

2011-12-19

a Bayesian test of independence of two categorical variables obtained from a small area : an application to BMD and BMI

jingran zhou

Worcester Polytechnic Institute

Follow this and additional works at: <https://digitalcommons.wpi.edu/etd-theses>

Repository Citation

zhou, jingran, "a Bayesian test of independence of two categorical variables obtained from a small area : an application to BMD and BMI" (2011). *Masters Theses (All Theses, All Years)*. 1129.
<https://digitalcommons.wpi.edu/etd-theses/1129>

This thesis is brought to you for free and open access by [Digital WPI](#). It has been accepted for inclusion in Masters Theses (All Theses, All Years) by an authorized administrator of Digital WPI. For more information, please contact wpi-etd@wpi.edu.

A Bayesian Test of Independence of Two Categorical Variables Obtained

From a Small Area: An Application to BMD and BMI

by

Jingran Zhou

A Thesis

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

December 2011

APPROVED:

Dr. Balgobin Nandram, Thesis Advisor

Abstract

Scientists usually need to understand the extent of the association of two attributes, and the data are typically presented in two-way categorical tables. In science, the chi-squared test is routinely used to analyze data from such tables. However, in many applications the chi-squared test can be defective. For example, when the sample size is small, the chi-squared test may not be applicable.

The terms “small area” and “local area” are commonly used to denote a small geographical area, such as a county. If a survey has been carried out, the sample size within any particular small area may be too small to generate accurate estimates from the data, and a chi-squared test may be invalid (i.e., expected frequencies in some cells of the table are less than five).

To deal with this problem we use Bayesian small area estimation. Because it is used to “borrow strength” from related or similar areas. It enhances the information of each area with common exchangeable information. We use a Bayesian model to estimate a Bayes factor to test the independence of the two variables.

We apply the model to test for the independence between bone mineral density (BMD) and body mass index (BMI) from 31 counties and we compare the results with a direct Bayes factor test. We have also obtained numerical and sampling errors; both the numerical and sampling errors of our Bayes factor are small. Our model is shown to be much less sensitive to the specification of the prior distribution than the direct Bayes factor test which is based on each area only.

Acknowledgement

I am heartily thankful to my supervisor Dr. Balgobin Nandram for everything that he has done for me. Without his great guidance, support, patience and understanding this thesis would not have been possible. It has been rewarding and challenging to work with such a remarkable statistician and to catch up with his endless ideas.

I owe my deepest gratitude to Dr. Joseph Petruccelli. Working for him as a TA, I gained a deeper understanding of how to motivate statistics concepts to beginners. His elated attitude towards statistics and deep understanding has always been encouraging.

Thanks are due to Dr. Dhiman Bhadra and Dilli Bhatta for their encouragement and assistance throughout the thesis and Liu Pan for providing such a beautiful and functional LATEX template.

I am also grateful to Dr. Higgins Huong for showing me the details of how to begin writing a paper.

A special thank you to my parents and my friends for their enduring love and support. Their presence has encouraged me to complete this Master's thesis.

Contents

I	Introduction	1
II	Bayesian Methodology	7
2.1	General Hypothesis	7
2.2	Hypothesis of Independent	10
2.3	Constructing Approximation Of Posterior Density	15
2.4	Errors	17
III	Data Analysis	19
3.1	Numerical Study	19
3.2	Sampling Error	21
3.3	Sensitivity Analysis	22
IV	Conclusion	30
V	Appendix	33

Part I

Introduction

In this article, we deal with the problem to test the independence of two attributes for small areas in which a chi-squared test is not applicable (i.e., when expected frequencies in some cells of the table are less than five).

“Small area” may also refer to a “small domain”, i.e. a particular demographic group within an area, in which the data may not be sufficient to deduce accurate estimation. We build a model to deal with the deficient information and test for independence of the variables.

We apply our model to the BMD and BMI data from the third National Health and Nutrition Examination Survey (NHANES III). The data is shown in Table 1, where each row represents a county. Data in each county are categorized into three levels of BMD and four levels of BMI (i.e., there are thirty one 3×4 categorical tables).

Traditionally, there are two ways to test for association in such a categorical table. First, we can use the well-known Pearson chi-squared statistic. We use SAS to obtain the chi-squared tests for each area. The chi-squared test results and the description of the data are in Table 2; we get 6 very strong significant P-values (< 0.01). However, all tables have larger than 30 percent of the cells with expected counts less than 5 implying that chi-squared test does not work well for all tables.

The second way is to do the direct Bayesian test (Kass and Raftery, 1995), an alternative test to chi-squared test. One defect about it is that it is sensitive to the prior specifications, especially when there are not enough data to estimate the parameters under the test (e.g., Sinharay and Stern, 2002). We will show the disadvantages of the direct Bayesian test in part III.

The Bayes factor is used to quantify the difference between a model with association and one without. It is a very useful tool to test for association between two categorical variables (Kass R.E., 1993). Suppose we build two models, M_0 and M_1 , for data \underline{n} . The Bayes factor for comparing models M_0 and M_1 is defined as the ratio of the marginal densities of the data \underline{n} under the two models as

$$B_{10} = \frac{p(\underline{n} | M_1)}{p(\underline{n} | M_0)}$$

with

$$p(\underline{n} | M_k) = \int p(\underline{n} | \underline{\theta}_k, M_k) p(\underline{\theta}_k | M_k) d\underline{\theta}_k, k = 0, 1,$$

where $\underline{\theta}_k$ is the parameter vector under M_k , $p(\underline{n} | \underline{\theta}_k, M_k)$ is the likelihood function and $p(\underline{\theta}_k | M_k)$ is the prior density. The Bayes factor summarizes the evidence provided by the data in favor of one scientific hypothesis M_1 relative to another M_0 . Kass and Raftery(1995) gives a comprehensive description of Bayes factors and its interpretation. For example, if $0 \leq \log(B_{10}) < 1$, the evidence against M_0 is “not worth more than a bare mention”; if $1 \leq \log(B_{10}) < 3$, the evidence against M_0 is “positive”; if $3 \leq \log(B_{10}) < 5$, the evidence against M_0 is “strong”; and if $\log(B_{10}) \geq 5$, the evidence against M_0 is “very strong” (log is the “natural” logarithm). In our work the Bayes factor is based on two models, one with association and the other with no association. In one example M_0 is a model that specifies no association between the two categorical variables and M_1 is a model that has no restriction (i.e., a model with association).

TABLE 1: BMD and BMI data (3×4 tables) for 31 counties from NHANES III

County	BMD	BMI											
		0			1			2			3		
		0	1	2	0	1	2	0	1	2	0	1	2
1		5	0	0	11	8	4	12	9	1	27	1	0
3		1	2	1	15	5	4	15	1	1	18	2	0
7		2	3	2	17	7	3	18	6	1	10	3	1
13		5	5	0	15	6	3	15	6	2	13	1	0
17		5	1	2	16	11	3	4	7	4	8	5	0
19		2	1	1	18	6	3	17	9	1	32	3	0
25		2	2	1	15	7	4	35	7	2	26	5	0
29		4	3	1	39	15	3	29	15	3	38	11	0
31		12	10	2	45	16	4	59	19	5	80	10	4
33		7	2	1	18	17	4	11	12	0	10	8	0
35		4	1	1	15	11	2	22	4	1	18	2	0
37		11	7	2	79	32	8	108	36	6	98	20	1
45		1	1	2	16	6	1	5	6	0	20	3	1
47		7	0	1	22	3	0	28	6	3	31	3	0
59		5	6	4	34	21	8	33	9	2	27	4	0
61		10	4	2	40	17	7	43	15	0	56	8	0
71		1	1	1	19	8	2	5	4	1	15	2	0
73		2	1	1	24	8	1	24	6	1	23	4	0
81		2	6	2	10	9	3	12	5	3	16	5	0
85		2	2	2	14	9	2	15	7	1	17	7	1
99		0	0	1	16	5	8	8	7	1	16	3	0
101		1	2	2	6	8	4	18	4	0	13	1	0
111		1	4	3	29	8	5	20	12	3	24	6	1
113		1	3	1	13	6	3	11	2	1	22	1	1
119		2	2	0	13	7	1	9	1	1	13	4	1
125		5	5	0	18	3	2	11	6	0	5	4	0
141		4	2	0	9	7	2	23	11	2	27	7	0
163		4	3	0	23	3	2	14	5	1	35	3	0
189		1	1	2	18	6	5	17	6	4	13	3	0
201		2	0	0	24	7	4	31	3	2	39	3	0
439		6	2	3	26	7	1	21	4	0	24	0	0
total													

Note: BMI: 0($\leq 20\text{kg}/\text{m}^2$; underweight), 1($>20\text{kg}/\text{m}^2, \leq 25\text{kg}/\text{m}^2$; optimal),

2($>25\text{kg}/\text{m}^2, \leq 30\text{kg}/\text{m}^2$; overweight), 3($>30\text{kg}/\text{m}^2$; obese);

BMD: 0($>0.82\text{mg}/\text{cm}^2$; normal), 1($>0.64, \leq 0.82\text{mg}/\text{cm}^2$; osteopenia),

2($\leq 0.64\text{mg}/\text{cm}^2$; osteoporosis).

TABLE 2: Results of the chi-squared test for the 31 categorical tables

County	P-value	Sample-size	Num. of 0 cells	Percentage
1	0.001	78	3	50
3	0.0549	65	1	75
7	0.3959	73	0	58
13	0.1782	71	2	50
17	0.1916	66	1	58
19	0.0343	93	1	50
25	0.0800	106	1	50
29	0.3873	161	1	42
31	0.0218	266	0	33
33	0.3335	90	2	42
35	0.0788	81	1	58
37	0.0326	408	0	33
45	0.0020	62	1	58
47	0.2392	104	2	67
59	0.0017	153	1	33
61	0.0022	202	2	42
71	0.1912	59	1	67
73	0.2164	95	1	50
81	0.0798	73	1	42
85	0.2870	79	0	50
99	0.0061	65	3	67
101	0.0008	58	2	67
111	0.0133	116	0	42
113	0.0302	65	0	75
119	0.7059	54	1	67
125	0.2077	59	3	50
141	0.3245	94	2	50
163	0.0919	93	2	58
189	0.2570	76	1	58
201	0.1310	115	3	75
439	0.0009	94	3	67

Note: Percentage refers to the percentage of the cells with expected counts less than

5.

NHANES is a periodic survey used to assess an aspect of health of the U.S. population. Details of the NHANES III sample design are available. The data from NHANES III were collected October 1998 to September 1994 from mobile examination centers (MECs) set up across the U.S.

We note that BMD is tested by dual energy X-ray absorptiometry method (DXA) and the DXA systems employ two X-ray beams with different levels of energy. After subtraction of soft-tissue absorption, the absorption of each beam by the bone is used

to calculate the BMD value. Because of its ability to accurately eliminate the soft-tissue absorption factor, DXA can measure BMD at central skeletal sites: the lumbar spine and the hip including the femoral neck, Ward's triangle, and greater trochanter. DXA measurements at peripheral body sites can be performed using small, portable devices, in the physicians' offices or in mobile centers, with little or no discomfort or inconvenience to the patient. DXA is the most widely used BMD measurement technology with high accuracy.

The World Health Organization (WHO) has defined three levels for BMD (a) normal: BMD less than 1 standard deviation (SD) below the young non-hispanic white (NHW) adult mean, (b) osteopenia: BMD from 1 to 2.5 SD below the young NHW adult mean, and (c) osteoporosis: BMD more than 2.5 SD below the young NHW adult mean, see Looker, Orwoll, Johnston, Lindsay, Wahner, Dunn Calvo and Harries (1998) defined the cutoff values of BMD in NHANES III based on the WHO criteria as normal(1: BMD greater than $0.82\text{mg}/\text{cm}^2$); osteopenia (2: BMD between $0.64\text{mg}/\text{cm}^2$ and $0.82\text{mg}/\text{cm}^2$) and osteoporosis(3: BMD less than $0.64\text{mg}/\text{cm}^2$). BMD is used to diagnose osteoporosis, a disease of elderly females, and in NHANES III it is measured for individuals at least twenty years old (i.e., we use the data on white females only with chronic conditions older than twenty years).

BMI is widely used as a measure of obesity. More easily obtained than skin fold thickness measurements, BMI is safe, simple, and inexpensive to obtain, and widely used to characterize childhood obesity in large-scale epidemiologic studies (Pietrobelli et al., 1998). As noted by Squires (2001), "Health care costs for overweight and obesity total an estimated 117 billion dollar annually."

The sample in NHANES III was selected from households across the United States during the period October 1988 through September 1994, but for confidentiality reasons, the final data set for this study uses only the 31 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). Although these counties represent only about 1% of all U.S. counties, they account for approximately 50% of the U.S. population.

BMI is calculated as body weight in kilograms divided by $[(\text{height in meters})^2]$ (kg/m^2). The current value settings are as follows: a BMI of 20 to 25 may indicate optimal weight; a BMI lower than 20 suggests the person is underweight while a number above 25 may indicate the person is overweight; a person may have a BMI below 20 due to disease; a number above 30 suggests the person is obese.

