Bayesian Analysis of Categorical Data- A Revisit

Samyajoy Pal
Department of Statistics, University of Madras, Chennai - 5
M. Subbiah
Department of Mathematics, Presidency College, Chennai-5
M R Srinivasan
Department of Statistics, University of Madras, Chennai - 5

Abstract:

A comprehensive study of two-dimensional categorical data from a Bayesian perspective has been attempted in this study. Two categorical variables with levels I and J are considered with the assumption that the underlying model follows multinomial distribution. The entire study embraces the essential components of Bayesian approach; prior construction, computations, and appropriate inferences using posterior distributions of the parameters. Conjugate prior distribution with symmetric and asymmetric hyper parameters are considered. Newly conceived asymmetric prior is based on perceived preferences of categories. Point and Interval estimation along with different types of probability computation from posterior distribution have been done using closed form integration, Monte-Carlo integration and MCMC methods. A notion of measuring association between the parameters has been shown using correlation matrix. Bayesian computation is done using R programming language and illustrated with appropriate data sets. Study has highlighted the application of Bayesian inference exploiting the distributional form of underlying parameters.

Keywords:

Bayesian Inference, Categorical Data,Closed Form Integration,MCMC Methods,Monte-Carlo Simulation, Preference Prior

1.Introduction:

Analysing categorical data is very common in statistical practice. But doing a Bayesian analysis gives some extra flexibility to explore the data. In a Bayesian analysis, there is a data model and most of the time the model follows some probability distribution. But the most interesting part is that, the unknown parameter is considered to be a random variable with a probability distribution. This is called prior distribution. We must have a state of knowledge before constructing a prior distribution. The prior can be convenient, informative, noninformative or sometimes improper. But selecting a suitable prior for a given problem is a statistician's responsibility. Then we have the likelihood of the data. And using Bayes formula we obtain a

posterior distribution of the unknown parameter. In simple words, we have some prior information (state of knowledge) about the unknown parameters, then we update that information using the data in hand and draw some inference about the unknown parameters. The complete Bayesian analysis from inferential aspect has been comprehensively described by Gelman et al (2002). With Bayesian analysis, we can directly answer questions regarding the population parameters and that is why it is a convenient way to deal with unknown parameters. The procedure is pretty much straight-forward for a categorical data, but beside estimating the parameters and testing independence between the factors, many modern-day problems demand something more. Sangeetha et al (2014) has provided an alternative way to test independence between two categorical variables in a I×J table, using Bayes Factor.

Now let us consider a data set on voters' preference discussed in Gelman et al (2002). In late October 1988, a survey was conducted by CBS News of 1447 adults in the United States to find out their preferences in the upcoming Presidential election. Out of 1447 persons, x_1=727 supported George Bush, x_2=583 supported Michael Dukakis, and x_3=137 supported other candidates or expressed no opinion. They have estimated the probability of the difference of the population proportion of Bush and Dukakis supporters ($\theta_1 - \theta_2$) by MCMC methods. But, in this situation other relevant questions may arise. So, the exercise can be extended to answer more probabilistic questions regarding the parameters. Suppose, the persons (or a fraction of them) who expressed no opinion decide to vote for Dukakis, then it changes the whole political equation. So, we may want to know what is the probability that $\theta_2 + \theta_3$ is greater than θ_1 or ($\theta_2 + (\theta_3 / a)$) is greater than θ_1. It can also happen that some people (probably, Dukakis supporters) are also interested to know that, given the condition Bush got more than 50% support, what is the probability that Dukakis has at least 40% support.

If we imagine the same situation in Indian political system and there are three parties viz. Party A, Party B and Party C, and $p(\theta_1 - \theta_2) = 0.999$ (almost certain), then also, it is not certain that Party A can form the Government. Because after election Party B and Party C can form alliance and if they have more support combined than Party A, they can form the Government. So, now it is of our interest to know that, given Party A has at least 40% support, what is the probability that party B has at least 30% support and party C has at least 20% support. Also, we would like to know, given Party A has at least 40% support what is the probability that Party B and Party C has at least 50% support combined. In other words, we would like to know $p(\theta_2 \geq 0.3, \theta_3 \geq 0.2 \mid \theta_1 \geq 0.4)$, $p(\theta_2 + \theta_3 \geq 0.5 \mid \theta_1 \geq 0.4)$ and $p(\theta_2 + \theta_3 \geq \theta_1)$. We may also want to know $p(0.2 \leq \theta_2 \leq 0.4, 0.1 \leq \theta_3 \leq 0.3)$.

Again, it can also be of our interest to know the association between the population parameters, where chi-square test of independence can only give an idea of overall independence between the categorical variables. In that case, we have used correlation matrix of the posterior distribution to measure the association in parameter level. We have also tried to give a notion of a prior based on perceived preference to stop the predominant use of symmetric priors where possible. In our study, we have attempted to provide a detailed analysis of a categorical data and estimated various quantities of interest, such as posterior mean, variance and probabilities involving respective parameters. Bayesian Computation has been done with closed form

integration, Monte-Carlo integration and MCMC methods using R programming language. R codes are made available in the web repository for quicker and easier future implementations.

2. Materials and Methods:

Data Model:

A detail study of categorical data is done by Agresti, A. (2002). We have considered general two-dimensional categorical data, consisting of two categorical variables for our study. A categorical variable has a measurement scale consisting of a set of categories. A general structure of categorical data is given below.

