Bayesian Statistics

Subjective & objective notions of probability –

Subjective probability is a measure of an individual's personal belief or degree of confidence about the likelihood of an event occuring. It is based on an individual's own subjective assessment of the situation, which can be influenced by personal experiences, biases & other subjective factors. Subjective prob. is often used in situations where there is little or no empirical evidence available or where the individual's personal experience or intution is considered more reliable than empirical evidence.

Objective prob. on the other hand, is a measure of the likelihood of an event based on emperical evidence or data. It is calculated using mathematical or statistical method based on observed frequenties or patterns of events. Objective prob. is often used in situations where emperical evidence is available, or where the accuracy and reliability of the prob. calculation is of primary importance.

→ For e.g. the prob. of heads occuring when tossing a coin logically must be 0.5. This can be proved by tossing the coin many times & observing the result.

→ In business decision, the prob. are often estimated based on managerial judgement. Prob. established in this way are known as subjective prob. because no two individuals will necessarily assign the same prob. to a particular outcome.

→ Subjective prob. are also known as uncertainty.

Bayes Theorem -

$$P(A_i|B) = \frac{P(B|A_i)\cdot P(A_i)}{P(A_1 \cap B) + P(A_2 \cap B) \cdots + P(A_n \cap B)}$$

$$= \frac{P(B|A_i)\cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i)\cdot P(A_i)}$$

$$P(\theta|x) = \frac{f(x|\theta)\cdot g(\theta)}{h(x)}$$

θ is fixed but in bayesian it is treated as random.

Decision Theory Vs Game Theory -

Decision Theory studies individual decision-making in situations in which an individual's choice neither affects nor is affected by individuals' choices; one-decision maker another-nature

while game theory studies decision-making in situations where individuals' choice do affect each other.

Classical def - $P(B) = \frac{m}{n} = \frac{\text{favourable no of outcomes}}{\text{Total no. of mutually exhaustive event}}$

Statistician def^n - $\lim_{n\to\infty} \frac{m}{n} \Rightarrow c$
$\downarrow$
$P(A)$

Utility. Utility is something that we get in loss

overestimates case is more serious than underestimated

Loss - $|\theta - \hat{\theta}|$ or $(\theta - \hat{\theta})^2$ is random

statistician task is to minimizes the average of loss.

Utility for different person matter different means

for e.g. 5 rupee for me does not means much but 200 rupee means a lot for me. & for rich person it does not matter

Statistical Games -

In statistical inference decisions about pop^n such as mean & variance of some characteristic are based on sample data. Statistical inference can therefore be regarded as a game bet^n nature which controls the relevent features of the pop^n & statistician, who is trying to make a decision about the pop^n

In a statistical game the statistician does not know the nature strategy but may have some information about the same through some sample ~~base~~ pop^n.

e.g. - A statistician is the told that a coin has either a head on one side & a tail on the other or it has two heads. The statistician can not inspect the coin but can observe a single toss of the coin & see whether it shows a head or tail. The statistician must then decide whether or not coin is two headed.

The state of nature $\begin{cases} \theta_1 - \text{The coin is of two head} \\ \theta_2 - \text{The coin is balanced (one side head \& one side tail)} \end{cases}$

Statistician's Decisions - Let. a_1 = The coin is of two head

a_2 = The coin is balanced

Statistician

		a_1	a_2
Nature's strategy	θ_1	0 $L(a_1, \theta_1)$	1 $L(a_2, \theta_1)$
	θ_2	1 $L(a_1, \theta_2)$	0 $L(a_2, \theta_2)$

→ Loss of function

After tossing the coin the statistician wants to use the information $X = \begin{cases} 0, & \text{head} \\ 1, & \text{tail} \end{cases}$ for taking the action a_1 & a_2 and needs a decision function sitting out the action to take in each case one possible decision function would be the decision function $d_1(X)$ ~~other~~ & d_2, d_3, & d_4.

Statistician decision $d_1(X) = \begin{cases} a_1, & X=0 \\ a_2, & X=1 \end{cases} \Rightarrow \begin{array}{l} d_1(0) = a_1 \\ d_1(1) = a_2 \end{array}$

$$d_2(X) = \begin{cases} a_1, & X=0 \\ a_1, & X=1 \end{cases} \Rightarrow \begin{array}{l} d_2(0) = a_1 \\ d_2(1) = a_1 \end{array}$$

$$d_3(X) = \begin{cases} a_2, & X=0 \\ a_1, & X=1 \end{cases} \Rightarrow \begin{array}{l} d_3(0) = a_2 \\ d_3(1) = a_1 \end{array}$$

$$d_4(X) = \begin{cases} a_2, & X=0 \\ a_2, & X=1 \end{cases} \Rightarrow \begin{array}{l} d_4(0) = a_2 \\ d_4(1) = a_2 \end{array}$$

These are the only possible function & some of these may not be very sensible in practice

Now, we consider our first decision f^n i.e., d_1 & find out the resulting expected loss.

$$R(d_1, \theta_j) = E(L(d_1(x), \theta_j))$$

	a_1	a_2
θ_1	$L(a_1, \theta_1)$	$L(a_2, \theta_1)$
θ_2	$L(a_1, \theta_2)$	$L(a_2, \theta_2)$

where $L(d_1(x), \theta_j)$ is the loss when nature's strategy & nature choice is θ_j & decision d_1 is taken by the statistician.

