Comparative Study of Bayesian Estimators and Maximum Likelihood Estimators

Name: Subhrangshu Bit

Roll No.: 0436Subhrangshu Bit

Supervisor: Dr. Surabhi Dasgupta

1.Introduction:

Probability plays an important role in almost every field-industry, commerce, physical and biological sciences etc. Whenever we make a decision under uncertainty, consciously or otherwise we actually make a probability statement. In several instances the frequency definition of probability defies logical interpretation and a 'subjective' or 'degree of belief' approach makes more sense. For instance, suppose John claims: "I am 80% certain that I will win the scholarship". In 'degree of belief' context John has the same degree of confidence in his winning the scholarship as he would in the proposition that when a ball is picked up at random form an urn containing 8 white and 2 black balls, the ball will turn out to be white. It is impractical to imagine that if the scholarship contest is repeated say 1000 times, John will be successful in 800 contests and will be unsuccessful in the remaining 200 contests. This gives rise to our purpose of study on the comparison between the Maximum Likelihood Estimators (MLE) and Bayesian Estimators of three standard theoretical distributions. Through this dissertation we have tried to derive the exact form of estimators of the parameters involved and compared them with varying sample size.

Unlike classical methodology, in Bayesian framework the parameter is justifiably regarded as a random variable and the data once obtained is given or fixed. It combines with prior information with information contained in the data collected to formulate the posterior distribution which in turn provides a measure of the probability to events hereafter.

2.Methodology:

Here for the comparison of Maximum Likelihood and Bayesian estimation of parameters we have considered three theoretical probability distributions namely: Poisson distribution, Binomial distribution, Normal distribution.

At the foremost we draw a random sample of size, say n, from a particular theoretical probability distribution (aforementioned).

We are given with a statistical model i.e. a family of distributions $\left\{f(\,.\,;\theta)\,|\,\theta \in \Omega\right\}$, where θ denotes the parameter for the model and Ω being the parameter space. Now utilizing the given random sample we construct the likelihood function $L\left(\theta;\underset{\smile}{x}\right)$. Using the method of maximum likelihood we find the values of the model parameter, θ that maximizes the likelihood function i.e. we select the parameter values that make the data most probable.

Turning over to Bayesian framework θ is a random variable and the data $\underset{\smile}{x}$ is given or fixed. To start with we have the pre-sample of prior information about θ summarized by the prior distribution g(θ). Then we proceed to collect the data $\underset{\smile}{x}$ and the corresponding likelihood function $L(\theta\,|\,\underset{\smile}{x})$ gives us the additional information about θ.

Using Bayes' theorem we combine the prior information and the information contained in the sample and calculate our new or revised degree of belief about θ given by the posterior distribution-

$$\prod\left(\theta \middle| \underset{\sim}{x}\right) = \frac{g(\theta) f\left(\underset{\sim}{x} \middle| \theta\right)}{h\left(\underset{\sim}{x}\right)} = c g(\theta) f(\underset{\sim}{x} \mid \theta)$$

where $c^{-1} = \int \prod (\theta \mid \underset{\sim}{x}) d\theta.$

The prior information is subjective and is based on person's own experience and judgement.

Here, we have considered two approaches-

- Jeffreys' Invariant Prior:

  When we are in a state of ignorance about the parameter we need to choose a prior known as the Non-informative or vague prior. Following the rules (1) as suggested by Jeffrey, we have constructed the prior in the following discussion.

  (1): Reference Book [B1, page 21] (Bibliography).

- Natural Conjugate Prior:

  The family of prior distributions g(θ), $\theta \epsilon \Omega$ is called natural conjugate prior (NCP) if the corresponding posterior distribution belongs to the same family as g(θ). Following DeGroot (2) we construct the NCP g(θ) observing the corresponding likelihood function of the given distribution.

  (2): Reference Book [B1, page 27] (Bibliography).

3.Results and Discussion:

3.1 Poisson Distribution:

$$f(x) = e^{-}\frac{x}{x!} \quad ; \quad x = 0,1,2,3,\ldots$$

Here λ is our parameter of interest.

- **Fisher Maximum Likelihood Estimator:**

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from f(x).Thus the likelihood function-

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)}$$

Or, $\ln L(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \ln(\prod_{i=1}^{n} x_i!)$

Partially deriving with respect to λ -

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda}$$

Now, equating to zero -

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Therefore, sample mean is the MLE of λ.

- **<u>Bayesian Estimation:</u>**

- <u>Consider Jeffrey's Prior -</u>

$$g(\lambda) \propto \left| I(\lambda) \right|^{\frac{1}{2}}$$

Now,

$$I(\lambda) = -E\left[\frac{\partial^2 \ln f(x)}{\partial \lambda^2}\right]$$

Here, $f(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!}; x = 0,1,2,\ldots \\ 0; \text{otherwise} \end{cases}$

$$\ln f(x) = -\lambda + x\ln\lambda - \ln(x!)$$

Partially deriving with respect to λ -

$$\frac{\partial \ln f(x)}{\partial \lambda} = -1 + \frac{x}{\lambda}$$

Again, partially deriving with respect to λ -

$$\frac{\partial^2 \ln f(x)}{\partial \lambda^2} = -\frac{x}{\lambda^2}\text{-}$$

$$E\left[\frac{\partial^2 \ln f(x)}{\partial \lambda^2}\right] = -\frac{E(x)}{\lambda^2} = -\frac{1}{\lambda}$$

$$\therefore I(\lambda) = \frac{1}{\lambda}$$

$$\therefore g(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

Now we know the posterior distribution:

$$\prod(\lambda \mid \underbrace{x}) \propto g(\lambda) L(\lambda \mid \underbrace{x})$$

$$\prod(\lambda \mid \underbrace{x}) \propto e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i + \frac{1}{2} - 1}}{\prod_{i=1}^{n}(x_i!)}$$

$$\prod\left(\lambda \middle| \underbrace{x}\right) = K e^{-n\lambda} \lambda^{(\sum_{i=1}^{n} x_i + \frac{1}{2}) - 1}$$, where K is the normalizing

constant.

