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

Introduction:

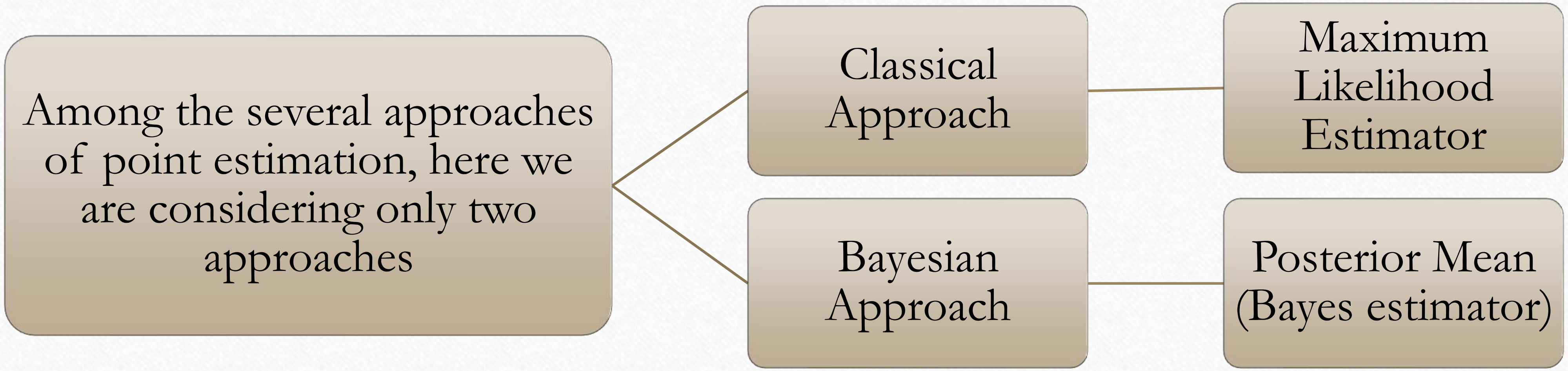
One of the main objectives of Statistics is to draw inferences about a population from the analysis of a sample drawn from that population.

- Statistical Inference
  - Theory of Estimation
    - Point Estimation
    - Interval Estimation
  - Testing of Hypothesis

❖ 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.

❖ Let $(x_1, x_2, \ldots., x_n)$ be the realisation of a random sample $(X_1, X_2, \ldots., X_n)$ from $f_X(.\,;\theta)$.

❖ Point Estimation takes into account to pick a suitable statistic, a function of sample observations, that best estimates the unknown population parameter θ.



Classical Approach

The unknown population parameter is assumed to be fixed quantity.

Bayesian Approach

The unknown population parameter is assumed to be random quantity or a random variable itself.

Now we will consider some standard distributions which will contain an unknown parameter. After that, we will try to estimate the unknown parameter by Maximum Likelihood Estimator in support of Classical approach and by Posterior Mean(Bayes Estimator) in support of Bayesian approach.

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

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

$$0 \qquad ; \quad otherwise$$

Here we draw a random sample of size n from Binomial (m , p) and consider m is known but p is unknown.

Finding an estimator of p in support of Classical approach:

- ❖ The estimate of the maximum likelihood estimator of p is given by,

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

- ❖ The estimate of the standard error of $\boldsymbol{p_{MLE}}$ is given by,

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

Finding an estimator of p in support of Bayesian approach:

- We consider p; $0<p<1$ to be a random quantity.
- We consider that prior distribution of p as $Beta(a, b)$ distribution of 1st kind which is actually a conjugate prior distribution.
- Thus the posterior distribution of p is following $\boldsymbol{Beta}(\sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{a}, \boldsymbol{mn} - \sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{b})$ of 1st kind.
- The estimate of the Bayes Estimator is given by,

