Bayesian Inference –

Students Teacher Interactive Approach

He said, “Son you are a Bayesian boy,
And that's the way to stay,
Son you'll be a Bayesian boy,
Until your dying day."
- M. Rimmer

Dr.Subbiah M
Visiting Faculty
Department of Statistics,
University of Madras
sisufive@gmail.com
Mobile: +91 9952726863

May 2017

PREFACE

“How absurdly simple!", I cried.
“Quite so!", said he, a little nettled. “Every problem becomes very childish when once it is explained to you."

Arthur Conan Doyle
The Adventure of the Dancing Men
(Inspired from Robert and Casella 2010)

The intended audience of this material is to those have an idea to learn Bayesian Inference. This work is not an attempt to replace a textbook but rather provides a road map for the journey with the Bayesian perspective of Statistical Inference. However, the presentation attempts to compile in an ordered way similar to first four chapters of Gelman et al (2002) or Lee (2012).

Further, preparation of the content is strictly based on the existing syllabus of Bayesian Inference course offered in the Department of Statistics, University of Madras. This course is aimed for the beginners and the material is an outcome of this course from a teaching-learning perspective. Hence, this material covers Student’s projects as illustrative examples based on simple models.

The students of the course (during December 2016 - April 2017) have conceived and organized illustrative problems, and four research students who also attended the course have mentored those students group. Efforts are made to highlight the course with the minimal use of computers. However, when there is a need, students are allowed to exploit the computational facility available to them. Nevertheless, the course is entirely free from any Computational Laboratory practices.

I would like to acknowledge Dr. M.R. Srinivasan, Dr. S. Sampath, Mrs. M.R. Sindhumol, Dr. M. Ramadurai and Dr. S. Suresh, the Faculty members of Department of Statistics,for providing the opportunity to teach this elective course. This is an opportunity to thank Dr K Suresh Chandra and Dr J Ranganathan for their eternal and persistent encouragement. I thank Harini S, Sripriya TP, Gokul T and Ghouse Basha MA, research scholars for their patience and help. Of course, the students also deserve special appreciations for their involvement in this teaching-learning process. For errors or omissions, I take the responsibility and please let me know if any error is noticed.

Dr. Subbiah M

Contents

Bayes! You're the very best,
Simpler than all the rest,
Just five words as clear as can be:
Post is prior times likelihood. See!
- D. Blackwell

1. INTRODUCTION

Statistical methods are tools of scientific investigation that are controlled by learning process in which various facets and dimension of the problem are illuminated as the study proceeds. A full treatise on the use of statistical methods in thescientific investigation would necessarily include consideration of statistical design as well as analysis. The analysis is concerned with the aspects of computation, developing appropriate models and to draw aninference of relevance through appropriate procedures. Statistical inference is concerned with drawing conclusions, from complete or incomplete information on the data under study.

The ability to reveal boundaries of scientific uncertainty is equivalent to the ability to reveal the sensitivity of conclusions from fixed data to various model specifications, all of which are scientifically acceptable. When sharp conclusions are not possible without obtaining more information, whether it be more data, new theory, or deeper understanding of existing data and theory, then it must be scientifically valuable and appropriate to expose this sensitivity and thereby direct efforts to seek the particular information needed to sharpen conclusions (Rubin, 1984).

The development of modern statistical theory has been characterized by three-sided approaches namely Bayesian, frequentist, and Fisherian with its own merits and demerits and divergent views. In Bayesian inference, known quantities are treated as observed values of random variables and unknown quantities are observed as random variables; the conditional distribution of unknowns given known follow from applying Bayes theorem to the model specifying the joint distribution of known and unknown quantities. The known refers to values that are both available and considered worthwhile to include in model specifications (Rubin, 1984).

Philosophical issues aside, Bayesian approaches provide a powerful tool for interpretation of study results and evaluation of hypotheses and allow considering a much broader class of conceptual and mathematical models than would have been possible using non-Bayesian approaches (Dunson, 2001).In recent years, the Bayesian inference has received more attention in applied and theoretical research. It is perceived by many statisticians as a natural paradigm for solving applied problems.

It is important to note that many of the advantages claimed for Bayesian approach follows from the ability to handle complex models and the three main aspects that reflect Bayesian modeling include; computation, incorporation of historical information, and inference on complex functions of parameters. It has been observed that a major goal of statistics is to find a completely coherent objective Bayesian methodology for learning from data; synthesize and communicate the uncertainties that arise in a specific situation.

Further, theBayesian analysis is considered as the core of statistics in that it conditions properly on the data and correctly processes the uncertainties arising from the wide variety of inputs that may go into a statistical analysis (Berger, 2006).Indeed, the growth and strength of Bayesian methodology lie in its wide applications to complicated problems that are not so obvious even to formulate for a more traditional analysis (Goldstein, 2006 and Wolport, 2004).

This presentation basically aims to exploit the inbuilt advantages of Bayesian approach in statistical inference. It provides a framework for the analysis of simple models without depending on computers, rather a calculus-based approach as an initial step of learning Bayesian inference. This includes the major areas of a Bayesian study design; the choice of appropriate priors and the summarizing strategies involved in the analysis of a problem in practice and few other specific advantages of Bayesian methodologies.

1.1 Phases in Bayesian

Step 1 – Calculus based approach and minimal use of computer or calculator (**Ccc)**.Deriving explicit mathematical structures/ use from the textbooks or related materials.

Step 2 – Computer intensive simulation techniques **(CIST)**-MC Integration and Optimization

Step 3 – **CIST** – Role of MCMC

*The syllabus of the course MSIE126 is confined to **Step 1**.*

1.2 Pre-requsites

Topics	Step 1	Step 2	Step 3
Understanding Probability	High	High	High
Distribution Theory (Mathematical Statistics)	High	High	High
Overview of inference (Purpose of Statistics)	High	High	High
Programming (R, WinBUGS or any other)	Low	High	High

Understanding Probability involves

- Various form of definitions
- Frequency
- Axiomatic
- Degree of belief

***Investigate**:*Which form would be more appealing, why?

This paves the way to unlock the uncertainty and helps to represent the state of knowledge about a problem/situation/scenario. There are different ways to infer and communicate such state of knowledge. Major methods could be

- A deterministic approach: Area of a floor, cost to paint a wall (if price is known)
- A probabilistic approach

Investigate: Area of a floor, theequation of motion $ut+1/2at^2$,cost to paint a wall.

All such questions are directly connected with the knowledge of the problem. For example, if one knows length and breadth of the room then the area can be ***computed*** (with a desired level of accuracy).Here an error due to measurement may be too small to worry. In some cases like whether we would reach our destination in time while boarding a bus/train/airplane etc, the state of knowledge is more extensive and compiling to a scientific approach may be tedious. But, once we gain this knowledge, ***estimating*** our desired quantity may be little easier. This may be more promising requirement to understand one of the ingredients of ***Bayesian Inference – Prior.***

1.3 Prior

Know the problem at hand. Understand the quantity that leads the uncertainty in the problem. Capitalize the knowledge about that quantity. Mathematical description of this knowledge is the final step in this process.

For example, maximum temperature in Chennai during this summer

Think(T): How much do we know about Chennai's weather?

Quantify(Q): At least, is it possible to specify two values within which maximum temperature would fall.

Maths(M): Any idea to communicate step 2 mathematically? Notation of probability would help.

Hint: *A uniform distribution is useful?*

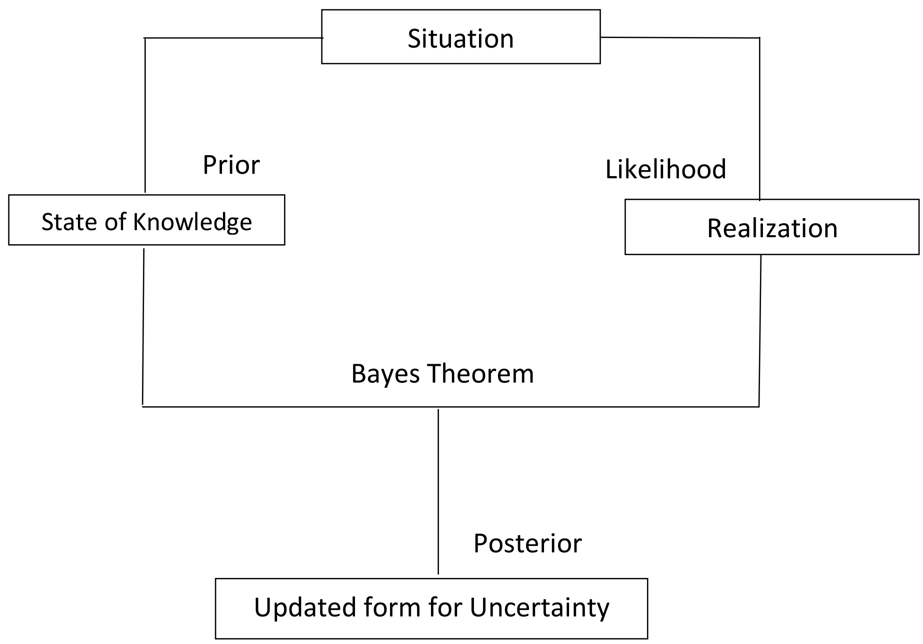
Alternately, **if someone collects the data from the meteorological department, what would be the answer to understand the uncertainty in the problem?**

Another example might be the probability of success of an event, say a power failure in the campus. Discussion for the problem is once again attempted with the above approach – T Q M. So, it is an active discussion to understand the notion of prior – knowledge to distribution. This exercise may be extended to understand the various terms found in books or literature in general

- Objective
- Subjective
- Non-Informative

- Reference
- Flat/Vague

Also, this discussion should connect the ideas drawn from mathematical distribution and statistical inference. Precisely, it is important to know the difference between Discrete and Continuous random variable in terms of nature of parameters, as well as the type of parameters – scale, shape or location. Once this is understood from a conceptual framework to devise a mathematical structure, then the typical road map of Bayesian inference can be extended using Bayes Theorem. More specifically the relation between prior and likelihood may be described as



1.4 Constructing Priors

The mechanism of the Bayesian approach to make inference consists of three basic steps; (i) assign priors to all unknown parameters, (ii) define the likelihood of the data given the parameters and (iii) determine the posterior distribution of the parameters given the data using Bayes' theorem. The ability to include prior information in the model is not only an attractive pragmatic feature of the Bayesian approach and is theoretically vital for a guaranteed coherent inference.

