UNIT - 1

PAGE NO.:
DATE: / /

14/7/25

DECISION THEORY

* **Fundamental elements of Decision Theory :-**

- **Action :-** Alternative courses of actions (strategies) that are available to the decision maker (statistician)

- **States of nature (or events) :-**

- These are the uncertain circumstances of external factors, beyond the control of decision-maker, that can influence the outcome of the action taken by a decision-maker.

- **Outcomes :-** These are the consequences that result from containing a specific action with a particular state of nature. *

- **Payoff (loss) :-** This is a numerical value (monetary gain, loss or utility) associated with each outcome.

- **Objective :-** Statisticians focus on minimizing loss, on the other hand, decision theorists in other fields might aim to maximize gains (utility)

* **Decision Rules :-**

These are the procedures or criteria used to select the best action from the available alternatives, considering the possible outcomes and their probabilities.

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

- **Framework for making decisions:**

- Given a situation with multiple alternatives courses of action, each potentially leading to various outcomes with certain probabilities, which course of action should the decision-maker choose?

Ans :- Statistical Inference

$\mathcal{A}$ = action space = $\{a_1, a_2, \ldots, a_m\}$ (finite countable)

Θ = parameter space = $\{\theta_1, \theta_2, \ldots, \theta_n\}$ (finite countable)

Loss function = $L(\theta, a)$ = Loss to statistician when he chooses action $a \in \mathcal{A}$ & nature chooses $\theta \in \Theta$

* **Travel Problem:-**

Action space = $\{a_1$ = no insurance, a_2 = insurance$\}$
Parameter space = $\{\theta_1$ = disease, θ_2 = no disease$\}$

Outcomes → Exposed to disease (Prob. = 0.03)
→ Not Exposed to disease (Prob. = 0.97)

Consequences → if exposed to disease, then treatment cost = \$ 1000

Take Insurance → Buying cost → \$ 50.

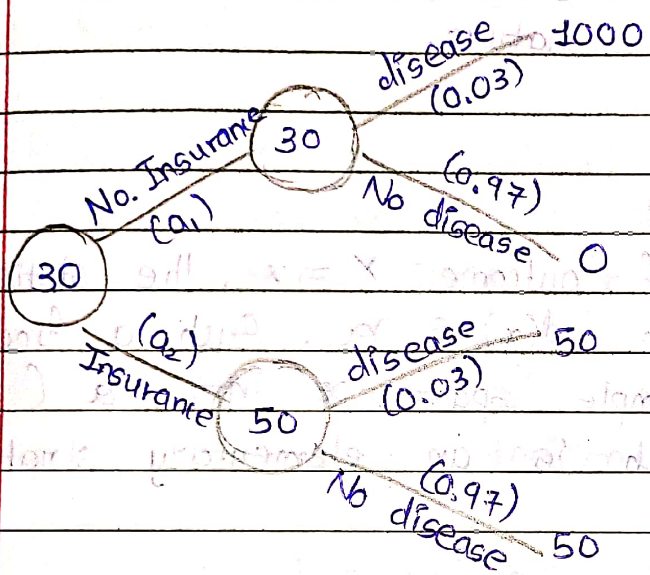
Doe not take Insurance → \$ 0.

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

$L(\theta, a)$

	Actions a_1 = do not take insurance	a_2 = take Insurance
disease θ_1	1000	50
no disease θ_2	0	50



min(30, 50) = 30

$\Rightarrow$ Do not take Insurance

Now, suppose there is a medical test available for knowing the chances of getting the disease.

Test Result → Positive (Prone to disease)
→ Negative (Not prone to disease)

(Used for Extra information)

Random Variable X : $\begin{cases} 1 & \text{if test is positive} \\ 0 & \text{o.w.} \end{cases}$

$P(X=1 / \theta_1) = 0.9 \Rightarrow P(X=0 / \theta_1) = 0.1$

$P(X=0 / \theta_2) = 0.77 \Rightarrow P(X=1 / \theta_2) = 0.23$

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

$X = 1$ if test is positive
$X = 0$ if test is negative

How to incorporate this information about r.v. X in the decision problem ?

When a r.v. X is involved in the Experiment whose prob. $dist^n$ P_θ depends on the state $\theta \in \Theta$ choosen by nature.

* Decision Rule :-

- On the basis of outcome $X = x$, the statistician choosen an action $d(x) \in \mathcal{A}$. Such a function d, which maps sample space $\mathcal{X}$ into $\mathcal{A}$ (i.e., $d : \mathcal{X} \to \mathcal{A}$) which is an elementary strategy for the statistician.

Decision Table

$\mathcal{X}$ \ D	Decision Rules			
	d_1	d_2	d_3	d_4
$x_1 = 0$	a_1	a_1	a_2	a_2
$x_2 = 1$	a_1	a_2	a_1	a_2

D = set of decision rules $= \{d_1, d_2, \ldots, d_{m^n}\}$

No. of decision rules $= \boxed{m^n}$ where
$\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$
$\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

Loss $L(\theta, d)$

$\mathfrak{X}$	$P_\theta(x)$		d_1		d_2		d_3		d_4	
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
$x_1 = 0$	0.1	0.77	1000	0	000	0	50	50	50	50
$x_2 = 1$	0.9	0.23	1000	0	50	50	1000	0	50	50
	1	1	$E_{\theta_1}(d_1(x))$	$E_{\theta_2}(d_1(x))$						
	Expected loss		1000	0	145	11.5	905	38.5	50	50

Risk Function $= R(\theta, d) = E_\theta[L(\theta; d(x))]$
= Expected value of loss when θ is the true state of nature.
= Average loss to the statistician when the true state of nature is θ & Statistician uses the decision rule d.

Risk Table $R(\theta, d)$

$\theta \backslash D$	d_1	d_2	d_3	d_4
$0.03 = \theta_1$	1000	145	905	50
$0.97 = \theta_2$	0	11.5	38.5	50

* **Non Randomized Decision Rules:-**

- Any function $d(x)$ that maps the sample space $\mathfrak{X}$ into $\mathfrak{a}$ is called a non-randomized decision rule (or function) provided the risk function $R(\theta, d)$ exists and finite $\forall\ \theta \in \Theta$.
- The class of all non-randomized decision rules is denoted by D.

PAGE NO.:
DATE: / /

Expected Overall Loss.

Under d_1 = $1000 \times 0.03 + 0 \times 0.97 = 30$
Under d_2 = $145 \times 0.03 + 11.5 \times 0.97 = \boxed{15.50}$
Under d_3 = $905 \times 0.03 + 38.5 \times 0.97 = 64.495 \simeq 64.5$
Under d_4 = $50 \times 0.03 + 50 \times 0.97 = 50$

$\Rightarrow$ d_2 is the $\boxed{\text{best}}$ decision rule.

17/7/25

$\mathcal{a} = \{a_1, a_2, \ldots, a_m\} \rightarrow$ Statistics
$\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\} \rightarrow$ nature (Parameters)
$\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$
$L(\theta, a)$ = loss to statistician when he chooses action $a \in \mathcal{a}$ & nature chooses $\theta \in \Theta$

d : decision rule : $d : \mathcal{X} \rightarrow \mathcal{a}$ $d(x) = a$
D : non randomized decision rule : $D = (d_1, d_2, \ldots, d_m)$

Decision problem : triplet $(\Theta, \mathcal{a}, L)$
Loss function = $L(\theta, d) = L(\theta, d(x))$

* Difference between Game theory Problem and Decision :-

(1) In a two-person game, both players are simultaneously trying to max the gain / min. loss; whereas in decision-theory problem nature chooses his strategies without having objective of maximization of gain / minimization of loss.

PAGE NO.:
DATE: / /

(2) In decision theory, it is assumed that nature chooses the 'true state' once and for all and statistician has possibility of gathering information on this choice by performing a suitable Experiment

payoff table
(Loss Table)
for nature (P-1)

	$L(\theta, a)$	
	a_1	a_2 (statistician)
(nature) θ_1	4	1
θ_2	-3	0

* Risk function = $R(\theta, d) = E_\theta[L(\theta, \overset{d(x)}{d})]$
$= \int_x L(\theta, d(x))$ or $\sum_x L(\theta, d(x)) \cdot f(x, \theta)$

$$E[g(x)] = \int_x g(x) \cdot f(x, \theta)\,dx \text{ or } \sum_x g(x)(f)_p(x;\theta)$$

Ex:- Consider a game in which nature (Player-1) and Statistician (player-2) to show up Head (H) or Tail (T) on a coin available to each one.

$\mathcal{A} = \{H = a_1, T = a_2\}$, $\Theta = \{H = \theta_1, T = \theta_2\}$

Rule :- The price for showing H is 1 and for T is 2. Nature wins if sum is odd and Statistician wins if sum is even

PAGE NO.:
DATE: / /

Loss table for Statistician $(L(\theta, a))$

$\Theta \backslash \mathcal{A}$	a_1 (H)	a_2 (T)
θ_1 (H)	-2	3
θ_2 (T)	3	-4

Statistician perform a random experiment by asking a group of people ~~to~~ about what they will show up and actually what they showed up (H or T)

Define X : answer given by an individual.

$P(\text{truth}) = \frac{3}{4}$, $P(\text{lie}) = \frac{1}{4}$.

X :		$H = x_1$	$T = x_2$
If $\theta = \theta_1 (H)$	$P_{\theta_1}(X = x)$	$3/4$	$1/4$
If $\theta = \theta_2 (T)$	$P_{\theta_2}(X = x)$	$1/4$	$3/4$

$\mathcal{X} = \{ x_1 = H , x_2 = T \} \Rightarrow n = 2$

$\mathcal{A} = \{ a_1 = H, a_2 = T \} \Rightarrow m = 2$

No. of decision rules $= m^n = 2^2 = 4$

$\Rightarrow$ Set of non-randomized decision rules D

$= \{ d_1 , d_2 , d_3 , d_4 \}$

Decision Table

$\mathcal{X} \backslash D$	d_1	d_2	d_3	d_4
$x_1 = H$	a_1	a_1	a_2	a_2
$x_2 = T$	a_1	a_2	a_1	a_2

PAGE NO.:
DATE: / /

$L(\theta, d)$:

x	$P_\theta(x)$ θ_1	$P_\theta(x)$ θ_2	d_1 θ_1	d_1 θ_2	d_2 θ_1	d_2 θ_2	d_3 θ_1	d_3 θ_2	d_4 θ_1	d_4 θ_2
$x_1 = H$	$3/4$	$1/4$	-2	3	-2	3	3	-4	3	-4
$x_2 = T$	$1/4$	$3/4$	-2	3	3	-4	-2	3	3	-4
	1	1	-2	3	$-3/4$	$-9/4$	$7/4$	$5/4$	3	-4

$R(\theta, d) \rightarrow$

Risk Table $R(\theta, d)$

$\theta \backslash d$	d_1	d_2	d_3	d_4
θ_1	-2	$-3/4$	$7/4$	3
θ_2	3	$-9/4$	$5/4$	-4

21/7/25

* 3 important categories of 'classical' mathematical statistics, in terms of decision theory problem :-

(i) Suppose $\mathcal{a} = \{a_1, a_2\}$
Decision theory problem in this case can be looked upon as a problem of 'testing of hypothesis'.

Let $\Theta = \mathbb{R}$ and loss function is

$$L(\theta, a_1) = \begin{cases} l_1 & \text{if } \theta > \theta_0 \\ 0 & \text{if } \theta \le \theta_0 \end{cases}$$

&

$$L(\theta, a_2) = \begin{cases} 0 & \text{if } \theta > \theta_0 \\ l_2 & \text{if } \theta \le \theta_0 \end{cases}$$

where $l_1, l_2 > 0$

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

Hence, statistician would like to take actions a_1 if $\theta \le \theta_0$ and a_2 if $\theta > \theta_0$

The set D of non-randomized decision rules consists of those functions 'd' from $\mathcal{X} \to \mathcal{A}$ with the property that $P_\theta\{d(X) = a_1\}$ is well-defined for all values of $\theta \in \Theta$.