The rest of the thesis is organized as follows. In part II we describe the methodology with two Bayesian models and we show how to compute the Bayes factor. This is done by using importance sampling. We use a method described by Nandram (1998) to obtain the proposal density which is needed in the importance sampling. We will show how to obtain the numerical and the sampling errors. In part III we apply our method to BMI and BMD data. Finally, in part IV we will give the conclusion. We also briefly discuss an application on the Kauffman Firm Survey in Appendix C.

Part II

Bayesian Methodology

We construct two models, one is under the null hypothesis and the other one is under alternative hypothesis. Under the null hypothesis, the two categorical variables are independent. Let p_{sij} denote the probability that a unit falls in the i^{th} row and j^{th} column of the s^{th} area (i.e., table). Here, $\sum_i \sum_j p_{sij} = 1, s = 1, 2, \dots, S$. The hypothesis of association has no further restriction on p_{sij} (i.e., the general hypothesis). The hypothesis of independence has $p_{sij} = p_{si} \times q_{sj}$, where $\sum_{i=1}^I p_{si} = 1, \sum_{j=1}^J q_{sj} = 1$ for each area. We describe the methodology for general $r \times c$ tables.

2.1 General Hypothesis

Let I be the number of cells in each table (i.e., $I = rc$). Let S be the number of tables we look at. Let n_{si} be the cell counts, p_{si} the corresponding cell probabilities, $s=1, 2, \dots, S, i=1, 2, \dots, I, n_{s\cdot} = \sum_{i=1}^I n_{si}, \underline{p}_s = (p_{s1}, p_{s2}, \dots, p_{sI})', \sum_{i=1}^I p_{si} = 1, \underline{n}_s = (n_{s1}, n_{s2}, \dots, n_{sI})'$.

We assume that

$$\underline{n}_s | \underline{p}_s \stackrel{i.i.d}{\sim} \text{Multinomial}(\underline{n}_s, \underline{p}_s), s = 1, 2, \dots, S, \quad (1)$$

$$\underline{p}_s | \underline{\mu}, \tau \stackrel{i.i.d}{\sim} \text{Dirichlet}(\underline{\mu}\tau), s = 1, 2, \dots, S, \quad (2)$$

where

$$f(\underline{n}_s | \underline{p}_s) = \underline{n}_s! \prod_{i=1}^I \frac{p_{si}^{n_{si}}}{n_{si}!}, i = 1, 2, \dots, I,$$

$$f(\underline{p}_s | \underline{\mu}, \tau) = \prod_{i=1}^I p_{si}^{\mu_i\tau-1} \frac{\Gamma(\tau)}{\prod_{i=1}^I \Gamma(\mu_i\tau)}, 0 < p_{si} < 1,$$

$$\sum_{i=1}^I p_{si} = 1 \text{ and } \underline{\mu} = (\mu_1, \mu_2, \dots, \mu_I)'$$

Without any prior information about $\underline{\mu}$ and τ we take $\underline{\mu} \sim \text{Dirichlet}(\underline{1})$, $i=1,2,\dots,I$, corresponding to a uniform prior on $\underline{\mu}, \tau$, $f(\underline{\mu}, \tau) = \frac{(I-1)!}{(1+\tau)^2}$, where $0 < \mu_i < 1$, $\sum_{i=1}^I \mu_i = 1$, $\tau > 0$. Here, the normalization constant is unity. That is, $k \int \int \frac{1}{(1+\tau)^2} d\underline{\mu} d\tau = 1$ we get $k \int 1 d\underline{\mu} = k D(\underline{1}) \int \frac{\prod_{i=1}^I \mu_i^{1-1}}{D(\underline{1})} = 1$, $k = \frac{1}{D(\underline{1})} = (I-1)!$. Assuming $\underline{\mu}$ and τ are independent, the joint prior distribution of $\underline{\mu}$ and τ is

$$f(\underline{\mu}, \tau) = \frac{(I-1)!}{(1+\tau)^2}. \quad (3)$$

Combining (1), (2) and (3) we get the joint density function:

$$\begin{aligned} f(\underline{n}, \underline{p}, \underline{\mu}, \tau) &= \left\{ \prod_{s=1}^S f(\underline{n}_s | \underline{p}_s) f(\underline{p}_s | \underline{\mu}, \tau) \right\} f(\underline{\mu}, \tau) \\ &= \frac{(I-1)!}{(1+\tau)^2} \prod_{s=1}^S \left\{ \underline{n}_s! \prod_{i=1}^I \frac{p_{si}^{n_{si}}}{n_{si}!} \prod_{i=1}^I p_{si}^{\mu_i \tau - 1} \frac{\Gamma(\tau)}{\prod_{i=1}^I \Gamma(\mu_i \tau)} \right\}. \end{aligned}$$

Integrating out \underline{p} from the joint density function $f(\underline{n}, \underline{p}, \underline{\mu}, \tau)$, we get the marginal joint density function for $\underline{n}, \underline{\mu}, \tau$

$$f(\underline{n}, \underline{\mu}, \tau) = \frac{(I-1)!}{(1+\tau)^2} \prod_{s=1}^S \left\{ \underline{n}_s! \left(\prod_{i=1}^I \frac{1}{n_{si}!} \left(\frac{\Gamma(n_{si} + \mu_i \tau)}{\Gamma(\mu_i \tau)} \right) \right) / \left(\frac{\Gamma(n_s + \tau)}{\Gamma(\tau)} \right) \right\}. \quad (4)$$

To obtain the Bayes factor, we need the marginal likelihood,

$$f(\underline{n}_s) = \int \int f(\underline{n}_s | \underline{\mu}, \tau) p(\underline{\mu}, \tau) d\underline{\mu} d\tau, s = 1, 2, \dots, S.$$

This integration is difficult to perform mathematically. To get an approximation with good accuracy we use the joint posterior density which is

$$f(\underline{\mu}, \tau | \underline{n}) \propto \frac{(I-1)!}{(1+\tau)^2} \prod_{s=1}^S \left\{ \left(\prod_{i=1}^I \left(\frac{\Gamma(n_{si} + \mu_i \tau)}{\Gamma(\mu_i \tau)} \right) \right) / \left(\frac{\Gamma(n_s + \tau)}{\Gamma(\tau)} \right) \right\}. \quad (5)$$

This posterior density does not exist in closed form because the constant of proportionality is intractable. We use the product of two simple distribution to approximate the posterior distribution, as

$$P_a(\underline{\mu}, \tau | \underline{n}) = f(\underline{\mu} | \tau, \underline{n}) f(\tau | \hat{\underline{\mu}}, \underline{n}).$$

We use a method from Nandram (1998) to set the two components $f(\underline{\mu} | \tau, \underline{n})$ and $f(\tau | \underline{\mu}, \underline{n})$. This process will be showed in section 2.3.

Now, $\int \int p(\underline{\mu}, \tau) d\underline{\mu} d\tau = 1$, and $f(\underline{n}_s) = \frac{f(\underline{n}_s)}{\int \int p(\underline{\mu}, \tau) d\underline{\mu} d\tau}$. Thus,

$$f(\underline{n}_s) = \frac{\int \int \frac{f(\underline{n}_s, \underline{\mu}, \tau) p(\underline{\mu}, \tau)}{P_a(\underline{\mu}, \tau | \underline{n})} P_a(\underline{\mu}, \tau | \underline{n}) d\underline{\mu} d\tau}{\int \int \frac{p(\underline{\mu}, \tau)}{P_a(\underline{\mu}, \tau | \underline{n})} P_a(\underline{\mu}, \tau | \underline{n}) d\underline{\mu} d\tau}. \quad (6)$$

By SLLN (Strong Law of Large Numbers), as M goes to infinity, $\lim_{n \rightarrow \infty} \frac{1}{M} \sum_{h=1}^M f(x) = \int f(x) p(x) dx$. Thus, an estimator of $f(\underline{n}_s)$ is

$$\widehat{f(\underline{n}_s)} = \left\{ \frac{1}{M} \sum_{h=1}^M f(\underline{n}_s | \underline{\mu}^{(h)}, \tau^{(h)}) \times \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} \right\} / \left\{ \frac{1}{M} \sum_{h=1}^M \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} \right\} \quad (7)$$

and in a simple form

$$\widehat{f(\underline{n}_s)} = \sum_{h=1}^M w_h f(\underline{n}_s | \underline{\mu}^{(h)}, \tau^{(h)}), \quad (8)$$

where,

$$w_h = \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} / \left\{ \sum_{h=1}^M \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} \right\}.$$

However, by performing the above model using the BMI and BMD data, we notice that the τ is so big that when the "small tables" are adjusted by common exchangeable $\underline{\mu}$ and τ , they are dominated by the common information and all the results for the areas are similar. The adjusted joint density for the s^{th} area is $D(n_s + \underline{\mu}\tau)$. If n_s is small, while τ is very big, the $\underline{\mu}$ and τ will kill the original n_s .

To fix that problem, we change part of the estimators. First, we get an approximation for $f(\underline{p}_s)$ and then the estimated marginal likelihood is $\widehat{f(\underline{n}_s)}$ which based on the approximated $f(\underline{p}_s)$ instead of using the joint distribution of $\underline{\mu}$ and τ . Since \underline{p} has a range from 0 to 1, the chance for the "dominating" case is much reduced. The adjusted process is shown below,

The marginal likelihood density is

$$f(\underline{n}_s) = \int \int f(\underline{n}_s | \underline{p}_s) f(\underline{p}_s) d\underline{p}_s, \quad (9)$$

where

$$f(\underline{p}_s) = \int \int f(\underline{p}_s | \underline{\mu}, \tau) p(\underline{\mu}, \tau) d\underline{\mu} d\tau. \quad (10)$$

Again, because $\int \int p(\underline{\mu}, \tau) d\underline{\mu} d\tau = 1$, $f(\underline{p}_s) = \frac{f(\underline{p}_s)}{\int \int p(\underline{\mu}, \tau) d\underline{\mu} d\tau}$,

$$f(\underline{p}_s) = \frac{\int \int \frac{f(\underline{p}_s | \underline{\mu}, \tau) p(\underline{\mu}, \tau)}{P_a(\underline{\mu}, \tau | \underline{n})} P_a(\underline{\mu}, \tau | \underline{n}) d\underline{\mu} d\tau}{\int \int \frac{p(\underline{\mu}, \tau)}{P_a(\underline{\mu}, \tau | \underline{n})} P_a(\underline{\mu}, \tau | \underline{n}) d\underline{\mu} d\tau}. \quad (11)$$

As the probability density function of \underline{p}_s does not depend on s , so for all the areas, \underline{p} have the same distribution, $f(\underline{p}_s) = f(\underline{p})$, $s = 1, 2, \dots, S$.

By SLLN,

$$\widehat{f(\underline{p})} = \left\{ \frac{1}{M} \sum_{h=1}^M f(\underline{p} | \underline{\mu}^{(h)}, \tau^{(h)}) \times \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} \right\} / \left\{ \frac{1}{M} \sum_{h=1}^M \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} \right\}. \quad (12)$$

and more simply

$$\widehat{f(\underline{p})} = \sum_{h=1}^M w_h f(\underline{p} | \underline{\mu}^{(h)}, \tau^{(h)}), \quad (13)$$

where,

$$w_h = \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} / \left\{ \sum_{h=1}^M \frac{p(\underline{\mu}^{(h)}, \tau^{(h)})}{P_a(\underline{\mu}^{(h)}, \tau^{(h)} | \underline{n})} \right\}.$$

By SLLN, an estimator of the marginal likelihood density $f(\underline{n}_s)$ is

$$\widehat{f(\underline{n}_s)} = \frac{1}{M} \sum_{h=1}^M f(\underline{n}_s | \underline{p}^{(h)}), \quad (14)$$

where $\underline{p}^{(h)}$ is from (12). Therefore, we get the general marginal likelihood density.

The next step is to get the independent marginal likelihood density.

2.2 Hypothesis of Independent

For a $r \times c$ table, Let J denote the number of rows and K the number of columns.