Most of the theoretical works on prior distribution has been on determining the conditions that need to be satisfied by the prior and data distributions so that the posterior distribution is well behaved. Conversely one might seek to avoid theoretical considerations entirely and simply pick a 'subjective' prior distribution that best represents one's scientific knowledge about the set of uncertain parameters in the problem. In practice, however, subjective knowledge is hard to specify precisely, and hence it is important to study the sensitivity of posterior inference.

Two basic interpretations can be given as a general approach for the problem of constructing prior distributions: (i) the prior distribution represents a population of possible parameter values from which the parameter of current interest has been drawn (Informative priors) and (ii) the investigator must express his knowledge and uncertainty about the parameter as if its value could be thought of as a random realization from the prior distribution.

Such prior distributions typically should include all plausible values of θ but the distribution need not be realistically concentrated around the true value. When prior distributions have no population basis, it is difficult to construct such a prior distribution which plays a minimal role in the posterior distribution. Such distributions are sometimes called 'reference prior distributions'. The prior density in this case is described as ***vague, flat, diffuse or non-informative***. The idea for using non-informative prior is based on "Principle of Insufficient Reason" (PIR) and the rationale is often said to be 'to let the data speak for themselves,' so that inference is not affected by information external to the current data.

A prior density $\pi(\theta)$ is ***proper***if it does not depend on data and integrates to 1. It is basic to the concept of a probability density that it integrates to 1, but sometimes it may be useful to extend the concept of a probability density to some cases like $\int \pi(\theta)\, d\theta = \infty$, which are called ***improper*** densities. It turns out that sometimes when an improper prior density is combined with likelihood proper posterior density results; that is $\int p(\theta/x) d\theta$ is finite for all x.

The non-informative and reference prior densities are often useful in many situations such as (i) when it does not seem to be worth the effort to quantify one's real prior knowledge as a probability distribution; (ii) when it is desired to perform the algebraic work to check the posterior density is proper and to determine the sensitivity of posterior inference; (iii) when the data analyst desires to describe the model more conveniently; (iv) to simplify computation or perhaps to avoid using a possibly unreliable source of information.

The property that the posterior distribution follows the same parametric form as the prior distribution is called ***conjugacy***and conjugate family is mathematically convenient to deal with for obtaining the posterior distribution. However, if the information available contradicts the conjugate parametric family, it may be necessary to use a more realistic prior distribution. A class F of prior distributions is said to form a conjugate family if the posterior density $\pi(\theta/x)$ is in class F for all x whenever the prior density is in F. Also if F is a conjugate family and $h(\theta)$ is any fixed function, then the family ψ of densities proportional to $h(\theta)\ \pi(\theta)$ for p belongs to F, is also a conjugate family. Much formal Bayesian analysis concentrates on situations where closed forms are available; such models are sometimes unrealistic, but their analysis often provides a useful starting point when it comes to constructing more realistic models.

The main motivation for using conjugate priors is their tractability of an explicit treatment of posterior distributions and mainly justified on technical grounds rather than for fitting properly the available prior information. Their role is then to provide a first approximation to the adequate prior distribution which should be followed by a robust analysis. They are more justified if considered as a basis for prior information modeling.

The conjugate priors are useful as approximations of the true prior distributions. However, when no prior information is available, their sole justification is analytical as they can lead to closed form expressions for some posterior distributions. On the other hand,non informative priors are to justify the choice of prior distributions on a subjective basis; the prior distributions must be derived from the sample distribution, since this is the only available "information".

The ***Jeffreys prior*** is based on a more intrinsic approach which indeed avoids the need to take the invariance structure into account, even though it is often compatible with it. Priors based on considering one-to-one transformations of the parameter $\eta = g(\theta)$ by transforming variables, as given by the prior density $h(\theta)$ is equivalent in terms of expressing the same beliefs to the prior density on η. The Jeffreys non-informative prior distributions are based on Fisher information given by $\pi(\theta) \propto [J(\theta)]^{1/2}$ where $J(\theta) = -E\left[\frac{d^2}{d\theta^2}\log\pi(x/\theta)\right]$.

However, Jeffreys is mainly emphasizing the use of such prior distributions in the one-dimensional case because Jeffreys non-informative approach may lead to incoherence or even paradoxes. Also, in multi-parameter cases, the choice of a prior for parameters which can be thought of as representing the location and scale it would be reasonable to think of these parameters as being independent a priori as illustrated in the normal case (Lee, 1989, Robert, 1994 and Gelman et al, 1995).

1.5 Posterior Distribution

The way of expressing the beliefs about θ is by taking into accounts both prior beliefs and the data. Though the prior beliefs may differ, there may be a common agreement on the way in which the data are related to θ. This eventually reflects in posterior but will turn out that if enough data is collected, then the posteriors will be less sensitive to the choice of priors; then Bayes theorem encapsulates the technical core of Bayesian inference. That is, if θ, a parameter in a likelihood $f(X \mid \theta)$ and with a prior $p(\theta)$ then Bayes theorem provides the posterior $\pi(\theta \mid X)$ as

$$\pi(\theta \mid X) = \frac{p(\theta)f(X \mid \theta)}{\int p(\theta)f(X \mid \theta)d\theta} \text{ assuming } \theta \text{ as a continuous random variable}$$

The posterior probability distribution contains all the current information about the parameter θ. The desirable numerical summaries of the distribution include; summaries of location such as mean, median and mode(s); variation is commonly summarized by the standard deviation, the interquartile range, and other quantiles. For example,

$$Mean = E(\theta \mid X) = \int \theta\pi(\theta \mid X)d\theta$$

$$Median = 50^{th} \text{ percentile of } \theta \mid X = \int_{-\infty}^{\theta_{0.50}} \pi(\theta \mid X)d\theta = 0.5$$

However, when the prior distribution has a closed form, such as the beta distribution, then summaries (mean, median and standard deviation) of the posterior distribution are often available in closed form expressions.

In addition to point summaries, it is also important to report the posterior uncertainty. The usual approach is to present quantiles of the posterior distribution of estimand of interest or, central Posterior Intervals. A central interval of posterior probability corresponds in the case of a 100(1- α)% interval, to the range of values above and below which lies exactly 100(α/2)% of the posterior probability and such interval estimates are sometimes referred to as aposterior interval. For example, the central 95% posterior interval [a, b] for the parameter θ for which Pr(θ< a) = Pr(θ> b) = 0.025.

This may be related to the following question

(i) $p(\theta < a) = \frac{\alpha}{2}$

(ii) $p(\theta > b) = \frac{\alpha}{2}$

Or, a probability can be computed that involves the parameter θ, say

$$p(\theta > a) = \int_a^\infty \pi(\theta \mid X) d\theta$$

Together provide, $p(a < \theta < b) = 1 - \alpha$ for a specified value of α in (0,1). Hence, the notion of interval estimate can directly be followed.

Calculus-based approach together with a reasonable knowledge of distribution can be followed with simple examples.

- θ : Proportion parameter
- X/θ : Realization of a Bernoulli trail
- θ/X : Direct computation is possible once a suitable prior is chosen
  (Example: Beta prior)

That is,

$$\theta \sim Beta(\alpha, \beta)$$

$$X \mid \theta \sim Bernoulli(\theta)$$

$\theta \mid X \sim Beta(x + \alpha, n - x + \beta)$ where n is the number of realization of the Bernoulli process

Now it is easy to follow that

$E(\theta \mid X) = \frac{x + \alpha}{n + \alpha + \beta}$ and similar other quantities in closed form.

However, while computing probability related questions, an attempt can be made with the relationship between Beta and F distribution. This helps in avoiding computers for the beginners.

Precisely, if Y ~ Beta (a,b) then

$$F = \frac{b}{a}\frac{Y}{1-Y} \sim \mathrm{F}(2\mathrm{a}, 2\mathrm{b})$$

This is useful to compute probabilities such as $p(\theta < \theta_1) = p\left(\frac{b}{a}\frac{\theta}{1-\theta} < \frac{b}{a}\frac{\theta_1}{1-\theta_1}\right)$ and subsequently, probabilities can be obtained by referring F-table

This exercise illustrates a complete road map

1) Setting up prior
2) Writing a likelihood
3) Deriving a posterior
4) Summarizing to answer suitable inferential objectives.

Other quantities that have much of Bayesian flavor are ***sequential update*** and ***predictive distributions***, using both prior and posterior distributions. First one is a sequential use of prior and posterior; that is posterior derived from one step becomes prior to the subsequent step and so on;

Secondly, predictive distribution is about unobserved data while inference has already been obtained from the available data; that is the distribution of a new data point (y), marginalized over the ***posterior*** is given by $p(y/x) = \int f(y/\theta)\ \pi(\theta/x)\ d\theta$ whereas the **prior** predictive distribution is the distribution of a new data point, marginalized over the ***prior*** m(x)=$\int f(x/\theta) p(\theta) d\theta$

1.6 Testing of Hypotheses

Statistical inference on parameter estimation based on Bayesian approach is conceptually straightforward in most of the problems. Once a prior distribution is defined, and with a reasonable likelihood function, inference about parameters of interest is obtained from marginal posterior distributions, and nuisance parameters are integrated.

The role of probability in measuring the uncertainty at each stage of Bayesian estimation process is well established. To arrive at the posterior distribution of possible states of the model, prior distribution of each model is updated using the information contained in the data. Now, the inference is drawn from the entire posterior distribution ofthe most plausible model**.**

However, in practice, it is impossible to have an agreement about the appropriateness of the model to be used, unless there is a well-established theoretical framework or a mechanism underlying the problem. Therefore, it is imperative to take into account uncertainties in the model-building process and so, to start with, a set of competing models has to be considered, with each model viewed as a different state of a random variable.

Johnson (2005) proposed methods based on Bayes factors for modeling the sampling distributions of standard test statistics. The study indicated the possible extensions to test statistics associated with categorical data. However, with the influence of priors and subjectivity involved, sensitivity of results are also significant issues for an active research.