We obtain,

$$K = \frac{n^{\sum_{i=1}^{n} x_i + \frac{1}{2}}}{\Gamma(\sum_{i=1}^{n} x_i + \frac{1}{2})}$$

Let, $\sum_{i=1}^{n} x_i = S$

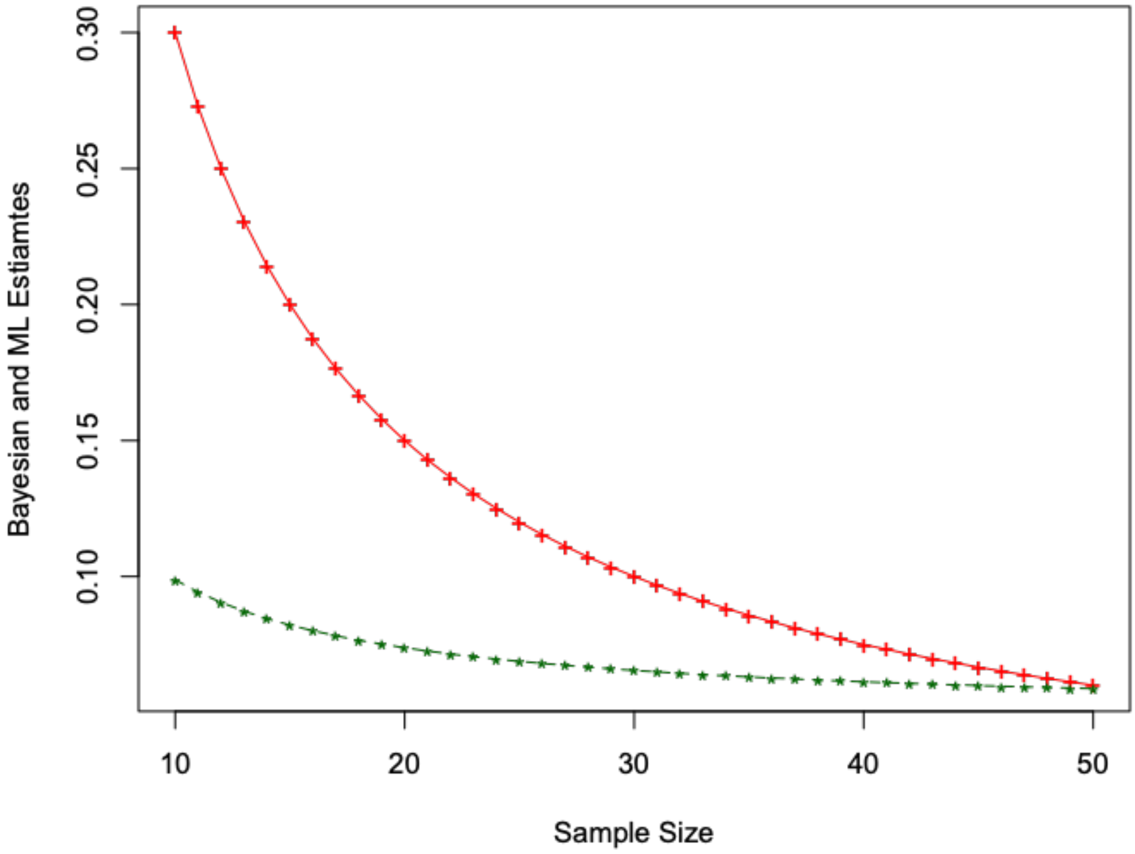
$$\therefore \prod(\lambda \mid \underbrace{x}) = \frac{n^{S+\frac{1}{2}}}{\Gamma(S+\frac{1}{2})} e^{-n\lambda} \lambda^{S+\frac{1}{2}-1} \quad \text{i.e. Gamma}(n,\ S+\frac{1}{2})$$

$$\therefore \lambda^* = E\left(\lambda \middle| \underbrace{x}\right) = \frac{S+\frac{1}{2}}{n} = \bar{x} + \frac{1}{2n}$$

(1) **Result:** *It is very clear that with increase in sample size λ^* tends to $\bar{x}$ i.e. both MLE and the Bayesian estimator of λ is the sample mean for considerably large sample size.*

Though $\lambda^*$is unbiased it is a consistent estimator for λ.

The plot clearly depicts the above result:



- <u>Consider Natural Conjugate Prior:</u>

$$g\left(\lambda \middle| a, b\right) = \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-b\lambda} \qquad ; a, b > 0 \ldots(1)$$

The posterior distribution is –

$$\prod\left(\lambda \middle| \underbrace{x}\right) \propto g\left(\lambda \middle| a, b\right)L\left(\lambda \middle| \underbrace{x}\right)$$

Or, $$\prod\left(\lambda \middle| \underbrace{x}\right) \propto \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-b\lambda}e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Let $\sum_{i=1}^{n} x_i = S,$

$$\prod\left(\lambda \middle| \underset{\smile}{x}\right) = Ke^{-\lambda(n+b)}\lambda^{S+a-1}$$, where K is the normalizing constant.

We obtain $K = \dfrac{(n+b)^{(S+a)}}{\Gamma(S+a)}$

$$\therefore \prod\left(\lambda \middle| \underset{\smile}{x}\right) = \frac{(n+b)^{(S+a)}}{\Gamma(S+a)} e^{-\lambda(n+b)}\lambda^{S+a-1}$$ i.e. Gamma $(n+b,\ S+a)$

$$\therefore \lambda^* = E\left(\lambda \middle| \underset{\smile}{x}\right) = \frac{S+a}{n+b}$$

The parameters (a, b) are usually unknown. We obtain the marginal distribution of X and estimate (a, b) by the method of moments. Equation (1) represents a Gamma prior. The joint distribution of (X, a, b) is given by-

$$h(x, a, b) = \frac{b^a}{\Gamma(a)} e^{-(b+1)\lambda} \frac{\lambda^{x+a-1}}{x!}$$

Integrating over λ -

$$h(x, a, b) = \frac{1}{x!}\frac{b^a}{\Gamma(a)} \int_0^\infty e^{-(b+1)\lambda}\lambda^{x+a-1} d\lambda$$

$$= \frac{b^a}{\Gamma(a)} \frac{1}{x!} \frac{\Gamma(x+a)}{(b+1)^{x+a}}$$

We know,

$$\sum_{x} h\left(x \middle| a, b\right) = 1$$

$$\sum_{x} \frac{b^a}{\Gamma(a)} \frac{1}{x!} \frac{\Gamma(x+a)}{(b+1)^{x+a}} = 1$$

Now,

$$E(X) = \sum_{x=0}^{\infty} x h\left(x \middle| a, b\right) = \sum_{x=1}^{\infty} \frac{b^a}{\Gamma(a)} \frac{1}{(x-1)!} \frac{\Gamma(x+a)}{(b+1)^{x+a}} = \sum_{x=1}^{\infty} \frac{b^a}{\Gamma(a)} \frac{1}{(x-1)!} \frac{\Gamma((x-1)+(a+1))}{(b+1)^{(x-1)+(a+1)}}$$

$$= \frac{a}{b} \sum_{x=1}^{\infty} \frac{b^{a+1}}{\Gamma(a+1)} \frac{1}{(x-1)!} \frac{\Gamma((x-1)+(a+1)}{(b+1)^{(x-1)+(a+1)}} = \frac{a}{b} \sum_{y=0}^{\infty} \frac{b^{a'}}{\Gamma(a')} \frac{1}{\left(y\right)!} \frac{\Gamma(y+a')}{(b+1)^{y+a'}}$$

, where $a' = a + 1$, $y = x - 1$.

