Journal of the Korean Data & Information Science Society
2012, **23**(3), 587–593

http://dx.doi.org/10.7465/jkdi.2012.23.3.587
한국데이터정보과학회지

Bayesian estimation for finite population proportions in multinomial data

Sanggyu Kwak[1] · Dalho Kim[2]

[12]Department of Statistics, Kyungpook National University

Received 5 March 2012, revised 18 April 2012, accepted 23 April 2012

Abstract

We study Bayesian estimates for finite population proportions in multinomial problems. To do this, we consider a three-stage hierarchical Bayesian model. For prior, we use Dirichlet density to model each cell probability in each cluster. Our method does not require complicated computation such as Metropolis-Hastings algorithm to draw samples from each density of parameters. We draw samples using Gibbs sampler with grid method. We apply this algorithm to a couple of simulation data under three scenarios and we estimate the finite population proportions using two kinds of approaches. We compare results with the point estimates of finite population proportions and their standard deviations. Finally, we check the consistency of computation using different samples drawn from distinct iterates.

Keywords: Contingency table, Dirichlet prior, finite population proportion, hierarchical model, hyper-parameters.

1. Introduction

We often obtain contingency tables in various fields where the cells contain frequency counts of outcomes. One of the frequently occurring problems in a contingency table is estimating the population proportions in finite population setup. Usually the contingency table consists of two binary variables. For example, if there is a contingency table between smoking and lung cancer, smoking is the response variable and lung cancer is the group variable. One can estimate the smoking proportions in each group. A more general problem is estimating proportions in multinomial case.

There are lots of literatures on Bayesian methods for data analysis of contingency table. A selective review of the literature for the Bayesian analysis of contingency table is proposed by Leonard and Hsu (1994). Usually, the Bayesian inference for contingency tables is based on the hierarchical linear model with normal priors (Lindley and Smith, 1972). For example, Leonard (1972) used the logit transformation for binomial case and Novick *et al.* (1973) used an arc-sine transformation. These approximations decrease the accuracy. Nandram (1998) took the three-stage hierarchical model to estimate the finite population proportion for contingency tables. He estimated the finite population proportions using the empirical

[1] Ph.D. candidate, Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

distributions of the proportion. He used Metropolis-Hastings algorithm to draw samples. In this paper we draw samples using Gibbs sampler with grid method and estimate the finite population proportions.

The rest of the paper is organized as follows. In Section 2, we propose the Bayesian hierarchical multinomial model and derive posterior densities to draw samples. We also describe how to compute the finite population proportions using Gibbs sampler with grid method. In Section 3, we provide numerical studies to illustrate our results. Section 4 presents some discussions.

2. Methodology

We will use a three-stage hierarchical mutinomial model to allow for the uncertainty in the estimation of all the hyper-parameters.

2.1. Modeling

Consider a $I \times J$ contingency table with cell counts $\{n_{ij}\},\ i = 1, ..., I,\ j = 1, ..., J$. Let p_{ij} denote the corresponding cell probability that a unit falls in the ith row and jth column and $\boldsymbol{p}_i = (p_{i1}, ..., p_{iJ})'$. For $\boldsymbol{n}_i = (n_{i1}, ..., n_{iJ})'$, we assume that

$$\boldsymbol{n}_i|\boldsymbol{p}_i \overset{ind}{\sim} \text{Multinomial}(n_{i\cdot}, \boldsymbol{p}_i), i = 1, ..., I, \tag{2.1}$$

where

$$f(\boldsymbol{n}_i|\boldsymbol{p}_i) = n_i! \prod_{j=1}^{J} \left\{ \frac{p_{ij}^{n_{ij}}}{n_{ij}!} \right\},\ 0 \le n_{ij} \le n_{i\cdot},\ j = 1, 2, \cdots, J.$$