Table 1: General Two-Dimensional Categorical Data with I Rows and J Columns

X \ Y	B_1	B_2	B_3	.	.	.	B_J	Total
A_1	x_{11}	x_{12}	x_{13}	.	.	.	x_{1J}	r_1
A_2	x_{21}	x_{22}	x_{23}	.	.	.	x_{2J}	r_2
A_3	x_{31}	x_{32}	x_{33}	.	.	.	x_{3J}	r_3
.								
.								
.								
A_I	x_{I1}	x_{I2}	x_{I3}	.	.	.	x_{IJ}	r_I
Total	c_1	c_2	c_3	.	.	.	c_J	n

The data is considered to be a multinomial model. The cell counts (x_{ij}) follow a multinomial distribution, where the unknown parameters are the population proportions(θ_{ij}).

In the case of general $I \times J$ tables, if x_{ij} $(i = 1,2, \dots, I, j = 1,2, \dots . J)$ denotes the observed cell counts, with $r_i = \sum_j x_{ij}$ is the row total $c_j = \sum_i x_{ij}$ is the column total and $n = \sum_i \sum_j x_{ij}$ is the grand total, then the Multinomial likelihood is

$f(x|\theta) = \frac{n!}{\prod_i \prod_j x_{ij}} \prod_i \prod_j \theta_{ij}^{\,x_{ij}}$ and $\sum_i \sum_j \theta_{ij} = 1$

Also, the conjugate prior (Gelman et al, 2002) for the proportion parameter vector $\theta = (\theta_{ij})$is a multivariate generalization of Beta distribution known as Dirichlet $(\alpha_{ij})$with $\alpha_{ij} > 0$ and density function is

$p(\theta) = \frac{\Gamma\alpha}{\prod_i \prod_j \Gamma(\alpha_{ij})} \prod_i \prod_j \theta_{ij}^{\,\alpha_{ij}-1}$, where $\alpha = \sum_i \sum_j \alpha_{ij}$, $\sum_i \sum_j \theta_{ij} = 1$ and $\alpha_{ij} > 0 \; \forall \; i, j$

As mentioned before, posterior is prior times likelihood,

$$\therefore \; \Pi(\theta|x) \propto f(x|\theta).p(\theta)$$

$$= \prod_{i=1}^{I} \prod_{j=1}^{J} {\theta_{ij}}^{x_{ij}+\alpha_{ij}-1}$$

$$\therefore \Pi(\theta|x) \sim Dirichilet\ (x_{11}+\alpha_{11}, x_{12}+\alpha_{12}, \dots, x_{IJ}+\alpha_{IJ})$$

Now, Jeffreys noninformative prior is,

$$p(\theta) \propto |J(\theta)|^{\left(\frac{1}{2}\right)}$$

Where, $J(\theta)$ is the determinant of Fisher Information Matrix.

For a multinomial model $J(\theta)$ is given by

$$J(\theta) = \frac{n^{I \times j-1}}{\prod_{i=1}^{I} \prod_{j=1}^{J} \theta_{ij}}$$

$$\therefore p(\theta) \propto |J(\theta)|^{\left(\frac{1}{2}\right)}$$

$$= \prod_{i=1}^{I} \prod_{j=1}^{J} {\theta_{ij}}^{\frac{1}{2}-1}$$

$$\therefore \alpha_{ij} = \frac{1}{2}\ \forall\ i = 1,2, \dots, I\ \forall\ j = 1,2, \dots, J$$

For uniform prior $\alpha_{ij} = 1\ \forall\ i = 1,2, \dots, I\ \forall\ j = 1,2, \dots, J$ and in general $\alpha_{ij} > 0\ \forall\ i, j$

Preference Based Prior:

Predominant use of symmetric prior in every problem sometimes limits the scope of Bayesian Inference. Alternatively, we can order the parameters of the prior distribution based on preference or choice. If there are m parameters in the prior distribution, we can form k groups ($k \leq m$) of size $m_1, m_2, \dots, m_k$, where, $m_1+m_2+\dots+m_k = m$ and order them from least preferred to highly preferred. So, higher preferred groups will be assigned more weights than less preferred groups. But, this is highly context based and will vary for different categorical variables. The preference or choice can be formed based on the type of the problem, relevant assumptions or some prior knowledge.

For example, if we have data on income groups (levels: low, medium, high) and expenditure (levels: low, medium, high), then the Dirichlet prior will have 9 parameters. From the problem, it can be easily perceived that diagonal cells are expected to have higher weights than the others. This assumption can be incorporated in the prior construction. We can form three groups viz. higher, moderate and low.

Table 2: General Structure of Income-Expenditure Data

		Expenditure		
		Low	Medium	High
Income	Low			
	Medium			
	High			

Hyper-parameters corresponding to the diagonal cells, α_1, α_5, α_9 can be assigned higher values, hyper-parameters corresponding to low income with high expenditure and high income with low expenditure α_3, α_7 can be assigned less values and the rest of the hyper-parameters can be assigned moderate values.

Higher Group: (α_1, α_5, α_9)

Lower Group: (α_3, α_7)

Moderate Group: (α_2, α_4, α_6, α_8)

In that way, the prior construction becomes more meaningful and the analysis can yield relevant results.

Joint and Marginal Distributions:

Now we would like to derive the marginal and joint distributions of θ_{ij}'s.