R is called the risk function which gives us the value of the ~~risk~~ loss expected from particular function & state of nature. The expectation is taken with respect to the r.v. X &

Under θ_1: $P(X=0)=1$

Since coin is of two head $P(X=1)=0$

Under θ_2: $P(X=0)=1/2$

coin is balanced $P(X=1)=1/2$

$$R(d_1,\theta_1)= 1\times L(d_1(0),\theta_1)+0\times L(d_1(1),\theta_1)$$
$$= L(a_1,\theta_1)+0$$
$$=0$$

$$R(d_1,\theta_2)=\frac{1}{2}\times L(d_1(0),\theta_2)+\frac{1}{2}\times L(d_1(1),\theta_2)$$
$$=\frac{1}{2}\times L(a_1,\theta_2)+\frac{1}{2}\times L(a_2,\theta_2)$$
$$=\frac{1}{2}\times 1+\frac{1}{2}\times 0 = 1/2$$

Again $R(d_2,\theta_1)= 1\times L(d_2(0),\theta_1)+0\times L(d_2(1),\theta_1)$
$$= L(a_1,\theta_1)$$
$$=0$$

& $R(d_2,\theta_2)=\frac{1}{2}\times L(d_2(0),\theta_2)+\frac{1}{2}\times L(d_2(1),\theta_2)$
$$=\frac{1}{2}\times L(a_1,\theta_2)+\frac{1}{2}L(a_1,\theta_2)$$
$$=\frac{1}{2}+\frac{1}{2}=1$$

Again $R(d_3,\theta_1)= 1\times L(d_3(0),\theta_1)+0\times L(d_3(1),\theta_1)$
$$= 1\times L(a_2,\theta_1)+0$$
$$=1$$

& $R(d_3,\theta_2)=\frac{1}{2}\times L(d_3(0),\theta_2)+\frac{1}{2}\times L(d_3(1),\theta_2)$
$$=\frac{1}{2}L(a_2,\theta_2)+\frac{1}{2}L(a_1,\theta_2)=\frac{1}{2}\times 0+\frac{1}{2}\times 1=\frac{1}{2}$$

$$R(d_4,\theta_1) = 1\times L(d_4(0),\theta_1) + 0\times L(d_4(1),\theta_1)$$
$$= L(a_2,\theta_1)+0$$
$$= 1$$

$$R(d_4,\theta_2) = \frac{1}{2}\times L(d_4(0),\theta_2) + \frac{1}{2}\times L(d_4(1),\theta_2)$$
$$= \frac{1}{2}\times L(a_2,\theta_2) + \frac{1}{2}\times L(a_2,\theta_2)$$
$$= 0+0=0$$

$p=2/3$ $(1-p)=1/3$

	d_1 p	d_2	d_3	d_4 $(1-p)$	maximum loss
θ_1	0	0	1	1	1
θ_2	1/2	1	1/2	0	1/2 (min)
		$\times$	$\times$		

d_2 loss > d_1 loss
discard d_2

d_3 loss > d_4 loss
discard d_3

Selection criterian-
1. Minimax criteria - Minimize the expected loss in θ_1 & θ_2
θ 1/2 has minimum loss in Maximum loss
2. Bayes criteria - First assign a probability of p to strategy d_1 and a probability of $1-p$ to strategy d_4. We can calculate expected loss when nature strategy is θ_1,

$$E(R,\theta_1) = 0\times P + (1-P)\times 1$$
$$= 1-P$$

when nature strategy is θ_2,

$$E(R,\theta_2) = \frac{1}{2}\times P + 0\times(1-P) = \frac{1}{2}P$$

Now, we want to equal the loss ~~them~~ we don't want to maximize the loss:

$$E(R,\theta_1) = E(R,\theta_2)$$
$$1-P = \frac{1}{2}P$$
$$\Rightarrow P+\frac{P}{2} = 1$$
$$\Rightarrow P = \frac{2}{3}$$

So, the randomized strategy that minimizes the maximum expected loss is to choose d_1 two third of the time & d_4 1/3 time.

Value of the Game - (Expected Risk) is 33p. This is the value of both of the expected loss function when p=2/...

Decision Criteria - In general it is possible to find the best decision f^n only in respect of some criteria the two important criteria which have been considered here, are

1- Minimax Criteria
2- Baye's

Minimax - Under the minimax criteria under the decision f^n d is chosen is that for which $R(d,\theta)$, maximize the θ, w.r.t. θ is the minimum.