In this case, the risk function will be:

$$R(\theta, d) = \begin{cases} l_1 P_\theta\{d(X) = a_1\} & \text{if } \theta > \theta_0 \\ l_2 P_\theta\{d(X) = a_2\} & \text{if } \theta \le \theta_0 \end{cases}$$

Thus, for $\theta > \theta_0$, $P_\theta\{d(X) = a_1) =$ Prob. of making an error in taking action a_1 when statistician should actually take action a_2 and θ is the true state of nature.

and

for $\theta > \theta_0$, $P_\theta\{d(X) = a_2\} =$ Prob of making an error in taking action a_2 when statistician should actually take action a_1 and θ is the true state of nature.

Suppose $H_0: \theta > \theta_0$ vs $H_1: \theta \le \theta_0$ | a_1 = Reject H_0, a_2 = Accept H_0

$$P_{H_0}\{d(X) = a_1\} = P[\text{Rej } H_0 / H_0 \text{ is true}] = P[\text{Type-I error}] = \alpha$$

&

$$P_{H_1}\{d(X) = a_2\} = P[\text{Accept } H_0 / H_1 \text{ true}] = P[\text{Type-II error}] = \beta$$

(ii) When $\mathcal{a} = \{a_1, \ldots, a_k\}$; $k > 2$.
In this case, the decision theory problem is called Multiple decision problem.
e.g. double sampling plan, SPRT, etc

(iii) When $\mathcal{a} = \mathbb{R}$
Such decision theory problems can be referred to as 'Point Estimation' of a real parameter.

e.g. Let us take $\Theta = \mathbb{R}$ and

$L(\theta, a) = c \cdot (\theta - a)^2$, $c > 0$ (squared error loss funct^n)

(other Loss function are :-
$L(\theta, a) = c \cdot |\theta - a|$ called absolute error loss funct^n.)

Then a decision funct^n can be defined as a real-valued funct^n defined on $\mathcal{X}$, which may be considered as an 'estimate' of the true state of nature θ.

So, statistician would like to choose funct^n d which minimizes the risk funct^n.

$$R(\theta, d) = c \cdot E_\theta[\theta - d(x)]^2$$
$$= c \cdot [\text{Mean squared error of estimate } d(x)]$$

* **Randomized Decision Rules** :-

Consider a decision problem $(\Theta, \mathcal{a}, L)$ with $\Theta = \{\theta_1, \theta_2\}$
$\mathcal{a} = \{a_1, a_2, a_3\}$

$\mathcal{a} = \{a_1, \ldots, a_m\}$, Θ,
$L(\theta, a)$
Random Expt $\to$ r.v X
$d : \mathcal{X} \to \mathcal{a}$
$\mathcal{X} = \{x_1, \ldots, x_n\}$

PAGE NO.:
DATE: / /

Loss table $L(\theta, a)$

$\theta \backslash a$	$1/2$	$1/2$	0
	a_1	a_2	a_3
θ_1	4	1	3
θ_2	1	4	3

$D = \{d_1, \ldots, d_{m^n}\}$
= set of non-randomized decision rule.

$a^* = \{P, P_j \geq 0 \ \& \ \sum_{j=1}^{m} P_j = 1\}$

- A randomized decision for the statistician in problem $(\Theta, \mathcal{A}, L)$ is a prob dist^n over $\mathcal{A}$.

- If P is a prob. dist^n over $\mathcal{A}$ & Z is a r.v. taking values in $\mathcal{A}$, whose prob. dist^n is given by P, then the Expected / avg. loss in using the randomized decision P is

$$L(\theta, P) = E[L(\theta, Z)] \qquad (1)$$

provided it exists

The space of randomized decisions P for which $L(\theta, P)$ exists and finite $\forall \ \theta \in \Theta$ is denoted by $\mathcal{A}^*$.

Ex: Let $P_1 = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$ & let δ_1 be the randomized decision rule corresponding to P_1, then the expected loss in using δ_1 will be :

$$\rightarrow L(\theta_1, P_1) = \sum_{j=1}^{m=3} l(\theta_1, a_j) \cdot P_j$$

$$= 4 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} + 3(0) = \frac{5}{2} = 2.5$$

&

Scanned with OKEN Scanner

$$L(\theta_2, P_1) = \sum_{j=1}^{m=3} l(\theta_2, a_j) \cdot P_j = 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} + 3(0) = \frac{5}{2}$$
$$= 2.5$$

Let δ_2 be another randomized decision rule which chooses a_1 with prob. $\frac{1}{4}$, a_2 with prob. $\frac{1}{2}$ & a_3 with prob. $\frac{1}{4}$ (ie. $P_2 = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$)

$$L(\theta_1, P_2) = 4 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = \frac{9}{4} = 2.25$$

$$L(\theta_2, P_2) = 1 \cdot \frac{1}{4} + 4 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = \frac{12}{4} = \boxed{3}$$

δ_3 with $P_3 = \left(\frac{4}{6}, \frac{1}{6}, \frac{1}{6}\right)$

δ_4 with $P_4 = \left(\frac{3}{8}, \frac{5}{8}, 0\right)$

(Pure Strategy $\rightarrow$ $P = (1,0,0)$ or $(0,1,0)$ or $(0,0,1)$
Mixed Strategy $\rightarrow$)

– Thus, the decision problem $(\Theta, \mathcal{A}^*, L)$ can be considered as a problem $(\Theta, \mathcal{A}, L)$ in which the statistician is allowed randomization on his actions.

– By analogy, we may extend this decision problem (Θ, D, R) to (Θ, D^*, R) where D^* is a space / set containing prob. distⁿ over D.

$D^* = \{ \delta = (p_1, p_2, \ldots, p_{m^n}) ;\ p_j \geq 0, \sum_{j=1}^{m} p_j = 1 \}$

= set of randomized decision rule.

* **Randomized Action :-**

- A prob. distⁿ P defined over $\mathcal{A}$ is called randomized action provided

$$L(\theta, P) = E[L(\theta, z)],$$

where z is a r.v. taking values provides $\mathcal{A}$ with prob. distⁿ P, exists & finite.

Thus, $L(\theta, P) = \sum_{j=1}^{m} L(\theta, a_j) \cdot P_j$

* **Randomized Decision Rule :-**

- Any prob. distⁿ δ on the space of non-randomized decision functⁿ D is called a 'randomized decision function / rule' provided the risk functⁿ $R(\theta, \delta) = E[R(\theta, z)]$ exists & finite $\forall\ \theta \in \Theta$ where z is a r.v. taking values in D with prob. distⁿ δ.

- The space / set of all randomized decision rules is denoted by δ^*.

Ex:- $\mathcal{X} = \{x_1, x_2\}$, $\mathcal{A} = \{a_1, a_2\}$, $\Theta = \{\theta_1, \theta_2\}$

$\Downarrow$ $n = 2$ $\quad$ $\Downarrow$ $m = 2$

$\Rightarrow$ No. of decision rules $= m^n = 2^2 = 4$

$\mathcal{X}$ \ D	
x_1	
x_2	

Θ \ D	
θ_1	
θ_2	

PAGE NO.:
DATE: / /

$L(\theta, a)$

Θ \ a	a_1	a_2
θ_1	1	2
θ_2	2	1

Decision rules

x \ D	d_1	d_2	d_3	d_4
θ_1	a_1	a_1	a_2	a_2
θ_2	a_1	a_2	a_1	a_2

x \ D	$P_\theta(x)$		d_1		d_2		d_3		d_4	
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
x_1	1/3	1/4	1	2	1	2	2	1	2	1
x_2	2/3	3/4	1	2	2	1	1	2	2	1
	1	1	1	2	5/3	5/4	4/3	5/4	2	1

Risk Table $R(\theta, d)$

Θ \ D	$\delta_1 \to \frac{1}{2}$ d_1	$1/10$ d_2	$3/10$ d_3	$1/10$ d_4
θ_1	1	5/3	4/3	2
θ_2	2	5/4	5/4	1

let $\delta_1 = \left(\frac{1}{2}, \frac{1}{10}, \frac{3}{10}, \frac{1}{10}\right)$

Then $R(\theta_1, \delta_1) = E[R(\theta_1, z)]$

$$= 1 \cdot \frac{1}{2} + \frac{5}{3} \cdot \frac{1}{10} + \frac{4}{3} \cdot \frac{3}{10} + 2 \cdot \frac{1}{10}$$

$$= \frac{15 + 5 + 12 + 6}{30} = \frac{38}{30} = \boxed{\frac{19}{15}} = 1.27$$

&

$$R(\theta_2, \delta_2) = E[R(\theta_2, z)]$$

$$= 2 \cdot \frac{1}{2} + \frac{5}{4} \cdot \frac{1}{10} + \frac{7}{4} \cdot \frac{3}{10} + 1 \cdot \frac{1}{10}$$

$$= \boxed{7/4}$$

PAGE NO. :
DATE : / /

24/7/25

* Behavioral Decision Rule / Function :-

- The term 'behavioral Strategy' refers to those strategies that tell the player how to randomize at each move.

- In decision theory, a behavioural decision rule tells the statistician.

- How to randomize after observing the outcome of the experiment whereas a randomized decision rule chooses at random a decision funct^n that tells him before observing the outcome of the experiment that Exactly what action to take as a result of the Experiment.

Ex:- $\mathcal{A} = \{a_1, a_2\}$, $\mathcal{X} = \{x_1, x_2\}$, $\Theta = \{\theta_1, \theta_2\}$

$\Downarrow$ $m = 2$ $\quad\quad$ $\Downarrow$ $n = 2$

$D = \{d_1, d_2, d_3, d_4\}$

$\delta_1 \to$ / D / $\mathcal{X}$	$P_1 = \frac{1}{2}$	$P_2 = \frac{3}{10}$	$P_3 = \frac{1}{10}$	$P_4 = \frac{1}{10}$
	d_1	d_2	d_3	d_4
x_1	a_1	a_1	a_2	a_2
x_2	a_1	a_2	a_1	a_2

$$D^* = \{\delta = (P_1, P_2, \ldots, P_{m^n}) ; P_j \geq 0, \sum_{j=1}^{m} P_j = 1\}$$

$$\mathcal{A}^* = \{P \text{ on } \mathcal{A} ; P_j \geq 0, \sum_{j=1}^{m} P_j = 1\}$$

Then from above table,

$$d_1(x_1) = a_1 \quad , \quad d_1(x_2) = a_1$$
$$d_2(x_1) = a_1 \quad , \quad d_2(x_2) = a_2$$
$$d_3(x_1) = a_2 \quad , \quad d_3(x_2) = a_1$$
$$d_4(x_1) = a_2 \quad , \quad d_4(x_2) = a_2$$

$\Rightarrow$ with $\delta_1 = \left(\frac{1}{2}, \frac{3}{10}, \frac{1}{10}, \frac{1}{10}\right)$

Then define π_1 = Prob. of choosing action a_1 when $X = x_1$

$= P_{x_1}(a_1)$

$\Rightarrow 1 - \pi_1 = P_{x_2}(a_2)$

& π_2 = Prob. of choosing action a_1 when $X = x_2$

$= P_{x_2}(a_1)$

$1 - \pi_2 = P_{x_2}(a_2)$

Then $\hat{\delta}_1 = (\pi_1, \pi_2)$: $0 \le \pi_1, \pi_2 \le 1$

(Sum of P_j's $= 1$ but Sum of π_1 & $\pi_2 \neq 1$)

$= \left(\pi_1 = \frac{4}{5}, \pi_2 = \frac{3}{5}\right)$

Under δ_1,

$\pi_1 = P_{x_1}(a_1) = \frac{1}{2} + \frac{3}{10} = P_1 + P_2$

$= \frac{8}{10} = \boxed{\frac{4}{5}}$

& $\pi_2 = P_{x_2}(a_1) = \frac{1}{2} + \frac{1}{10} = \frac{6}{10} = \boxed{\frac{3}{5}}$

Def^n:- Behavioral decision rule :-

$\rightarrow$ A funct^n $\hat{\delta}(x)$, from $\mathcal{X}$ into $\mathcal{A}^*$ is called a Behavioral decision rule / funct^n provided

PAGE NO.:
DATE: / /

$$\hat{R} = (\theta, \delta) = E_{\theta}[L(\theta, \delta(x))]$$

exists and finite, and set of all behavioral decision rules is denoted by $\mathcal{D}$.

