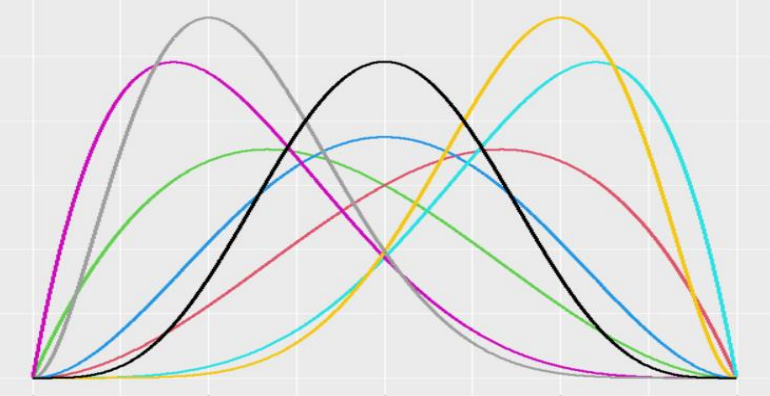
ST. XAVIER'S COLLEGE(AUTONOMOUS), KOLKATA

DEPARTMENT OF STATISTICS

COMPARISON BETWEEN CLASSICAL AND BAYESIAN APPROACH TO ESTIMATE UNKNOWN POPULATION PARAMETER



NAME: ROHIT DUTTA

ROLL: 19-300-4-07-0464

REGISTRATION NUMBER: A01-1112-0855-19

SEMESTER: 6

SESSION: 2019-2022

SUPERVISOR: PROF. PALLABI GHOSH

DECLARATION

I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.

Rohit Dutta

ROHIT DUTTA

Acknowledgement

I would like to express my thanks and acknowledgement to **Rev. Dr. Dominic Savio. S.J.**, the Principal of my college for giving me the opportunity to work on this project. I would like to thank the Vice-Principal of Arts and Science, **Prof. Betram Da'Silva** and the Dean of Science **Dr. Tapati Dutta**. I would also like to thank the Head of the Department of Statistics, **Prof. Dr. Durba Bhattacharya** for guiding me for this project.

Specially, I would like to express my thanks and gratitude to my project supervisor **Prof. Pallabi Ghosh** for her immense support, guidance and valuable advice to complete this dissertation paper, titled as – "Comparison Between Classical and Bayesian Approach to Estimate Unknown Population Parameter". This project has helped me to explore so many things related to the topic. It has been a great experience to work under the supervision of **Prof. Pallabi Ghosh,** whose valuable lessons and suggestions have really enriched the content of my dissertation work.

I would also like to take a moment to acknowledge my other professors in the Department of Statistics at St. Xavier's College (Autonomous), Kolkata, **Prof. Dr. Surabhi Dasgupta**, **Prof. Dr. Ayan Chandra**, **Prof. Dr. Surupa Chakraborty**, **Prof. Debjit Sengupta** and **Prof. Madhura Dasgupta**, who all have helped me to develop the mindset prone to research, which has made it possible for me to complete this project.

Lastly, I would like to thank my parents and friends for their supporting nature and constant guidance throughout my dissertation work.

Rohit Dutta

St. Xavier's College (Autonomous), Kolkata

Date: April, 2022

Abstract:

The parameter of interest, say θ, in statistical inferential problem is often regarded as fixed a constant. In these cases, we try to find a point estimator of θ by classical approach. But, if in some practical situation, when some additional information is available for the parameter θ, it is sometimes reasonable to consider the parameter of interest to be a random variable having some known distribution. In this project we study how we can incorporate this additional information about θ to propose a point estimator by Bayesian approach and we compare the estimators obtained with respect to the two approaches mentioned above.

Introduction:

One of the main objectives of Statistics is to draw inferences about a population from the analysis of a sample drawn from that population. Two important problems of statistical inference are 'Theory of Estimation' and 'Testing of Hypothesis'. One type of estimation, namely ***point estimation*** will be considered throughout this project. Assume that some characteristic of the elements in a population can be represented by a random variable X whose probability mass function or the probability density function is $f_X(.;\theta)$, where the form of $f_X(.;\theta)$ is assumed to be known except that it contains an unknown parameter θ. Further assume that the values $(x_1, x_2, \ldots, x_n)$ of a random sample $(X_1, X_2, \ldots, X_n)$ from $f_X(.;\theta)$ can be observed. On, the basis of the observed sample values $x_1, x_2, \ldots, x_n$ it is desired to estimate the value of the unknown parameter θ. One of the ways of this estimation is ***point estimation***, which takes into account to pick a suitable statistic, a function of sample observations, that best estimates the unknown population parameter θ. Among the several approaches of point estimation, here we are considering only two approaches, namely, i) **Classical approach** and ii) **Bayesian approach**, to estimate the unknown population parameter. In support of classical approach, where we consider the unknown parameter to be a fixed quantity, we can consider **Maximum Likelihood Estimator** of the parameter of interest. But if it is known that the considered population parameter is not a fixed quantity rather a random variable with some known distribution, then Bayesian approach can provide a way to utilize this additional information about the parameter to incorporate in our inference to propose an estimator namely, **Bayes Estimator** (here we will consider only Posterior Mean).

Some useful definitions and results:

1. ***Likelihood Function:*** The likelihood function of n random variables $X_1, X_2, \ldots, X_n$ is defined to be the joint density of n random variables, say $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n; \theta)$, which is considered as a function of θ. In particular, if $X_1, X_2, \ldots, X_n$ is a random sample from the density $f(x;\theta)$, then the likelihood function is $f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta)$.

   Notation: To think of the likelihood function as a function of θ, we shall use the notation $L(\theta; x_1, x_2, \ldots, x_n)$ for the likelihood function.

2. ***Maximum Likelihood Estimator:*** Let $L(\theta) = L(\theta; x_1, x_2, \ldots, x_n)$ be the likelihood function for the random variables $X_1, X_2, \ldots, X_n$. If $\hat{\theta}$, a function of the sample observations $x_1, x_2, \ldots, x_n$, is the value of θ in $\Theta(parameter\ space)$ which maximizes $L(\theta)$, then $\hat{\theta}$ is called the maximum likelihood estimator of θ.

3. ***Large Sample Property of Maximum Likelihood Estimator***: If $\hat{\theta}$ be the maximum likelihood estimator of θ corresponding to a random sample $X_1, X_2, \ldots, X_n$ of size n from the density $f(x; \theta)$, under certain regularity condition for large n we have,

$$\hat{\theta} \quad \widetilde{a} \quad N(\theta, \frac{1}{I(\theta)})$$

where, $I(\theta) = E\left(\frac{\partial}{\partial\theta}\ln\left(f(x_1, x_2, \ldots, x_n; \theta)\right)\right)^2$, which is known as Fisher information about θ in the random sample $X_1, X_2, \ldots, X_n$.

Suppose, we have an additional information or evidence that Θ acts as a random variable. Let, θ is the value of the random variable Θ. If Θ is a random variable, it has a distribution. Let, $g_\Theta(.)$ denote the density function of Θ. Here we will assume that $g(.)$ is completely known, that is, it contains no unknown parameters. Since θ is a value of Θ, we denote the density of X as $f_{(X|\Theta=\theta)}(x|\theta)$ instead of $f(x; \theta)$.

4. ***Prior Distribution:*** The distribution $g_\Theta(.)$ is called the prior distribution of Θ.

5. ***Posterior Distribution:*** The conditional density of Θ given $X_1 = x_1, X_2 = x_2, .. X_n = x_n$
denoted by $f_{\Theta|X_1=x_1, X_2=x_2, .., X_n=x_n}(\theta \mid x_1, x_2, \ldots, x_n)$ is called the posterior distribution of Θ.

By Bayes Theorem, we can write,

$$f_{\Theta|X_1=x_1, X_2=x_2, .., X_n=x_n}(\theta \mid x_1, x_2, \ldots, x_n) = \frac{f_{(X_1, X_2, \ldots, X_n \mid \Theta=\theta)}(x_1, x_2, \ldots, x_n \mid \theta)\ g_\Theta(\theta)}{f_{(X_1, X_2, \ldots, X_n)}(x_1, x_2, \ldots, x_n)}$$

$$= \frac{[\prod_{i=1}^n f(x_i|\ \theta)]\ g_\Theta(\theta)}{\int [\prod_{i=1}^n f(x_i|\ \theta)] g_\Theta(\theta)\, d\theta}$$

where, $\theta \in \Theta$, a set of all possible values of θ

6. ***Bayes Estimator:*** Among different choices of Bayes estimator, here we are considering Posterior Mean as a Bayes Estimator which is given by,

$$E(\Theta|X_1 = x_1, X_2 = x_2, .., X_n = x_n) = \frac{\int \theta \, [\prod_{i=1}^{n} f(x_i|\, \theta)] \, g_\Theta(\theta) \, d\theta}{\int [\prod_{i=1}^{n} f(x_i|\, \theta)] g_\Theta(\theta) \, d\theta}$$

where, $\theta \in \Theta$, a set of all possible values of θ

7. ***Conjugate Prior:*** $g_\Theta(.)$ is said to the conjugate prior family if the corresponding posterior distribution $f_{\Theta|X_1=x_1, X_2=x_2, .., X_n=x_n}(\theta \mid x_1, x_2, \ldots., x_n)$ belongs to the same family as $g_\Theta(.)$.

8. ***Bootstrap:*** The bootstrap is a widely applicable and extremely powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator or statistical learning method. The power of the bootstrap lies in the fact that it can be easily applied to a wide range of statistical learning methods, including some for which a measure of variability is otherwise difficult to obtain and is not automatically output by statistical software. In various practical situations, we can only collect a sample of a particular size from a population. In such cases to estimate the standard error of an estimator of some unknown parameter of the population we use the Bootstrap method. The idea is, we consider the drawn sample as the population for the time being and we draw, say B (generally more than 100) number of random samples from it of size less than or equal to the original sample. These are called the Bootstrap samples. Now, for each such Bootstrap samples, we compute the value of the estimator, say $\hat{\alpha}_r$, the estimate of the $r-th$ Bootstrap sample, of the parameter of interest and ultimately, we calculate the standard deviation of the B such estimates and get the estimate of the standard error ($\widehat{SE(\alpha)}$) of the estimator of the population parameter.

$$\widehat{SE(\alpha)} = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}(\hat{\alpha}_r - \frac{1}{B}\sum_{r=1}^{B}\hat{\alpha}_r)^2}$$

Here, in this project, we have considered three distributions namely, Binomial Distribution, Poisson Distribution and Normal Distribution. For the distributions, we have computed the Maximum Likelihood Estimator in support of Classical approach and Posterior Mean (Bayes Estimator) in support of Bayesian approach to estimate some unknown population parameter of interest. We have also computed the standard errors of the estimates for comparison between the two approaches. We have also discussed how the choice of appropriate conjugate prior distributions can lower the standard deviation of the Bayes estimator (posterior mean) and give us more accurate results.

When the population distribution follows $\boldsymbol{Binomial(m,p)}$, $\boldsymbol{0<p<1; m\epsilon\mathbb{N}}$:

Here we are considering the characteristic of the elements of the population, on which we have the interest, can be represented by a random variable X whose p.m.f, $f_X(.;p)$, is given by,

$$f_X(x;\mathrm{p}) = \binom{m}{x} p^x (1-p)^{(m-x)}, \qquad x = 1(1)m;\ \ 0<p<1$$

$$0 \quad ; \quad otherwise$$

Here, m is considered to be known but p is an unknown quantity, the parameter of interest.

Now, let us draw a random sample $(X_1, X_2, \ldots., X_n)$ of size n from the population. Let us consider, $(x_1, x_2, \ldots., x_n)$ to be the realisation of $(X_1, X_2, \ldots., X_n)$.

Finding an Estimator of p by Classical Approach:

In classical approach we consider p to be an unknown but fixed quantity. Here we will use the method of maximum likelihood to find an estimator of p in support of classical approach.

Likelihood Function:

The likelihood function of p is given by,

$$\begin{aligned} L(p) = L(p; x_1, x_2, \ldots., x_n) &= f(x_1, x_2, \ldots., x_n; p) \\ &= \prod_{i=1}^{n} \binom{m}{x_i} p^{\sum_{i=1}^n x_i} (1-p)^{(mn-\sum_{i=1}^n x_i)}, where\ 0<p<1 \end{aligned}$$

The estimate of the maximum likelihood estimator of p is obtained by maximizing the likelihood function with respect to p, which is obtained as,

$$\widehat{\boldsymbol{p_{MLE}}} = \frac{\bar{\boldsymbol{x}}}{\boldsymbol{m}} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i}}{\boldsymbol{mn}}$$

As the p.m.f $f_X(x;\mathrm{p})$ follows certain regularity conditions and p_{MLE} is the Maximum Likelihood Estimator of p for a random sample of size n from $f_X(x;\mathrm{p})$, then for large n,

$$p_{MLE} \ \ \widetilde{a}\ \ N(p, 1/I(p))$$

$$\begin{aligned} where, I(p) &= E\left(\frac{\partial}{\partial p}\ln\big(f(x_1, x_2, \ldots., x_n; p)\big)\right)^2 \\ &= -E\left(\frac{\partial^2}{\partial p^2}\ln\big(f(x_1, x_2, \ldots., x_n; p)\big)\right) \\ &= -E\left(\frac{\partial^2}{\partial p^2}\left\{\sum_{i=1}^{n} ln\binom{m}{x_i} + \sum_{i=1}^{n} x_i \ln p + \left(mn - \sum_{i=1}^{n} x_i\right)\ln(1-p)\right\}\right) \end{aligned}$$

$$= E\left(\frac{\sum_{i=1}^{n} x_i}{p^2} + \frac{(mn - \sum_{i=1}^{n} x_i)}{(1-p)^2}\right)$$

$$= \frac{\sum_{i=1}^{n} mp}{p^2} + \frac{(mn - \sum_{i=1}^{n} mp)}{(1-p)^2}$$

$$= \frac{mn}{p} + \frac{mn}{1-p}$$

$$= \frac{mn}{p(1-p)}$$

Thus, the variance of p_{MLE} is given by,

$$var(p_{MLE}) = \frac{p(1-p)}{mn}$$

Now to find an unbiased estimator of $var(p_{MLE})$, we need an unbiased estimator of $p(1-p)$.

To find an unbiased estimator of $p(1-p)$:

Here we have, $X_i's$ are independently and identically distributed as $Bin(m,p)$ distribution, $\forall i = 1(1)n$.

Therefore $T = \sum_{i=1}^{n} x_i \sim Bin(mn, p)$ distribution.

Now, from the result of factorial moments of order r about zero for a $Bin(mn, p)$ distribution, we have,

$$E((T)_r) = (mn)_r p^r$$

$$where, (T)_r = T(T-1)\ldots.(T-r+1)$$
$$(mn)_r = mn(mn-1)\ldots.(mn-r+1)$$

Therefore, an unbiased estimator for p is $\frac{T}{mn}$ and an unbiased estimator for p^2 is $\frac{T(T-1)}{mn(mn-1)}$.

Thus, an unbiased estimator for $p(1-p)$ is $\frac{T(mn-T)}{mn(mn-1)}$.

Therefore, estimate of the standard error of p_{MLE} is given by,

$$\widehat{SE(p_{MLE})} = \sqrt{\frac{\sum_{i=1}^{n} x_i\,(mn - \sum_{i=1}^{n} x_i)}{(mn)^2(mn-1)}}$$

Finding an Estimator of p by Bayesian Approach:

Unlike classical approach, here in Bayesian approach we consider the parameter p to be an unknown but a random quantity. Therefore, p itself is a random variable and it will have some probability distribution. As, p can take any values between 0 and 1, we can consider p has a $Beta(a, b)$ distribution(prior) of 1st kind. Then the p.d.f of p, $g_P(.)$, is given by,

$$g_P(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} \ , 0 < p < 1; a, b > 0$$

$$0 \qquad\qquad\qquad , otherwise$$

Here, a and b both are considered to be known.

The posterior distribution of p is given by,

$$f_P(p \mid x_1, x_2, \dots., x_n) = \frac{f_{(X_1,X_2,\dots.,X_n \mid P=p)}(x_1, x_2, \dots., x_n \mid p)\ g_P(p)}{f_{(X_1,X_2,\dots.,X_n)}(x_1, x_2, \dots., x_n)}$$

$$= \frac{\prod_{i=1}^{n} f(x_i \mid p)\ g_P(p)}{\int_0^1 \prod_{i=1}^{n} f(x_i \mid p)\ g_P(p)\, dp} \quad \dots\dots\dots.(i)$$

Now,

$$\int_0^1 \prod_{i=1}^{n} f(x_i \mid p)\ g_P(p)\, dp = \int_0^1 \prod_{i=1}^{n} \binom{m}{x_i} p^{\sum_{i=1}^n x_i} (1-p)^{(mn-\sum_{i=1}^n x_i)} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp$$

$$= \prod_{i=1}^{n} \binom{m}{x_i} \frac{1}{B(a,b)} \int_0^1 p^{\sum_{i=1}^n x_i + a - 1} (1-p)^{mn-\sum_{i=1}^n x_i + b - 1}\, dp$$

$$= \prod_{i=1}^{n} \binom{m}{x_i} \frac{1}{B(a,b)} B\left(\sum_{i=1}^{n} x_i + a, mn - \sum_{i=1}^{n} x_i + b\right)$$

Now, from (i) we get,

$$f_P(p \mid x_1, x_2, \dots., x_n) = \frac{\prod_{i=1}^{n} \binom{m}{x_i} p^{\sum_{i=1}^n x_i} (1-p)^{(mn-\sum_{i=1}^n x_i)} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}}{\prod_{i=1}^{n} \binom{m}{x_i} \frac{1}{B(a,b)} B(\sum_{i=1}^n x_i + a, mn - \sum_{i=1}^n x_i + b)}$$

$$= \frac{p^{(\sum_{i=1}^n x_i + a) - 1} (1-p)^{(mn-\sum_{i=1}^n x_i + b) - 1}}{B(\sum_{i=1}^n x_i + a, mn - \sum_{i=1}^n x_i + b)}$$

Thus, the posterior distribution of $\boldsymbol{p}$ follows $\boldsymbol{Beta}(\sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{a}, \boldsymbol{mn} - \sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{b})$ of 1st kind.

