Order-Restricted Bayesian Estimation of Multinomial Counts for Small Areas

Xinyu Chen* Balgobin Nandram†

Abstract

We consider making inference about several small areas with data obtained in the form of multinomial counts. The cell probabilities have the same unimodal order restriction across areas, and these cell probabilities share a common effect. Therefore, a hierarchical multinomial-Dirichlet model, used to model the cell probabilities and the cell counts, allows a borrowing of strength across areas. We used the Gibbs sampler to make inference about the order-restricted multinomial parameters. We show how to perform the computations because there are difficulties poised by the order restrictions. An application on body mass index showed that the order restriction is necessary and it provides increased precision over the scenario without the restriction.

Key Words: Bayesian computation, Gibbs sampler, Multinomial counts, Monte Carlo Integration, Unimodal order restrictions.

1. Introduction

Traditional sample surveys are not designed for the estimations of small areas such as counties. They can not provide enough samples to produce reliable estimates. Running a specific well-designed survey for small areas will be expensive and even impossible. The use of some suitable statistical models on those samples from traditional sample surveys can provide improved precision for the estimations of small areas. Hierarchical Bayesian models, such as multinomial-Dirichlet model (e.g. Nandram 1998), are used to make inference about the finite population proportion of each small-area. Nandram (1997) provided a great discussion of the methods, techniques, and approaches for the Multinomial-Dirichlet model under Bayesian framework. In this paper, we discuss the hierarchical multinomial-Dirichlet model associated with the analysis of body mass index data, which is multilevel survey data.

Body mass index (BMI) is a ratio of body weight(kg) and the square of the body height(m). This is important measure the public health since overweight individuals may have a higher risk for some diseases, from Nandram, Kim, and Zhou(2019). Our research concerns improved estimation of the BMI composition of the United States. BMI data constructed from several counties have five levels (underweight, normal, overweight, obese1 and obese2).

Nandram, Sedransk, and Smith (1997) provided an improved estimation of the age composition of a population of fish under a Bayesian framework. Kim, Reiter, Wang, Cox, and Karr (2014) presented a fully Bayesian model for missing values under linear constraints. Heck and Davis-Stober (2019) provided Bayesian analyses for the linear inequality constraints problem. In our application, it is reasonable to believe most people in different counties will fall in the second level of BMI, which is a normal level. Our key contribution is to bring this restriction into our model to hopefully gain a higher accuracy.

We show how to perform more efficient and accurate computations using Monte Carlo integration and Gibbs samples Gelfand, Smith, and Lee (1992). We use trace plot, auto

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

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

correlation and effective sample size to assess the convergence. Then we compare our proposed model with Dirichlet multinomial Bayesian hierarchical model without any order restriction. In our future work, we will test that the order restriction is necessary using the Bayes factor.

2. Multinomial Dirichlet Model

In this section, we present a brief review of Multinomial-Dirichlet model and its extension with the order restriction.

2.1 General Multinomial Dirichlet Model

Nandram, Kim, and Zhou(2019) had a good discussion of the general multinomial Dirichlet model and their methodology. The Dirichlet distribution, $\boldsymbol{x} \sim$ Dirichlet($\boldsymbol{\alpha}$), is parameterized by positive scalar $\alpha_j > 0$ for $j = 1, 2, \ldots, K$, where $K \geqslant 2$. The probability density of $\boldsymbol{x}$ is

$$f(\boldsymbol{X}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} \prod_{j=1}^{K} x_j^{\alpha_j - 1}.$$

The Multinomial distribution, $\boldsymbol{n} \sim$ Multinomial($n_{\cdot}$, $\boldsymbol{\theta}$), is a discrete distribution over K dimensional non-negative integer vectors $\boldsymbol{n}$ where $\sum_{j=1}^{K} n_j = n_{\cdot}$. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$. The probability mass function is given as

$$f(\boldsymbol{n}|\boldsymbol{\theta}) = \frac{\Gamma(n_{\cdot} + 1)}{\prod_{j=1}^{K} \Gamma(n_j + 1)} \prod_{j=1}^{K} \theta_j^{n_j}, \qquad \sum_{j=1}^{K} n_j = n_{\cdot}, \qquad n_i \geqslant 0.$$