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

Equating to the sample mean we have the estimating equation-

$$\bar{x} = \frac{a}{b}$$

Substituting in (2) we have –

$$\lambda^* = \frac{n\bar{x} + b\bar{x}}{n + b} = \bar{x}$$

(2) **Result:** _Considering Natural Conjugate Prior, the Bayesian estimator_ λ^* _is same as the MLE_ $\bar{x}$_. Hence,_ λ^* _may be looked upon as an improvement over MLE as along with possessing the properties of the MLE, it can be updated for a revised degree of belief about_ λ _as successive samples are drawn hereafter._

3.2 Binomial Distribution:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; x=0,1,2,\ldots,n \quad \ldots\ldots(1)$$

Here p is our parameter of interest.

- **Fisher Maximum Likelihood Estimator:**

Let $X_1, X_2, X_3, \ldots, X_N$ be a random sample from f(x).Thus the likelihood function

$$L(p) = \prod_{i=1}^{N} \left\{ \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \right\} = p^{\sum_{i=1}^{N} x_i}(1-p)^{Nn-\sum_{i=1}^{N} x_i} \prod_{i=1}^{N} \binom{n}{x_i}$$

Or, $lnL(p) = \sum_{i=1}^{N} \ln\binom{n}{x_i} + \sum_{i=1}^{N} x_i \ln(p) + (Nn - \sum_{i=1}^{N} x_i)\ln(1-p))$

Partially deriving with respect to p -

$$\frac{\partial lnL(p)}{\partial p} = \frac{\sum_{i=1}^{N} x_i}{p} - \frac{Nn - \sum_{i=1}^{N} x_i}{1-p}$$

Equating to zero we get-

$$\hat{p} = \frac{1}{Nn} \sum_{i=1}^{N} x_i = \frac{\bar{x}}{n}$$

Thus we obtain the MLE of p as the ratio of the sample mean to the population size. Further,

let $\sum_{i=1}^{N} X_i = S$

$$\therefore L(p) = p^S (1-p)^{Nn-S} \prod_{i=1}^{N} \binom{n}{x_i}$$

- **Bayesian Estimation:**

- Consider Jeffrey's Prior -

$$g(p) \propto \left|I(p)\right|^{1/2}$$

Now, $$I(p) = -E[\frac{\partial^2 lnf(p)}{\partial p^2}]$$

Here,

$$\ln f(p) = ln\binom{n}{x} + xln(p) + (n-x)\ln(1-p)$$

Partially deriving with respect to p -

$$\frac{\partial \ln f(p)}{\partial p} = 0 + \frac{x}{p} - \frac{n-x}{1-p}$$

Again, partially deriving with respect to p -

$$\frac{\partial^2 lnf(p)}{\partial p^2} = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}$$

$$\therefore -E\left[\frac{\partial^2 lnf(p)}{\partial p^2}\right] = -E[-\frac{x}{p} - \frac{n-x}{(1-p)^2}]$$

$$= \frac{np}{p^2} - \frac{n-np}{(1-p)^2} = n[\frac{1}{p} + \frac{1}{1-p}] = \frac{n}{p(1-p)}$$

$$\therefore g(p) \propto \frac{1}{\sqrt{p(1-p)}}$$

Or, $g(p) \propto p^{\frac{1}{2}-1}(1-p)^{\frac{1}{2}-1}$

$\therefore g(p) = Kp^{\frac{1}{2}-1}(1-p)^{\frac{1}{2}-1}$

, where $K = B\left(\frac{1}{2}, \frac{1}{2}\right)$

The posterior distribution of p is given by:

$$\prod(p\,|\,x) = cg(p)L(p) = cp^{S+\frac{1}{2}-1}(1-p)^{Nn-S+\frac{1}{2}-1}$$

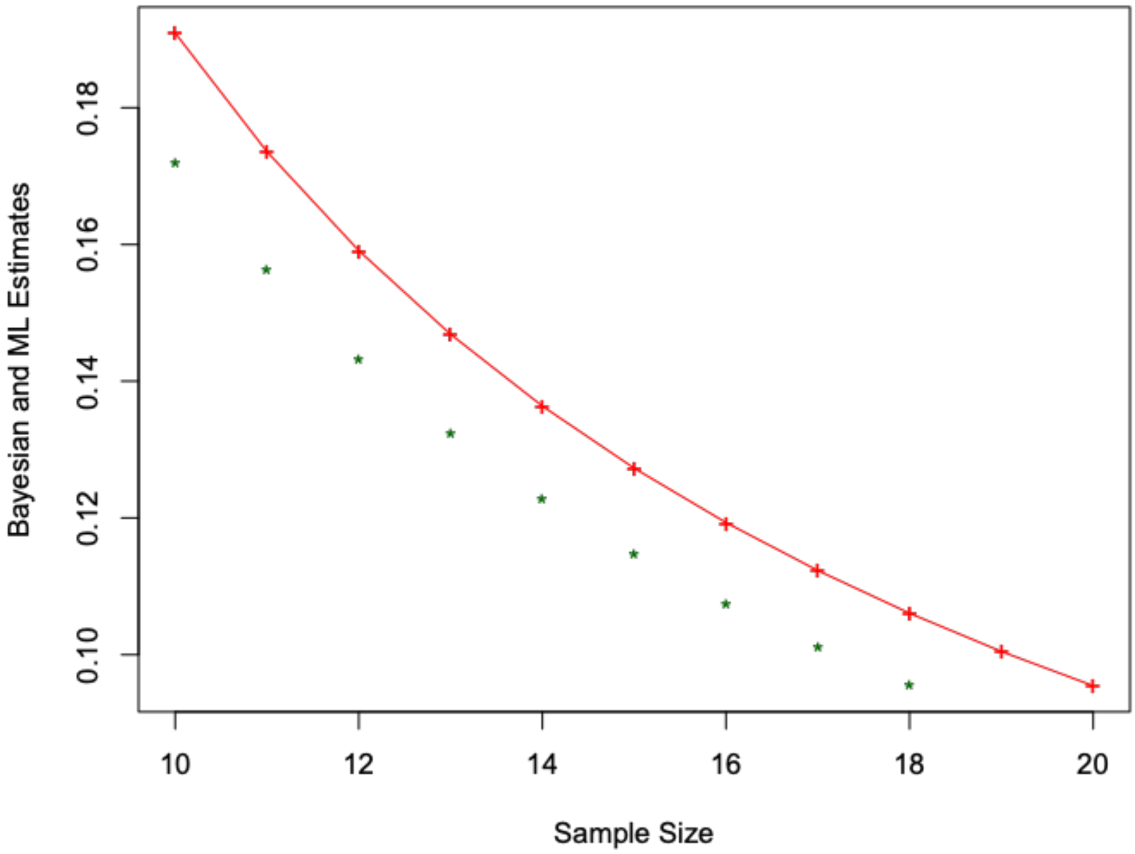
where $c^{-1} = \int_0^1 \prod\left(p\,\middle|\,x\right)dp = B\left(S+\frac{1}{2}, Nn+\frac{1}{2}-S\right)$