As, the posterior distribution and prior distribution of p belongs to the same family, $g_P(.)$ is called the conjugate prior family.

Now, the estimate of the Bayes Estimator of p is given by the posterior mean, which is given by,

$$\widehat{\boldsymbol{p_b}} = \frac{\boldsymbol{a} + \sum_{i=1}^{n} \boldsymbol{x_i}}{\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{mn}}$$

To Obtain the Standard Error of the Bayes Estimator p_b:

We obtain the standard error of p_b by Bootstrap method, which is a method of re-sampling. Here we consider the drawn random sample $(X_1, X_2, \ldots, X_n)$ of size n as a population and we draw B number of samples of size n with replacement from the random sample $(X_1, X_2, \ldots, X_n)$. For each such B bootstrap samples we get an estimate of p_b. The standard deviation of the B values of the estimates $\widehat{p_b}$ can be considered as the standard error of the Bayes estimator p_b for estimating p.

Choice of Prior Distributions:

Bayes theorem gives us a method to revise our (belief) distribution about the parameter, given the data. In order to use it, one must have a distribution that represents our belief about the parameter, before we look at the data. This is our prior distribution. Here we will propose some methods to choose the prior, as well as things to consider in prior choice.

Choosing a Conjugate Prior When We Have Vague Prior Knowledge:

The shape of the $Beta(a, b)$ distribution of 1st kind changes with respect to different choices of the parameters a and b.

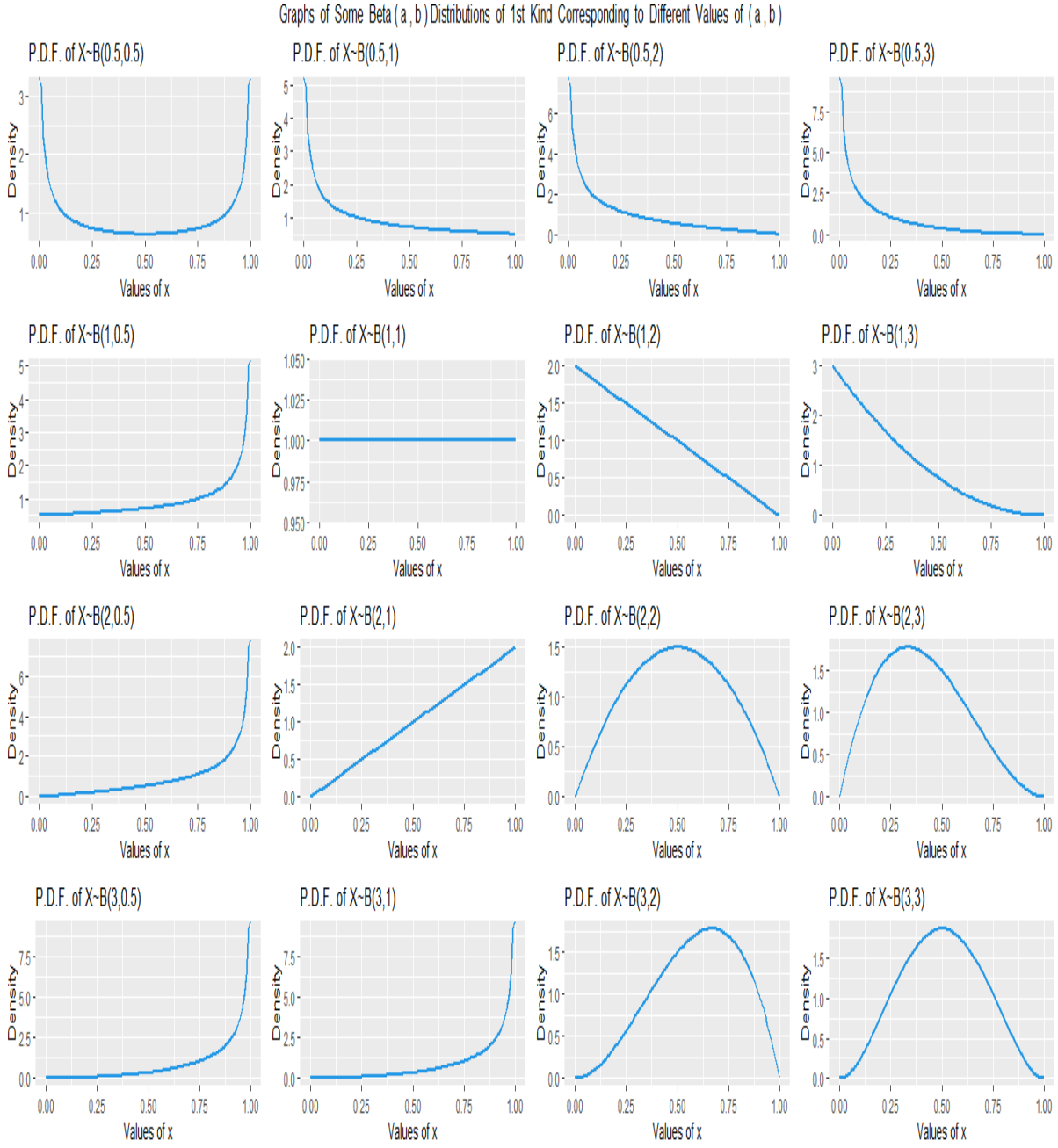
When $a < b$, the density has more weight in the lower half with respect to the range of the random variable following $Beta(a, b)$. That means, the values of the random variable close to 0 will have high probability density compared to that of the values close to 1. So, if we have the prior belief that the parameter under study $p, 0 < p < 1$ can be considered as a random variable and on an average, it takes the lower value, then we should consider such a prior distribution $Beta(a, b)$ which assigns high density towards the lower values of p, that is we should take a $Beta(a, b)$ prior distribution such that $a < b$.

On the other hand, when $a > b$, the density has more weight in the upper half with respect to the range of the random variable following $Beta(a, b)$. That means, the values of the random variable close to 1 will have high probability density compared to that of the values close to 0. So, if we have the prior belief that the parameter under study $p, 0 < p < 1$ can be considered as a random variable and on an average, it takes the higher value, then we should consider such

a prior distribution $Beta(a, b)$ which assigns high density towards the higher values of p, that is we should take a $Beta(a, b)$ prior distribution such that $a > b$.

The following figure shows the shapes of $Beta(a, b)$ densities for values of $a = 0.5, 1, 2, 3$ and $b = 0.5, 1, 2, 3$. This shows the some of the shapes that the members of the $Beta(a, b)$ family can take.

Figure 1:



Note: When both a and b are 1, then the prior is actually the uniform distribution. When we do not have any prior knowledge about p then one can not make any preferences about the

different values of p. In that case we should assume that all the values of p are equally probable and to incorporate this, we use $Beta(1,1)$ distribution of 1st kind (uniform prior).

Choosing a Conjugate Prior When We Have Real Prior Knowledge:

Beta distribution can take various shapes depending on the values of its parameters (shown in the ***figure 1***). The prior chosen should correspond to our belief about the parameter of interest p. Now, by saying we have real prior knowledge, we mean that, from some previous survey or from some other sources, we can strongly belief that the mean of p is p_0 and the standard deviation of p is σ_0. Now, we suggest choosing a $Beta(a,b)$ of 1st kind that matches our prior belief about the (location) mean and (scale) standard deviation.

The mean of $Beta(a,b)$ distribution is $\frac{a}{a+b}$. We, set this equal to our prior belief about the mean of the proportion p. So, we have,

$$p_0 = \frac{a}{a+b} \dots\dots\dots\dots\dots (1)$$

Again, the standard deviation of $Beta(a,b)$ is $\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$. We, set this equal to our prior belief about the standard deviation of the proportion p. So, we have,

$$\sigma_0 = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}} \dots\dots\dots\dots\dots\dots (2)$$

Solving equations (1) and (2), we can get a proper choice of a and b.

Precautions before using the conjugate prior:

- First, we have to graph our $Beta(a,b)$ prior. If the shape looks reasonably close to what we believe, we will use it. Otherwise, we can adjust p_0 and σ_0 until we find a prior whose graph approximately corresponds to our belief. As long as the prior has reasonable probability over the whole range of values that we think the parameter could possibly be in, it will be a satisfactory prior.
- ***Calculating the equivalent sample size of the prior:*** We note that, the sample proportion $p = \frac{\sum_{i=1}^{n} X_i}{mn}$ from a $Bin(m,p)$ distribution has variance equal to $\frac{p(1-p)}{mn}$. Now, we equate this variance (at p_0, prior mean) to the prior variance. So, we will have,

  $$\frac{p_0(1-p_0)}{mn_{eq}} = \frac{ab}{(a+b)^2(a+b+1)}$$

  Since, $p_0 = \frac{a}{a+b}$ and $(1-p_0) = \frac{b}{a+b}$, the equivalent sample size $n_{eq} = \frac{(a+b+1)}{m}$. It says that the amount of information about the parameter from our prior is equivalent to the amount from a random sample of that size. One should always check if this is unrealistically high. If it is so, we should increase the prior standard deviation and recalculate the prior. Otherwise, we would be putting too much prior information about the parameter relative to the amount of information that will come from the data.

Illustrating Example:

Let us consider X to be a random variable following a $Bin\ (m = 10, p)$ distribution, where $0 < p < 1$ but p is unknown to us (parameter of interest). To understand the performance of the estimators corresponding to Classical and Bayesian approach, we considered $p = 0.768$ and then after drawing a random sample from $Bin\ (15, 0.768)$ we find the estimates of the estimators obtained corresponding to the above-mentioned approaches and their standard errors for comparison.

We draw a random sample of size $n = 10$ from $Bin\ (15, 0.768)$ distribution. The sample comes out to be $(8, 7, 8, 9, 7, 9, 8, 8, 7, 7)$.

The estimate of the maximum likelihood estimators $\widehat{p_{MLE}} = \frac{\bar{x}}{m} = \frac{\sum_{i=1}^{n} x_i}{mn}$ comes out to be 0.78

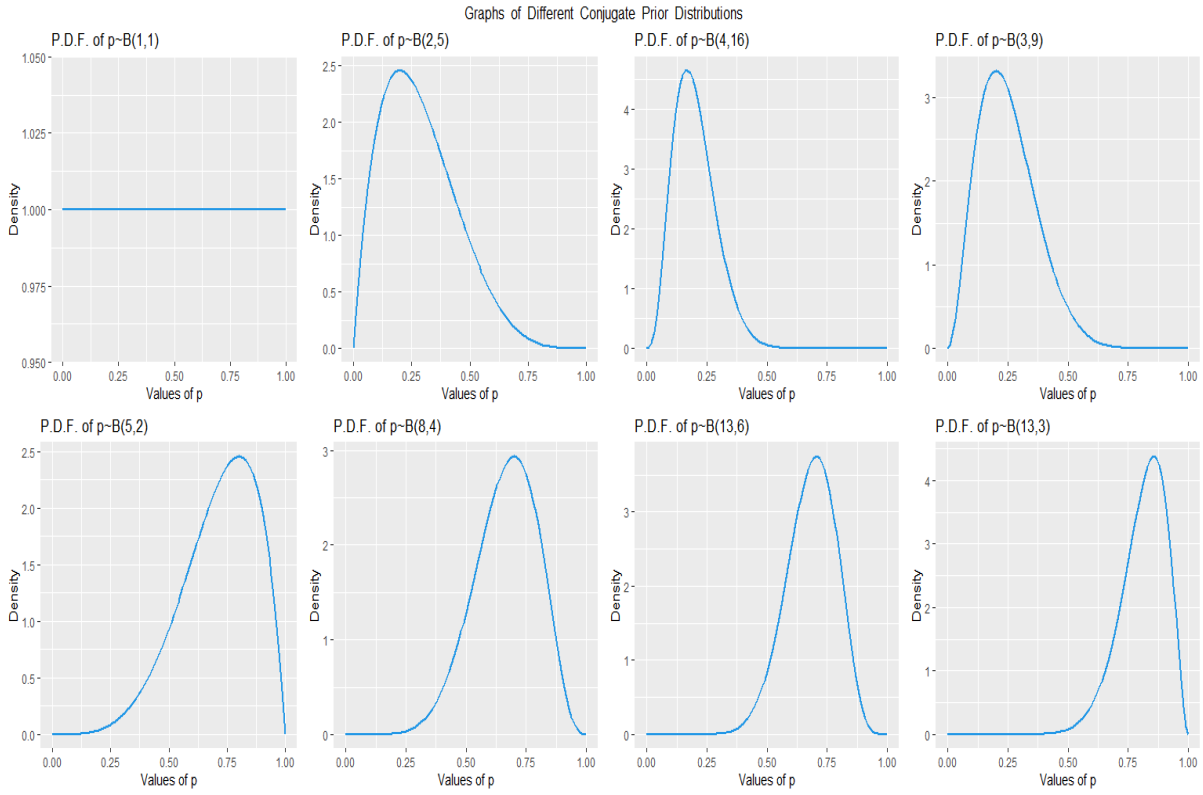
The standard error of p_{MLE} comes out to be $SE\widehat{(p_{MLE})} = \sqrt{\frac{\sum_{i=1}^{n} x_i (mn - \sum_{i=1}^{n} x_i)}{(mn)^2 (mn-1)}} = 0.04163332$

To illustrate the proper choice of conjugate priors we considered eight conjugate prior distributions which are given by,

$Beta(1,1), Beta(0.5,5), Beta(4,16), Beta(3,9), Beta(5,2), Beta(8,4), Beta(13,6), Beta(13,3)$

The graphs of the different conjugate priors are given below,

Figure 2:



Now, for each conjugate prior distribution we obtained the corresponding posterior distributions and obtained the posterior means which corresponds to the Bayes Estimator for estimating P.

To obtain the standard errors of the Bayes Estimators, we used a method of resampling, bootstrap. We used B=1000 Bootstrap samples to estimate the standard errors of posterior means corresponding to the different priors.

Suppose we have an additional information that the random quantity p takes values around 0.8 on an average. So, we can consider that on an average p assumes high values (around 0.6 to 0.9). In that case, we should choose such conjugate priors which takes high probability density towards the values closer to 1. Otherwise, the Bayes estimator will give estimates not closer to the actual value of the parameter or even if it gives close estimates, the standard errors will be comparatively high. This scenario can be observed through the graphs of probability densities corresponding to the different priors and also from the computations shown in the table below.

The estimate of maximum likelihood estimator, its standard error and the estimates of the Bayes estimator along with its standard error for different conjugate priors are represented in a tabular form which is given below.

Table 1

Value of the parameter of interest (p)	Estimate of the maximum likelihood estimator $(\widehat{p_{MLE}})$	Standard error of maximum likelihood estimator $SE(\widehat{p_{MLE}})$	Estimate of the Bayes Estimator (posterior mean) $(\widehat{p_b})$	Standard error of the Bayes Estimator $SE(\widehat{p_b})$	Conjugate Priors
0.768	0.780000	0.04163332	0.7745098	0.02378058	$Beta(1,1)$
0.768	0.780000	0.04163332	0.7440758	0.02192890	$Beta(0.5,5)$
0.768	0.780000	0.04163332	0.6833333	0.01991244	$Beta(4,16)$
0.768	0.780000	0.04163332	0.7232143	0.02101257	$Beta(3,9)$
0.768	0.780000	0.04163332	0.7757009	0.02286907	$Beta(5,2)$
0.768	0.780000	0.04163332	0.7678571	0.02112012	$Beta(8,4)$
0.768	0.780000	0.04163332	0.7647059	0.01969165	$Beta(13,6)$
0.768	0.780000	0.04163332	0.7844828	0.02048502	$Beta(13,3)$

Findings:

- If we incorporate the additional information about p that it is itself a random variable, then Bayes estimator performs better than Maximum Likelihood estimator in terms of standard errors of the estimators.
- For all the conjugate priors, the standard error of the Bayes estimators is lower than that of the Maximum Likelihood estimators.
- As the value of p is 0.768 which is close to 1, the conjugate priors that have more weight in the upper half or for that $Beta(a, b)$ prior for which $a > b$, are more appropriate than the others. From the graphs of different conjugate priors, it is seen that

$Beta(5,2), Beta(8,4), Beta(13,6), Beta(13,3)$ conjugate prior distributions have high density in the region $0.6 < p < 0.9$. So, by considering these priors we can get better Bayes estimators than the others, which can be seen from the ***Table 1*** given above.

- If we do not prefer any values of p over the others, which means if we are assuming that the all-possible values of p are equally probable ($Beta(1,1)$ prior), then the corresponding Bayes estimator yields largest standard error than the other Bayes estimators corresponding to the different conjugate priors.

