Bayesian Model Diagnostics with Order Restrictions on Cell Probabilities

Xinyu Chen* Balgobin Nandram†

Abstract

Estimating cell probabilities for small areas can be difficult, due to a lack of available data from national surveys. One of statistical techniques for small area estimation is using multinomial Dirichlet models to borrow information among small areas. We study Bayesian diagnostics for multinomial counts from small areas. Within each area, the cell probabilities are ordered (e.g. unimodal ordering). Specifically we consider Bayesian diagnostics for a multinomial Dirichlet model with order restriction which shares a common effect among areas. The log pseudo marginal likelihood (LPML) is a well-known Bayesian criterion for comparing models. Since the order restriction significantly increases the difficulty, we develop an algorithm to compute LPML. We use a special-designed importance function to increase the efficiency of Monte Carlo integration, thereby gaining a higher precision for estimations of LPML. The proposed methodology is applied to a case study of body mass index (BMI).

Keywords: Bayesian computation, LPML, Multinomial, Monte Carlo method, Small areas, Unimodal order restrictions.

1. Introduction

In many surveys, questionnaires have items that are categorized into several cells. These items may be filled in by people from different areas or groups, which may be small. Estimates of cell probabilities for individual areas may not be reliable and a statistician might need to pool data from different small areas (Rao and Molina 2015). Furthermore, there may be important information over the cells from each area and this information can be incorporated into a model to provide additional improvement. So our problem is to obtain a methodology to pool information across areas and to incorporate information across the cells in each area. The Bayesian paradigm is attractive for this problem, and we start with the hierarchical Bayesian multinomial Dirichlet model, and then we incorporate the order restrictions over the cell probabilities into this model.

There are extensive researches to consider different techniques for Small Area Estimation(SAE) with different order restrictions. 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. They assumed constrained estimators satisfying the monotonicity. Malinovsky and Rinott (2010) presented predictors with an appropriate amount of shrinkage for the particular problem of ordered parameters in the context of Small Area Estimation. Their performance is close to that of the optimal predictors. 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. They also described a general Gibbs sampler for drawing posterior samples. A suitable order restriction assumption can increase model precision. Li (2008) made a great overview about statistical inference under order restrictions. He also showed the inference of ordered binomial probabilities in frequentist statistics. From Wu, Meyer

*Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609
†Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609

and Opsomer's research about order restriction to Li's review, they proved that the order constraints should be considered in order to improve efficiency and minimize bias, which can be done in different aspects.

In the small area context, most of these papers cover order restrictions across areas (e.g., Wu, Meyer and Opsomer, 2016). However, we are not interested in order restriction across areas, but rather we are interested in order restriction across the cell probabilities within each area. For this type of order restrictions problem, Nandram (1997) provided a good discussion about a hierarchical Bayesian approach for taste-testing experiment and appropriate methods for the model. To select the best population, he studied three criteria based on the distribution of random variables representing values on a hedonic scale using the simple tree order (See also Nandram 1998). Nandram, Sedransk, and Smith (1997) improved estimation of the age composition of a population of fish with the help of order restrictions. They proposed different order restrictions for different fish length strata. With the help of the Gibbs sampler, they showed that order restrictions provided large gains in precision for estimating the proportion of fish in each age class. Their research was motivated by Gelfand, Smith and Lee (1992) and earlier Sedransk, Monahan, and Chiu (1985).

But our interest is not only order restrictions across the cell probabilities within each area, but also similar unimodal structure within each area. Chen and Nandram (2019), which appeared the Proceedings of the American Statistical Association, proposed a multinomial Dirichlet model with order restrictions. They considered similar unimodal structure within each area. They showed how to use Gibbs sampler for posterior distribution. A great improvement for estimating the cell probabilities has been shown in their model application.

The work in this paper is a large step forward from their work. We not just presented more details for model comparison with better visualization. We also show the Bayesian diagnostics. In the Bayesian framework, the log pseudo marginal likelihood (LPML) is a well-known Bayesian criteria for comparing models. Since the order restriction significantly increases the difficulty, we develop an algorithm to compute LPML for the Bayesian diagnostic without increasing computation time. We discuss an illustrative example on body mass index (BMI) data. Since people have similarity that the majority in each county will be in the same level of BMI, it is reasonable to assume that the cell probabilities share a common effect and have the same order restrictions in each county. Actually it seems that most people will have a third level BMI, which is overweight, among those counties. So it is reasonable to believe that the cell probabilities are unimodal in each county and the third level is the mode. With this information, our estimates for each county can be improved using a multinomial Dirichlet model with order restrictions such as $\theta_{i1} \leqslant \theta_{i2} \leqslant \theta_{i3} \geqslant \theta_{i4} \geqslant \theta_{i5}$ for the i^{th} area. One feature of our approach is that Dirichlet distribution with parameters $\boldsymbol{\mu}$ and τ embodies the common effect and the same order restriction. At the second stage of model, parameter $\boldsymbol{\mu}$ has a similar order restriction as cell probabilities $\boldsymbol{\theta}_i$. It has more flexibility without increasing computation difficulty. Then we compute LPML of our proposed model and the model without order restriction.

Further, Chen and Nandram(2020) will have an overview for this type order restriction problem in SAE. Their overview will cover model selection, sampling from posterior distribution, model diagnostics, model illustration example, and problem discussions.

The article is organized as follows. In Section 2, we present the hierarchical Bayesian multinomial Dirichlet model with order restrictions. In Section 3, we use the same techniques from Chen and Nandram (2019). In Section 4, we present a Bayesian diagnostic for the model with order restrictions and we discuss difficulties associated with a standard Bayesian diagnostic measure that may not be appropriate. In Section 5, we show how to

analyze the BMI data in our application. We demonstrate how much improvement there is under the order restrictions. Section 6 has a summary of our work.