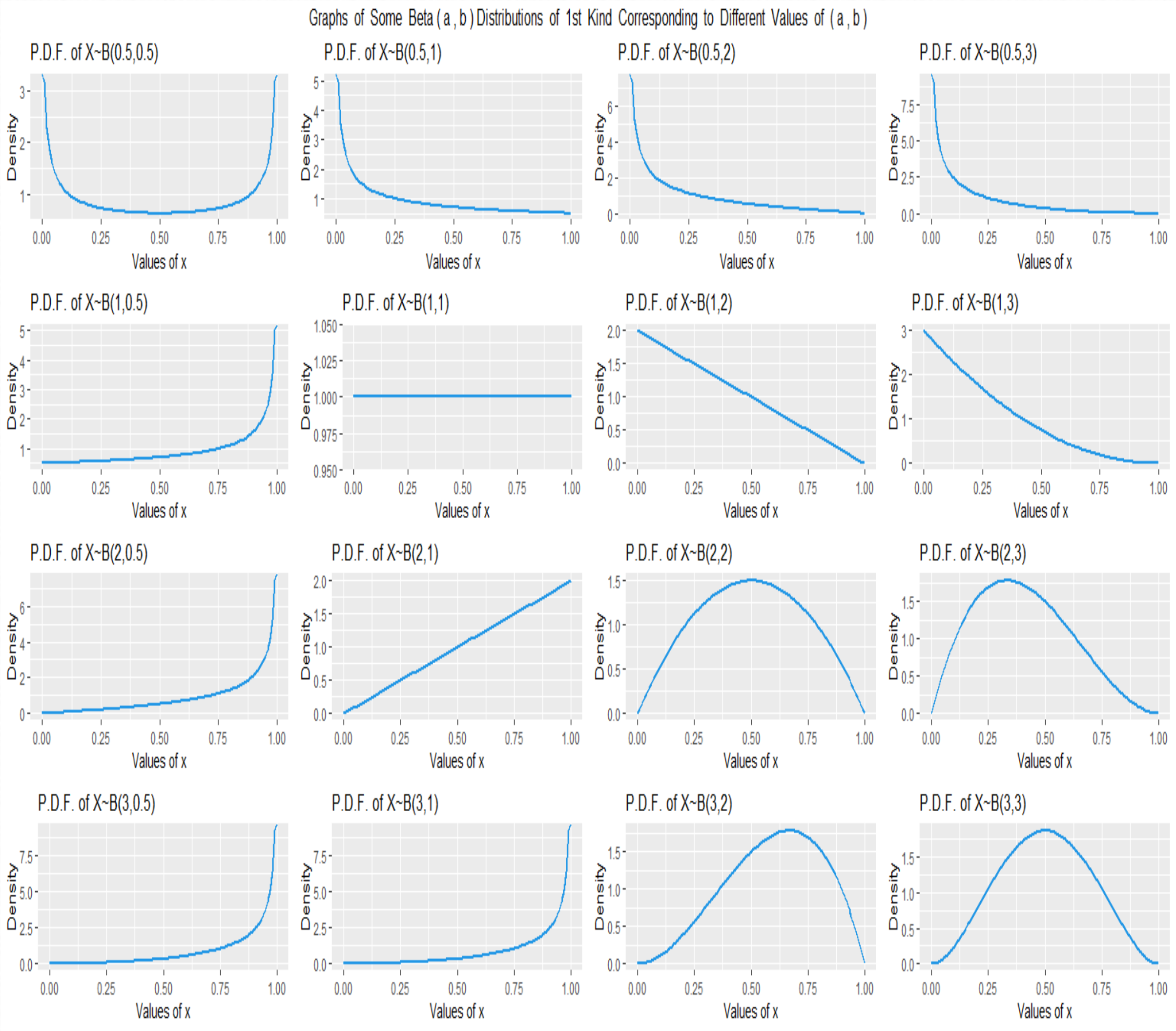
$$\widehat{\boldsymbol{p_b}} = \frac{\boldsymbol{a} + \sum_{i=1}^{n} \boldsymbol{x_i}}{\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{mn}}$$

- The standard error of the Bayes Estimator is computed by the method of bootstrap.

Choice of Prior($\boldsymbol{Beta(a, b)}$) Distributions:

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 $a < b$.

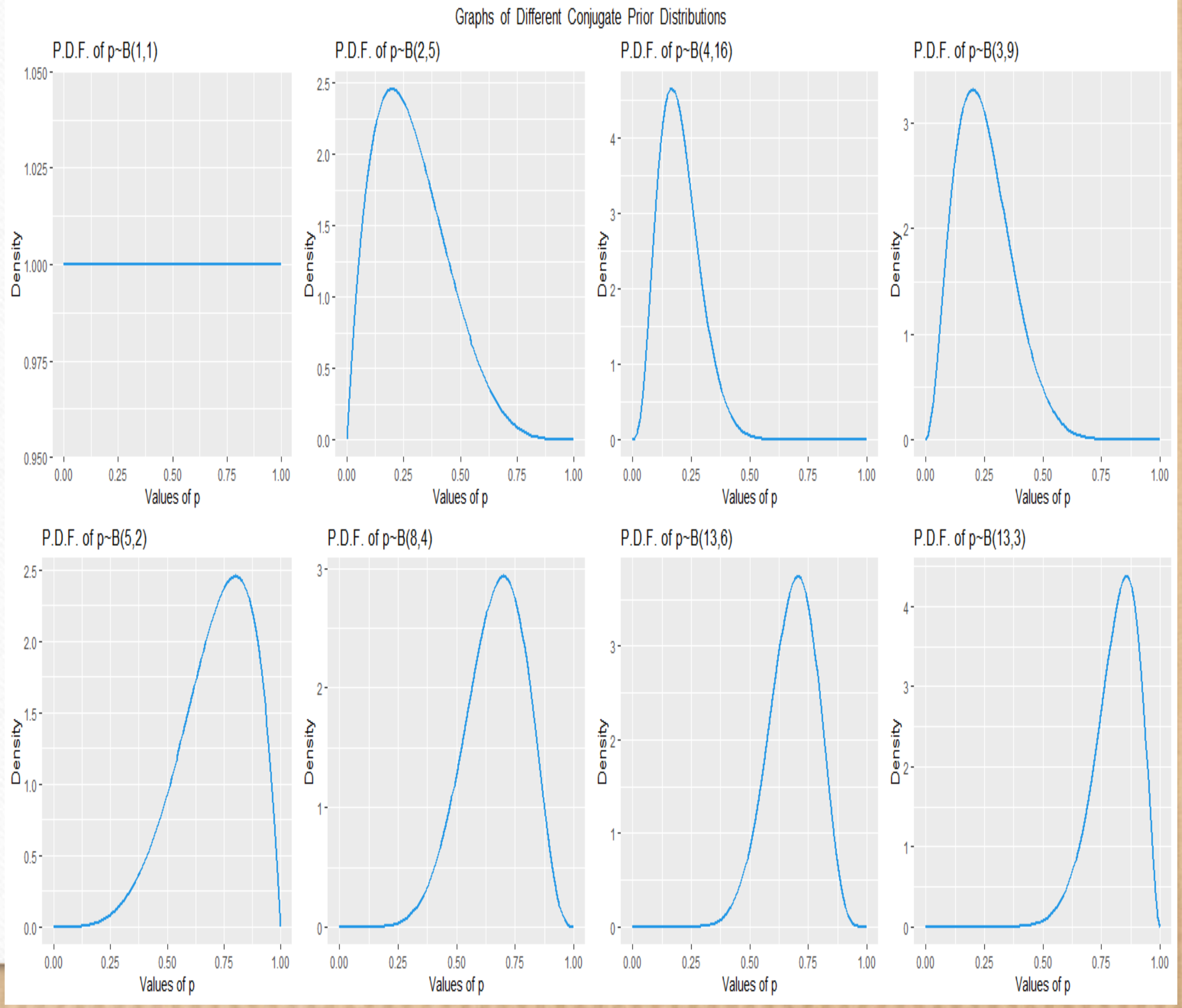
On the other hand, if we have prior belief that, on an average p takes higher values, then we should consider such a prior distribution $Beta(a, b)$ which assigns high density towards the higher values of p, that is we should take a $Beta(a, b)$ prior distribution such that $a > b$.