<u>To choose a conjugate prior when we have real prior knowledge</u>:

Suppose, from some previous survey or from some other sources, we can strongly belief that the mean of p is $p_0 = 0.71$ and the standard deviation of p is $\sigma_0 = 0.0569$. Now, we need to choose a $Beta(a, b)$ of 1st kind that matches our prior belief about the (location) mean and (scale) standard deviation. Now, we equate,

$$0.71 = \frac{a}{a+b} \dots\dots\dots\dots\dots (1)$$

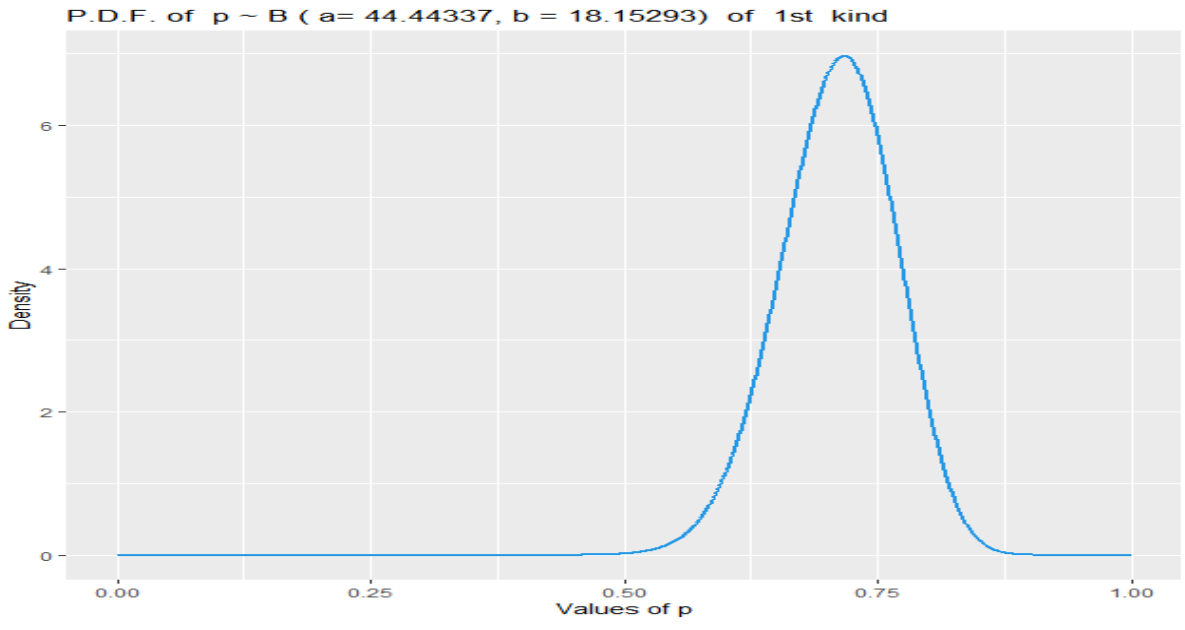
$$0.0569 = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}} \dots\dots\dots\dots\dots (2)$$

Solving the above two equations (1) and (2), we get the estimates of a and b as,

$$a = 44.44337 \ and \ b = 18.15293$$

The graph of the probability density function of $Beta(a = 44.44337, b = 18.15293)$ is given below,

<u>Figure 3:</u>



Considering the conjugate prior distribution $Beta(a = 44.44337, b = 18.15293)$ of 1st kind, the estimate of the Bayes Estimator (posterior mean) is given by,

$$\widehat{p_b} = 0.7530514$$

Also, the estimate of the standard error of p_b is given by,

$$SE(\widehat{p_b}) = 0.01463382$$

Calculating the equivalent sample size:

As mentioned in the theory, the expression of the equivalent sample size is,

$$n_{eq} = \frac{(a + b + 1)}{m}$$

From the above relation, the equivalent sample size is obtained as,

$$n_{eq} = 6.359629$$

It says that the amount of information about the parameter p from our prior is equivalent to the amount from a random sample of size approximately equal to 6. As, the value of equivalent sample size is not unrealistically high, we can say that our prior belief seems to be appropriate.

When the population distribution follows $\boldsymbol{Poisson}\ (\boldsymbol{\lambda});\ \boldsymbol{\lambda > 0}$:

Here we are considering the characteristic of the elements of the population, on which we have the interest, can be represented by a random variable X whose p.m.f $f_X(.;\lambda)$, is given by,

$$f_X(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!};\ \ x > 0, \lambda > 0$$

$$0 \quad ; \quad otherwise$$

Here, λ is an unknown quantity, the parameter of interest.

Now, let us draw a random sample $(X_1, X_2, \ldots, X_n)$ of size n from the population. Let us consider, $(x_1, x_2, \ldots, x_n)$ to be the realisation of $(X_1, X_2, \ldots, X_n)$.

Finding an Estimator of λ by Classical Approach:

In classical approach we consider λ to be an unknown but fixed quantity. Here we will use the method of maximum likelihood to find an estimator of λ in support of classical approach.

Likelihood Function:

The likelihood function of λ is given by,

$$L(\lambda) = L(\lambda; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \lambda)$$

$$= e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\ ; where\ \lambda > 0$$

The estimate of the maximum likelihood estimator of λ is obtained by maximizing the likelihood function with respect to λ, which is obtained as,

$$\widehat{\boldsymbol{\lambda_{MLE}}} = \boldsymbol{\bar{x}} = \frac{\boldsymbol{\sum_{i=1}^n x_i}}{\boldsymbol{n}}$$

As the p.m.f $f_X(x;\lambda)$ follows certain regularity conditions and λ_{MLE} is the Maximum Likelihood Estimator of p for a random sample of size n from $f_X(x;\lambda)$, then for large n,

$$\lambda_{MLE}\ \widetilde{a}\ N(\lambda, 1/I(\lambda))$$

$$where, I(\lambda) = E\left(\frac{\partial}{\partial\lambda}\ln\big(f(x_1, x_2, \ldots, x_n;\lambda)\big)\right)^2$$

$$= -E\left(\frac{\partial^2}{\partial\lambda^2}\ln\big(f(x_1, x_2, \ldots, x_n;\lambda)\big)\right)$$

$$= -E\left(\frac{\partial^2}{\partial\lambda^2}\left\{-n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln x_i !\right\}\right)$$

$$= E\left(\frac{\sum_{i=1}^{n} x_i}{\lambda^2}\right)$$

$$= \frac{\sum_{i=1}^{n} E(x_i)}{\lambda^2}$$

$$= \frac{n\lambda}{\lambda^2}$$

$$= \frac{n}{\lambda}$$

Thus, the variance of λ_{MLE} is given by,

$$var(\lambda_{MLE}) = \frac{\lambda}{n}$$

Now to find an unbiased estimator of $var(\lambda_{MLE})$, we need an unbiased estimator of λ.

Again, we know that, $E(\bar{X}) = \lambda$

Therefore, the estimate of the standard error of λ_{MLE} is given by,

$$\boldsymbol{\widehat{SE(\lambda_{MLE})}} = \sqrt{\frac{\bar{x}}{n}} = \frac{\sqrt{\sum_{i=1}^{n} x_i}}{n}$$

Finding an Estimator of λ by Bayesian Approach:

Unlike classical approach, here in Bayesian approach we consider the parameter λ to be an unknown but a random quantity. Therefore, λ itself is a random variable and it will have some probability distribution. As, λ can take any values greater than 0, we can consider λ has a $Gamma(m, \theta)$ distribution(prior). Then the p.d.f of λ, $g_\lambda(.)$, is given by,

$$g_\lambda(\lambda) = \frac{\theta^m e^{-\lambda\theta}\lambda^{m-1}}{\Gamma(m)} \; ; \quad \lambda > 0; \theta, m > 0$$

$$0 \qquad\qquad\qquad , otherwise$$

Here, m and θ both are considered to be known.

The posterior distribution of λ is given by,

$$f_{\boldsymbol{\lambda}}(\lambda \mid x_1, x_2, \ldots., x_n) = \frac{f_{(X_1,X_2,\ldots.,X_n \mid \boldsymbol{\lambda}=\lambda)}(x_1, x_2, \ldots., x_n \mid \lambda)\, g_{\lambda}(\lambda)}{f_{(X_1,X_2,\ldots.,X_n)}(x_1, x_2, \ldots., x_n)}$$

$$= \frac{\prod_{i=1}^{n} f(x_i \mid \lambda)\, g_{\boldsymbol{\lambda}}(\lambda)}{\int_0^{\infty} \prod_{i=1}^{n} f(x_i \mid \lambda)\, g_{\boldsymbol{\lambda}}(\lambda)\, d\lambda} \quad \ldots\ldots\ldots (i)$$

Now,

$$\int_0^{\infty} \prod_{i=1}^{n} f(x_i \mid \lambda)\, g_{\boldsymbol{\lambda}}(\lambda)\, d\lambda = \int_0^{\infty} e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \frac{\theta^m e^{-\lambda\theta} \lambda^{m-1}}{\Gamma(m)}\, d\lambda$$

$$= \frac{\theta^m}{\prod_{i=1}^{n} x_i!\, \Gamma(m)} \int_0^{\infty} e^{-\lambda(n+\theta)} \lambda^{(\sum_{i=1}^{n} x_i + m)-1}\, d\lambda$$

$$= \frac{\theta^m}{\prod_{i=1}^{n} x_i!\, \Gamma(m)} \frac{\Gamma(\sum_{i=1}^{n} x_i + m)}{(n+\theta)^{(\sum_{i=1}^{n} x_i + m)}}$$

Now, from (i) we get,

$$f_{\lambda}(\lambda \mid x_1, x_2, \ldots., x_n) = \frac{e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \frac{\theta^m e^{-\lambda\theta} \lambda^{m-1}}{\Gamma(m)}}{\frac{\theta^m}{\prod_{i=1}^{n} x_i!\, \Gamma(m)} \frac{\Gamma(\sum_{i=1}^{n} x_i + m)}{(n+\theta)^{(\sum_{i=1}^{n} x_i + m)}}}$$

$$= \frac{(n+\theta)^{(\sum_{i=1}^{n} x_i + m)} e^{-\lambda(n+\theta)} \lambda^{(\sum_{i=1}^{n} x_i + m)-1}}{\Gamma(\sum_{i=1}^{n} x_i + m)}$$

Thus, the posterior distribution of λ follows $\boldsymbol{Gamma}(\sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{m}, \boldsymbol{n} + \boldsymbol{\theta})$.

As, the posterior distribution and prior distribution of λ belongs to the same family, $g_{\lambda}(.)$ is called the conjugate prior family.

Now, the estimate of the Bayes Estimator of λ is given by the posterior mean, which is given by,

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{b}} = \frac{\sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{m}}{\boldsymbol{n} + \boldsymbol{\theta}}$$

To Obtain the Standard Error of the Bayes Estimator λ_b:

We obtain the standard error of λ_b by Bootstrap method, which is a method of re-sampling. Here we consider the drawn random sample $(X_1, X_2, \ldots, X_n)$ of size n as a population and we draw B number of samples of size n with replacement from the random sample $(X_1, X_2, \ldots, X_n)$. For each such B bootstrap samples we get an estimate of λ_b. The standard deviation of the B number of estimates $\widehat{\lambda_b}$ can be considered as the standard error of the Bayes estimator λ_b for estimating λ.

Choice of Prior Distributions:

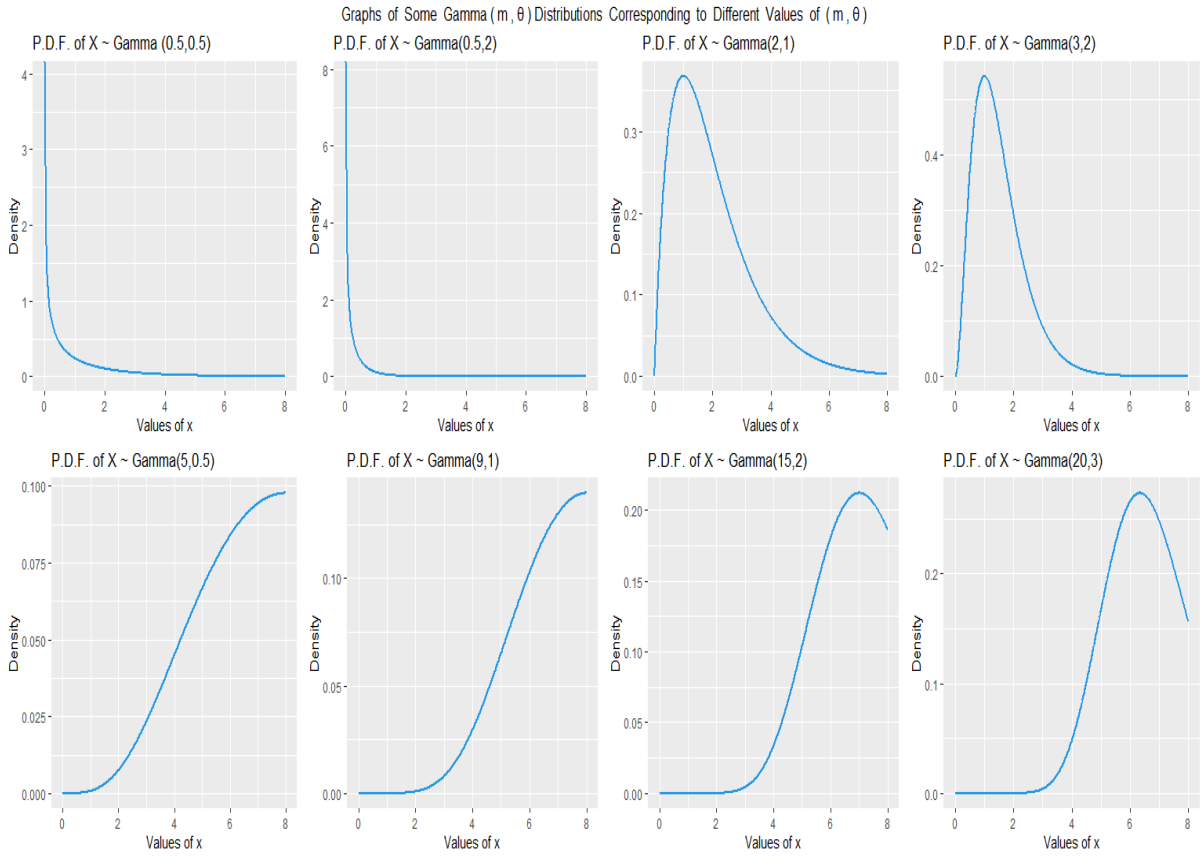
Bayes theorem gives us a method to revise our (belief) distribution about the parameter, given the data. In order to use it, one must have a distribution that represents our belief about the parameter, before we look at the data. This is our prior distribution. Here we will propose some methods to choose the prior, as well as things to consider in prior choice.

Choosing a Conjugate Prior When We Have Vague Prior Knowledge:

The shape of the $Gamma(m, \theta)$ distribution changes with respect to different choices of the parameters m and θ. One should consider an appropriate conjugate prior $Gamma(m, \theta)$ that matches with our prior belief. For example, if we have the prior belief that the parameter under study λ; $\lambda > 0$ can be considered as a random variable and on an average, it takes the lower value, then we should consider such a prior distribution $Gamma(m, \theta)$ which assigns high density towards the lower values of λ. Similarly, if we have the prior belief that the parameter under study λ; $\lambda > 0$ can be considered as a random variable and on an average, it takes the higher value, then we should consider such a prior distribution $Gamma(m, \theta)$ which assigns high density towards the higher values of λ.

The following figure shows the shapes of $Gamma(m, \theta)$ densities for values of (m, θ): $(0.5,0.5), (0.5,2), (2,1), (3,2), (5,0.5), (9,2), (15,2), (20,3)$. This shows some of the shapes that the members of the $Gamma(m, \theta)$ family can take.

Figure 4:



Note: Suppose we have no idea what is the value of λ in prior to looking at the data. In that case, we should give all the values of λ equal weight. So, we consider the uniform prior density, which is given by,

$$g_\lambda(\lambda) = 1; for\ \lambda > 0$$

Clearly this prior density is improper since its integral over all possible values is infinite. Nevertheless, the posterior will be proper in this case and we can use it for making inference about λ. In this case, the posterior distribution will follow $Gamma(\sum_{i=1}^{n} x_i + 1\ , n)$.

Choosing a Conjugate Prior When We Have Real Prior Knowledge:

Gamma distribution can take various shapes depending on the values of its parameters (shown in the ***figure 4***). The prior chosen should correspond to our belief about the parameter of interest λ. Now, by saying we have real prior knowledge, we mean that, from some previous survey or from some other sources, we can strongly belief that the mean of λ is λ_0 and the standard deviation of λ is σ_0. Now, we suggest choosing a $Gamma(m, \theta)$ that matches our prior belief about the (location) mean and (scale) standard deviation.

The mean of $Gamma(m, \theta)$ distribution is $\frac{m}{\theta}$. We, set this equal to our prior belief about the mean of the proportion λ. So, we have,

$$\lambda_0 = \frac{m}{\theta} \dots\dots\dots\dots\dots (1)$$

Again, the standard deviation of $Gamma(m, \theta)$is $\sqrt{\frac{m}{\theta^2}}$. We, set this equal to our prior belief about the standard deviation of the proportion λ. So, we have,

$$\sigma_0 = \sqrt{\frac{m}{\theta^2}} \dots\dots\dots\dots\dots\dots\dots (2)$$

Solving equations (1) and (2), we can get a proper choice of m and θ.

Precautions before using the conjugate prior:

- First, we have to graph our $Gamma(m, \theta)$ prior. If the shape looks reasonably close to what we believe, we will use it. Otherwise, we can adjust λ_0 and σ_0 until we find a prior whose graph approximately corresponds to our belief. As long as the prior has reasonable probability over the whole range of values that we think the parameter could possibly be in, it will be a satisfactory prior.

- ***Calculating the equivalent sample size of the prior:*** This is the size of a random sample of $Poisson(\lambda)$ variables that matches the amount of prior information about λ that we are putting in with our prior. We note that if $(x_1, x_2, \dots, x_n)$ is a random sample from $Poisson(\lambda)$ then $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ will have mean λ and variance $\frac{\lambda}{n}$. The equivalent sample size will be the solution of,

  $$\frac{\lambda}{n_{eq}} = \frac{m}{\theta^2}$$

  Setting the mean equal to the prior mean $\lambda = \frac{m}{\theta}$ the equivalent sample size of the $Gamma(m, \theta)$ prior for λ is $n_{eq} = \theta$. It says that the amount of information about the parameter from our prior is equivalent to the amount from a random sample of that size. One should always check if this is unrealistically high. If it is so, we should increase the prior standard deviation and recalculate the prior. Otherwise, we would be putting too much prior information about the parameter relative to the amount of information that will come from the data.

Illustrating Example:

Let us consider X to be a random variable following a $Poisson\,(\lambda)$ distribution, where $\lambda > 0$ but λ is unknown to us (parameter of interest). To understand the performance of the estimators corresponding to Classical and Bayesian approach, we considered $\lambda = 2$ and then after drawing a random sample from $Poisson\,(2)$ we find the estimates of the estimators obtained corresponding to the above-mentioned approaches and their standard errors for comparison.

We draw a random sample of size $n = 15$ from $Poisson\ (2)$ distribution. The sample comes out to be $(1, 3, 2, 1, 2, 0, 2, 1, 3, 3, 2, 2, 3, 2, 2)$.