2. Multinomial Dirichlet Models

In this section, we describe the Bayesian methodology for the cell counts over the small areas. Nandram, Kim and Zhou (2019) has a useful discussion of hierarchical Bayesian multinomial Dirichlet model without order restriction (M1) and the methodology needed to fit it.

We incorporate the order restriction into the hierarchical Bayesian Dirichlet multinomial model (M2). We use a grid method in Gibbs sampler. This is more efficient than the method by Nandram (1998). Letting n_{ij} be the cell counts, θ_{ij} the corresponding cell probabilities,$i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, K$,$\boldsymbol{n}_{i.} = \sum_{j=1}^{K} n_{ij}$ and we believe the mode of $\boldsymbol{\theta}_i s$ is $\theta_{im}, 1 \leqslant m \leqslant K$.

Specifically, we take

$$\boldsymbol{n}_i|\boldsymbol{\theta}_i \overset{ind}{\sim} \text{Multinomial}(\boldsymbol{n}_i. , \boldsymbol{\theta}_i), \qquad \boldsymbol{\theta}_i \in C \quad i = 1, \ldots, I,$$
$$pdf : f(\boldsymbol{n}_i|\boldsymbol{\theta}_i) = \frac{\Gamma(n_{i.}+1)}{\prod_{j=1}^{K}\Gamma(n_{ij}+1)} \prod_{j=1}^{K} \theta_{ij}^{n_{ij}}, \qquad \sum_{j=1}^{K} n_{ij} = n_{i.}, \qquad n_i \geqslant 0.$$

where $C = \{\boldsymbol{\theta}_i : \theta_{i1} \leqslant \ldots \leqslant \theta_{im} \geqslant \ldots \geqslant \theta_{iK}, i = 1, \ldots, I\}$, and assume C is known. As mentioned above, in our BMI study, $C = \{\boldsymbol{\theta}_i : \theta_{i1} \leqslant \theta_{i2} \leqslant \theta_{i3} \geqslant \theta_{i4} \geqslant \theta_{i5}, i = 1, 2, \ldots, 35\}$.

At the second stage, we take

$$\boldsymbol{\theta}_i|\boldsymbol{\mu}, \tau \overset{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), i = 1, \ldots, I,$$
$$pdf : f(\boldsymbol{\theta}_i|\boldsymbol{\mu}, \tau) = \frac{\Gamma(\sum_{j=1}^{K}\mu_j\tau)}{\prod_{j=1}^{K}\Gamma(\mu_j\tau)} \prod_{j=1}^{K} \theta_{ij}^{\mu_j\tau-1}, \sum_{j=1}^{K} \theta_{ij} = 1, \theta_{ij} \geq 0, j = 1, \ldots, K.$$
$$\pi(\boldsymbol{\mu}, \tau) = \frac{K(m-1)!(K-m)!}{(1+\tau)^2}, \qquad \mu_j > 0, \quad \sum_{j=1}^{K} \mu_j = 1, \quad \boldsymbol{\mu} \in C_{\boldsymbol{\mu}}.$$

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

$$C_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} : \mu_1 \leqslant \ldots \leqslant \mu_m \geqslant \ldots \geqslant \mu_K\}.$$

Using Bayes' theorem, the joint posterior distribution of all variables is

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

where I_C and $I_{C_{\boldsymbol{\mu}}}$ are the indicator functions under those order restrictions, and

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

$$D(\boldsymbol{\mu}\tau) = \frac{\prod_{j=1}^{K}\Gamma(\mu_j\tau)}{\Gamma[\sum_{j=1}^{K}\mu_j\tau]}.$$

3. Computations

It is straight forward to generate samples from M1; see Nandram (1998). In fact, using the griddy Gibbs sampler, it can be done easier than the method in Nandram (1998). Chen and Nandram (2019) presented a new method for the order restrictions of $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ into two parts for model M2. They used Gibbs sampling, a Markov chain Monte Carlo (MCMC) algorithm, for $\boldsymbol{\mu}$ with an order restriction and τ. Instead of sampling directly from the posterior of $\boldsymbol{\theta}$, they sampled from a set of truncated Gamma distributions.

4. Bayesian Diagnostics

In the Bayesian framework, the logarithm of the pseudo-marginal likelihood (LPML) is a well-known Bayesian criterion for comparing models. A ratio of LPML's is a surrogate for the Bayes factor. The best model among competing models have the largest LPML,

$$LPML = \sum_{i=1}^{I}\log(CPO_i),$$

where $CPO_i = P(n_i \mid n_i \text{ is deleted})$ for the i^{th} county. Essentially the i^{th} county is deleted and then its cell counts are predicted from the remaining counties.

Conditional predictive ordinate (CPO) can be estimated using the harmonic mean of the likelihood of the vectors of the n_{ij}. (M is the number of converged posterior samples from Gibbs sampling in Section 3.)

$$\begin{aligned}\hat{CPO}_i &= \left[\frac{1}{M}\sum_{h=1}^{M}\frac{1}{f(n_i|\boldsymbol{\mu}^{(h)},\tau^{(h)})}\right]^{-1}\\ &= \left[\frac{1}{M}\sum_{h=1}^{M}\frac{\prod_{j=1}^{K}n_{ij}!}{n_{i.}!}\frac{\int_{\boldsymbol{\theta}_i\in C}\prod_{j=1}^{K}\theta_{ij}^{\mu_j^{(h)}\tau^{(h)}-1}d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i\in C}\prod_{j=1}^{K}\theta_{ij}^{\mu_j^{(h)}\tau^{(h)}+n_{ij}-1}d\boldsymbol{\theta}_i}\right]^{-1}.\end{aligned}$$