Illustrating Example:

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

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*** given in the previous slide.

- 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.

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

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

$$0 \qquad ;\quad otherwise$$

Here we draw a random sample of size n from $Poisson\ (\lambda)$ and consider λ is unknown.

Finding an estimator of $\boldsymbol{\lambda}$ in support of Classical approach:

- The estimate of the maximum likelihood estimator of λ is given by,

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

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

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

Finding an estimator of λ in support of Bayesian approach:

- We consider λ; λ>0 to be a random quantity.
- We consider that prior distribution of λ as $Gamma(m, \theta)$ distribution which is actually a conjugate prior distribution.
- Thus the posterior distribution of λ follows $\boldsymbol{Gamma(\sum_{i=1}^{n} x_i + m, n + \theta)}$ distribution.
- The Bayes Estimator is given by,

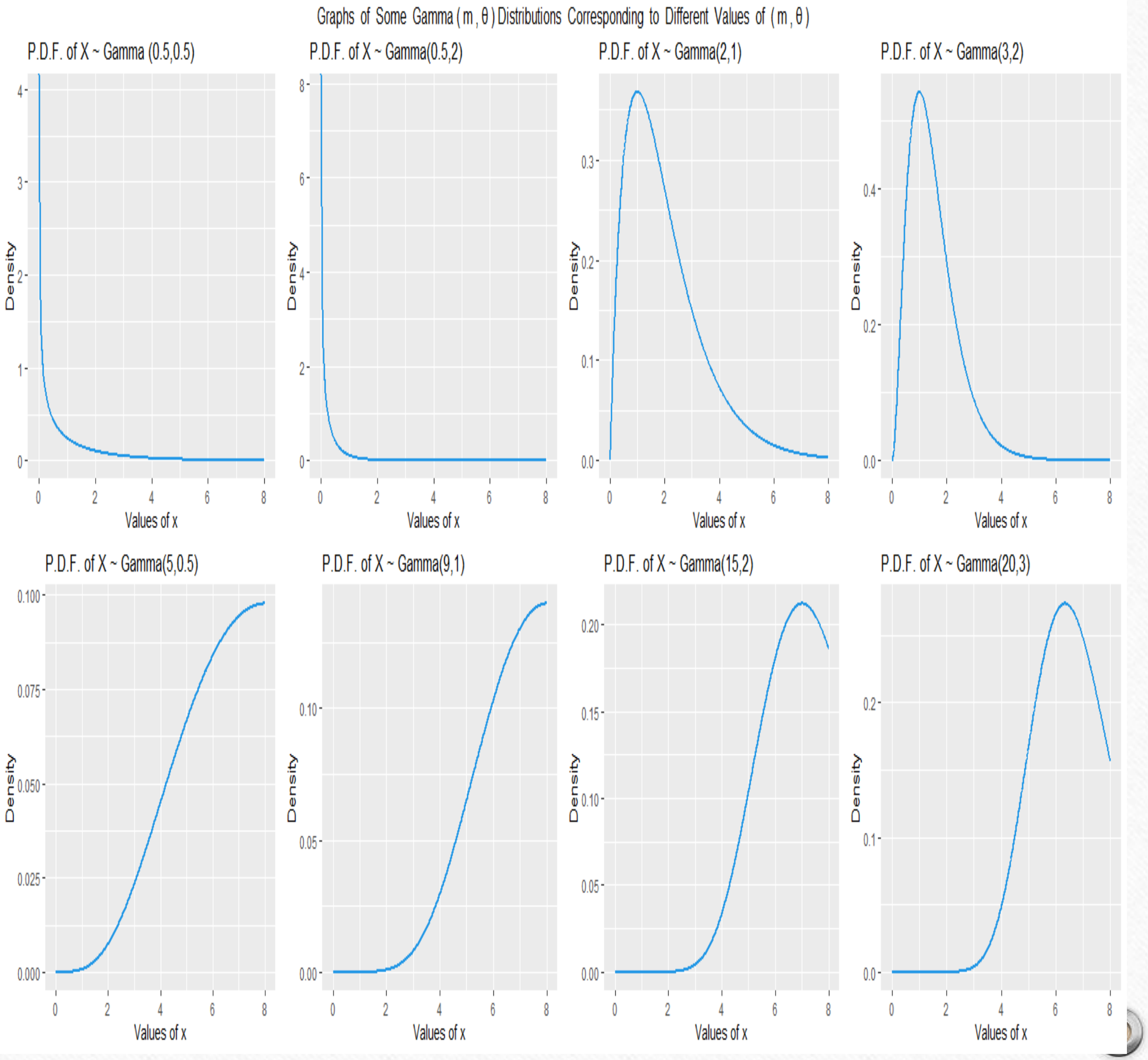
$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{b}} = \frac{\sum_{i=1}^{n} \boldsymbol{x_i} + \boldsymbol{m}}{\boldsymbol{n} + \boldsymbol{\theta}}$$