The Bayesian hierarchical multinomial Dirichlet model: Letting n_{ij} be the cell counts , θ_{ij} the corresponding cell probabilities, $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, K$, and $n_{i\cdot} = \sum_{j=1}^{K} n_{ij}$, The general hierarchical Bayesian model is

$$\begin{aligned}
&\boldsymbol{n}_i|\boldsymbol{\theta}_i \overset{ind}{\sim} \text{Multinomial}(\boldsymbol{n}_{i\cdot}, \boldsymbol{\theta}_i), \\
&\boldsymbol{\theta}_i|\boldsymbol{\mu}, \tau \overset{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau), \\
&\pi(\boldsymbol{\mu}, \tau) = \frac{(K-1)!}{(1+\tau)^2}, \\
&\qquad \mu_j > 0, \quad \sum_{j=1}^{K} \mu_j = 1 \quad , \tau > 0,
\end{aligned}$$

where, without any prior information, we take $\boldsymbol{\mu}$ and τ to be independent.

2.2 Multinomial Dirichlet Model with Order Restriction

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

$$\boldsymbol{n}_i|\boldsymbol{\theta}_i \overset{ind}{\sim} \text{Multinomial}(\boldsymbol{n}_{i\cdot}, \boldsymbol{\theta}_i), \qquad \boldsymbol{\theta}_i \in C \quad i = 1, \ldots, I,$$

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. At the second stage we take

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

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

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

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

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

A posteriori $\boldsymbol{\theta}_i|\boldsymbol{\mu}, \tau, \boldsymbol{n}_i \overset{ind}{\sim} Dirichlet(\boldsymbol{n}_i + \boldsymbol{\mu}\tau), \qquad \boldsymbol{\theta}_i \in C_i, i = 1, \ldots, I$ where

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

3. Methodology with Order Restriction

3.1 Gibbs Sampling

Liu and Sabatti (2000) presented a comprehensive discussion of the general Gibbs sampler which is more efficient the Markov chain Monte Carlo method for Bayesian inference. We take advantage of Gibbs sampler to generate the posterior samples for the Bayesian inference.

In this section, we present the modified Gibbs sampler for $\boldsymbol{\mu} \in C_{\boldsymbol{\mu}}$ and τ.

Algorithm:

1. Draw τ from $\pi(\tau|\boldsymbol{\mu}, \boldsymbol{n})$,
2. For j from m-1 to 1,

   Draw μ_j from $\pi(\mu_j|\boldsymbol{\mu}^{(-j)}, \tau, \boldsymbol{n})$, where $0 < \mu_j < min\{\mu_{j+1}, \frac{1 - \sum_{t=1, t\neq m, t\neq j}^{K} \mu_t}{2}\}$,
3. For j from m+1 to K,

   Draw μ_j from $\pi(\mu_j|\boldsymbol{\mu}^{(-j)}, \tau, \boldsymbol{n})$, where $0 < \mu_j < min\{\mu_{j-1}, \frac{1 - \sum_{t=1, t\neq m, t\neq j}^{K} \mu_t}{2}\}$,
4. Get $\mu_m = 1 - \sum_{j=1, j\neq m}^{K} \mu_j$

3.2 Dirichlet Distribution of $\boldsymbol{\theta}$ with Order Restriction

$\boldsymbol{\theta}$ has a recognizable distribution, which is the Dirichlet distribution with the order restriction. We can generate samples from its distribution using the exact method. Sedransk, Monahan, and Chiu (1985) provided an efficient algorithm to generate random vectors from the constrained density. However instead of drawing samples directly from the Dirichlet distribution with the order restriction, we present an alternative method related to drawing samples from the Gamma distributions.

To simplify the problem about how to draw samples from the constrained Dirichlet distribution, let $\boldsymbol{\alpha}_i = \boldsymbol{n}_i + \boldsymbol{\mu}\tau$

$$\boldsymbol{\theta}_i|\boldsymbol{\alpha}_i \overset{ind}{\sim} Dirichlet(\boldsymbol{\alpha}_i), \qquad \boldsymbol{\theta}_i \in C.$$

Then we draw samples from the Gamma distribution to construct the samples from the constrained Dirichlet distribution.