As a prior we assume that

$$\boldsymbol{p}_i|\boldsymbol{\mu}, \tau \overset{ind}{\sim} \text{Dirichlet}(\boldsymbol{\mu}\tau),\ i = 1, 2, \cdots, I, \tag{2.2}$$

where

$$\pi(\boldsymbol{p}_i|\boldsymbol{\mu}, \tau) \frac{\prod_{j=1}^{J} p_{ij}^{\mu_j\tau - 1}}{D(\boldsymbol{\mu}\tau)},\ 0 < p_{ij} < 1,\ \sum_{j=1}^{J} p_{ij} = 1 \text{ and } \boldsymbol{\mu} = (\mu_1, \cdots, \mu_J)'$$

with $D(\boldsymbol{\mu}\tau) = \{\Gamma(\mu_j\tau)\}/\Gamma(\tau),\ 0 < \mu_j < 1$ and $\sum_{j=1}^{J} \mu_j = 1$. The model (2.1) and (2.2) are usually used for many contingency tables in multinomial problems. Finally, we assume that

$$\boldsymbol{\mu} \sim \text{Dirichlet}(\boldsymbol{\mu}^{(0)}\tau^{(0)}),\ 0 < \mu_j < 1,\ \sum_{j=1}^{J} \mu_j = 1,\ \boldsymbol{\mu}^{(0)} = (\mu_1^{(0)}, \cdots, \mu_J^{(0)})' \tag{2.3}$$

and

$$\tau \sim \Gamma(\eta^{(0)}, \nu^{(0)}),\ \tau > 0, \tag{2.4}$$

where

$$\pi(\tau) = (\nu^{(0)})^{\eta^{(0)}} \tau^{\eta^{(0)}-1} e^{-\nu^{(0)}\tau} / \Gamma(\eta^{(0)}), \ \tau > 0.$$

Hence we specify a three-stage hierarchical Bayesian model for multinomial data from (2.1) to (2.4). For the hyper-parameters $\boldsymbol{\mu}^{(0)}$, $\tau^{(0)}, \eta^{(0)}$ and $\nu^{(0)}$, we assume that $\boldsymbol{\mu}^{(0)} = \mathbf{1}$, $\tau^{(0)} = 1$, $\eta^{(0)}$=1 and $\nu^{(0)}$=0. This is a uniform prior for $\boldsymbol{\mu}$ and a noninformative prior for τ. Obviously these priors are proper.

2.2. Computation

Using Bayes' theorem, the joint posterior density given the data is

$$\begin{aligned}\pi(\boldsymbol{p}, \boldsymbol{\mu}, \tau|\boldsymbol{n}) &\propto \prod_{i=1}^{I} f(\boldsymbol{n}_i|\boldsymbol{p}_i) \cdot \pi(\boldsymbol{p}_i|\boldsymbol{\mu}, \tau) \cdot \pi(\boldsymbol{\mu}) \cdot \pi(\tau) \\ &\propto \prod_{i=1}^{I} \left(\frac{\prod_{j=1}^{J} p_{ij}^{n_{ij}+\mu_j\tau-1}}{D(\boldsymbol{\mu}\tau)} \right) \cdot \pi(\boldsymbol{\mu}) \cdot \pi(\tau),\end{aligned}$$

where $\boldsymbol{p} = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_I)'$ and $\boldsymbol{n} = (\boldsymbol{n}_1, \cdots, \boldsymbol{n}_I)'$. To assess the computations we use the Rao-Blackwellizations to get the posterior density of $\boldsymbol{p}$. The joint posterior density given the data can be expressed by $\pi(\boldsymbol{p}, \boldsymbol{\mu}, \tau|\boldsymbol{n}) = \pi(\boldsymbol{p}|\boldsymbol{\mu}, \tau, \boldsymbol{n}) \times \pi(\boldsymbol{\mu}, \tau|\boldsymbol{n})$. By integrating out $\boldsymbol{p}$ from the joint posterior, the marginal joint posterior density of $\boldsymbol{\mu}, \boldsymbol{\tau}$ given $\boldsymbol{n}$ is given by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\tau}|\boldsymbol{n}) \propto \int_0^1 \pi(\boldsymbol{p}, \boldsymbol{\mu}, \tau|\boldsymbol{n}) d\boldsymbol{p} \propto \prod_{i=1}^{I} \left\{ \frac{D(\boldsymbol{n}_i + \boldsymbol{\mu}\tau)}{D(\boldsymbol{\mu}\tau)} \right\} \cdot \pi(\boldsymbol{\mu}) \cdot \pi(\tau).$$