- The standard error of the Bayes Estimator is computed by the method of bootstrap.

Choice of Prior(Gamma($\boldsymbol{m, \theta}$)) Distributions:

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.

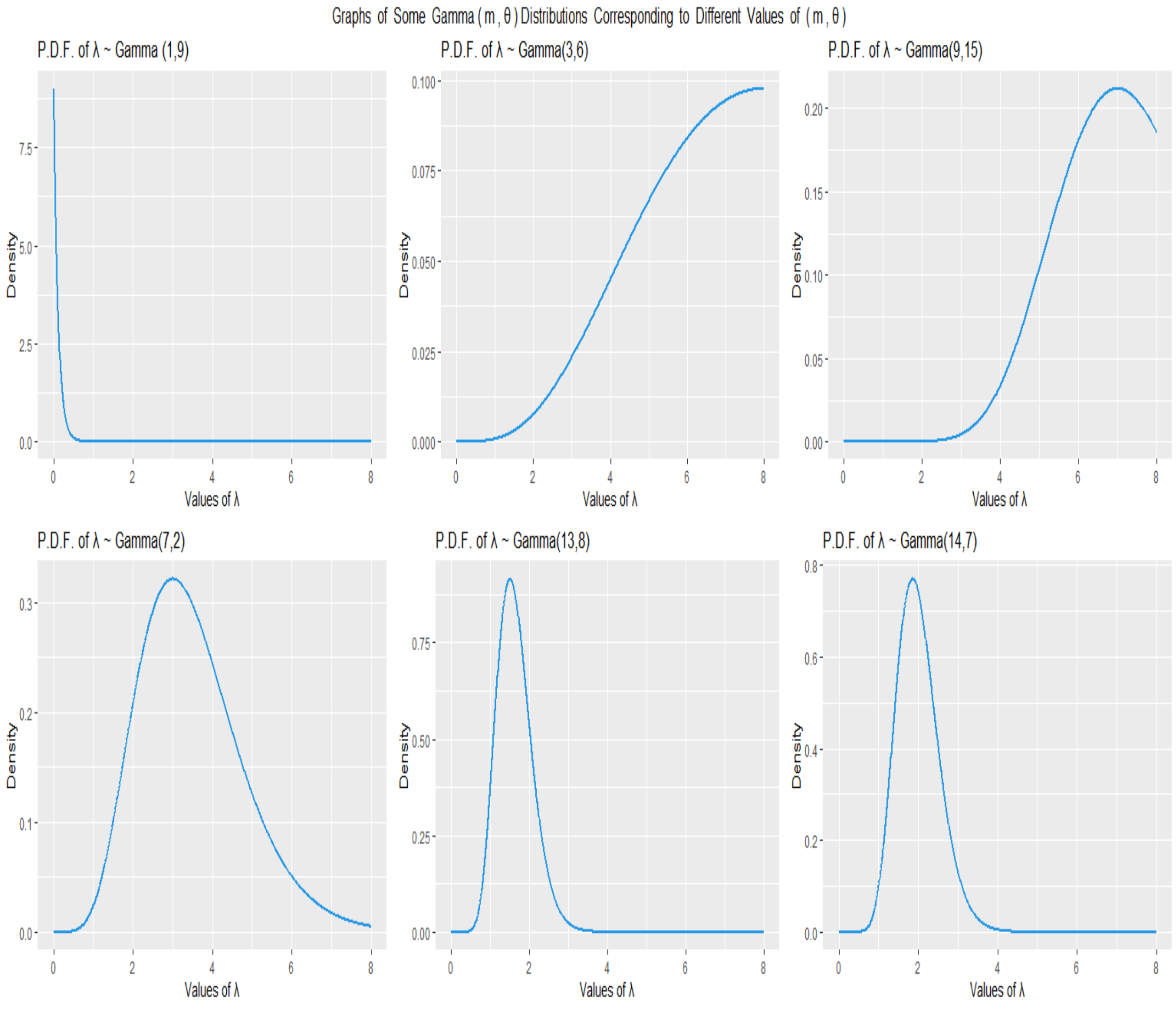
On the other hand, if we have prior belief that, on an average λ takes higher values, then we should consider such a prior distribution $Gamma(m, \theta)$ which assigns high density towards the higher values of λ.