Vanpaemel (2010) has provided an exhaustive list of studies that address the sensitivity of priors and has emphasized the importance of prior and its sensitivity analysis. Particularly Hashemi (1997) and Nandram and Choi (2007), separated by a decade but similar in approach, have motivated the researcher to investigate deeply the Bayes factor for categorical data under multinomial design, which is ubiquitous in many social science survey sampling designs and problems.

1.7 Bayes factor

If there are several competing hypotheses or models about a system, then the set of models can be considered as mutually exclusive and exhaustive. A prior probability $p(H_i)$ $(i=1,2,\dots,N)$ can be assigned to each hypothesis such that $\Sigma p(H_i) = 1$, with N denoting the number of hypotheses**.** After observing data y, the posterior probability of hypothesis H_i is

$$p(H_i \mid y) = \frac{p(H_i)p(y \mid H_i)}{\sum_{i=1}^{N} p(H_i)p(y \mid H_i)}$$

where $p(y|H_i)$ is the marginal density, which is the expected value of all possible likelihoods. Then hypothesis i relative to j is of the form

$$\frac{p(H_i \mid y)}{p(H_j \mid y)} = \frac{\dfrac{p(H_i)p(y \mid H_i)}{\sum_{i=1}^{N} p(H_i)p(y \mid H_i)}}{\dfrac{p(H_j)p(y \mid H_j)}{\sum_{i=1}^{N} p(H_i)p(y \mid H_i)}} = \frac{p(H_i)}{p(H_j)} \frac{p(y \mid H_i)}{p(y \mid H_j)}$$

It could be observed that the posterior odds ratio is the product of the prior odds ratio and theratio of the marginal probabilities under each of the hypotheses. Then Bayes factor B_{ij} is defined as

$$B_{ij} = \frac{p(y \mid H_i)}{p(y \mid H_j)} = \frac{\dfrac{p(H_i \mid y)}{p(H_j \mid y)}}{\dfrac{p(H_i)}{p(H_j)}} = \frac{\text{posterior odds}}{\text{prior odds}}$$

Berger and Delampady (1987), Kass (1993), Bernardo and Smith (1994), Kass and Raftery (1995), Goodman (1999), Delampady and Berger (1990), Lavine and Schervish (1999), and Ghosh et al (2006) provide a better insight into the concept of Bayes factor. The explicit forms of Bayes factor for the multinomial model is derived in the following section.

The interpretation of B_{ij} as per the recommendation of Kass and Raftery (1995) about the degree of evidence for H_i is as follows:

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

Remaining section deals with implementation of these details in practical situations. All illustrations are conceived by the students who have enrolled in the course. Emphasize is to understand Bayesian formulation and computation, mainly calculus based, though the students have involved in setting objectives and data collection (in-person / online). Nevertheless, few computations are carried out using R. Three sets are provided due to space restriction and to avoid redundancy.

I. Bayesian Inference on Categorical Data for An Insurance Problem

Abstract:

A $I \times J$ categorical data follows multinomial model. The conjugate prior distribution is Dirichlet. Posterior probabilities, point and interval estimates have been calculated from the posterior distribution, which also follows Dirichlet distribution. Marginal distributions of the population proportions follow Beta distribution and joint distributions of the population proportions follow Dirichlet distribution. Test of independence has been conducted through Bayes factor. And further comparative analysis between population parameters has been carried out using the marginal and joint distributions.

Keywords: *Bayesian Inference, Categorical Data, Test of Independence, Bayes Factor, Multinomial Model, Conjugate Dirichlet Prior, Comparative Analysis, Joint Distribution, Conditional Probability*

1. Introduction:

Analyzing categorical data is very common in statistical practice. But doing a Bayesian analysis gives some extra scope to explore the data. In a Bayesian analysis, there is a data model and 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 the prior distribution.

We must have a state of knowledge before constructing a prior distribution. The prior can be convenient, informative, non-informative 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. In simple words, we have some information (state of knowledge) about the unknown parameter then we update that information using the data in hand and draw some inference about that unknown parameter.

The procedure is pretty much straight-forward for a univariate data, but for a categorical data, beside drawing inference about the unknown parameter, it is also of our interest to measure association or independence between the factors. Further, we may want to do some comparative study between the unknown parameters.

In this article, we have an insurance data where we have counts on how many people from rural, urban and outskirts going for three different types of insurance plan. To eliminate the factor of population difference between cities, villages and outskirts, we have collected data from 30 people from each category. **The survey was done online in a Facebook group, where the members are from different parts of India and the age group varies from 21-60. A brief description of three different types of insurance products was given. The collected data and further analysis are as follows.**

2. Data

Table Showing Counts for Different Products

	Term Assurance	Pure Endowment	Endowment Assurance
Rural	21	2	7
Urban	10	5	15
Outskirts	8	12	10

3. Methodology

The data is considered to be a multinomial model. The cell counts $\left(x_{ij}\right)$ follow a multinomial distribution, where the unknown parameters are the population proportions of customers having different insurance products.In the case of general $I \times J$ tables, if $x_{ij}(i = 1,2, \ldots, I, j = 1,2, \ldots . J)$ denotes the observed cell counts, with $r_i = \sum_i x_{ij}$is the row total $c_j = \sum_j x_{ij}$ 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 $\theta = \left(\theta_{ij}\right)$ could be a multivariate generalization of Beta distribution known as Dirichlet $\left(\alpha_{ij}\right)$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 $\alpha = \sum_i \sum_j \alpha_{ij}$

As mentioned before, posterior is prior times likelihood,

$$\therefore \Pi(\theta|x) \propto f(x|\theta)$$

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

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

For Jeffrey's non-informative prior,

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

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

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

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

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

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

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

3.1 Marginal Distribution

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

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

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

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

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

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

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

$\therefore$ Marginal distributions of based on the posterior distribution is

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

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 $(1-y_1-y_2)\mathrm{u} = y_3$

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

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

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

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

$\therefore y_1, y_2, \dot{y} \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3 + \alpha_4 = 1-\alpha_1-\alpha_2)$ and $\dot{y} = 1-y_1-y_2$

In general, if $(y_1, y_2, \dots, y_k) \sim Dirichlet(\alpha_1, \alpha_2, \dots, \alpha_k)$ then any r (1 < r < k) combination of y's will have joint Dirichlet distribution; that is

$y_1, y_2, \dots, y_r, \dot{y} \sim Dirichlet\left(\alpha_1, \alpha_2, \dots, \alpha_r, \alpha - \sum_{1<j<r} \alpha_j\right)$ and $\dot{y} = 1 - y_1 - y_2 \dots \dots \dots - y_r$

This form would be helpful in determining desired probabilities that involve the parameters of Multinomial likelihood. Posterior distribution on these parameters provides the scope to compute these probabilities.

3.2 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 $\theta = (\theta_{ij})$ isbased on the law of independence $\theta_{ij} = \Pi_i \Psi_j$ where $\Pi_i = \text{Dirichlet}(\gamma_i)$ 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,

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

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

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'

4. Analysis

4.1 Point Estimate

Now let us find the point estimates of the unknown parameters.We'd 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 $X \sim Beta(a,b), E(X) = \frac{a}{a+b}, Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Table showing posterior mean and variance of the parameters

Parameter	Mean	Variance
$\boldsymbol{\theta_{11}}$	0.2275	0.0018
$\boldsymbol{\theta_{12}}$	0.0265	0.0003
$\boldsymbol{\theta_{13}}$	0.0794	0.0008
$\boldsymbol{\theta_{21}}$	0.1111	0.0010
$\boldsymbol{\theta_{22}}$	0.0582	0.0006
$\boldsymbol{\theta_{23}}$	0.1640	0.0014
$\boldsymbol{\theta_{31}}$	0.0899	0.0009
$\boldsymbol{\theta_{32}}$	0.1323	0.0012
$\boldsymbol{\theta_{33}}$	0.1111	0.0010

4.2 Credible Interval

For a specific γ Credible intervals can be obtained as follows.

$$p[\theta_L < \theta < \theta_U] = 1 - \gamma$$

$$\therefore p[\theta > \theta_U] = \frac{\gamma}{2}$$

$$and\ p[\theta < \theta_L] = \frac{\gamma}{2}$$

$$\Rightarrow 1 - p[\theta > \theta_L] = \frac{\gamma}{2}$$

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,

$$p[\theta_L < \theta < \theta_U] = 1-\gamma$$

$$\Rightarrow p\left[\frac{b}{a}\frac{\theta_L}{1-\theta_L} < \frac{b}{a}\frac{\theta}{1-\theta} < \frac{b}{a}\frac{\theta_U}{1-\theta_U}\right] = 1-\gamma$$

$$\Rightarrow p\left[F_L < F_{2a,2b} < F_U\right] = 1-\gamma$$

$$\therefore p\left[F_{2a,2b} > F_U\right] = \frac{\gamma}{2}$$

$$and\, p\left[F_{2a,ab} < F_L\right] = \gamma$$

$$\Rightarrow p\left[\frac{1}{F_{2a,2b}} > \frac{1}{F_L}\right] = \frac{\gamma}{2}$$

$$\Rightarrow p\left[F_{2b,2a} > \frac{1}{F_L}\right] = \frac{\gamma}{2}$$

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

For our problem let γ = 0.05 then 95% credible interval is

Parameters	**Credible Intervals**	
	Lower Limit	Upper Limit
$\boldsymbol{\theta_{11}}$	0.1492	0.3167
$\boldsymbol{\theta_{12}}$	0.0045	0.0668
$\boldsymbol{\theta_{13}}$	0.0341	0.1414
$\boldsymbol{\theta_{21}}$	0.0563	0.1814
$\boldsymbol{\theta_{22}}$	0.0207	0.1132
$\boldsymbol{\theta_{23}}$	0.0969	0.2446
$\boldsymbol{\theta_{31}}$	0.0413	0.1549
$\boldsymbol{\theta_{32}}$	0.0721	0.2071
$\boldsymbol{\theta_{33}}$	0.0563	0.1814

4.3 Bayes Factor

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

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

$$\therefore B_{01} = 4.7945$$

For our problem, we have calculated the Bayes Factor using R programming language. The code is mentioned at the end of this article. Now, B_{01}= 4.7945 (3< B_{01}< 20) indicates that H_0 is positive. In other words, there is not much support for the model H_0.