Baye's - If Θ is regarded as a random variable under the Baye's criterian the decision f^n choosen is that for which $E[R(d,\Theta)]$ is a minimum where the expectation is taken with respect to θ. The criterian needs θ to be regarded as a random variable with a given dist^n.

Applying Bayes' creterion requires probabilities to be attached to nature's two strategies θ_1 & θ_2. If $P(\theta_1)=p$ & $P(\theta_2)=1-p$ then the Bayes risk for d_1 is $0.p+\frac{1}{2}.(1-p) = \frac{1}{2}(1-p)$
& d_4 is $1.p+0.(1-p) = p$

Thus $p>\frac{1}{3}$ the Bayes risk of d_1 is less than the Bayes risk of d_4 & d_1 is preferred to d_4.

When $p<\frac{1}{3}$ the Bayes risk of d_4 is less than the Bayes risk of d_1 & d_4 is preferred to d_1.

When $p=\frac{1}{3}$ the two Bayes risks are equal & either d_1 or d_4 can be chosen.

Q.1- A statistician is observing when use from Bino$(2, p)$ distribution he know $p = 1/4$ or $1/2$ & he is trying to choose b/w these value he observe a single value x from the distribution & proposes to use one of the following four decision f^n.

$x = 0, 1, 2,$ $\quad x \sim B(2, p)$

The nature strategy :-

$\theta_1 = 1/4 \quad , \quad \theta_2 = 1/2$

$$\therefore d_1(x) = \begin{cases} p = \frac{1}{4} , & x = 0 \\ p = \frac{1}{2} , & x = 1 \text{ or } 2 \end{cases}$$

$$d_2(x) = \begin{cases} p = \frac{1}{4} , & x = 0 \text{ or } 1 \\ p = \frac{1}{2} , & x = 2 \end{cases}$$

$$d_3(x) = \begin{cases} P = \frac{1}{4} , & x = 0 \text{ or } 1 \text{ or } 2 \\ \cancel{P = \frac{1}{2} , x =} \end{cases}$$

$$d_4(x) = \{ P = \frac{1}{2} , \quad x = 0, \text{ or } 1, \text{ or } 2$$

If we incorrectly concludes that $P = \frac{1}{4}$ he suffers a loss of 1 if he incorrectly concludes that $P = \frac{1}{2}$, he suffers a loss of 2. Find the risk function for the each decision f^n & find the decision function a/c to minimax & baye's ~~cis~~ criteria. If the statistician has the prior feels that it is

Statistician

		$1/4 = a_1$	$1/2 = a_2$
Nature	$\theta_1 = 1/4$	0	2
	$\theta_2 = 1/2$	1	0

if $p = 1/4$ $\quad B(2, 1/4)$

$P(x=0) = {}^nC_x p^x q^{n-x} = {}^2C_0 p^0 q^2 = \left(\frac{3}{4}\right)^2 = \frac{9}{16}$

if $(P = 1 \text{ or } 2) = 1 - \frac{9}{16} = \frac{7}{16}$

$= L(d_1(0), \theta_1) \times P(X=0) + L(d_1(1 \text{ or } 2), \theta_1) \times P(X = 1 \text{ or } 2)$

$R[d_1, \frac{1}{4}] = 0 \times P(X=0) + 2 \times P(X = 1 \text{ or } 2)$

$= 2 \times 7/16 = 7/8$

$P(X=0) = {}^2C_0 p^x q^{n-x} = \left(1 - \frac{1}{2}\right)^2 = \frac{1}{4}$

$R\left[d_1, \frac{1}{2}\right] = L(d_1(0), \theta_2) \times P(X=0) + L(d_1(1 \text{ or } 2), \theta_2) \times P(x = 1 \text{ or } 2)$

$= 1 \times \frac{1}{4} + 0 \times \frac{3}{4}$

$= \frac{1}{4}$

d_2	$\frac{1}{8}$	$3/4$
d_3	0	1
$d_4 =$	2	0

$R\left[d_2, \frac{1}{4}\right] = L(d_2(0 \text{ or } 1), \theta_1) \times P(X = 0 \text{ or } 1) + L(d_2(2), \theta_1)\, P(X=2)$

$= \frac{15}{16} \times 0 + \frac{1}{16} \times 2 = \frac{1}{8}$

$R\left[d_2, \frac{1}{2}\right] = L(d_2(0 \text{ or } 1), \theta_2) \times P(X = 0 \text{ or } 1) + L(d_2(2), \theta_1) \cdot P(X=2)$

$= \cancel{1 \times \frac{15}{16} \frac{1}{4} + \frac{1}{16} \times 0 \times \frac{3}{4}}$

$= 3/4 \times 1 + \frac{1}{4} \times 0$

$= 3/4$

At θ_1 $\quad B(2, \frac{1}{4})$

$P(X=2) = {}^2C_2 p^2 q^0 = \left(\frac{1}{4}\right)^2 = \frac{1}{16}$ $\qquad P(X = 1 \text{ or } 0) = 1 - \frac{1}{16} = \frac{15}{16}$

