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.

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

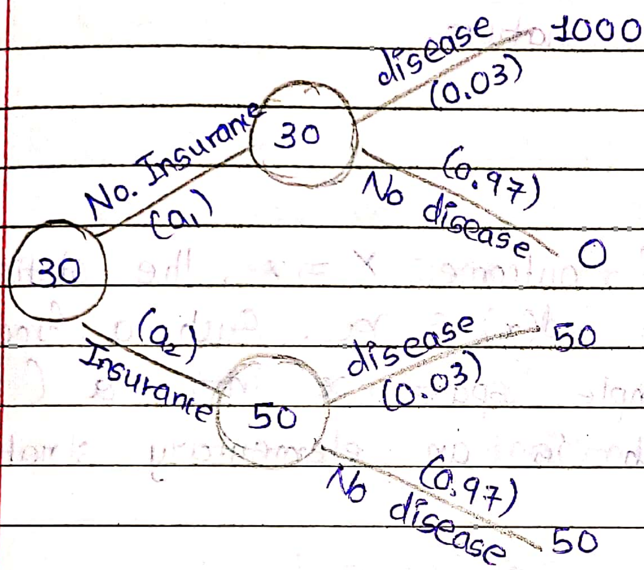
Take Insurance → Buying cost → \$ 50.

Doe not take Insurance → \$ 0.

Scanned with OKEN Scanner

$L(\theta, a)$

	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 $\rightarrow$ Positive (Prone to disease)
$\rightarrow$ Negative (Not prone to disease)

(Used for Extra Information)

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

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

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

$X = 1$ 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. distⁿ 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

D / $\mathcal{X}$	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 $= \{d_1, d_2, \dots, d_{m^n}\}$

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

Loss $L(\theta, d)$

$\mathcal{X}$	$P_\theta(x)$		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	1000	0	50	50	150	50
$x_2 = 1$	0.9	0.23	1000	0	50	50	1000	0	150	50
	1	1	$E_{\theta_1}(d_1(x))$	$E_{\theta_2}(d_1(x))$						
	$\boxed{\text{Expected loss}}$		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 $R(\theta, d)$

$\theta \diagdown D$	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 $\mathcal{X}$ 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 $d_3 = 905 \times 0.03 + 38.5 \times 0.97 = 64.495 \approx 64.5$
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.

(2)

*

Ex:-

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 = $R(\theta, d) = E_\theta[L(\theta, d)]$, where $d \to d(x)$

$$= \int_x L(\theta, d(x)) \quad \text{or} \quad \sum_x L(\theta, d(x)) \cdot f(x;\theta)$$

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

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

≈ 64.5

m^n)

oss ;

es

tion

PAGE NO.:
DATE: / /

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

Θ \ $\mathcal{a}$	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

$\mathcal{X}$ \ 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

$R(\theta, d)$

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

Hence, statistician would like to take action 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.

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

* <u>Randomized Decision Rules</u> :-

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 → 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

$$L(\theta, P) = E[L(\theta, z)] \quad \text{———} \quad (1)$$

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:

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

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

&

PAGE NO.:
DATE: / /

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

$$= 2.5$$

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 → $P = (1,0,0)$ or $(0,1,0)$ or $(0,0,1)$
Mixed Strategy →)

- 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. $dist^n$ 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 $\mathfrak{a}$ is called randomized action provided

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

where z is a r.v. taking values provides $\mathfrak{a}$ 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 funcⁿ D is called a 'randomized decision function / rule' provided the risk funcⁿ $R(\theta, \delta) = E[R(\theta, z)]$ 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:- $\mathcal{X} = \{x_1, x_2\}$, $\mathfrak{a} = \{a_1, a_2\}$, $\Theta = \{\theta_1, \theta_2\}$

$\Downarrow$ $n = 2$ $\qquad$ $\Downarrow$ $m = 2$

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

$L(\theta, a)$

$\Theta \backslash a$	a_1	a_2
θ_1	1	2
θ_2	2	1

Decision rules

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

$x \backslash D$	$P_\theta(x)$		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 $R(\theta, d)$