Letting n_{sjk} be the cell count in the j^{th} row and k^{th} column in the s^{th} table, p_{sjk}

the corresponding cell probabilities, $s=1, 2, \dots, S$, $j=1,2,\dots,J$, $k=1,2,\dots,K$. Let $n_s =$

$$\sum_{j=1}^J \sum_{k=1}^K n_{sjk}, \underline{p}_{\cdot s} = (p_{s1}, p_{s2}, \dots, p_{sJ})', \underline{q}_{\cdot s} = (q_{s1}, q_{s2}, \dots, q_{sK})', \sum_{j=1}^J \sum_{k=1}^K p_{sj} q_{sk} = 1, \underline{n}_s = (n_{s11}, n_{s12}, \dots, n_{sJK})', p_{sj} = \sum_{k=1}^K p_{sjk}, q_{sk} = \sum_{j=1}^J p_{sjk}, n_{sj} = \sum_{k=1}^K n_{sjk}, n_{s.k} = \sum_{j=1}^J n_{sjk}$$

$$\underline{n}_s | \underline{p}_{\cdot s}, \underline{q}_{\cdot s} \sim \text{multinomial}(\underline{n}_s, \underline{p}_{\cdot s}, \underline{q}_{\cdot s}), s = 1, 2, \dots, S, \quad (15)$$

$$\underline{p}_{\cdot s} | \underline{\mu}_{\cdot 1}, \tau_1 \sim \text{dirichlet}(\underline{\mu}_{\cdot 1}, \tau_1), s = 1, 2, \dots, S, \quad (16)$$

$$\underline{q}_{\cdot s} | \underline{\mu}_{\cdot 2}, \tau_2 \sim \text{dirichlet}(\underline{\mu}_{\cdot 2}, \tau_2), s = 1, 2, \dots, S, \quad (17)$$

where

$$f(\underline{n}_s | \underline{p}_{\cdot s}, \underline{q}_{\cdot s}) = \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{(p_{sj} q_{sk})^{n_{sjk}}}{n_{sjk}!}, 0 < n_{sjk} < n_s, j = 1, 2, \dots, J, k = 1, 2, \dots, K,$$

$$f(\underline{p}_{\cdot s} | \underline{\mu}_{\cdot 1}, \tau_1) = \prod_{j=1}^J p_{sj}^{\mu_{1j} \tau_1 - 1} \frac{\Gamma(\tau_1)}{\sum_{j=1}^J \Gamma(\mu_{1j} \tau_1)}, 0 < p_{sj} < 1,$$

$$\sum_{j=1}^J p_{sj} = 1 \text{ and } \underline{\mu}_{\cdot 1} = (\mu_{11}, \mu_{12}, \dots, \mu_{1J}),$$

and

$$f(\underline{q}_{\cdot s} | \underline{\mu}_{\cdot 2}, \tau_2) = \prod_{k=1}^K p_{sk}^{\mu_{2k} \tau_2 - 1} \frac{\Gamma(\tau_2)}{\sum_{k=1}^K \Gamma(\mu_{2k} \tau_2)}, 0 < p_{sk} < 1,$$

$$\sum_{k=1}^K p_{sk} = 1 \text{ and } \underline{\mu}_{\cdot 2} = (\mu_{21}, \mu_{22}, \dots, \mu_{2K}).$$

Here without any prior information we also use uniform distribution as prior distribution for $\underline{\mu}_{\cdot 1}, \tau_1, \underline{\mu}_{\cdot 2}, \tau_2 : f(\underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2) \propto \frac{1}{(1+\tau_1)^2} \frac{1}{(1+\tau_2)^2}$, where $0 < \mu_{tj} < 1, \sum_{j=1}^J \mu_{tj} \tau_t = 1, \tau_t > 0, t = 1, 2$, because $\int \int \int \int f(\underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2) d\underline{\mu}_{\cdot 1} d\tau_1 d\underline{\mu}_{\cdot 2} d\tau_2 = 1$, we get

$$f(\underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2) = \frac{(J-1)! (K-1)!}{(1+\tau_1)^2 (1+\tau_2)^2}. \quad (18)$$

Combing (15), (16), (17), (18) we get the joint density function of $\underline{n}, \underline{p}, \underline{q}, \underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2$:

$$\begin{aligned} & f(\underline{n}, \underline{p}, \underline{q}, \underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2) \quad (19) \\ &= \prod_{s=1}^S f(\underline{n}_s | \underline{p}_{\cdot s}, \underline{q}_{\cdot s}, \underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2) \\ &= \left\{ \prod_{s=1}^S f(\underline{n}_s | \underline{p}_{\cdot s}, \underline{q}_{\cdot s}) f(\underline{p}_{\cdot s} | \underline{\mu}_{\cdot 1}, \tau_1), f(\underline{q}_{\cdot s} | \underline{\mu}_{\cdot 2}, \tau_2) \right\} f(\underline{\mu}_{\cdot 1}, \underline{\mu}_{\cdot 2}, \tau_1, \tau_2). \end{aligned}$$

So that

$$f(\underline{n}, \underline{p}, \underline{q}, \underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2) = \frac{J-1}{(1+\tau_1)^2} \frac{K-1}{(1+\tau_2)^2} \prod_{s=1}^S \left\{ \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{(p_{sj}q_{sk})^{n_{sjk}}}{n_{sjk}!} \right. \quad (20)$$

$$\left. \times \prod_{j=1}^J p_{sj}^{\mu_{1j}\tau_1-1} \frac{\Gamma(\tau_1)}{\prod_{j=1}^J \Gamma(\mu_{1j}\tau_1)} \prod_{k=1}^K p_{sk}^{\mu_{2k}\tau_2-1} \frac{\Gamma(\tau_2)}{\prod_{k=1}^K \Gamma(\mu_{2k}\tau_2)} \right\}.$$

Integrating out $\underline{p}, \underline{q}$ from the above joint density function, we get the joint density function of $\underline{n}, \underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2$:

$$f(\underline{n}, \underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2) = \prod_{s=1}^S \left\{ \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{1}{n_{sjk}!} \left(\prod_{j=1}^J \frac{\Gamma(n_{sj} + \mu_{1j}\tau_1)}{\mu_{1j}\tau_1} \right) / \left(\frac{\Gamma(n_s + \tau_1)}{\tau_1} \right) \right. \quad (21)$$

$$\left. \times \left(\prod_{k=1}^K \frac{\Gamma(n_{s,k} + \mu_{2k}\tau_2)}{\mu_{2k}\tau_2} \right) / \left(\frac{\Gamma(n_s + \tau_2)}{\tau_2} \right) \right\} \frac{(J-1)!}{(1+\tau_1)^2} \frac{(K-1)!}{(1+\tau_2)^2}.$$

The posterior function of $\underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2$ given \underline{n} :

$$f(\underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2 | \underline{n}) \propto \prod_{s=1}^S \left\{ \left(\prod_{j=1}^J \frac{\Gamma(n_{sj} + \mu_{1j}\tau_1)}{\mu_{1j}\tau_1} \right) / \left(\frac{\Gamma(n_s + \tau_1)}{\tau_1} \right) \right. \quad (22)$$

$$\left. \times \left(\prod_{k=1}^K \frac{\Gamma(n_{s,k} + \mu_{2k}\tau_2)}{\mu_{2k}\tau_2} \right) / \left(\frac{\Gamma(n_s + \tau_2)}{\tau_2} \right) \right\} \frac{(J-1)!}{(1+\tau_1)^2} \frac{(K-1)!}{(1+\tau_2)^2}.$$

For the next step, we need to get the marginal likelihood density $f(\underline{n}_s)$, which is

$$f(\underline{n}_s) = \int \int \int \int f(\underline{n}_s | \underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2) f_1(\underline{\mu}_1, \tau_1) f_2(\underline{\mu}_2, \tau_2) d\mu_1 d\mu_2 d\tau_1 d\tau_2. \quad (23)$$

Also, as the difficulty to perform this integration mathematically, we introduce $P_a(\underline{\mu}, \tau | \underline{n})$, which is to be determined later, to get an approximation.

As $\int \int p(\underline{\mu}_t, \tau_t) d\underline{\mu}_t d\tau_t = 1$ where $t=1,2$, so $f(\underline{n}_s) = \frac{f(\underline{n}_s)}{\int \int p(\underline{\mu}_t, \tau_t) d\underline{\mu}_t d\tau_t}$, thus,

$$f(\underline{n}_s) = \frac{\int \int \int \int \frac{f(\underline{n}_s | \underline{\mu}_1, \underline{\mu}_2, \tau_1, \tau_2) p_1(\underline{\mu}_1, \tau_1) p_2(\underline{\mu}_2, \tau_2)}{P_a(\underline{\mu}_1, \tau_1 | \underline{n}) P_a(\underline{\mu}_2, \tau_2 | \underline{n})} P_a(\underline{\mu}_1, \tau_1 | \underline{n}) P_a(\underline{\mu}_2, \tau_2 | \underline{n}) d\mu_1 d\mu_2 d\tau_1 d\tau_2}{\int \int \int \int \frac{p(\underline{\mu}_1, \tau_1)}{P_a(\underline{\mu}_1, \tau_1 | \underline{n})} P_a(\underline{\mu}_1, \tau_1 | \underline{n}) \frac{p(\underline{\mu}_2, \tau_2)}{P_a(\underline{\mu}_2, \tau_2 | \underline{n})} P_a(\underline{\mu}_2, \tau_2 | \underline{n}) d\mu_1 d\tau_1 d\mu_2 d\tau_2} \quad (24)$$

By SLLN, M goes to infinity, $\int f(x)p(x)dx = \frac{1}{M} \sum_{h=1}^M f(x)$, we get an estimator of the marginal likelihood density,

$$\widehat{f(\underline{n}_s)} = \frac{\sum_{h=1}^M f(\underline{n}_s | \underline{\mu}_1^{(h)}, \tau_1^{(h)}, \underline{\mu}_2^{(h)}, \tau_2^{(h)}) \times \frac{p(\underline{\mu}_1^{(h)}, \tau_1^{(h)})}{P_a(\underline{\mu}_1^{(h)}, \tau_1^{(h)} | \underline{n})} \frac{p(\underline{\mu}_2^{(h)}, \tau_2^{(h)})}{P_a(\underline{\mu}_2^{(h)}, \tau_2^{(h)} | \underline{n})}}{\sum_{h=1}^M \frac{p(\underline{\mu}_1^{(h)}, \tau_1^{(h)})}{P_a(\underline{\mu}_1^{(h)}, \tau_1^{(h)} | \underline{n})} \frac{p(\underline{\mu}_2^{(h)}, \tau_2^{(h)})}{P_a(\underline{\mu}_2^{(h)}, \tau_2^{(h)} | \underline{n})}} \quad (25)$$

Thus, we have

$$\widehat{f(\underline{n}_s)} = \sum_{h=1}^M w_h f(\underline{n}_s | \mu_1^{(h)}, \tau_1^{(h)}, \mu_2^{(h)}, \tau_2^{(h)}), \quad (26)$$

where

$$w_h = \left\{ \frac{p(\mu_1^{(h)}, \tau_1^{(h)})}{P_a(\mu_1^{(h)}, \tau_1^{(h)} | \underline{n})} \frac{p(\mu_2^{(h)}, \tau_2^{(h)})}{P_a(\mu_2^{(h)}, \tau_2^{(h)} | \underline{n})} \right\} / \left\{ \sum_{h=1}^M \frac{p(\mu_1^{(h)}, \tau_1^{(h)})}{P_a(\mu_1^{(h)}, \tau_1^{(h)} | \underline{n})} \frac{p(\mu_2^{(h)}, \tau_2^{(h)})}{P_a(\mu_2^{(h)}, \tau_2^{(h)} | \underline{n})} \right\}.$$

Theorem 1. *To make a uniform format in general hypothesis we can rewrite the function(24) as:*

$$f(\underline{n}_s) = c \sum_{h=1}^M w_h^{(1)} f(\underline{n}_{1s} | \mu_1, \tau_1) \sum_{h=1}^M w_h^{(2)} f(\underline{n}_{2s} | \mu_2, \tau_2), \quad (27)$$

where

$$c = \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{1}{n_{sjk}!} \frac{\prod_{j=1}^J n_{sj} \cdot \prod_{k=1}^K n_{s,k}}{n_s! n_s!},$$

$$w_h^{(1)} = \frac{p(\mu_1^{(h)}, \tau_1^{(h)})}{P_a(\mu_1^{(h)}, \tau_1^{(h)} | \underline{n})} / \sum_{h=1}^M \frac{p(\mu_1^{(h)}, \tau_1^{(h)})}{P_a(\mu_1^{(h)}, \tau_1^{(h)} | \underline{n})}, \underline{n}_{1s} = (n_{s1}, n_{s2}, \dots, n_{sJ}),$$

$$w_h^{(2)} = \frac{p(\mu_2^{(h)}, \tau_2^{(h)})}{P_a(\mu_2^{(h)}, \tau_2^{(h)} | \underline{n})} / \sum_{h=1}^M \frac{p(\mu_2^{(h)}, \tau_2^{(h)})}{P_a(\mu_2^{(h)}, \tau_2^{(h)} | \underline{n})}, \underline{n}_{2s} = (n_{s,1}, n_{s,2}, \dots, n_{s,K}).$$

Proof.

$$f(\underline{n}_s | \underline{p}_s, \underline{q}_s) = \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{(p_{sj} q_{sk})^{n_{sjk}}}{n_{sjk}!}$$

$$= \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{1}{n_{sjk}!} \frac{\prod_{j=1}^J n_{sj} \cdot \prod_{k=1}^K n_{s,k}}{n_s! n_s!} n_s! \prod_{j=1}^J \frac{(p_{sj})^{n_{sj}}}{n_{sj}!} n_s! \prod_{k=1}^K \frac{(q_{sk})^{n_{s,k}}}{n_{s,k}!}.$$