$p = \frac{1}{4}$

At θ_2 $\quad B(2, \frac{1}{2})$

$P(X=2) = {}^nC_2 p^2 q^0 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$ $\qquad P(X = 1 \text{ or } 0) = 1 - \frac{1}{4} = \frac{3}{4}$

$P(X = 0 \text{ or } 1 \text{ or } 2)$

$R(d_3, \tfrac{1}{4}) = L(d_3, \tfrac{1}{4}) \times P(x = 0 \text{ or } 1 \text{ or } 2)$
$= 0 \times 1$
$= 0$

At θ_1, $p = 1/4$

$P(x = 0 \text{ or } 1 \text{ or } 2) = \cancel{P(x=0)}\ 1$

$R(d_3, \tfrac{1}{4}) = L(d_3, \theta_2) \times P(x = 0 \text{ or } 1 \text{ or } 2)$
$= 1 \times 1 = 1$

$R(d_4, \tfrac{1}{2}) = L(d_4, \theta_1) \times P(X = 0 \text{ or } 1 \text{ or } 2)$
$= 2 \times 1$
$= 2$

$R(d_4, \tfrac{1}{4}) = L(d_4, \theta_2) \times P(X = 0 \text{ or } 1 \text{ or } 2)$
$= 0 \times 1$
$= 0$

Minimax Criteria

Now, we have to compare decision rule

	d_1	d_2	d_3	d_4	Prob.	
$p = 1/4 = \theta_1$	$7/8$	$1/8$	0	2	$1/2$	} equally likely
$p = 1/2 : \theta_2$	$1/4$	$3/4$	1	0	$1/2$	

min (max↑ max↑ max↑ max↑) = min $(7/8, 3/4, 1, 2)$

$= 3/4 = d_2$

Baye's Criteria

$E[R(d_1, \theta)] = \frac{7}{8} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{9}{16}$

$E[R(d_2, \theta)] = \frac{1}{8} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2} = \frac{7}{16}$ (min)

$E[R(d_3, \theta)] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$

$E[R(d_4, \theta)] = 2 \times \frac{1}{2} + 0 = 1$

Q.

	0.25 θ_1	0.15 θ_2	0.60 θ_3
D_1	23	34	16
D_2	30	19	18
D_3	23	27	20
D_4	32	19	19 ×

→ Discrad D_4 because if we compare it with D_2, D_4 is always greater than D_2.

→ $\min(34, 30, 27, 32) = 27 = D_3$

→ $E[R(d_1, \theta)] = 23 \times 0.25 + 34 \times 0.15 + 0.6 \times 16$
$= 5.75 + 5.10 + 9.60$
$= 20.45$ (min)

$E[R(d_2, \theta)] = 30 \times 0.25 + 19 \times 0.15 + 0.6 \times 18$
$= 7.50 + 2.85 + 10.80$
$= 31.15$

$E[R(d_3, \theta)] = 23 \times 0.25 + 0.15 \times 27 + 0.6 \times 20$
$= 5.75 + 4.05 + 12$
$= 21.80$

$E[R(d_4, \theta)] = 32 \times 0.25 + 19 \times 0.15 + 19 \times 0.60$

$$L(\theta) = \prod_{i=1}^{n} (x_i|.\theta)$$

$$p(\theta|x) = \frac{L(\theta) \times g(\theta)}{\int L(\theta) \times g(\theta)\, d\theta}$$

Q. If $L(\lambda) \sim Pois(\lambda)$ & $g(\lambda) = Exp(\lambda')$

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

$$p(\theta|\underline{x}) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \cdot \lambda' e^{-\lambda'\lambda}}{\prod_{i=1}^{n} x_i! \int_0^\infty \frac{e^{-n\lambda} \lambda^{\sum x_i} \lambda' e^{-\lambda'\lambda}}{\prod_{i=1}^{n} x_i!} d\lambda}$$

$$= \frac{e^{-n\lambda} \lambda^{\sum x_i} \lambda' e^{-\lambda'\lambda}}{\int_0^\infty \lambda' \lambda^{\sum x_i} e^{-(n+\lambda')\lambda} d\lambda}$$

$$= \frac{e^{-(n+\lambda')\lambda} \lambda^{\sum x_i}}{\int_0^\infty \lambda^{\sum x_i + 1 - 1} e^{-(n+\lambda')\lambda} d\lambda}$$

$$p(\lambda|\underline{x}) = \frac{e^{-(n+\lambda')\lambda} \lambda^{\sum x_i}}{\frac{\Gamma(\sum x_i + 1)}{(n+\lambda')^{(\sum x_i + 1)}}}$$

$$\left\{ \int_0^\infty e^{-\beta x} x^{\alpha-1} dx = \frac{\Gamma\alpha}{\beta^\alpha} \right.$$

2 - $X \sim Bin(m, p)$; $p \sim Beta(\alpha, \beta)$

Find posterior distribution.