The estimate of the maximum likelihood estimator $\widehat{\lambda_{MLE}} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ comes out to be 1.933333

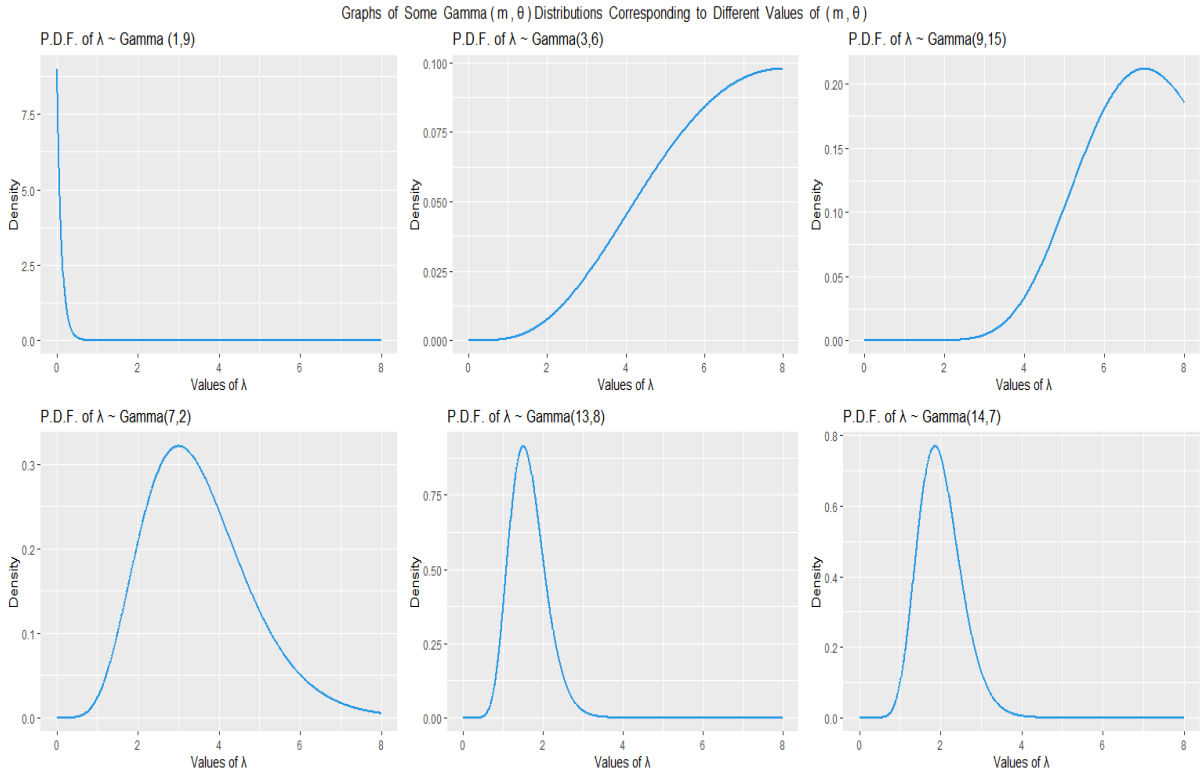
The estimate of the standard error of λ_{MLE} comes out to be $SE(\widehat{\lambda_{MLE}}) = \sqrt{\frac{\bar{x}}{n}} = \frac{\sqrt{\sum_{i=1}^{n} x_i}}{n} =$ 0.359011

To illustrate the proper choice of conjugate priors we considered six conjugate prior distributions which are given by,

$Gamma(1,9), Gamma(5,0.5), Gamma(15,2), Gamma(7,2), Gamma(13,8), Gamma(14,7)$

The graphs of the different conjugate priors are given below,

Figure 5:



Now, for each conjugate prior distribution we obtained the corresponding posterior distributions and obtained the posterior means which corresponds to the Bayes Estimator for estimating λ.

To obtain the standard errors of the Bayes Estimators, we used a method of resampling, bootstrap. We used B = 1000 Bootstrap samples to estimate the standard errors of posterior means corresponding to the different priors.

Suppose we have an additional information that the random quantity λ takes values around 1.75 on an average. So, we can consider that on an average λ assumes high densities (around

1 to 3). In that case, we should choose such conjugate priors which takes high probability density towards the values closer to 2. Otherwise, the Bayes estimator will give estimates not closer to the actual value of the parameter or even if it gives close estimates, the standard errors will be comparatively high. This scenario can be observed through the graphs of probability densities corresponding to the different conjugate priors and also from the computations shown in the table below.

The estimate of maximum likelihood estimator, its standard error and the values of the Bayes estimator along with its standard error for different conjugate priors are represented in a tabular form which is given below.

Table 2

Value of the parameter of interest (λ)	**Estimate of the maximum likelihood estimator ($\widehat{\lambda_{MLE}}$)**	**Standard error of maximum likelihood estimator $SE(\widehat{\lambda_{MLE}})$**	**Estimate of the Bayes Estimator (posterior mean) ($\widehat{\lambda_b}$)**	**Standard error of the Bayes Estimator $SE(\lambda_b)$**	**Conjugate Priors**
2	1.933333	0.359011	1.250000	0.1410430	$Gamma(1,9)$
2	1.933333	0.359011	2.193548	0.2183891	$Gamma(5,0.5)$
2	1.933333	0.359011	2.588235	0.1991195	$Gamma(15,2)$
2	1.933333	0.359011	2.117647	0.1991195	$Gamma(7,2)$
2	1.933333	0.359011	1.826087	0.1471753	$Gamma(13,8)$
2	1.933333	0.359011	1.954545	0.1538650	$Gamma(14,7)$
2	1.933333	0.359011	2	0.2256687	Uniform Prior: $g_\lambda(\lambda) = 1; for\ \lambda > 0$

Findings:

- If we incorporate the additional information about λ that it is itself a random variable, then Bayes estimator performs better than Maximum Likelihood estimator in terms of standard errors of the estimators.
- For all the conjugate priors, the standard error of the Bayes estimators is lower than that of the Maximum Likelihood estimators.
- As the actual value of λ is 2, the conjugate priors that have more weight in the lower half (around 2), are more appropriate than the others. From the graphs of different conjugate priors, it is seen that $Gamma(7,2), Gamma(13,8), Gamma(14,7)$ conjugate prior distributions have high density in the region $1 < \lambda < 3$. So, by considering these priors we can get better Bayes estimators than the others, which can be seen from the ***Table 2*** also, given above.
- If we do not prefer any values of λ over the others, which means if we are assuming that the all-possible values of λ are equally probable ($Uniform$ prior), then the corresponding Bayes estimator yields largest standard error than the other Bayes estimators corresponding to the different conjugate priors.

To choose a conjugate prior when we have real prior knowledge:

Suppose, from some previous survey or from some other sources, we can strongly belief that the mean of λ is $\lambda_0 = 1.75$ and the standard deviation of λ is $\sigma_0 = 0.0891$. Now, we need to choose a $Gamma(m, \theta)$ that matches our prior belief about the (location) mean and (scale) standard deviation. Now, we equate,

$$1.75 = \frac{m}{\theta} \dots\dots\dots\dots\dots (1)$$

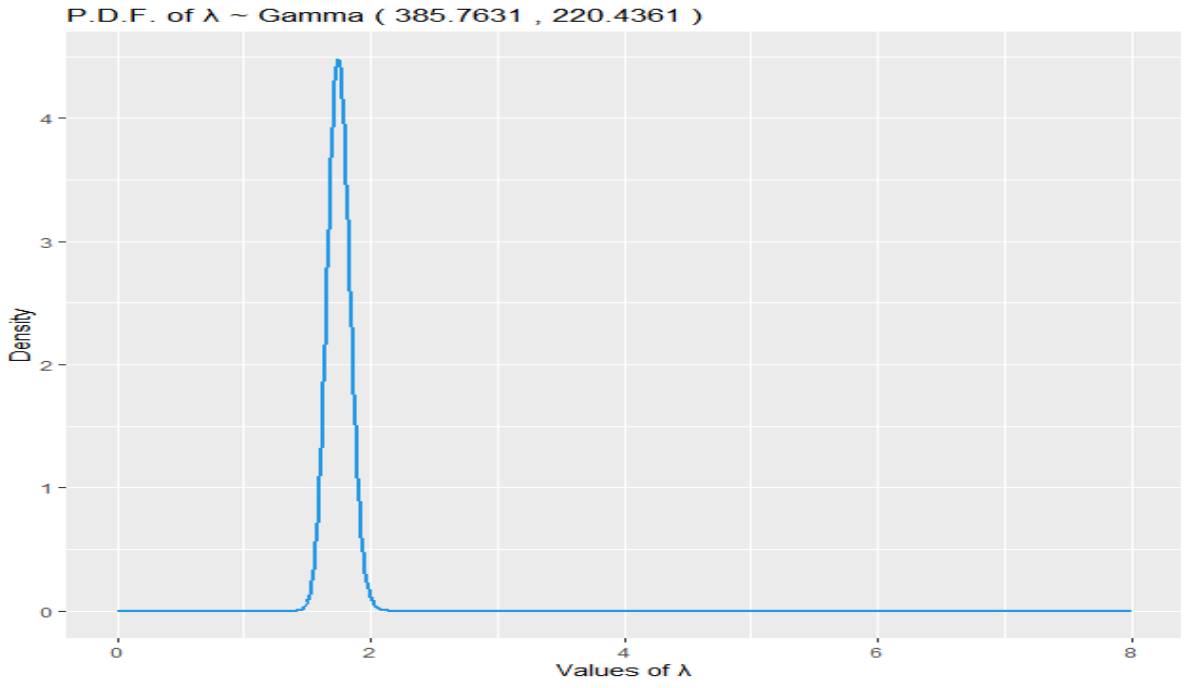
$$0.0891 = \sqrt{\frac{m}{\theta^2}} \dots\dots\dots\dots\dots (2)$$

Solving the above two equations (1) and (2), we get the estimates of m and θ as,

$$m = 385.7631 \ and \ \theta = 220.4361$$

The graph of the probability density function of $Gamma(m = 385.7631, \theta = 220.4361)$ is given below,

Figure 6:



Considering the conjugate prior distribution $Gamma(m = 385.7631, \theta = 220.4361)$, the estimate of the Bayes Estimator (posterior mean) is given by,

$$\widehat{\lambda_b} = 1.76168$$

Also, the estimate of the standard error of λ_b is given by,

$$\widehat{SE(\lambda_b)} = 0.01366563$$

Calculating the equivalent sample size:

As mentioned in the theory, the expression of the equivalent sample size is,

$$n_{eq} = \theta$$

From the above relation, the equivalent sample size is obtained as,

$$n_{eq} = 220.4361$$

It says that the amount of information about the parameter λ from our prior is equivalent to the amount from a random sample of size approximately equal to 221. As, the value of equivalent sample size is unrealistically high, we can say that our prior belief seems to be inappropriate as we are putting in too much prior information relative to the amount we will be getting from the data. So, we should suspect our prior belief and we need to increase the standard error and again need to obtain a proper conjugate prior.

To choose a conjugate prior after increasing the standard deviation corresponding to our prior belief:

Suppose, we take the mean of λ is $\lambda_0 = 1.75$ and the standard deviation of λ is $\sigma_0 = 1.991$. Now, we need to choose a $Gamma(m, \theta)$ that matches our prior belief about the (location) mean and (scale) standard deviation. Now, we equate,

$$1.75 = \frac{m}{\theta} \dots\dots\dots\dots\dots (1)$$

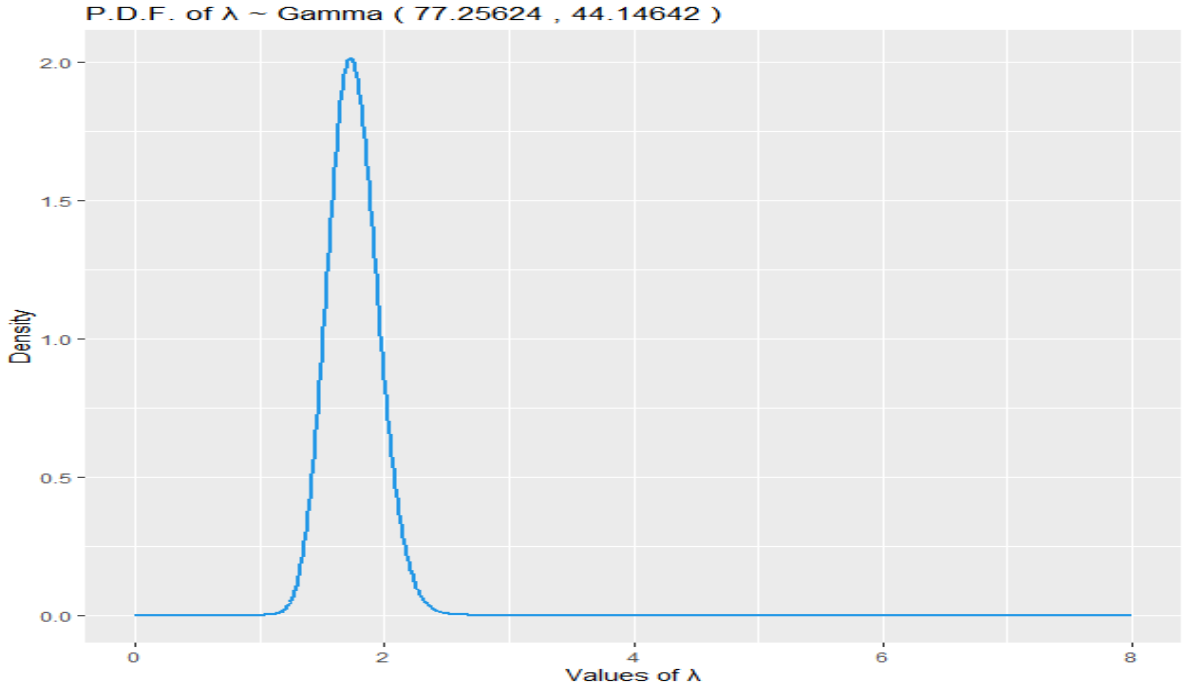
$$1.991 = \sqrt{\frac{m}{\theta^2}} \dots\dots\dots\dots\dots (2)$$

Solving the above two equations (1) and (2), we get the estimates of m and θ as,

$$m = 77.25624 \ and \ \theta = 44.14642$$

The graph of the probability density function of $Gamma(m = 77.25624 , \theta = 44.14642)$ is given below,

Figure 7:



Considering the conjugate prior distribution $Gamma(m = 77.25624 , \theta = 44.14642)$, the estimate of the Bayes Estimator (posterior mean) is given by,

$$\widehat{\lambda_b} = 1.796495$$

Also, the estimate of the standard error of λ_b is given by,

$$\widehat{SE(\lambda_b)} = 0.05605694$$

Calculating the equivalent sample size:

As mentioned in the theory, the expression of the equivalent sample size is,

$$n_{eq} = \theta$$

From the above relation, the equivalent sample size is obtained as,

$$n_{eq} = 44.14642$$

It says that the amount of information about the parameter λ from our prior is equivalent to the amount from a random sample of size approximately equal to 44. As, the value of equivalent sample size is not unrealistically high, we can say that our prior belief seems to be appropriate.

When the population distribution follows $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma^2}); -\infty < \boldsymbol{\mu} < \infty;\ \boldsymbol{\sigma} > 0$ *where variance* $\boldsymbol{\sigma^2}$ *is known but* $\boldsymbol{\mu}$ *is unknown*:

Here we are considering the characteristic of the elements of the population, on which we have the interest, can be represented by a random variable X whose p.d.f $f_X(.;\mu,\sigma^2)$, is given by,

$$f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mathbf{1}}{\mathbf{2}}\frac{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathbf{2}}}{\boldsymbol{\sigma^2}}}, where\ -\infty < x < \infty; -\infty < \mu < \infty;\ \sigma > 0$$

$$0 \quad ; \quad otherwise$$

Here, μ is an unknown quantity, the parameter of interest, but σ is known.

Now, let us draw a random sample $(X_1, X_2, \ldots, X_n)$ of size n from the population. Let us consider, $(x_1, x_2, \ldots, x_n)$ to be the realisation of $(X_1, X_2, \ldots, X_n)$.

Finding an Estimator of μ by Classical Approach:

In classical approach we consider μ to be an unknown but fixed quantity. Here we will use the method of maximum likelihood to find an estimator of μ in support of classical approach.

Likelihood Function:

The likelihood function of μ is given by,

$$L(\mu) = L(\mu; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \mu)$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\mathbf{1}}{\mathbf{2}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\frac{(\boldsymbol{x_i}-\boldsymbol{\mu})^{\mathbf{2}}}{\boldsymbol{\sigma^2}}}, where\ -\infty < x < \infty; -\infty < \mu < \infty;\ \sigma > 0$$

The estimate of the maximum likelihood estimator of μ is obtained by maximizing the likelihood function with respect to μ, which is obtained as,

$$\widehat{\boldsymbol{\mu_{MLE}}} = \bar{\boldsymbol{x}} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i}}{\boldsymbol{n}}$$

As we know, if $X_1, X_2, \ldots, X_n$ are independently or identically distributed as $\mathrm{N}(\mu, \sigma^2)$ distribution, then,

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Thus, the variance of the maximum likelihood estimator $\mu_{MLE} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is given by,

$$var(\mu_{MLE}) = var(\bar{X}) = \frac{\sigma^2}{n}$$

Now, an unbiased estimator of σ^2 is given by,

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Thus, estimate of the standard error of the maximum likelihood estimator of μ is given by,

$$\boldsymbol{SE\widehat{(\mu_{MLE})}} = \sqrt{\frac{\widehat{\sigma^2}}{n}} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Finding an Estimator of μ by Bayesian Approach:

Unlike classical approach, here in Bayesian approach we consider the parameter μ to be an unknown but a random quantity. Therefore, μ itself is a random variable and it will have some probability distribution. As, μ can take any values between $-\infty$ to ∞, we can consider μ has a $Normal(\mu_0, {\sigma_0}^2)$ distribution(prior). Then the p.d.f of μ, $g_\mu(.)$, is given by,

$$g_\mu(\mu) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1(\mu-\mu_0)^2}{2\ \ \sigma_0^2}}, where\ -\infty < \mu < \infty; -\infty < \mu_0 < \infty;\ \sigma_0 > 0$$

$$0 \quad ; \quad otherwise$$

Here, μ_0 and σ_0 both are considered to be known.