$\Theta \backslash D$	$\delta_1 \to \frac{1}{2}$	1/10	3/10	1/10
	d_1	d_2	d_3	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:- $\mathfrak{a} = \{a_1, a_2\}$, $\mathcal{X} = \{x_1, x_2\}$, $\Theta = \{\theta_1, \theta_2\}$

$\Downarrow$ $m = 2$, $\Downarrow$ $n = 2$

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

$\mathcal{X}$ \ D, δ	$P_1 = \frac{1}{2}$	$P_2 = \frac{3}{10}$	$P_3 = \frac{1}{10}$	$P_4 = \frac{1}{10}$
	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

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

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

Then from above table,

$d_1(x_1) = a_1$, $d_1(x_2) = a_1$

$d_2(x_1) = a_1$, $d_2(x_2) = a_2$

$d_3(x_1) = a_2$, $d_3(x_2) = a_1$

$d_4(x_1) = a_2$, $d_4(x_2) = a_2$

$\Rightarrow$ 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 $X = x_1$

$= P_{x_1}(a_1)$

$\Rightarrow 1 - \pi_1 = P_{x_1}(a_2)$

& 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 : / /

28/7/25

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

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 distⁿ 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:

28/7/25

Machines → 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)

Method - 2 : Ordering the decision rules

(i) The Bayes Principle :-

Prior Distribution :-

- The prob. distⁿ defined over Θ 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 Θ 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 $\tau \in \Theta^*$ if

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

"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.

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 $\tau \in \textcircled{H}^*$ if

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

(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

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

minimax value / upper value of game

Notes :-

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

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

(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.

PAGE NO.:
DATE: / /

<u>Def^n</u>:- Let $\varepsilon > 0$. A decision rule δ_0 is said to be ε- <u>minimax</u> iff

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

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

<u>Def^n</u>:- <u>Least</u> <u>Favourable</u> <u>Prior</u> <u>distribution</u> :-

- A prior dist^n $\tau_0 \in \Theta^*$ is said to be the <u>least favourable</u> if

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

* <u>Geometric</u> <u>Interpretation</u> <u>of</u> <u>a</u> <u>Decision</u> <u>Problem</u>

- 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.

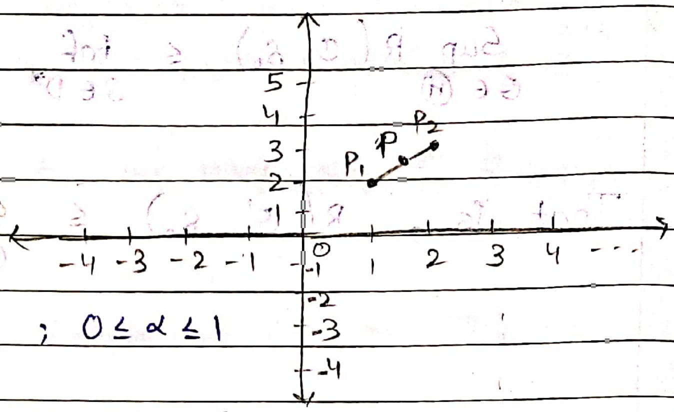
- Suppose $P_1, \ldots, P_K$ are K points in A, then their convex combination will be

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

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

— 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)

e.g. :- $A = R^2$



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

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

$\alpha = 0.4$

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

$\Rightarrow P = (1.6, 2.6)$

31/7/25

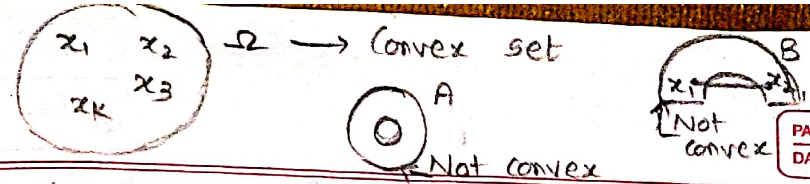
* Geometric Interpretation of Decision Problem

Def^n :- Convex Hull :-

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 void on its exterior.