As mentioned by Sedransk, Monahan, and Chiu (1985), a different importance sampling could be used to estimate the ratio,

$$\frac{\int_{\boldsymbol{\theta}_i\in C}\prod_{j=1}^{K}\theta_{ij}^{\mu_j^{(h)}\tau^{(h)}-1}d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i\in C}\prod_{j=1}^{K}\theta_{ij}^{\mu_j^{(h)}\tau^{(h)}+n_{ij}-1}d\boldsymbol{\theta}_i}.$$

It is more precise but slower than the importance function used in Gibbs sampling. More details can be found in Appendix. CPO_i can be estimated as

$$\begin{aligned}\hat{CPO}_i &= \left[\frac{1}{M}\sum_{h=1}^{M}\frac{\prod_{j=1}^{K}n_{ij}!}{n_{i.}!}\frac{\int_{\boldsymbol{\theta}_i\in C}\prod_{j=1}^{K}\theta_{ij}^{\mu_j^{(h)}\tau^{(h)}-1}d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i\in C}\prod_{j=1}^{K}\theta_{ij}^{\mu_j^{(h)}\tau^{(h)}+n_{ij}-1}d\boldsymbol{\theta}_i}\right]^{-1}\\ &= \left[\frac{1}{M}\sum_{h=1}^{M}\frac{\prod_{j=1}^{K}n_{ij}!}{n_{i.}!}\frac{\Gamma(n_{i.}+\tau^{(h)})}{\Gamma(\tau^{(h)})}\frac{\sum_{q=1}^{M'}(\prod_{j=1}^{K}x_{ij}^{(q)\mu_j^{(h)}\tau^{(h)}-\alpha^*}e^{-(1-r^*)\sum_{j=1}^{K}x_{ij}^{(q)}})}{\sum_{q=1}^{M'}(\prod_{j=1}^{K}x_{ij}^{(q)n_{ij}+\mu_j^{(h)}\tau^{(h)}-\alpha^*}e^{-(1-r^*)\sum_{j=1}^{K}x_{ij}^{(q)}})}\right]^{-1},\end{aligned}$$

M' is the Monte Carlo sample size for integrating over x_{ij}, and

$$LPML \approx \sum_{i=1}^{I} \log(C\hat{P}O_i).$$

In our illustrative example, $M = M' \approx 1000$, but this is not necessary.

5. Application to BMI

5.1 Body Mass Index

In our application, we use a selected subset of the female 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. Our goal is to estimate the proportions of the BMI levels. Table 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 to form the unimodal order restriction. Indeed, there are violations in some counties in the earliest and latest cells.

For large population counties, we consider that people randomly fall into five BMI categorical levels, which are underweight, normal, overweight, obese1, and obese2. Thus, 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 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 restrictions holds, a way to improve posterior inference. 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.

Table 1: US Female BMI data

State ID	County ID	BMI_lvl1	BMI_lvl2	BMI_lvl3	BMI_lvl4	BMI_lvl5
4	13	3	40	37	13	4
6	1	1	36	38	15	1
6	19	3	20	49	13	5
6	37	2	145	174	77	14
6	59	1	29	31	16	3
...	...	...	...	...	...	...

5.2 MCMC Convergence

We run 20,000 MCMC iterations, take 10,000 as a 'burn in' and use every 10th to obtain 1,000 converged posterior samples. Table 2 gives the effective sample size of the parameters $\boldsymbol{\mu}$, τ for the model with the order restriction and the general model. The effective sample sizes are almost 1,000. Table 3 gives the p-values of the Geweke test for the parameters (Cowles and Carlin 1996). The p-values are all large so we can not reject that null hypothesis that the MCMC is stationary. Then posterior samples can be used for the further inference.

Figure 1: Total Counts for 35 Counties

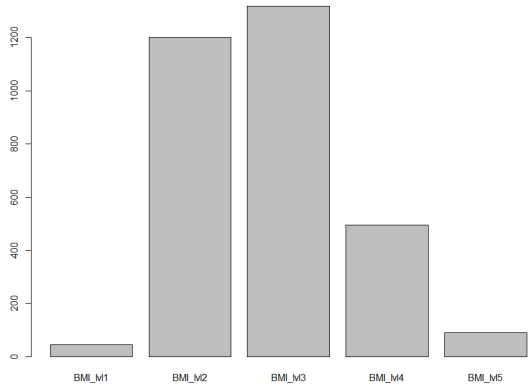


Table 2: Effective Sizes

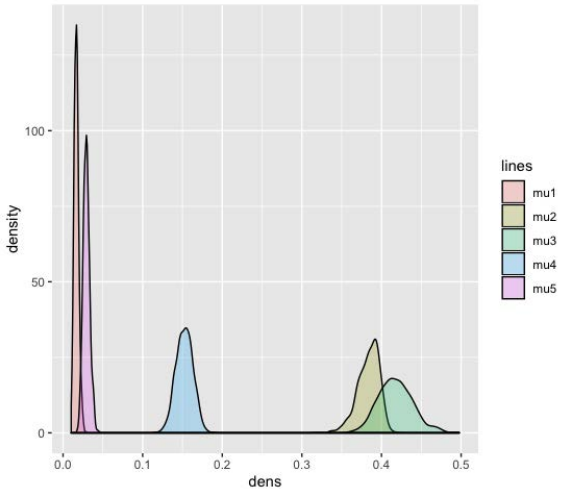
Models	μ_1	μ_2	μ_3	μ_4	μ_5	τ
W. Order	974	1000	1000	1000	1000	1000
W/O Order	859	1000	1000	971	1000	1032

Table 3: Geweke Diagnostics

Models	μ_1	μ_2	μ_3	μ_4	μ_5	τ
W. Order	0.4275	0.3221	0.2376	0.0895	0.3784	0.1393
W/O Order	0.8352	0.785	0.6931	0.4425	0.3692	0.8983

In Figure 2, posterior densities of μ show a nice pattern and μ_3 is centered at the largest value. It means that our samples from μ posterior densities have an order restriction. It matches our model assumptions. But we notice that there is an overlap between μ_2 and μ_3. It is apparent that $\mu_2 \leqslant \mu_3$ may not be appropriate for BMI counts. The order restriction assumption may be too strong in this case.