Let us consider the pdf of Dirichlet distribution for 3 cases for quicker understanding.

$$f(y|\alpha) = \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}{y_3}^{\alpha_3-1}$$

$$\therefore f(y) = \int_{y_2=0}^{1-y_1} \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}(1-y_1-y_2)^{\alpha_3-1}\, dy_2$$

$$= \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3} {y_1}^{\alpha_1-1} \int_{y_2=0}^{1-y_1} {y_2}^{\alpha_2-1}(1-y_1-y_2)^{\alpha_3-1}\, dy_2$$

$= \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3} {y_1}^{\alpha_1-1} \int_0^1 (1-y_{1)}{}^{\alpha_2-1} u^{\alpha_2-1}(1-y_1)^{\alpha_3-1}(1-u)^{\alpha_3-1}(1-y_1)du$,

By using the substitution $(1-y_1)\mathrm{u} = y_2$

$$= \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3} {y_1}^{\alpha_1-1}(1-y_1)^{\alpha_2+\alpha_3-1} \int_0^1 u^{\alpha_2-1}(1-u)^{\alpha_3-1}du$$

$$= \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3} {y_1}^{\alpha_1-1}(1-y_1)^{\alpha_2+\alpha_3-1} Beta(\alpha_2,\alpha_3)$$

$$= \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma(\alpha_2+\alpha_3)} {y_1}^{\alpha_1-1}(1-y_1)^{\alpha_2+\alpha_3-1}$$

$\therefore\ y_1 \sim Beta(\alpha_1, \alpha_2+\alpha_3)$

In general, if $f(x|\theta) \sim Dirichlet(\alpha_1, \alpha_2, \dots, \alpha_k)$

Then $X_i \sim Beta(\alpha_i, \sum_{\substack{j=1 \\ i \neq j}}^{k} \alpha_j)$

∴ Marginal distributions of $\theta_{ij}|\alpha \sim Beta(\alpha_{ij} + x_{ij}, \sum_{\substack{l=1 \\ l \neq i}}^{I} \sum_{\substack{m=1 \\ m \neq j}}^{J} \alpha_{lm} + x_{lm})$

Now we shall derive the joint distribution of these variables.

Let us consider the pdf of Dirichlet distribution for 4 cases.

$$\therefore f(y|\alpha) = \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma\alpha_3\Gamma\alpha_4} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}{y_3}^{\alpha_3-1}{y_4}^{\alpha_4-1}$$

$$f(y_1, y_2) = \int_0^{1-y_1-y_2} \frac{\Gamma\alpha}{\prod_{i=1}^{4}\Gamma\alpha_i} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}{y_3}^{\alpha_3-1}(1-y_1-y_2-y_3)^{\alpha_4-1}dy_3$$

$$= \frac{\Gamma\alpha}{\prod_{i=1}^{4}\Gamma\alpha_i} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1} \int_0^{1-y_1-y_2} {y_3}^{\alpha_3-1}(1-y_1-y_2-y_3)^{\alpha_4-1}dy_3$$

$= \frac{\Gamma\alpha}{\prod_{i=1}^{4}\Gamma\alpha_i} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1} \int_0^1 (1-y_1-y_2)^{\alpha_3-1}\, u^{\alpha_3-1}[1-y_1-y_2-(1-y_1-y_2)u](1-y_1-y_2)du,$

By using the substitution $(1-y_1-y_2)\mathrm{u} = y_3$

$$= \frac{\Gamma\alpha}{\prod_{i=1}^{4}\Gamma\alpha_i} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}(1-y_1-y_2)^{\alpha_3-1}(1-y_1-y_2) \int_0^1 [(1-y_1-y_2)\,(1-u)]^{\alpha_4-1}u^{\alpha_3-1}du$$

$$= \frac{\Gamma\alpha}{\prod_{i=1}^{4}\Gamma\alpha_i} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}(1-y_1-y_2)^{\alpha_3-1}(1-y_1-y_2)(1-y_1-y_2)^{\alpha_4-1} \int_0^1 u^{\alpha_3-1}(1-u)^{\alpha_4-1}du$$

$$= \frac{\Gamma\alpha}{\prod_{i=1}^{4}\Gamma\alpha_i} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}(1-y_1-y_2)^{\alpha_3+\alpha_4-1} \frac{\Gamma\alpha_3\Gamma\alpha_4}{\Gamma(\alpha_3+\alpha_4)}$$

$$= \frac{\Gamma\alpha}{\Gamma\alpha_1\Gamma\alpha_2\Gamma(\alpha_3+\alpha_4)} {y_1}^{\alpha_1-1}{y_2}^{\alpha_2-1}(1-y_1-y_2)^{\alpha_3+\alpha_4-1}$$

$$\therefore f(y_1, y_2) \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3+\alpha_4)$$

In general,

If $(y_1, y_2, \dots, y_k) \sim Dirichlet(\alpha_1, \alpha_2, \dots, \alpha_k)$

Then $\left(y_1, y_2, \dots, y_j\right) \sim Dirichlet\left(\alpha_1, \alpha_2, \dots, \alpha_j, \alpha - \sum_{j<k}\ \alpha_j\right), 1 \leq j < k$

Bayes Factor for I × J multinomial model

The pervasive inferential problem related to a categorical data summarized in contingency tables is testing the statistical independence of two categories of the categorical data. Model H_0 corresponds to the null hypothesis that there is no association between the two categories whereas Model H_1 takes that there is an association between the categories constituting I x J contingency table.