Illustrating Example:

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

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 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 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*** given in the previous slide.

- 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.

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

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

$$0 \quad ; \quad otherwise$$

Here we draw a random sample of size n from $N(\mu\,,\sigma^2)$.

Finding an estimator of μ in support of Classical approach:

- ❖ The estimate of the maximum likelihood estimator of μ is given by,

$$\widehat{\boldsymbol{\mu}_{\boldsymbol{MLE}}} = \overline{\boldsymbol{x}} = \frac{\sum_{i=1}^{n} \boldsymbol{x}_i}{\boldsymbol{n}}$$

- ❖ The estimate of the standard error of μ_{MLE} is given by,

$$\boldsymbol{SE}(\widehat{\boldsymbol{\mu}_{\boldsymbol{MLE}}}) = \sqrt{\frac{\widehat{\boldsymbol{\sigma}^2}}{\boldsymbol{n}}} = \sqrt{\frac{\frac{\mathbf{1}}{\boldsymbol{n}-\mathbf{1}}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^{\mathbf{2}}}{\boldsymbol{n}}}$$

Finding an estimator of $\boldsymbol{\mu}$ in support of Bayesian approach:

- We consider $\mu; -\infty < \mu < \infty$ to be a random quantity.
- We consider that prior distribution of μ as $Normal(\mu_0, {\sigma_0}^2)$ distribution which is actually a conjugate prior distribution.
- Thus the posterior distribution of μ follows $\boldsymbol{Normal}\left(\frac{\frac{\sum_{i=1}^{n} x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)$ distribution.
- The Bayes Estimator is given by,

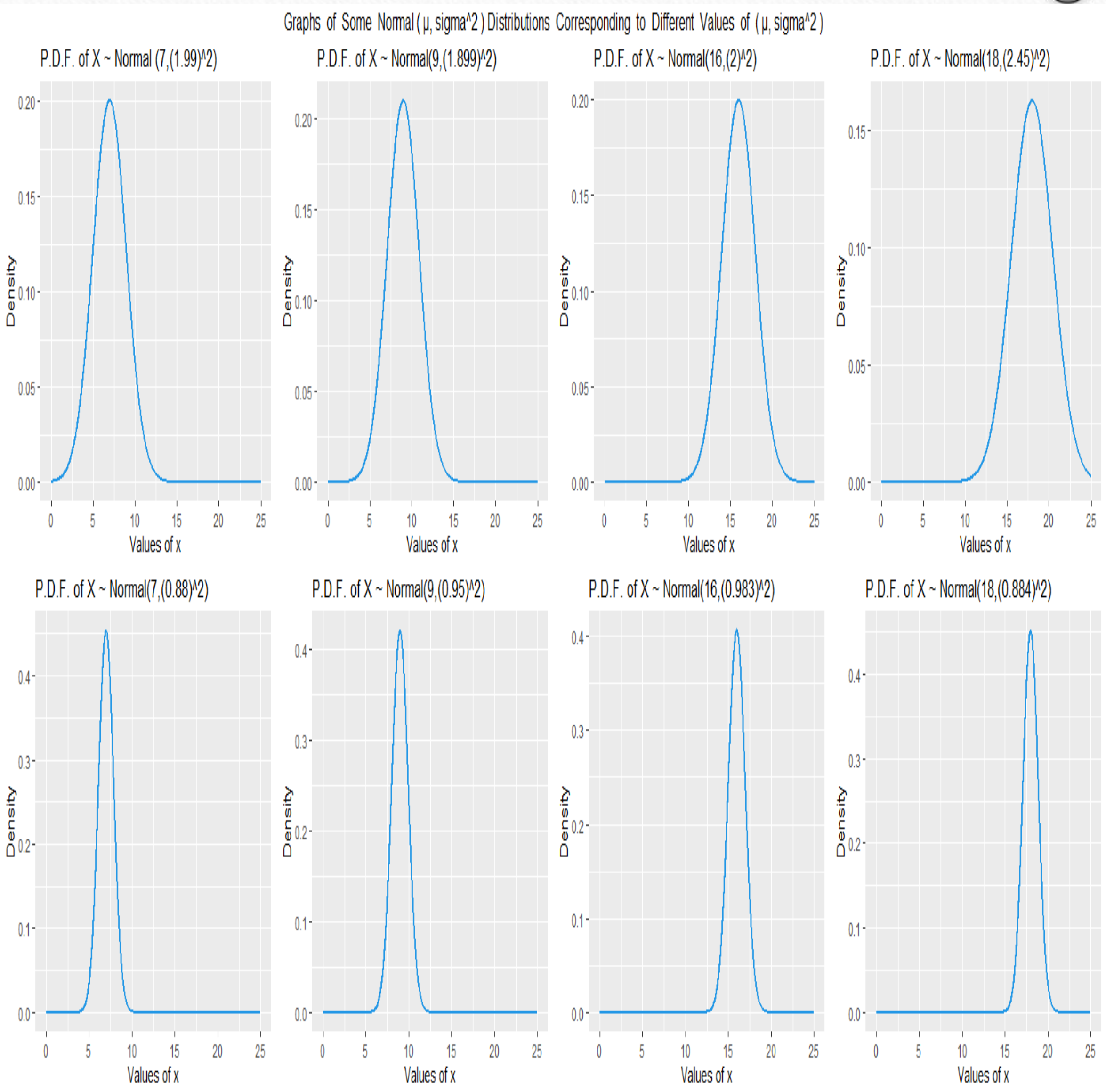
$$\widehat{\mu_b} = \frac{\frac{\sum_{i=1}^{n} x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}$$

- The standard error of the Bayes Estimator is computed by the method of bootstrap.