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	$\cdots$	P_m
$\theta \backslash a$	a_1	a_2	$\cdots$	a_m
θ_1	$l(\theta_1, a_1)$	$l(\theta_1, a_2)$	$\cdots$	$l(\theta_1, a_m)$
θ_2	$l(\theta_2, a_1)$	$l(\theta_2, a_2)$	$\cdots$	$l(\theta_2, a_m)$
$\vdots$	$\vdots$			
θ_K	$l(\theta_K, a_1)$	$l(\theta_K, a_2)$	$\cdots$	$l(\theta_K, a_m)$

(Column heading: $L(\theta, a)$)

$$L_1 = E[l(\theta_1, a)]$$
$$= l(\theta_1, a_1)\cdot P_1 + l(\theta_1, a_2) P_2 + \cdots + l(\theta_1, a_m) P_m$$
$$\vdots$$
$$L_K = E[l(\theta_K, a)]$$
$$= l(\theta_K, a_1)\cdot P_1 + l(\theta_K, a_2) P_2 + \cdots + l(\theta_K, a_m) P_m$$

$$\Rightarrow \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_K \end{bmatrix} = P_1 \begin{bmatrix} l(\theta_1, a_1) \\ \vdots \\ l(\theta_i, a_1) \\ \vdots \\ l(\theta_K, a_1) \end{bmatrix} + P_2 \begin{bmatrix} l(\theta_1, a_2) \\ \vdots \\ l(\theta_i, a_2) \\ \vdots \\ l(\theta_K, a_2) \end{bmatrix} + \cdots + P_m \begin{bmatrix} l(\theta_1, a_m) \\ \vdots \\ l(\theta_i, a_m) \\ \vdots \\ l(\theta_K, a_m) \end{bmatrix}$$

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

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

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

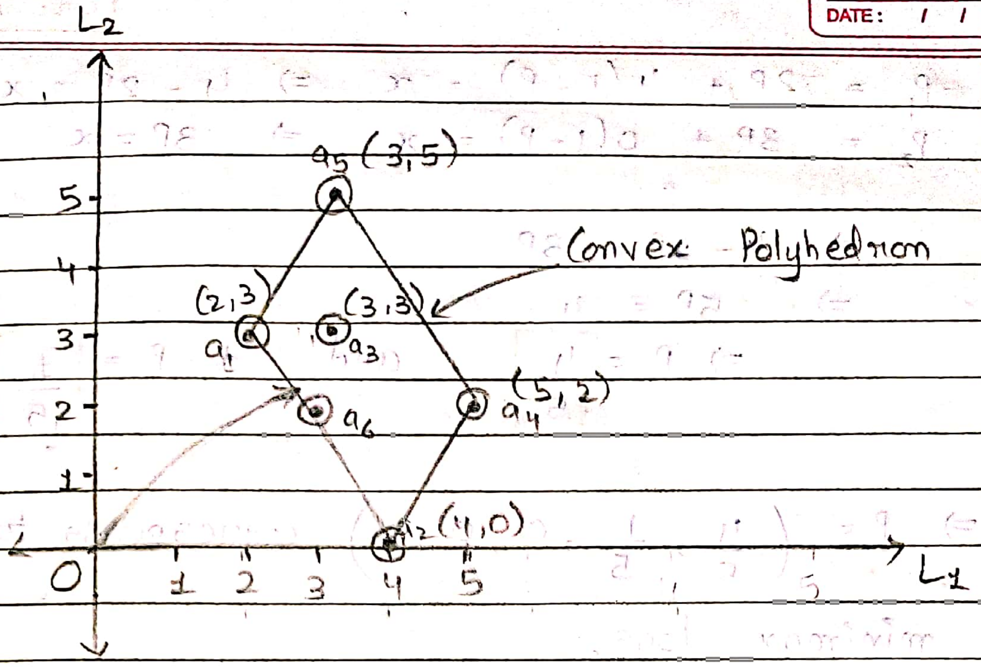
$\Rightarrow$ Loss Point corresponding to
$P_1 = (L_1, L_2) = (2, 3) = a_1$

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

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

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

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



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

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

* Loss Table :-

$\theta \backslash a$	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

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

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.