Then under H_0, the prior distribution $\pi_0(\theta)$ for the parameter $\theta = (\theta_{ij})$ is based on the law of independence $\theta_{ij} = \Pi_i \Psi_j$ where

$$\Pi_i = \text{Dirichlet}(\gamma_i) \text{ and } \Psi_j = \text{Dirichlet}(\delta_j)$$

Also, for the prior $\pi_1(\theta)$ for model H_1 is $\theta = (\pi_{ij}) \sim$ Dirichlet (α_{ij}). Hence the marginal likelihood under the model M_t (t = 0, 1) is $p(X \mid H_t) = \int f(X \mid \theta)\, \pi_t(\theta)\, d\theta$

After suitable integration,

$$p(X \mid H_1) = \frac{n!}{\Pi\Pi x_{ij}!} \frac{\Pi\Pi \Gamma(n_{ij} + \alpha_{ij})}{\Gamma(n + \alpha)} \frac{\Gamma(\alpha)}{\Pi\Pi\Gamma(\alpha_{ij})}$$

$$p(X \mid H_0) = \frac{n!}{\Pi\Pi x_{ij}!} \frac{\Gamma(\gamma)}{\prod \Gamma(\gamma_i)} \frac{\Gamma(\delta)}{\prod \Gamma(\delta_j)} \frac{\Pi\Gamma(r_i + \gamma_i)\, \Pi\Gamma(c_j + \delta_j)}{\Gamma(n + \gamma)\, \Gamma(n + \delta)},$$

$$\text{where } \gamma = \Sigma\gamma_i ; \delta = \Sigma\delta_j$$

Hence, the Bayes factor for comparing these two models is

$$B_{01} = \frac{p(X \mid H_0)}{p(X \mid H_1)}$$

$$= \frac{\Pi\Pi\Gamma(\alpha_{ij})\, \Gamma\gamma\, \Gamma\delta\, \Gamma(n + \alpha)\, \Pi\Gamma(r_i + \gamma_i)\, \Pi\Gamma(c_j + \delta_j)}{\Pi\Pi\Gamma(n_{ij} + \alpha_{ij})\, \Gamma\alpha\, \Pi\Gamma\gamma_i \Pi\Gamma\delta_j\, \Gamma(n + \gamma)\, \Gamma(n + \delta)}$$

However computing B_{01} on log scale will alleviate the problem of overflow that may occur if it is computed directly. Kass and Raftery (1995) have provided appropriate guidelines for interpreting B_{01} and $\log(B_{01})$ as the degree of evidence for H_0 and is as follows;

- $1 < B_{01} < 3$ indicates 'H_0 is not worth more than a bare mention'
- $3 < B_{01} < 20$ indicates 'H_0 is positive'
- $20 < B_{01} < 150$ indicates 'strong evidence for H_0'
- $150 < B_{01}$ indicates 'very strong evidence for H_0'

Point Estimate:

Now let us find the point estimates of the unknown parameters.

We would like to consider the mean and variance of the marginal distributions for the point estimates. Similarly, median or any quantile measure can also be taken.

Marginal distributions of $\theta_{ij}|\alpha \sim Beta(\alpha_{ij} + x_{ij}, \sum_{\substack{l=1 \\ l\neq i}}^{I} \sum_{\substack{m=1 \\ m\neq j}}^{J} \alpha_{lm} + x_{lm})$

We know that if $X \sim Beta(a,b), E(X) = \frac{a}{a+b}, Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Credential Interval:

For a specific α, credential intervals can be obtained as follows.

$$p[\theta_L < \theta < \theta_U] = 1 - \alpha$$
$$\therefore p[\theta > \theta_U] = \frac{\alpha}{2}$$
$$and\ p[\theta < \theta_L] = \frac{\alpha}{2}$$
$$\Rightarrow 1 - p[\theta > \theta_L] = \frac{\alpha}{2}$$

Using these two equations we can find out the lower and upper bounds.

Now marginal distributions of θ_{ij}'s follow beta distribution. So, for computational convenience, we can make use of a transformation.

If $X \sim Beta(a,b)$,

Then $\frac{b}{a}(\frac{X}{1-X}) \sim F(2a, 2b)$

Now,

$$p[\theta_L < \theta < \theta_U] = 1 - \alpha$$
$$\Rightarrow p\left[\frac{b}{a}\frac{\theta_L}{1-\theta_L} < \frac{b}{a}\frac{\theta}{1-\theta} < \frac{b}{a}\frac{\theta_U}{1-\theta_U}\right] = 1 - \alpha$$
$$\Rightarrow p\left[F_L < F_{2a,2b} < F_U\right] = 1 - \alpha$$

$$\therefore p\left[F_{2a,2b} > F_U\right] = \frac{\alpha}{2}$$
$$and\ p\left[F_{2a,ab} < F_L\right] = \frac{\alpha}{2}$$
$$\Rightarrow p\left[\frac{1}{F_{2a,2b}} > \frac{1}{F_L}\right] = \frac{\alpha}{2}$$
$$\Rightarrow p\left[F_{2b,2a} > \frac{1}{F_L}\right] = \frac{\alpha}{2}$$

Now using F table, we can calculate the lower and upper bounds.

Comparative Analysis:

The correlation matrix of the posterior distribution is given by,

$$\Sigma_{n\times n} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & r_{33} & \dots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & r_{nn} \end{bmatrix}$$

Now, we have correlation coefficient of any pair of θ_{ij}'s by which we get the measure of association between the parameters. And if the correlation coefficient is non-zero then we can certainly say that the parameters are not independent. For a Dirichlet posterior distribution we are expected to have negative covariance. But we can measure the extent of association using correlation matrix.

Probability from Posterior Distribution:

We have the posterior distribution as well as joint and marginal distributions of θ_{ij}'s. So now we can estimate probabilities like,

i) $p[\theta_{ij} > \theta_{kl}]$

ii) $p[a < \theta_{ij} < b, c < \theta_{kl} < d]$

iii) $p[a < \theta_{ij} < b | c < \theta_{kl} < d]$

iv) $p\left[a < \theta_{ij} < b, c < \theta_{kl} < d | e < \theta_{mn} < f\right]$

v) $p\left[\theta_{ij} + \theta_{kl} > a | \theta_{mn} > b\right]$

vi) $p\left[\theta_{ij} + \theta_{kl} > \theta_{mn}\right]$

and many more.