4.4 Comparative Analysis

The major advantage of Bayesian inference is the option to compute probabilities that involve the parameters of multinomial distribution. Posterior distribution provides the updated version of those parameters. That is, joint and marginal distributions of θ_{ij}'s will be useful to compute probabilities like,

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

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

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

It is of our interest to compare the population proportion of rural people going for term insurance and urban people going for term insurance.

So, we'd like to know $p[\theta_{11} > \theta_{21}]$

Now $\theta_{11}, \theta_{21}, \dot{\theta} \sim \text{Dirichlet}\ (21.5,10.5,62.5)\ \dot{\theta} = 1 - \theta_{11} - \theta_{21}$

$$\therefore p[\theta_{11} > \theta_{21}] = \int_{\theta_{21}=0}^{\theta_{21}=0.5} \int_{\theta_{11}=\theta_{21}}^{\theta_{11}=1-\theta_{21}} f(\theta_{11}, \theta_{21})\ d\theta_{11} d\theta_{21} = 0.9768$$

So, we see that the probability of proportion of rural people going for term insurance more than urban people is significantly high. We assume that rural people are more concerned about uncertain death due to their profession and socio-economic condition than urban people and that is being reflected in this result.

Similarly, we calculate $p[\theta_{11} > \theta_{23}] = 0.8413$ and we can see that though urban people are more concerned about retirement plans and survival benefit, still the probability of proportion of rural people going for term insurance than urban people going for endowment assurance is quite high. So, the concern of uncertain death and due to that, the unavoidable circumstances of life have turned out to be a serious factor for rural people to choose their insurance product. And we have also calculated $p[\theta_{32} > \theta_{21}] = 0.6645$, $p[\theta_{33} > \theta_{13}] = 0.7659$, $p[\theta_{11} > 0.33, \theta_{23} > 0.33] = 0.1043$

Further, conditional probability is based on the fact that (X,Y) follows Dirichlet Distribution and Y follows Beta Distribution. The formula is Conditional Probability = NR / DR where

NR = P[a< X < b,c < Y < d] = F(b,d) - F(a,d) - F(b,c) + F(a,c) and

DR = P[c < Y < d] = F(d) - F(c)

For example conditional Probability $p[\theta_{11} > 0.33 | \theta_{21} < 0.5] = 0.013$; which is the probability that rural population prefers term assurance with probability at least 0.33 when urban population does so at most 0.5; whereas $p[\theta_{11} > 0.11 \mid \theta_{21} > 0.11] = 0.999$ provides more insight to the preference among rural and urban populations.

5. Summary

The idea was doing Bayesian analysis on categorical data. And we chose a data about people from different categories going for the different type of insurance products. Our objective was to draw inference on population proportions and we have point and interval estimations of the same. We have also conducted a test of independence to test if socio-economic condition affects the choice of different insurance products. And we saw that the choice of products and the socio-economic condition is not independent. So, they do affect each other. Again, we have carried out some comparative analysis to compare different population proportions and we calculated their probabilities.

The theory we derived and hence used can be applied for any kind of categorical data. Our experiment was very simple, yet it gives strong results. We believe that the same procedure can be followed to solve many problems and answer many types of question. Because now we can go beyond the point and interval estimates and play with probability. It is quite exciting to explore the joint and conditional probabilities of population parameters. We hope to carry on our exercise further and do more analysis on homogeneity and group-wise comparison.

II. Bayesian Inference on a Binomial Proportion

1. Introduction

So far we have been using frequentist (or classical) methods. In the frequentist approach, the probability is interpreted as long run frequencies. The goal of frequentist inference is to create procedures with long run guarantees. Indeed, a better name for frequentist inference might be procedural inference. Moreover, the guarantees should be uniform over θ if possible. For example, a confidence interval traps the true value of θ with probability 1−α, no matter what the true value of θ is. In frequentist inference, procedures are random while parameters are fixed, unknown quantities.

In the Bayesian approach, the probability is regarded as a measure of the subjective degree of belief. A Bayesian thought would be regarding the experimental outcome as fixed and treating the parameter of interest as random. In this case, the parameter has a probability distribution before the data is collected which is referred as prior distribution and a probability distribution after the data is collected is posterior distribution.

In the light of many paradigms in the statistical inference, there are many debates still exist such as "Are you a Bayesian or frequentist?" However, this made us enlighten our own problem with a question "What is the probability that M.Sc. Students in the department are Bayesian after the elective course?"

Secondly, death reports are observed to be high due to selfies. Hence we are also interested in observing what is the proportion of youngsters, between the age group of 20-25, taking selfies? The question is subjective which facilitates the -introduction of the prior distribution. However, it needs to be noted the objectives are totally different from each other and independent. Both the objectives are observed separately for which we have enriched with Bayesian thoughts. The objective of this study has been in two different aspects

(1) What is the probability that students are Bayesian after the elective course?

(2) How often do you take selfies?

2. Data Description

The data observed in the study is collected directly, it is a primary data.

DATASET1: The samples are collected from batch 2016-2018, Department of Statistics, University of Madras. Out of the 40 students (Actuarial Science, Statistics) enrolled in the department, 17 students opted for Bayesian Inference Elective course. Among the 17 students, 3 of them are Ph.D. students.

The data consists of their course, gender and their interest in Bayesian (1).

Sl.No.	Course	Gender	Bayesian
1	M.Sc. Statistics	F	1
2	M.Sc. Statistics	F	1
3	M.Sc. Statistics	F	1

4	M.Sc. Statistics	F	1
5	M.Sc. Actuarial Science	M	1
6	M.Sc. Statistics	M	1
7	M.Sc. Actuarial Science	M	1
8	M.Sc. Actuarial Science	F	0
9	M.Sc. Actuarial Science	M	1
10	M.Sc. Actuarial Science	F	1
11	M.Sc. Actuarial Science	F	1
12	Ph. D	F	1
13	Ph. D	M	1
14	M.Sc. Statistics	F	1
15	M.Sc. Statistics	M	1
16	M.Sc. Statistics	M	1
17	Ph.D	F	1

<u>DATASET 2</u>:

The samples are collected from 100 people between the age group of 20-25 (Male and Female). The data consists of gender and their interest in taking selfies (1 refers Yes, 0 refers No).

Sl.No.	Gender	Taking selfies
1	F	1
2	M	0
3	F	0
4	F	1
:	:	:
100	M	1

For these two datasets, Bayesian analysis is discussed in Section 4.

3. Methodology

For both data setlikelihood is given by, $l(\theta/x) = {}^{n}C_{x}\theta^{x}(1-\theta)^{n-x}$; x=0,1,2,.....n, $0\leq\theta\leq1$. Thus both the data follow Binomial Distribution with parameter θ. The prior is taken as Beta distribution 1st kind, i.e. θ ~Beta(α,β)

$$P(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$; α,β >0 and 0<θ <1and zero otherwise; where,

B(α,β) is the beta function $=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Hence the posterior is defined as

$$\pi(\theta/x) = \frac{\frac{nCx\theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}}{\int_0^1 \frac{nCx\theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}d\theta} = \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}d\theta}$$

$=\frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{Beta(x+\alpha,n-x+\beta)}$ and hence θ/x~Beta(x+α,n - x+β)

3.1 Summary Measures

If X~Beta(a,b) then closed form for few summary measures are readily available

Mean $= \frac{a}{a+b}$

Variance$= \frac{ab}{(a+b)^2\ (a+b+1)}$

Median $= \frac{a-1/3}{a+b-2/3}$

Mode $= \frac{a-1}{a+b-2}$

Here, since θ/x~Beta(x+α, n – x+β), a = x+α and b = n - x+β

3.2 Credible Interval

A set (θ_L, θ_U), such that P(θ_L<θ<θ_U) = 1-α is called the credible interval.i.e. P(θ< θ_L) = α/2 and P(θ_U< θ) = α/2. Using the relation between F and Beta can be used to derive a (1- α) % credible interval

Here if θ~B(α,β), then $\frac{\beta\theta}{\alpha(1-\theta)}\sim$ F(m.n) where, m = 2α , n = 2β. Also following property of F-distribution is helpful; if F~F(m,n) , then 1/F ~ F(n,m)

3.3 Sensitivity of priors

Jeffrey's prior: P(θ) α $[J(\theta)]^{1/2}$where, J(θ) = $-E[\partial^2 l/\partial\theta^2] = E[\partial l/\partial\theta]^2$

l = ln f(x/θ) =ln l(θ/x)=ln$^nC_x\theta^x(1-\theta)^{n-x}$ = k + x lnθ + (n-x) ln (1-θ).

$\frac{\partial l}{\partial\theta} = \frac{x}{\theta} + \frac{(n-x)(-1)}{(1-\theta)}and\frac{\partial^2 l}{\partial\theta^2} = -\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}$ So that expectation with respect to X ~ B(n, θ) and subsequent square root implies α = ½ ; β = ½ . Also, it could be observed that other vague priors are (1,1) , (0,1) and (1,0).