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

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

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

Proof:- Let $\textcircled{H} = \{\theta_1, \theta_2, \ldots, \theta_K\}$. 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

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

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

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

Then,

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

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

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

$$r(\tau_0, \delta') \geq \inf_{\delta \in D^*} r(\tau, \delta) \quad \forall\ \delta \in D^* \text{ and } \tau \in \Theta \qquad \text{--- (1)}$$

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

Def^n $\Rightarrow$ $\boxed{\inf_{\delta \in D^*} r(\tau_0, \delta) = \underset{\tau \in \Theta^*}{\overset{\text{max.}}{\sup}}\ \underset{\delta \in D^*}{\overset{\text{min}}{\inf}}\ r(\tau, \delta)}$

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

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

* Obtaining the Baye's Estimate :-

(Prior distⁿ of τ) $\rightarrow$ 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. distⁿ τ.

- In a problem of estimation if $\exists$ a randomized Baye's rule w.r.t a prior distⁿ 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.

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

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

– 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,

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

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

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

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

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

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

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

⇒ 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.

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 $\Theta = \alpha = (0, \infty)$; $L(\theta, a) = c.(\theta - a)^2$

$=$ Square error loss

Suppose $(X|\theta) \sim U(0, \theta)$

↳ (nature's action space) :

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

Let the prior $dist^n$ τ of θ is

→ 1 - Parameter Gamma

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

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

Take $x = \theta$

$\beta = 1, \alpha = 2$

$\Rightarrow e^{-\theta}\theta, \quad \theta > 0$

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

$= f(x, \theta) \cdot \tau(\theta)$

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

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

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

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

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 $x > 0 \Rightarrow \theta > x$

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

PAGE NO.:
DATE: / /

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

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

$\Rightarrow$ 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)

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

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

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

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

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

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

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

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

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 $G(2)$.

Rule 2

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

Rule 1

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

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

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

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

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

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

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

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

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

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

Rule 2 :- 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,

$$\text{Bayes estimate } d(x) = E\left[\theta / X = x\right]$$

$$= \int_{\Theta} \theta \, d\tau(\theta \mid X = x)$$

Rule 1 :- 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 $(X|\theta) \sim B(n, \theta)$, $\theta \sim \beta_I(\alpha, \beta)$, $L(\theta, a) = (\theta - a)^2$. Find Bayes estimator of θ under the given loss function.

known

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

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

& $0 \le \theta \le 1$

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

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

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

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

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

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

PAGE NO.:
DATE: / /

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

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

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

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

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

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

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

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

= Mean of Conditional $dist^n$ of $\theta | X = x$

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

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

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

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

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 $\tau(\theta)$.

PAGE NO.:
DATE: / /

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

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

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

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

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

Let $f(x)$ be a pdf of X & $g(x)$ be any other funct^n of X, then
$E[g(x)] = \int_x g(x)\cdot f(x)\, dx$

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

Scanned with OKEN Scanner

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

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

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

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

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

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

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

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

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

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

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

<u>Ex:</u> If $L(\theta, a) = |\theta - a|$ = absolute error loss

Then $d(x) = a$ = median of dist^n $(\theta \mid X=x)$

= median of Posterior dist^n $\tau_1(\theta \mid X=x)$

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

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

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

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

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

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

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

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

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

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

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

PAGE NO. :
DATE : / /

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

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

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

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

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

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

(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 $\Theta = \mathcal{A} = R$, $L(\theta, a) = (\theta - a)^2$.
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

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

PAGE NO.:
DATE: / /

It 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 $x = h(x) = \int_\theta f_1(x,\theta)\,d\theta$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

PAGE NO.:
DATE: / /

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

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

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

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

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

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

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

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

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

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.