The posterior distribution of μ is given by,

$$f_\mu(\mu \mid x_1, x_2, \ldots, x_n) = \frac{f_{(X_1, X_2, \ldots, X_n \mid \mu=\mu)}(x_1, x_2, \ldots, x_n \mid \mu)\ g_\mu(\mu)}{f_{(X_1, X_2, \ldots, X_n)}(x_1, x_2, \ldots, x_n)}$$

$$= \frac{\prod_{i=1}^{n} f(x_i \mid \mu)\ g_\mu(\mu)}{\int_{-\infty}^{\infty}\prod_{i=1}^{n} f(x_i \mid \mu)\ g_\mu(\mu)\, d\mu} \quad \ldots\ldots\ldots (i)$$

Now,

$$\int_{-\infty}^{\infty}\prod_{i=1}^{n} f(x_i|\,\mu)\; g_\mu(\mu)\, d\mu = \int_{-\infty}^{\infty}\frac{1}{\sigma^n\sigma_0(2\pi)^{n+1}}\, e^{-\frac{1}{2}\left[\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}+\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right]}d\mu$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sigma^n\sigma_0(2\pi)^{n+1}}\, e^{-\frac{1}{2}\left[\mu^2\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)-2\mu\left(\frac{\sum_{i=1}^{n}x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)+\left(\frac{\sum_{i=1}^{n}{x_i}^2}{\sigma^2}+\frac{{\mu_0}^2}{\sigma_0^2}\right)\right]}d\mu$$

$$= \frac{e^{\frac{1}{2}\left[\frac{\left(\frac{\sum_{i=1}^{n}x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)^2}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}-\left(\frac{\sum_{i=1}^{n}{x_i}^2}{\sigma^2}+\frac{{\mu_0}^2}{\sigma_0^2}\right)\right]}}{\sigma^n\sigma_0(2\pi)^n}\sqrt{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}\;\frac{1}{\sqrt{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{\left(\mu-\frac{\left(\frac{\sum_{i=1}^{n}x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}\right)^2}{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}}d\mu$$

$$= \frac{e^{\frac{1}{2}\left[\frac{\left(\frac{\sum_{i=1}^{n}x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)^2}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}-\left(\frac{\sum_{i=1}^{n}{x_i}^2}{\sigma^2}+\frac{{\mu_0}^2}{\sigma_0^2}\right)\right]}}{\sigma^n\sigma_0(2\pi)^n}\sqrt{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}$$

Now, from (i) we get,

$$f_\mu(\mu\mid x_1, x_2, \ldots., x_n) = \frac{\frac{1}{\left(\sigma\sqrt{2\pi}\right)^n}e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}\frac{1}{\sigma_0\sqrt{2\pi}}e^{-\frac{1(\mu-\mu_0)^2}{2\sigma_0^2}}}{\frac{e^{\frac{1}{2}\left[\frac{\left(\frac{\sum_{i=1}^{n}x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)^2}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}-\left(\frac{\sum_{i=1}^{n}{x_i}^2}{\sigma^2}+\frac{{\mu_0}^2}{\sigma_0^2}\right)\right]}}{\sigma^n\sigma_0(2\pi)^n}\sqrt{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}}$$

$$= \frac{1}{\sqrt{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}\sqrt{2\pi}} e^{-\frac{1}{2}\left[\mu^2\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)-2\mu\left(\frac{\sum_{i=1}^n x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)+\frac{\left(\frac{\sum_{i=1}^n x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)^2}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}\right]}$$

$$= \frac{1}{\sqrt{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\left(\mu-\frac{\left(\frac{\sum_{i=1}^n x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}\right)^2}{\frac{1}{\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)}}}$$

Thus, the posterior distribution of μ follows $\boldsymbol{Normal}\left(\frac{\frac{\sum_{i=1}^n x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)$.

As, the posterior distribution and prior distribution of μ belongs to the same family, $g_\mu(.)$ is called the conjugate prior family of μ.

Now, the estimate of the Bayes Estimator of μ is given by the posterior mean, which is given by,

$$\widehat{\boldsymbol{\mu_b}} = \frac{\frac{\sum_{i=1}^n \boldsymbol{x_i}}{\boldsymbol{\sigma^2}} + \frac{\boldsymbol{\mu_0}}{\boldsymbol{\sigma_0^2}}}{\frac{\boldsymbol{n}}{\boldsymbol{\sigma^2}} + \frac{\boldsymbol{1}}{\boldsymbol{\sigma_0^2}}}$$

To Obtain the Standard Error of the Bayes Estimator μ_b:

We obtain the standard error of μ_b by Bootstrap method, which is a method of re-sampling. Here we consider the drawn random sample $(X_1, X_2, \ldots, X_n)$ of size n as a population and we draw B number of samples of size n with replacement from the random sample $(X_1, X_2, \ldots, X_n)$. For each such B bootstrap samples we get an estimate of μ_b. The standard deviation of the B values of μ_b can be considered as the standard error of the Bayes estimator μ_b for estimating μ.

Choice of Conjugate Prior Distributions:

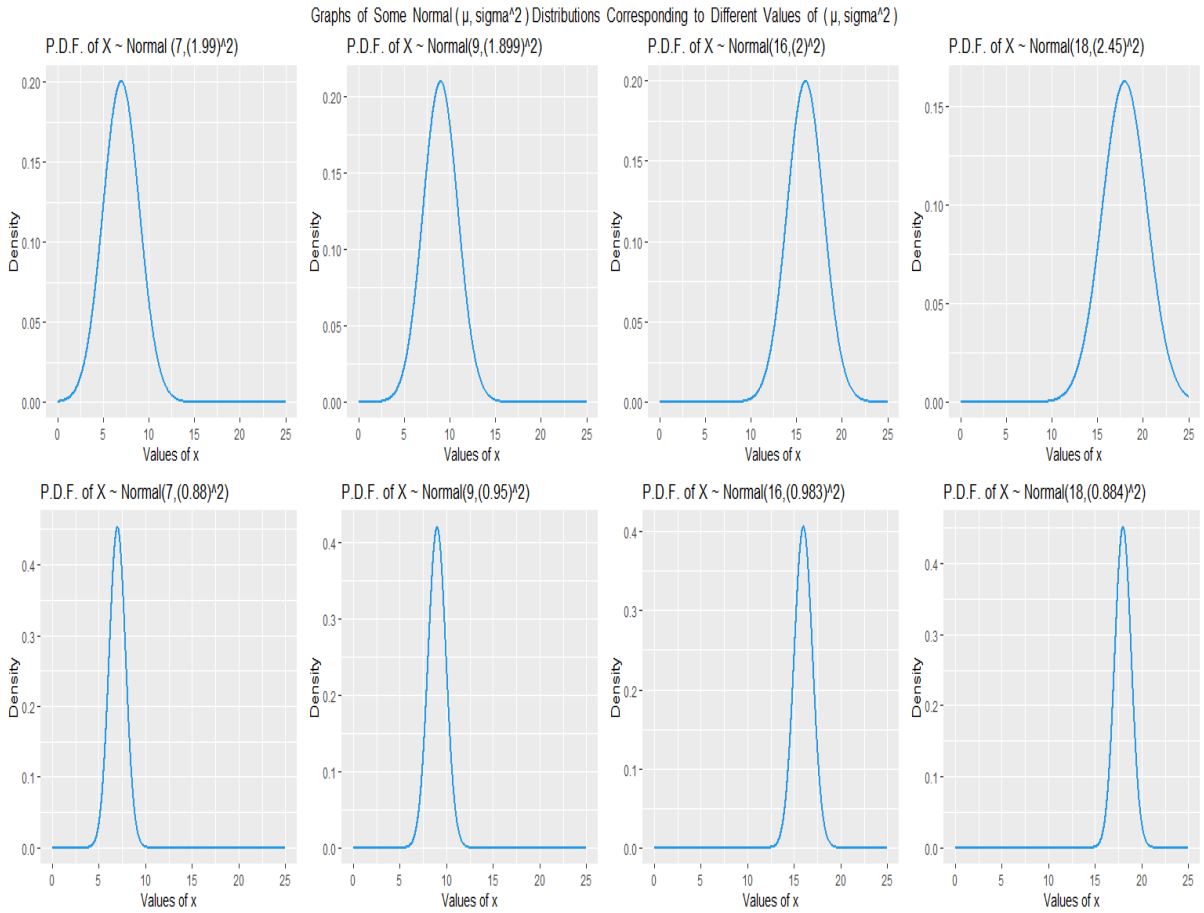
Bayes theorem gives us a method to revise our (belief) distribution about the parameter, given the data. In order to use it, one must have a distribution that represents our belief about the parameter, before we look at the data. This is our prior distribution. In this section we propose some methods to choose the prior, as well as things to consider in prior choice.

Choosing a Conjugate Prior When We Have Vague Prior Knowledge:

The shape of the $Normal(\mu_0, {\sigma_0}^2)$ distribution changes with respect to different choices of the parameters μ_0 and σ_0. One should consider an appropriate conjugate prior $Normal(\mu_0, {\sigma_0}^2)$ that matches with our prior belief. For example, if we have the prior belief that the parameter under study $\mu;\ -\infty < \mu < \infty$ can be considered as a random variable and on an average, it takes values around a particular value, say a, with a moderate concentration. Then we should choose a prior $Normal(\mu_0, {\sigma_0}^2)$ such that the prior distribution assigns high probability density close to that mentioned particular value a with a standard deviation not too small. On the other hand, if we have the prior belief that the parameter under study $\mu;\ -\infty < \mu < \infty$ on an average takes values around a particular value, say b, with a high concentration. Then we should choose a prior $Normal(\mu_0, {\sigma_0}^2)$ such that the prior distribution assigns high probability density close to that mentioned particular value b with a small standard deviation.

The following figure shows the shapes of some $Normal(\mu_0, {\sigma_0}^2)$ densities for values of (μ_0, σ_0): $(7,1.99), (9,1.899), (16,2), (18,2.45), (7,0.88), (9,0.95), (16,0.983), (18,0.884)$. This shows the some of the shapes that members of the $Normal(\mu_0, {\sigma_0}^2)$ family can take.

Figure 8:



In figure 8, it is illustrated that, if X following $Normal(\mu, \sigma^2)$, assumes values around a particular value with moderate concentration, then the graphs of its probability density will look like the graphs presented in the 1st row. Whereas, X assumes values around a particular value with high concentration, then the graphs of its probability density will look like the graphs presented in the 2nd row. So, a prior should be chosen accordingly.

Choosing a Conjugate Prior When We Have Real Prior Knowledge:

Normal distribution can take various shapes depending on the values of its parameters (shown in the ***figure 8***). The prior chosen should correspond to our belief about the parameter of interest μ. Now, by saying we have real prior knowledge, we mean that, from some previous survey or from some other sources, we can strongly belief that the mean of μ is μ_0. This is the value our prior belief is centred on. Then we have to decide on our prior standard deviation σ_0.Now, we think of the points above and below that we consider to be the upper and lower bounds of possible values of μ. Dividing the distance between these two points by 6 we get our prior standard deviation σ_0. This way we can get reasonable probability over all the region we believe possible.

Precautions before using the conjugate prior:

- First, we have to graph our $Normal(\mu_0, {\sigma_0}^2)$ prior. If the shape looks reasonably close to what we believe, we will use it. Otherwise, we can adjust μ_0 and σ_0 until we find a prior whose graph approximately corresponds to our belief. As long as the prior has reasonable probability over the whole range of values that we think the parameter could possibly be in, it will be a satisfactory prior.

- ***Calculating the equivalent sample size of the prior:*** This is the size of a random sample of $Normal(\mu, \sigma^2)$ variables that matches the amount of prior information about μ that we are putting in with our prior. We note that if $(X_1, X_2, \ldots, X_n)$ is a random sample from $Normal(\mu, \sigma^2)$ then $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ will follow Normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. The equivalent sample size will be the solution of,

$$\frac{\sigma^2}{n_{eq}} = {\sigma_0}^2$$

So, the equivalent sample size for $Normal(\mu_0, {\sigma_0}^2)$ prior for μ is $n_{eq} = \frac{\sigma^2}{\sigma_0^2}$. It says that the amount of information about the parameter from our prior is equivalent to the amount from a random sample of that size from $Normal(\mu, \sigma^2)$ distibution. One should always check if this is unrealistically high. If it is so, it shows we have very strong prior belief about μ. It will take a lot of sample data to move our posterior belief far from our prior belief. So, we should increase the prior standard deviation and recalculate the prior. Otherwise, we would be putting too much prior information about the parameter relative to the amount of information that will come from the data. If the equivalent sample size is small, it means our prior belief is not so strong, and our posterior belief will be strongly influenced by a more modest amount of sample data.

Illustrating Example:

Let us consider X to be a random variable following a $Normal(\mu, \sigma^2)$ distribution, where $-\infty < \mu < \infty$ but μ is unknown to us (parameter of interest) and σ is known. To understand the performance of the estimators corresponding to Classical and Bayesian approach, we considered $\mu = 20 \; and \; \sigma = 2$ and then after drawing a random sample from $Normal(20, 2^2)$ distribution, we find the estimates of the estimators obtained corresponding to the above-mentioned approaches and their standard errors for comparison.

We draw a random sample of size $n = 15$ from $Normal(20, 2^2)$ distribution.

The random sample comes out to be (19.33036, 20.24134, 20.79421, 20.12592, 21.06675, 20.59965, 21.35308, 20.79693, 19.00265, 17.15015, 19.97033, 17.45221, 19.74741, 18.47389, 22.07296)

The estimate of the maximum likelihood estimator μ_{MLE} is given by $\widehat{\boldsymbol{\mu_{MLE}}} = \bar{\boldsymbol{x}} = \frac{\sum_{i=1}^{n} \boldsymbol{x_i}}{\boldsymbol{n}}$, which comes out to be 19.87852

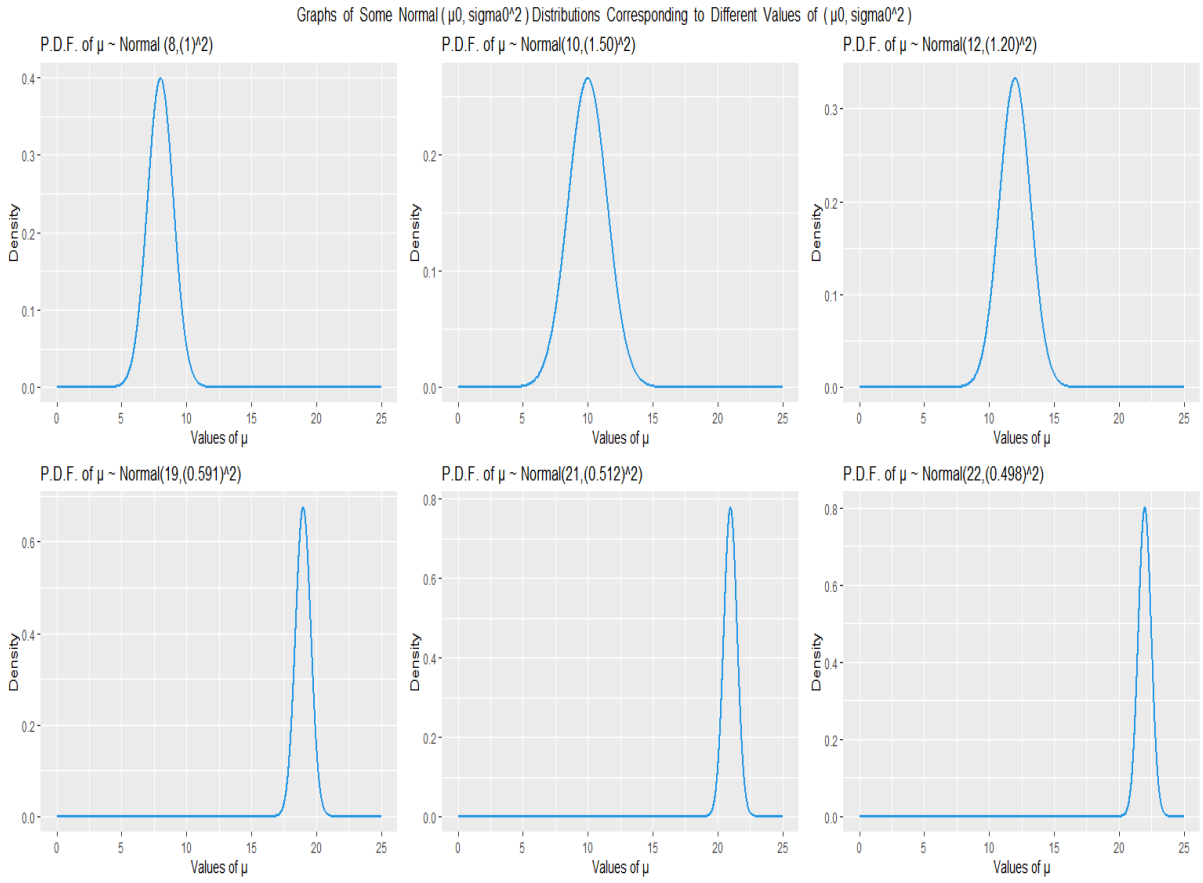
The estimate of the standard error of μ_{MLE} comes out to be $\widehat{SE(\mu_{MLE})} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}} = 0.3598533$

To illustrate the proper choice of conjugate priors we considered six conjugate prior distributions which are given by,

$$Normal(8, 1^2), Normal(10, 1.50^2), Normal(12, 1.20^2), Normal(19, 0.591^2),$$
$$Normal(21, 0.512^2), Normal(22, 0.498^2)$$

The graphs of the different conjugate priors are given below,

Figure 9:



Now, for each conjugate prior distribution we obtained the corresponding posterior distributions and obtained the posterior means which corresponds to the Bayes Estimator for estimating μ.