Figure 2: Posterior Density of μ

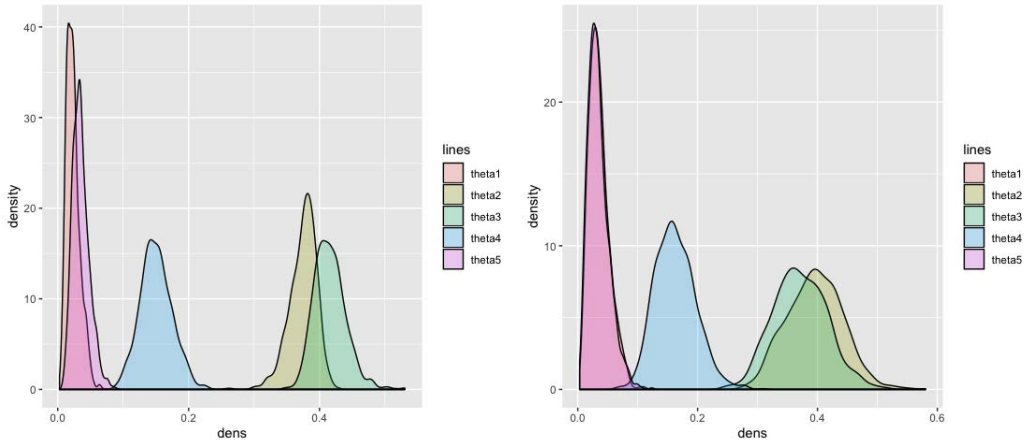


5.3 Model Comparison

We compute the estimated cell probabilities for each county and their variances, which are the posterior sample means and posterior standard deviations of parameter $\boldsymbol{\theta}$. In Table 6(Appendix), we show their posterior means (PM), posterior standard deviations (PSD), and coefficients of variation (CV). We notice that the posterior means from the model with order restrictions (M2) have lower variances compared with the general model (M1). Generally we have higher accuracy for the estimation of the cell probabilities $\boldsymbol{\theta}$. But for parameters θ_1 and θ_5 in some counties, such as the second county in Table 6 (PSD: 0.0106 vs 0.0089), the model with order restriction does not gain precision on them. This is expected because the extreme cells are generally sparse. In general, many of the coefficients of variation are small enough to declare that the posterior means are reliable. In Table 6, we also present coefficient of variation, also known as relative standard deviation, for the model comparison. It is defined as the ratio of the posterior standard deviation σ to the posterior mean μ, $CV = \frac{\sigma}{\mu}$. In Table 6, the model with order restrictions (M2) has lower CV than the model without order restriction. Specially for θ_2, θ_3, and θ_4, CV reduced to almost half of the previous.From this aspect, our model M2 is more suitable for BMI data.

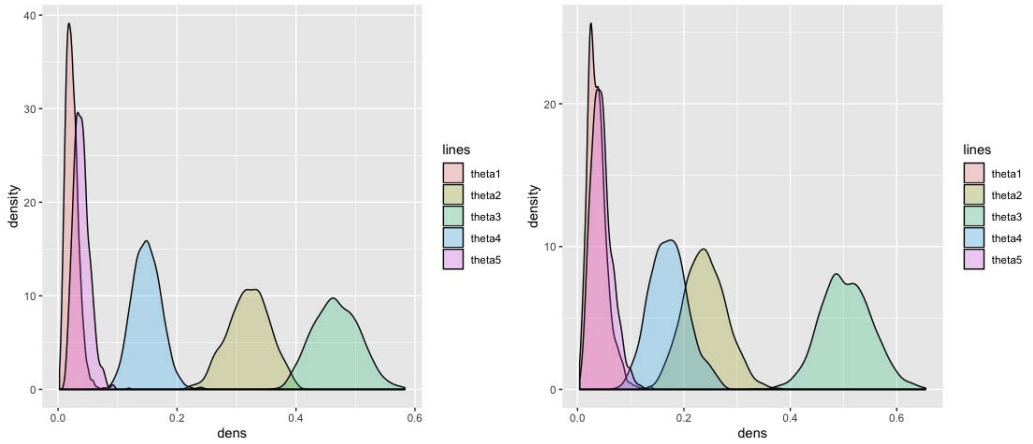
In Figure 3, the top panel is the model with order restrictions and the bottom panel is the model without any order restriction for the same county, County 1. It can be seen from the plots of the posterior densities of the θ's that $\boldsymbol{\theta}$ in this county has an order restriction. Our unimodal assumption for this county holds. However in the first density (top panel) and the second density (bottom panel), there are overlaps between θ_2 and θ_3. It means that the order restriction may not hold for this county. The overlap between θ_1 and θ_5 is acceptable, since there is no direct comparison between them. Specially in the bottom panel, the densities from the model without order restriction show that θ_2 is even larger than θ_3. Our unimodal assumption may not be proper in this county.