Letting $f(\underline{n}_{1s} | \underline{p}_s) = n_s! \prod_{j=1}^J \frac{(p_{sj})^{n_{sj}}}{n_{sj}!}$, $f(\underline{n}_{2s} | \underline{q}_s) = n_s! \prod_{k=1}^K \frac{(q_{sk})^{n_{s,k}}}{n_{s,k}!}$, we rewrite the function as

$$f(\underline{n}_s | \underline{p}_s, \underline{q}_s) = \left\{ \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{1}{n_{sjk}!} \frac{\prod_{j=1}^J n_{sj} \cdot \prod_{k=1}^K n_{s,k}}{n_s! n_s!} \right\} f(\underline{n}_{1s} | \underline{p}_s) f(\underline{n}_{2s} | \underline{q}_s).$$

This is just a combination of two general models. For each of them we do the same process as we did in the the general model, first we have

$$f(\underline{n}_{1s}, \underline{p}_s | \mu_1, \tau_1) = f(\underline{n}_{1s} | \underline{p}_s) f(\underline{p}_s | \mu_1, \tau_1) = n_s! \prod_{j=1}^J \frac{(p_{sj})^{n_{sj}}}{n_{sj}!} \prod_{j=1}^J p_{sj}^{\mu_{1j} \tau_1 - 1} \frac{\Gamma(\tau_1)}{\prod_{j=1}^J \Gamma(\mu_{1j} \tau_1)}$$

Then we integrate the function by \underline{p}_s , which yields

$$f(\underline{n}_{1s} | \underline{\mu}_1, \tau_1) = \underline{n}_s! \prod_{j=1}^J \frac{1}{n_{sj}!} \left(\prod_{j=1}^J \frac{\Gamma(n_{sj} + \mu_{1j}\tau_1)}{\mu_{1j}\tau_1} \right) / \left(\frac{\Gamma(n_s + \tau_1)}{\tau_1} \right) \frac{(J-1)!}{(1+\tau_1)^2}.$$

By the same process, we get

$$f(\underline{n}_{2s} | \underline{\mu}_2, \tau_2) = \underline{n}_s! \prod_{k=1}^K \frac{1}{n_{sk}!} \left(\prod_{k=1}^K \frac{\Gamma(n_{sk} + \mu_{2k}\tau_2)}{\mu_{2k}\tau_2} \right) / \left(\frac{\Gamma(n_s + \tau_2)}{\tau_2} \right) \frac{(K-1)!}{(1+\tau_2)^2}.$$

Thus function (24) is equal to:

$$c \int \int \frac{f(\underline{n}_{1s} | \underline{\mu}_1, \tau_1) p_1(\underline{\mu}_1, \tau_1) P_a(\underline{\mu}_1, \tau_1 | \underline{n})}{P_a(\underline{\mu}_1, \tau_1 | \underline{n})} d\underline{\mu}_1 d\tau_1 \int \int \frac{f(\underline{n}_{2s} | \underline{\mu}_2, \tau_2) p_2(\underline{\mu}_2, \tau_2) P_a(\underline{\mu}_2, \tau_2 | \underline{n})}{P_a(\underline{\mu}_2, \tau_2 | \underline{n})} d\underline{\mu}_2 d\tau_2 /$$

$$\left\{ \int \int \frac{p(\underline{\mu}_1, \tau_1) P_a(\underline{\mu}_1, \tau_1 | \underline{n})}{P_a(\underline{\mu}_1, \tau_1 | \underline{n})} d\underline{\mu}_1 d\tau_1 \frac{p(\underline{\mu}_2, \tau_2) P_a(\underline{\mu}_2, \tau_2 | \underline{n})}{P_a(\underline{\mu}_2, \tau_2 | \underline{n})} d\underline{\mu}_2 d\tau_2 \right\},$$

where, $c = \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{1}{n_{sjk}!} \frac{\prod_{j=1}^J n_{sj} \prod_{k=1}^K n_{sk}}{n_s! n_s!}$.

By SLLN, M goes to infinity, $\int f(x)p(x)dx = \frac{1}{M} \sum_{h=1}^M f(x)$

$$\widehat{f(\underline{n}_s)} = c \sum_{h=1}^M w_h^{(1)} f(\underline{n}_{1s}, | \underline{\mu}_1, \tau_1) \sum_{h=1}^M w_h^{(2)} f(\underline{n}_{2s}, | \underline{\mu}_2, \tau_2).$$

◇

Below is the adjusted process:

The marginal likelihood density

$$f(\underline{n}_s) = \int f(\underline{n}_s | \underline{p}_s, \underline{q}_s) f(\underline{p}_s) f(\underline{q}_s) d\underline{p}_s d\underline{q}_s, \quad (28)$$

$$f(\underline{p}_s) = f(\underline{p}_s | \underline{\mu}_1, \tau_1) f(\underline{\mu}_1, \tau_1) d\underline{\mu}_1 d\tau_1. \quad (29)$$

Because $\int \int p(\underline{\mu}_1, \tau_1) d\underline{\mu}_1 d\tau_1 = 1$, $f(\underline{p}_s) = \frac{f(\underline{p}_s)}{\int \int p(\underline{\mu}_1, \tau_1) d\underline{\mu}_1 d\tau_1}$,

$$f(\underline{p}_s) = \frac{\int \int \frac{f(\underline{p}_s | \underline{\mu}_1, \tau_1) p(\underline{\mu}_1, \tau_1)}{P_a(\underline{\mu}_1, \tau_1 | \underline{n})} P_a(\underline{\mu}_1, \tau_1 | \underline{n}) d\underline{\mu}_1 d\tau_1}{\int \int \frac{p(\underline{\mu}_1, \tau_1)}{P_a(\underline{\mu}_1, \tau_1 | \underline{n})} P_a(\underline{\mu}_1, \tau_1 | \underline{n}) d\underline{\mu}_1 d\tau_1}. \quad (30)$$

The probability density function $f(\underline{p}_s)$ does not depend on s, thus $f(\underline{p}_s) = f(\underline{p})$

By SLLN,

$$\widehat{f(\underline{p})} = \left\{ \frac{1}{M} \sum_{h=1}^M f(\underline{p} | \underline{\mu}_1^{(h)}, \tau_1^{(h)}) \times \frac{p(\underline{\mu}_1^{(h)}, \tau_1^{(h)})}{P_a(\underline{\mu}_1^{(h)}, \tau_1^{(h)} | \underline{n})} \right\} / \left\{ \frac{1}{M} \sum_{h=1}^M \frac{p(\underline{\mu}_1^{(h)}, \tau_1^{(h)})}{P_a(\underline{\mu}_1^{(h)}, \tau_1^{(h)} | \underline{n})} \right\}, \quad (31)$$

which is simply

$$\widehat{f(\underline{p})} = \sum_{h=1}^M w_{1h} f(\underline{p} \mid \mu_{\sim 1}^{(h)}, \tau_1^{(h)}), \quad (32)$$

where,

$$w_{1h} = \frac{p(\mu_{\sim 1}^{(h)}, \tau_1^{(h)})}{P_a(\mu_{\sim 1}^{(h)}, \tau_1^{(h)} \mid \underline{n})} \Big/ \left\{ \sum_{h=1}^M \frac{p(\mu_{\sim 1}^{(h)}, \tau_1^{(h)})}{P_a(\mu_{\sim 1}^{(h)}, \tau_1^{(h)} \mid \underline{n})} \right\}.$$

Perform the same process to $f(\underline{q}_s)$, we get

$$\widehat{f(\underline{q})} = \sum_{h=1}^M w_{2h} f(\underline{q} \mid \mu_{\sim 2}^{(h)}, \tau_2^{(h)}), \quad (33)$$

where,

$$w_{2h} = \frac{p(\mu_{\sim 2}^{(h)}, \tau_2^{(h)})}{P_a(\mu_{\sim 2}^{(h)}, \tau_2^{(h)} \mid \underline{n})} \Big/ \left\{ \sum_{h=1}^M \frac{p(\mu_{\sim 2}^{(h)}, \tau_2^{(h)})}{P_a(\mu_{\sim 2}^{(h)}, \tau_2^{(h)} \mid \underline{n})} \right\}.$$

Thus, by SLLN, the estimation of the marginal likelihood density is

$$\widehat{f(\underline{n}_s)} = \frac{1}{M} \sum_{h=1}^M f(\underline{n}_s \mid \underline{p}_s^{(h)}, \underline{q}_s^{(h)}). \quad (34)$$

It is easy to show that the above equation can also have the same format as in the general case

$$f(\underline{n}_s) = c \int f(\underline{n}_s \mid \underline{p}_s) f(\underline{p}_s) d\underline{p}_s f(\underline{n}_s \mid \underline{q}_s) f(\underline{q}_s) d\underline{q}_s, \quad (35)$$

where:

$$c = \underline{n}_s! \prod_{j=1}^J \prod_{k=1}^K \frac{1}{n_{sjk}!} \frac{\prod_{j=1}^J n_{sj} \cdot \prod_{k=1}^K n_{s.k}}{n_{s.!} n_{s.!}}.$$

2.3 Constructing Approximation Of Posterior Density

From Nandram (1998), a reasonable probability density function for the conditional posterior density of $\underline{\mu}$ given τ, \underline{n} , $(\underline{\mu} \mid \tau, \underline{n})$ is an approximately Dirichlet $(\mu_1^a \tau^a, \mu_2^a \tau^a, \dots, \mu_j^a \tau^a)$.

Also a reasonable probability density function for the conditional posterior density of τ given $\underline{n}, \hat{\underline{\mu}}$, $(\tau \mid \hat{\underline{\mu}}, \underline{n})$ is approximately $\Gamma(\eta^a, \nu^a)$. In

$$\underline{\mu} \mid \tau, \underline{n} \sim D(\mu_1^a \tau^a, \mu_2^a \tau^a, \dots, \mu_j^a \tau^a), \quad (36)$$

$\mu_j^a = \hat{\mu}_j, \tau^a = \{ \sum_{j=1}^J \sum_{j'=1}^j d_{jj'} I_{jj'} \} / \sum_{j=1}^J \sum_{j'=1}^j I_{jj'}$, in which:

$$E(\mu_j | \tau, n) = \frac{n_{.j+1}}{\sum_{s=1}^J (n_{.s+1})} = \underline{\theta}_j, \quad (37)$$

$$A_j = \tau \sum_{i=1}^I \ln(1 + n_{i,j} / \tau \underline{\theta}_j), \quad (38)$$

$$B_j = \tau \sum_{i=1}^I \{ \underline{\theta}_j^{-1} - (\underline{\theta}_j + n_{i,j} / \tau)^{-1} \}. \quad (39)$$

We are assuming $n_{ij} \geq 1$ for each j and at least one i . (i.e, $B^{-1} \neq 0$)

$$\bar{A} = \sum_{j=1}^J B_j^{-1} A_j / \sum_{j=1}^J B_j^{-1}, \quad (40)$$

$$v_j = B_j^{-1} / (\sum_{s=1}^J B_s^{-1})^{1/2}, \quad (41)$$

$$\hat{\mu}_j = \hat{\theta}_j + B_j^{-1} (A_j - \bar{A}), \quad (42)$$

$$\tilde{\mu}_j = \begin{cases} \hat{\mu}_j & 0 < \hat{\mu}_j < 1 \\ \hat{\theta}_j & \hat{\mu}_j \leq 0 \text{ or } \hat{\mu}_j \geq 1, \end{cases}$$

$$\hat{\mu}_j = \tilde{\mu} / \sum_{s=1}^J \tilde{\mu}_s,$$

$$d_{jj'} = \begin{cases} \mu_j^{a^2} / (B_j^{-1} - v_j^2), & j = j' \\ \mu_j^a \mu_{j'}^a / v_j v_{j'} - 1, & j \neq j', \end{cases}$$

where $I_{jj'} = 1$ if $d_{jj'} > 0$ and $I_{jj'} = 0$ if $d_{jj'} \leq 0$ see Appendix A. Then

$$\tau | \underline{\mu}, \tilde{n} \sim \Gamma(\eta^a, \nu^a) \quad (43)$$

$\eta^a = \{ \frac{\tau_*}{2\sigma_*} + \sqrt{(\frac{\tau_*}{2\sigma_*})^2 + 1} \}^2$, and $\nu^a = \sqrt{\eta^a} / \sigma_*$ in which: $\sigma_*^2 = [\sum_{i=1}^I \{ (\frac{1}{\tau_*} - \frac{1}{\tau_* + n_i}) + \sum_{j=1}^J \hat{\mu}_j^2 (\frac{1}{\hat{\mu}_j \tau_*} - \frac{1}{n_{i,j} + \hat{\mu}_j \tau_*}) \}]^{-1}$. τ_* is the maximum likelihood estimator of

$$\pi(\tau | \hat{\mu}, \tilde{n}) \propto \prod_{i=1}^I \{ \prod_{j=1}^J \frac{\Gamma(n_{i,j} + \hat{\mu}_j)}{\Gamma(\hat{\mu}_j \tau)} \} / \{ \Gamma(n_{i.} + \tau) / \Gamma(\tau) \} \frac{(I-1)!}{(1+\tau)^2} \quad (44)$$

is given by Nelder-Mead algorithm, see Appendix B.