To obtain the standard errors of the Bayes Estimators, we used a method of resampling, bootstrap. We used $B = 1000$ Bootstrap samples to estimate the standard errors of posterior means corresponding to the different conjugate priors.

Suppose we have an additional information that the random quantity μ takes values around 19 on an average. So, we can consider that on an average μ assumes high densities (around 17 to 22). In that case, we should choose such conjugate priors which takes high probability density towards the values in the interval (17,22). Otherwise, the Bayes estimator will give estimates not closer to the actual value of the parameter or even if it gives close estimates, the standard errors will be comparatively high. This scenario can be observed through the graphs of probability densities corresponding to the different conjugate priors and also from the computations shown in the table below.

The estimate of maximum likelihood estimator, its standard error and the values of the Bayes estimator along with its standard error for different conjugate priors are represented in a tabular form which is given below.

Table 3

Value of the parameter of interest (μ)	**Estimate of the maximum likelihood estimator (μ_{MLE})**	**Standard error of maximum likelihood estimator $\widehat{SE(\mu_{MLE})}$**	**Estimate of the Bayes Estimator (posterior mean) (μ_b)**	**Standard error of the Bayes Estimator $\widehat{SE(\mu_b)}$**	**Conjugate Priors**
20	19.87852	0.3598533	17.37778	0.2800558	$Normal(8, 1^2)$
20	19.87852	0.3598533	18.83179	0.3228861	$Normal(10, 1.50^2)$
20	19.87852	0.3598533	18.64750	0.2910656	$Normal(12, 1.20^2)$
20	19.87852	0.3598533	19.49818	0.2003469	$Normal(19, 0.591^2)$
20	19.87852	0.3598533	20.44406	0.1672664	$Normal(21, 0.512^2)$
20	19.87852	0.3598533	20.97772	0.1708604	$Normal(22, 0.498^2)$

Findings:

- If we incorporate the additional information about μ that it is itself a random variable, then Bayes estimator performs better than Maximum Likelihood estimator in terms of standard errors of the estimators.
- For all the conjugate priors, the standard error of the Bayes estimators is lower than that of the Maximum Likelihood estimators.
- As the actual value of μ is 20, the conjugate priors that have more weight around the value 20, are more appropriate than the others. From the graphs of different conjugate priors, it is seen that $Normal(19, 0.591^2)$, $Normal(21, 0.512^2)$ and $Normal(22, 0.498^2)$ conjugate prior distributions have high density in the region $17 < \mu < 22$. So, by considering these priors we can get better Bayes estimators than the others, which can be seen from the ***Table 3*** also, given above.
- Here, we are interested with the parameter μ, the mean of a $Normal(\mu, \sigma^2)$ distribution with σ known. We know that, sample mean is a good representative of population mean. So, we can assume that the variance of sample mean will be a good representative of variance of population mean μ. Now, under this assumption if we take the prior mean

around 20 and prior standard deviation around 0.5163978, which is the standard deviation of the sample mean with respect to our example, we can observe from the table that we can obtain a better Bayes estimator whose estimates are close enough to the actual value of the parameter of interest and standard errors of the estimators get reduced comparatively.

To choose a conjugate prior when we have real prior knowledge:

Suppose, from some previous survey or from some other sources, we can strongly belief that the mean of μ is $\mu_0 = 18$ and standard deviation σ_1. Now, we need to choose a $Normal(\mu_0, {\sigma_0}^2)$ that matches our prior belief about the (location) mean and (scale) standard deviation. Then to decide on our prior standard deviation σ_0, we think of the points 18 and 21 to be the upper and lower bounds of possible values of μ respectively. Dividing the distance between these two points by 6 to get our prior standard deviation σ_0.

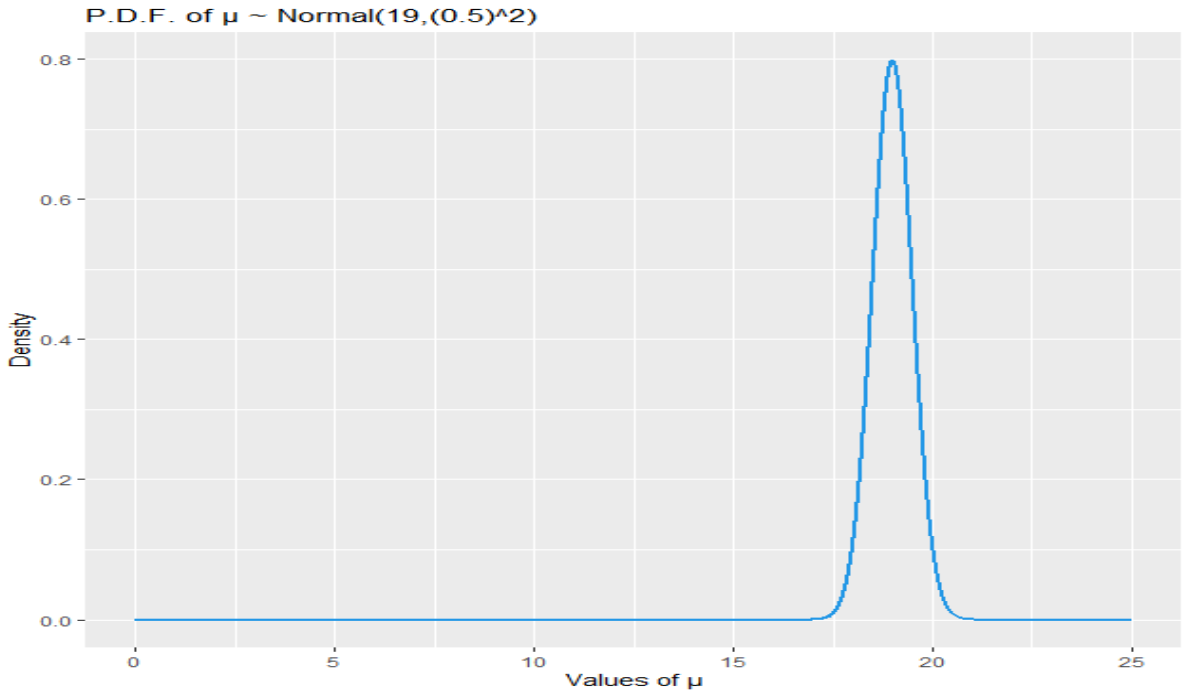
So, we get,

$$\sigma_0 = \frac{21 - 18}{6} = 0.5$$

So, we are getting the appropriate conjugate prior as $Normal(19, 0.5^2)$

The graph of probability density function of the conjugate prior $Normal(\mu_0 = 19, {\sigma_0}^2 = 0.5^2)$ is given below,

Figure 10:



Considering the conjugate prior distribution $Normal(\mu_0 = 19, {\sigma_0}^2 = 0.5^2)$, the estimate of the Bayes Estimator (posterior mean) is given by,

$$\widehat{\mu_b} = 19.42509$$

Also, the estimate of the standard error of μ_b is given by,

$$\widehat{SE(\mu_b)} = 0.1674867$$

<u>Calculating the equivalent sample size</u>:

As mentioned in the theory, the expression of the equivalent sample size is,

$$n_{eq} = \frac{\sigma^2}{\sigma_0^2}$$

From the above relation, the equivalent sample size is obtained as,

$$n_{eq} = 16$$

It says that the amount of information about the parameter μ from our prior is equivalent to the amount from a random sample of size approximately equal to 16. As, the value of equivalent sample size is not unrealistically high, we can say that our prior belief seems to be appropriate.

When the population distribution follows $N(\mu, \sigma^2); -\infty < \mu < \infty;\ \sigma > 0$ *where μ is known but variance σ^2 is unknown*:

Here we are considering the characteristic of the elements of the population, on which we have the interest, can be represented by a random variable X whose p.d.f $f_X(.;\mu,\sigma^2)$, is given by,

$$f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, where\ -\infty < x < \infty; -\infty < \mu < \infty;\ \sigma > 0$$

$$0 \quad ; \quad otherwise$$

Here, σ^2 is an unknown quantity, the parameter of interest, but μ is known.

Now, let us draw a random sample $(X_1, X_2, \ldots, X_n)$ of size n from the population. Let us consider, $(x_1, x_2, \ldots, x_n)$ to be the realisation of $(X_1, X_2, \ldots, X_n)$.

Finding an Estimator of σ^2 by Classical Approach:

In classical approach we consider σ^2 to be an unknown but fixed quantity. Here we will use the method of maximum likelihood to find an estimator of σ^2 in support of classical approach.

Likelihood Function:

The likelihood function of σ^2 is given by,

$$L(\sigma^2) = L(\mu; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \sigma^2)$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}, where\ -\infty < x < \infty; -\infty < \mu < \infty;\ \sigma > 0$$

The maximum likelihood estimate of σ^2 is obtained by maximizing the likelihood function with respect to σ^2, which is obtained as,

$$\widehat{\sigma^2_{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

As we know, if $X_1, X_2, \ldots, X_n$ are independently or identically distributed as $N(\mu, \sigma^2)$ distribution, then,

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Thus, the standard deviation of the maximum likelihood estimator ${\sigma^2}_{MLE}$ is given by,

$$SE({\sigma^2}_{MLE}) = \frac{\sqrt{2(n-1)}\sigma^2}{n}$$

To find an unbiased estimator of σ^2 :

We have,

$$E\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}\right) = (n-1)$$

$$\Rightarrow E\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\right) = \sigma^2$$

So, the estimate of the unbiased estimator of σ^2 is given by, $\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Thus, the estimate of the standard error of the maximum likelihood estimator of σ^2 is given by,

$$SE(\widehat{{\sigma^2}_{MLE}}) = \frac{\sqrt{2(n-1)}}{n}\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sqrt{\frac{2}{n-1}}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

Finding an Estimator of σ^2 by Bayesian Approach:

Unlike classical approach, here in Bayesian approach we consider the parameter σ^2 to be an unknown but a random quantity. Therefore, σ^2 itself is a random variable and it will have some probability distribution. As, σ^2 can take any values between 0 to ∞, we can consider σ^2 has a $Inverse\ Gamma(m, \theta)$ distribution(prior). Then the p.d.f of σ^2, $g_{\sigma^2}(.)$, is given by,

$$g_{\sigma^2}(\sigma^2) = \frac{\theta^m e^{-\frac{\theta}{\sigma^2}}(\sigma^2)^{-1-m}}{\Gamma(m)}\ ; \quad \sigma^2 > 0; \theta, m > 0$$

$$0 \qquad\qquad\qquad ,otherwise$$

Here, m and θ both are considered to be known.

Now, the posterior distribution of σ^2 is given by,

$$f_{\sigma^2}(\sigma^2 \mid x_1, x_2, \dots., x_n) = \frac{f_{(X_1,X_2,\dots.,X_n \mid \sigma^2=\sigma^2)}(x_1, x_2, \dots., x_n \mid \sigma^2)\; g_{\sigma^2}(\sigma^2)}{f_{(X_1,X_2,\dots.,X_n)}(x_1, x_2, \dots., x_n)}$$

$$= \frac{\prod_{i=1}^{n} f(x_i \mid \sigma^2)\; g_{\sigma^2}(\sigma^2)}{\int_0^{\infty} \prod_{i=1}^{n} f(x_i \mid \sigma^2)\; g_{\sigma^2}(\sigma^2)\, d\sigma^2} \quad \dots\dots\dots.(i)$$

Now,

$$\int_0^{\infty} \prod_{i=1}^{n} f(x_i \mid \sigma^2)\; g_{\sigma^2}(\sigma^2)\, d\sigma^2 = \int_0^{\infty} \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}} \frac{\theta^m e^{-\frac{\theta}{\sigma^2}} (\sigma^2)^{-1-m}}{\Gamma(m)} d\sigma^2$$

$$= \frac{\theta^m}{\left(\sqrt{2\pi}\right)^n \Gamma(m)} \int_0^{\infty} (\sigma^2)^{-1-m-\frac{n}{2}}\, e^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)} d\sigma^2$$

$$= \frac{\theta^m}{\left(\sqrt{2\pi}\right)^n \Gamma(m)} \int_{\infty}^{0} y^{1+m+\frac{n}{2}} e^{-y\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)} \left(-\frac{dy}{y^2}\right); \text{ by making the transformation } y = \frac{1}{\sigma^2}$$

$$= \frac{\theta^m}{\left(\sqrt{2\pi}\right)^n \Gamma(m)} \int_0^{\infty} y^{\left(m+\frac{n}{2}\right)-1} e^{-y\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)} dy$$

$$= \frac{\theta^m}{\left(\sqrt{2\pi}\right)^n \Gamma(m)} \frac{\Gamma\left(m+\frac{n}{2}\right)}{\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)^{\left(m+\frac{n}{2}\right)}}$$

Now, from (i) we get,

$$f_{\sigma^2}(\sigma^2 \mid x_1, x_2, \dots., x_n) = \frac{\frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}} \frac{\theta^m e^{-\frac{\theta}{\sigma^2}} (\sigma^2)^{-1-m}}{\Gamma(m)}}{\frac{\theta^m}{\left(\sqrt{2\pi}\right)^n \Gamma(m)} \frac{\Gamma\left(m+\frac{n}{2}\right)}{\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)^{\left(m+\frac{n}{2}\right)}}}$$

$$= \frac{\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)^{-m-\frac{n}{2}} e^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta\right)} (\sigma^2)^{-1-\left(m+\frac{n}{2}\right)}}{\Gamma\left(m+\frac{n}{2}\right)}$$

Thus,

the posterior distribution of σ^2 follows $\boldsymbol{Inverse\ gamma}\left(\boldsymbol{m+\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta}\right)$

As, the posterior distribution and prior distribution of σ^2 belongs to the same family $g_{\sigma^2}(.)$, the prior is called the conjugate prior family of σ^2.

Now, the estimate of the Bayes Estimator of σ^2 is given by the posterior mean, which is given by,

$$\widehat{\sigma^2{}_b} = \frac{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\theta}{m+\frac{n}{2}-1} \; ; \; m+\frac{n}{2}>1$$

To Obtain the Estimate of the Standard Error of the Bayes Estimator $\sigma^2{}_b$:

We obtain the standard error of $\sigma^2{}_b$ by Bootstrap method, which is a method of re-sampling. Here we consider the drawn random sample $(X_1, X_2, \ldots, X_n)$ of size n as a population and we draw B number of samples of size n with replacement from the random sample $(X_1, X_2, \ldots, X_n)$. For each such B bootstrap samples we get an estimate of $\sigma^2{}_b$. The standard deviation of the B values of $\sigma^2{}_b$ can be considered as the standard error of the Bayes estimator $\sigma^2{}_b$ for estimating σ^2.

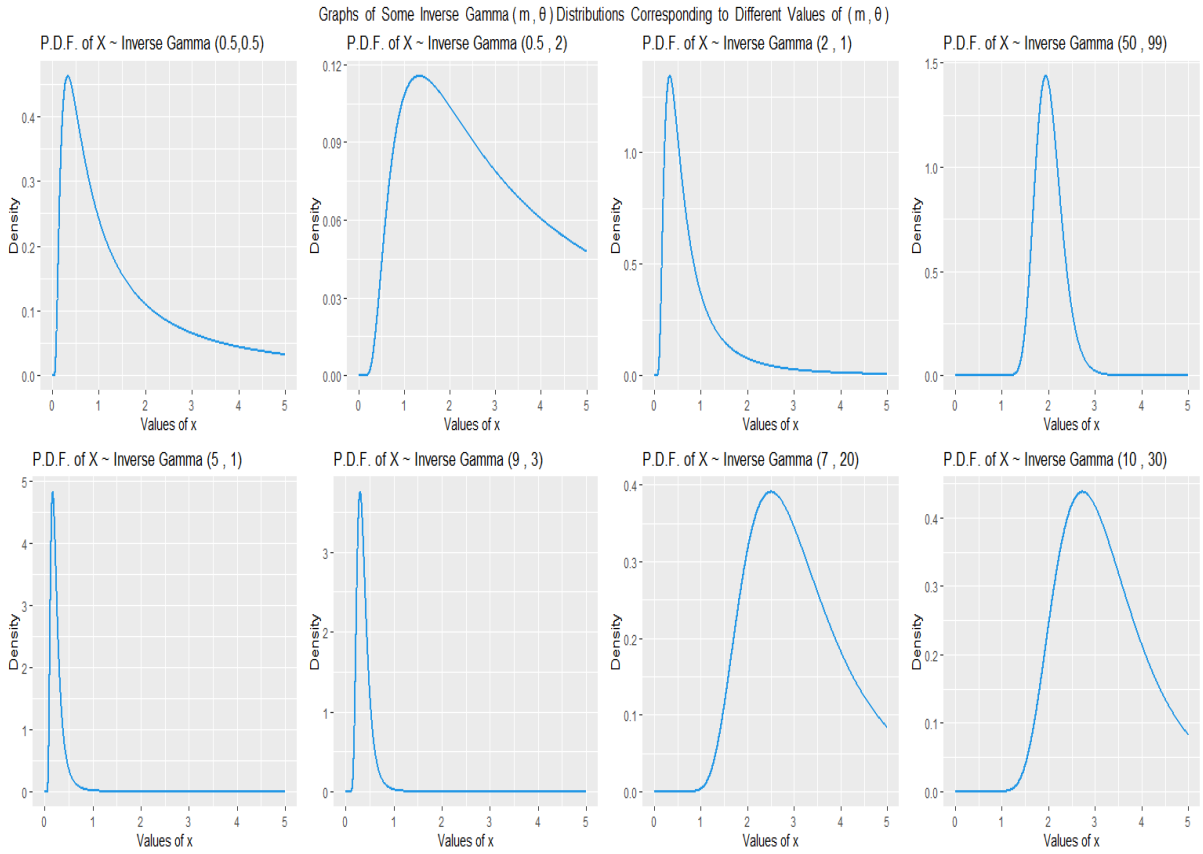
Choice of Conjugate Prior Distributions:

Bayes theorem gives us a method to revise our (belief) distribution about the parameter, given the data. In order to use it, one must have a distribution that represents our belief about the parameter, before we look at the data. This is our prior distribution. In this section we propose some methods to choose the prior, as well as things to consider in prior choice.