3.4 Predictive Distributions

Here, since X ~B(n,θ)

$\theta\sim Beta(\alpha,\beta)and\theta/x \sim Beta(x+\alpha, n-x+\beta)$

Then $P(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\Gamma\beta}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ and

$$\pi\left(\frac{\theta}{x}\right) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

3.4.1 Prior Predictive

The prior predictive density of x or marginal likelihood of x is given as

$$M(x) = \int f(x/\theta).P(\theta)d\theta = \int_0^1 \mathrm{nCx}\,\theta^{x}(1-\theta)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\Gamma\beta}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$$

$$= {}^{n}C_{x}\frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\Gamma\beta}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

3.4.2 Posterior Predictive

Let $Y \sim B(m,\theta)$. The posterior predictive of Y=y is given by,

$$P(Y/x) = \int f(y/\theta).\pi(\theta/x)d\theta$$

$$= \int_0^1 \mathrm{mCy}\,\theta^{y}(1-\theta)^{m-y}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}d\theta$$

$$= {}^{m}C_{y}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}\frac{\Gamma(\Gamma x+y+\alpha)\Gamma(m+n-x-y+\beta)}{\Gamma(m+n+\alpha+\beta)}$$

3.5 Bayes Factor

The expressions for Bayes factor (BF_{01}) corresponding to five different testing scenario involving binomial proportion (indicated as Hypothesis 1 to 5) have been listed based on the Bayesian model, $x \sim Binomial(n,\theta)\ \theta \sim Beta(a,b)$

Hypothesis 1: $H_0 : \theta = \theta_0$ Vs $H_1 : \theta \neq \theta_0$

$$BF_{01} = \frac{\Gamma(a)\Gamma(b)\Gamma(n+a+b)}{\Gamma(a+b)\Gamma(x+a)\Gamma(n-x+b)}\theta_0^x(1-\theta_0)^{n-x}$$ where Γ is the Gamma integral in $(0,\infty)$

Hypothesis 2: $H_0 : \theta = \theta_0$ Vs $H_1 : \theta > \theta_0$

$$BF_{01} = \frac{Beta(a,b) - LIB(a,b,\theta_0)}{Beta(x+a,n-x+b) - LIB(x+a,n-x+b,\theta_0)}\theta_0^x(1-\theta_0)^{n-x}$$ where $Beta(a,b)$ is the Beta integral in $(0,1)$ and LIB is the lower incomplete Beta integral in $(0,\theta_0)$

Hypothesis 3: $H_0 : \theta = \theta_0$ Vs $H_1 : \theta < \theta_0$

$$BF_{01} = \frac{LIB(a,b,\theta_0)}{LIB(x+a,n-x+b,\theta_0)}\theta_0^x(1-\theta_0)^{n-x}$$

Hypothesis 4: $H_0 : \theta \le \theta_0$ Vs $H_1 : \theta > \theta_0$

$$BF_{01} = \frac{\left(Beta(a_1, b_1) - LIB(a_1, b_1, \theta_0)\right) LIB(x + a_0, n - x + b_0, \theta_0)}{LIB(a_0, b_0, \theta_0)\left(Beta(x + a_1, n - x + b_1) - LIB(x + a_1, n - x + b_1, \theta_0)\right)}$$

Hypothesis 5: $H_0 : \theta \ge \theta_0$ Vs $H_1 : \theta < \theta_0$

$$BF_{01} = \frac{LIB(a_1, b_1, \theta_0)\left(Beta(x + a_0, n - x + b_0) - LIB(x + a_0, n - x + b_0)\right)}{\left(Beta(a_0, b_0) - LIB(a_0, b_0, \theta_0)\right) - LIB(x + a_1, n - x + b_1, \theta_0)}$$

where a_0, b_0, a_1, b_1 are the hyperparameters related to null and alternative hypothesis respectively.

The explicit derivations of Bayes factor for testing single binomial proportion based on three different hypotheses are presented. Details of hypotheses III and V are omitted from this presentation, due to the similarity in their approaches.

Hypothesis I

$H_0 : \theta = \theta_0$ Vs $H_1 : \theta \ne \theta_0$

$g(\theta_1) = \mathrm{Beta}(a, b)$

$$BF_{01} = \frac{f(x/\theta_0)}{\int f(x/\theta_1) g(\theta_1) d\theta_1} \text{------------------------} (1)$$

Denominator of (1):

$$\int f(x/\theta_1) g(\theta_1) d\theta_1 = \int \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} d\theta$$

$$= \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int \theta^{x+a-1} (1-\theta)^{n-x+b-1} d\theta$$

$$= \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mathrm{Beta}(x+a, n-x+b) \frac{1}{\mathrm{Beta}(x+a, n-x+b)} \int_0^1 \theta^{x+a} (1-\theta)^{n-x+b} d\theta$$

$$= \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}$$

Numerator of (1): $f(x/\theta_0) = \binom{n}{x} \theta_0^x (1-\theta_0)^{n-x}$

Therefore $$BF_{01} = \frac{\binom{n}{x} \theta_0^x (1-\theta_0)^{n-x}}{\binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}}$$

$$BF_{01} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)} \theta_0^x (1-\theta_0)^{n-x}$$

Hypothesis II

$H_0:\theta=\theta_0$ Vs $H_1:\theta>\theta_0$

Numerator of (1) $\binom{n}{x}\theta_0^x(1-\theta_0)^{n-x}$

For Alternative test, $\theta>\theta_0$, the entire space is $(\theta_0,1)$ and the pdf is such that

$$K\int_{\theta_0}^{1}\theta^{a-1}(1-\theta)^{b-1}=1$$

$$K=\left[\int_{\theta_0}^{1}\theta^{a-1}(1-\theta)^{b-1}d\theta\right]^{-1}$$

$$K=[1-LIB(a,b,\theta_0)]^{-1}$$

Denominator of (1):

$$\int f(x/\theta(x/\theta)d\theta$$

$$\int f\left(x/\theta\right)=\binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad ;\ 0<\theta<1$$

$$g(\theta)=K\theta^{a-1}(1-\theta)^{b-1} \qquad ;\ a,b>0,\ \ \theta_0<\theta<1$$

Therefore $\int_{\theta_0}^{1}\binom{n}{x}\theta^x(1-\theta)^{n-x}K\theta^{a-1}(1-\theta)^{b-1}d\theta$

$$=\binom{n}{x}K\int_{\theta_0}^{1}\theta^{x+a+1}(1-\theta)^{n-x+b-1}d\theta$$

$=\binom{n}{x}KUIB(x+a,n-x+b,\theta_0)$ where UIB is the Upper Incomplete Beta Function defined as

$$UIB(a,b,\theta_0)=\int_{\theta_0}^{1}x^{a-1}(1-x)^{b-1}dx$$

Therefore $UIB(a,b,\theta_0)=Beta(a,b)-LIB(a,b,\theta_0)$

$$BF_{01}=\frac{\binom{n}{x}\theta_0{}^x(1-\theta_0)^{n-x}}{\binom{n}{x}KUIB(x+a,n-x+b,\theta_0)}$$

Substitute K in BF_{01}

$$BF_{01}=\frac{Beta(a,b)-LIB(a,b,\theta_0)}{Beta(x+a,n-x+b)-LIB(x+a,n-x+b,\theta_0)}\theta_0{}^x(1-\theta_0)^{n-x}$$

Hypothesis IV

$H_0:\theta\le\theta_0$ Vs $H_1:\theta>\theta_0$

$$BF_{01}=\frac{\int f(x/\theta_0)g_0(\theta)d\theta}{\int f(x/\theta_0)g_1(\theta)d\theta}------------------------(2)$$

$$g_0(\theta)=K_0\theta^{a_0-1}(1-\theta)^{b_0-1} \quad ;\ 0<\theta<\theta_0$$

$$K_0=\left[\int_0^{\theta_0}\theta^{a_0-1}(1-\theta)^{b_0-1}d\theta\right]^{-1}$$

$$K_0=[LIB(a_0,b_0,\theta_0)]^{-1}$$

$$g_1(\theta) = K_1\theta^{a_1-1}(1-\theta)^{b_1-1} \quad ; \qquad \theta_0 < \theta < 1$$

$$K_1 = \left[\int_{\theta_1}^{1}\theta^{a_1-1}(1-\theta)^{b_1-1}d\theta\right]^{-1}$$

$$K_1 = [UIB(a_1, b_1, \theta_0)]^{-1}$$

Numerator of (2):

$$\int f(x/\theta)g_0(\theta)d\theta = \int_0^{\theta_0}\binom{n}{x}\theta^x(1-\theta)^{n-x}K_0\theta^{a_0-1}(1-\theta)^{b_0-1}d\theta$$

$$= \binom{n}{x}K_0\int_0^{\theta_0}\theta^{x+a_0-1}(1-\theta)^{n-x+b_0-1}d\theta$$

$$= \binom{n}{x}K_0 LIB(x+a_0, n-x+b_0, \theta_0)$$

Denominator of (2):

$$\int f(x/\theta)g_1(\theta)d\theta = \int_{\theta_0}^{1}\binom{n}{x}\theta^x(1-\theta)^{n-x}K_1\theta^{a_1-1}(1-\theta)^{b_1-1}d\theta$$

$$= \binom{n}{x}K_1\int_{\theta_0}^{1}\theta^{x+a_1-1}(1-\theta)^{n-x+b_1-1}d\theta$$

$$= \binom{n}{x}K_1 UIB(x+a_1, n-x+b_1, \theta_0)$$

$$BF_{01} = \frac{K_0 LIB(x+a_0, n-x+b_0, \theta_0)}{K_1(Beta(x+a_1, n-x+b_1) - LIB(x+a_1, n-x+b_1, \theta_0))}$$

Substitute K_0 and K_1 in BF_{01}

$$BF_{01} = \frac{(Beta(a_1, b_1) - LIB(a_1, b_1, \theta_0))LIB(x+a_0, n-x+b_0, \theta_0)}{LIB(a_0, b_0, \theta_0)(Beta(x+a_1, n-x+b_1) - LIB(x+a_1, n-x+b_1, \theta_0))}$$

4. Analysis

4.1. Data set 1:

n = 17, x = 16 and n-x = 17-16 = 1 and X $\sim B(17, \theta)$

The likelihood is given by,

$l(\theta/x) = {}^{17}C_{16}\theta^{16}(1-\theta)^1$; x= 0,1,2,.....17

The prior is $\theta \sim Beta(\alpha, \beta)$

Posterior is $\theta/x \sim Beta(x+\alpha, n-x+\beta)$ and hence $\theta/x \sim Beta(16+\alpha, 1+\beta)$. For example Jeffrey's prior (1/2,1/2) yields a prior $\theta \sim Beta(1/2,1/2)$ so that the posterior is $\theta/x \sim Beta(\frac{33}{2}, \frac{3}{2})$

4.1.1 Summary measures

For different values of α and β, the mean, variance, mode and median and credible intervals are given in the table.