Figure 3: Posterior Densities of $\boldsymbol{\theta}$ for County 1



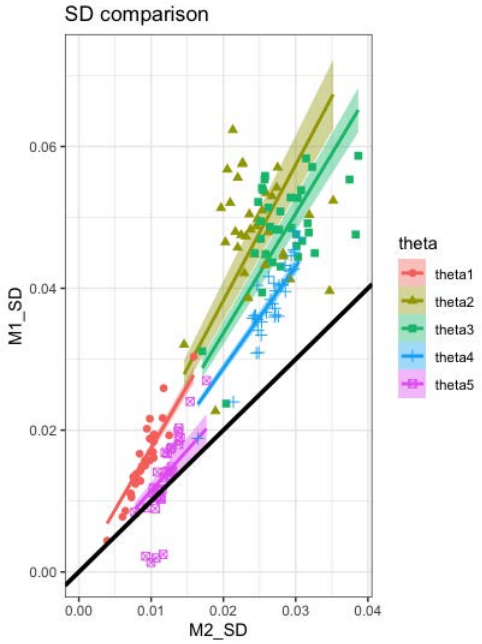
In Figure 4, the top panel is the model with order restrictions and the bottom panel is the model without any order restriction for another county, County 3. Plots of the posterior densities of the θ's without any order restriction show that $\boldsymbol{\theta}$s in each county may have an order restriction. It can be seen from the second density (bottom) that θ_3 is the mode for the cell probabilities even without order restriction assumption. It means that our unimodal assumption in this county is valid. Like in Figure 3, the overlap between θ_1 and θ_5 is acceptable, since there is no direct comparison between them.

Figure 4: Posterior Densities of $\boldsymbol{\theta}$ for County 3



In Figure 5, we use posterior standard deviations (PSD) to generate regression lines. Those regression lines show the overall PSD comparison between the model with order restrictions (M2) and the model without order restriction (M1). If the slope of regression line is larger than the black reference line whose slope is one, it means that M2 has smaller PSDs than M1. For each cell probability θ shown in different color, the slope is larger than the reference line's. Therefore we gain higher precision on estimation of cell probabilities among 35 counties.

Figure 5: Posterior standard Deviation Comparisons of $\boldsymbol{\theta}$s'



The LPML of Multinomial Dirichlet model with order restrictions is -977.102, and the LPML of Multinomial Dirichlet model without any order restriction is -471.821. The LPMLs show that the model without order restriction is better than the model with order restriction. This result is puzzling because we found better precision for the model with order restriction. To look at this issue more closely, in Table 4, we present the $\log(CPO)$ for the individual counties. Most of the $\log(CPO)$ are larger under Model M1 than Model M2, except for county 21 (smallest sample size) where the order restriction is satisfied.

Table 4: $\log(CPOi)$ for each county by model

County	size	M2	M1	County	size	M2	M1
1	97	-22.957	-16.485	18	61	-20.853	-13.894
2	91	-29.87	-13.948	19	52	-19.404	-7.793
3	90	-22.368	-17.386	20	64	-27.941	-13.975
4	412	-109.684	-21.155	21	49	-11.848	-16.226
5	80	-20.223	-14.695	22	77	-26.737	-15.169
6	66	-16.642	-16.232	23	50	-16.403	-13.573
7	62	-18.532	-13.434	24	70	-19.817	-14.386
8	53	-20.073	-13.669	25	64	-30.908	-10.063
9	73	-22.255	-13.11	26	60	-16.544	-14.578
10	81	-32.178	-8.714	27	48	-14.371	-13.135
11	98	-17.48	-16.144	28	52	-16.706	-12.687
12	84	-38.694	-16.25	29	75	-28.701	-13.678
13	217	-67.18	-19.439	30	82	-23.83	-13.209
14	72	-25.363	-14.297	31	75	-19.882	-14.452
15	87	-21.267	-16.952	32	102	-33.623	-14.868
16	101	-35.126	-14.356	33	129	-37.243	-15.985
17	99	-22.601	-15.412	34	84	-29.009	-15.338
				35	92	-40.79	-10.593

[1]Note: Shaded Area: The model with order restrictions (M2)
Unshaded Area: The model without any order restriction (M1)

For the county with the largest sample size, the $\log(CPO)$ under model M2 is much too small even though the order restriction is satisfied. There are a few counties in which the $\log(CPO)$ are comparable (e.g., counties 6, 11, 26 and 27). There are two explanations for these findings. First, for the few counties with larger sample sizes, because the $\log(CPO)$ is based on deletion and prediction, the model is doing a bad job in predicting for the counties with larger sample sizes. Second, if the order restrictions are not fully satisfied (see Figure 3), then the fitted values will not agree with the observed values, thereby causing the discrepancy.

Next we split 35 counties into 3 groups based on the ratio of people in level 2 and level 3. We expect better performances of our proposed model in each subgroup after splitting. In Group 1, the number of people in BMI level 3 is at least 10% more than people in BMI level 2. The mode is more likely at the third position in Group 1. Controversially in Group 3, the number of people in BMI level 2 is at least 10% more than people in BMI in level 3. Then we fit the model with order restrictions for each group, seen in Table 5. As we can see, we have some improvement in Group 3 but not in Group 1. One possible reason is that some counties with large sizes, eg County 4 and County 13, will cause the problem. Since CPO is one kind of leave-one-out methods, it will be hard for a large county to borrow information from other counties.

Table 5: $\log(CPOi)$ for each county

	Group 1				Group 3		
	County ID	M1	M2		County ID	M1	M2
1	27	-11.2098	-14.4154	1	31	-11.9776	-19.3504
2	3	-16.38	-22.0339	2	14	-11.6669	-23.6478
3	20	-10.0768	-28.1033	3	28	-10.2338	-15.476
4	32	-12.6318	-34.1183	4	13	-19.8629	-63.2454
5	30	-12.1297	-24.5537	5	22	-11.4118	-25.6294
6	15	-14.3668	-22.9273	6	25	-11.1472	-28.6044
7	17	-13.7607	-22.8095	7	18	-10.928	-19.3792
8	34	-11.9051	-29.7782	8	19	-10.6955	-17.8958
9	16	-12.3094	-36.3469	9	21	-12.7922	-9.61966
10	26	-11.7787	-16.8221				
11	4	-23.8692	-113.144				
12	8	-10.4124	-19.9604				
13	23	-10.5611	-16.801				
14	33	-13.6766	-39.3604				
	LPML	-185.067	-441.175		LPML	-110.716	-222.848
	Previous LPML	-202.112	-390.841		Previous LPML	-129.929	-251.255

[2]Note: Shaded Area: The model with order restrictions (M2)
Unshaded Area: The model without any order restriction (M1)

Perhaps one can consider other Bayesian diagnostics (e.g., deviance information criterion - DIC - and Bayes factors). Mode uncertainty can be considered in future study to create a more flexible model such as the one inspired by Nandram (1997). That may help to get a larger LPML for the model with order restriction than the model without any order restriction.