18/05/20..

Likelihood is the occurrence of $f(x_i, \theta)$. It is function of θ. $L(\theta|x)$

Joint density function is the function of $x_1, x_2, \ldots x_n$

$f(x_1, x_2, \ldots x_n, \theta)$, $\prod_{i=1}^{n} f(x_i, \theta)$.

<u>Bayes</u> – $$p(\theta|\underset{\sim}{x}) = \frac{L(\theta) \cdot g(\theta)}{\int L(\theta) \cdot g(\theta) d\theta} \begin{matrix} \rightarrow \text{joint} \\ \rightarrow \text{marginal} \end{matrix}$$

Proper prior – integration of proper prior is 1.

Improper prior – integration of improper prior is not 1.

Prior & Posterior Distribution –

Let $x_1, x_2, \ldots x_n$ be a random sample from a popn specified by the density $f(x;\theta)$ & it is required to estimate θ. In Baye's paradigne The parameter θ is considered as a random variable therefore it will have a prob. distribution. This allows the use of any knowledge available about possible values for θ before the collection of any data.

This knowledge is quantified by expressing it as a prior distribution of θ.

Then after collecting appropriate data the posterior distn of θ is determine & this founds the bases of all inferences conserning θ. The information from the random sample contain in the likelihood sample for that sample so the Bayesian approach combine the information obtain from the likelihood fn with the information in the prior distribution & hence obtained posterior estimate for the requiring parameter.

if $g(\Theta)$ is prior & $L(\Theta)$ is likelihood then

posterior $$p(\theta|\underset{\sim}{x}) = \frac{L(\Theta)\cdot g(\theta)}{\int L(\Theta)\cdot g(\theta)d\theta}$$

$$p(\theta|\underset{\sim}{x}) \propto L(\theta)\cdot g(\theta)$$

Q.1

Shortest life times-

mean life - 500 hours $\sim exp(1/500)$

longer-

mean life - 2500 hours $\sim exp(1/2500)$

5 bulbs continuous lightening 300 h if life is exponential dist^n. If ~~ten~~ 5 bulbs are alive after 300 h the long life bulbs prob.?

P (Long Life bulb$_{B_2}$ | 5 bulb are alive after 300 h$_{A}$)

$$P(A|B_1) = \left[\int_{300}^{\infty} \frac{1}{500} e^{-\frac{1}{500}x} dx\right]^5 = \left(\frac{1}{500}\right)^5\left[\frac{-1}{\frac{1}{500}} e^{-\frac{1}{500}x}\right]_{300}^{\infty}{}^{5} = \left[e^{-3/5}\right]^5 = e^{-3}$$

$$P(A|B_2) = \left[\int_{300}^{\infty} \frac{1}{2500} e^{-\frac{1}{2500}x} dx\right]^5$$

$$= \left[\frac{-1}{2500} \times 2500 \left[e^{-\frac{1}{2500}x}\right]_{300}^{x}\right]^5$$

$$= e^{-\frac{300\times5}{2500}}$$

$$= e^{-3/5}$$

$$= e^{-0.6}$$

$$P(B_2|A) = \frac{P(A|B_2)\,P(B_2)}{P(A|B_1)P(B_1) + P(A|B_2)\cdot P(A|B_2)} \qquad P(B_1) = 1/2 = P(B_2)$$

$$= \frac{\frac{1}{2}\times e^{-0.6}}{\frac{1}{2}e^{-0.6} + \frac{1}{2}e^{-3}}$$

$$= \frac{e^{-0.6}}{e^{-0.6} + e^{-3}}$$

Axioms – 3 –

If there is consistency & coherency then ~~there~~ if there is subjective prob. $P(A)$ then it converges to $1-P(A) = P(A^c)$ but not equal to $[1-P(A)] \cdot P(A^c)$.

Gamma

$$G(\alpha, \lambda) = \frac{\lambda^{\alpha}}{\sqrt{\alpha}} e^{-\lambda x} x^{\alpha-1} \quad := \quad \frac{\lambda^{\alpha}}{\Gamma\alpha} \int_0^{\infty} e^{-\lambda x} x^{\alpha-1} dx = 1$$

$$\Rightarrow \int_0^{\infty} e^{-\lambda x} x^{\alpha-1} dx = \frac{\Gamma\alpha}{\lambda^{\alpha}}$$

$$G(\alpha, 1) = \frac{1}{\Gamma\alpha} e^{-x} x^{\alpha-1} \qquad G(1, \lambda) = \lambda e^{-\lambda x} \cdot x^0$$

$$= \lambda e^{-\lambda x}$$

Baye's Risk – $E_X[f(x)] = \int_{R_X} f(x) \cdot g(x)\, dx$

$p(\theta|x)$

$\therefore$ Baye's Risk $\cdot$ $\int_{R_\theta} L(\theta) \cdot p(\theta|x)\, d\theta$