Algorithm:

Denote $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$,
If $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_m \geqslant \ldots \geqslant \theta_K$, the mode is θ_m,
$0 \leqslant \beta_1 \leqslant \beta_2 \leqslant \ldots \leqslant \beta_m \geqslant \ldots \geqslant \beta_K$, the mode is β_m,

1. Draw $\beta_m \sim Gamma(\alpha_m, 1)$, where $0 \leqslant \beta_m < \infty$;
2. Draw $\beta_{m-1} \sim Truncated\, Gamma(\alpha_{m-1}, 1)$, where $0 \leqslant \beta_{m-1} \leqslant \beta_m$,
   $\ldots$ $\beta_1 \sim Truncated\, Gamma(\alpha_1, 1)$, where $0 \leqslant \beta_1 \leqslant \beta_2$;
3. Draw $\beta_{m+1} \sim Truncated\, Gamma(\alpha_{m+1}, 1)$, where $0 \leqslant \beta_{m+1} \leqslant \beta_m$,
   $\ldots \beta_K \sim Truncated\, Gamma(\alpha_K, 1)$, where $0 \leqslant \beta_K \leqslant \beta_{K-1}$.

Then,

$$\theta_1 = \frac{\beta_1}{\beta_1 + \beta_2 + \ldots + \beta_K},$$
$$\theta_2 = \frac{\beta_2}{\beta_1 + \beta_2 + \ldots + \beta_K},$$

$$\vdots$$

$$\theta_{K-1} = \frac{\beta_{K-1}}{\beta_1 + \beta_2 + \ldots + \beta_{K-1}},$$

$$\theta_K = 1 - \sum_{i=1}^{K-1} \theta_i.$$

4. Application to Body Mass Index Data

4.1 Body Mass Index

As an illustrative example, we use a set of body mass index data from a national survey, where we use only the female BMI data from the 35 largest counties with a population at least 500,000. Our goal is to predict the percentage of the normal BMI level. Table 1 gives an illustration of the female BMI data.

Table 1: US Female BMI data

State	County	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
...	...	...	...	...	...	...

4.2 MCMC Convergence

We ran 50,000 MCMC iterations, used 10,000 as a 'burn in' and used every 40th 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. The p-values are all large so we can not reject that null hypothesis that the MC is stationary. Then posterior samples can be used for the further inference.

Table 2: Effective Sizes

Models	μ_1	μ_2	μ_3	μ_4	μ_5	τ
W Order	973.7572	1000.0000	1000.0000	1000.0000	1000.0000	1000
W/O Order	859.3293	1000.0000	1000.0000	970.8221	1000.0000	1032.3

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

4.3 Model Comparison

We compute the estimated cell probabilities for each county and their variances, which are the sample mean of parameter $\boldsymbol{\theta}$ and the sample variances. In Table 4, the second level and the third level of BMI have lower variances compared with the general model. We have

higher accuracy for the estimation of the second level proportion. But for parameters θ_1 and θ_5, the model with order restriction lost some precision for the estimates of them.

As expected, Multinomial Dirichelt model with the order restriction has higher accuracy for the estimation of mode in parameters.

Table 4: Two Models Means and Variances Comparison

X	Θ_1.mean	SD1	Θ_2.mean	SD2	Θ_3.mean	SD3	Θ_4.mean	SD4	Θ_5.mean	SD5
1	0.02623	0.01282	0.37278	0.02312	0.41610	0.02723	0.14772	0.02940	0.03717	0.01445
1	0.02253	0.01038	0.39584	0.03403	0.40331	0.03267	0.14471	0.02362	0.03361	0.01242
2	0.01455	0.00956	0.37427	0.02585	0.43005	0.02955	0.16197	0.03053	0.01918	0.01105
2	0.01394	0.00774	0.38789	0.03210	0.41940	0.03224	0.15740	0.02490	0.02137	0.01000
3	0.02583	0.01223	0.30158	0.03724	0.48296	0.04170	0.14645	0.02840	0.04318	0.01622
3	0.02238	0.00968	0.32111	0.03459	0.46908	0.03570	0.14879	0.02337	0.03864	0.01324
4	0.00688	0.00378	0.35565	0.01947	0.42290	0.02074	0.18093	0.01786	0.03364	0.00808
4	0.00775	0.00392	0.36033	0.02005	0.42155	0.02090	0.17809	0.01720	0.03228	0.00746