(long term behaviour so we use limit)

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: / /

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

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

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

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

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
NOTE
Def^n

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

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

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

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

$\Theta \backslash D$	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 ~~tha~~ 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.

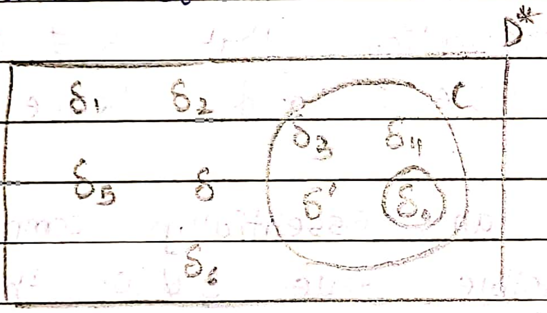
d_6 is better than $d_8 \Rightarrow d_8$ is inadmissible
$\nexists$ any rule better than d_6 and d_9
$\Rightarrow d_6$ and d_9 are admissible.

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 complete 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 essentially complete 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

PAGE NO.:
DATE: / /

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 $A \subset B$

$\Rightarrow \delta$ 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 defn 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 $\theta \in \Theta$.
$\Rightarrow \delta'$ is better than δ

which means δ is inadmissible which is a contradiction.

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

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

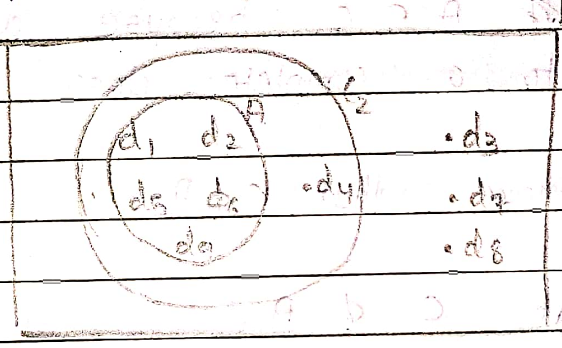
$\Rightarrow \delta'$ is equivalent to δ.

Def^n (4): A class C of decision rules is said to be "minimal complete" if

(i) C is complete and

(ii) no proper subset of C is complete.

Ex:



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 "minimal essentially complete"

(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

$\delta_1 > \delta_0$

$\delta_1 > \delta$

$\delta > \delta_0$ $\Rightarrow$ $\delta_1 > \delta_0$

$\delta_1 > \delta > \delta_0$

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.

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

$\Rightarrow$ a proper ~~subset~~ (subject) C_1 of C is complete which violates the second ~~condn~~ 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 subclass, 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.

21

(ii) To prove that $\because$ 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

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

PAGE NO.:
DATE: / /

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

Suppose we take prior $dist^n$ of θ.

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

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 $dist^n$ $\tau(\theta)$.

* Important theorems on decision theory & Game Theory :-

Recall Θ^* = set of all prior $dist^{ns}$ = set of all prob. dist defined over Θ

PAGE NO.:
DATE: / /

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

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

(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,

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

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

(2) represents the 'minimax' value 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 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 :- $\sup_{\tau \in \Theta^*} r(\tau, \delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$

Proof :- Let Θ^* = set of all prior $dist^n$ defined over Θ.
ω = set of all degenerate $dist^n$ defined over Θ.
Then,
$$\omega \subset \Theta^*$$
$$\Rightarrow \sup_{\tau \in \omega} r(\tau, \delta) \le \sup_{\tau \in \Theta^*} r(\tau, \delta) \quad \text{--- (1)}$$