Let $\boldsymbol{\mu}_{(j)}$ denote the vector of all components of $\boldsymbol{\mu}$ excluding the jth elements, $j = 1, \cdots, J$. Then we obtain the conditional posterior densities of $(\mu_j|\boldsymbol{\mu}_{(j)}, \tau, \boldsymbol{n})$, $j = 1, \cdots, J-1$. And the conditional posterior density of $(\mu_J|\boldsymbol{\mu}_{(J)}, \tau, \boldsymbol{n})$ is obtained using $\mu_J = 1 - \sum_{j=1}^{J-1} \mu_j$ because $\sum_{j=1}^{J} \mu_j = 1$. So each conditional posterior density of μ_j to draw the iterative Gibbs sampler is given by

$$\pi(\mu_j|\boldsymbol{\mu}_{(j)}, \tau, \boldsymbol{n}) \propto \prod_{i=1}^{I} \left\{ \frac{D(\boldsymbol{n}_i + \boldsymbol{\mu}\tau)}{D(\boldsymbol{\mu}\tau)} \right\}.$$

For τ, we transform τ to $\rho = 1/(1+\tau)$ because τ is expected to be small, so that drawing ρ using a grid method is efficient. Then, the conditional posterior density of $\rho|\boldsymbol{\mu}, \boldsymbol{n}$ to execute the iterative Gibbs sampler is as follows.

$$\pi(\rho|\boldsymbol{\mu}, \boldsymbol{n}) \propto \prod_{i=1}^{I} \left\{ \frac{D(\boldsymbol{n}_i + \boldsymbol{\mu}\frac{1-\rho}{\rho})}{D(\boldsymbol{\mu}\frac{1-\rho}{\rho})} \right\} \cdot \frac{1}{\rho^2}.$$

We draw a sample $(\boldsymbol{\mu}, \tau)$ from the above full conditionals, each in turn, and iterate the procedure by using Gibbs sampler with grid method. First, we consider μ_1 where $0 \leq \mu_1 \leq 1 - \sum_{j'=1, j'\neq 1}^{J-1} \mu_{j'}$. We draw μ_1 using grid method by following these steps.

· Step 1. Divide 100 intervals between 0 and $1-\sum_{j'=1,j'\neq 1}^{J-1}\mu_{j'}$ and take the point values $I_k, k=1,\cdots,101$ (i.e., $I_1=0$ and $I_{101}=1-\sum_{j'=1,j'\neq 1}^{J-1}\mu_{j'}$).

· Step 2. Calculate 100 mid-points ($M_k, k=1,\cdots,100$) for each intervals.

· Step 3. Input the mid-points to the conditional posterior density and calculate the values ($a_k, k=1,\cdots,100$) according to mid-points.

· Step 4. Calculate $b_k=a_k/A$, where $A=\sum_{k=1}^{100}a_k$.

· Step 5. Generate $u_1\sim$ Uniform$(0,1)$.

· Step 6. Select kth interval made by Step 1, which satisfies $b_k\leq u_1<b_{k+1}$.

· Step 7. Generate $u_2\sim$ Uniform(I_k,I_{k+1}) and $\mu_1=u_2$.