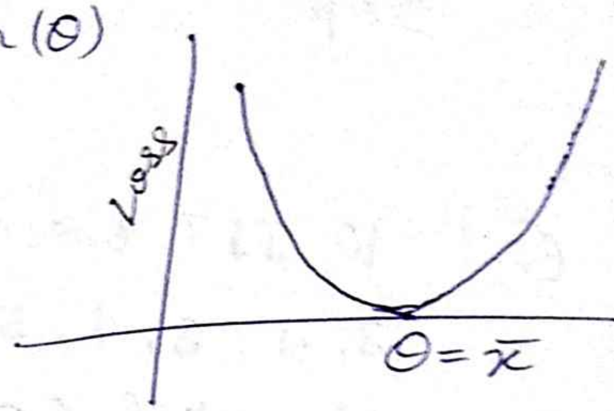
Loss Function – To obtain an estimator of unknown parameter θ a loss function must be specified. This ~~is~~ the major of the loss incurred when $g(X)$ is used as a estimator θ.

When $g(X) = \theta$ the loss is zero and it is positive & does not decrease. as $g(X)$ get further away from θ. The commonly used loss functions are –

1 – quadratic error loss function (square error loss function) –

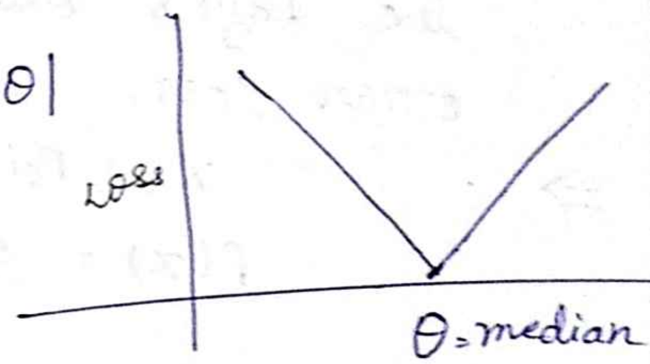
$$L[g(\underline{x}), \theta] = [g(x) - \theta]^2$$

The square error loss function is minimum when $g(x)$ is equal to the parameter (θ)



2- Absulute error loss function –

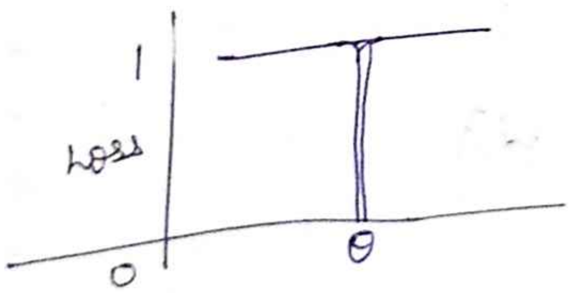
$$L[g(\underline{x}), \theta] = |g(\underline{x}) - \theta|$$



3- Zero – One loss function –
All or Nothing loss f^n -

$$L[g(\underline{x}, \theta)] = \begin{cases} 0 , & g(x) = \theta \\ 1 , & g(x) \neq \theta \end{cases}$$

{If $g(x)$ is not equal to θ then there is always loss



The Baye's estimator that arises by minimizing the expected loss for ~~eq~~ each of these loss f^n in turn is the mean, median & mode respectively of the posterior distribution. The expected posterior loss:—

Expected value of ~~expe~~ Loss w.r.t. θ

$$E_{\Theta}[L(g(\underline{x}), \theta)] = \int L g(x, \theta) \cdot p(\theta|x) d\theta$$

~~Excep~~

Q:1 10 IIT observation from a λ-distribution 3, 4, 3, 1, 5, 5, 2, 3, 3, 2. Assumming an exponential (0.2) prior distribution for λ find the baye's estimator for λ under square error loss.

$\Rightarrow$

$$X \sim Poi(\lambda)$$

$$f(x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \quad , x = 0, 1, 2, \dots \quad \lambda > 0$$

$$L(\lambda) = \frac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\pi x_i!}$$

$$\lambda \sim exp(\lambda')$$

$$g(\lambda) = \lambda' e^{-\lambda' \lambda} \quad ; \lambda > 0$$

$$p(\lambda|\underline{x}) = \frac{\frac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\pi x_i!} \cdot \lambda' e^{-\lambda' \lambda}}{\int_0^{\infty} \frac{e^{-n\lambda} \lambda^{\Sigma x_i}}{\pi x_i!} \cdot \lambda' e^{-\lambda' \lambda} d\lambda}$$

$$= \frac{e^{-(n+\lambda')\lambda} \lambda^{\Sigma x_i}}{\int_0^{\infty} e^{-(n+\lambda')\lambda} \lambda^{\Sigma x_i + 1 - 1} d\lambda}$$

$$\left\{ \text{by Gamma } f^n \quad \int_0^{\infty} e^{-\beta x} x^{\alpha - 1} dx = \frac{\Gamma\alpha}{\beta^{\alpha}} \right.$$

$$= \frac{e^{-(n+\lambda')\lambda} \lambda^{\Sigma x_i}}{\frac{\Gamma(\Sigma x_i + 1)}{(n+\lambda')^{(\Sigma x_i + 1)}}}$$