[1]Note: Shaded Area: Our model with the order restriction
Unshaded Area: The model without any order restriction

5. Concluding Remarks and Future Work

Bayesian hierarchical multinomial Dirichlet models can be used to make inference for small areas, which is hard for Non-Bayesian models. We have proposed the model with the order restriction to increase the accuracy of the estimation for parameters. We increased the precision of the estimation of cell probabilities for level 2 and level 3 in the BMI data.

However, since one of assumptions for cell probabilities is that the sum has to be equal to one, our model can not increase the over-all accuracy of the estimation for all parameters. Our proposed model may be used for the estimation of some specific levels only.

We want to compare the multinomial Dirichlet model without any order restriction with the order restricted one using the Bayes factor.

The marginal likelihood for the model with the order restriction is given by

$$f(\boldsymbol{n}_1, \boldsymbol{n}_2, \ldots, \boldsymbol{n}_I | M_1) = \int_{\boldsymbol{\theta},\boldsymbol{\mu},\tau} \prod_{i=1}^{I} \{ \frac{n_{i.}!}{\prod_{j=1}^{K} n_{ij}!} \frac{\prod_{j=1}^{K} \theta_{ij}^{n_{ij}} \prod_{j=1}^{K} \theta_{ij}^{\mu_j \tau - 1}}{D(\boldsymbol{\mu}\tau) C(\boldsymbol{\mu}\tau)} \} \frac{I_{C_\mu}}{((1+\tau)^2} d\boldsymbol{\theta} d\boldsymbol{\mu} d\tau$$

Integral with respect to $\boldsymbol{\theta}$

$$= \int_{\boldsymbol{\mu},\tau} \prod_{i=1}^{I} \{ \frac{n_{i.}!}{\prod_{j=1}^{K} n_{ij}!} \frac{D(\boldsymbol{\mu}\tau + n_{i.}) C(\boldsymbol{\mu}\tau + n_{i.})}{D(\boldsymbol{\mu}\tau) C(\boldsymbol{\mu}\tau)} \} \frac{I_{C_\mu}}{((1+\tau)^2} d\boldsymbol{\mu} d\tau.$$

The marginal likelihood for the model without the order restriction is given by

$$f(\boldsymbol{n}_1, \boldsymbol{n}_2, \ldots, \boldsymbol{n}_I | M_2) = \int_{\boldsymbol{\mu},\tau} \prod_{i=1}^{I} \{ \frac{n_{i.}!}{\prod_{j=1}^{K} n_{ij}!} \frac{D(\boldsymbol{\mu}\tau + n_{i.}))}{D(\boldsymbol{\mu}\tau)} \} \frac{(K-1)!}{((1+\tau)^2} d\boldsymbol{\mu} d\tau.$$

We can not have a close form for $f(\boldsymbol{n}_1, \boldsymbol{n}_2, \ldots, \boldsymbol{n}_I | M_1)$ because of the order restriction of $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$. It will be harder to compute the Bayes factor for the model comparison. We will use Monte Carlo integration to estimate the marginal likelihood in the future.

REFERENCES

Gelfand A.E., Smith, A.F. 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.

Kim, H.J. , Reiter, J.P., Wang, Q., Cox, L.H., and Karr, A.F. (2014), "Multiple Imputation of Missing or Faulty Values Under Linear Constraints", *Journal of Business & Economic Statistics*, 32(3):375-386.

Liu, J.S. and Sabatti, C. (2000), "Generalised Gibbs sampler and multigrid Monte Carlo for bayesian computation", *Biometrika*, 87(2):353–369.

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

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.