The Bayes' estimator –

$$p^* = E\left[p\,\middle|\,\underbrace{x}\right] = \frac{S+\frac{1}{2}}{Nn+\frac{1}{2}+\frac{1}{2}} = \frac{S+\frac{1}{2}}{Nn+1} = \frac{N\bar{x}+\frac{1}{2}}{Nn+1} = \frac{N(\bar{x}+\frac{1}{2N})}{N(n+\frac{1}{N})} = \frac{(\bar{x}+\frac{1}{2N})}{(n+\frac{1}{N})}$$

(3.2.1) Result: *For considerably large sample size p^* tends to $\frac{\bar{x}}{n}$ which is same as the MLE.*

The following plot-



- Consider the conjugate prior distribution of p-

$g(p) \propto p^{a-1}(1-p)^{b-1}$; a, b > 0 .... (2)

The posterior distribution of p is given by-

$\prod(p|x) = Kp^{S+a-1}(1-p)^{Nn+b-S-1}$, where

$$K^{-1} = \int_0^1 \prod(p|x)dp = B(S+a,\ Nn+b-S)$$

Restoring the normalizing constant, the posterior distribution of p is given by-

$$\prod\left(p\middle|x\right) = \frac{1}{B\left(S+a,\ Nn+b-S\right)}p^{S+a-1}(1-p)^{Nn+b-S-1} \qquad ;0<p<1$$

Hence the Bayes estimator is:

$$p* = E\left(p\middle|\underset{\sim}{x}\right) = \frac{B\left(S+a-1,Nn+b-S\right)}{B\left(S+a,Nn+b-a\right)} = \frac{S+a}{Nn+a+b} \quad \ldots (2.1)$$

The parameters (a, b) are usually unknown. One may obtain the marginal distribution of X and estimate (a, b) by the method of moments. (2) represents a Beta-prior. Restoring the normalizing constant and combining with (1) we get-

$$h\left(x,p\right) = \frac{\binom{n}{x}}{B\left(a,b\right)}p^{x+a-1}\left(1-p\right)^{n-x+b-1}$$

Or,

$$h\left(x\middle|p\right) = \frac{\binom{n}{x}}{B\left(a,b\right)}B\left(x+a,N-x+b\right) = \frac{\binom{n}{x}(x+a-1)!\left(n-x+b-1\right)!\left(a+b-1\right)!}{(a-1)!\left(b-1\right)!\left(n+a+b-1\right)!}$$

$$= \frac{\binom{x+a-1}{a-1}\binom{n-x+b-1}{b-1}}{\binom{n+a+b-1}{a+b-1}} \ldots\ldots..(3)$$

Then it follows:

$$K = \binom{n+a+b-1}{a+b-1} = \sum_{x=0}^{n}\binom{x+a-1}{a-1}\binom{n-x+b-1}{b-1} = \frac{(a+x-1)!\left(n-x+b-1\right)!}{(a-1)!\left(b-1\right)!x!(n-x)!}$$

∴

$$E(X) = \sum_{x=0}^{n} xh(x|p) = K\sum_{x=1}^{n} \frac{(a+x-1)!(b+n-x-1)!}{(a-1)!(b-1)!(x-1)!(n-x)!} = Ka\sum_{y=0}^{n-1} \frac{(a+y)!(b+n-y-2)!}{a!(b-1)!y!(n-y-1)!}$$

$$= Ka\sum_{y=0}^{m} \frac{(a+1+y-1)!(b+m-y-1)!}{(a+1-1)!(b-1)!(m-y)!y!}$$

and comparing with (3) we have-

$$E(X) = Ka\binom{m+a+1+b-1}{a+1+b-1} = Ka\binom{n+a+b-1}{a+b} = \frac{na}{a+b}$$

…….(4)

Equating (4) to the sample mean x̄ we have the estimating equation:

$$\bar{x} = \frac{na}{a+b}$$

Substituting in (2.1)

$$p* = \frac{N\bar{x}+a}{Nn+\frac{na}{\bar{x}}} = \frac{\bar{x}\left(N+\frac{a}{\bar{x}}\right)}{n\left(N+\frac{a}{\bar{x}}\right)} = \frac{\bar{x}}{n}$$

(3.2.2) Result: *Independent of the parameters (a, b) of the Beta prior (2), Bayes estimator of p identifies itself with the MLE of p.*

3.3 Normal Distribution:

Pdf of X: $f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ $\quad -\infty < x, \mu < \infty \quad \sigma > 0$

Here μ and σ are our parameters of interest.

- **Fisher Maximum Likelihood Estimator:**

$$L\left(\mu, \sigma \middle| \underbrace{x}\right) \propto \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Then-

$$lnL\left(\mu, \sigma^2 \middle| \underbrace{x}\right) = -\frac{n}{2}(\ln(2\pi) + \ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$$

Partially deriving with respect to μ -

$$\frac{\partial lnL(\mu, \sigma^2 | \underbrace{x})}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{\sigma^2}n(\bar{x} - \mu)$$

Equating to zero we have- $\hat{\mu} = \bar{x}$

Again, partially deriving with respect to σ^2 -

$$\frac{\partial lnL(\mu, \sigma^2 | \underbrace{x})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^{2^2}}\sum_{i=1}^{n}(x_i - \mu)^2 = -\frac{n}{2\sigma^{2^2}}(\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Equating to zero and recalling that $\hat{\mu} = \bar{x}$ we get-

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

- **Bayesian Estimation:**

(i) σ known: Conjugate prior for μ

We write the likelihood function as:

$$L\left(\mu,\sigma\middle|\underbrace{x}\right)=\frac{1}{\left(\sigma\sqrt{2\pi}\right)^{n}}e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right)}=\frac{1}{\left(\sigma\sqrt{2\pi}\right)^{n}}e^{-\frac{1}{2\sigma^2}\left(A+n(\bar{x}-\mu)^2\right)}$$

$$=e^{-\frac{n(\bar{x}-\mu)}{2\sigma^2}}\left(\frac{1}{(\sigma\sqrt{2\pi})^{n}}\right)e^{-\frac{A}{2\sigma^2}}$$