These probabilities can be estimated using closed form integrations. But sometimes, depending on the problem and size of the data, it may not be convenient to carry on closed form integration. We face difficulty to evaluate the integral using R. In that case Monte-Carlo integration or MCMC methods are lot easier to apply. Let us first discuss the probabilities in the light of closed form integration.

Closed Form Integration:

Now let us say, $(\theta_{11}, \theta_{21}) \sim \text{Dirichlet (p, q, r)}$

$$\therefore p[\theta_{11} > \theta_{21}] = \int_{\theta_{21}=0}^{\theta_{21}=0.5} \int_{\theta_{11}=\theta_{21}}^{\theta_{11}=1-\theta_{21}} f(\theta_{11}, \theta_{21}) d\theta_{11} d\theta_{21}$$

$$p[a < \theta_{11} < b, c < \theta_{21} < d] = \int_{\theta_{21}=c}^{\theta_{21}=d} \int_{\theta_{11}=a}^{\theta_{11}=b} f(\theta_{11}, \theta_{21}) d\theta_{11} d\theta_{21}$$

$$p[a < \theta_{11} < b | c < \theta_{21} < d] = \frac{\int_{\theta_{21}=c}^{\theta_{21}=d} \int_{\theta_{11}=a}^{\theta_{11}=b} f(\theta_{11}, \theta_{21}) d\theta_{11} d\theta_{21}}{\int_c^d f(\theta_{21}) d\theta_{21}}$$

Let us say, $(\theta_{11}, \theta_{21}, \theta_{31}) \sim \text{Dirichlet (p, q, r, s)}$

$$\therefore p\left[a < \theta_{ij} < b, c < \theta_{kl} < d | e < \theta_{mn} < f\right] = \frac{\int_{\theta_{31}=e}^{\theta_{31}=f} \int_{\theta_{21}=c}^{\theta_{21}=d} \int_{\theta_{11}=a}^{\theta_{11}=b} f(\theta_{11}, \theta_{21}, \theta_{31}) d\theta_{11} d\theta_{21} d\theta_{31}}{\int_e^f f(\theta_{31}) d\theta_{31}}$$

$$p[\theta_{11} + \theta_{21} > a | \theta_{31} > b] = \frac{\int_{\theta_{31}=b}^{\theta_{31}=1} \int_{\theta_{21}=a}^{\theta_{21}=1-\theta_{31}} \int_{\theta_{11}=a-\theta_{21}}^{\theta_{11}=1-\theta_{21}-\theta_{31}} f(\theta_{11}, \theta_{21}, \theta_{31}) d\theta_{11} d\theta_{21} d\theta_{31}}{\int_b^1 f(\theta_{31}) d\theta_{31}}$$

$$p[\theta_{11} + \theta_{21} > \theta_{31}] = \int_{\theta_{31}=0}^{\theta_{31}=0.5} \int_{\theta_{21}=\theta_{31}}^{\theta_{21}=1-\theta_{31}} \int_{\theta_{11}=\theta_{31}-\theta_{21}}^{\theta_{11}=1-\theta_{21}-\theta_{31}} f(\theta_{11}, \theta_{21}, \theta_{31}) d\theta_{11} d\theta_{21} d\theta_{31}$$

These integrations can be easily done using R programming languages. The programming codes are attached in the Appendix section.

Monte-Carlo Integration:

In Monte-Carlo integration, we simulate a large number of random samples from the given function and see how many of them fall under the required region. So, that ratio multiplied by the total area gives the value of the integration. For our problem, we are going to simulate random samples from the posterior distribution and see how many of them fall under the given region. For a probability distribution, the total area is always 1. So, the ratio will give the required probability.

This procedure can be done with some simple programming. The code is attached in the Appendix section.

MCMC Methods:

To apply MCMC method we do not need supply the posterior distribution. Markov chain Monte Carlo draws those samples by running a cleverly constructed Markov chain over a long period.

Robert, C. P. (2004) has discussed the simulation process in detail.

There are two basic methods of MCMC:

Gibbs sampler is a technique for generating random variables from a marginal distribution indirectly, without having to calculate the density, but it sequentially samples from the collection of full conditional distributions. In addition to an impact and theory, these calculation methodologies focus on the statistical aspects of a problem, thereby freeing statisticians from dealing with complicated calculations.

The second method is applicable when simulation from the full conditionals becomes difficult. The Metropolis-Hastings algorithm simulates from a different Markov chain, having some other stationary distributions, but then modifies it in such a way that a new Markov chain is constructed having the posterior as its stationary distribution.

In our problem, the simulations have been done using R programming language. And like before the probabilities have been calculated with the ratio. The code is attached in the Appendix section.

Now, let us consider a real-life problem and see how these techniques can be useful.

3. Data Analysis:

A survey was conducted by GSS to know the political affiliation of the people of united states in the year 2016. The data is given below.

Table 3: Political Affiliation of People of USA, 2016

	Strong Democrat	Strong Republican	Independent
Male	5203	3966	5610
Female	3862	1685	3288

The data model is Multinomial, and the conjugate prior is Dirichlet. According to Jeffrey's non-informative prior α_{ij}= 0.5 for all i=1,2 and all j= 1,2,3. The posterior distribution is also Dirichlet.

$$\therefore\ (\theta|\alpha) \sim Dirichlet(5203.5,3966.5,5610.5,3862.5,1685.5,3288.5)$$

Marginal distributions of θ_{ij}'s are Beta distribution.