* **Optimal (Best) Decision Rules :-**

- Given a decision problem $(\Theta, \mathcal{A}, L)$ and r.v. X whose dist^n is $P_\theta(x)$ / $f(x; \theta)$; what decision decision rule δ should the statistician use ?

Method-1 :- Restricting the available rules OR

Reduction of size of randomized decision D^* according to some criteria :-

(i) Unbiasedness :- We know that, an estimate $\hat{\theta}$ of a parameter θ is said to be unbiased if

$E_\theta[\hat{\theta}] = \theta \quad \forall \ \theta$	$T(x) \rightarrow$ estimator of θ $\Rightarrow \hat{\theta} = T(x)$
- A decision rule is said to be unbiased if	If $E_\theta[T(x)] = \theta$ $\rightarrow T(x) = \hat{\theta}$ is UE.

$$E_\theta[L(\theta, \delta(x))] \leq E_\theta[L(\theta', \delta(x))] \; ; \quad \theta \neq \theta'$$

(ii) Invariance :- To be discussed in Unit-3.

'a priori'
'a posteriori'
A : item is defective

$P(A) = \frac{\text{No. of defectives}}{\text{Total items}}$

PAGE NO.: 'a priori'
DATE:

Machines $\to$ M_1 40%, M_2 35%, M_3 25%

B_1 : item produced by M_1
B_3 : - - - M_3

$P(A/B_1) =$, $P(A/B_2) =$, ...

$P(B_1/A) = \frac{P(B_1)P(A/B_1)}{P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \cdots}$
(a Posteriori)
(Bayes Theory)

28/7/25

Method - 2 : Ordering the decision rules

(i) The Bayes Principle :-

Prior Distribution :-

- The prob. distⁿ defined over Ⓗ is called prior distⁿ. Suppose, prior distⁿ is denoted by 'τ'.

Define $r(\tau, \delta) = E[R(T, \delta)]$, where T is a r.v. over Ⓗ having prob. distⁿ τ.

$r(\tau, \delta)$ = Bayes risk of a decision rule δ w.r.t. prior distⁿ τ.

Defⁿ: Baye's Rule :-

- A decision rule δ_0 is said to be Bayes decision rule w.r. to the prior distⁿ $\tau \in$ Ⓗ* if

$$r(\tau, \delta_0) = \underset{\delta \in D^*}{\inf}\, r(\tau, \delta) \quad \text{--- (1)}$$

"min. Bayes risk" w.r. to prior distⁿ τ

Note :- It is possible that min. Bayes risk w.r. to τ exists but ∄ a decision rule δ_0 whose Bayes risk w.r. to τ is equal to min. Bayes risk. In this case, Bayes rule may not exist. Then Statistician may have to be satisfied with a rule whose Bayes risk is closer to the min Bayes risk.

Scanned with OKEN Scanner

PAGE NO. :
DATE : / /

Def^n:- **ε - Bayes Rule :-**

- Let $\varepsilon > 0$. A decision rule δ_0 is said to be ε - Bayes w.r. to prior dist^n $\tau \in Ⓗ^*$ if

$$r(\tau, \delta_0) \leq \inf_{\delta \in D^*} r(\tau, \delta) + \varepsilon \quad \text{——— (2)}$$

(ii) The minimax Principle :-

- According to this principle, the statistician would arrange decision rules in order of the max risk involved with them.

Def^n:- A decision rule δ_0 is said to be minimax if

$$\sup_{\theta \in Ⓗ} R(\theta, \delta_0) = \inf_{\delta \in D^*} \sup_{\theta \in Ⓗ} R(\theta, \delta)$$

minimax value / upper value of game

Notes :-

(1) Decision rule δ_0 is minimax if & only if

$$R(\theta', \delta_0) \leq \sup_{\theta \in Ⓗ} R(\theta, \delta) \quad \forall \ \theta' \in Ⓗ \text{ and } \delta \in D^*$$

(2) Even if minimax value is finite, there may not be a minimax decision rule, so the statistician may have to be satisfied with a rule whose max. risk is within ε of the minimax value.

Scanned with OKEN Scanner

Defⁿ:- Let $\varepsilon > 0$. A decision rule δ_0 is said to be ε-minimax iff

$$\sup_{\theta \in \Theta} R(\theta, \delta_0) \leq \inf_{\delta \in D^*} \sup_{\theta \in \Theta} R(\theta, \delta) + \varepsilon$$

That is, $R(\theta', \delta_0) \leq \sup_{\theta \in \Theta} R(\theta, \delta) + \varepsilon \quad \forall\, \theta' \in \Theta \;\&\; \delta \in D^*$

Defⁿ:- Least Favourable Prior distribution :-

- A prior distⁿ $\tau_0 \in \Theta^*$ is said to be the least favourable if

$$\inf_{\delta \in D^*} r(\tau_0, \delta) = \underbrace{\sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta)}_{\text{maximin value / lower value of game.}}$$

* Geometric Interpretation of a Decision Problem

- Given set A. Let P_1 & P_2 be the two points in A. Then $\alpha P_1 + (1-\alpha) P_2$ is called Convex Combination of P_1 & P_2.

- Suppose $P_1, \ldots, P_K$ are K points in A, then their convex combination will be

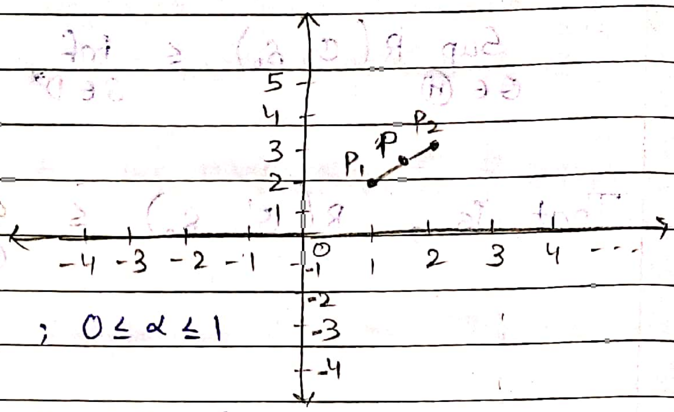
$$P = \alpha_1 P_1 + \alpha_2 P_2 + \cdots + \alpha_K P_K$$

where $\alpha_i \geq 0$ and $\sum_{i=1}^{K} \alpha_i = 1$.

PAGE NO.:
DATE: / /

— Set A is said to be a Convex set, if $P \in A$ whenever $P_1, P_2, \dots, P_k \in A$ (convex set)

<u>e.g.</u> :- $A = R^2$



$P_1 = (x_1, y_1) = (1, 2)$
$P_2 = (x_2, y_2) = (2, 3)$

$P = \alpha P_1 + (1-\alpha) P_2 \quad ; \quad 0 \le \alpha \le 1$

$\alpha = 0.4$

$\Rightarrow P = (0.4)(1,2) + (0.6)(2,3)$

$\Rightarrow P = (1.6, 2.6)$

31/7/25

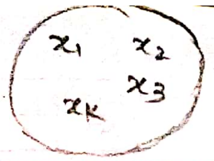
* Geometric Interpretation of Decision Problem

<u>Def^n</u> :- <u>Convex Hull</u> :-

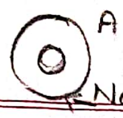
If $\{x_1, x_2, \dots, x_k\}$ is a ~~set~~ sequence of points in R^K and $0 \le \alpha_i \le 1$ are number s.t. $\sum_{i=1}^{K} \alpha_i = 1$ then $\sum_{i=1}^{K} \alpha_i x_i$ is called a Convex Combination.

→ The convex hull of a set Ω is the set of all points which are convex combinations of points in Ω.

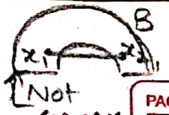
→ Convex Hull is a set with no holds in its interior and no [illegible] on its exterior.



$\Omega \longrightarrow$ Convex set



Not convex



Not convex

PAGE NO. :
DATE : / /

Also Convex hull of a set Ω is defined as the smallest convex set containing Ω i.e. it is intersection of all convex sets containing Ω.

* Consider a decision problem with follow Loss Table :-

P	p_1	p_2	- - - - -	p_m
$\theta \backslash a$	a_1	a_2	- - - - -	a_m
θ_1	$l(\theta_1, a_1)$	$l(\theta_1, a_2)$	- - - -	$l(\theta_1, a_m)$
θ_2	$l(\theta_2, a_1)$	$l(\theta_2, a_2)$	- - - -	$l(\theta_2, a_m)$
$\vdots$	$\vdots$			
θ_k	$l(\theta_k, a_1)$	$l(\theta_k, a_2)$	- - - -	$l(\theta_k, a_m)$

(columns under $L(\theta, a)$)

$$L_1 = E[l(\theta_1, a)]$$
$$= l(\theta_1, a_1)\cdot p_1 + l(\theta_1, a_2) p_2 + \dots + l(\theta_1, a_m) p_m$$
$$\vdots$$
$$L_k = E[l(\theta_k, a)]$$
$$= l(\theta_k, a_1)\cdot p_1 + l(\theta_k, a_2) p_2 + \dots + l(\theta_k, a_m) p_m$$

$$\Rightarrow \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ \vdots \\ L_k \end{bmatrix} = p_1 \begin{bmatrix} l(\theta_1, a_1) \\ \vdots \\ l(\theta_i, a_1) \\ \vdots \\ l(\theta_k, a_1) \end{bmatrix} + p_2 \begin{bmatrix} l(\theta_1, a_2) \\ \vdots \\ l(\theta_i, a_2) \\ \vdots \\ l(\theta_k, a_2) \end{bmatrix} + \dots + p_m \begin{bmatrix} l(\theta_1, a_m) \\ \vdots \\ l(\theta_i, a_m) \\ \vdots \\ l(\theta_k, a_m) \end{bmatrix}$$

– $(L_1, L_2, \dots, L_k)$ is a point in R^k which is a convex combination points $(l(\theta_i, a_1), l(\theta_i, a_2), \dots, l(\theta_i, a_m))$ $i = 1, 2, \dots, k$

PAGE NO.:
DATE: / /

- The set of Loss Points $(L_1, L_2, \ldots, L_k)$ defined by all possible randomized actions $(P_1, P_2, \ldots P_m)$ is a convex set in R^k. ~~If~~ It forms a $\boxed{\text{Convex Polyhedron}}$ where the extreme points are pure action / strategies.
(e.g. $P = (1, 0, \ldots, 0)$ or $(0, 1, 0, \ldots, 0)$ etc)

$L(\theta, a)$

P_1	1	0	0	0	0
$\theta \backslash a$	a_1	a_2	a_3	a_4	a_5
θ_1	2	4	3	5	3
θ_2	3	0	3	2	5

$$P_1 = (1, 0, 0, 0, 0) \Rightarrow L_1 = E[L(\theta_1, a)]$$
$$= l(\theta_1, a_1) \cdot P_1 + \cdots + l(\theta_1, a_5) P_5$$
$$= \boxed{2}$$

and

$$L_2 = E[L(\theta_2, a)]$$
$$= l(\theta_2, a_1) P_1 + \cdots + l(\theta_2, a_5) \cdot P_5$$
$$= \boxed{3}$$