, where $A=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=e^{-\frac{n(\bar{x}-\mu)}{2\sigma^2}}\left(\frac{1}{(\sigma\sqrt{2\pi})^{n}}\right)e^{-\frac{A}{2\sigma^2}}$

We know $\bar{x}$ is sufficient for μ. Replacing $\bar{x}$ by m and $\frac{\sigma^2}{n}$ by δ^2.

$$g(\mu)\propto e^{-\frac{(\mu-m)^2}{2\delta^2}}\propto N(m,\delta^2)$$

Normalizing we get:

$$g(\mu)=\frac{1}{\delta\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu-m}{\delta}\right)^2}$$

Combining with the Likelihood function we have-

$$\pi\left(\mu\middle|\underbrace{x}\right)\propto e^{-\frac{1}{2}\left[\frac{(\bar{x}-\mu)^2}{\frac{\sigma^2}{n}}+\frac{(\mu-m)^2}{\delta^2}\right]}$$

$$\propto e^{-\frac{1}{2}\left[\mu^2\left\{\frac{1}{\delta^2}+\frac{n}{\sigma^2}\right\}-2\mu\left\{\frac{m}{\delta^2}+\frac{n\bar{x}}{\sigma^2}\right\}+\left\{\frac{n\bar{x}^2}{\sigma^2}+\frac{m^2}{\delta^2}\right\}\right]}\propto e^{-\frac{1}{2}\left[B\mu^2-2\mu C+D\right]}$$

$$\propto e^{-\frac{B}{2}\left(\mu-\frac{C}{B}\right)^2}$$

, where $B=\frac{n}{\sigma^2}+\frac{1}{\delta^2}$

$$C = \frac{m}{\delta^2} + \frac{n\bar{x}}{\sigma^2}$$

$$D = \frac{n\bar{x}^2}{\sigma^2} + \frac{m^2}{\delta^2}$$

$\pi\left(\mu \middle| \underset{\smile}{x}\right) = Ke^{-\frac{B}{2}\left(\mu - \frac{C}{B}\right)^2}$, where $K = \sqrt{B/2\pi}$

Then the Bayesian estimate of μ is-

$$\mu^* = E\left(\mu \middle| \underset{\smile}{x}\right) = \frac{C}{B} = \frac{\frac{m}{\delta^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\delta^2}} = \frac{m\sigma^2 + n\bar{x}\delta^2}{n\delta^2 + \sigma^2} = \frac{n\bar{x} + \frac{m\sigma^2}{\delta^2}}{\frac{\sigma^2}{\delta^2} + n} = \frac{n\bar{x} + \lambda m}{n + \lambda}$$

, where $\lambda = \dfrac{\sigma^2}{\delta^2}$

And we also obtain –

$$\sigma^{2*} = \frac{1}{B} = \frac{\sigma^2\delta^2}{\sigma^2 + n\delta^2} = \frac{\sigma^2}{n + \lambda}$$

(3.3.1) Result: *As n→∞ $\mu^* \rightarrow \bar{x}$, the maximum likelihood estimator (MLE) of μ. Further, as $\sigma^{2*} \rightarrow \infty$ which implies that as our prior information about μ becomes vaguer and vaguer, the posterior mean μ^* tends more to the sample mean $\bar{x}$ independent of the sample size.*

(ii) σ known: Jeffreys' prior for μ:

Suppose we are in a state of complete ignorance about μ and we represent our prior ignorance by Jeffreys' prior. In such a case, the likelihood and posterior distribution must be the same and hence we have-

$$\pi\left(\mu \middle| \underbrace{x}\right) = K\, e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}$$

$$K = \int_{-\infty}^{\infty} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} d\mu = \sqrt{(\frac{2\pi}{n})}\sigma$$

Restoring the normalizing constant, we have-

$$\pi\left(\mu \middle| \underbrace{x}\right) = \frac{1}{\frac{\sqrt{2\pi}\sigma}{\sqrt{n}}} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}, \quad -\infty < \mu < \infty$$

which implies the posterior distribution of μ is $N\left(\bar{x}, \frac{\sigma^2}{n}\right)$ and

$$\mu = E\left(\mu \middle| \underbrace{x}\right) = \bar{x}$$

(3.3.2) Result: *Clearly in the above case as the likelihood and the posterior distribution are same. Thus we have obtained the same estimator i.e. the sample mean for* μ.

(iii) Conjugate prior(μ and σ unknown):

The prior is hierarchical.

First, we assign the following prior to the mean, conditional on the variance:

$$p\left(\mu \middle| \sigma^2\right) = \sqrt{\frac{1}{\frac{2\pi\sigma^2}{\kappa_0}}}\, e^{-(\frac{1}{\frac{2\sigma^2}{\kappa_0}}(\mu_0-\mu)^2)}$$

that is, μ has a standard normal distribution with mean μ_0 and variance $\frac{\sigma^2}{\kappa_0}$.

Note that the variance of the parameter μ is assumed to be proportional to the unknown variance σ^2 of the data points. The constant of proportionality κ_0 determines how tight the prior is, that is, how probable we deem that μ is very close to the prior mean μ_0.

Then, we assign the following prior to the variance:

$$p\left(\sigma^2\right) = \frac{(\gamma_0 \sigma_0{}^2)^{\frac{\gamma_0}{2}}}{2^{\frac{\gamma_0}{2}} \Gamma \frac{\gamma_0}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{\gamma_0}{2}+1} e^{-\left(\frac{\gamma_0 \sigma_0{}^2}{2\sigma^2}\right)}$$

that is, has an inverse-chi square distribution with parameters γ_0, $\sigma_0{}^2$ and

We can think of $\frac{1}{\sigma_0{}^2}$ as our best guess of the precision of the data generating distribution. γ_0 is the parameter that we use to express our degree of confidence in our guess about the precision. The greater γ_0, the tighter our prior about $\frac{1}{\sigma^2}$ is, and the more we believe that $\frac{1}{\sigma^2}$ is close to $\frac{1}{\sigma_0{}^2}$.