6. Conclusion

Hierarchical Bayesian multinomial Dirichlet models can be used to make inference for small areas. We have proposed the model with order restrictions to increase the accuracy of the estimation for the parameters. We have also shown how to generate samples from posterior distributions with order restrictions. We have significantly increased the precision of the estimation of cell probabilities for cells 2, 3 and 4 for the female BMI data. This is true for most of the counties. We have also shown difficulties in assessing model fit using Bayesian diagnostic measures ($\log(CPO)$) under order restriction; we believe that this is an open problem.

However, as shown in Figure 3, the same unimodal assumption may be too strong. Some counties have more people in BMI level 2 than level 3, for instance County 1; some counties have opposite situations. It seems that the mode of cell probabilities in each county is not fixed. Nandram and Sedransk (1995) and Nandram, Sedransk and Smith (1997) presented a good discussion about unimodal order restriction in a stratified population. They made inference about the proportion of firms belonging to each of several classes when there are unimodal order relations among the proportions. In that paper, the hyperparameters are specified and they did not have a small area estimation problem; our problem is much more difficult. They discussed an extension of their approach, which is the uncertain modal positions for their case. So for our model, one possible solution is considering uncertainty of modal position of the cell probabilities. We can consider a random variable to

indicate the type of order restriction for each county. In other words, even counties have different order restriction structure, we can borrow strenght among them with the help of uncertainty. Uncertain modal positions may be more suitable for our BMI data. But again this is a much more difficult problem. If order restriction obviously exist and each small area has similar size, the multinomial Dirichlet model with order restrictions will work well.

In future, we might want to make the correlation structure of the Dirichlet distribution (all components are negatively correlated) more flexible. This can be done by using multivariate logistic models with similar unimodal order restrictions as are studied in this paper. Also, we can use more flexible prior distributions such as Dirichlet process on the cell probabilities; this is a difficult problem with the unimodal order restrictions.