$\Rightarrow$ Loss Point corresponding to
$P_1 = (L_1, L_2) = (2, 3) = a_1$

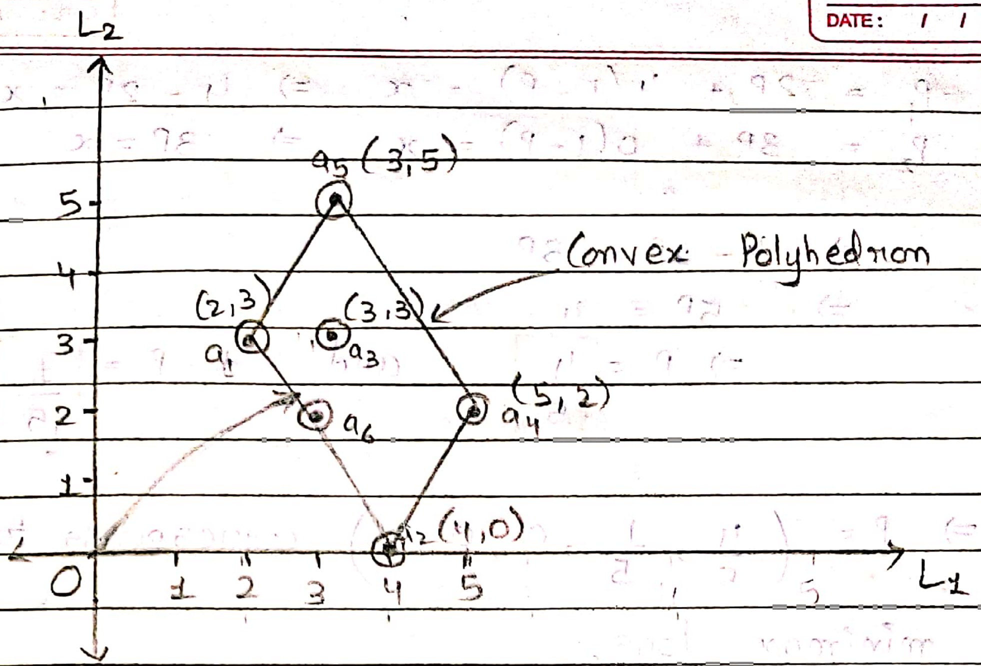
Similarly, $P_2 = (0, 1, 0, 0, 0) \Rightarrow L_1 = 4, L_2 = 0$
$\Rightarrow (L_1, L_2) = (4, 0) = a_2$

$P_3 = (0, 0, 1, 0, 0) \Rightarrow (L_1, L_2) = (3, 3) = a_3$

$P_4 = (0, 0, 0, 1, 0) \Rightarrow (L_1, L_2) = (5, 2) = a_4$

$P_5 = (0, 0, 0, 0, 1) \Rightarrow (L_1, L_2) = (3, 5) = a_5$

Scanned with OKEN Scanner



Let $P_6 = (0.6, 0.4, 0, 0, 0)$

$\Rightarrow (L_1, L_2) = (2.8, 1.8) = F$

* <u>Loss</u> <u>Table</u> :-

$\theta \backslash a$	a_1	a_2	a_3	a_4	a_5
θ_1	2	4	3	5	3
θ_2	3	0	3	2	5
$\max_\theta l(\theta, a)$	3	4	3	5	5

$$\boxed{\min_a \max_{\theta} L(\theta, a) = \boxed{3}}$$

Bisector falls on $\overline{AB}$, so if the statistician choose a mixed strategy that involves only actions a_1 and a_2 then the point (P_1, P_2) of interest will be a Convex Combination of points A and B given by

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = PA + (1-P)B = P\begin{pmatrix} 2 \\ 3 \end{pmatrix} + (1-P)\begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

PAGE NO.:
DATE: / /

$P_1 = 2P + 4(1-P) = x \Rightarrow 4 - 2P = x$
$P_2 = 3P + 0(1-P) = x \Rightarrow 3P = x$

$\Rightarrow 4 - 2P = 3P$
$\Rightarrow 5P = 4$
$\Rightarrow P = \boxed{\frac{4}{5}}$ and $1 - P = \boxed{\frac{1}{5}}$

$\Rightarrow P = \left(\frac{4}{5}, \frac{1}{5}, 0, 0, 0\right)$ corresponds to minimax loss.

$L_1 = \frac{8}{5} + \frac{4}{5} = \frac{12}{5}$

$L_2 = \frac{12}{5} + \frac{0}{5} = \frac{12}{5}$

4/8/25

Lemma:- The risk set S is a convex subset of E_K.

Proof:- Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K\}$. Let S = risk set

Let $\underline{y}^{(1)}$ and $\underline{y}^{(2)}$ be two points in S

$\Rightarrow \underline{y}^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \ldots, y_K^{(1)}\right)$ and

$\underline{y}^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \ldots, y_K^{(2)}\right)$

Since $\underline{y}^{(1)} \in S \Rightarrow \exists\ \delta_1 \in D^*$ s.t.

$R(\theta_j, \delta_1) = y_j^{(1)}, \ j = 1, 2, \ldots, K$ and

* Result:
Proof:-
De

$\underline{y}^{(2)} \in S \Rightarrow \exists\ \delta_2 \in D^*$ s.t.

$$R(\theta_j, \delta_2) = y_j^{(2)}, \quad j = 1, 2, \dots, k$$

Let $0 \le \alpha \le 1$ and $\underline{y} = \alpha \underline{y}^{(1)} + (1-\alpha)\underline{y}^{(2)}$

Then,

$$\begin{aligned} y_j &= j^{th} \text{ coordinate} \\ &= \alpha y_j^{(1)} + (1-\alpha) y_j^{(2)} \qquad j = 1, 2, \dots, k \\ &= \alpha R(\theta_j, \delta_1) + (1-\alpha) R(\theta_j, \delta_2) \\ &= R(\theta_j, \delta) \quad \text{where } \delta = \alpha\delta_1 + (1-\alpha)\delta_2 \\ & \qquad\qquad\qquad\qquad \text{and } \delta \in D^* \end{aligned}$$

$\Rightarrow \underline{y} = (y_1, y_2, \dots y_j \dots y_k) \in S$

Moreover S is convex and $S \subset E_k$. Hence S is convex subset of E_k.

4/8/25

* Baye's Rule in Estimation :-

Result:- A prior dist^n τ_0 is least favourable if and only if

$$\pi(\tau_0, \delta') \ge \inf_{\delta \in D^*} \pi(\tau, \delta) \quad \forall\ \delta \in D^* \text{ and } \tau \in \text{Ⓗ} \qquad \text{---}(1)$$

Proof:- (i) Let τ_0 be a least favourable prior dist^n

Def^n $\Rightarrow$ $\boxed{\inf_{\delta \in D^*} \pi(\tau_0, \delta) = \underset{\text{max.}}{\sup_{\tau \in \text{Ⓗ}^*}} \underset{\text{min}}{\inf_{\delta \in D^*}} \pi(\tau, \delta)}$

$$\Rightarrow \pi(\tau_0, \delta') \ge \sup_{\tau \in \text{Ⓗ}^*} \inf_{\delta \in D^*} \pi(\tau, \delta) \quad \forall\ \delta' \in D^* \qquad \text{---}(2)$$

PAGE NO.:
DATE: / /

$\Rightarrow r(\tau_0, \delta') \geq \inf_{\delta \in D^*} r(\tau, \delta) \quad \forall \; \delta' \in D^*$ and $\tau \in \Theta^*$

and hence (1)

(ii) Suppose (2) holds, that is

$\Rightarrow r(\tau_0, \delta') \geq \inf_{\delta \in D^*} r(\tau, \delta) \quad \forall \; \delta' \in D^*$ and $\tau \in \Theta^*$

$\Rightarrow r(\tau_0, \delta') \geq \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta) \quad \forall \; \delta' \in D^*$

$\Rightarrow \inf_{\delta \in D^*} r(\tau_0, \delta') = \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta)$

$\Rightarrow \tau_0$ is least favourable.

* <u>Obtaining the Baye's Estimate</u> :-

(Prior dist^n of τ) → random variable Z

- We know that $r(\tau, \delta) = E[R(z, \delta)]$ where z is a r.v. taking values in Θ with prob. dist^n τ.

- In a problem of estimation if ∃ a randomized Baye's rule w.r.t a prior dist^n τ, then there will also exist a non-randomized Baye's rule w.r.t the same prior τ.

$\therefore r(\tau, d) = E[R(z, d)]$ where d is a non-randomized decision rule.

$$f_1(x|y) = \frac{g(x,y)}{g_2(y)} \implies g(x,y) = f_1(x|y) \cdot g_2(y)$$

$$f_2(y|x) = \frac{g(x,y)}{g_1(x)} \quad \& \quad g(x,y) = f_2(y|x) g_1(x)$$

- Suppose d_0 is a non-randomized Baye's rule w.r.t. prior distⁿ τ then

$$r(\tau, d_0) = \inf_{d \in D} r(\tau, d)$$

Now,

Parameter i.e. we say parameter as r.v.

$$r(\tau, d) = E[R(Z, d)]$$

$$= \int_{\Theta} R(\theta, d)\, d\tau(\theta)$$

where

$$R(\theta, d) = E_{\theta}[L(\theta, d(x))]$$

$$= \int_{x} L(\theta, d(x))\, dF(x/\theta)$$

where $F(x/\theta)$ is a conditional distⁿ function of x given θ

$$\implies r(\tau, d) = \int_{\Theta}\int_{x} L(\theta, d(x))\, dF(x|\theta)\, d\tau(\theta)$$

$$= \int_{x}\left[\int_{\Theta} L(\theta, d(x))\, d\tau(\theta|x)\right] dF(x)$$

* <u>Objective</u> : To find Baye's decision rule and Baye's risk

$\implies$ We need to find decision ruled which ~~minimum~~ minimizes $r(\tau, d)$, that is to minimize $\int_{\Theta} L(\theta, d(x))\, d\tau(\theta/x)$, for each given x.

Scanned with OKEN Scanner

If $X \sim G(\alpha, \beta)$

$$f(x) = \frac{e^{-x/\beta} x^{\alpha-1}}{\beta^{\alpha} \Gamma\alpha}, \quad x>0$$

If $X \sim \exp(\theta)$

$$f(x) = \theta e^{-\theta x}, \quad x>0$$

PAGE NO. :
DATE : / /

Ex:- Let $\Theta = \mathcal{a} = (0, \infty)$; $L(\theta, a) = c \cdot (\theta - a)^2$
= Square error loss

Suppose $(X|\theta) \sim U(0, \theta)$
↳ (nature's action space) :

$$\Rightarrow f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x < \theta \\ 0 & \text{o.w.} \end{cases}$$

Let the prior dist^n τ of θ is
→ 1 - Parameter Gamma

$$\tau(\theta) = \begin{cases} \theta e^{-\theta}, & \theta > 0 \\ 0 & \text{o.w.} \end{cases}$$

$\Rightarrow \theta \sim G(2)$

Take $x = \theta$
$\beta = 1, \alpha = 2$
$\Rightarrow e^{-\theta}\theta, \ \theta > 0$

If Pdf of X and $\theta = f_1(x, \theta)$
$= f(x, \theta) \cdot \tau(\theta)$

$\left[\because f_1(x|y) = \frac{f(x,y)}{g_1(y)} \right.$

$\left. f(x,y) = f_1(x|y) \cdot g(y) \right]$

$$= \begin{cases} \frac{1}{\theta} \theta e^{-\theta}, & 0 < x < \theta \text{ and } \theta > 0 \\ 0, & \text{o.w.} \end{cases}$$

$$= \begin{cases} e^{-\theta}, & 0 < x < \theta \\ 0 & \text{o.w.} \end{cases}$$

Marginal Pdf of $X = h(x)$

$$= \int_{\theta} f_1(x, \theta)\, d\theta$$

$$= \int_{\theta = x}^{\infty} e^{-\theta}\, d\theta$$

$\theta > 0$ and $\theta > x$
$\Rightarrow \theta > \max(0, x)$
but $x > 0 \Rightarrow \theta > x$

To

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g$$

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

$$= \left[-e^{-\theta}\right]_x^\infty = e^{-x}.$$

$\Rightarrow h(x) = e^{-x}, \quad x > 0 \Rightarrow X \sim \exp(1)$

$\Rightarrow$ Conditional Dist^n of θ/x (Posterior) $= \tau_1(\theta/x)$

$$= \frac{f_1(x,\theta)}{h(x)}$$

$$= \frac{e^{-\theta}}{e^{-x}}$$

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

$$= \begin{cases} e^{x-\theta} & \theta > x \\ 0 & \text{o.w.} \end{cases}$$

$\therefore$ Conditional Expected Loss (S)

$$= \int_\theta L(\theta, d(x)) \cdot d\tau_1(\theta/x)$$

$$= c\int_x^\infty (\theta - a)^2 \cdot e^{x-\theta}\, d\theta$$

To minimize S, $\frac{dS}{da} = 0$

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g$$

$$\Rightarrow c\int_x^\infty 2(\theta - a)(-1)e^{x-\theta}\, d\theta = 0$$

$$\Rightarrow \int_x^\infty (\theta - a)e^{x-\theta}\, d\theta = 0$$

$$\Rightarrow \int_x^\infty \theta e^{x-\theta}\, d\theta - a\int_x^\infty e^{x-\theta}\, d\theta = 0$$

$$\Rightarrow e^x \int_{\theta=x}^\infty \theta e^{-\theta} = a e^x \left[\int_x^\infty e^{-\theta}\, d\theta\right]$$

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

$$= ae^{x}\left[-e^{-\theta}\right]_{x}^{\infty}$$

$$= ae^{x}e^{-x} = \boxed{a}$$

$$\Rightarrow e^{x}\left[\left[\theta(-e^{-\theta})\right]_{\theta=x}^{\infty} - \int_{\theta=x}^{\infty} -e^{-\theta}\right] = a$$

$$\Rightarrow e^{x}\left[xe^{-x} + (-e^{-\theta})_{x}^{\infty}\right] = a$$

$$\Rightarrow e^{x}\left[xe^{-x} + e^{-x}\right] = a$$

$$\Rightarrow e^{x}e^{-x}(x+1) = a$$

$$\Rightarrow \boxed{a = x+1}$$

$\Rightarrow$ $d(x) = x+1$ is Baye's estimate of θ wrt prior distⁿ $G(2)$.