Point Estimate:

Table 4: Point Estimates for The Proportion Parameters ($\boldsymbol{\theta_{ij}}$; $\boldsymbol{i = 1, 2; j = 1, 2, 3}$) in The Multinomial Model Related to GSS Data Using Three Different Methods and Three Different Priors

Parameters	Jeffrey	Uniform	Arbitrary
Closed Form:			
$\boldsymbol{\theta_{11}}$	0.2203	0.2203	0.2203
$\boldsymbol{\theta_{12}}$	0.1679	0.1679	0.1679
$\boldsymbol{\theta_{13}}$	0.2375	0.2375	0.2375
$\boldsymbol{\theta_{21}}$	0.1635	0.1635	0.1635
$\boldsymbol{\theta_{22}}$	0.0713	0.0713	0.0713
$\boldsymbol{\theta_{23}}$	0.1392	0.1392	0.1392
Monte-Carlo Simulation:			
$\boldsymbol{\theta_{11}}$	0.2203	0.2203	0.2203
$\boldsymbol{\theta_{12}}$	0.1679	0.1680	0.1680
$\boldsymbol{\theta_{13}}$	0.2376	0.2376	0.2376
$\boldsymbol{\theta_{21}}$	0.1635	0.1635	0.1635
$\boldsymbol{\theta_{22}}$	0.0714	0.0714	0.0714
$\boldsymbol{\theta_{23}}$	0.1392	0.1392	0.1392
MCMC Methods:			
$\boldsymbol{\theta_{11}}$	0.2203	0.2203	0.2203
$\boldsymbol{\theta_{12}}$	0.1679	0.1679	0.1679
$\boldsymbol{\theta_{13}}$	0.2375	0.2375	0.2375
$\boldsymbol{\theta_{21}}$	0.1635	0.1635	0.1635
$\boldsymbol{\theta_{22}}$	0.0713	0.0713	0.0713
$\boldsymbol{\theta_{23}}$	0.1392	0.1392	0.0713

Interval Estimate:

Table 5: 95% Confidence Intervals for The Proportion Parameters ($\boldsymbol{\theta_{ij}}$; $\boldsymbol{i = 1, 2; j = 1, 2, 3}$) in The Multinomial Model Related to GSS Data Using Three Different Methods and Three Different Priors

Parameters	Jeffrey	Uniform	Arbitrary
Closed Form:			
$\boldsymbol{\theta_{11}}$	(0.2152,0.2256)	(0.2152,0.2256)	(0.2152,0.2256)
$\boldsymbol{\theta_{12}}$	(0.1633,0.1727)	(0.1633,0.1727)	(0.1633,0.1727)
$\boldsymbol{\theta_{13}}$	(0.2323,0.2430)	(0.2323,0.2430)	(0.2323,0.2430)
$\boldsymbol{\theta_{21}}$	(0.1590,0.1683)	(0.1590,0.1683)	(0.1590,0.1683)
$\boldsymbol{\theta_{22}}$	(0.0682,0.0747)	(0.0682,0.0747)	(0.0682,0.0747)
$\boldsymbol{\theta_{23}}$	(0.1349,0.1437)	(0.1350,0.1437)	(0.1349,0.1437)
Monte-Carlo Simulation:			
$\boldsymbol{\theta_{11}}$	(0.2151,0.2256)	(0.2151,0.2256)	(0.2150,0.2256)
$\boldsymbol{\theta_{12}}$	(0.1632,0.1728)	(0.1632,0.1727)	(0.1632,0.1728)
$\boldsymbol{\theta_{13}}$	(0.2322,0.2430)	(0.2321,0.2430)	(0.2321,0.2430)
$\boldsymbol{\theta_{21}}$	(0.1589,0.1683)	(0.1589,0.1683)	(0.1589,0.1683)
$\boldsymbol{\theta_{22}}$	(0.0681,0.0747)	(0.0681,0.0747)	(0.0681,0.0747)
$\boldsymbol{\theta_{23}}$	(0.1349,0.1437)	(0.1349,0.1437)	(0.1348,0.1437)
MCMC Methods:			
$\boldsymbol{\theta_{11}}$	(0.2151, 0.2257)	(0.2150,0.2257)	(0.2150, 0.2257)
$\boldsymbol{\theta_{12}}$	(0.1632, 0.1728)	(0.1632, 0.1727)	(0.1632, 0.1727)
$\boldsymbol{\theta_{13}}$	(0.2321, 0.2430)	(0.2322, 0.2430)	(0.2322, 0.2430)
$\boldsymbol{\theta_{21}}$	(0.1588, 0.1683)	(0.1589, 0.1682)	(0.1589, 0.1682)
$\boldsymbol{\theta_{22}}$	(0.0680, 0.0746)	(0.0681, 0.0747)	(0.0681, 0.0747)
$\boldsymbol{\theta_{23}}$	(0.1348, 0.1436)	(0.1349, 0.1437)	(0.1349, 0.1437)