Next, we similarly draw μ_2 with the above algorithm. But we should use the updated μ_1 when we calculate $1-\sum_{j'=1,j'\neq 2}^{J-1}\mu_{j'}$. Finally, we obtain $\mu_J=1-\sum_{j=1}^{J-1}\mu_j$, where $\mu_1,\cdots,\mu_{J-1}$ are updated values. For ρ, we divide 100 intervals between 0 and 1 for the grid method.

2.3. Finite population proportions

Our finite population consists of I clusters with $N_{i\cdot}$ units in the ith cluster. The response of each unit will fall in one of J categories. Let N_{ij} be the total number of units which is unknown, responding in the jth category and the ith cluster. Our objective is to estimate finite population proportion for the jth category, namely

$$P_j=N^{-1}\sum_{i=1}^{I}N_{ij},\ j=1,\cdots,J \tag{2.5}$$

where $N=\sum_{i=1}^{I}\sum_{j=1}^{N_i}N_{ij}$. We assume that N is known for simplicity. We can rewrite (2.5) using seen part and unseen part, namely

$$P_j=\left\{\left(\sum_{i=1}^{I}n_{ij}\right)+\left(\sum_{i=1}^{I}(N_{ij}-n_{ij})\right)\right\}N^{-1},\ j=1,\cdots,J \tag{2.6}$$

The seen part n_{ij} is known, but the unseen part $N_{ij}-n_{ij}$ is unknown. Thus $\sum_{i=1}^{I}n_{ij}$ is known from the observed data where $\boldsymbol{n}$ is given.

We obtain estimates of finite population proportions (P_j) by two methods. The first one is implemented as follows. We generate p_{ij} from Dirichlet($\boldsymbol{\mu}\tau$) and draw $(N_{i1}-n_{i1},\cdots,N_{iJ}-n_{iJ})$ from Mulinomial($N_{i\cdot}-n_{i\cdot},\boldsymbol{p}_i$) at each iterate of the grid method. Then we compute P_j using (2.6). The posterior means and standard deviations of the P_j are obtained using the estimated empirical distribution of the P_j. Next, the second method is implemented as follows. The posterior mean of $P_j|\boldsymbol{n}$ is given by

$$E(P_j|\boldsymbol{n})=\left\{\sum_{i=1}^{I}n_{ij}+\sum_{i=1}^{I}(N_{i\cdot}-n_{i\cdot})E(p_{ij}|\boldsymbol{n})\right\}N^{-1}. \tag{2.7}$$

And the posterior variance of $P_j|\boldsymbol{n}$ is given by

$$Var(P_j|\boldsymbol{n})=\left\{\sum_{i=1}^{I}(N_{i\cdot}-n_{i\cdot})^2\cdot Var(p_{ij}|\boldsymbol{n})+\sum_{i=1}^{I}(N_{i\cdot}-n_{i\cdot})E(p_{ij}(1-p_{ij})|\boldsymbol{n})\right\}N^{-2}. \quad (2.8)$$

So we generate p_{ij} from Dirichlet($\boldsymbol{\mu}\tau$) and calculate $E(p_{ij}|\boldsymbol{n})$ and $Var(p_{ij}|\boldsymbol{n})$. Then we calculate posterior mean and posterior variance of $P_j|\boldsymbol{n}$ using (2.7) and (2.8), respectively.

3. Numerical studies

We consider three scenarios: (1) both $\boldsymbol{\mu}$ and τ are unknown, (2) $\boldsymbol{\mu}$ is unknown but τ is known, and (3) both $\boldsymbol{\mu}$ and τ are known to implement the computation. We use 500 iterates to "burn out" the grid method because all parameters drawn seem to be sufficiently stable after 500th iterate. And we take every fifth value to obtain 2,000 iterates. This is a very conservative algorithm, and it virtually does not have the autocorrelation. The result based on 10,000 iterates show no changes for the relevant parameters.