<u>Note:</u> $E[\theta / X = x]$ = mean of Conditional (Posterior) distⁿ of θ/x

$$= \int_{\theta=x}^{\infty} \theta \cdot \tau_1(\theta|x)\, d\theta$$

$$= \int_{\theta=x}^{\infty} \theta e^{x-\theta}\, d\theta$$

$$= e^{x}\int_{\theta=x}^{\infty} \theta e^{-\theta}\, d\theta$$

$$= e^{x}\left[\left(\theta(-e^{-\theta})\right)_{x}^{\infty} - \int_{\theta=x}^{\infty} -e^{-\theta}\, d\theta\right]$$

Rule 2

Rule 1

Scanned with OKEN Scanner

If f is pdf of X over (a, b) then median M of X will be obtained on

$$\int_a^M F(x)\,dx = \boxed{0.5}$$

PAGE NO.: DATE: / /

$$= e^x \left[x e^{-x} + (-e^{-\theta})_x^{\infty} \right]$$

$$= e^x \left[x e^{-x} + e^{-x} \right]$$

$$= e^x e^{-x} [x + 1]$$

$$= \boxed{x + 1}$$

<u>Rule 2</u> :- In problem of estimating a real parameter θ, with loss proportional to squared error (i.e. $L(\theta, a) = c(\theta - a)^2$), a Bayes decision rule w.r.t a given prior dist^n $\tau(\theta)$ is :- To estimate θ as the mean of the posterior dist^n of θ given data (i.e. X or observation) that is,

$$\text{Bayes estimate } d(x) = E[\theta / X = x]$$
$$= \int_{\Theta} \theta \, d\tau(\theta \mid X = x)$$

<u>Rule 1</u> :- In problem of estimating a real parameter θ with loss proportional to absolute error (i.e. $L(\theta, a) = c|\theta - a|$) a Bayes decision rule w.r.t a given prior dist^n $\tau(\theta)$ is : To estimate θ as the median of the Posterior dist^n of θ given data (i.e. given X or abs^n X_i) that is

$d(x)$ = median of conditional dist^n of $\theta | X = x$

Posterior dist^n $(\theta | x)$

PAGE NO.:
DATE: / /

7/8/25

Ex: Let $(X|\theta) \sim B(\overset{\text{known}}{n}, \theta)$, $\theta \sim \beta_I(\alpha, \beta)$, $L(\theta, a) = (\theta - a)^2$. Find Bayes estimator of θ under the given loss function.

Sol^n: Here $\Theta = [0, 1]$

$\Rightarrow f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$; $x = 0, 1, 2, \ldots, n$, $0 \le \theta \le 1$

&

$$\theta \sim \beta_I(\alpha, \beta) \Rightarrow \tau(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \le \theta \le 1, \ \alpha, \beta > 0$$

$\Rightarrow$ Jt pdf of X and θ = $f_1(x, \theta) = f(x|\theta)\tau(\theta)$

$$= \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}; \quad x = 0, \ldots, n, \ 0 \le \theta \le 1, \ \alpha, \beta > 0$$

$$\Rightarrow f_1(x, \theta) = \frac{\binom{n}{x} \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}}{B(\alpha, \beta)}$$

Marginal density of X = $h(x) = \int_\Theta f_1(x, \theta)\, d\theta$

$$= \frac{\binom{n}{x}}{B(\alpha, \beta)} \int_{\theta=0}^{1} \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}\, d\theta.$$

$$= \frac{\binom{n}{x} B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)} \underbrace{\int_{\theta=0}^{1} \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{B(x+\alpha, n-x+\beta)}\, d\theta}_{=1}$$

If $X \sim \beta_I(\alpha, \beta)$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \le x \le 1, \ \alpha, \beta > 0$$

$$\Rightarrow \int_0^1 f(x)\, dx = 1$$

$$\Rightarrow \int_{x=0}^{1} x^{\alpha-1}(1-x)^{\beta-1} = B(\alpha, \beta)$$

$$h(x) = \frac{\binom{n}{x} B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)}; \quad x = 0, \ldots, n, \ \alpha, \beta > 0$$

Conditional distⁿ of $\theta | X = x$ = Posterior distⁿ of θ

$$= \tau_1(\theta | X = x) = \frac{f_1(x, \theta)}{h(x)}$$

$$= \frac{\cancel{\binom{n}{x}}\, \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}}{\cancel{B(\alpha, \beta)}} \cdot \frac{\cancel{B(\alpha, \beta)}}{\cancel{\binom{n}{x}} B(x+\alpha, n-x+\beta)}$$

$$\Rightarrow \tau_1(\theta | x) = \frac{\theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}}{B(x+\alpha, n-x+\beta)} \; , \quad 0 \le \theta \le 1, \; \alpha, \beta > 0$$

$$= \text{Pdf of } \beta_I(x+\alpha, n-x+\beta)$$

Given $L(\theta, a) = (\theta - a)^2$

Bayes estimate of $\theta = d(x) = a = E[\theta | X = x]$

= Mean of Conditional distⁿ of $\theta | X = x$

= Mean of Posterior distⁿ $\tau_1(\theta | X = x)$

= Mean of $\beta_I(x+\alpha, n-x+\beta)$

$$= \frac{x+\alpha}{\cancel{x} + \alpha + n - \cancel{x} + \beta}$$

If $X \sim \beta_I(\alpha, \beta)$ then $E(X) = \dfrac{\alpha}{\alpha+\beta}$

$$\Rightarrow d(x) = \frac{x+\alpha}{\alpha+\beta+n}$$

Ex: If $X \sim B(5, \theta)$, $\theta \sim \beta_I(1, 9)$

then $d(x) = \dfrac{x+1}{15}$

* **Bayes Estimate in case of weighted squared error loss:**

Let $L(\theta, a) = \omega(\theta) \cdot (\theta - a)^2$

Then to find Bayes estimate w.r. to $\tau(\theta)$.

Scanned with OKEN Scanner

we minimize $\int L(\theta, a)\, d\tau(\theta|X=x)$

i.e. minimize $\int \omega(\theta)(\theta-a)^2\, d\tau(\theta|X=x)$

$$\Rightarrow \frac{\partial}{\partial a}\left[\int \omega(\theta)\cdot(\theta-a)^2\, d\tau(\theta|X=x)\right] = 0$$

$$\Rightarrow -2a\int \omega(\theta)\cdot(\theta-a)\, d\tau(\theta|X=x) = 0$$

$$\Rightarrow \int \omega(\theta)\,\theta\, d\tau(\theta|X=x) = a\int \omega(\theta)\, d\tau(\theta|X=x)$$

$\Rightarrow$ Bayes estimate of θ

$= d(x) = a$

$$= \frac{\int \theta\cdot\omega(\theta)\, d\tau(\theta|X=x)}{\int \omega(\theta)\, d\tau(\theta|x)} = \frac{E[\theta\cdot\omega(\theta)|X=x]}{E[\omega(\theta)|X=x]}$$

Now,

Suppose $L(\theta, a) = \dfrac{(\theta-a)^2}{\theta(1-\theta)} \Rightarrow \omega(\theta) = \dfrac{1}{\theta(1-\theta)}$

Then Bayes estimate of $\theta = d(x) = a$

$$= \frac{E[\theta\cdot\omega(\theta)|X=x]}{E(\omega(\theta)|X=x]} \quad \text{——(1)}$$

$$E[\theta\cdot\omega(\theta)|X=x] = \int_\theta \theta\cdot\omega(\theta)\cdot\tau_1(\theta|x)\,dx$$

$$= \int_{\theta=0}^{1} \cancel{\theta}\cdot\frac{1}{\cancel{\theta}(1-\theta)} \times \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{B(x+\alpha, n-x+\beta)}\, d\theta$$

Let $f(x)$ be a pdf of X & $g(x)$ be any other funct^n of X, then

$E[g(x)] = \int_x g(x)\cdot f(x)\,dx$

$$= \frac{B(x+\alpha, n-x+\beta-1)}{B(x+\alpha, n-x+\beta)} \underbrace{\int_{\theta=0}^{1} \frac{\theta^{x+\alpha-1}(1-\theta)^{(n-x+\beta-1)-1}}{B(x+\alpha, n-x+\beta-1)}\, d\theta}_{=1}$$

PAGE NO.:
DATE: / /

$$= \frac{B(x+\alpha,\ n-x+\beta-1)}{B(x+\alpha,\ n-x+\beta)} \quad \text{——— (2)}$$

$$E[w(\theta) \mid X=x] = \int_\theta w(\theta)\,\tau_1(\theta|x)\,d\theta$$

$$= \int_\theta \frac{1}{\theta(1-\theta)} \cdot \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{B(x+\alpha,\ n-x+\beta)}\,d\theta$$

$$= \frac{B(x+\alpha-1,\ n-x+\beta-1)}{B(x+\alpha,\ n-x+\beta)} \quad \text{——— (3)}$$

Using (2) & (3) in (1)

$$\Rightarrow d(x) = \frac{B(x+\alpha,\ n-x+\beta-1)}{B(x+\alpha,\ \cancel{n-x+\beta})} \cdot \frac{B(x+\alpha,\ n-x+\beta)}{B(x+\alpha-1,\ n-x+\beta-1)}$$

$$\cancel{\Rightarrow d(x) = \frac{\Gamma(x+\alpha)\,\Gamma(n-x+\beta-1)}{\Gamma(x+\alpha+n-x+\beta-1)} \cdot \frac{\Gamma(x+\alpha-1)\,\Gamma(n-x+\beta-1)}{\Gamma(x+\alpha-1+n-x+\beta-1)}}$$

$$\Rightarrow d(x) = \frac{\Gamma(x+\alpha)\,\cancel{\Gamma(n-x+\beta-1)}}{\Gamma(\cancel{x}+\alpha+n-\cancel{x}+\beta-1)} \cdot \frac{\Gamma(\cancel{x}+\alpha-1+n-\cancel{x}+\beta-1)}{\Gamma(x+\alpha-1)\,\cancel{\Gamma(n-x+\beta-1)}}$$

$$= \frac{\Gamma(x+\alpha)\,\Gamma(\alpha+\beta+n-2)}{\Gamma(\alpha+\beta+n-1)\,\Gamma(x+\alpha-1)}$$

$$\Rightarrow d(x) = \frac{(x+\alpha-1)\,\cancel{\Gamma(x+\alpha-1)}\,\cancel{\Gamma(\alpha+\beta+n-2)}}{(\alpha+\beta+n-2)\,\cancel{\Gamma(\alpha+\beta+n-2)}\,\cancel{\Gamma(x+\alpha-1)}}$$

$$d(x) = \boxed{\frac{x+\alpha-1}{\alpha+\beta+n-2}}$$

Ex: If $L(\theta, a) = |\theta - a|$ = absolute error loss
Then $d(x) = a$ = median of distⁿ $(\theta \mid X=x)$
= median of Posterior distⁿ $\tau_1(\theta \mid X=x)$

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

Thus, the problem will be to find $d(x) = a$ s.t.