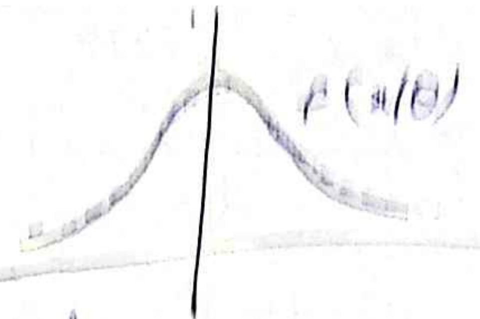
$\lambda' = 0.2. \quad n = 10$

$\Sigma x_i = 3 + 4 + 3 + 1 + 5 + 5 + 2 + 3 + 3 + 2$
$= 31$

$$\therefore \; p(\lambda \mid \underline{x})' = \frac{e^{-(10+0.2)\lambda} \cdot \lambda^{31}}{\frac{\Gamma 31+1}{(10+0.2)^{32}}}$$

$$= \frac{\lambda^{31} \cdot e^{-10.2\lambda} \cdot (10.2)^{32}}{\Gamma 32}$$

Predictive distribution -



Suppose we are given a random sample $x_1, x_2, \dots x_n$ from $f(x|\theta)$ & suppose $z_1, z_2 \dots z_n$ be the predictive observation from same distribution $f(z|\theta)$. Then previous sample is called informative sample & latter is called future or predictive data.

We have to find out $p(z|x)$ means we are given that previous data & want to find future data.

$$h(\theta|x) = \frac{f(x|\theta)\cdot g(\theta)}{\int f(x|\theta)\cdot g(\theta)\, d\theta} \quad \begin{matrix} \text{— Joint density} \\ \text{— Marginal} \end{matrix}$$

$$\therefore\ p(z|x) = \frac{l(z,x)}{f_1(x)} = \frac{p(z,x,\theta)}{p(x)}$$

$$\Rightarrow\ p(z|x) = \frac{p(z,x|\theta)\cdot g(\theta)}{\int f(x|\theta)\, g(\theta)\, d\theta} \qquad \left\{ \begin{matrix} p - \text{predictive} \\ h - \text{posterior} \end{matrix} \right.$$

$$= \frac{f(z|\theta)\cdot f(x|\theta)\cdot g(\theta)}{\int f(x|\theta)\cdot g(\theta)\, d\theta}$$

$$= \int f(z|\theta)\cdot h(\theta|x)\, d\theta$$

$$= \underset{\theta|x}{E}\left(f(z|\theta)\right) \qquad \left\{ \begin{matrix} \because \int \theta\, f(\theta|x)\, d\theta \\ = \underset{\theta|x}{E}(\theta) \end{matrix} \right.$$

= Posterior mean

Here $p(z|x)$ is the predictive informati distribution which gives the information about future data in the light of given data.

Q. $x_1, x_2, \cdots x_n$ are iid's $\sim B(\theta)$ Bernaulli
$\theta \sim Beta(\alpha, \lambda)$ also given posterior dist^n $\theta/x \sim Beta(\Sigma x_i + \alpha, n - \Sigma x_i + \lambda)$
find predictive dist^n

$x_1, x_2, \ldots x_n \sim B(\theta)$

$$B(\theta) = \theta^{x_i}(1-\theta)^{1-x_i} \quad \text{—— (1)}$$

$$\theta \sim B(\alpha, \lambda)$$

$$B(\alpha, \lambda) = \frac{1}{B(\alpha, \lambda)} \cdot \theta^{\alpha-1}(1-\theta)^{\lambda-1}$$

Note

$X \sim B(a, b)$

$$f(x) = \frac{1}{B(a,b)} x^{(a-1)}(1-x)^{b}$$

$0 < x < 1$
$a, b > 0$

$x = \theta$
$a = \alpha$
$b = \lambda$

Joint Distribution,

$$P(\theta, D) = P(\theta) \cdot P(D|\theta)$$

& $P(\theta|D) \propto \underset{\text{Prior}}{P(\theta)} \cdot \underset{\text{Posterior}}{P(D|\theta)}$

$$P(\theta|D) = \frac{1}{B(\alpha,\lambda)} \theta^{\alpha-1}(1-\theta)^{\lambda-1} \cdot \prod_{i=0}^{n-1}\left(\theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$$P(\theta|x) = \frac{1}{B(\alpha,\lambda)} \theta^{\alpha-1} (1-\theta)^{\lambda-1} \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i}$$

$$= \frac{1}{B(\alpha,\lambda)} \theta^{(\alpha+\Sigma x_i - 1)} \cdot (1-\theta)^{(\alpha+n-\Sigma x_i - 1)}$$

$$\approx \frac{1}{B(\alpha+\Sigma x_i, \alpha+n-\Sigma x_i)} \cdot \theta^{(\alpha+\Sigma x_i-1)} \cdot (1-\theta)^{(\alpha+n-\Sigma x_i-1)}$$