The shape of the $Inverse\ Gamma(m, \theta)$ distribution changes with respect to different choices of the parameters m and θ. One should consider an appropriate conjugate prior $Inverse\ Gamma(m, \theta)$ that matches with our prior belief about σ^2. For example, if we have the prior belief that the parameter under study σ^2; $\sigma^2 > 0$ can be considered as a random variable and on an average, it takes the lower value, then we should consider such a prior distribution $Inverse\ Gamma(m, \theta)$ which assigns high density towards the lower values of σ^2. Similarly, if we have the prior belief that the parameter under study σ^2; $\sigma^2 > 0$ can be considered as a random variable and on an average, it takes the higher value, then we should consider such a conjugate prior distribution $Inverse\ Gamma(m, \theta)$ which assigns high density towards the higher values of σ^2.

The following figure shows the shapes of $Inverse\ Gamma(m,\theta)$ densities for values of (m,θ): $(0.5,0.5),(0.5,2),(2,1),(50,99),(5,0.5),(9,1),(7,20),(10,30)$. This shows some of the shapes that the members of the $Inverse\ Gamma(m,\theta)$ family can take.

Figure 11:



Illustrating Example:

Let us consider X to be a random variable following a $Normal(\mu,\sigma^2)$ distribution, where $-\infty<\mu<\infty$ but μ is known to us but σ^2 is unknown (parameter of interest). To understand the performance of the estimators corresponding to Classical and Bayesian approach, we considered $\mu=20\ and\ \sigma^2=2$ and then after drawing a random sample from $Normal(20,(\sqrt{2})^2)$ distribution, we find the estimates of the estimators obtained corresponding to the above-mentioned approaches and their standard errors for comparison.

We draw a random sample of size $n=15$ from $Normal\big(20,(\sqrt{2})^2\big)$ distribution.

The random sample comes out to be (19.52649, 20.17065, 20.56159, 20.08904, 20.75431, 20.42401, 20.95677, 20.56351, 19.29477, 17.98485, 19.97902, 18.19844, 19.82139, 18.92088, 21.46581)

The estimate of the maximum likelihood estimator of σ^2 is given by $\widehat{\sigma^2_{MLE}}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ which comes out to be 0.9064608

The estimate of the standard error of ${\sigma^2}_{MLE}$ comes out to be,

$$SE(\widehat{{\sigma^2}_{MLE}}) = \sqrt{\frac{2}{n-1}}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \ 0.34261$$

To illustrate the proper choice of conjugate priors we have considered eight conjugate prior distributions which are given by,

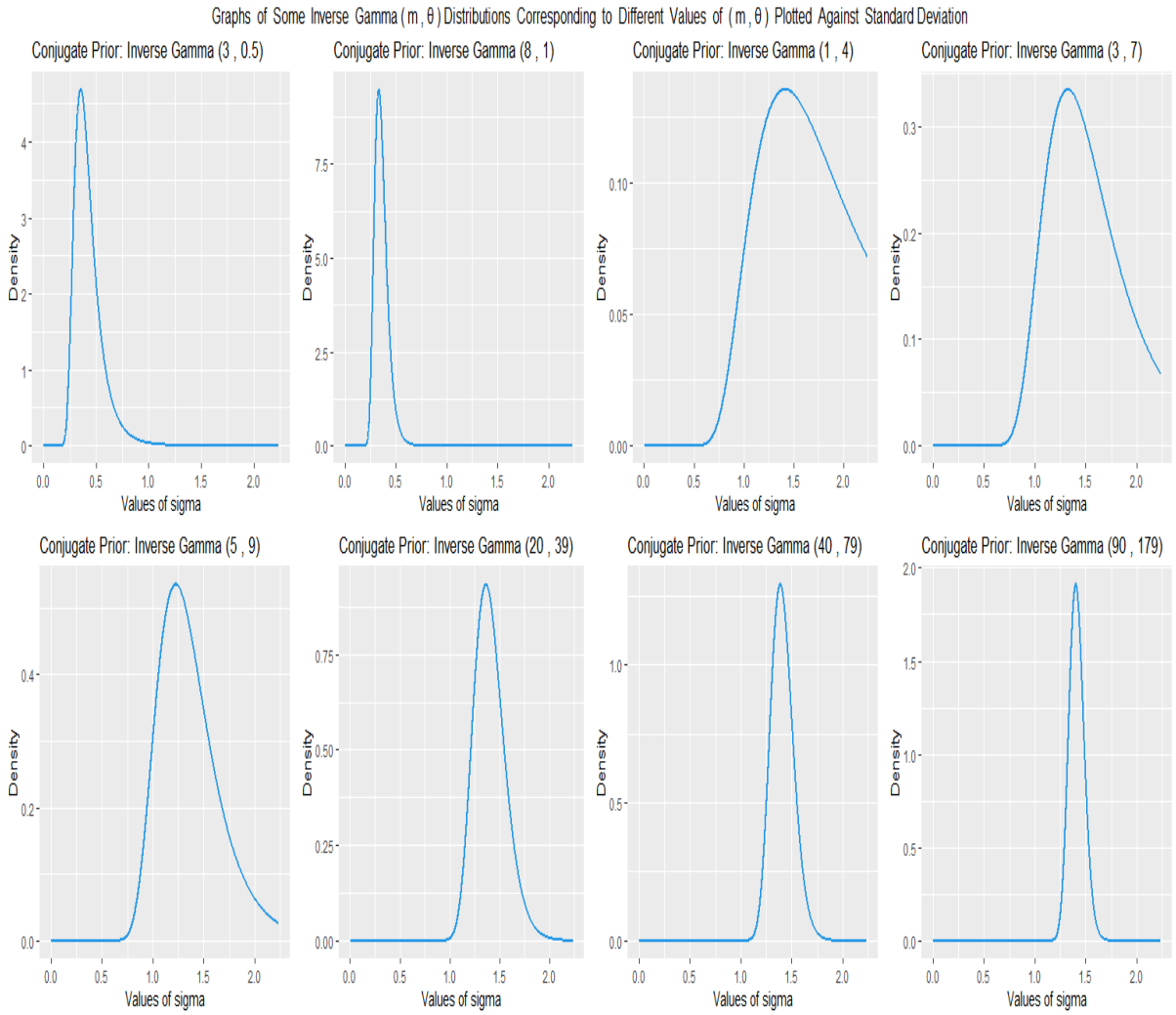
$$Inverse\ Gamma(3,0.5), Inverse\ Gamma(8,1), Inverse\ Gamma(1,4), Inverse\ Gamma(3,7)$$

$$Inverse\ Gamma(5,9), Inverse\ Gamma(20,39), Inverse\ Gamma(40,79), Inverse\ Gamma(90,179)$$

As the variance σ^2 is in squared units, it is hard to visualize our belief about it. So, for graphical presentation, we will plot the densities of the corresponding conjugate priors against the values of standard deviation σ.

The graphs of the different conjugate priors are given below,

Figure 12:



Now, for each conjugate prior distribution we obtained the corresponding posterior distributions and the posterior means which corresponds to the Bayes Estimator for estimating σ^2 .

To obtain the standard errors of the Bayes Estimators, we used the method of resampling, bootstrap. We used $B = 1000$ Bootstrap samples to estimate the standard errors of posterior means corresponding to the different priors.

Suppose we have an additional information that the random quantity σ^2 takes values around 1.4 on an average. So, we can consider that on an average μ assumes high densities (around 1 to 2). In that case, we should choose such conjugate priors which takes high probability density corresponding to the values in the interval (1,2). Otherwise, the Bayes estimator will give estimates not closer to the actual value of the parameter or even if it gives close estimates, the standard errors will be comparatively high. This scenario can be observed through the graphs of probability densities corresponding to the different conjugate priors and also from the computations shown in the table below.

The estimate of maximum likelihood estimator, its standard error and the values of the Bayes estimator along with its standard error for different conjugate priors are represented in a tabular form which is given below.

Table 4

Value of the parameter of interest (σ^2)	**Value of the maximum likelihood estimate ($\widehat{\sigma^2_{MLE}}$)**	**Standard error of maximum likelihood estimator $SE(\widehat{\sigma^2_{MLE}})$**	**Value of the Bayes Estimator (posterior mean) ($\widehat{\sigma^2_b}$)**	**Standard error of the Bayes Estimator $\widehat{SE(\sigma^2_b)}$**	**Conjugate Priors**
2	0.9064608	0.34261	0.7740836	0.24725838	$Inverse\ Gamma(3,0.5)$
2	0.9064608	0.34261	0.5416410	0.16433659	$Inverse\ Gamma(8,1)$
2	0.9064608	0.34261	1.4471726	0.31751664	$Inverse\ Gamma(1,4)$
2	0.9064608	0.34261	1.4582942	0.24594242	$Inverse\ Gamma(3,7)$
2	0.9064608	0.34261	1.3785908	0.20140338	$Inverse\ Gamma(5,9)$
2	0.9064608	0.34261	1.7303319	0.09016895	$Inverse\ Gamma(20,39)$
2	0.9064608	0.34261	1.8463182	0.05152346	$Inverse\ Gamma(40,79)$
2	0.9064608	0.34261	1.9259461	0.02361808	$Inverse\ Gamma(90,179)$

Findings:

- If we incorporate the additional information about σ^2 that it is itself a random variable, then Bayes estimator performs better than Maximum Likelihood estimator in terms of standard errors of the estimators and the Bayes estimates are close to the actual value of the parameter $\sigma^2 = 2$ than that of Maximum likelihood estimates.
- For all the conjugate priors, the standard error of the Bayes estimators is lower than that of the Maximum Likelihood estimators.
- As the actual value of σ^2 is 2 i.e., $\sigma \simeq 1.414$, the conjugate priors that have more weight around the value 1.414 in the ***figure 12***, are more appropriate than the others.

From the graphs of different conjugate priors, it is seen that $Inverse\ Gamma(5,9)$, $Inverse\ Gamma(20,39)$, $Inverse\ Gamma(40,79)$ and the conjugate prior distribution $Inverse\ Gamma(90,179)$ have high density in the region $1 < \sigma < 2$. So, by considering these priors we can get better Bayes estimates than the others, which can be seen from the ***Table 3*** also, given above.

- Those Bayes estimators corresponding to the conjugate priors (2^{nd} row's 2^{nd}, 3^{rd}, 4^{th} graphs; ***figure 12***), which ensures high concentration of the values of standard deviation σ around the value 1.414 gives better Bayes estimates with low standard deviations than the Bayes estimators corresponding to the conjugate priors (1^{nd} row's 3^{rd}, 4^{th} graphs and 2^{nd} row's 1^{st} graph; ***figure 12***), which ensures moderate concentration of the values of standard deviation σ around the value 1.414.

References:

1. Introduction To The Theory Of Statistics – Alexander M. Mood, Franklin A. Graybill, Duane C. Boes

2. Introduction To Bayesian Statistics – William M. Bolstad, James M. Curran

3. An Introduction to Statistical Learning – Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

4. https://en.wikipedia.org/wiki/Conjugate_prior

Appendix for R codes:

Here, we have used R software (version: 4.0.5)

1. R code For Binomial Distribution:

```
rm(list=ls())
set.seed(seed=987654321)

p=rep(0.768,8);p   #probability of success

m=10    # number of trials
n=10    # sample size
B=1000  #bootstrap number

pb.hat.mean=array(0);pb.hat.median=array(0);
pb.hat.mode=array(0);post.1=array(0);post.2=array(0);
sd.err.pb.boot=array(0);pb.hat.boot=array(0);

a = c(1,0.5,4,3,5,8,13,13); b = c(1,5,16,9,2,4,6,3)
# "a" : 1st parameter of prior distribution of p
# "b" : 2nd parameter of prior distribution of p

#Drawing the random sample from Bin(m,p)
obs = rbinom(n,m,p[1])

#----------------------Classical Approach
mle=rep(mean(obs)/m,8)
sderr.mle=rep(sqrt(mle[1]*(1-mle[1])/((m*n)-1)),8)

#----------------------Bayesian Approach

for(i in 1:length(p)){


post.1[i]=sum(obs)+a[i]          #1st parameter of posterior distribution
```

```
post.2[i]=(m*n)+b[i]-sum(obs)     #2nd parameter of posterior distribution

pb.hat.mean[i] = (post.1[i]/(post.1[i]+post.2[i]))


for(j in 1:B){

 boot.obs=sample(obs,n,replace=T)
 pb.hat.boot[j]=(sum(boot.obs)+a[i])/(a[i]+(m*n)+b[i])
}
sd.err.pb.boot[i]=sd(pb.hat.boot)

}

vauge=data.frame(p,mle,sderr.mle,pb.hat.mean,sd.err.pb.boot,a,b)
vauge
View(vauge)
#sderr.mle = Standard error of MLE
#pb.hat.mean = Posterior Mean
#sd.err.pb.boot = Standard Error of pb.hat.mean obtained by Bootstrap method
# a = 1st parameter of prior distribution of p
# b = 2nd parameter of prior distribution of p
# 1000 Bootstrap sample has been considered here


#when we have real prior knowledge(B(a0,b0))
p0=0.71
sig0=0.0569

#solving p0=a0/(a0+b0) and sig0=sqrt((a0*b0)/(a0+b0)^2*(a0+b0+1))

a0=p0*((p0*(1-p0)/(sig0^2))-1);a0
b0=((1/p0)-1)*a0;b0

pb.hat.mean.real=array(0);pb.hat.boot.real=array(0);

pb.hat.mean.real=(sum(obs)+a0)/((m*n)+b0+a0);pb.hat.mean.real

for(j in 1:B){

 boot.obs.real=sample(obs,n,replace=T)
 pb.hat.boot.real[j]=(sum(boot.obs.real)+a0)/(a0+(m*n)+b0)
}
sd.err.pb.boot.real=sd(pb.hat.boot.real);sd.err.pb.boot.real


#Calculating equivalent sample size

n.eq=(a0+b0+1)/m;n.eq

#It says that the amount of information about the parameter from
```

```
#our prior is equivalent to the amount from a random sample of that size.
#If the equivalent sample size is unrealistically high then it means that
#we are imposing too much prior information about relative to the amount
#of information that will come from the data.

data.frame(p0,sig0,a0,b0,pb.hat.mean.real,sd.err.pb.boot.real,n.eq)
```

2. ***R code for figure 1:***

```
rm(list=ls())
x=seq(0,1,0.01)
library(ggplot2)
library(patchwork)

y1=dbeta(x,0.5,0.5)
g1=ggplot(NULL,aes(x,y1))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(0.5,0.5)")

y2=dbeta(x,0.5,1)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(0.5,1)")

y3=dbeta(x,0.5,2)
g3=ggplot(NULL,aes(x,y3))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(0.5,2)")

y4=dbeta(x,0.5,3)
g4=ggplot(NULL,aes(x,y4))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(0.5,3)")

y5=dbeta(x,1,0.5)
g5=ggplot(NULL,aes(x,y5))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(1,0.5)")

y6=dbeta(x,1,1)
g6=ggplot(NULL,aes(x,y6))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(1,1)")

y7=dbeta(x,1,2)
g7=ggplot(NULL,aes(x,y7))+
```

```
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(1,2)")

y8=dbeta(x,1,3)
g8=ggplot(NULL,aes(x,y8))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(1,3)")

y9=dbeta(x,2,0.5)
g9=ggplot(NULL,aes(x,y9))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(2,0.5)")

y10=dbeta(x,2,1)
g10=ggplot(NULL,aes(x,y10))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(2,1)")

y11=dbeta(x,2,2)
g11=ggplot(NULL,aes(x,y11))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(2,2)")

y12=dbeta(x,2,3)
g12=ggplot(NULL,aes(x,y12))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(2,3)")

y13=dbeta(x,3,0.5)
g13=ggplot(NULL,aes(x,y13))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(3,0.5)")

y14=dbeta(x,3,1)
g14=ggplot(NULL,aes(x,y13))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(3,1)")

y15=dbeta(x,3,2)
g15=ggplot(NULL,aes(x,y15))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X~B(3,2)")
```

```
y16=dbeta(x,3,3)
g16=ggplot(NULL,aes(x,y16))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
            title="P.D.F. of X~B(3,3)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,g7,g8,g9,g10,g11,g12,g13,g14,g15,g16,nrow=4,top = paste("Graphs of Some Beta ( a , b ) Distributions of 1st Kind Corresponding to Different Values of ( a , b )"))
```

3. R code for figure 2:

```
rm(list=ls())
x=seq(0,1,0.01)
library(ggplot2)
library(patchwork)

y1=dbeta(x,1,1)
g1=ggplot(NULL,aes(x,y1))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
            title="P.D.F. of p~B(1,1)")

y2=dbeta(x,2,5)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
            title="P.D.F. of p~B(2,5)")

y3=dbeta(x,4,16)
g3=ggplot(NULL,aes(x,y3))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
            title="P.D.F. of p~B(4,16)")

y4=dbeta(x,3,9)
g4=ggplot(NULL,aes(x,y4))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
            title="P.D.F. of p~B(3,9)")

y5=dbeta(x,5,2)
g5=ggplot(NULL,aes(x,y5))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
            title="P.D.F. of p~B(5,2)")

y6=dbeta(x,8,4)
g6=ggplot(NULL,aes(x,y6))+
    geom_line(lwd=1,col=4)+
```

```
    labs(x="Values of p",y="Density",
         title="P.D.F. of p~B(8,4)")

y7=dbeta(x,13,6)
g7=ggplot(NULL,aes(x,y7))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
         title="P.D.F. of p~B(13,6)")

y8=dbeta(x,13,3)
g8=ggplot(NULL,aes(x,y8))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
         title="P.D.F. of p~B(13,3)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,g7,g8,nrow=2,top = paste("Graphs of Different
Conjugate Prior Distributions"))
```