Then the joint prior distribution is-

$$g(\mu, \sigma^2) = NI\chi^2(\mu_0, \kappa_0, \gamma_0, {\sigma_0}^2) = N\left(\mu \middle| \mu_0, \frac{\sigma^2}{\kappa_0}\right)\chi^{-2}(\sigma^2 | \gamma_0, \sigma_0)$$

$$= K^{-1}(\frac{1}{\sigma})(\frac{1}{\sigma^2})^{\frac{\gamma_0}{2}+1} e^{-(\frac{1}{2\sigma^2})(\gamma_0{\sigma_0}^2+\kappa_0(\mu_0-\mu)^2)}$$

, where $K = \frac{\sqrt{2\pi}}{\sqrt{\kappa_0}} \left\lceil (\frac{\gamma_0}{2}) \frac{1}{(\frac{\gamma_0{\sigma_0}^2}{2})^{\frac{\gamma_0}{2}}}\right.$

Now observe that –

$$Q_0(\mu) = S_0 + \kappa_0(\mu_0 - \mu)^2 = \kappa_0\mu^2 - 2(\kappa_0\mu_0)\mu + (\kappa_0{\mu_0}^2 + S_0)$$

$S_0 = \gamma_0{\sigma_0}^2$: Prior Sum of Squares

The posterior distribution is then obtained as:

$$\pi\left(\mu, \sigma^2 \middle| \underbrace{x}\right) = N\left(\mu \middle| \mu_0, \frac{\sigma^2}{\kappa_0}\right)I\chi^2(\sigma^2 | \gamma_0, {\sigma_0}^2)L\left(\mu, \sigma^2 \middle| \underbrace{x}\right)$$

$$\propto \{\left(\frac{1}{\sigma}\right)\left(\frac{1}{\sigma^2}\right)^{\frac{\gamma_0}{2}+1} e^{-\frac{1}{2\sigma^2}\left[\gamma_0{\sigma_0}^2 + \kappa_0(\mu_0-\mu)^2\right]}\}\{\frac{1}{\sigma^2}^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left[ns^2 + n(\bar{x}-\mu)^2\right]}\}$$

$$= NI\chi^2(\mu_n, \kappa_n, \gamma_n, {\sigma_n}^2)$$

where-

$\gamma_n = \gamma_0 + n$

Since ${S_0}^2 = \gamma_0{\sigma_0}^2$ then grouping the terms within the exponential:

Let ${S_n}^2 = \gamma_n{\sigma_n}^2$

$$S_0 + \kappa_0(\mu_0 - \mu)^2 + ns^2 + n(\bar{x} - \mu)^2 = \left(S_0 + \kappa_0\mu_0{}^2 + ns^2 + n\bar{x}^2\right) + \mu^2\left(n + \kappa_0\right) - 2(\kappa_0\mu_0 + n\bar{x})\mu$$

Comparing with $Q_0(u)$–

$$\kappa_n = \kappa_0 + n$$

$$\kappa_n\mu_n = \kappa_0\mu_0 + n\bar{x}$$

$$S_n + \kappa_n\mu_n{}^2 = S_0 + \kappa_0\mu_0{}^2 + ns^2 + n\bar{x}^2$$

$$\therefore S_n = S_0 + \kappa_0\mu_0{}^2 + ns^2 + n\bar{x}^2 - \kappa_n\mu_n{}^2$$

Rearranging the terms we have-

$$S_n = S_0 + ns^2 + \kappa_0\mu_0{}^2 + n\bar{x}^2 - \frac{\left(\kappa_0\mu_0 + n\bar{x}\right)^2}{\left(\kappa_0 + n\right)^2} = S_0 + ns^2 + \frac{n\kappa_0}{n + \kappa_0}(\mu_0 - \bar{x})^2$$

$\therefore$ Posterior Sum of squares = $\mathrm{S_n} = \gamma_n\sigma_n{}^2$ combines prior sum of squares $\mathrm{S_0} = \gamma_0\sigma_0{}^2$, the sample sum of squares ns^2 and a term due to uncertainty in the mean.

$$\therefore \mu_n = \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_n}$$

$$\kappa_n = \kappa_0 + n$$

$$\gamma_n = \gamma_0 + n$$

$$\sigma_n{}^2 = \frac{1}{\gamma_n}[\gamma_0\sigma_0{}^2 + \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 + \frac{n\kappa_0}{n + \kappa_0}\left(\mu_0 - \bar{x}\right)^2]$$

The posterior mean i.e. the Bayesian estimate is obtained as:

$$\mu^* = E\left(\mu \,\middle|\, \underbrace{x}\right) = \mu_n = \frac{1}{\kappa_0 + n}[n\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \kappa_0\mu_0]$$

$$\sigma^{2^*} = E\left(\sigma^2 \middle| \underbrace{x}\right) = \frac{\gamma_n \sigma_n{}^2}{\gamma_n - 1}$$

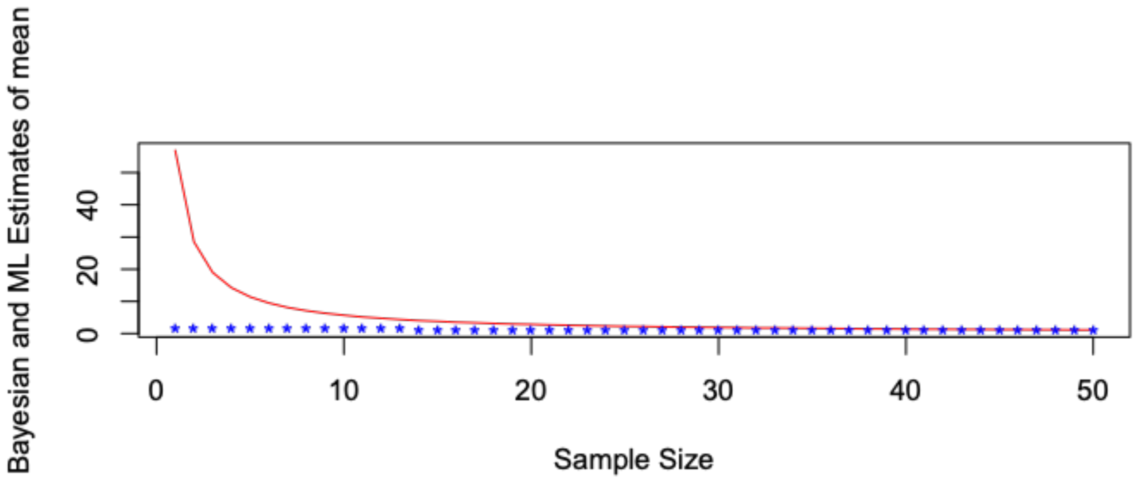
(3.3.3) Result: *The posterior mean μ_n is the weighted average of two information:*