PAGE NO.:
DATE: / /

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

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

Thus (1) and (2)

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

Now consider

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

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

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

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

$$\Rightarrow r(\tau, \delta) \leq \sup_{\theta \in \Theta} R(\theta, \delta) \quad \forall \tau \in \Theta^*$$
$$\left(\because \sum_j p_j = 1\right)$$

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

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

$$\Rightarrow r(\tau, \delta) \le \sup_{\theta \in \Theta} R(\theta, \delta) \quad \forall \; \tau \in \Theta^*$$
$$\left(\because \sum_j p_j = 1\right)$$

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

Hence (3) and (5)

$$\Rightarrow \sup_{\tau \in \Theta^*} \underbrace{r(\tau, \delta)}_{\text{Bayes risk}} = \sup_{\theta \in \Theta} \underbrace{R(\theta, \delta)}_{\text{Randomise}}$$

$\boxed{25 | 8}$

Thm :- If minimax thm holds and if $\exists$ least favourable prior distⁿ τ_0, then any minimax dec. rule δ_0 will be Bayes w.r.to prior distⁿ τ_0.

Proof :- Given minimax thm holds, therefore

$$\sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta) = \inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta) \quad \text{--- (1)}$$

Also given that τ_0 is least favourable prior therefore

$$\inf_{\delta \in D^*} r(\tau_0, \delta) = \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta) \quad \text{--- (2)}$$

Date :

Page No. :

Let δ_0 be a minimax dec rule

$$\Rightarrow \sup_{\tau \in \Theta^*} R(\theta, \delta_0) = \sup_{\tau \in \Theta^*} r(\tau, \delta_0) = \inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta)$$

$\hookrightarrow$ (3)

To P.T δ_0 is Bayes w.r.to prior τ_0 we can P.T $r(\tau_0, \delta_0) = \inf_{\delta \in D^*} r(\tau_0, \delta)$

Consider $r(\tau_0, \delta_0) \geq \inf_{\delta \in D^*} r(\tau_0, \delta)$

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

$(\because (2))$

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

$(\because (1))$

$$= \sup_{\tau \in \Theta^*} r(\tau, \delta_0) \quad (\because (3))$$

$$\geq r(\tau_0, \delta_0)$$

Hence $r(\tau_0, \delta_0) = \inf_{\delta \in D^*} r(\tau_0, \delta)$

($\because$ equality must hold, because $a > a$ is not possible, that is $a = a$)

Hence δ_0 is Bayes rule.

* **The complete class thm:-**
(The fundamental thm of dec theory)

Que:- When are Bayes rule admissible?

Thm-1:- If for a given prior dist^n τ, a Bayes rule w.r.to τ is unique upto equivalence, then this Bayes rule is admissible.

proof:- Suppose δ_0 is given Bayes rule w.r.to prior dist^n τ and is unique upto equivalence

Then to P.T δ_0 is admissible.

Suppose δ_0 is inadmissible

Then $\exists$ a dec. rule, say δ_1, which is better than δ_0, that is

$$R(\theta, \delta_1) \leq R(\theta, \delta_0) \ \forall \ \theta \in \Theta$$

and $R(\theta, \delta_1) < R(\theta, \delta_0)$ for atleast one $\theta \in \Theta$

Let $\Theta = \{\theta_1, \theta_2 - - - \theta_k\}$ (finite)

Then,

$$r(\tau, \delta_1) \leq r(\tau, \delta_0) \quad \text{--- (2)}$$

Inequality hold in (2) if τ attaches positive prob. for those θ for which

inequality holds in (1)
If inequality hold, δ_0 is not Bayes rule.
If equality hold, both δ_0 and δ_1 are Bayes,
and δ_0 and δ_1 are equivalent, which is a ~~e~~ contradiction in either case.
— Hence, δ_0 must be admissible.