$$\int_{\theta=0}^{a} \tau_1(\theta|x)\,d\theta = 0.5$$

Incomplete beta Integral

13/8/25

Ex: Let $\Theta = (0, \infty)$, $\mathcal{A} = \mathbb{R}$, $L(\theta, a) = (\theta - a)^2$

$x|\theta \sim P(\theta)$, $\theta \sim G(\alpha, \beta)$

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

and $$\tau(\theta) = \frac{e^{-\theta/\beta}\,\theta^{\alpha-1}}{\beta^{\alpha}\Gamma\alpha}; \quad \theta > 0 \text{ and } \alpha, \beta > 0$$

Jt Pdf of x and θ = $f_1(x, \theta) = f(x|\theta)\cdot\tau(\theta)$

$$= \frac{e^{-\theta}\theta^{x}}{x!}\,\frac{e^{-\theta/\beta}\,\theta^{\alpha-1}}{\beta^{\alpha}\Gamma\alpha}, \quad x = 0, 1, 2, \ldots; \ \theta > 0; \ \alpha, \beta > 0$$

Marginal Pdf of x = $h(x) = \int_{\theta} f_1(x, \theta)\,d\theta$

$$= \frac{1}{x!\,\beta^{\alpha}\Gamma\alpha}\int_{\theta=0}^{\infty} e^{-\theta\left(1+\frac{1}{\beta}\right)}\theta^{x+\alpha-1}\,d\theta$$

$$= \frac{1}{\beta^{\alpha}\Gamma\alpha\, x!}\int_{\theta=0}^{\infty}\frac{e^{-\theta/\left(\frac{\beta}{1+\beta}\right)}\,\theta^{x+\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{x+\alpha}\Gamma x+\alpha}{\left(\frac{\beta}{1+\beta}\right)^{x+\alpha}\Gamma x+\alpha}\,d\theta$$

$$= \frac{\beta^{x+\alpha}\,\Gamma x+\alpha}{(\beta+1)^{x+\alpha}\,\beta^{\alpha}\Gamma\alpha\, x!}\underbrace{\int_{\theta=0}^{\infty} G\left(x+\alpha, \frac{\beta}{1+\beta}\right)}_{=1}$$

$$\Rightarrow h(x) = \frac{\beta^{x}\,\Gamma x+\alpha}{(\beta+1)^{x+\alpha}\,\Gamma\alpha\, x!}, \quad x = 0, 1, 2, \ldots$$

Posterior $dist^n$ of $\theta | X = x$

$= $ Conditional $dist^n$ of $\theta | X = x = \tau_1(\theta|x)$

$$= \frac{f_1(x,\theta)}{h(x)}$$

$$\Rightarrow \tau_1(\theta|x) = \frac{e^{-\theta}\theta^x}{\cancel{x!}} \frac{e^{-\theta/\beta}\theta^{\alpha-1}(\beta+1)^{x+\alpha}\cancel{\Gamma\alpha}\,\cancel{x!}}{\beta^\alpha\cancel{\Gamma\alpha}\,\beta^x\Gamma{x+\alpha}}$$

$$= \frac{e^{-\theta/\left(\frac{\beta}{1+\beta}\right)}\theta^{x+\alpha-1}}{\left(\frac{\beta}{\beta+1}\right)^{x+\alpha}\Gamma{x+\alpha}} \quad ; \theta > 0$$

$$= G\left(x+\alpha, \frac{\beta}{\beta+1}\right)$$

(squared error loss)

with $L(\theta, a) = (\theta - a)^2$, Bayes estimate $= d(x)$

$= a = E(\theta | X = x)$

$= $ Mean of $\tau_1(\theta|x)$

$= (x+\alpha)\cdot\left(\frac{\beta}{\beta+1}\right)$

$$= \frac{(x+\alpha)\beta}{\beta+1}$$

If $X \sim G(\alpha, \beta)$
$\Rightarrow E(X) = \alpha\beta$

Ex:- Let $\textcircled{H} = \mathscr{A} = \mathbb{R}$, $L(\theta, a) = (\theta - a)^2$.
Let $X|\theta \sim N(\theta, 1)$. Find Bayes estimate of θ w.r. to a prior $dist^n$ of θ, where $\theta \sim N(0, \sigma^2)$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} \quad ; -\infty < x < \infty \text{ and } -\infty < \theta < \infty.$$

and

$$\tau(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{\theta^2}{\sigma^2}} \quad ; -\infty < \theta < \infty \quad ; \sigma > 0$$

PAGE NO.:
DATE: / /

If pdf of x and θ = $f_1(x,\theta) = f(x|\theta) \cdot \tau(\theta)$

$$= \frac{1}{2\pi\sigma} e^{-\frac{1}{2}\left[(x-\theta)^2 + \frac{\theta^2}{\sigma^2}\right]}$$

Then marginal distⁿ of x = $h(x) = \int_{\theta} f_1(x,\theta)\,d\theta$

$$= \frac{1}{2\pi\sigma} \int_{\theta=-\infty}^{\infty} e^{-\frac{1}{2}\left[(x-\theta)^2 + \frac{\theta^2}{\sigma^2}\right]} d\theta$$

$$= \frac{1}{2\pi\sigma} \int_{\theta=-\infty}^{\infty} e^{-\frac{1}{2}\left[x^2 - 2x\theta + \theta^2\left(1+\frac{1}{\sigma^2}\right)\right]} d\theta$$

$$\Bigg(-\frac{1}{2}\left[x^2 - 2x\theta + \theta^2\left(1+\frac{1}{\sigma^2}\right)\right]$$

$$= -\frac{1}{2}\left[x^2 - 2x\theta + \frac{\theta^2(\sigma^2+1)}{\sigma^2}\right]$$

$$= -\frac{1}{2}\frac{(\sigma^2+1)}{\sigma^2}\left[\theta^2 - \frac{2x\theta\sigma^2}{\sigma^2+1} + \frac{x^2\sigma^2}{\sigma^2+1}\right]$$

$$= -\frac{1}{2}\frac{(\sigma^2+1)}{\sigma^2}\left[\theta^2 - \frac{2x\theta\sigma^2}{\sigma^2+1} + \frac{x^2\sigma^4}{(\sigma^2+1)^2} - \frac{x^2\sigma^4}{(\sigma^2+1)^2} + \frac{x^2\sigma^2}{(\sigma^2+1)}\right]$$

$$= -\frac{1}{2}\frac{(\sigma^2+1)}{\sigma^2}\left[\left(\theta - \frac{x\sigma^2}{\sigma^2+1}\right)^2 - \frac{x^2\sigma^4}{(\sigma^2+1)^2} + \frac{x^2\sigma^2}{(\sigma^2+1)}\right]$$

$$= -\frac{1}{2}\frac{(\sigma^2+1)}{\sigma^2}\left[\left(\theta - \frac{x\sigma^2}{\sigma^2+1}\right)^2 - \frac{x^2\sigma^2}{(\sigma^2+1)}\left(\frac{\sigma^2}{\sigma^2+1} - 1\right)\right]$$

$$= -\frac{1}{2}\frac{(\sigma^2+1)}{\sigma^2}\left[\left(\theta - \frac{x\sigma^2}{\sigma^2+1}\right)^2 - \frac{x^2\sigma^2}{(\sigma^2+1)^2}\right] \Bigg)$$

$$= \frac{1}{2\pi\sigma}\int_{\theta=-\infty}^{\infty} e^{-\frac{1}{2}\frac{\sigma^2+1}{\sigma^2}\left(\theta-\frac{x\sigma^2}{\sigma^2+1}\right)^2} e^{-\frac{1}{2}\frac{(\cancel{\sigma^2+1})}{\cancel{\sigma^2}}\frac{x^2\sigma^2}{(\sigma^2+1)^{\cancel{2}}}}\,d\theta$$

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{\frac{-x^2}{2(\sigma^2+1)}} \cdot \frac{\sigma}{\sqrt{1+\sigma^2}} \int_{\theta=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\cdot\frac{\sigma}{\sqrt{1+\sigma^2}}}\, e^{-\frac{1}{2}\frac{\left(\theta-\frac{x\sigma^2}{\sigma^2+1}\right)^2}{\left(\frac{\sigma^2}{1+\sigma^2}\right)}}\,d\theta$$

$$h(x) = \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}}\, e^{-\frac{x^2}{2(\sigma^2+1)}} \cdot \frac{\cancel{\sigma}}{\sqrt{1+\sigma^2}} \underbrace{\int_{\theta=-\infty}^{\infty} N\left(\frac{x\sigma^2}{\sigma^2+1}, \frac{\sigma^2}{1+\sigma^2}\right)}_{=1}$$

$$\Rightarrow h(x) = \frac{1}{\sqrt{2\pi}\sqrt{1+\sigma^2}}\, e^{-\frac{x^2}{2(\sigma^2+1)}} \;; \quad -\infty < x < \infty.$$

$\Rightarrow$ Posterior dist^n of $\theta \mid X = x$ $= \tau_1(\theta \mid x) = \dfrac{f_1(x,\theta)}{h(x)}$

$$= \frac{\dfrac{1}{2\pi\sigma}\, e^{\frac{-1}{2}\left[(x-\theta)^2 + \frac{\theta^2}{\sigma^2}\right]}}{\dfrac{1}{\sqrt{2\pi}\sqrt{1+\sigma^2}}\, e^{-\frac{x^2}{2(\sigma^2+1)}}}$$

$$= \frac{\sqrt{1+\sigma^2}}{\sqrt{2\pi}\,\sigma}\, \frac{e^{\frac{-1}{2}\left(\frac{1+\sigma^2}{\sigma^2}\right)\left(\theta-\frac{x\sigma^2}{1+\sigma^2}\right)^2} \cdot \cancel{e^{\frac{-1}{2}\frac{x^2}{(1+\sigma^2)}}}}{\cancel{e^{\frac{-1}{2}\frac{x^2}{(1+\sigma^2)}}}}$$

$$= \frac{1}{\sqrt{2\pi}\,\frac{\sigma}{\sqrt{1+\sigma^2}}}\, e^{\frac{-1}{2}\frac{\left(\theta-\frac{x\sigma^2}{1+\sigma^2}\right)^2}{\frac{\sigma^2}{1+\sigma^2}}}$$

$$= N\left(\frac{x\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$$

PAGE NO.:
DATE: / /

$\Rightarrow$ Bayes Estimate w.r. to prior $\tau(\theta)$ for squared error loss

$= E[\theta \mid X = x)$

$= \text{Mean of } N\left(\frac{x\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$

$= \frac{x\sigma^2}{1+\sigma^2} = d_\sigma(x) = a$

Also, Bayes risk $= r(\tau_\sigma, d_\sigma)$

$= E\left[E\{(\theta - d_\sigma(x))\}^2 \mid x\right]$

$\Rightarrow r(\tau_\sigma, d_\sigma) = E\left[\left(\theta - \underbrace{\frac{x\sigma^2}{1+\sigma^2}}_{E(\theta/x)}\right)^2 \Big| x\right]$ $\left(E[x - E(x)]^2 = V(x) \right)$

$= \text{Var of } \tau(\theta \mid x = x)$

$= \frac{\sigma^2}{1+\sigma^2}$

NOTE:- If we let $d(x) = x$, and we know that

$d_\sigma(x) = \frac{x\sigma^2}{1+\sigma^2} = \frac{x}{\frac{1}{\sigma^2}+1}$ ($\because$ dividing N & D by σ^2)

and it can be observed that $d_\sigma(x) \to d(x)$ as $\sigma \to \infty$.

* Useful Extension in the definition of Bayes rule:-

Def^n:- A rule δ is said to be a limit of Bayes rules δ_n ($n \to$ sample size), if for almost all x, $\delta_n(x) \to \delta(x)$ in distribution.

Scanned with OKEN Scanner

(long term behaviour so we use limit)

PAGE NO.:
DATE: / /

Note:- For non-randomized decision rules, this defⁿ becomes $d_n \to d$ for almost all x.

Ex:- In the previous Example, where $X|\theta \sim N(0,1)$ and $\theta \sim N(0,\sigma^2) = \tau_{\sigma^2}$, we have seen that $d_\sigma(x) \to d(x)$ if $\sigma^2 \to \infty$ where $d_\sigma(x) = \frac{x\sigma^2}{1+\sigma^2}$ and $d(x) = x$.