(α,β)	Mean	Variance	Median	Mode	95%credible interval
(1/2,1/2)	0.9167	0.0041	0.9327	0.9688	(0.7918,0.9936)
(1,1)	0.8947	0.0047	0.9091	0.9412	(0.7624,0.9862)
(0,1)	0.8889	0.0052	0.9038	0.9375	(0.7497,0.9854)
(1,0)	0.9445	0.0027	0.9615	1	(0.8383,0.9985)

4.1.2 Predictive Densities

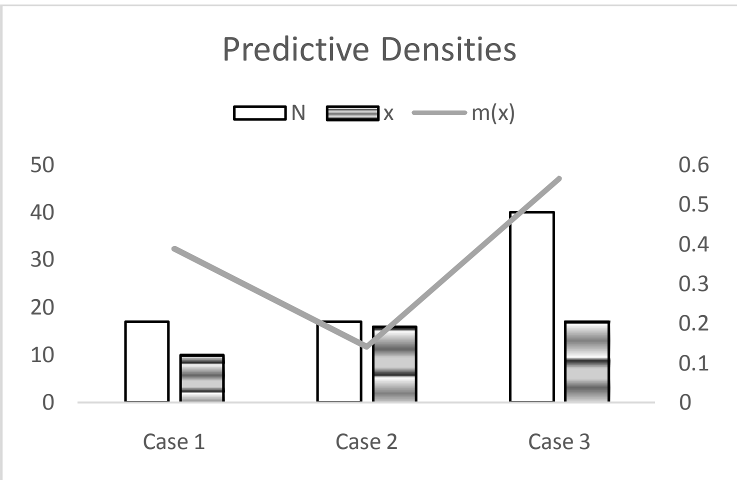
The marginal likelihood of X is given by,

$$M(x) = {}^{n}C_{x}\frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\Gamma\beta}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$M(16)= {}^{17}C_{16}\frac{\Gamma(\frac{1}{2}+\frac{1}{2})}{\frac{\Gamma 1\Gamma 1}{2\ \ 2}}\frac{\Gamma\left(16+\frac{1}{2}\right)\Gamma(1+\frac{1}{2})}{\Gamma(17+\frac{1}{2}+\frac{1}{2})}= \frac{17!}{16!1!}\frac{\Gamma 1}{\frac{\Gamma 1\ \Gamma 1}{2\ \ 2}}\frac{\frac{\Gamma 33\Gamma 3}{2\ \ 2}}{\Gamma 18}= 0.1399$$

Few more illustration

Case	N	x	m(x)
1	17	10	0.3875
2	17	16	0.1399
3	40	17	0.5654



As an example for **posterior predictive** we find the probability that 10 students out of 17 might be Bayesian after the present data,

$$P(Y=10/x=16) = {}^{17}C_{10}\frac{\Gamma(17+\frac{1}{2}+\frac{1}{2})}{\Gamma\left(16+\frac{1}{2}\right)\Gamma(1+\frac{1}{2})}\frac{\Gamma\left(26+\frac{1}{2}\right)\Gamma(17+17-16-10+\frac{1}{2})}{\Gamma(17+17+\frac{1}{2}+\frac{1}{2})}$$

$$= \frac{17!}{10!7!}\frac{\Gamma 18}{\Gamma\left(\frac{33}{2}\right)\Gamma(\frac{3}{2})}\frac{\Gamma\left(\frac{53}{2}\right)\Gamma(\frac{17}{2})}{\Gamma(35)} = 0.14913$$

To test the percentage of students who are Bayesians after the elective course let us take,

$H_0: \theta = 0.75\ vs.\ \theta \neq 0.75$

For convenience, we take the values of α,β as (1,1)

4.1.3 Bayes Factor

$$B_{01} = \frac{B(\alpha,\beta)}{B(x+\alpha,n-x+\beta)}\theta_0^{x}(1-\theta_0)^{n-x}$$

$$= \frac{B(1,1)}{B(17,2)}(0.75)^{16}(0.25)^{17-16}$$

$$= 306\times 0.25\times 0.75^{16}$$

$$= 0.7768$$

$B_{10} = (B_{01})^{-1} = 1.304$

H_1 is not worth more than a bare mention. H_0 is favoured. Also, we have tested H_0: $\theta = 0.75\ vs.\ \theta > 0.75$ which yields $B_{01} = 6.5 \times 10^{-4}$; that is strong evidence against H_0.

4.2 Dataset 2:

Here n = 100 and X = 1, if taking selfies; 0, otherwise. $\overline{x = 61, \text{ then } n\text{-}x}$ = 39X∼B(100,θ). Prior is $\theta \sim Beta(\alpha,\beta)$ and Posterior is $\theta/x \sim Beta(61+\alpha, 39+\beta)$

4.2.1 Summary Measures

For different values of α and β, the mean, variance, mode, median and credible intervals are given in the table.

(α, β)	**Mean**	**Variance**	**Median**	**Mode**	**95%credible interval**
(1/2,1/2)	0.6089	0.00233	0.6096	0.6111	(0.5281, 0.6872)
(1,1)	0.6078	0.00231	0.6086	0.6100	(0.5274, 0.6858)
(1,0)	0.6139	0.00232	0.6146	0.6162	(0.5332, 0.6919)
(0,1)	0.6039	0.00230	0.6046	0.6061	(0.5231, 0.6825)

To test the percentage of youngsters taking selfies, we take

$$H_0: \theta_0 = 0.65 \text{ vs. } \theta_0 \neq 0.65$$

For convenience, let us take α,β as (1,1)

4.2.2 Bayes Factor

$B_{01} = \frac{B(1,1)}{B(62,40)} (0.65)^{61} (0.35)^{39} = 5.8292$ and a positive evidence favouring H_0

5. Summary

In this work, Bayesian analysis has been studied for two different aspects which are independent of each other.

(a) The probability of students to prefer Bayesian after the elective course and the other dataset how often selfies are taken are observed with Binomial distribution as likelihood.

(b) For both illustrations, the prior and posterior distributions are Beta distributions.

(c) The summary measures such as mean, variance, median, mode are studied for both illustrations by using Jeffrey's prior and vague priors. The sensitivity of prior reveals that Jeffrey's prior (1/2, 1/2) and vague prior (1,0) provide a larger mean, median and mode with a lesser variance for dataset 1. However, for dataset 2, the vague priors and Jeffrey's priors are closer to each other.

(d) The 95% confidence interval for the dataset 1 which gives the probability of students turns out to be Bayesian is (0.8383, 0.9985) for the vague pair. We have considered this is most suitable because Jeffrey's prior reveals 95% credible interval with wide interval (0.7918, 0.9936).The 90% credible interval for dataset 2 which gives the probability of how often selfies are taken by the age group 20-25 is (0.5332,0.6919), more or less all others are going to be same.

(e) In the case of Bayes factor for the calculative purpose, the vague prior (1, 1) is considered for both data sets.

It is revealed that the hypothesis of 75% of students turns out to be Bayesian after the elective course. Similarly, the hypothesis of 65% between the age group 20-25, irrespective of gender is prone to selfies has been favoured.

III. Bayesian Inference on Most Worrying Issue in India

Abstract

This project emphasizes on the Bayesian analysis of a multinomial dataset. The prior distribution of a multinomial data will follow a Dirichlet distribution and so as the posterior. The marginals of the posterior variables follow the beta distribution of 1^{st} kind. F distribution can be obtained from Beta distribution with suitable transformations. This will be helpful to calculate the credibility intervals of the parameters. The concept of Gauss hypergeometric function has been discussed on the notion of evaluating the ratio of two independent beta distributions. Apart from numerical analysis, theoretical support to the calculations has been provided as per necessity.

1. Introduction

Analyzing multinomial data is very challenging in statistical practice. Our Bayesian project deals with the worrying issues of our country. We select few issues like Terrorism, Unemployment, Corruption, Poverty/Inequality, Environmental Threats, Education, Moral decline, and Rise of Extremism. Moreover, we select a particular age (20-25) group as our population to work with our project. Individuals can select exactly one issue which they consider to be the most worrying issue and we collect the data via social media. Our data model is multinomial dealing with eight variables.

We are interested in the comparative study of the parameters to examine which issue is more sensible to be considered as a most worrying issue. In order to fulfill the interest, we go for ratio test. There are little hardships in the numerical evaluation of the integrations. In that case, we took help of R-language to get the proper result. The necessary R codes have been provided in the appendix portion.

2. Data

Survey on the most worrying issue in India (among age group of 20-25)										
Serial No			Issues under consideration							
	Occupation	Gender	Terrorism	Unemployment	Corruption	Poverty/ Inequality	Environmental Threats	Education	Moral Decline	Rise of Extremism
1	Student	M								•
2	Working professional	F		•						
3	Student	F		•						
4	Political activist	F		•						
5	Student	F		•						
6	Working professional	F					•			
7	Working professional	M					•			
8	Student/Political activist	F		•						
9	Housewife	M		•						
10	Student	M								•
11	Student	F						•		
12	Student	F							•	
13	Student	F				•				
14	Student	F			•					
15	Student	F					•			
16	Working professional	M							•	
17	Student	F			•					
18	Working professional	F			•					
19	Student	F						•		
20	Working professional	M							•	
21	Student	M						•		
22	Student	F	•							
23	Student	F							•	
24	Student	M						•		
25	Student	F	•							
26	Student	F			•					
27	Student	M		•						
28	Student	M			•					
29	Student	M			•					
30	Student	F								•
31	Student	F		•						
32	Student	M		•						
33	Student	M						•		
34	Student	F						•		
35	Student	F						•		
36	Working professional	F		•						
37	Student	F							•	
38	Student	F							•	
39	Student	M								•
40	Unemployed	F							•	
Total			**2**	**10**	**6**	**1**	**3**	**7**	**7**	**4**

3. Methodology

The data is considered to be a multinomial model. To construct a prior distribution we consider the parameter(s) of the prior distribution as the random variable(s) of the prior distribution. Likewise, in the present scenario, θ_i's will be treated as random variables in the prior distribution with range obviously 0 to 1.