4. ***R code for figure 3:***

```
rm(list=ls())
x=seq(0,1,0.001)
library(ggplot2)
y8=dbeta(x,44.44337,18.15293)
g8=ggplot(NULL,aes(x,y8))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of p",y="Density",
         title="P.D.F. of  p ~ B ( a= 44.44337, b = 18.15293)  of  1st  kind")
g8
```

5. ***R code For Poisson Distribution:***

```
rm(list=ls())

lamda=rep(2,6);lamda   #Poisson  parameter

n=15  # sample size
B=1000  #Bootstrap number

m=c(1,5,15,7,13,14)
theta=c(9,0.5,2,2,8,7);theta

mle=array(0);sderr.mle=array(0);lamda.hat.mean=array(0);
lamda.hat.boot=array(0);
sd.err.lamda.boot=array(0);


for(i in 1:length(lamda)){

set.seed(seed=987654321)

obs=rpois(n,lamda[i])
```

```
mle[i]=mean(obs)
sderr.mle[i]=(sqrt(sum(obs)))/n

lamda.hat.mean[i]=(sum(obs)+m[i])/(n+theta[i])

for(j in 1:B){

 boot.obs=sample(obs,n,replace=T)
 lamda.hat.boot[j]=(sum(boot.obs)+m[i])/(n+theta[i])
}
sd.err.lamda.boot[i]=sd(lamda.hat.boot)

}

data.frame(lamda,mle,sderr.mle,lamda.hat.mean,sd.err.lamda.boot,m,theta)


#when we have real prior knowledge(Gamma(m0,theta0))
lamda0=1.75
sig0=0.0891

#solving lamda0=m0/theta0  and  sig0^2=m0/(theta0^2)

theta0=lamda0/(sig0^2);theta0
m0=theta0*lamda0;m0

lamda.hat.mean.real=array(0);lamda.hat.boot.real=array(0);

lamda.hat.mean.real=(sum(obs)+m0)/(n+theta0);lamda.hat.mean.real

for(j in 1:B){

 boot.obs.real=sample(obs,n,replace=T)
 lamda.hat.boot.real[j]=(sum(boot.obs.real)+m0)/(n+theta0)
}
sd.err.lamda.boot.real=sd(lamda.hat.boot.real);sd.err.lamda.boot.real


#Calculating equivalent sample size

n.eq=theta0;n.eq

#It says that the amount of information about the parameter from
#our prior is equivalent to the amount from a random sample of that size.
#If the equivalent sample size is unrealistically high then it means that
#we are imposing too much prior information about relative to the amount
#of information that will come from the data.

data.frame(lamda0,sig0,m0,theta0,lamda.hat.mean.real,sd.err.lamda.boot.real,n.eq)

#Here the equivalent sample size is unrealistically high, so we should
```

```
#increase the standard deviation


#----------------------increasing the standard deviation of lamda w.r.t prior belief
lamda0=1.75
sig01=0.1991

#solving lamda0=m0/theta0  and  sig01^2=m0/(theta0^2)

theta01=lamda0/(sig01^2);theta01
m01=theta01*lamda0;m01

lamda.hat.mean.real1=array(0);lamda.hat.boot.real1=array(0);

lamda.hat.mean.real1=(sum(obs)+m01)/(n+theta01);lamda.hat.mean.real1

for(j in 1:B){

 boot.obs.real=sample(obs,n,replace=T)
 lamda.hat.boot.real[j]=(sum(boot.obs.real)+m01)/(n+theta01)
}
sd.err.lamda.boot.real1=sd(lamda.hat.boot.real);sd.err.lamda.boot.real1



#Calculating equivalent sample size

n.eq1=theta01;n.eq1
data.frame(lamda0,sig01,m01,theta01,lamda.hat.mean.real1,sd.err.lamda.boot.real1,n.eq1
)


#----------------------------------------(for Uniform Prior)
rm(list=ls())
set.seed(seed=987654321)

lamda=2;lamda   #Poisson  parameter

n=15  # sample size
B=1000

obs=rpois(n,lamda);obs

lam.b=(sum(obs)+1)/n
lamda.hat.boot=array(0)
for(j in 1:B){

 boot.obs=sample(obs,n,replace=T)
 lamda.hat.boot[j]=(sum(boot.obs)+1)/n
}
sd.err.lamda.boot=sd(lamda.hat.boot)
```

6. R code for figure 4:

```
rm(list=ls())
x=seq(0,8,0.01)
library(ggplot2)
library(patchwork)

y1=dgamma(x,0.5,0.5)
g1=ggplot(NULL,aes(x,y1))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma (0.5,0.5)")

y2=dgamma(x,0.5,2)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(0.5,2)")

y3=dgamma(x,2,1)
g3=ggplot(NULL,aes(x,y3))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(2,1)")

y4=dgamma(x,3,2)
g4=ggplot(NULL,aes(x,y4))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(3,2)")

y5=dgamma(x,5,0.5)
g5=ggplot(NULL,aes(x,y5))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(5,0.5)")

y6=dgamma(x,9,1)
g6=ggplot(NULL,aes(x,y6))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(9,1)")

y7=dgamma(x,15,2)
g7=ggplot(NULL,aes(x,y7))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(15,2)")

y8=dgamma(x,20,3)
g8=ggplot(NULL,aes(x,y8))+
```

```
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
         title="P.D.F. of X ~ Gamma(20,3)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,g7,g8,nrow=2,top = paste("Graphs of Some Gamma ( m , Θ ) Distributions Corresponding to Different Values of ( m , Θ )"))
```

7. ***R code for figure 5:***

```
rm(list=ls())
x=seq(0,8,0.01)
library(ggplot2)
library(patchwork)

y1=dgamma(x,1,9)
g1=ggplot(NULL,aes(x,y1))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
         title="P.D.F. of Θ ~ Gamma (1,9)")

y2=dgamma(x,5,0.5)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
         title="P.D.F. of Θ ~ Gamma(3,6)")

y3=dgamma(x,15,2)
g3=ggplot(NULL,aes(x,y3))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
         title="P.D.F. of Θ ~ Gamma(9,15)")

y4=dgamma(x,7,2)
g4=ggplot(NULL,aes(x,y4))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
         title="P.D.F. of Θ ~ Gamma(7,2)")

y5=dgamma(x,13,8)
g5=ggplot(NULL,aes(x,y5))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
         title="P.D.F. of Θ ~ Gamma(13,8)")

y6=dgamma(x,14,7)
g6=ggplot(NULL,aes(x,y6))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
         title="P.D.F. of Θ ~ Gamma(14,7)")

library(gridExtra)
```

```
grid.arrange(g1,g2,g3,g4,g5,g6,nrow=2,top = paste("Graphs of Some Gamma ( m , Θ ) Distributions Corresponding to Different Values of ( m , Θ )"))
```

8. R code for figure 6:

```
rm(list=ls())
x=seq(0,8,0.01)
library(ggplot2)
y8=dgamma(x,385.7631,220.4361)
g8=ggplot(NULL,aes(x,y8))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
            title="P.D.F. of Θ ~ Gamma ( 385.7631 , 220.4361 )")
g8
```

9. R code for figure 7:

```
rm(list=ls())
x=seq(0,8,0.01)
library(ggplot2)
y9=dgamma(x,77.25624,44.14642)
g9=ggplot(NULL,aes(x,y9))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of Θ ",y="Density",
            title="P.D.F. of Θ ~ Gamma ( 77.25624 , 44.14642 )")
g9
```

10. R code for Normal Distribution with unknown mean but known variance:

```
rm(list=ls())
set.seed(seed=987654321)

mu=rep(20,6)   #Normal  parameter
sig=2                        #Considering----->>N(mu=20, sig=2)
n=15  # sample size
B=1000  #Bootstrap number

mu0=c(8,10,12,19,21,22)
sig0=c(1,1.5,1.2,0.591,0.512,0.498)

mle=array(0);sderr.mle=array(0);mu.hat.mean=array(0);mu.hat.boot=array(0);
sd.err.mu.boot=array(0);

#Drawing the random sample from Normal
obs = rnorm(n,mu[1],2)

#---------------------- Classical Approach
mle=rep(mean(obs),6)
sderr.mle=rep(sqrt((var(obs))/n),6)

#----------------------Bayesian Approach

for(i in 1:length(mu)){
```

```
mu.hat.mean[i] = ((sum(obs)/sig^2) + (mu0[i]/sig0[i]^2))/((n/sig^2) + (1/sig0[i]^2))

for(j in 1:B){

 boot.obs=sample(obs,n,replace=T)
 mu.hat.boot[j]= ((sum(boot.obs)/sig^2) + (mu0[i]/sig0[i]^2))/((n/sig^2) + (1/sig0[i]^2))
}
sd.err.mu.boot[i]=sd(mu.hat.boot)

}

data.frame(mu,mle,sderr.mle,mu.hat.mean,sd.err.mu.boot,mu0,sig0)

#when we have real prior knowledge----------
mu.real=19
sig.real=0.5

mu.hat.mean.real=array(0);mu.hat.boot.real=array(0);

mu.hat.mean.real=((sum(obs)/sig^2) + (mu.real/sig.real^2))/((n/sig^2) +
(1/sig.real^2));mu.hat.mean.real

for(j in 1:B){

 boot.obs.real=sample(obs,n,replace=T)
 mu.hat.boot.real[j]=((sum(boot.obs.real)/sig^2) + (mu.real/sig.real^2))/((n/sig^2) +
(1/sig.real^2))
}
sd.err.mu.boot.real=sd(mu.hat.boot.real);sd.err.mu.boot.real

neq=(sig/sig.real)^2;neq
```

11. ***R code for figure 8:***

```
rm(list=ls())
x=seq(0,25,0.001)
library(ggplot2)
library(patchwork)

y1=dnorm(x,7,1.99)
g1=ggplot(NULL,aes(x,y1))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
            title="P.D.F. of X ~ Normal (7,(1.99)^2)")

y2=dnorm(x,9,1.899)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
            title="P.D.F. of X ~ Normal(9,(1.899)^2)")

y3=dnorm(x,16,2)
```

```
g3=ggplot(NULL,aes(x,y3))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of x",y="Density",
               title="P.D.F. of X ~ Normal(16,(2)^2)")

y4=dnorm(x,18,2.45)
g4=ggplot(NULL,aes(x,y4))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of x",y="Density",
               title="P.D.F. of X ~ Normal(18,(2.45)^2)")

y5=dnorm(x,7,0.88)
g5=ggplot(NULL,aes(x,y5))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of x",y="Density",
               title="P.D.F. of X ~ Normal(7,(0.88)^2)")

y6=dnorm(x,9,0.95)
g6=ggplot(NULL,aes(x,y6))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of x",y="Density",
               title="P.D.F. of X ~ Normal(9,(0.95)^2)")

y7=dnorm(x,16,0.983)
g7=ggplot(NULL,aes(x,y7))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of x",y="Density",
               title="P.D.F. of X ~ Normal(16,(0.983)^2)")

y8=dnorm(x,18,0.884)
g8=ggplot(NULL,aes(x,y8))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of x",y="Density",
               title="P.D.F. of X ~ Normal(18,(0.884)^2)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,g7,g8,nrow=2,top = paste("Graphs of Some Normal ( μ,
sigma^2 ) Distributions Corresponding to Different Values of ( μ, sigma^2 )"))
```

12. <u>R code for figure 9:</u>

```
rm(list=ls())
x=seq(0,25,0.001)
library(ggplot2)
library(patchwork)

y1=dnorm(x,8,1)
g1=ggplot(NULL,aes(x,y1))+
     geom_line(lwd=1,col=4)+
     labs(x="Values of μ",y="Density",
               title="P.D.F. of μ ~ Normal (8,(1)^2)")
```

```
y2=dnorm(x,10,1.50)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of μ",y="Density",
            title="P.D.F. of μ ~ Normal(10,(1.50)^2)")

y3=dnorm(x,12,1.20)
g3=ggplot(NULL,aes(x,y3))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of μ",y="Density",
            title="P.D.F. of μ ~ Normal(12,(1.20)^2)")

y4=dnorm(x,19,0.591)
g4=ggplot(NULL,aes(x,y4))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of μ",y="Density",
            title="P.D.F. of μ ~ Normal(19,(0.591)^2)")

y5=dnorm(x,21,0.512)
g5=ggplot(NULL,aes(x,y5))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of μ",y="Density",
            title="P.D.F. of μ ~ Normal(21,(0.512)^2)")

y6=dnorm(x,22,0.498)
g6=ggplot(NULL,aes(x,y6))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of μ",y="Density",
            title="P.D.F. of μ ~ Normal(22,(0.498)^2)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,nrow=2,top = paste("Graphs of Some Normal ( μ0, sigma0^2 ) Distributions Corresponding to Different Values of ( μ0, sigma0^2 )"))
```

13. ***R code for figure 10:***

```
#-----------------------------------------------(real prior knowledge)
rm(list=ls())
x=seq(0,25,0.001)
library(ggplot2)
y8=dnorm(x,19,0.5)
g8=ggplot(NULL,aes(x,y8))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of μ",y="Density",
            title="P.D.F. of μ ~ Normal(19,(0.5)^2)")
g8
```

14. ***R code for Normal Distribution with unknown variance but known mean:***

```
rm(list=ls())
set.seed(seed=987654321)
mu=20
sig=rep(sqrt(2),8)
```

```
sig2=2
n=15
B=1000
m=c(3,8,1,3,5,20,40,90)
the=c(0.5,1,4,7,9,39,79,179)

mle=array(0);sderr.mle=array(0);sig2.hat.mean=array(0);sig2.hat.boot=array(0);
sd.err.sig2.boot=array(0);

#Drawing the random sample from Normal
obs = rnorm(n,mu,sig[1])

#----------------------Classical Approach
mle=rep(((n-1)/n)*(var(obs)),8)
sderr.mle=rep(mle[1]*sqrt(2/(n-1)),8)


#----------------------Bayesian Approach

for(i in 1:length(sig)){

sig2.hat.mean[i] = (0.5*sum((obs-mu)^2)+the[i])/(m[i]+(n/2)-1)

for(j in 1:B){

 boot.obs=sample(obs,n,replace=T)
 sig2.hat.boot[j]= (0.5*sum((boot.obs-mu)^2)+the[i])/(m[i]+(n/2)-1)
}
sd.err.sig2.boot[i]=sd(sig2.hat.boot)

}

data.frame(sig2,mle,sderr.mle,sig2.hat.mean,sd.err.sig2.boot,m,the)
```

15. ***R code for figure 11:***

```
rm(list=ls())
library("invgamma")
x=seq(0.000001,5,0.001)
library(ggplot2)
library(patchwork)

y1=dinvgamma(x,0.5,0.5)
g1=ggplot(NULL,aes(x,y1))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
            title="P.D.F. of X ~ Inverse Gamma (0.5,0.5)")

y2=dinvgamma(x,0.5,2)
g2=ggplot(NULL,aes(x,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of x",y="Density",
```

```
        title="P.D.F. of X ~ Inverse Gamma (0.5 , 2)")

y3=dinvgamma(x,2,1)
g3=ggplot(NULL,aes(x,y3))+
  geom_line(lwd=1,col=4)+
  labs(x="Values of x",y="Density",
       title="P.D.F. of X ~ Inverse Gamma (2 , 1)")

y4=dinvgamma(x,50,99)
g4=ggplot(NULL,aes(x,y4))+
  geom_line(lwd=1,col=4)+
  labs(x="Values of x",y="Density",
       title="P.D.F. of X ~ Inverse Gamma (50 , 99)")

y5=dinvgamma(x,5,1)
g5=ggplot(NULL,aes(x,y5))+
  geom_line(lwd=1,col=4)+
  labs(x="Values of x",y="Density",
       title="P.D.F. of X ~ Inverse Gamma (5 , 1)")

y6=dinvgamma(x,9,3)
g6=ggplot(NULL,aes(x,y6))+
  geom_line(lwd=1,col=4)+
  labs(x="Values of x",y="Density",
       title="P.D.F. of X ~ Inverse Gamma (9 , 3)")

y7=dinvgamma(x,7,20)
g7=ggplot(NULL,aes(x,y7))+
  geom_line(lwd=1,col=4)+
  labs(x="Values of x",y="Density",
       title="P.D.F. of X ~ Inverse Gamma (7 , 20)")

y8=dinvgamma(x,10,30)
g8=ggplot(NULL,aes(x,y8))+
  geom_line(lwd=1,col=4)+
  labs(x="Values of x",y="Density",
       title="P.D.F. of X ~ Inverse Gamma (10 , 30)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,g7,g8,nrow=2,top = paste("Graphs of Some Inverse
Gamma ( m , θ ) Distributions Corresponding to Different Values of ( m , θ )"))
```

16. ***R code for figure 12:***

```
rm(list=ls())
library("invgamma")
x=seq(0.000001,5,0.001)
x1=sqrt(x)
library(ggplot2)
library(patchwork)
y1=dinvgamma(x,3,0.5)
g1=ggplot(NULL,aes(x1,y1))+
```

```
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (3 , 0.5)")

y2=dinvgamma(x,8,1)
g2=ggplot(NULL,aes(x1,y2))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (8 , 1)")

y3=dinvgamma(x,1,4)
g3=ggplot(NULL,aes(x1,y3))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (1 , 4)")

y4=dinvgamma(x,3,7)
g4=ggplot(NULL,aes(x1,y4))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (3 , 7)")

y5=dinvgamma(x,5,9)
g5=ggplot(NULL,aes(x1,y5))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (5 , 9)")

y6=dinvgamma(x,20,39)
g6=ggplot(NULL,aes(x1,y6))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (20 , 39)")

y7=dinvgamma(x,40,79)
g7=ggplot(NULL,aes(x1,y7))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (40 , 79)")

y8=dinvgamma(x,90,179)
g8=ggplot(NULL,aes(x1,y8))+
    geom_line(lwd=1,col=4)+
    labs(x="Values of sigma",y="Density",
         title="Conjugate Prior: Inverse Gamma (90 , 179)")

library(gridExtra)
grid.arrange(g1,g2,g3,g4,g5,g6,g7,g8,nrow=2,top = paste("Graphs of Some Inverse
Gamma ( m , θ ) Distributions Corresponding to Different Values of ( m , θ ) Plotted
Against Standard Deviation"))
```