Choice of Prior($\boldsymbol{Normal(\mu_0, \sigma_0{}^2)}$) Distributions:

If we have the prior belief that the parameter under study μ; $-\infty < \mu < \infty$ 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.

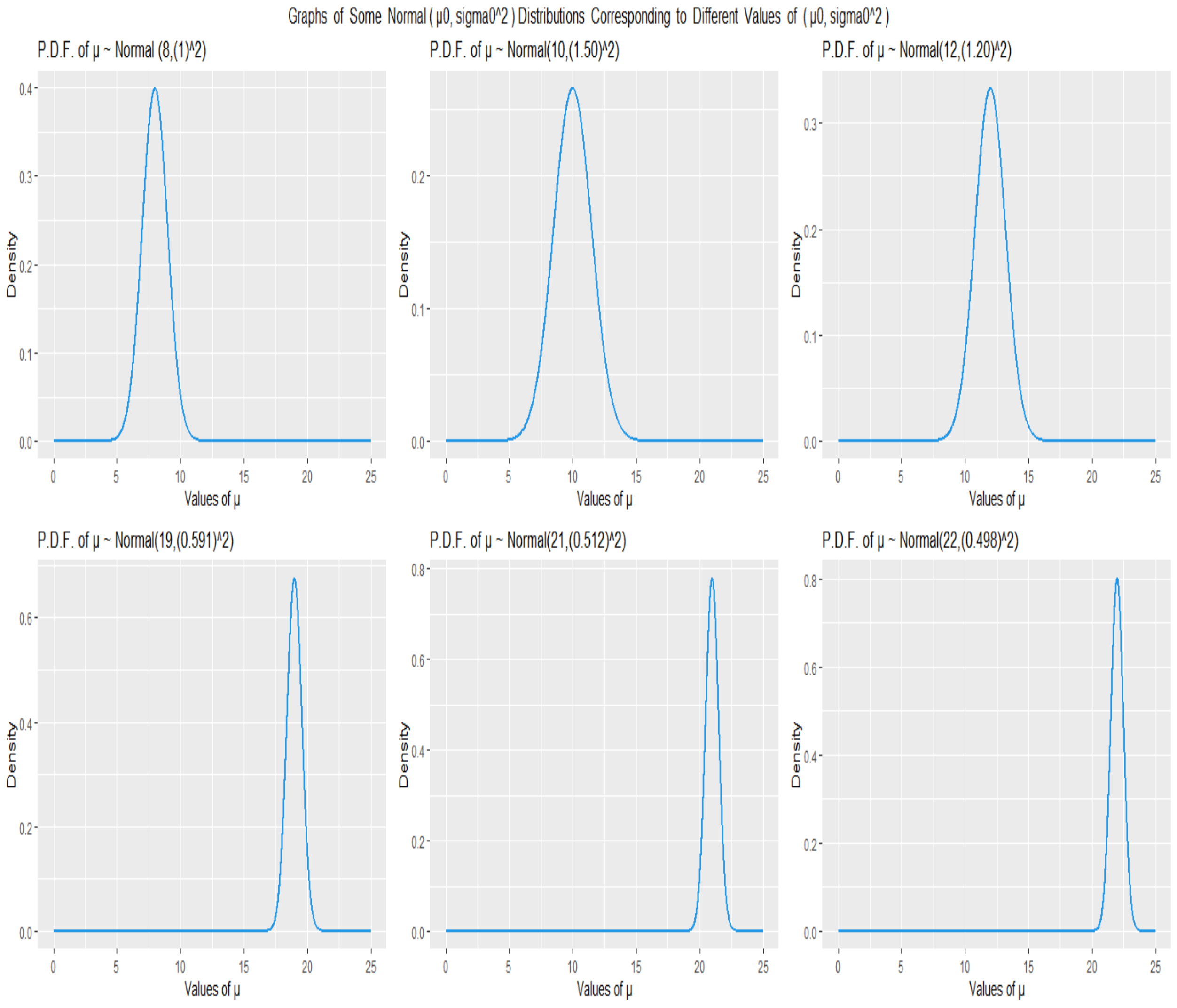
On the other hand, if we have prior belief that, on an average μ takes values around a particular value, say b, with a high concentration. Then we should choose a prior $Normal(\mu_0, \sigma_0{}^2)$ such that the prior distribution assigns high probability density close to that mentioned particular value b with a small standard deviation.



Illustrating Example:

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

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 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 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*** given in the previous slide.
- 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.

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

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

$$0 \quad ; \quad otherwise$$

Here we draw a random sample of size n from $N(\mu, \sigma^2)$.