7. Appendix

7.1 Model Comparison

Table 6: Posterior Mean, Standard Deviation and Coefficient of Variation

	θ_1			θ_2			θ_3			θ_4			θ_5		
Counties	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV	PM	PSD	CV
1	0.0234	0.0103	0.4402	0.3751	0.0203	0.0541	0.4157	0.0243	0.0585	0.1505	0.0248	0.1648	0.0354	0.0128	0.3616
	0.0356	0.0169	0.4747	0.3959	0.0465	0.1175	0.3709	0.0449	0.1211	0.163	0.0341	0.2092	0.0345	0.0175	0.5072
2	0.0152	0.0085	0.5592	0.3726	0.0243	0.0652	0.4275	0.0279	0.0653	0.162	0.026	0.1605	0.0227	0.0106	0.4670
	0.0198	0.0128	0.6465	0.3807	0.0448	0.1177	0.4011	0.0431	0.1075	0.1892	0.0366	0.1934	0.0092	0.0089	0.9674
3	0.0233	0.0104	0.4464	0.3203	0.0347	0.1083	0.4678	0.0383	0.0819	0.1489	0.0241	0.1619	0.0397	0.0138	0.3476
	0.0374	0.0184	0.4920	0.24	0.0396	0.1650	0.5051	0.0476	0.0942	0.1719	0.0358	0.2083	0.0455	0.0198	0.4352
4	0.0079	0.0039	0.4937	0.3583	0.0189	0.0527	0.4224	0.0204	0.0483	0.1787	0.0165	0.0923	0.0327	0.0077	0.2355
	0.0078	0.0044	0.5641	0.35	0.0227	0.0649	0.4182	0.0237	0.0567	0.1916	0.0188	0.0981	0.0324	0.0084	0.2593
5	0.0158	0.0087	0.5506	0.3625	0.0231	0.0637	0.4151	0.0264	0.0636	0.174	0.027	0.1552	0.0326	0.0123	0.3773
	0.0213	0.0149	0.6995	0.3544	0.0473	0.1335	0.3778	0.0487	0.1289	0.216	0.0407	0.1884	0.0306	0.0167	0.5458
6	0.012	0.0085	0.7083	0.3662	0.0254	0.0694	0.4195	0.0303	0.0722	0.1515	0.0277	0.1828	0.0507	0.0177	0.3491
	0.0146	0.0128	0.8767	0.3649	0.051	0.1398	0.3798	0.0527	0.1388	0.1722	0.0404	0.2346	0.0685	0.027	0.3942
7	0.0118	0.0079	0.6695	0.3634	0.0259	0.0713	0.4192	0.0293	0.0699	0.1749	0.0298	0.1704	0.0307	0.0127	0.4137
	0.0149	0.0138	0.9262	0.3594	0.0534	0.1486	0.3737	0.0528	0.1413	0.2276	0.0451	0.1982	0.0245	0.0173	0.7061
8	0.0174	0.0094	0.5402	0.3743	0.0274	0.0732	0.4419	0.0323	0.0731	0.1392	0.0281	0.2019	0.0272	0.013	0.4779
	0.0304	0.0202	0.6645	0.3748	0.0542	0.1446	0.4314	0.0571	0.1324	0.1488	0.0405	0.2722	0.0145	0.0142	0.9793
9	0.0166	0.0092	0.5542	0.3694	0.0236	0.0639	0.4184	0.0279	0.0667	0.1704	0.0292	0.1714	0.0253	0.0114	0.4506
	0.0242	0.0158	0.6529	0.3806	0.0504	0.1324	0.3731	0.0508	0.1362	0.2115	0.0424	0.2005	0.0105	0.0102	0.9714
10	0.0106	0.0072	0.6792	0.3743	0.0217	0.0580	0.4202	0.0252	0.0600	0.1758	0.0272	0.1547	0.0189	0.0093	0.4921
	0.0116	0.0109	0.9397	0.3883	0.048	0.1236	0.3847	0.047	0.1222	0.215	0.0396	0.1842	0.0004	0.0022	5.5000
11	0.023	0.0104	0.4522	0.3575	0.0229	0.0641	0.4041	0.0268	0.0663	0.1751	0.0272	0.1553	0.0402	0.0138	0.3433
	0.0351	0.0161	0.4587	0.3545	0.0421	0.1188	0.3595	0.0437	0.1216	0.209	0.0363	0.1737	0.0419	0.018	0.4296
12	0.0108	0.0071	0.6574	0.3898	0.0237	0.0608	0.4419	0.0272	0.0616	0.1297	0.0252	0.1943	0.0278	0.0118	0.4245
	0.0119	0.011	0.9244	0.4146	0.0483	0.1165	0.4307	0.0483	0.1121	0.1238	0.0334	0.2698	0.0189	0.0133	0.7037
13	0.0127	0.006	0.4724	0.3894	0.0146	0.0375	0.4099	0.0171	0.0417	0.1524	0.0214	0.1404	0.0357	0.0103	0.2885
	0.0138	0.0078	0.5652	0.4367	0.0321	0.0735	0.364	0.0311	0.0854	0.1507	0.024	0.1593	0.0348	0.0119	0.3420
14	0.0215	0.0105	0.4884	0.3845	0.0209	0.0544	0.4243	0.0262	0.0617	0.1444	0.0268	0.1856	0.0254	0.0113	0.4449
	0.0357	0.0193	0.5406	0.4264	0.0521	0.1222	0.3747	0.0514	0.1372	0.1524	0.0378	0.2480	0.0108	0.0104	0.9630
15	0.0194	0.0098	0.5052	0.3604	0.0283	0.0785	0.4381	0.0309	0.0705	0.1405	0.0243	0.1730	0.0417	0.014	0.3357
	0.03	0.0157	0.5233	0.3411	0.0449	0.1316	0.4355	0.0467	0.1072	0.1475	0.0361	0.2447	0.046	0.0189	0.4109
16	0.0141	0.0078	0.5532	0.3686	0.0258	0.0700	0.4378	0.0301	0.0688	0.1581	0.0253	0.1600	0.0214	0.0094	0.4393
	0.0182	0.0124	0.6813	0.3573	0.0447	0.1251	0.4345	0.046	0.1059	0.1814	0.0352	0.1940	0.0086	0.009	1.0465
17	0.0309	0.0125	0.4045	0.3559	0.0258	0.0725	0.4281	0.0303	0.0708	0.1591	0.0247	0.1552	0.026	0.0109	0.4192
	0.0511	0.0193	0.3777	0.3313	0.0433	0.1307	0.4183	0.0444	0.1061	0.1827	0.0356	0.1949	0.0165	0.0117	0.7091
18	0.012	0.0079	0.6583	0.378	0.022	0.0582	0.4141	0.0248	0.0599	0.1638	0.0293	0.1789	0.0321	0.0128	0.3988
	0.0154	0.0134	0.8701	0.425	0.0556	0.1308	0.3388	0.0522	0.1541	0.1956	0.0426	0.2178	0.0252	0.0175	0.6944
19	0.0181	0.0103	0.5691	0.3769	0.0226	0.0600	0.4162	0.0257	0.0617	0.1664	0.0305	0.1833	0.0223	0.0117	0.5247
	0.0304	0.0193	0.6349	0.4237	0.0575	0.1357	0.3426	0.0554	0.1617	0.2027	0.0472	0.2329	0.0005	0.0025	5.0000
20	0.0114	0.0075	0.6579	0.3547	0.0319	0.0899	0.4644	0.0374	0.0805	0.1445	0.0247	0.1709	0.025	0.0115	0.4600
	0.0143	0.0134	0.9371	0.3042	0.0504	0.1657	0.5057	0.0553	0.1094	0.1644	0.0404	0.2457	0.0115	0.0116	1.0087
21	0.0355	0.0159	0.4479	0.3745	0.0213	0.0569	0.4071	0.0254	0.0624	0.1428	0.0289	0.2024	0.04	0.0154	0.3850
	0.0745	0.0303	0.4067	0.4398	0.0623	0.1417	0.2993	0.0539	0.1801	0.1428	0.0412	0.2885	0.0437	0.024	0.5492