We begin with (37), $\hat{\mu} = \underline{\theta}_j$. Then we use formulas from (43) to (44) to generate an approximated τ . For the next step, we plug τ into formulas (36) to (42) to get an approximation of $\underline{\mu}$. We do the above process for $\tilde{M} = 10000$ times to get a sample of $\underline{\mu}$ and τ . Then based on the approximated $\underline{\mu}$ s and τ s we generate \tilde{M} \underline{p} s and \underline{q} s. Then using (13) we subsample $M = 1000$ of \underline{p} s and \underline{q} s which we use in (14). For independence case we use a similar procedure in (34).

2.4 Errors

We will look at two different errors, numerical error and sampling error.

Numerical Errors

Suppose, we apply our model to a data set has S $r \times c$ tables. To get marginal likelihood density for the Bayes factor test we use equation (14) and (33) and set M . For each application we get M general marginal likelihood densities, denoted as M_1 , and M independent marginal likelihood densities, denoted as M_0 . Thus, we get two samples with sample size M . The log BF (Bayes Factor) is the log of the ratio of the two marginal likelihood sample means, which is given by $\log(\frac{\bar{M}_1}{\bar{M}_0})$. The numerical standard error (NSE) is the standard deviation of the log of the ratio of the two marginal likelihood ratio samples, which is given as

$$NSE = \sqrt{Var(\log(\frac{\bar{M}_1}{\bar{M}_0}))},$$

where

$$\begin{aligned} Var(\log(\frac{\bar{M}_1}{\bar{M}_0})) &= Var(\log(\bar{M}_1) - \log(\bar{M}_0)) \\ &= (Var(\log(M_1)) + Var(\log(M_0)) - 2Cov(\log(M_0), \log(M_1)))/M. \end{aligned}$$

For the derivation we have

$$Var(\bar{y} - \bar{x}) = Var(\bar{y}) + Var(\bar{x}) - 2Cov(\bar{y}, \bar{x}),$$

where,

$$Cov(\bar{y}, \bar{x}) = \frac{1}{n} Cov(\underline{y}, \underline{x}) \frac{1}{n}' = \frac{\sum \sum Cov(\underline{y}_i, \underline{x}_j)}{n^2}.$$

In our case, if $i = j$, x and y are correlated, if not they are independent. Thus:

$$Cov(\bar{y}, \bar{x}) = \frac{\sum_1^n Cov(\underline{y}, \underline{x})}{n^2} = \frac{Cov(\underline{y}, \underline{x})}{n},$$

and,

$$Var(\bar{y} - \bar{x}) = (Var(\underline{y}) + Var(\underline{x}) - 2Cov(\underline{y}, \underline{x}))/n.$$

Sampling Errors

For the sampling standard error (SSE), we use a bootstrap algorithm. In each area, the cells follow a multinomial distribution. First, we get the values of parameters for these multinomial distributions based on the original data set (i.e., $p_{si} = n_{si}/n_s$). Then, we sample the tables in the data set separately from multinomial distribution with the sample size equal to the original sample size of the table (i.e., n_s), probabilities of the cells being to p_{si} . Thus, each time we bootstrap a new data set with S tables. We do it M times and get M data sets. We apply our model on those M new data sets to get M log BFs. The sampling error is the standard deviation of these M log BFs. These two errors will be showed in part III.

Part III

Data Analysis

In these section, we

- (a) Apply our methodology on BMI and BMD data;
- (b) Use bootstrapping to obtain sampling errors of log Bayes factors;
- (c) Study the sensitivity of pooled model and direct model by changing the prior distribution of $\underline{\mu}, \tau$;
- (d) Compare the results of chi-squared test, pooled model and direct model.

For direct model, we use the formula from (Nandram, (2007))

$$BF = \frac{D_{rc}(\underline{n} + \underline{k})/D_r(\underline{n}^{(1)} + \underline{k}^{(1)})D_c(\underline{n}^{(2)} + \underline{k}^{(2)})}{D_{rc}(\underline{k})/D_r(\underline{k}^{(1)})D_c(\underline{k}^{(2)})}, \quad (45)$$

where the prior of general hypothesis $\underline{\pi} \sim \text{Dirichlet}(\underline{k})$, the priors of independent hypothesis $\underline{\pi}^{(1)} \sim \text{Dirichlet}(\underline{k}^{(1)})$, $\underline{\pi}^{(2)} \sim \text{Dirichlet}(\underline{k}^{(2)})$, $n_j^{(1)} = \sum_{k=1}^K n_{jk}$, $j = 1, 2, \dots, J$, $n_k^{(2)} = \sum_{j=1}^J n_{jk}$. $D_k(\underline{\mu}\tau) = \prod_{j=1}^J \Gamma(\mu_j\tau)/\Gamma(\tau)$ is the Dirichlet function with $\tau > 0$, $\mu_j > 0$, $\sum_{j=1}^J \mu_j = 1$.

3.1 Numerical Study

Set M=10,000 for equation (13), (32), (33), M=1,000 for equation (14), (34) and perform the program step by step, we get Table 3. Log BF is the log Bayes factor. NSE is the numerical standard deviation. Instead of using county number as the index we use normal index, 1 to 31.

TABLE 3: Numerical results for pooled Bayes factor test

county	log BF	NSE	result
1	6.82	0.10	VR
2	4.57	0.10	SR
3	1.43	0.12	PR
4	2.83	0.11	PR
5	-2.12	0.21	N
6	6.36	0.11	VR
7	4.78	0.15	SR
8	2.75	0.20	PR
9	4.64	0.34	SR
10	-3.39	0.28	N
11	4.24	0.09	SR
12	5.60	0.47	VR
13	3.03	0.10	SR
14	-1.06	0.21	N
15	9.77	0.24	VR
16	9.65	0.21	VR
17	2.82	0.10	PR
18	2.43	0.13	PR
19	3.93	0.18	SR
20	2.92	0.12	PR
21	4.79	0.18	SR
22	7.67	0.11	VR
23	4.86	0.19	SR
24	4.48	0.11	SR
25	-0.70	0.08	N
26	-1.23	0.14	N
27	1.60	0.16	PR
28	2.54	0.17	PR
29	2.47	0.18	PR
30	1.92	0.23	PR
31	4.14	0.18	SR

Note: VR: Very strongly reject, SR: Strongly reject PR: Positively reject,

BR: Barely reject

We get 6 very strong rejections, 10 strong rejections, 10 positive rejections and 5 negative log Bayes factors. The NSE of the log Bayes factor are all less than 0.5, the biggest one is 0.34. Note that the Bayes factors for the direct model are obtained without numerical errors.

3.2 Sampling Error

For the next step, we obtain sampling standard errors of the our model based on BMI and BMD data. We set $M=100$ and perform the bootstrap procedure for both pooled and direct models.

TABLE 4: Sampling errors for pooled and direct models

	pooled log BF	pooled SSE	direct log BF	direct SSE
1	5.6329	0.0193	2.5690	0.0317
2	3.9773	0.0287	-1.1268	0.0320
3	1.4032	0.0262	-2.5090	0.0269
4	2.0562	0.0262	-2.3849	0.0251
5	-5.5391	0.0340	-0.2420	0.0257
6	5.9668	0.0273	-1.9546	0.0341
7	4.0609	0.0358	-2.9406	0.0325
8	3.1335	0.0373	-6.3505	0.0304
9	3.4688	0.0506	-2.9234	0.0422
10	-8.6923	0.0485	-3.8737	0.0234
11	4.4520	0.0259	-2.5214	0.0314
12	5.9873	0.0568	-5.9712	0.0381
13	1.9759	0.0350	0.1258	0.0359
14	-3.9146	0.0387	-7.1817	0.0275
15	8.0607	0.0405	1.9651	0.0410
16	9.8016	0.0363	-1.1488	0.0381
17	1.3922	0.0281	-2.4630	0.0277
18	2.2151	0.0335	-6.4556	0.0277
19	2.3704	0.0373	1.0612	0.0325
20	2.9293	0.0298	-2.6694	0.0314
21	2.8927	0.0325	1.1938	0.0305
22	7.1385	0.0186	4.6406	0.0386
23	4.8365	0.0368	0.4421	0.0380
24	4.4936	0.0305	0.4487	0.0285
25	-1.0433	0.0265	-3.8931	0.0222
26	-3.6414	0.0279	-2.7833	0.0262
27	-0.1041	0.0321	-4.8186	0.0289
28	0.8363	0.0356	-4.5220	0.0288
29	1.0864	0.0323	-2.3887	0.0273
30	-1.4810	0.0367	-6.7982	0.0289
31	2.2471	0.0432	-1.6210	0.0370

The SSE of log Bayes factors for the pooled model are similar to those from the direct model. The sampling errors in pooled model are smaller than in direct model for 12 areas. The highest SSE for the models is 0.06 and the lowest one is 0.0193. However the log Bayes factors differ greatly. In the pooled model, there are 6 very strong rejections, 7 strong rejections, 10 positive rejections, 1 weak rejection and 7 negative log Bayes Factors. In the direct model, there are no very strong rejections, only 1 strong rejection, 4 positive rejections, 3 weak rejections.

3.3 Sensitivity Analysis

In our model, without any prior information τ has a distribution $p(\tau) = \frac{1}{(1+\tau)^2}, \tau > 0$, which is a special case for symmetric F distribution $f(x | \nu) = \frac{\Gamma(\nu)}{\Gamma(\frac{\nu}{2})^2} \frac{x^{(\nu-2)/2}}{(1+x)^\nu}$ with $\nu = 2$. $\underline{\mu}$ has a distribution Dirichlet(1).

Now, we set $\nu = 1.5, 2, 3, 5, 10$, and $\underline{\mu}$ has a distribution Dirichlet(\underline{k}), where $k = 0.5, 1, 2, 5, 10$. We still assume independence of $\underline{\mu}$ and τ to test the sensitivity of our model to the prior information. Thus,

$$f(\underline{\mu}, \tau) = \frac{\prod_i^I \mu_i^{k_i-1}}{D(k)} \frac{\Gamma(\nu)}{\Gamma(\frac{\nu}{2})^2} \frac{\tau^{(\nu-2)/2}}{(1+\tau)^\nu}.$$

To compare the sensitivity with direct Bayes factor model, we also change the prior distribution of direct model which is Dirichlet(\underline{k}), with the same set of values, $k = 0.5, 1, 2, 5, 10$.

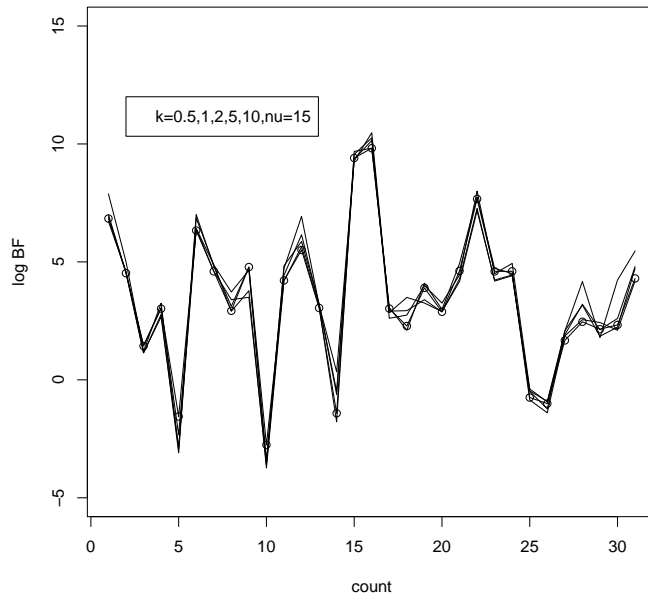


FIGURE 1: Sensitivity of Bayes factors for the pooled model with $\nu=1.5$ and $k=0.5, 1, 2, 5, 10$

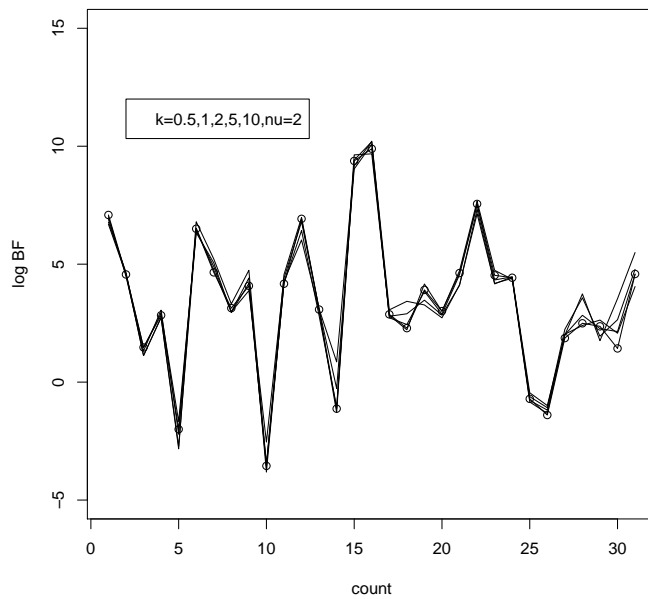


FIGURE 2: Sensitivity of Bayes factors for the pooled model with $\nu=2$ and $k=0.5, 1, 2, 5, 10$

