500,000 for selected age categories by sex (male, female) and race (white non-Hispanic, black non-Hispanic, Hispanic, other). We can easily apply our method to the small domains formed by on age, race, and sex, such as the young Hispanic-male BMI data. But the cells of the multinomial tables will become sparse. We can eliminate some counties that become small or we can combine some counties. However, due to the structures of multinomial-Dirichlet models with order restrictions, we cannot add race, age and sex as covariates into the model.

Since the BMI data are from the survey sampling and individuals are selected with different probabilities, we should not ignore the survey weights. It is possible to incorporate the survey weights into our model as well. Let  $W_{ig}$  denote the survey weights, adding up to the population size within each county,  $i = 1, \dots, \ell$ , sample index  $g = 1, \dots, n_i$  and cell index  $j = 1, \dots, K$ . Yang (2021) provided adjusted weights are  $\omega_{ig} = \hat{n}_i \frac{W_{ig}}{\sum_{g=1}^{n_i} W_{ig}}$ , where  $\hat{n}_i = \frac{(\sum_{g=1}^{n_i} W_{ig})^2}{\sum_{g=1}^{n_i} W_{ig}^2}$  and  $\sum_{g=1}^{n_i} \omega_{ig} = \sum_{g=1}^{n_i} \omega_{ig}^2 = \hat{n}_i$ . Yang (2021) used weighted likelihood distributions for a single multinomial model, see also Nandram, Choi, and Liu (2021). Yang (2021) found out there is a very small difference between normalized and unnormalized weighed likelihood.

We can transform BMI data using the adjusted weights into adjusted counts. Let  $I_{igj}$  be the BMI category indicator for individual  $g$  in county  $i$ ,  $i = 1, \dots, \ell$  at cell  $j$ ,  $j = 1, \dots, K$ , where define  $I_{igj} = 0$  or 1 with  $\sum_{j=1}^K I_{igj} = 1$ , for example, if a person responds in cell  $j$ , a one is scored and all other cells have zeros. For simplification, we can have the unnormalized weighted joint posterior distribution as

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\mu}, \tau, \mathbf{p}, \phi | \mathbf{n}) &\propto \prod_{i=1}^{\ell} \left\{ \frac{(\sum_{j=1}^K \sum_{g=1}^{n_i} I_{igj} \omega_{ig})!}{\prod_{j=1}^K (\sum_{g=1}^{n_i} I_{igj} \omega_{ig})!} \prod_{j=1}^K \theta_{ij}^{\sum_{g=1}^{n_i} I_{igj} \omega_{ig}} \right. \\ &\quad \left[ p_i \frac{\text{Dirichlet}(\boldsymbol{\mu}\tau)}{\int_{\theta_i \in C} \text{Dirichlet}(\boldsymbol{\mu}\tau) d\theta_i} \frac{(K-1)!}{(1+\tau)^2} + \right. \\ &\quad \left. \left. (1-p_i) \text{Dirichlet}(1, \dots, 1) \right] \frac{p_i^{\phi\tau_0-1} (1-p_i)^{(1-\phi)\tau_0-1}}{B(\phi\tau_0, (1-\phi)\tau_0)} \right\}. \end{aligned}$$

Our approaches can be applied to the adjusted counts directly.

The computational burden to fit each of the models is enormous. In Chapter 3, we would like to know how to provide more efficient algorithms to make posterior inference and compute the Bayesian diagnostics criterion. Operationalizing the

Gibbs sampler will make our work more accessible to a much larger audience (e.g. government agencies).

In Chapter 4, we would like to explore how to incorporate the partial exchangeability assumption into the model with order restrictions differently to have different levels of borrowing information from all areas. We also would like to see the performance of the multinomial-Dirichlet EXNEX models in the simulation under different scenarios.

In the posterior summary of mixture probabilities  $p_i$ , we can use posterior coefficients of variations (PCV), but these are misleading for the mixture probabilities  $p_i$ . Because  $p_i$  and  $1 - p_i$  have different PCVs even they are essentially the same variables. PCVs do not make sense when the variables are correlated, such as  $\theta_{ij}, j = 1, \dots, K$  or  $p_i$  and  $1 - p_i$ . In the Bayesian paradigm, one needs to present the entire posterior density graphically. But when they are too many of these, we need summaries. Perhaps we can use robust measures of standard deviations and center, e.g. median and median absolute deviation (MAD) and we can take  $PCV = MAD/median$ .

# Bibliography

- Binder, D. A. (1982). Non-parametric Bayesian models for samples from finite populations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(3):388–393.
- Chen, M.-H. and Shao, Q.-M. (1997). On Monte Carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics*, 25(4):1563–1594.
- Chen, X. and Nandram, B. (2019). Order-restricted Bayesian estimation of multinomial cell counts for small areas. In *JSM 2019*, Section on Bayesian Statistical Science, pages 1024–1030. Alexandria, VA: American Statistical Association.
- Chen, X. and Nandram, B. (2020). Bayesian model diagnostics with order restrictions on cell probabilities. In *JSM 2020*, Section on Statistical Computing, pages 1754–1766. Alexandria, VA: American Statistical Association.
- Chen, X. and Nandram, B. (2021a). Bayesian inference for multinomial data from small areas incorporating uncertainty about order restriction. *Survey Methodology*. (revising).
- Chen, X. and Nandram, B. (2021b). Bayesian order-restricted inference of multinomial counts from small areas. *Recent Advances in Applied Statistics*. David

Hangal, Raosahed Laptate, Hukum Chandra and Girish Chandra, Springer Nature(in press).