Finding an estimator of σ^2 in support of Classical approach:

- The estimate of the maximum likelihood estimator of σ^2 is given by,

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

- The estimate of the standard error of $\sigma^2{}_{MLE}$ is given by,

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

Finding an estimator of σ^2 in support of Bayesian approach:

- We consider $\sigma^2;\ 0 < \sigma^2 < \infty$ to be a random quantity.

- We consider that prior distribution of σ^2 as $Inverse\ Gamma(m, \theta)$distribution which is actually a conjugate prior distribution.

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

$$0 \qquad\qquad , otherwise$$

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

- The Bayes Estimator is given by,

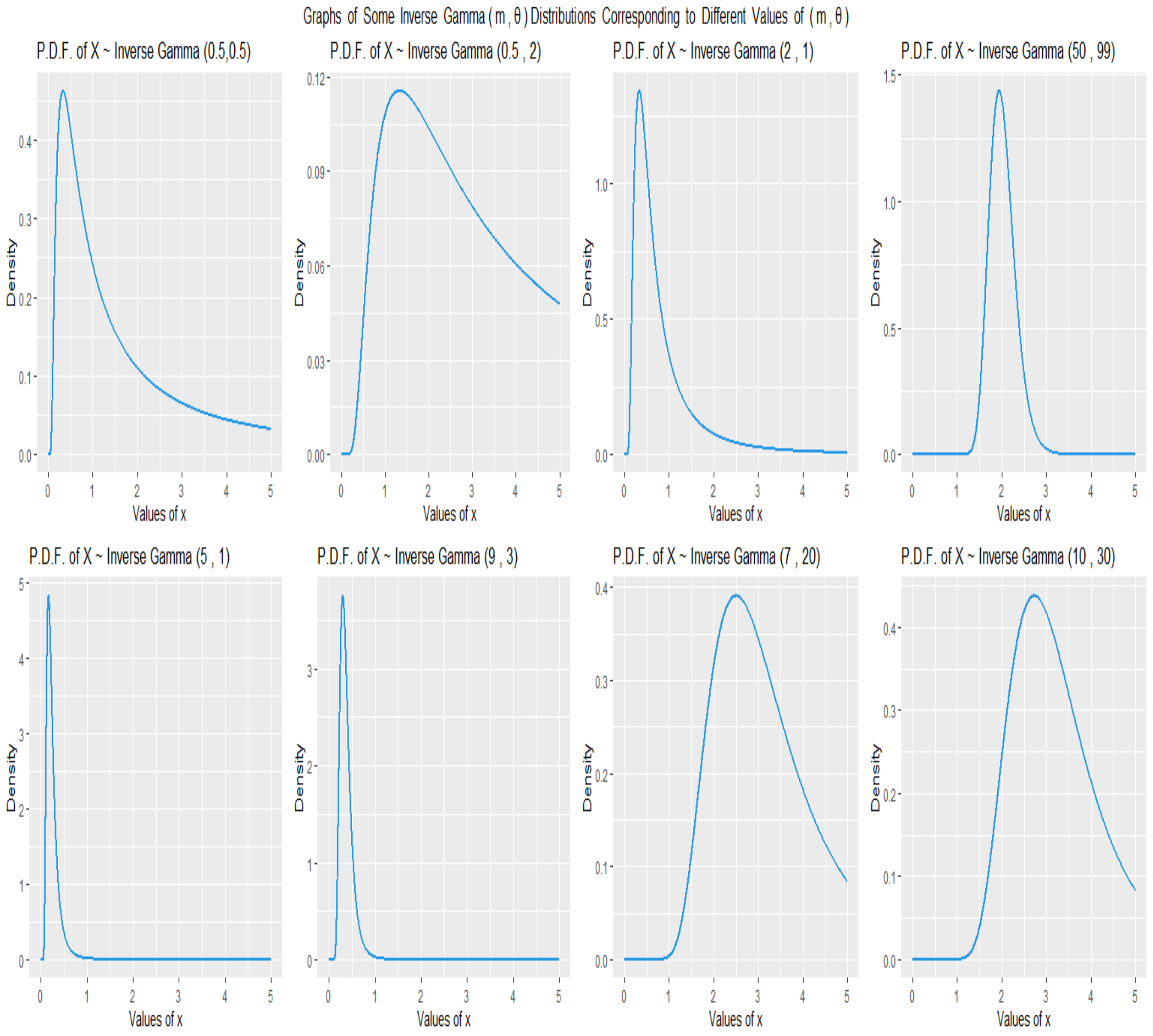
$$\widehat{\boldsymbol{\sigma^2}}_{\boldsymbol{b}} = \frac{\boldsymbol{\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \theta}}{\boldsymbol{m + \frac{n}{2} - 1}}\ ;\ \boldsymbol{m + \frac{n}{2} > 1}$$

- The standard error of the Bayes Estimator is computed by the method of bootstrap.

Choice of Prior($\boldsymbol{Inverse\ Gamma(m,\theta)}$) Distributions:

If we have the prior belief that the parameter under study σ^2; $0 < \sigma^2 < \infty$ 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.