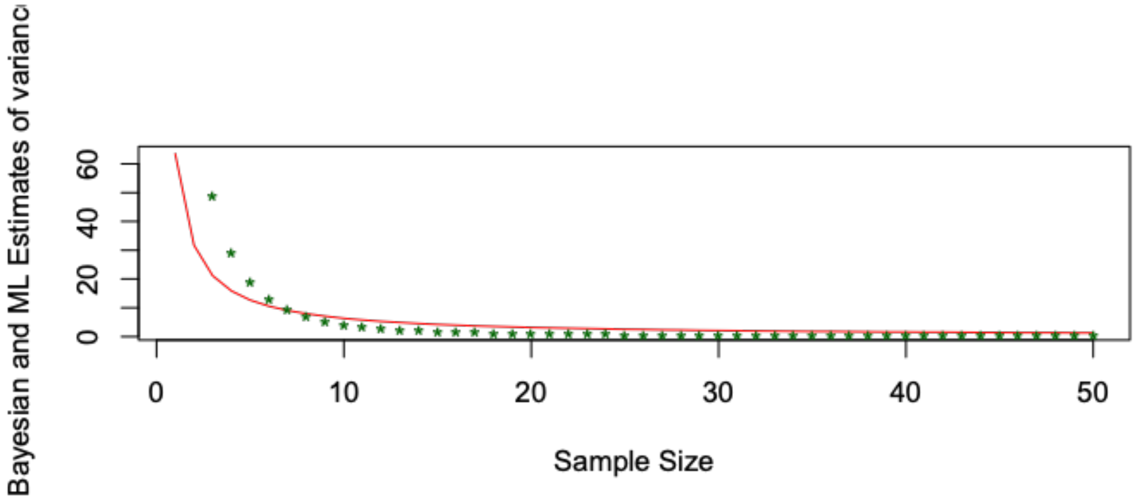
1. The sample mean $\bar{x}$ of the observed data;

2. The prior mean μ_0.

The greater the precision of one, the higher it's weight is. Both the prior and the sample mean convey some information about μ_n. These two are combined (linearly), but more weight is given to the signal that has higher precision (smaller variance).

The weight given to the sample mean increases with n, while the weight given to the prior mean does not. As a consequence, when the sample size becomes large, more and more weight is given to the sample mean. In the limit, all weight is given to the information coming from the sample and no weight is given to the prior.





(iv) <u>Jeffreys' prior(μ and σ unknown):</u>

$$g(\mu,\sigma) \propto \frac{1}{\sigma^c} \qquad ; c > 0$$

Then the joint posterior distribution of (μ,σ) is –

$$\prod\left(\mu,\sigma \middle| \underset{\sim}{x}\right) = K\frac{1}{\sigma^{n+c}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \quad ....(1)$$

$$= K\frac{1}{\sigma^{n+c}}e^{-\frac{1}{2\sigma^2}\left\{A+n(\bar{x}-\mu)^2\right\}}$$

$$\therefore K\int_{0}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sigma^{n+c}}e^{-\frac{1}{2\sigma^2}\{A+n(\bar{x}-\mu)^2\}}d\mu\, d\sigma = 1$$

$$K\int_0^\infty \frac{1}{\sigma^{n+c}}e^{-\frac{A}{2\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)^2}d\mu\, d\sigma = 1$$

$$K\int_0^\infty \frac{e^{-\frac{A}{2\sigma^2}}}{\sigma^{n+c}}\frac{\sqrt{(2\pi)}\sigma}{\sqrt{n}}d\sigma = 1$$

$$K\frac{\sqrt{2\pi}}{\sqrt{n}}\int_0^\infty \frac{e^{-\frac{A}{2\sigma^2}}}{\sigma^{n+c-1}}d\sigma = 1$$

$$K\sqrt{\frac{\pi}{2n}}\left\lceil\frac{n+c-2}{2}\left(\frac{2}{A}\right)^{\frac{n+c-2}{2}} = 1\right.$$

$$K^{-1} = \sqrt{\frac{\pi}{2n}}\left\lceil\frac{n+c-2}{2}\left(\frac{2}{A}\right)^{\frac{n+c-2}{2}}\right.$$

<u>Marginal posterior of μ:</u>

∴Restoring K in (1) and obtaining the marginal distribution of μ we have-

$$\prod\left(\mu\middle|\underbrace{x}\right) = \sqrt{\frac{2n}{\pi}}\left(\frac{A}{2}\right)^{\frac{n+c-2}{2}}\frac{1}{\left\lceil\frac{n+c-2}{2}\right.}\frac{1}{2}\int_0^\infty \frac{e^{-\frac{1}{2\sigma^2}\left\{A+n(\bar{x}-\mu)^2\right\}}}{\sigma^{2\frac{n+c-1}{2}+1}}$$

$$= \sqrt{\frac{2n}{\pi}} \left(\frac{A}{2}\right)^{\frac{n+c-2}{2}} \frac{\Gamma\frac{n+c-1}{2}}{\Gamma\frac{n+c-2}{2}} \left(\frac{1}{2}\right) \{\frac{2}{A+n\left(\bar{x}-\mu\right)^2}\}^{\frac{n+c-1}{2}}$$

$$= \sqrt{\frac{2n}{\pi}} \left(\frac{A}{2}\right)^{\frac{n+c-2}{2}} \frac{\Gamma\frac{n+c-1}{2}}{\Gamma\frac{n+c-2}{2}} \left(\frac{1}{2}\right) \left\{\frac{2}{1+\frac{n\left(\bar{x}-\mu\right)^2}{A}}\right\}^{\frac{n+c-1}{2}} \frac{1}{A}^{\frac{n+c-1}{2}}$$

$$= \sqrt{\frac{n}{\pi}} \left(\frac{A}{2}\right)^{\frac{n+c-2}{2}} \frac{\Gamma\frac{n+c-1}{2}}{\Gamma\frac{n+c-2}{2}} \left\{\frac{1}{1+\frac{n\left(\bar{x}-\mu\right)^2}{A}}\right\}^{\frac{n+c-1}{2}} \frac{1}{A}^{\frac{1}{2}}$$

$$= \sqrt{\frac{n}{A}} \left(\frac{A}{2}\right)^{\frac{n+c-2}{2}} \frac{\Gamma\frac{n+c-1}{2}}{\Gamma\frac{1}{2}\Gamma\frac{n+c-2}{2}} \left\{\frac{1}{1+\frac{n\left(\bar{x}-\mu\right)^2}{A}}\right\}^{\frac{n+c-1}{2}}$$

$$= \sqrt{\frac{n}{A}} \frac{1}{B(\frac{1}{2}, \frac{n+c-2}{2})} \left\{\frac{1}{1+\frac{n\left(\bar{x}-\mu\right)^2}{A}}\right\}^{\frac{n+c-1}{2}}$$

Now,

$$\mu^* = E\left(\mu \middle| \underbrace{x}\right) = \sqrt{\frac{n}{A}} \frac{1}{B(\frac{1}{2}, \frac{n+c-2}{2})} \int_{-\infty}^{\infty} \frac{\mu}{(1+\frac{n(\bar{x}-\mu)^2}{A})^{\frac{n+c-1}{2}}} d\mu$$