Posterior dist^n $= B(\alpha+\Sigma x_i, \alpha+n-\Sigma x_i)$

Now,

Predictive distribution $p(z|x) = \int_0^1 f(z|\theta) \cdot p(\theta|x)\, d\theta$

$$p(z|x) = \int_0^1 \theta^{z} \cdot (1-\theta)^{1-z} \cdot \frac{1}{B(\alpha+\Sigma x_i, \alpha+n-\Sigma x_i)} \cdot \theta^{(\alpha+\Sigma x_i-1)} \cdot (1-\theta)^{(\alpha+n-\Sigma x_i)} d\theta$$

$$= \frac{1}{B(\alpha+\Sigma x_i, \alpha+n-\Sigma x_i)} \int_0^1 \theta^{(\alpha+z+\Sigma x_i-1)} \cdot (1-\theta)^{\alpha+n-z-\Sigma x_i+1-1} d\theta$$

$$= \frac{B(\Sigma x_i + z + \alpha, n+\lambda-\Sigma x_i - z + 1)}{B(\Sigma x_i + \alpha, n + \lambda - \Sigma x_i)}$$

Bayesian Interval Estimation

The estimator of θ is $\hat{\theta} = E(x_1, \dots x_n)$ is one which minimizes the posterior expected loss. So, we need to work out for the probability that θ lies in the interval $[\theta_1, \theta_2]$ where $\theta_1 < \theta_2$ this interval which is based on the posterior distribution θ / x is called a credible interval.

$$1-\alpha = P[\theta_1 < \theta < \theta_2]$$

$$1-\alpha = \int_{\theta_1}^{\theta_2} p(\theta|x)\, d\theta \quad \text{——} \; (1)$$

We can have infinite such solutions which satisfies eqn (1) so we have to make some strategy to choose the best among them.

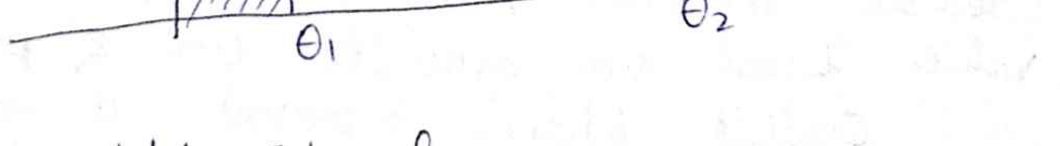
1- Equal tail credible Interval-

An equal tail $(1-\alpha)$ credible Interval is given by

$$1-\alpha = \int_{\theta_1}^{\theta_2} p(\theta|x)\, d\theta$$

$$\frac{\alpha}{2} = \int_{-\infty}^{\theta_1} p(\theta|x)\, d\theta = \int_{\theta_2}^{\infty} p(\theta|x)\, d\theta$$

=



2) Shortest credible Interval-

To obtain Shortest $(1-\alpha)$ credible Interval one has to minimize $I = \theta_2 - \theta_1$ such that condition (1) is satisfies which requires $p(\theta_2|x) = p(\theta_1|x)$ —— (2)

The interval $[\theta_1, \theta_2]$ which simultaneously satisfies (1) & (2) is called shortest $(1-\alpha)$ credible interval.

3→ Highest Posterior Density Interval (H.P.D Interval) –

An interval I which satisfies following condition simultaneously.

a) The interval is shortest.

b) $p(\theta|x)$ s.t. $\theta \in I > p(\theta|x)$ s.t. $\theta \notin I$

i.e., the posterior density inside at each point of interval is greater than the posterior density at every point outside the interval this ofcourse implies that the interval includes more probable values of θ and excludes the lesser ones.

Note If the posterior density is unimodel (not necessary symmetric) the shortest credible & hpd intervals are same.

Credible Interval vs Confidence Interval –

Credible interval are a concept used in statistical inference and Bayesian Statistics to estimate the uncertaintity associated with a parameter or a set of parameters. Unlike frequentist confidence intervals which provide a range of ~~parameters. Unti~~ pausible values for a parameter based on repeated sampling, credible interval provide a range of pausible values based on available data & prior knowledge.

Credible intervals represent a range of values within which a true values for the parameter is believed to lie with a certain level of confidence based on available data.

In frequentist terms, the parameter is fixed and the confidence interval is random. A credible interval is simply an interval in the domain of the posterior distribution within which an unobserved parameter value falls with a particular probability