The prior distribution will be,

$\left(\theta_{2,}\theta_{3,}\theta_{4,}\theta_5,\theta_{6,}\theta_7,\theta_8\right) \sim dirichilet(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8)$

The probability density function will be,

$$p(\theta) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_2)\Gamma(\alpha_2)\Gamma(\alpha_3)\Gamma(\alpha_4)\Gamma(\alpha_5)\Gamma(\alpha_6)\Gamma(\alpha_7)\Gamma(\alpha_8)} \times \quad \theta_1^{\alpha_1 - 1} \ldots\ldots \theta_8^{\alpha_8 - 1},\ \alpha_i \geq 0,\ 0 \leq \theta_i \leq 1$$

The posterior distribution will be $\Pi(\theta|x)\ \alpha\ p(\theta) \times f(x|\theta)$. In general, If $(X_1, X_2, X_3, \ldots\ldots, X_k) \sim$ Multinomial $(N, \theta_1, \theta_2, \theta_3, \ldots\ldots, \theta_k)$ Then, $\theta \sim$ diri $(\alpha_1, \alpha_2, \alpha_3, \ldots\ldots, \alpha_k)$ And, $\theta|X \sim$ diri$(\alpha_1+x_1, \alpha_2+x_2, \alpha_3+x_3, \ldots\ldots\ldots\ldots, \alpha_k+x_k)$ Jeffreys prior for θ is, $\alpha_i = 1/2$, $i=1,2,3,\ldots,k$

We would like to derive the marginal distribution ,Let us consider the pdf of Dirichlet distribution for 8 cases

$$(\theta_1, \theta_2, .., \theta_8 \mid X) \sim Diri(x_1+\alpha_1, x_2+\alpha_2, x_3+\alpha_3, x_4+\alpha_4, x_5+\alpha_5, x_6+\alpha_6, x_7+\alpha_7, x_8+\alpha_8)$$

$$\Pi(\theta|X) = \frac{\Gamma(N+\alpha)}{\prod_{i=1}^{8}\Gamma(x_i+\alpha_i)} \prod_{i=1}^{8} \theta_i^{x_i+\alpha_i-1}, \quad \sum_{i=1}^{8}(x_i+\alpha_i) = N+\alpha$$

Now, the marginal distribution of $(\theta_1|X)$ will be,

$$\iiint \underset{\sum\theta_i \leq 1}{\iiiint} \frac{\Gamma(N+\alpha)}{\prod_{i=1}^{8}\Gamma(x_i+\alpha_i)} \prod_{i=1}^{8} \theta_i^{x_i+\alpha_i-1}\ \partial\theta_1 \ldots\ldots \partial\theta_8$$

Consider, $\vartheta_7 = (1-\theta_1-\theta_2-\theta_3-\theta_4-\theta_5-\theta_6)\, u_1$

$$\partial\theta_7 = (1-\theta_1-\theta_2-\theta_3-\theta_4-\theta_5-\theta_6)\ \partial u_1$$

Then the integral will be,

$$\int \underset{\sum\theta_i < 1}{\iint} \iint \frac{\Gamma(N+\alpha)}{\prod_{i=1}^{8}\Gamma(x_i+\alpha_i)} \theta_1^{x_1+\alpha_1-1}\theta_2^{x_2+\alpha_2-1}\theta_3^{x_3+\alpha_3-1}\theta_4^{x_4+\alpha_4-1}\theta_5^{x_5+\alpha_5-1}\theta_6^{x_{61}+\alpha_{61}-1} (1-\theta_1-\theta_2-\theta_3-\theta_4-\theta_5-\theta_6)^{x_7+\alpha_7+x_8+\alpha_8-1} \times B(x_7+\alpha_7, x_8+\alpha_8)\ \partial\theta_2\partial\theta_3\partial\theta_4\partial\theta_5\partial\theta_6$$

Proceeding in this way, we get,

$$\prod_1(\theta_1 \mid X) = \frac{\Gamma(N+\alpha)}{\Gamma(x_1+\alpha_1)\Gamma(x_2+\alpha_2+x_3+\alpha_3+x_4+\alpha_4+x_5+\alpha_5+x_6+\alpha_6+x_7+\alpha_7+x_8+\alpha_8)}$$

$$\times \quad \theta_1^{x_1+\alpha_1-1}(1-\theta_1)^{\sum_{j=2}^{8} x_j+\alpha_j-1} \qquad , \text{j=2,…,8}$$

Therefore,

$$\prod_1(\theta_1 \mid X) \sim B(x_1+\alpha_1, N+\alpha-x_1-\alpha_1)$$

In general,

$$\prod_k(\theta_k \mid X) \sim B(x_k+\alpha_k, N+\alpha-x_k-\alpha_k)$$

Then under H_0, the prior distribution $\pi_0(\theta)$ for the parameter $\theta = (\theta_{ij})$ isbased on the law of independence $\theta_{ij} = \Pi_i \Psi_j$ where $\Pi_i = \text{Dirichlet}(\gamma_i)$ and $\Psi_j = \text{Dirichlet}(\delta_j)$

Also for the prior $\pi_1(\theta)$ for model H_1 is $\theta = (\pi_{ij}) \sim \text{Dirichlet}(\alpha_{ij})$. 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,

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

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

where $\gamma = \Sigma\gamma_i; \delta = \Sigma\delta_j$

3.1 Bayes Factor

Bayes factor for comparing two models is

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

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

3.2 Comparative Study

Comparative study ϑ_i*'s. i.e.* in that case, we will try to find out $P(\vartheta_i > \vartheta_j)$ where $i,j=1,2,3…k$ *and* $i \neq j$ has been attempted. It has been already proved that each ϑ_i follows the beta distribution of 1st kind. The probability $P(\vartheta_i > \vartheta_j)$ Can be redefined as, $p\left(\frac{\theta_i}{\theta_j} > 1\right)$, then to find this type of probabilities, we need to check the probability distribution of the ratio of two beta distribution. Consider the standard beta distribution with density,

X~Beta(a, b)

Y~Beta(α, β)

X and Y are independent.

$$f_1(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad , 0 < x < 1$$

= 0 , otherwise

$$f_2(y) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} , 0 < y < 1$$

= 0, otherwise

Since X and Y are independent, the joint pdf of X and Y will be,

$$f(x,y) = \frac{1}{A} x^{a-1}(1-x)^{b-1} y^{\alpha-1}(1-y)^{\beta-1} \quad , 0 < x < 1 ; 0 < y < 1$$

Where, A=B(a, b).B(α, β)

Jacobian of the transformation is:

$$|J| = \frac{\partial(x,y)}{\partial(w,v)} = \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} v & w \\ 0 & 1 \end{vmatrix}$$

= v

The joint PDF of w and v is

$$\Phi(w,v) = \frac{1}{A}(wv)^{a-1} v^{\alpha-1}(1-wv)^{b-1}(1-v)^{\beta-1} v$$

$$= \frac{1}{A} w^{a-1} v^{a+\alpha-1}(1-wv)^{b-1}(1-v)^{\beta-1}$$

$w > 0$, $0 < v < 1$ and $wv \le 1$

Or equivalently the joint PDF of w and y is:

$$\Phi(w, y) = \frac{1}{A} w^{a-1} y^{a+\alpha-1} (1-wy)^{b-1} (1-y)^{\beta-1}$$

If w<1,

$$f(w) = \int_0^1 \phi(w, y) dy$$

$$= \frac{w^{a-1}}{A} \int_0^1 y^{a+y-1} (1-wy)^{b-1} (1-y)^{\beta-1} dy \qquad \text{(i)}$$

Consider the Gauss hypergeometric function in three parameters.

$${}_2F_1(m;n;k;x) = \frac{1}{B(m, k-m)} \int_0^1 u^{m-1} (1-u)^{k-m-1} (1-xu)^{-n} du \qquad \text{(ii)}$$

Comparing (i) and (ii), equation (i) becomes,

$$f(w) = \frac{w^{a-1}.B(a+\alpha, \beta)}{A} {}_2F_1(a+\alpha; 1-b; \beta+\alpha+a; w)$$

If $w \geq 1$, $let\ wy = t$

$$y = \frac{t}{w}$$

$$dy = \frac{dt}{w}$$

$$f(w) = \frac{w^{a-1}}{A} \int_0^1 \left(\frac{t}{w}\right)^{a+\alpha-1} (1-t)^{b-1} \left(1-\frac{t}{w}\right)^{\beta-1} \frac{dt}{w}$$

$$= \frac{w^{-(\alpha+1)}.B(a+\alpha, \beta)}{A} {}_2F_1(a+\alpha; 1-\beta; a+b+\alpha; \frac{1}{w})$$

Hence the pdf of $w = \frac{x}{y}$ is,

$$f(w) = \begin{cases} \frac{w^{a-1}.B(a+\alpha, \beta)}{A} {}_2F_1(a+\alpha; 1-b; a+\alpha+\beta; w), 0 < w < 1 \\ \frac{w^{-(1+\alpha)}.B(a+\alpha, b)}{A} {}_2F_1(a+\alpha; 1-\beta; a+\alpha+b; w), w \geq 1 \end{cases}$$

4. Analysis

$$\prod\nolimits_1 (\theta_1 \mid X) \sim B(x_1 + \alpha_1, N + \alpha - x_1 - \alpha_1)$$

i.e. $\prod_1(\theta_1 \mid X) \sim B(2.5,41.5)$ (putting the values of α_i's and x_i's)

4.1 Posterior Mean and Variance of the Parameters

Parameter	Mean	Variance
$\boldsymbol{\theta_1}$	0.057	0.0012
$\boldsymbol{\theta_2}$	0.237	0.0044
$\boldsymbol{\theta_3}$	0.148	0.0028
$\boldsymbol{\theta_4}$	0.034	0.0007
$\boldsymbol{\theta_5}$	0.080	0.0016
$\boldsymbol{\theta_6}$	0.170	0.0031
$\boldsymbol{\theta_7}$	0.170	0.0031
$\boldsymbol{\theta_8}$	0.102	0.0020

4.2 Credible Interval

Parameters	**Credible Intervals**	
	Lower Limit	Upper Limit
$\boldsymbol{\theta_1}$	0.065	0.140
$\boldsymbol{\theta_2}$	0.120	0.370
$\boldsymbol{\theta_3}$	0.006	0.265
$\boldsymbol{\theta_4}$	0.001	0.103
$\boldsymbol{\theta_5}$	0.019	0.174
$\boldsymbol{\theta_6}$	0.074	0.292
$\boldsymbol{\theta_7}$	0.074	0.292
$\boldsymbol{\theta_8}$	0.032	0.206

Comparative analysis has been carried for $P(\theta_2 \geq \theta_6)$.