Defⁿ:- Generalized Bayes Rule:-

- A rule δ_0 is said to be generalized Bayes rule if $\exists$ a measure τ on Θ such that $\int L(\theta,\delta) f_x(x|\theta)\, d\tau(\theta)$ takes on a finite minimum value when $\delta = \delta_0$ (Note that $\tau(\theta)$ is not prior distⁿ)

Ex:- Consider Example where $x|\theta \sim N(\theta,1)$, ~~$\theta \sim N(0,\sigma^2)$~~.
Let $d(x) = x$ be generalized Bayes rule when $d\tau(\theta) = d\theta$ i.e. $\tau(\theta) = \theta$.

To verify this, let $S = \int L(\theta,d) f_x(x|\theta)\, d\tau(\theta)$

$$= \int_\theta (\theta - a)^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}\, d\theta.$$

For S to be minimum, $\frac{\partial S}{\partial a} = 0$

$$\Rightarrow \frac{-2}{\sqrt{2\pi}} \int_\theta (\theta - a) e^{-\frac{1}{2}(x-\theta)^2}\, d\theta = 0$$

$$\Rightarrow \underbrace{\frac{1}{\sqrt{2\pi}} \int_\theta \theta e^{-\frac{1}{2}(x-\theta)^2}\, d\theta}_{E(\theta) =} = a \underbrace{\frac{1}{\sqrt{2\pi}} \int_\theta e^{-\frac{1}{2}(x-\theta)^2}\, d\theta}_{\int N(0,1) = 1}$$

PAGE NO.:
DATE: / /

$\Rightarrow E(\theta) = a$ ~~d(x)~~

$\Rightarrow$ If $\theta \sim N(x, 1)$, then $E(\theta) = x$ and therefore $d(x) = x$ is a generalized Bayes rule.

14/8/2

Def^n:- Extended Bayes Rule :-

- A rule δ_0 is said to be Extended Bayes if δ_0 in ε- Bayes for every $\varepsilon > 0$.

OR

A rule δ_0 is said to be Extended Bayes if for ~~every~~ every $\varepsilon > 0$, $\exists$ a prior dist^n τ such that δ_0 is ε - Bayes w.r. to τ, that is,

$$\boxed{r(\tau, \delta_0) \le \inf_{\delta} r(\tau, \delta) + \varepsilon}$$

Mathematical form

Ex:- Consider Ex. where $x|\theta \sim N(\theta, 1)$, $\theta \sim N(0, \sigma^2)$ ie $\tau_\sigma(\theta) = N(0, \sigma^2)$. To show that $d(x) = x$ is Extended Bayes rule.

Compute $r(\tau_\sigma, d) = E[(\theta - x)^2] = E[E\{(\theta - x)^2 | \theta\}]$

$= E[E\{(x - \theta)^2 | \theta\}]$

$= E[V(x|\theta)] = \boxed{1}$ $E[x|\theta]$

Now, $\inf_{d} (\tau_\sigma, d) = r(\tau_\sigma, d_\sigma) = \dfrac{\sigma^2}{1+\sigma^2}$ ← Min. risk.

$\Rightarrow \underbrace{r(\tau_\sigma, d)}_{=1} = \inf_{d} r(\tau_\sigma, d) + \varepsilon$ for $\varepsilon = \dfrac{1}{1+\sigma^2}$

$\left(\because \dfrac{\sigma^2}{1+\sigma^2} + \varepsilon = 1 \right.$

$\left. \Rightarrow \varepsilon = 1 - \dfrac{\sigma^2}{1+\sigma^2} = \dfrac{1}{1+\sigma^2} \right)$

Def^n

NOT

Def^n

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

$\Rightarrow$ d is E - Bayes rule for choice of $\varepsilon = \frac{1}{1+\sigma^2}$

14/8/25

UNIT - 2

* Admissibility and Completeness :-

Def$^n_{(1)}$:- Natural Ordering

- A decision rule δ_1 is said to be as good as a rule δ_2 if

$$R(\theta, \delta_1) \leq R(\theta, \delta_2) \quad \forall \theta \in \Theta$$

[If $R(\theta, \delta_1) = R(\theta, \delta_2)$ a decision rule of δ_1 is equivalent as rule δ_2.]

A rule δ_1 is said to be better than a rule δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2) \quad \forall \theta \in \Theta$.

and $R(\theta, \delta_1) < R(\theta, \delta_2)$ for atleast one $\theta \in \Theta$.

A rule δ_1 is said to be equivalent to a rule δ_2 if

$$R(\theta, \delta_1) = R(\theta, \delta_2) \quad \forall \theta \in \Theta$$

NOTE :- If δ_1 is as good as δ_2 then δ_2 is not to be preferred over δ_1.

Def$^n_{(2)}$:- A rule δ is said to be "admissible" if $\exists$ no rule better than δ.

Scanned with OKEN Scanner

A rule is said to be "**inadmissible**" if it is not admissible that is $\exists$ a better decision rule than that rule.

Ex :- Let $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\mathcal{A} = \{a_1, a_2, a_3\}^{(m)}$, $\mathcal{X} = \{x_1, x_2\}_{(n)}$

Θ \ D	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
θ_1	0	$\frac{1}{64}$	$\frac{4}{64}$	$\frac{3}{64}$	$\frac{1}{16} = \frac{4}{64}$	$\frac{7}{64}$	$\frac{3}{16} = \frac{12}{64}$	$\frac{1}{4} = \frac{16}{64}$	$\frac{1}{4} = \frac{16}{64}$
θ_2	$\frac{1}{16} = \frac{4}{64}$	$\frac{2}{64}$	$\frac{4}{64}$	$\frac{2}{64}$	0	$\frac{2}{64}$	$\frac{1}{16} = \frac{4}{64}$	$\frac{2}{64}$	$\frac{1}{16} = \frac{4}{64}$
θ_3	$\frac{1}{4} = \frac{16}{64}$	$\frac{7}{64}$	$\frac{4}{64}$	$\frac{13}{64}$	$\frac{1}{16} = \frac{4}{64}$	$\frac{1}{64}$	$\frac{3}{16} = \frac{12}{64}$	$\frac{3}{64}$	0
	✓	✓	✗	✗	✓	✓	✗	✗	✓

There does not exist any decision rule (among $(d_2, \ldots, d_9)$) which is better than d_1.
$\Rightarrow$ d_1 is admissible

d_2 is better than d_4
$\Rightarrow$ d_4 is inadmissible

d_2 is better ~~the~~ than d_7 $\Rightarrow$ d_7 is inadmissible
$\nexists$ any rule better than d_2
$\Rightarrow$ d_2 is admissible.

d_5 is better than d_3 $\Rightarrow$ d_3 is inadmissible
$\nexists$ any rule better than d_5
$\Rightarrow$ d_5 is admissible.

Scanned with OKEN Scanner

PAGE NO.:
DATE: / /

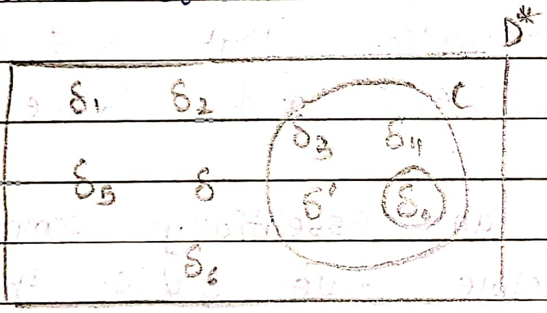
d_6 is better than d_8 $\Rightarrow$ d_8 is inadmissible
$\nexists$ any rule better than d_6 and d_9
$\Rightarrow$ d_6 and d_9 are admissible.

Hence, class of admissible rules $= \{d_1, d_2, d_5, d_6, d_9\}$
Class of inadmissible rules $= \{d_3, d_4, d_7, d_8\}$

Def^n(3) :- A class C of decision rules, $C \in D$ (or D^*) is said to be <u>complete</u> if given any rule $\delta \in D^*$ and $\delta \notin C$, $\exists$ a rule $\delta_0 \in C$ such that δ_0 is better than δ.

- A class C of decision rules is said to be <u>essentially complete</u> if given any rule $\delta \in C$, $\exists$ a rule $\delta_0 \in C$ that is as good as δ.

Ex :-



$\delta \in D^*$ but $\delta \notin C$.
Suppose δ_0 is better than δ and $\delta_0 \in C$

Suppose $C = \{d_1, d_2, d_3, d_5, d_6, d_9\}$
$\rightarrow$ $d_4 \in D^*$ but $d_4 \notin C$, but $\exists$ $d_2 \in C$ which is better than d_4.

$\rightarrow$ $d_7 \in D^*$ but $d_7 \notin C$ but $\exists$ $d_2 \in C$ which is better than d_7.

$\rightarrow$ $d_8 \in D^*$ but $d_8 \notin C$ but $\exists$ $d_6 \in C$ which is better than d_8.

$\Rightarrow$ C is complete class.

18/8/25

Lemma 1 :- If C is a complete class and A denotes the class of all admissible rules then $A \subset C$.

Proof :- Let A be the set of all admissible decision rules and C is a complete class.

Let δ be any admissible decision rule $\Rightarrow \delta \in A$ but suppose $\delta \notin C$. Since C is a complete class, and $\delta \notin C$ $\Rightarrow \exists$ a decision rule say $\delta' \in C$ such that δ' is better than δ.

if $x \in A$ $\Rightarrow x \in B$ Then $A \subset B$

$\Rightarrow$ δ is inadmissible and $\delta \notin A$ which is a contradiction.

Hence, our assumption that $\delta \notin C$ is wrong.

Thus $\delta \in A \Rightarrow \delta \in C$ and hence $A \subset C$.

Lemma 2 :- If C is an essentially complete class and $\exists$ an admissible rule $\delta \notin C$ then $\exists$ a rule $\delta' \in C$ which is equivalent to δ.

Proof :- Since C is essentially complete class δ is admissible and $\delta \notin C$ $\exists$ $\delta' \in C$ such that δ' is as good as δ, by def^n of essentially complete class

$$\Rightarrow R(\theta, \delta') \le R(\theta, \delta) \quad \forall \ \theta \in \Theta$$

If $R(\theta, \delta') < R(\theta, \delta)$ for some $\theta \in \Theta$ then

Since $R(\theta, \delta') \le R(\theta, \delta) \ \forall \ \theta \in \Theta$

and $R(\theta, \delta') < R(\theta, \delta)$ for some $\theta \in \Theta$.

$\Rightarrow$ δ' is better than δ

which means δ is inadmissible which is a contradiction.

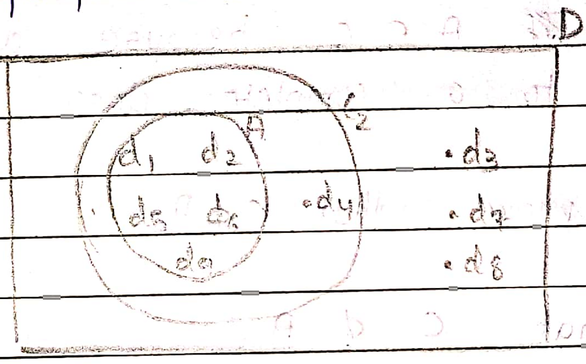
$\therefore \nexists$ any $\theta \in \Theta$, for which $R(\theta, \delta') < R(\theta, \delta)$

$\Rightarrow R(\theta, \delta') = R(\theta, \delta) \quad \forall \theta \in \Theta$

$\Rightarrow \delta'$ is equivalent to δ.

<u>Def^n (4)</u>: A class C of decision rules is said to be "<u>minimal</u> <u>Complete</u>" if
(i) C is complete and
(ii) no proper subset of C is Complete.

<u>Ex</u>:



C_1 = Complete class = $\{d_1, d_2, d_3, d_5, d_6, d_9\}$
C_2 = Complete class = $\{d_1, d_2, d_4, d_5, d_6, d_9\}$
C_3 = $\{d_1, d_4, d_5, d_6, d_9\}$ $\rightarrow$ Not a Complete class.