Let $\frac{\sqrt{n}(\bar{x}-\mu)}{\sqrt{A}} = \frac{t}{\sqrt{n+c-2}}$

$$\therefore d\mu = -\sqrt{\frac{A}{n(n+c-2)}} dt$$

$$\therefore \mu^* = \frac{1}{\sqrt{n+c-2}} \frac{1}{B\left(\frac{1}{2}, \frac{n+c-2}{2}\right)} \int_{-\infty}^{\infty} \frac{\bar{x} - \sqrt{\frac{A}{n(n+c-2)}} t}{\left(1+\frac{t^2}{n+c-2}\right)^{\frac{(n+c-2)+1}{2}}}$$

$$= \frac{1}{\sqrt{n+c-2}} \frac{1}{B(\frac{1}{2}, \frac{n+c-2}{2})} \bar{x} \int_{-\infty}^{\infty} \frac{dt}{(1+\frac{t^2}{n+c-2})^{\frac{n+c-2+1}{2}}} - \sqrt{\frac{A}{n(n+c-2)}} E(t)$$

$$= \bar{x} - \sqrt{\frac{A}{n(n+c-2)}} E(t) = \bar{x}$$ [t is distributed as student's t with (n+c-2) degrees

of freedom.]

(3.3.4) Result: *Bayes estimator of* μ *is same as MLE of* μ *(independent of c).*

Marginal Posterior of σ^2:

$$\prod\left(\sigma^2\middle|\underbrace{x}\right)=\int_{-\infty}^{\infty}\prod\left(\mu,\sigma\middle|\underbrace{x}\right)d\mu=K\sqrt{\frac{\pi}{2n}}\frac{e^{-\frac{A}{2\sigma^2}}}{\sigma^{2\frac{n+c}{2}}}=\frac{\frac{A}{2}^{\frac{n+c-2}{2}}}{\left\lceil\frac{n+c-2}{2}\right.}\frac{e^{-\left(\frac{A}{2\sigma^2}\right)}}{\sigma^{2\frac{n+c}{2}}}$$

$$...0<\sigma<\infty$$

$$\therefore \sigma^{*2}=E\left(\sigma^2\middle|\underbrace{x}\right)=\frac{\frac{A}{2}^{\frac{n+c-2}{2}}}{\left\lceil\frac{n+c-2}{2}\right.}\int_0^{\infty}\frac{\sigma^2e^{-\frac{A}{2\sigma^2}}}{\sigma^{2\frac{n+c}{2}}}d\sigma^2=\frac{\frac{A}{2}^{\frac{n+c-2}{2}}}{\left\lceil\frac{n+c-2}{2}\right.}\int_0^{\infty}\frac{1}{\sigma^2}^{\frac{n+c-2}{2}}e^{-\frac{A}{2\sigma^2}}d\sigma^2$$

$$=\frac{\frac{A}{2}^{\frac{n+c-2}{2}}}{\left\lceil\frac{n+c-2}{2}\right.}\int_0^{\infty}\frac{1}{\sigma^{2\frac{n+c-4}{2}+1}}e^{-\frac{A}{2\sigma^2}}d\sigma^2$$

$$=\frac{\frac{A}{2}^{\frac{n+c-2}{2}}}{\left\lceil\frac{n+c-2}{2}\right.}\frac{\left\lceil\frac{n+c-4}{2}\right.}{\frac{A}{2}^{\frac{n+c-4}{2}}}=\frac{A}{2}\frac{\left(\frac{n+c-4}{2}-1\right)!}{\left(\frac{n+c-2}{2}-1\right)!}=\frac{A}{2}\frac{2}{n+c-4}=\frac{A}{n+c-4}$$

$$=\frac{1}{n+c-4}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

(3.3.5) Result: *The variance estimator would be same as the MLE in case c=4. Also if c=3 the above is same as the Uniformly Minimum Variance Unbiased Estimator of* σ^2 *.*

6. Concluding Remarks:

The present dissertation explores Maximum Likelihood and Bayesian estimation of parameters in Poisson, Binomial and Normal distribution under Natural

Conjugate Prior and Jeffreys' Invariant Prior and demonstrates that the Bayes estimator performs better than the ML estimator under certain cases.

As the sample size increases both the approaches tend to yield a similar kind of estimator of the parameter of interest.

In certain situations, Bayesian estimators outperform ML estimates as the former approach takes into consideration the pre-sample or prior information based on a person's own experience and judgement.

The Bayesian estimators although biased (in some cases) are consistent.

Given the sample and the corresponding likelihood function, Bayes' estimator is unique unlike MLE(s) sometimes obtained.

When the underlying distribution depends on, say r number of parameters $\underbrace{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_k, \theta_{k+1}, \ldots, \theta_r\right)$ and we are interested only in a subset

$\underbrace{\propto} = (\theta_1, \theta_2, \ldots, \theta_k)$, estimation of $\underbrace{\propto}$ may be based upon the marginal posterior of $\underbrace{\propto}$

obtained by integrating out $\underbrace{\beta} = \left(\theta_{k+1}, \ldots, \theta_r\right)$ from the posterior of $\underbrace{\theta}$.

\

7.Limitations:

The most important limitation is the need for a prior probability distribution over the model parameters. Any domain information that can help in choosing the prior can have a big influence on the accuracy of the posterior. Estimation comes from the different resource (dataset), may add noise to model. When no reliable information is present, typically a uniform prior or the uninformed Jeffreys' prior is used.

A second limitation is that the posterior distribution often does not have a neat, closed form. Only when a conjugate prior exists for the distribution we avoid this problem (e.g. a normal-inverse chi square distribution for a normally-distributed variable). When such a prior does not exist, techniques such as sampling or Laplace estimators are often used to obtain a closed-form distribution close to the posterior

8.Scope for further study:

We have only considered 3 specific theoretical probability distributions and takes into account two approaches to obtain prior distributions. There is further scope for various other approaches to prior distributions which may yield a better explanation for the purpose of comparison in the above discussed cases.

Moreover, the Bayesian approach can be implemented over realistic situations in our daily life. Here, we have compared Bayesian estimators and Maximum Likelihood Estimators over varying sample size, which can be further extended over various other factors and studied.

9.Acknowledgement:

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to my supervisor Dr. Surabhi Dasgupta for her expert guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & professors of Statistics Department of St. Xavier's College for their kind co-operation and encouragement which helped me in completion of this project.

My thanks and appreciations also go to my friends in developing the project and people who have willingly helped me out with their abilities.

10.Bibliography:

Books:

[B1] Bayesian Estimation S.K. Sinha

[B2] The Bayesian Choice Christian P. Robert

Articles:

[A1] https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss

[A2] https://www.statlect.com/fundamentals-of-statistics/normal-distribution-Bayesian-estimation

[A3] http://www.ams.sunysb.edu/~zhu/ams570/Bayesian_Normal

[A4] https://www.statlect.com/fundamentals-of statistics/Hierarchical-Bayesian-models