$P(\theta_2 \geq \theta_6)$= 0.791 and $P(\theta_2 < \theta_6) = 0.201$ i.e., unemployment is being more emphasised than education as the most worrying issue with probability 0.791 and the probability of "unemployment" being less emphasised than education is 0.201. Next, we consider corruption and moral decline. The required estimates will be $P(\theta_3 \geq \theta_7)$= 0.3822 equivalently $P(\theta_3 < \theta_7) = 0.6177$

5. Summary

In this project, we calculated the expectations, variances and the credibility intervals of the posterior variables. The credibility intervals have been calculated along with the pictorial representation. We have conducted the comparative analysis for two pairs of variables, viz.,

θ_2, θ_6 , and θ_3, θ_7. Though further analysis can be conducted within the other variables, we confine our analysis within these two pairs.

Form the case 1, (i.e., comparison between θ_2 and θ_6) we observe that the probability of θ_2 being greater than θ_6 is sufficiently large. We can conclude that, with respect to education as a worrying issue, people think unemployment is more burning one. On the other hand, if we investigate the case 2, we see, the probability of θ_7 being greater than θ_3 is greater than the reverse scenario.

This implies according to the respondents' mindset, moral decline is more serious than the political/social corruption. Statistical operations should be done ultimately for the betterment of our society. These inferences significantly focus on those notions. If the most worrying issues, according to the people's mindset, become the prior targets to be eradicated, that would lead the nation in a better way.

Apart from these three examples, following list indicates other problems and the respective student names

S.No	Name	Topic
1	Dayasri Ravi, Vaishnavi&Revathi	Bayesian Inference for Studying Fluctuation of Gold Price in India
2	Dhruv&Dhanavenden	Bayesian Binomial Model for finding preference for Life Insurance policy
3	Himani Mehta &SandhyaRenganathan	Bayesian Inference on Danish Claim Dataset
4	Syamily M & Rahul Nair K	Perception of online and in-line shopping

5. Conclusion

Something in your prior
was so exciting
Something in your data
was so inviting
Something in my model
told me I must have you.

B. Natvig and M. DeGroot

With a motivation to present an introduction of Bayesian inference from a learner's perspective, especially with less use of programming, this material provides some nuances of learning Bayesian inferential procedures. Nevertheless, an additional course that deals with CIST (Simulation, Markov chain Monte Carlo methods) will definitely enhance the use of Bayesian methods for more complex situations. However, the starting point is the only intention of this presentation. A list of references (surely, not exhaustive) is provided to know the other studies from literature that will be more helpful for prospective learners.

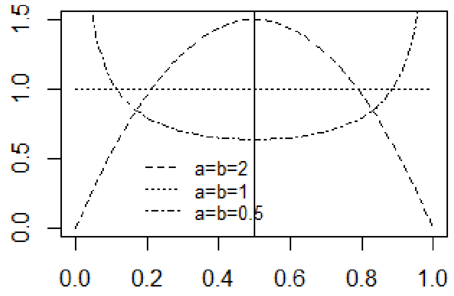
Also, a set of questions that helps an active discussion in the classroom

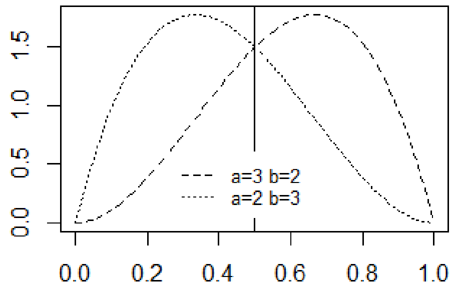
1. Example: any two situations in your daily life using the word "probably".
2. What is the probability that Chennai will receive rain today evening?
3. Extend Question number 5, for the amount of rain
4. What is the probability that Tsunami will hit Chennai coast tomorrow evening?
5. What is the probability that Tsunami will hit Chennai coast on Dec 24, 2017?
6. What is the probability that CSK will play next edition of IPL?
7. What is the probability that epidemic of smallpox will break next month?
8. Based on Question 9, find the probability that no case will be registered till Dec 1, 2017?
9. What is the probability that a movie will be a superhit?
10. What is the probability that Jio of Reliance will be successful in the Indian market?
11. What is the probability that you will have a meal at 11 AM, today?
12. Answer Question number 11, assuming that you missed your breakfast today.
13. What is the probability that metro will be extended to NIE from a place of your choice?
14. Based on an answer to question 13, what is the probability that you will commute in the metro?
15. For understanding discrete probabilities and associated Bayesian inference
    a. What is the probability of getting a head from a coin?
    b. <u>What is the probability of getting a head from an ***unbiased/fair*** coin?</u>
    c. This summer is too hot. This statement has a probability {0.1, 0.25, 0.5, 0.75, and 0.9}. There is a state of knowledge for these values respectively {0.1, 0.2, 0.2, 0.3, and 0.2}. Find the posterior distributions if the maximum temperature (in Celsius) in Chennai is 37, 35, 35, 33, 36, 38, and 39 during the second week of April

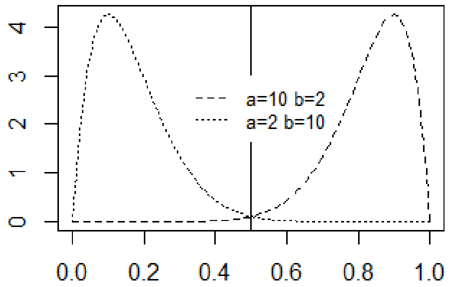
More examples of learning activities

Activity 1: Thinking or Formula - to understand Bayesian Inference -

An illustration at Subbiah et al (2013) *http://www.jenvstat.org/v04/i07*







Activity 2: Application of Bayes Formula

"extraordinary claims require extraordinary evidence**"**

Suppose a patient has one of three diseases (X, Y, Z) whose prevalence is 0.6, 0.3 or 0.1, respectively—X is relatively common, whereas Z is rare. We have access to a diagnostic test that measures the presence of protein markers (A, B). Both markers can be present, and the probabilities of observing a given marker for each disease are known (refer the display) and independent of each other in each disease state.

	X	Y	Z
A	0.2	0.9	0.2
B	0.2	0.2	0.9

If we see marker A, can we predict the state of the patient? Also, how do our predictions change if we subsequently assay for B?*http://www.nature.com/nmeth/journal/v12/n4/pdf/nmeth.3335.pdf*

Activity 3: Reading from Ellison, A. M. (2004)

http://www.uvm.edu/~bbeckage/Teaching/DataAnalysis/AssignedPapers/Ellison_2004.pdf

"Bayes Theorem is also iterative. An investigator may start with little or no information with which to construct the prior, but the posterior derived from the first experiment can then be used as a prior for the next experiment."

Try to **update your belief** in any case of proportion of success from a sequence of Bernoulli trials

Activity 4:

Attempt a Bayesian approach for inference about Poisson parameter arising from any situation of your interest. An illustrative data set (Horse kick data) may be accessed from
http://www.math.uah.edu/stat/data/HorseKicks.html

Activity 5:

Attempt a Bayesian approach for inference about Normal mean parameter arising from any situation of your interest.

References

1. Rubin DB (1984).Bayesianly justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics: 12; 1151 – 1172.
2. Dunson DB (2001). Practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology: 153; 1222 – 1226.
3. Berger J (2006). The case for objective Bayesian analysis. Bayesian Analysis: 3; 385 – 402.
4. Goldstein M (2006). Subjective Bayesian analysis: principles and practice. Bayesian Analysis: 3; 403 – 420.
5. Wolport RL (2004). A conversation with James O. Berger: Statistical Science: 19; 205 – 218.
6. Lee PM (1989). Bayesian Statistics: An introduction. Oxford university press, Oxford.
7. Robert CP (1994). The Bayesian choice. A decision-theoretic motivation. Springer-Verlag, New York.
8. Gelman A (1996). Inference and monitoring convergence. In: Gilks WR, Richardson S and Spiegelhalter DJ, ed. Markov Chain Monte Carlo in practice. Boca Raton, Chapman & Hall/CRC.
9. Johnson (2005) Bayes factors based on test statistics. J.R.Statist.Soc. B 67: 689-701.
10. Vanpaemel (2010) Prior sensitivity in theory testing: An apologia for the Bayes factor. Journal of Mathematical Psychology; 54: 491-498
11. Hashemi L, Nandram B and Goldberg R (1997). Bayesian analysis for a single 2 × 2 table. Statistics in Medicine: 16; 1311 – 1328.
12. Nandram and Choi (2007) Bayesian analysis of a two-way categorical table incorporating intraclass correlation. Journal of Statistical Computation and Simulation 76: 233 – 249.
13. Berger and Delampady (1987) Bayesian testing of precise hypotheses (with discussion). *Statistical science*, 2, 317-348.
14. Kass (1993) Bayes Factors in Practice. *The Statistician*, 42, 551-560.
15. Bernardo and Smith (1994) *Bayesian Theory.* New York: John Wiley & Sons.
16. Campbell, I. (2007). Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26, 3661-3675.
17. Kass and Raftery (1995) Bayes Factors. *Journal of the American Statistical Association*, 90, 773-795.
18. Goodman (1999) Toward Evidence – Based Medical Statistics. 2: The Bayes Factor. *Annals of Internal Medicine*, 130, 1005-1013.
19. Delampady and Berger (1990) Lower bounds on Bayes factors for multinomial distributions with application to chi-squared tests of fit. *The Annals of Statistics*, 18, 1295-1316.
20. Lavine and Schervish (1999) Bayes factor: What they are and what they are not. *The American Statistician*, 53, 119-122.
21. Ghosh, J.K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis,* New York: Springer-Verlag.
22. Robert, C., & Casella, G. (2009). Introducing Monte Carlo Methods with R. Springer Science & Business Media.
23. Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2002). Bayesian Data Analysis. London, UK: Chapman & Hall.
24. Lee, P. M. (2012). Bayesian Statistics: an introduction. John Wiley & Sons.

Quotes Courtesy

Carlin, B. P. (2006). The Bayesian Songbook:
http://stat-athens.aueb.gr/~grstats/grbayes/val8/songbook.pdf