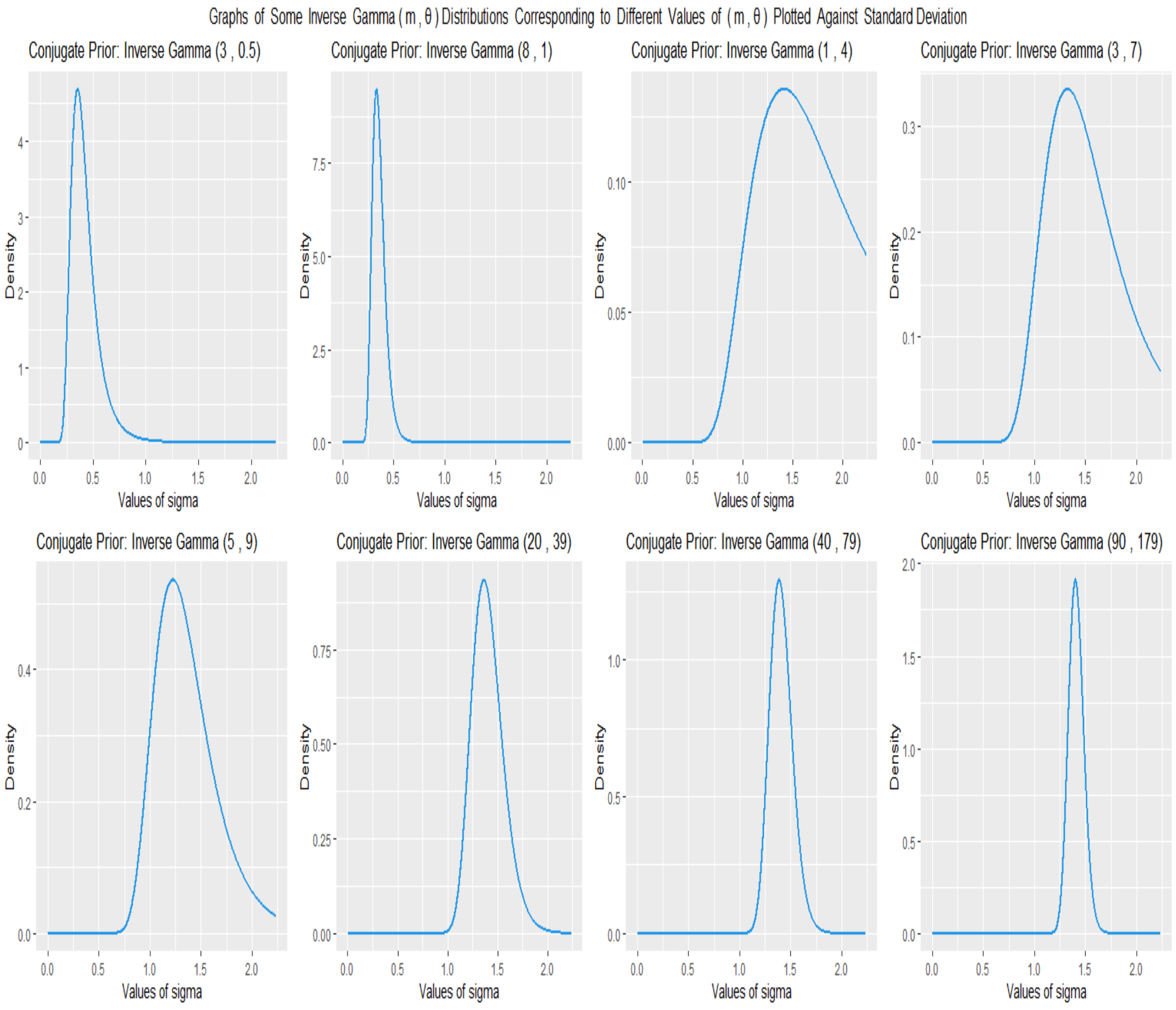
On the other hand, if we have the prior belief that the parameter under study σ^2; $\sigma^2 > 0$ can be considered as a random variable and on an average, it takes the higher value, then we should consider such a conjugate prior distribution $Inverse\ Gamma(m, \theta)$ which assigns high density towards the higher values of σ^2.



Illustrating Example:

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

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 $SE(\widehat{\sigma^2}_b)$	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, 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

- Those Bayes estimators corresponding to the conjugate priors (2nd row's 2nd, 3rd, 4th graphs), 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 (1nd row's 3rd, 4th graphs and 2nd row's 1st) which ensures moderate concentration of the values of standard deviation σ around the value 1.414.

References:

- Introduction To The Theory Of Statistics – Alexander M. Mood, Franklin A. Graybill, Duane C. Boes
- Introduction To Bayesian Statistics – William M. Bolstad, James M. Curran
- An Introduction to Statistical Learning – Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
- https://en.wikipedia.org/wiki/Conjugate_prior

Acknowledgement

First of all I would like to thank and acknowledge **Reverent Father Principle Dr. Dominic Savio. S.J.** for his blessings throughout the years in this college.

I must acknowledge my project supervisor **Prof. Pallabi Ghosh** for her immense support, guidance and valuable advice to complete this dissertation paper which have really enriched the content of my dissertation work.

I would also like to thank all the other professors of the **Department of Statistics**, who all have helped me to develop the mindset prone to research, which has made it possible for me to complete this project.

Finally, I must thank my **parents** and **friends** for their constant support throughout my undergraduate days.

Thank You