Thm-2: Assume that $\Theta = \{\theta_1, \theta_2 \ldots \theta_K\}$ and a Bayes rule δ_0 w.r.to the prior dist $\tau = (P_1, P_2 \ldots P_K)$ exists.

If $P_j > 0 \ \forall \ j = 1, 2 \ldots K$ then δ_0 is admissible.

proof :- Suppose δ_0 is admissible.
Then $\exists$ a dec rule $\delta' \in D^*$ which is better then δ_0 that is

$$R(\theta_j, \delta') \leq R(\theta_j, \delta_0) \ \forall j$$
and
$$R(\theta_j, \delta') < R(\theta_j, \delta_0) \text{ for some } j$$

Since all $P_j > 0$.

$$\sum_{j=1}^{K} P_j R(\theta_j, \delta') < \sum_{j=1}^{P} P_j R(\theta_j, \delta_0)$$

$$\Rightarrow r(\tau, \delta') < r(\tau, \delta_0)$$

$\Rightarrow \delta_0$ is not Bayes w.r.to Prior.

$\tau = (P_1, P_2, \ldots P_K)$
This is contradiction to the statement that δ_0 is Bayes rule

Hence δ_0 is admissible.

e.g To demonstrate that if $P_j > 0$ $\forall$ $j = 1, 2 \ldots K$ is violated in thm-2 then δ_0 will not necessarily be admissible.

Let $\Theta = \{\theta_1, \theta_2\}$, $\mathcal{a} = \{a_1, a_2, a_3, a_4\}$
suppose the dist^n of r.v. X is degenerate at $x = 0$ $\forall$ θ

$\Rightarrow \mathcal{X} = \{x = 0\}$

$\Rightarrow P(X=0) = 1$ and $P(X \neq 0) = 0$

suppose the Loss table is $L(\theta, a)$

Θ \ $\mathcal{a}$	a_1	a_2	a_3	a_4
θ_1	1	1	2	2
θ_2	0	1	0	1

No. of non-randomized dec rule
$= (\text{no. of elements in } \mathcal{a})^{(\text{no. of elements in } \mathcal{X})}$

$= 4^1 = \boxed{4}$

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

Decision table

$x \setminus \mathcal{D}$	d_1	d_2	d_3	d_4
$x=0$	a_4	a_2	a_3	a_4

$L(\theta, d)$

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
$x=0$	$P(X=0) = 1$	1	0	1	1	2	0	2	1
	$R(\theta, d)$	1	0	1	1	2	0	2	1

* Risk table $R(\theta, d)$

	Θ	d_1	d_2	d_3	d_4
$P_1=1$	θ_1	1	1	2	2
$P_2=0$	θ_2	0	1	0	1
	$\max_\theta R(\theta,d)$	1	1	2	2

$\min_{d \in \mathcal{D}} \max_{\theta \in \Theta} R(\theta, d) = 1$, which corresponds to d_1 and d_2

$r(\tau_0, d_1) = E[R(\theta, z)] = 1$
$r(\tau_0, d_2) = 1$
$r(\tau_0, d_3) = 2$
$r(\tau_0, d_4) = 2$

d_2 is inadmissible
d_3 is inadmissible
d_4 is inadmissible
d_1 is admissible.

If τ_1; $P_1 = \frac{2}{3}$, $P_2 = \frac{1}{3}$

$\boxed{r(\tau_1, d_1) = \frac{2}{3}}$ → Bayes → admissible

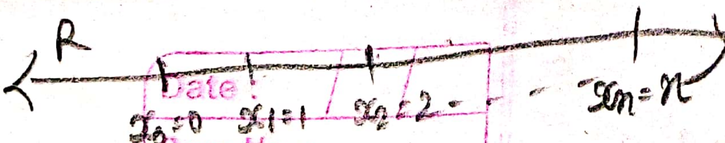
$r(\tau_1, d_2) = 1$

$r(\tau_1, d_3) = \frac{4}{3}$

$r(\tau_1, d_4) = \frac{5}{3}$

* Support of a distribution :-

1) Let $\Theta = E_1 = \mathbb{R}$ that is Θ is infinite
A point $\theta_0 \in E_1$ is said to be in the support of a dist^n τ on the real

e.g. 