(minimal Complete class)
$C = \{d_1, d_2, d_5, d_6, d_9\} \rightarrow$ Complete class
$\rightarrow C' \subset C$, say $C' = \{d_2, d_5, d_6, d_9\}$

$\boxed{C \subset C_2}$

– A class C of decision rules is said to be "<u>minimal</u> <u>essentially</u> <u>complete</u>"
(i) C is essentially complete and
(ii) no proper subset of C is essentially complete.

PAGE NO.:
DATE: / /

NOTE: It is not necessary that minimal complete or minimal essentially complete classes exist.

Theorem (1): If a minimal complete class exists it consists of exactly the admissible decision rules.

Proof: Let C denote a minimal complete class and let A denote the class of all admissible rules.

To show that: $C = A$.

Lemma - 1: $A \subset C$ because a minimal complete class has to a complete class ——— (i)

Now to prove that $C \subset A$.

Assume that $C \not\subset A$.

Let $\delta_0 \in C$ but $\delta_0 \notin A$

We assume that $\exists$ $\delta_1 \in C$ that is better than δ_0 $\Rightarrow$ δ_0 is inadmissible

Because δ_0 is inadmissible $\exists$ a rule say δ is better than δ_0.

If $\delta \in C$ we may take $\delta_1 = \delta$

If $\delta \notin C$ then because C is complete $\exists$ $\delta_1 \in C$ that is better than δ and hence better than δ_0

Thus, in either case $\delta_1 \in C$ is better than δ_0

$$\left.\begin{array}{l} \delta_1 > \delta_0 \\ \delta_1 > \delta \\ \delta > \delta_0 \end{array}\right\} \quad \delta_1 > \delta > \delta_0 \quad \Rightarrow \delta_1 > \delta_0$$

Now,

let $C_1 = C - \{\delta_0\}$

let δ be an arbitrary rule not in C_1.

If $\delta = \delta_0$ then $\delta_1 \in C_1$ which is better than δ.

Scanned with OKEN Scanner

and if $\delta \neq \delta_0$ $\exists$ $\delta' \in C$ which is better than δ.

$\Rightarrow$ If $\delta' = \delta_0$, then $\delta_1 \in C$ is better than δ and if $\delta' \neq \delta_0$, then $\delta' \in C$ is better than δ.

Thus in any case $\exists$ a decision rule of C_1 is better than δ which means $C_1 \subset C$ is complete (subject)

$\Rightarrow$ a proper ~~subset~~ C_1 of C is complete which violates the second ~~condi~~ condition for a class C to be minimal complete which contradicts our assumption then $C \not\subset A$.

$\Rightarrow$ $C \not\subset A$ is incomplete

$\Rightarrow$ $C \subset A$ —— (ii)

Hence (i) and (ii) $\Rightarrow$ $C = A$

Result:- (Converse of the Previous Theorem)

If the class of admissible rules is complete then it is minimal complete.

Proof:- Let A be the class of admissible rules, which is complete.

To Prove that :- A is minimal complete

Let us assume that A is not minimal complete. Since minimal complete class is intersection of all complete classes,

Let A_1 be the proper subset of A which is complete.

Let $\delta_0 \in A$ but $\delta_0 \notin A_1$ and A_1 is complete

$\Rightarrow$ $\exists$ a decision rule say δ_1 such that $\delta_1 \in A_1$ and δ_1 is better than δ_0.

PAGE NO.:
DATE: / /

$\Rightarrow \delta_0$ is inadmissible which is a contradiction because A is class of admissible rules and $\delta_0 \in A$ which means δ_0 has to be admissible

Hence A has to be a minimal complete class

Result:- Show that if C is complete and contains no proper essentially complete sub class, then C is minimal complete and minimal essentially complete.

Proof:- Given C is complete. Let $C_1 \subset C$ and therefore C_1 is not essentially complete.

(i) Prove that C is minimal complete
That is, to prove that $C_1 \subset C$ is not complete

Suppose C_1 is complete.
$\Rightarrow$ if $\delta_0 \notin C_1$ then $\exists\ \delta \in C_1$ which is better than δ_0
that is $R(\theta, \delta) \leq R(\theta, \delta_0)\ \forall\ \theta \in \Theta$ — (1)
and $R(\theta, \delta) < R(\theta, \delta_0)$ for some $\theta \in \Theta$

But C_1 is not essentially complete
$\therefore$ For $\delta_0 \notin C_1\ \nexists\ \delta \in C_1$ which is as good as δ_0
ie. $R(\theta, \delta) \nleq R(\theta, \delta_0)\ \forall\ \theta \in \Theta$ —— (2)
Thus (1) & (2) contradicts each other.

$\Rightarrow C_1$ is not complete and $C_1 \subset C$.
$\Rightarrow$ C is minimal complete class.

Scanned with OKEN Scanner

(ii) To prove that : C is minimal essentially complete

First, we shall prove that C is essentially complete.

Suppose, we assume C is not essentially complete.

$\therefore$ For $\delta_0 \notin C$, $\nexists$ $\delta_1 \in C$ such that δ_1 is as good as δ_0 that is $R(\theta, \delta_1) \nleq R(\theta, \delta_0)$ $\forall \theta \in \Theta$

$\therefore$ $\nexists$ $\delta_1 \in C$ which is better than δ_0.

$\Rightarrow$ δ_0 is admissible and $\delta_0 \notin C$ but C is minimal complete ($\because$ Part (i))

$\Rightarrow$ $\delta_0 \notin C$ is a contradiction

$\Rightarrow$ $\delta_0 \in C$ $\Rightarrow$ C must contain all admissible rules.

Also $\exists$ $\delta_1 \in C$ such that $R(\theta, \delta_1) \leq R(\theta, \delta_0)$ $\forall \theta \in \Theta$

$\Rightarrow$ δ_1 is as good as δ_0

$\Rightarrow$ C is essentially complete

Let $C_1 \subset C$ and as per statement $C_1 \subset C$ is not essentially complete

$\Rightarrow$ C is minimal essentially complete.

Hence the proof.

21/8/25

Risk table

Ex:

Θ \ D	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
θ_1	0	1/64	4/64	3/64	4/64	7/64	12/64	16/64	16/64
θ_2	4/64	2/64	4/64	2/64	0	2/64	4/64	2/64	4/64
θ_3	16/64	7/64	4/64	13/64	4/64	1/64	12/64	3/64	0

Scanned with OKEN Scanner

$\tau'(\theta) = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$

PAGE NO.:
DATE: / /

Bayes rule wrt Prior $\tau(\theta)$?

Suppose we take prior distⁿ of θ

$$\left.\begin{array}{lccc} \theta : & \theta_1 & \theta_2 & \theta_3 \\ P(\theta) : & 1/3 & 1/3 & 1/3 \end{array}\right\} \text{——} \ \tau(\theta)$$

Bayes risk :- $r(\tau, d) = E[R(\theta, d)]$

$$= \sum_{\theta} R(\theta, d) \cdot P(\theta)$$

$r(\tau, d_1) = \frac{20}{64 \times 3} = \frac{20}{192}$

$r(\tau, d_2) = \frac{10}{64 \times 3} = \frac{10}{192}$

$r(\tau, d_3) = \frac{12}{64 \times 3} = \frac{12}{192}$

$r(\tau, d_4) = \frac{18}{172}$, $r(\tau, d_5) = \frac{8}{192}$

$r(\tau, d_6) = \frac{10}{192}$, $r(\tau, d_7) = \frac{28}{192}$

$r(\tau, d_8) = \frac{21}{192}$, $r(\tau, d_9) = \frac{20}{192}$

Bayes dec. rule is the one for which Bayes risk is minimum

$\Rightarrow$ Here, d_5 is Bayes rule w.r.t Prior distⁿ $\tau(\theta)$.

* Important theorems on decision theory & Game Theory :-

Recall Θ^* = set of all prior distⁿˢ = set of all prob. dist defined over Θ

PAGE NO.:
DATE: / /

Maximin value $= \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta)$

$= \text{lower value of the game} = \underline{V}$ —(1)

(1) represents the 'maximin' value of the game. If 'nature' is an opponent who is playing to ruin the statistician then it could use a least favourable dist^n if it existed which guarantees that the statistician expected loss would be atleast equal to $\underline{V}$ No matter what decision rule he might use whereas,

minimax value $= \inf_{\delta \in D^*} \sup_{\theta \in \Theta^*} r(\tau, \delta)$

$= \text{Upper value of game} = \bar{V}$ —(2)

(2) represents the 'minimax' value ~~of~~ or 'upper value of game'. Since, for every randomised rule $\delta \in D^*$, $\sup_{\tau \in \Theta^*} r(\tau, \delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$ —(3)

Thus, statistician has a rule which ensures him that his expected loss will not be greater than any pre-assigned number larger than the upper value no matter what prior dist^n nature decides to used.

- **The minimax Theorem** :- (The fundamental thm of game theory) :-

- Consider a game and assume that the risk set S is bounded from below. Then the game has value $\underline{V} = \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta) = \inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta) = \bar{V}$ —(4) $(\underline{V} = \bar{V}$

PAGE NO.:
DATE: / /

and a maximin strategy (that is, a least favourable prior $dist^n$ τ) exists. Moreover, if S is closed from below, then a minimax strategy (δ) exists, and $r(\tau_0, \delta_0) = V$

NOTES: (1) It is important to know, when (4) holds

(2) In game theory, where nature is also a thinking player, (4) holds under general conditions for 2-person zero-sum games.

(3) In decision theory, the minimax thm is helpful in helping the statistician to find minimax decision rule

(4) Also, the minimax thm is useful in answering the questions :- when are the ~~#~~ minimax rules also Bayes rules w.r. to some prior $dist^n$?

(Ans :- If the minimax thm holds and if ~~the~~ a least favourable prior $dist^n$, say τ_0, exists, then any minimize rule, say δ_0, is Bayes w.r. to τ_0.)

Result: Prove that :- $\sup\limits_{\tau \in \Theta^*} r(\tau, \delta) = \sup\limits_{\theta \in \Theta} R(\theta, \delta)$

Proof: Let Θ^* = set of all prior $dist^n$ defined over Θ.

ω = set of all degenerate $dist^n$ defined over Θ.

Then,

$$\omega \subset \Theta^*$$

$$\Rightarrow \sup_{\tau \in \omega} r(\tau, \delta) \leq \sup_{\tau \in \Theta^*} r(\tau, \delta) \quad \text{---} (1)$$

Scanned with OKEN Scanner

Since ω is the set of all degenerate dist^{ns} for $R(\theta_j, \delta)$, $\tau = (0, \ldots, 0, \underset{j^{th} \text{ position}}{1}, 0, \ldots, 0) \; \forall j$

$$\therefore \sup_{\tau \in \omega} r(\tau, \delta) = \sup_{\theta \in \Theta^*} R(\theta, \delta) \quad \text{---} \quad (2)$$

Thus (1) and (2)

$$\Rightarrow \sup_{\theta \in \Theta} R(\theta, \delta) \le \sup_{\tau \in \Theta^*} r(\tau, \delta) \quad \text{---} \quad (3)$$

Now consider

$$R(\theta', \delta) \le \sup_{\theta \in \Theta} R(\theta, \delta) \quad \forall \theta' \in \Theta \quad \text{---} \quad (4)$$

Taking Expectation on both the sides w.r to prior dist^n τ.

$$E[R(\theta', \delta)] \le \sup_{\theta \in \Theta} E[R(\theta, \delta)]$$

$$\Rightarrow \sum_j R(\theta', \delta) \cdot p_j \le \sum_j \underbrace{\sup_{\theta \in \Theta} R(\theta, \delta)}_{\text{constant}} \cdot p_j$$

$$\Rightarrow r(\tau, \delta) \le \sup_{\theta \in \Theta} R(\theta, \delta) \quad \forall \tau \in \Theta^*$$

$$\left(\because \sum_j p_j = 1\right)$$

$$\Rightarrow \sup_{\tau \in \Theta^*} r(\tau, \delta) \le \sup_{\theta \in \Theta} R(\theta, \delta) \quad \text{---} \quad (5)$$

Hence, (3) and (5) $\Rightarrow$ $\sup_{\tau \in \Theta^*} r(\tau, \delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$.

Scanned with OKEN Scanner