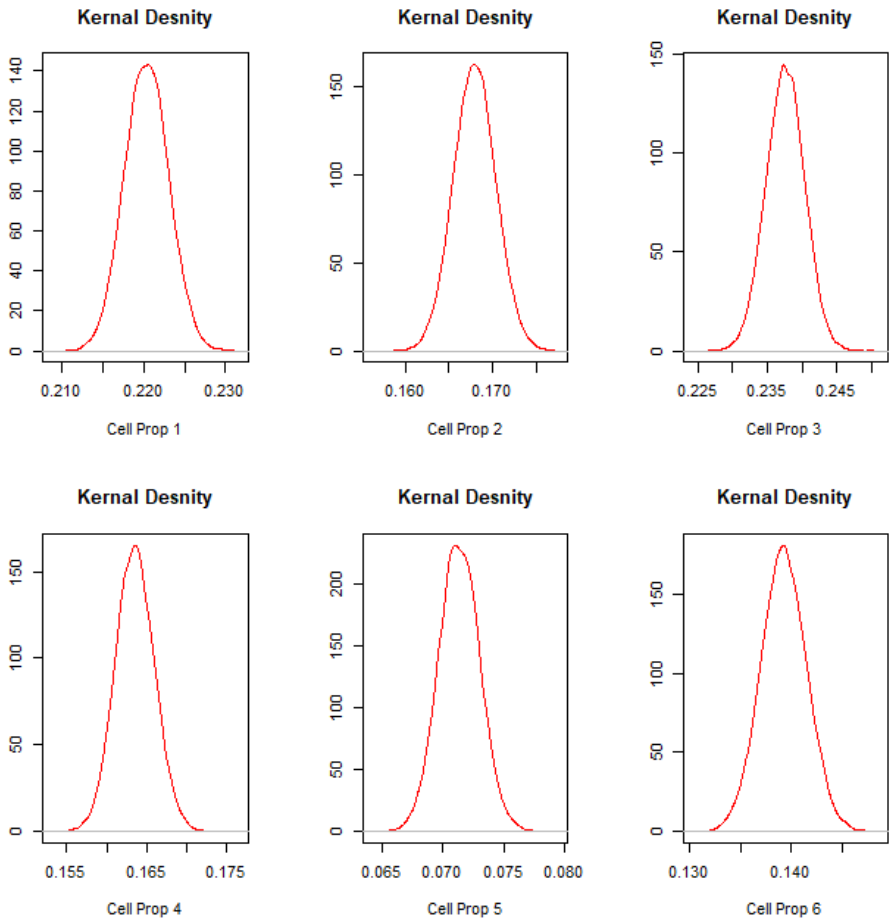

Figure 1, Kernel Density of The Marginal Distributions

Comparative Analysis:

The correlation matrix of the posterior distribution is given by,

$\sum_{5\times5}=$

$$\begin{pmatrix} 1. & -0.238834 & -0.296733 & -0.235061 & -0.14737 \\ -0.238834 & 1. & -0.250786 & -0.198664 & -0.124551 \\ -0.296733 & -0.250786 & 1. & -0.246824 & -0.154745 \\ -0.235061 & -0.198664 & -0.246824 & 1. & -0.122583 \\ -0.14737 & -0.124551 & -0.154745 & -0.122583 & 1. \end{pmatrix}$$

Probability from Posterior Distribution:

For our chosen problem, the values of the gamma functions are too large to be evaluated in R and hence the integration also gives some error. But for a smaller data closed form integration seems to work smoothly. Now let us list out the results using Monte-Carlo integration and MCMC methods.

Table 6: Table Showing, Probabilities from Posterior Distribution

Probabilities from Posterior Distribution	Monte-Carlo Integration	MCMC Methods
$\boldsymbol{p[\theta_1 > \theta_2]}$	1.0000	1.0000
$\boldsymbol{p[\theta_4 > \theta_5]}$	1.0000	1.0000
$\boldsymbol{p[\theta_1 > \theta_4]}$	1.0000	1.0000
$\boldsymbol{p[\theta_2 > \theta_5]}$	1.0000	1.0000
$\boldsymbol{p[\theta_3 > \theta_6]}$	1.0000	1.0000
$\boldsymbol{p[\theta_2 + \theta_3 > \theta_1]}$	1.0000	1.0000
$\boldsymbol{p[\theta_5 + \theta_6 > \theta_4]}$	1.0000	1.0000
$\boldsymbol{p[\theta_2 + \theta_3/5 > \theta_1]}$	0.0845	0.1175
$\boldsymbol{p[\theta_2 + \theta_3/4 > \theta_1]}$	0.9747	0.9545
$\boldsymbol{p[\theta_5 + \theta_6/1.66 > \theta_4]}$	0.0040	0.0096
$\boldsymbol{p[\theta_5 + \theta_6/1.428 > \theta_4]}$	0.9508	0.9295
$\boldsymbol{p[\theta_2 + \theta_3 > 0.40 \mid \theta_1 > 0.23]}$	0.4800	0.5000
$\boldsymbol{p[\theta_5 + \theta_6 > 0.21 \mid \theta_4 > 0.16]}$	0.5780	0.5735
$\boldsymbol{p[0.16 < \theta_1 < 0.32, 0.16 < \theta_4 < 0.32]}$	0.9310	0.9263
$\boldsymbol{p[0.16 < \theta_1 < 0.32 \mid 0.16 < \theta_4 < 0.32]}$	1.0000	1.0000
$\mathbf{p}[\boldsymbol{\theta_2 > 0.16, \theta_3 > 0.2 \mid \theta_1 > 0.2}]$	0.9996	0.9994

4. Discussion:

From table 4 and table 5 it can be seen that point and interval estimates obtained by closed form, Monte-Carlo Integration and MCMC methods don't show much difference even using different priors. Simulations have been done several times, but again the results don't differ much, and it was consistent. So, the three methods give stable results when it comes to point and interval estimates. And non-informative priors don't influence much the result in this problem.