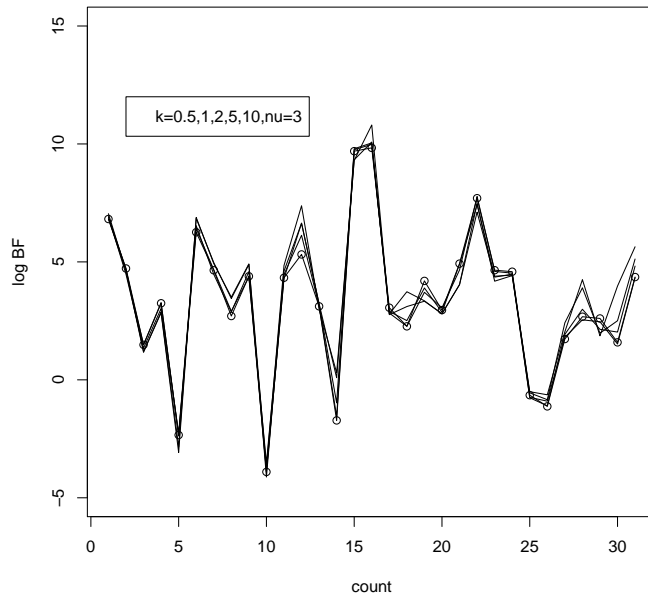


FIGURE 3: Sensitivity of Bayes factors for the pooled model with $\nu=3$ and $k=0.5, 1, 2, 5, 10$

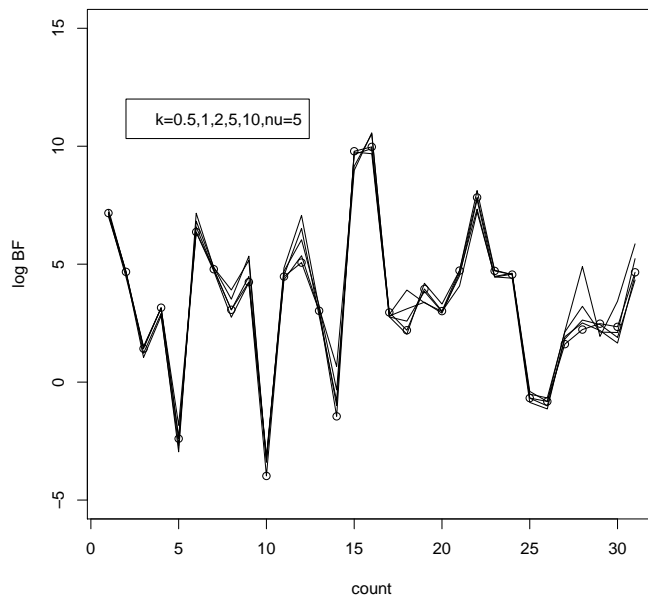


FIGURE 4: Sensitivity of Bayes factors for the pooled model with $\nu=5$ and $k=0.5, 1, 2, 5, 10$

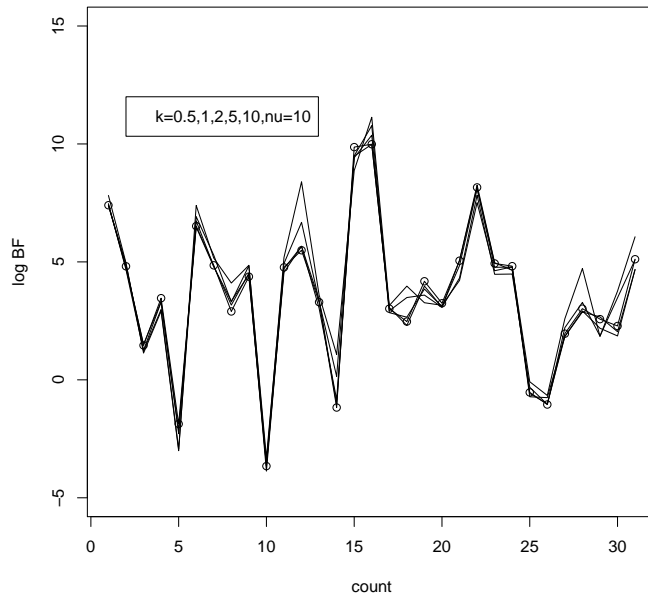


FIGURE 5: Sensitivity of Bayes factors for the pooled model with $\nu=10$ and $k=0.5, 1, 2, 5, 10$

In Figures 1 to 5, the log BFs are similar to each other for different k , which means that our model is not sensitive to prior distribution of the μ .

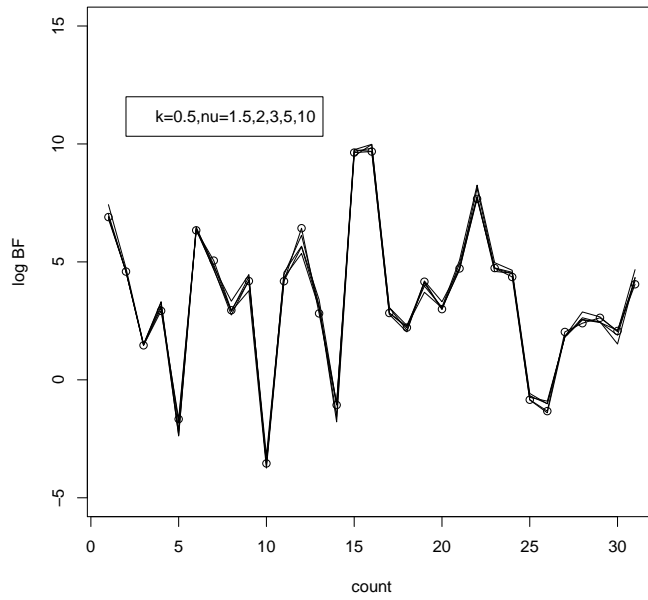


FIGURE 6: Sensitivity of Bayes factors for the pooled model with $k=0.5$ and $\nu=1.5, 2, 3, 5, 10$

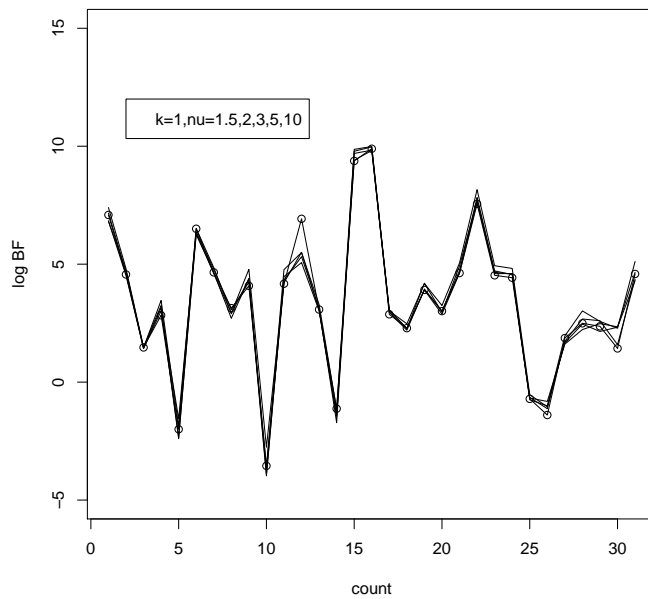


FIGURE 7: Sensitivity of Bayes factors for the pooled model with $k=1$ and $\nu=1.5, 2, 3, 5, 10$

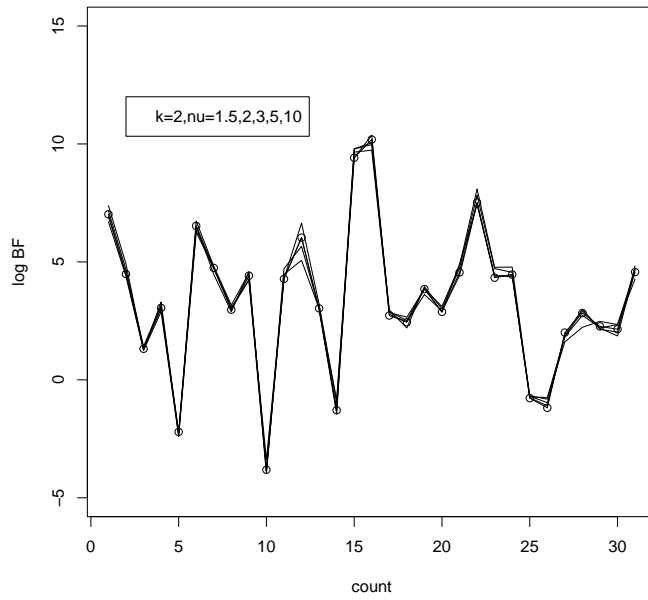


FIGURE 8: Sensitivity of Bayes factors for the pooled model with $k=2$ and $\nu=1.5, 2, 3, 5, 10$

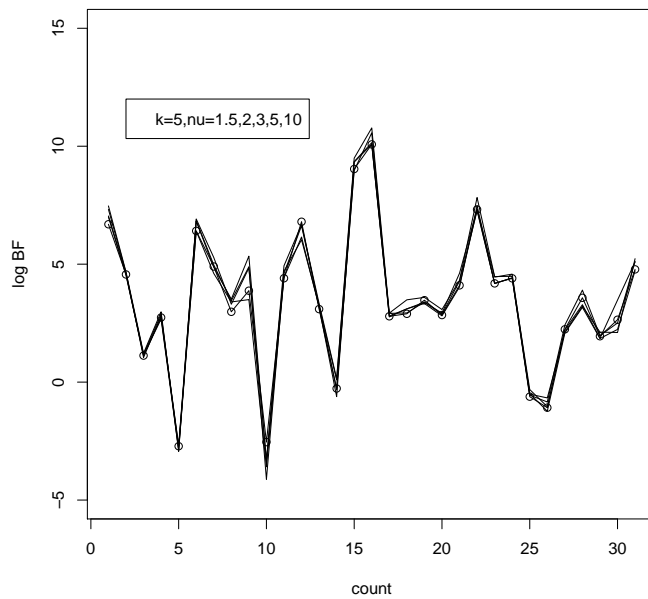


FIGURE 9: Sensitivity of Bayes factors for the pooled model with $k=5$ and $\nu=1.5, 2, 3, 5, 10$

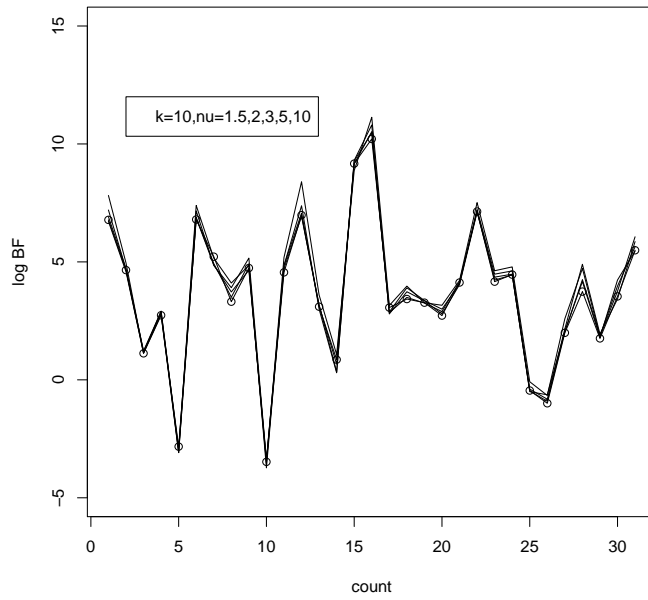


FIGURE 10: Sensitivity of Bayes factors for the pooled model with $k=10$ and $\nu=1.5$, 2, 3, 5, 10

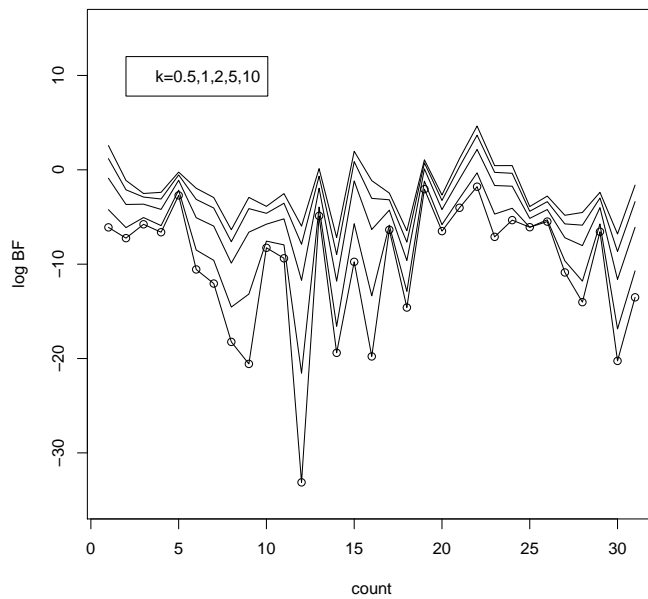


FIGURE 11: Sensitivity of Bayes factors for direct model with $k=0.5, 1, 2, 5, 10$

In Figures 6 to 10, the log Bayes factors are close to each other for different values of ν , which means that our pooled model is not sensitive to the prior distribution of the τ . Thus, the Bayes factors are not sensitive to its prior $f(\underline{\mu}, \tau)$.