Chen, X. and Nandram, B. (2021c). A Bayesian small area model with order restrictions for contingency table. *Communications in Statistics - Theory and Methods*. (in press).

Dunson, D. B. and Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, 59(2):286–295.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992a). Model determination using predictive distributions with implementation via sampling-based methods. Technical report, Stanford University, Department of Statistics.

Gelfand, A. E., Smith, A. F. M., and Lee, T.-M. (1992b). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics*, pages 169–193. University Press.

Ghosh, M. and Kim, D. H. (2002). Robust Bayesian analysis with partially exchangeable priors. *Journal of Statistical Planning and Inference*, 102(1):99–107.

Heck, D. W. and Davis-Stober, C. P. (2019). Multinomial models with linear inequality constraints: Overview and improvements of computational methods for Bayesian inference. *Journal of Mathematical Psychology*, 91:70–87.



- Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473.
- Kaizer, A. M., Koopmeiners, J. S., and Hobbs, B. P. (2017). Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics*, 19(2):169–184.
- Li, Z. (2008). *Some Problems in Statistical Inference Under Order Restrictions*. PhD thesis, University of Michigan.
- Mair, P., Hornik, K., and de Leeuw, J. (2009). Isotone optimization in r: pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of statistical software*, 32(5):1–24.
- Malec, D. and Sedransk, J. (1992). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain. *Biometrika*, 79(3):593–601.
- Malinovsky, Y. and Rinott, Y. (2010). Prediction of ordered random effects in a simple small area model. *Statistica Sinica*, 20(2):697–714.
- Nadarajah, S. and Kotz, S. (2006). R programs for truncated distributions. *Journal of Statistical Software*, 16(1):1–8.
- Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, 61(1-2):97–126.
- Nandram, B. (2005). A Bayesian subset analysis of sensory evaluation data. *Journal of Modern Applied Statistical Methods*, 4(2):13.

- Nandram, B. and Choi, J. W. (2004). Nonparametric Bayesian analysis of a proportion for a small area under nonignorable nonresponse. *Journal of Nonparametric Statistics*, 16(6):821–839.
- Nandram, B. 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(489):120–135.
- Nandram, B., Choi, J. W., and Liu, Y. (2021). Integration of nonprobability and probability samples via survey weights. *International journal of statistics and probability*, 10(6):5–21.
- Nandram, B., Kim, D., and Zhou, J. (2019). A pooled Bayes test of independence for sparse contingency tables from small areas. *Journal of Statistical Computation and Simulation*, 89(5):899–926.
- Nandram, B. and Peiris, T. B. (2018). Bayesian analysis of a roc curve for categorical data using a skew-binormal model. *Statistics and Its Interface*, 11(2):369–384.
- Nandram, B. and Sedransk, J. (1995). Bayesian inference for the mean of a stratified population when there are order restrictions. In Gatsonis, C., Hodges, J. S., Kass, R. E., and Singpurwalla, N. D., editors, *Case Studies in Bayesian Statistics, Volume II*, pages 309–322, New York, NY. Springer New York.
- Nandram, B., Sedransk, J., and Smith, S. J. (1997). Order-restricted Bayesian estimation of the age composition of a population of Atlantic cod. *Journal of the American Statistical Association*, 92(437):33–40.

- Nandram, B. and Yin, J. (2016a). Bayesian predictive inference under a Dirichlet process with sensitivity to the normal baseline. *Statistical Methodology*, 28:1–17.
- Nandram, B. and Yin, J. (2016b). A nonparametric Bayesian prediction interval for a finite population mean. *Journal of Statistical Computation and Simulation*, 86(16):3141–3157.
- Nandram, B. and Yin, J. (2019). Hierarchical Bayesian models for small areas with Dirichlet processes. In *JSM 2019, Section on Survey Research Methods*, pages 2594–2613. Alexandria, VA: American Statistical Association.
- Neuenschwander, B., Wandel, S., Roychoudhury, S., and Bailey, S. (2016). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics*, 15(2):123–134.
- Quintana, F. A. (1998). Nonparametric Bayesian analysis for assessing homogeneity in  $k \times i$  contingency tables with fixed right margin totals. *Journal of the American Statistical Association*, 93(443):1140–1149.
- Rao, J. and Molina, I. (2015). *Small Area Estimation*. Wiley Series in Survey Methodology. Wiley.
- Robertson, T. T., Wright, F., and Dykstra, R. (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics. Wiley, Chichester.
- Roy, V. (2020). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7:387–412.

- Sedransk, J., Monahan, J., and Chiu, H. Y. (1985). Bayesian estimation of finite population parameters in categorical data models incorporating order restrictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):519–527.
- Silvapulle, M. and Sen, P. (2004). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley Series in Probability and Statistics. Wiley.
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809.
- Wu, J., Meyer, M. C., and Opsomer, J. D. (2016). Survey estimation of domain means that respect natural orderings. *Canadian Journal of Statistics*, 44(4):431–444.
- Yang, L. (2021). Bayesian predictive inference with survey weights. Master’s thesis, Department of Mathematical Sciences, Worcester Polytechnic Institute.
- Yin, J. and Nandram, B. (2020a). A Bayesian small area model with Dirichlet processes on the responses. *Statistics in Transition*, 21(3):1–19.
- Yin, J. and Nandram, B. (2020b). A nonparametric Bayesian analysis of response data with gaps, outliers and ties. *Statistics and Applications*, 18(2):121–141.