The correlation matrix shows relevant association between the parameters. The correlation between (θ_{11}, θ_{13}), (θ_{12}, θ_{13}), (θ_{13}, θ_{21}), (θ_{11}, θ_{21}) are little high. So, change in population proportions of those categories will affect significantly to the corresponding categories and the extent is directly interpretable from the result. One interesting result that came out from the

analysis is that, if the proportion of male independent voters increases or decreases then the proportion of female democrat voters decreases or increases significantly. And the reason is left to be found out by the experts of the concerned domain. So, the association between parameters can reveal meaningful results regarding the population. When dimension of the categorical data increases and chi square test of independence shows some limitation, correlation matrix of the posterior distribution can give some idea about the associations between the parameters.

For the above problem we have estimated several types of probabilities from the posterior distribution. Depending on the problem many more probabilistic questions regarding the parameters can be formed and those can be directly answered by the above methods. Our analysis shows some interesting results regarding the problem. From table 6, it can be seen that, it is almost certain that the proportion of male democrats being more than male republicans or female democrats. And proportion of female democrats being more than female republicans is also almost certain. Again, proportion of male republicans being more than female republicans and proportion of male independent voters being more than female independent voters are also certain. Now, if 20% of male independent voters, vote for republicans, then also probability of losing of democrats are very less. But, if 25% male independent voters, vote for republicans then the probability of losing of democrats become very high. Similarly, if 60% of female independent voters vote for republicans, probability of losing of democrats is very less. But, if 70% female independent voters, vote for republicans then the probability of losing of democrats becomes very high. So, even though the probability of winning of democrats seems almost certain, a little shift of independent voters towards republicans can totally reverse the scenario.

Joint and conditional probabilities obtained from the posterior probability is also showing some meaningful results. From table 6, it can be seen that given the male proportion of democrats is more than 23%, the probability of having proportion of male republican and independent voters more than 40% is not so high. But the good news for the democrats is that the probability of the proportions of male democrats and female democrats both being in the range 16% to 32% is very high. And that is again confirmed by the probability of proportions of male democrats being in the range 16% to 32% given that female proportion of democrats is in the range 16% to 32%: which is almost certain.

Conclusion:

The study has attempted to give an alternative approach to carry on analysis of a categorical data from a Bayesian perspective. Using these techniques, we can go deeper in a categorical data and reveal results directly related to the population parameters. In Bayesian inference, it is possible to form different types of probabilistic questions regarding the population parameters, and the answers can serve the purpose of comparative analysis, where chi square test impendence can only give some idea about the overall independence of the categorical variables. Again, when the dimension of the data increases, chi square test of independence may have limitations, but correlation matrix of the posterior distribution can give some idea about the association between the parameters. Our objective was to do the whole exercise using

closed form integration, Monte-Carlo Integration and MCMC methods. But, for large data, closed form integration becomes troublesome and it is easier to use other two methods. The study has shown the analysis of two-dimensional categorical data, but it can be extended to higher dimensions using same techniques as the underlying data model will still be Multinomial. This approach can also be applied for other data models, especially multivariate normal models, but the exercise has not been shown in this study. We leave that exercise for future study. Krushke (2010) showed his concern about the need to construct mildly informed or consensually informed prior distributions rather than objective priors. So,it was also of our interest to stop the practice of using symmetric priors for every problem and an idea of using preference-based priors has been illustrated in this article. These techniques can be very useful in modern day problems, especially in the fields of bio-statistics, social sciences and management studies.

Acknowledgement:

The authors would like to thank the Department of Statistics, University of Madras for providing all necessary support and guidance to carry on this research.

References:

1. Agresti, A. (1992). A survey of exact inference for contingency tables. Statistical science, 131-153.
2. Agresti, A. (2002). Categorical Data Analysis 2/e. New Jersey: John Wiley & Sons, Inc.Berger,J.O. and Delampady, M. (1987). Bayesian testing of precise hypotheses (with discussion). Statistical science, 2, 317-348.
3. Agresti, A., & Hitchcock, D. B. (2005). Bayesian inference for categorical data analysis. Statistical Methods & Applications, 14(3), 297-330.
4. Congdon, P. (2005). Bayesian models for categorical data. John Wiley & Sons.
5. Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2002). Bayesian Data Analysis. London, UK: Chapman & Hall.
6. Gilks, W. R., Richardson, S., &Spiegelhalter, D. J. (1996). Introducing markov chain monte carlo. Markov chain Monte Carlo in practice, 1, 19.
7. Kass, Robert E., and Adrian E. Raftery. "Bayes factors." Journal of the american statistical association 90.430 (1995): 773-795.
8. Kruschke, J.K. (2010). What to believe: Bayesian methods for data analysis, Trends in Cognitive Sciences 14, 293–300.
9. Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press.
10. Nandram, B. and Choi, J.W. (2007). Alternative Tests of Independence in Two-Way Categorical Tables. Journal of Data Science, 5, 217-237.
11. Robert, C. P. (2004). Monte carlo methods. John Wiley & Sons, Ltd.
12. Sangeetha, U., Subbiah, M., Srinivasan, M. R., &Nandram, B. (2014). Sensitivity Analysis of Bayes Factor for Categorical Data with Emphasis on Sparse Multinomial Data. Journal of Data Science, 12(2), 339-357.

13. Subbiah, M. and Srinivasan, M.R. (2008). Classification of 2 × 2 sparse data with zero cells. Statistics & Probability Letters, 78, 3212 – 3215.

Appendix:

All the R codes used in this study can be found in the online repository with the following link.