Figure 11 represents the direct model with 5 different priors. The scale of direct model (-30 to 20) is longer than the scale (-5 to 35) of pooled model. The log Bayes factors with different prior are much different from each other. Thus the direct model is sensitive to prior distribution. Thus, it is easy to say that direct model is much sensitive to prior information than pooled model.

Part IV

Conclusion

In this paper, we construct a pooled model to perform a Bayesian test of independence for small areas where some areas have too limited data (i.e., cell frequencies are less than 5) to generate any accurate estimations based on the data from the each of them.

We put chi-squared test, pooled Bayesian test, and direct Bayesian test results of BMI and BMD data together (shown in Table 5) to compare them. Set $P\text{-value} < 0.01$, very strongly reject; $0.01 < p\text{-value} < 0.25$, strongly reject; $0.25 < p\text{-value} < 0.05$, positively reject; $0.05 < p\text{-value} < 0.1$, weakly reject. A summary of is shown in Table 6.

There are 6 very strong rejections in both chi-squared and pooled test, while there are no very strong rejections in the direct test. The chi-squared test has 3 strong rejections, the pooled test has 7 strong rejections, the direct test has 1 rejection. The results from the direct test mostly differ from the pooled and chi-squared test. Thus we do further comparison to the results of chi-square test and pooled test.

TABLE 5: Test results from chi-squared test, pooled Bayesian test, and direct Bayesian test

	chi-squared	pooled	direct
1	0.00	5.63	2.57
2	0.05	3.98	-1.13
3	0.40	1.40	-2.51
4	0.18	2.06	-2.38
5	0.19	-5.54	-0.24
6	0.03	5.97	-1.95
7	0.08	4.06	-2.94
8	0.39	3.13	-6.35
9	0.02	3.47	-2.92
10	0.33	-8.69	-3.87
11	0.08	4.45	-2.52
12	0.03	5.99	-5.97
13	0.00	1.98	0.13
14	0.24	-3.91	-7.18
15	0.00	8.06	1.97
16	0.00	9.80	-1.15
17	0.19	1.39	-2.46
18	0.22	2.22	-6.46
19	0.08	2.37	1.06
20	0.29	2.93	-2.67
21	0.01	2.89	1.19
22	0.00	7.14	4.64
23	0.01	4.84	0.44
24	0.03	4.49	0.45
25	0.71	-1.04	-3.89
26	0.21	-3.64	-2.78
27	0.32	-0.10	-4.82
28	0.09	0.84	-4.52
29	0.26	1.09	-2.39
30	0.13	-1.48	-6.80
31	0.00	2.25	-1.62

TABLE 6: Summary of results in table 5

Test	Results			
	Weak	Positive	Strong	Very
Chi-square	5	3	3	6
Pooled	1	10	7	6
Direct	3	4	1	0

TABLE 7: Comparison of chi-squared and pooled tests results

chi-squared test	pooled Bayesian test			
	Barely	Positive	Strong	Very
Weak	1(28)	1(19)	3(2,7,11)	0
Positive	0	0	1(24)	2(6,12)
Strong	0	1(21)	2(9,23)	0
Very	0	2(13,31)	0	4(1,15,16,22)

In table 7, the rows represent categories of pooled test results, columns represent the categories of chi-squared test results. In the diagonal of the table (i.e., county 28, 9, 23, 1, 15, 16, 22) the two test results are same. In most counties they are not accordant. Although chi-squared test are not invalid, some of them are still compatible with the pooled Bayes factor test.

Instead of using only the information within a single area, our model borrow strength from other similar areas. Both NSE and SSE for our model are small. Our model is shown not to be sensitivity to prior distribution. However, the construction of our model is under the assumption of multinomial distribution of the data. In the future, we may deal with the test when there is an intra-class correlation, thereby violating the assumptions in the multinomial distribution.

Part V

Appendix

Appendix A

Proposal density for the conditional posterior density of $\underline{\mu}$ given τ

Starting with the prior $\pi(\underline{\mu}, \tau) = 1$, the conditional posterior distribution of $\underline{\mu} \mid \tau, \underline{n}$ is

$$\pi(\underline{\mu} \mid \tau, \underline{n}) \propto \prod_{i=1}^I \prod_{j=1}^J \left\{ \frac{\Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\mu_j \tau)} \right\}, 0 < \mu_j < 1 \text{ and } \sum_{s=1}^J \mu_s = 1.$$

We use Stirling's formula to obtain a convenient approximation for $\pi(\underline{\mu} \mid \tau, \underline{n})$. Stirling's formula states that for large a

$$\Gamma(a) \approx \sqrt{\frac{2\pi}{a}} \left(\frac{a}{e}\right)^a$$

and Stirling's formula implies that for any h

$$\frac{\Gamma(a+h)}{\Gamma(a)} \approx a^h \text{ and } \frac{\Gamma(a+h)}{\Gamma(a)} \approx h^h e^{-h} \left(1 + \frac{h}{a}\right)^a \left(1 + \frac{a}{h}\right)^h.$$

See Feller (1968, Ch. 2, Sec. 9 and Exe. 22).

By assuming the $\mu_j \tau$ are large, the first implication of Stirling's formula gives

$$\frac{\Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\mu_j \tau)} \approx (\mu_j \tau)^{n_{ij}} \tag{A.1}$$

while the second implication gives

$$\frac{\Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\mu_j \tau)} \approx n_{ij}^{n_{ij}} e^{-n_{ij}} \left(1 + \frac{n_{ij}}{\mu_j \tau}\right)^{\mu_j \tau} \left(1 + \frac{\mu_j \tau}{n_{ij}}\right)^{n_{ij}}, \tag{A.2}$$

$i = 1, 2, \dots, I$, $j = 1, 2, \dots, J$. By (A.1) we can approximate $\pi(\underline{\mu} \mid \tau, \underline{n})$ by

$$p_1(\underline{\mu} \mid \underline{n}, \tau) = \frac{\prod_{j=1}^J \mu_j^{n_{\cdot j}}}{D(\underline{a})}, 0 < \mu_j < 1 \text{ and } \sum_{j=1}^J \mu_{ij} = 1$$

where $D(\underline{a}) = \prod_{j=1}^J \Gamma(a_j) / \Gamma\left(\sum_{j=1}^J a_j\right)$ and $a_j = n_{\cdot j} = \sum_{i=1}^I n_{ij}$, $j = 1, 2, \dots, J$. Note that $p_1(\underline{\mu} | \underline{n}, \tau)$ has a distribution independent of τ and

$$E(\mu_j | \tau, \underline{n}) = \frac{n_{\cdot j} + 1}{\sum_{s=1}^J (n_{\cdot s} + 1)} = \hat{\theta}_j, \quad j = 1, 2, \dots, J. \quad (\text{A.3})$$

By (A.2) we can approximate $\pi(\underline{\mu} | \tau, \underline{n})$ by

$$p_2(\underline{\mu} | \underline{n}, \tau) \propto \prod_{i=1}^I \prod_{j=1}^J \left(1 + \frac{n_{ij}}{\mu_j \tau}\right)^{\mu_j \tau} \left(1 + \frac{\mu_j \tau}{n_{ij}}\right)^{n_{ij}}, \quad 0 < \mu_j < 1 \text{ and } \sum_{j=1}^J \mu_j = 1.$$

Now, letting $\underline{\mu}_{(J)} = (\mu_1, \mu_2, \dots, \mu_{J-1})'$, the posterior distribution of $\underline{\mu}_{(J)} | \tau, \underline{n}$ can be approximated by

$$p_3(\underline{\mu}_{(J)} | \tau, \underline{n}) \propto \exp(D_1 + D_2)$$

where

$$D_1 = \sum_{j=1}^{J-1} \sum_{i=1}^I \{\tau \mu_j \ln(1 + n_{ij}/\tau \mu_j) + n_{ij} \ln(1 + \tau \mu_j/n_{ij})\}$$

and

$$\begin{aligned} D_2 &= \sum_{i=1}^I \left[\tau \left(1 - \sum_{j=1}^{J-1} \mu_j\right) \ln \left\{1 + n_{iJ}/\tau \left(1 - \sum_{j=1}^{J-1} \mu_j\right)\right\} \right. \\ &\quad \left. + n_{iJ} \ln \left\{1 + \tau \left(1 - \sum_{j=1}^{J-1} \mu_j\right)/n_{iJ}\right\} \right], \quad 0 < \mu_j < 1 \text{ and } j = 1, 2, \dots, J-1. \end{aligned}$$

Next, we approximate $p_3(\underline{\mu}_{(J)} | \tau, \underline{n})$ by a second order multivariate Taylor's series expansion about $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_{J-1})$ in (A.3). Then letting k_1 and k_2 be constants,

$$D_1 \approx k_1 + \sum_{j=1}^{J-1} (\mu_j - \hat{\theta}_j) A_j - \frac{1}{2} \sum_{j=1}^{J-1} (\mu_j - \hat{\theta}_j)^2 B_j$$

and

$$D_2 \approx k_2 - A_J \sum_{j=1}^{J-1} (\mu_j - \hat{\theta}_j) - \frac{1}{2} B_J \left(\sum_{j=1}^{J-1} (\mu_j - \hat{\theta}_j) \right)^2$$

where

$$A_j = \tau \sum_{i=1}^I \ln(1 + n_{ij}/\tau \hat{\theta}_j)$$

and

$$B_j = \tau \sum_{i=1}^I \{ \hat{\theta}_j^{-1} - (\hat{\theta}_j + n_{ij}/\tau)^{-1} \}$$

$j = 1, 2, \dots, J$. Thus, we approximate $p_3(\underline{\mu}_{(J)} | \tau, \underline{n})$ by

$$p_4(\underline{\mu}_{(J)} | \underline{n}, \tau) \propto \exp \left[-\frac{1}{2} \left\{ B_J \left\{ \sum_{j=1}^{J-1} (\mu_j - \hat{\theta}_j) \right\}^2 + \sum_{j=1}^{J-1} (\mu_j - \hat{\theta}_j)^2 B_j - 2 \sum_{j=1}^{J-1} (A_j - A_J) (\mu_j - \hat{\theta}_j) \right\} \right], \quad 0 < \mu_j < 1. \quad (\text{A.4})$$

It follows from (A.4) that $\underline{\mu} | \underline{n}, \tau$ is approximately normally distributed with mean

$$\hat{\mu}_j = \hat{\theta}_j + B_j^{-1} (A_j - \bar{A})$$

where $\bar{A} = \sum_{j=1}^J B_j^{-1} A_j / \sum_{j=1}^J B_j^{-1}$ and letting $\nu_j = B_j^{-1} / \left(\sum_{s=1}^J B_s^{-1} \right)^{1/2}$

$$\text{cov}(\mu_j, \mu_{j'} | \underline{n}, \tau) = \begin{cases} B_j^{-1} - \nu_j^2, & j = j' \\ -\nu_j \nu_{j'}, & j \neq j' \end{cases} \quad (\text{A.5})$$

$j, j' = 1, 2, \dots, J$. As $\hat{\mu}_j$ can be outside $[0, 1]$ we take

$$\tilde{\mu}_j = \begin{cases} \hat{\mu}_j, & 0 < \hat{\mu}_j < 1 \\ \hat{\theta}_j, & \hat{\mu}_j \leq 0 \text{ or } \hat{\mu}_j \geq 1. \end{cases}$$

Then, we approximate $E(\mu_j | \underline{n}, \tau)$ by $\hat{\mu}_j$ where

$$\hat{\mu}_j = \tilde{\mu}_j / \sum_{s=1}^J \tilde{\mu}_{sJ} = 1, 2, \dots, J. \quad (\text{A.6})$$

Note that if $\tau > 1$, $\text{cov}(\mu_j, \mu_{j'} | \underline{n}, \tau)$ is well defined.

Finally, a Dirichlet approximation for the conditional posterior distribution of $\underline{\mu} | \underline{n}, \tau$ is utilized. We equate the covariance matrix of a Dirichlet distribution with parameters $(\mu_1^{(a)} \tau^{(a)}, \mu_2^{(a)} \tau^{(a)}, \dots, \mu_J^{(a)} \tau^{(a)})$ with that in (A.4) where $\mu_j^{(a)} = \hat{\mu}_j, j = 1, 2, \dots, J$. Letting

$$d_{jj'} = \begin{cases} \mu_j^{(a)2} / (B_j^{-1} - \nu_j^2), & j = j' \\ \mu_j^{(a)} \mu_{j'}^{(a)} / (\nu_j \nu_{j'} - 1), & j \neq j' \end{cases}$$

$j, j' = 1, 2, \dots, J$, we take

$$\tau^{(a)} = \left\{ \sum_{j=1}^J \sum_{j'=1}^j d_{jj'} I_{jj'} \right\} / \sum_{j=1}^J \sum_{j'=1}^j I_{jj'} \quad (\text{A.7})$$

where $I_{jj'} = 1$ if $d_{jj'} > 0$ and $I_{jj'} = 0$ if $d_{jj'} \leq 0$.