In our simulation, we take $N_{i\cdot} = 100$ and $n_{i\cdot} = 20$, $i = 1, \cdots, I$ where $I = 10$. We consider two multinomial models, the first with two cells ($\mu_1 = 0.4 = 1 - \mu_2$) and the second with five cells ($\mu_1 = \mu_2 = 0.1$, $\mu_3 = \mu_4 = 0.2$ and $\mu_5 = 0.4$). In two models, we start with $\tau = 5, 10, 25$.

Table 3.1 Comparison of the posterior distributions of finite population proportions for two cell multinomial model by scenarios with $\mu_1 = 0.4$

τ	Scenario	P_1				
		Method 1			Method 2	
		Mean	STD	NSE	Mean	STD
5	1	0.3967	0.0271	0.0029	0.3968	0.0161
	2	0.3946	0.0253	0.0020	0.3949	0.0145
	3	0.4073	0.0198	0.0031	0.4072	0.0155
10	1	0.3944	0.0314	0.0029	0.3941	0.0136
	2	0.4023	0.0352	0.0030	0.4022	0.0137
	3	0.3936	0.0191	0.0028	0.3935	0.0150
25	1	0.4076	0.0282	0.0031	0.4072	0.0146
	2	0.4138	0.0271	0.0039	0.4135	0.0138
	3	0.3898	0.0159	0.0025	0.3899	0.0154

Table 3.1 presents the simulation results for the binomial case with $\mu_1 = 0.4$. The estimates of finite population proportions are very similar for two methods in each scenario. The differences are quite little for the three kinds of τ in both methods. Scenario 3 has smaller standard deviations than those of scenario 1 and 2 in method 1. The increase of standard deviation from scenario 3 to scenario 2 are 28%, 84% and 70% for $\tau = 5$, $\tau = 10$ and $\tau = 25$, respectively. But method 2 has similar standard deviations for scenario 1,2 and 3. Moreover, the standard deviations of method 2 are smaller than those of method 1. The numerical standard errors (NSE) are not much different in method 1.

We present the five cell multinomial case with $\mu_1 = \mu_2 = 0.1$, $\mu_3 = \mu_4 = 0.2$, $\mu_5 = 0.4$ and $\tau = 25$ in Table 3.2. The estimates of finite proportions are very similar for two methods again. The standard deviations of scenario 2 are smaller than those of scenario 1 and 3 in

Table 3.2 Comparison of the posterior distributions of finite population proportions for five cell multinomial model by scenarios with $\mu_1 = \mu_2 = 0.1$, $\mu_3 = \mu_4 = 0.2$, $\tau = 25$

Method	Scenario	Quantity	P_1	P_2	P_3	P_4	P_5
1	1	Mean	0.0973	0.0938	0.2626	0.1721	0.3740
		STD	0.0210	0.0232	0.0249	0.0251	0.0313
		NSE	0.0025	0.0023	0.0029	0.003	0.0041
	2	Mean	0.1002	0.1112	0.2077	0.2223	0.3584
		STD	0.0099	0.0093	0.0122	0.0138	0.0161
		NSE	0.0020	0.0021	0.0022	0.0025	0.0026
	3	Mean	0.1072	0.0854	0.1880	0.2224	0.3968
		STD	0.0110	0.0097	0.0147	0.0151	0.0168
		NSE	0.0011	0.0009	0.0014	0.0015	0.0016
2	1	Mean	0.0976	0.0943	0.2627	0.1721	0.3730
		STD	0.0068	0.0073	0.0083	0.0099	0.0162
	2	Mean	0.1012	0.1114	0.2067	0.2233	0.3573
		STD	0.0072	0.0067	0.0104	0.0105	0.0142
	3	Mean	0.1071	0.0857	0.1885	0.2221	0.3964
		STD	0.0064	0.0072	0.0101	0.0104	0.0151

method 1. And the standard deviations of method 2 are smaller than those of method 1 in all scenarios. The P_2 and P_3 of scenario 3 seem to be underestimated which is compared with scenario 1 and 2 in both methods again.