22	0.0162	0.0092	0.5679	0.3853	0.0197	0.0511	0.4222	0.0251	0.0595	0.1467	0.0271	0.1847	0.0295	0.0126	0.4271
	0.0231	0.015	0.6494	0.4384	0.0513	0.1170	0.3673	0.0495	0.1348	0.1509	0.036	0.2386	0.0202	0.0141	0.6980
23	0.0124	0.0084	0.6774	0.3689	0.0275	0.0745	0.4336	0.0314	0.0724	0.1515	0.0294	0.1941	0.0336	0.0138	0.4107
	0.0173	0.0167	0.9653	0.3652	0.057	0.1561	0.4053	0.0583	0.1438	0.1839	0.0446	0.2425	0.0284	0.0203	0.7148
24	0.0117	0.0076	0.6496	0.3629	0.0247	0.0681	0.4171	0.0294	0.0705	0.173	0.0285	0.1647	0.0353	0.0138	0.3909
	0.0141	0.0133	0.9433	0.364	0.0498	0.1368	0.3731	0.0485	0.1300	0.2158	0.0432	0.2002	0.033	0.0185	0.5606
25	0.0175	0.0099	0.5657	0.3944	0.0205	0.0520	0.4304	0.0253	0.0588	0.1366	0.0276	0.2020	0.021	0.0107	0.5095
	0.0255	0.0168	0.6588	0.4673	0.0568	0.1215	0.3769	0.0542	0.1438	0.13	0.0361	0.2777	0.0004	0.002	5.0000
26	0.0226	0.0116	0.5133	0.3676	0.0267	0.0726	0.4336	0.0308	0.0710	0.1457	0.0263	0.1805	0.0306	0.0119	0.3889
	0.0401	0.0217	0.5411	0.3571	0.053	0.1484	0.4178	0.0538	0.1288	0.1608	0.0417	0.2593	0.0242	0.0169	0.6983
27	0.0233	0.0117	0.5021	0.3356	0.0352	0.1049	0.4516	0.0387	0.0857	0.162	0.0298	0.1840	0.0275	0.0125	0.4545
	0.0464	0.0259	0.5582	0.2428	0.0524	0.2158	0.4785	0.0587	0.1227	0.2177	0.048	0.2205	0.0146	0.0136	0.9315
28	0.018	0.0098	0.5444	0.3758	0.0227	0.0604	0.4179	0.0258	0.0617	0.1605	0.0297	0.1850	0.0277	0.0128	0.4621
	0.0311	0.0216	0.6945	0.4109	0.0577	0.1404	0.3531	0.0559	0.1583	0.1903	0.0455	0.2391	0.0147	0.0148	1.0068
29	0.0114	0.0072	0.6316	0.3756	0.0246	0.0655	0.4338	0.0301	0.0694	0.1542	0.0271	0.1757	0.025	0.0112	0.4480
	0.0129	0.0111	0.8605	0.3832	0.0488	0.1273	0.4146	0.0476	0.1148	0.1784	0.0372	0.2085	0.0109	0.0106	0.9725
30	0.0153	0.0085	0.5556	0.3484	0.0283	0.0812	0.4254	0.0318	0.0748	0.1877	0.0297	0.1582	0.0231	0.0107	0.4632
	0.0219	0.0144	0.6575	0.3101	0.0445	0.1435	0.4074	0.0479	0.1176	0.2504	0.0427	0.1705	0.0102	0.0104	1.0196
31	0.0208	0.0098	0.4712	0.373	0.0225	0.0603	0.4159	0.0277	0.0666	0.1608	0.0286	0.1779	0.0295	0.0122	0.4136
	0.0337	0.0188	0.5579	0.3953	0.0475	0.1202	0.3633	0.0483	0.1329	0.1858	0.0396	0.2131	0.0218	0.0142	0.6514
32	0.0097	0.0064	0.6598	0.3412	0.0292	0.0856	0.439	0.0326	0.0743	0.184	0.0265	0.1440	0.026	0.0111	0.4269
	0.0102	0.0086	0.8431	0.2919	0.0413	0.1415	0.4469	0.045	0.1007	0.2346	0.0384	0.1637	0.0163	0.0114	0.6994
33	0.0128	0.0073	0.5703	0.3653	0.0234	0.0641	0.4228	0.0254	0.0601	0.172	0.0249	0.1448	0.0271	0.0103	0.3801
	0.015	0.0105	0.7000	0.3586	0.0386	0.1076	0.4082	0.0394	0.0965	0.1982	0.0309	0.1559	0.02	0.0111	0.5550
34	0.0156	0.0085	0.5449	0.37	0.0261	0.0705	0.4451	0.0317	0.0712	0.1414	0.0244	0.1726	0.0278	0.0109	0.3921
	0.0217	0.0142	0.6544	0.3584	0.0479	0.1336	0.4466	0.0492	0.1102	0.1532	0.0362	0.2363	0.0201	0.0141	0.7015
35	0.015	0.0082	0.5467	0.3914	0.022	0.0562	0.4409	0.0258	0.0585	0.1342	0.0246	0.1833	0.0184	0.01	0.5435
	0.02	0.0129	0.6450	0.4227	0.0457	0.1081	0.4236	0.0448	0.1058	0.1334	0.0309	0.2316	0.0003	0.0013	4.3333

[3]Note: Shaded Area: The model with order restrictions (M2)
Unshaded Area: The model without any order restriction (M1)

References

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 order-restricted inference of multinomial counts from small areas. *Tentatively accepted by Modern Age Statistics*.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904.

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

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.

Malinovsky, Y. and Rinott, Y. (2010). Prediction of ordered random effects in a simple small area model. *Statistica Sinica*, pages 697–714.

Nandram, B. (1997). Bayesian inference for the best ordinal multinomial population in a taste test. In Gatsonis, C., Hodges, J. S., Kass, R. E., McCulloch, R., Rossi, P., and Singpurwalla, N. D., editors, *Case Studies in Bayesian Statistics*, pages 399–418, New York, NY. Springer New York.

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., 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 Sedransk, J. (1995). Bayesian inference for the mean of a stratified population when there are order restrictions. In *Case Studies in Bayesian Statistics, Volume II*, pages 309–322. Springer.

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.

Rao, J. and Molina, I. (2015). *Small Area Estimation*. Wiley Series in Survey Methodology. Wiley.

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.

Wu, J., Meyer, M. C., and Opsomer, J. D. (2016). Survey estimation of domain means that respect natural orderings. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 44(4):431–444.