Thus, $\underline{\mu} \mid \tau, \underline{n} \sim \text{Dirichlet}(\mu_1^{(a)} \tau^{(a)}, \mu_2^{(a)} \tau^{(a)}, \dots, \mu_J^{(a)} \tau^{(a)})$ approximately. Note that the $\mu_j^{(a)}$ and $\tau^{(a)}$ are functions of τ .

Appendix B

Proposal density for the conditional posterior density of τ given $\underline{\mu}$

Starting with the prior $\pi(\underline{\mu}, \tau) = 1$, the conditional posterior distribution of $\tau \mid \underline{\mu}, \underline{n}$ is

$$\pi(\tau \mid \underline{\mu}, \underline{n}) \propto \prod_{i=1}^I \left\{ \prod_{j=1}^J \frac{\Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\mu_j \tau)} \right\} / \{\Gamma(n_{i\cdot} + \tau) / \Gamma(\tau)\}, \quad \tau > 0.$$

Using Stirling's formula, we approximate $\pi(\tau \mid \underline{\mu}, \underline{n})$ by

$$p_1(\tau \mid \underline{\mu}, \underline{n}) \propto \exp(E_1 - E_2) \tag{B.1}$$

where

$$E_1 = \sum_{i=1}^I \sum_{j=1}^J \left\{ \mu_j \tau \ln \left(1 + \frac{n_{ij}}{\mu_j \tau} \right) + n_{ij} \ln \left(1 + \frac{\mu_j \tau}{n_{ij}} \right) \right\},$$

$$E_2 = \sum_{i=1}^I \left\{ \tau \ln \left(1 + \frac{n_{i\cdot}}{\tau} \right) + n_{i\cdot} \ln \left(1 + \frac{\tau}{n_{i\cdot}} \right) \right\}$$

and $n_{i\cdot} = \sum_{j=1}^J n_{ij}$, $i = 1, 2, \dots, I$.

Next, we approximate E_1 and E_2 by second order univariate Taylor's series expansions about τ_* , where τ_* is the posterior mode of $\pi(\tau \mid \underline{n}, \underline{\mu})$ for each $\underline{\mu}$. We have

$$E_1 \approx k_1 + (\tau - \tau_*) \sum_{i=1}^I \sum_{j=1}^J \mu_j \ln \left(1 + \frac{n_{ij}}{\mu_j \tau_*} \right) - \frac{1}{2} (\tau - \tau_*)^2 \sum_{i=1}^I \sum_{j=1}^J \mu_j^2 \left\{ \frac{1}{\mu_j \tau_*} - \frac{1}{n_{ij} + \mu_j \tau_*} \right\} \tag{B.2}$$

where k_1 is a constant. An important first step is to approximate E_2 by

$$E_2^* = \sum_{i=1}^I \left\{ \tau_* \ln \left(1 + \frac{n_{i\cdot}}{\tau_*} \right) + n_{i\cdot} \ln \left(1 + \frac{\tau_*}{n_{i\cdot}} \right) \right\}$$

and expanding E_2^* about τ_* to the first order, we have

$$E_2 \approx k_2 + \frac{1}{2} (\tau - \tau_*)^2 \sum_{i=1}^I \left(\frac{1}{\tau_*} - \frac{1}{\tau_* + n_{i\cdot}} \right) \tag{B.3}$$

where k_2 is a constant. The posterior mode τ_* is obtained by an optimization routine (e.g., Nelder-Mead algorithm.)

It follows easily from (B.1), (B.2) and (B.3) that the posterior mean and variance of $\tau \mid \underline{\mu}, \underline{n}$ can be approximated by μ_* and σ_*^2 respectively where

$$\mu_* = \tau_* + \sigma_*^2 \sum_{i=1}^I \sum_{j=1}^J \mu_j \ln \left(1 + \frac{n_{ij}}{\mu_j \tau_*} \right)$$

and

$$\sigma_*^2 = \left[\sum_{i=1}^I \left\{ \left(\frac{1}{\tau_*} - \frac{1}{\tau_* + n_i} \right) + \sum_{j=1}^J \mu_j^2 \left(\frac{1}{\mu_j \tau_*} - \frac{1}{n_{ij} + \mu_j \tau_*} \right) \right\} \right]^{-1}. \quad (\text{B.4})$$

As $\tau > 0$ we approximate the posterior distribution of $\tau \mid \underline{\mu}, \underline{n}$ by a gamma distribution with index $\eta^{(a)}$ and scale $\nu^{(a)}$. Thus, equating the mode of this candidate with the true mode τ_* we have $(\eta^{(a)} - 1)/\nu^{(a)} = \tau_*$, and equating the variances we have $\eta^{(a)}/(\nu^{(a)})^2 = \sigma_*^2$. It follows that approximately

$$\tau \mid \underline{\mu}, \underline{n} \sim \Gamma(\eta^{(a)}, \nu^{(a)}) \quad (\text{B.5})$$

where

$$\eta^{(a)} = \left\{ \frac{\tau_*}{2\sigma_*} + \sqrt{\left(\frac{\tau_*}{2\sigma_*} \right)^2 + 1} \right\}^2 \quad \text{and} \quad \nu^{(a)} = \sqrt{\eta^{(a)}}/\sigma_*.$$

Note that $\eta^{(a)}$ and $\nu^{(a)}$ are functions of $\underline{\mu}$.

Appendix C

An application to data from the Kauffman Firm Survey

To test the feasibility of our method, we apply it to a different data set obtained from the Kauffman Firm Survey (KFS). For the KFS, the target population was all new businesses that were started in the 2004 calendar year in the United States (the 50 states plus the District of Columbia). The KFS is a panel study of 4,928 businesses founded in 2004 and tracked over their early years of operation. The survey focuses on the nature of new business formation activity, characteristics of the strategy, offerings, and employment patterns of new businesses, the nature of the financial and organizational arrangements of these businesses, and the characteristics of their founders.

The sample for the first follow-up survey consisted of the 4,928 businesses that completed the baseline survey. The first follow-up was conducted between June 2006 and January 2007, during which 3,998 interviews were completed and 369 companies were identified as “out of business”. Most of them are small companies. In 2004 more than 57.6 percent of the companies do not have employees, 27 percent of the companies have 1 to 3 employees, 15.4 percent of the companies have more than 3 employees. In 2004, 1132 companies have 1 financial expert and 8 companies have more than 5 financial experts. In 2008, these numbers change to 968 and 7 respectively.

We focus on companies with 1 to 3 employees and group them by primary locations (5 levels). We are interested in the factors that affect the companies’ exist life and will test the association of Average Number of Financial Experts and Exist Life and Number of Owners and Exist Life.

We classify the Exist Life into four categories. As all data in the database are companies starting from 2004, Exist Life = 1 if the company is out of business in 2005, Exist life = 2 if out in 2006, Exist Life = 3 if out in 2007, else Exist Life = 4. With the median is 0.75 for the average number of financial experts (add financial experts up from 2004 to 2008 then divided by 4) bigger than 0, we classify Average Number of Experts into 3 levels. ANOFE=0, if average number of financial experts = 0, ANOFE=1, if $0 < \text{average number of financial experts} < 0.75$, else ANOFE=2. The table is shown below:

TABLE 8: Classification of businesses by number of financial experts and exist life for 5 primary locations

PL	FE	Exist life			
		1	2	3	4
1	1	10	8	5	66
	2	0	22	29	250
	3	35	14	19	160
2	1	4	5	2	20
	2	0	8	22	218
	3	21	17	14	247
3	1	0	0	0	2
	2	0	1	1	14
	3	3	3	2	42
4	1	0	0	0	1
	2	0	2	4	29
	3	1	0	0	18
5	1	1	0	0	1
	2	0	1	6	7
	3	0	0	3	3

Note: PL: primary location, FE: Financial experts

We use chi-squared, direct model and pooled model to test the association.

TABLE 9: Test results of Table 8 from chi-squared, direct model and pooled model

PL	pooled log BF	pooled NSE	direct log BF	Direct NSE	p-value
1	254.807574	4.99	240.011247	18.69	0.001
2	261.216171	4.68	240.98321	16.7	0.001
3	19.618089	0.669	11.370242	5.29	0.959
4	30.577243	0.37	18.743516	3.196	0.477
5	2.297748	0.101	1.840688	1.69	0.021

Note: p-value refers to the chi-squared test.

Bayes factors are considerably different for locations. Results from the direct model are similar to those from the pooled model. The results from the chi-squared test differ from those of the Bayesian models, only for the 1st, 2nd and 5th primary locations and for the other locations the chi-squared test agree with the Bayesian models.

In the whole data set, we have 4928 observations. Among them 2973 companies were individually owned and 1334 companies were owned by two people, the sum of which is 87.56 percent of the total companies. The average owner of the companies is 1.75. As individually owned and two partners owned companies possess 91.6 percent of the total companies. We categorize the data into 3 levels: Owner=1, if individually owned, Owner=2, if two partners owned, else Owner=3.

TABLE 10: Classification of businesses by number of owners and exist life for 5 primary locations

PL	Owner	Exist life			
		1	2	3	4
1	1	34	35	40	347
	2	9	5	12	107
	3	2	4	1	22
2	1	14	18	17	257
	2	7	6	15	162
	3	4	6	6	66
3	1	1	2	3	29
	2	2	1	0	18
	3	0	1	0	11
4	1	1	1	3	33
	2	0	1	1	11
	3	0	0	0	4
5	1	1	1	0	4
	2	0	0	0	3
	3	0	0	0	1

Note: p-value refers of the chi-squared test.

TABLE 11: Test results of Table 8 from chi-squared, direct model and pooled model

PL	pooled log BF	pooled NSE	direct log BF	Direct NSE	p-value
1	289.30839	4.325	268.3	19.5	0.49
2	353.443747	4.809	335.82	19.4	0.676
3	49.99798	0.713	37.014	4.87	0.538
4	38.773211	0.359	23.52	4.82	0.946
5	8.180922	0.087	2.69	1.201	0.797

Again, the Bayes factors are considerably different. Only for the 5th location, the direct model does not agree with the pooled model. The chi-squared test fails to reject the independent hypothesis at all five locations; these differ from the Bayesian models.

References

- Ghosh, M. and Rao, J.N.K. (1994), "Small Area Estimation: An Appraisal" *Institute of Mathematical Statistics*, **9**, 55-76.
- Kass, R.E. and Raftery, A.E. (1995), "Bayes factors" *Journal of the American Statistical Association*, **90**, 773-795.
- Kass, R.E. (1993), "Bayesian factors in practise" *Statistician*, **42**, 551-560.
- Looker, A.C., Wahner, H.W., Dunn, W.L., Calvo, M.S., Harris, R.R., Heyse, S.P., Johnston, C.C. and Lindsay, R. (1998), "Updated data on proximal femur bone mineral levels of US adults" *Osteoporosis International*, **8**, 468-489.
- National Center for Health Statistics (1994), "Plan and Operation of the third National Health and Nutrition Examination Survey" *Vital and Health Statistics Series 1*, **32**
- Nandram, B. (1998), "A Bayesian analysis of the three-stage hierarchical multinomial model" *Journal of Statistical Computation and Simulation*, **61**, 97-12.
- Nandram, B., Liu, N., and Choi, J.W. (2005), "Bayesian non-response models for categorical data from small areas: an application to BMD and age" *Statistics in Medicine*, **24**, 1047-1074.
- Nandram, B. and Choi, J.W. (2007), "Alternative Test of Independent in Two-Way Categorical Tables" *Journal of Data Science*, **5**, 217-237.
- Nandram, B. and Choi, J. W. (2010), "A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection" *Journal of the American Statistical Association*, **105**, 120-135.

Pietrobelli, A., Faith, M., Allison, D., Gallagher, D., Chiumello, G., and Heymsfield, S. (1998), "Body mass index as a measure of adiposity among children and adolescents: A validation study" *Journal of Pediatrics*, **132**, 204-210.

Romero-Corral, A., Somers, V. K., Sierra-Johnson, J., Thomas, R. J., Collazo-Clavell, M. L., Korinek, J. and Allison, T. G. (2008), "Accuracy of Body Mass Index to Diagnose Obesity In the US Adult Population" *International Journal of Obesity*, **32**, 959-966.

Squires, S. (2001), "National Plan Urges to Combat Obesity: Weight-Related Illnesses kill 300,000 Americans Annually, Surgeon General Says," *The Washington Post*, December 14.

Sinharay, S., and Stern, H. S. (2002), "on the Sensitivity of Bayes Factors to The Prior Distributions" *The American Statistician*, **56**, 196-201