According to Tables 3.1 and 3.2, two methods are very similar in the sense of point estimates of finite population proportions. But the standard deviations of method 1 are smaller than those of method 2.

We choose every fifth value thereafter to obtain 2,000 iterates. To check the consistency of computation, we change the choosing number in iteration. We choose every ℓth value in iteration, $\ell = 5, 10$ and 20. In Table 3.3 we present the posterior distributions of finite population proportions for five cells model by ℓ with $\mu_1 = \mu_2 = 0.1$, $\mu_3 = \mu_4 = 0.2$, $\tau = 10$, under the scenario 1. The posterior means are not change by different ℓ. But the standard deviations and numerical standard errors are quite different in both methods. Specially, NSE of estimation tends to increase when ℓ become large.

4. Discussion

In this paper, we studied Bayesian estimates for finite population proportions in multinomial data. And we implemented 3 scenarios using two simulation data. Finally, we checked the consistency for our algorithm by choosing different ℓth value.

It appears that the point estimates of finite population proportions do not differ between method 1 and 2. But the standard deviations of method 2 is smaller than those of method 1. And our computation does not have any fluctuation.

Our computation has several valuable results. First, we compute the estimates of finite population proportions using grid method. Our algorithm doesn't need to use Metropolis-Hastings (Chib and Greenberg, 1995), although the density of parameter doesn't have any closed form. This is a quite useful way to draw parameters because we can avoid finding out some candidate densities and calculate the accepting probabilities. Second, the algorithm can be processed for any number of I (≥ 1) clusters and J (≥ 2) columns. Third, we show

Table 3.3 Comparison of the posterior distributions of finite population proportions for five cell multinomial model by $\ell = 5, 10, 20$ with $\mu_1 = \mu_2 = 0.1$, $\mu_3 = \mu_4 = 0.2$, $\tau = 10$, under the scenario 1

Method	Quantity	ℓ	P_1	P_2	P_3	P_4	P_5
1	Mean	5	0.0973	0.0938	0.2626	0.1721	0.3740
		10	0.0973	0.0947	0.2577	0.1724	0.3785
		20	0.0979	0.0953	0.2507	0.1726	0.3748
	STD	5	0.0210	0.0228	0.0251	0.0252	0.031
		10	0.0231	0.0249	0.0301	0.0291	0.037
		20	0.0233	0.0241	0.0299	0.0310	0.039
	NSE	5	0.0025	0.0023	0.0029	0.0030	0.0041
		10	0.0033	0.0036	0.0042	0.0041	0.0056
		20	0.0046	0.0048	0.0061	0.0063	0.0079
2	Mean	5	0.0974	0.0937	0.2627	0.1725	0.3735
		10	0.0963	0.0951	0.2579	0.1722	0.3789
		20	0.0978	0.0963	0.2507	0.1737	0.3728
	STD	5	0.0069	0.0072	0.0092	0.0102	0.0152
		10	0.0068	0.0072	0.0093	0.0103	0.0151
		20	0.0065	0.0072	0.0093	0.0106	0.0150

that method 2 is better than method 1 because method 2 has small standard deviations. Finally, when we estimate the finite population proportions, we can reduce the time for computation using method 2. When we use method 1, we should compute every P_j in each iterate. But in method 2, we just compute p_{ij} in each iterate.

References

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistocian*, **49**, 327-335.

Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika*, **59**, 581 -589.

Leonard, T. and Hsu, J. S. J. (1994). The Bayesian analysis of categorical data - A selective review. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, edited by Smith, A. F. M. and Freeman, P., New York, Wiley, 283-310.

Lindley, D. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society B*, **34**, 1-41.

Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, **61**, 97-12.

Novick, M. R., Lewis, C. and Jackson, P. H. (1973). The estimation of proportions in m groups. *Psychometrika*, **38**, 19-46.