line if for every $\varepsilon > 0$, the interval $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ has positive prob. that is $\tau(\theta_0 - \varepsilon, \theta_0 + \varepsilon) > 0$ or that is,

$$\text{if } P(\theta_0 - \varepsilon < \theta < \theta_0 + \varepsilon) > 0$$

2) If X is a r.v F is a $dist^n$ f^n, then $x_0 \in \mathcal{X}$ is said to be in support of $dist^n$ f^n F if

$$F(x_0 + \varepsilon) - F(x_0 - \varepsilon) > 0 \text{ for some } \varepsilon > 0$$

that is if $P(X \le x_0 + \varepsilon) - P(X \le x_0 - \varepsilon) > 0$ for $\varepsilon > 0$

that is if $P(x_0 - \varepsilon < x \le x_0 + \varepsilon) > 0$ for $\varepsilon > 0$.

e.g. Let $X \sim B(n, \theta)$, $0 < \theta < 1$

we know, $x = 0, 1, 2 \ldots n$

Let us take $x_0 = 0$, then

$$P(0 - \varepsilon < x \le 0 + \varepsilon) > 0$$

for $x_1 = 1$, $P(1 - \varepsilon < x \le 1 + \varepsilon) > 0$

and so on --

for $x_n = n$, $P(n - \varepsilon < x \le n - \varepsilon) > 0$

$\Rightarrow$ support of $B(n, \theta) = (0, 1, 2 \ldots n)$

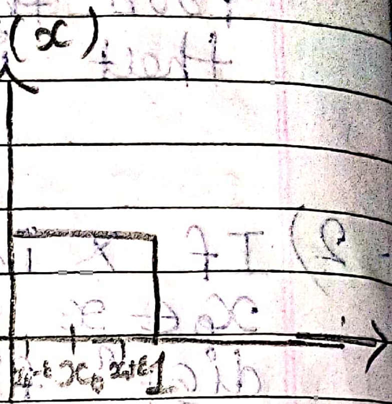
~~Let $X \sim U[0$~~

e.g. Let $X \sim U[0, 1]$

Then $\forall$ $x \in [0, 1]$

$P(x_0 - \varepsilon < x < x_0 + \varepsilon)$

$$= \int_{x = x_0 - \varepsilon}^{x_0 + \varepsilon} f(x)\,dx = [x]_{x_0 - \varepsilon}^{x_0 + \varepsilon}$$



$(\varepsilon > 0)$

$$\boxed{f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{o.w} \end{cases}}$$

$= x_0 + \varepsilon - x_0 + \varepsilon$

$= 2\varepsilon > 0$

$\Rightarrow$ support of $U[0, 1] = [0, 1]$

$\boxed{28 | 8}$ Thm 3 :- Let $\Theta = E_1 = R$ and assume that $R(\theta, \delta)$ is continuous function of θ for all $\delta \in D^*$. If δ_0 is Bayes rule w.r.to a prob. dist^n τ on the real line, for which $r(\tau, \delta_0)$ is finite and if the support of τ is the whole real line then δ_0 is admissible.

Proof :- Assume that δ_0 is not admissible.

Then it is possible to get a dec rule $\delta' \in D^*$ which is better than δ_0 that is,

$$R(\theta, \delta') \leq R(\theta, \delta_0) \quad \forall \; \theta \in \Theta$$

and $R(\theta, \delta') < R(\theta, \delta_0)$ for some $\theta \in \Theta$

Suppose the second condition hold. for some $\theta_0 \in \Theta$ that is

$$R(\theta_0, \delta') < R(\theta_0, \delta_0)$$

Since $R(\theta, \delta)$ is continuous in θ for all δ, $\exists \; \varepsilon > 0$ for which

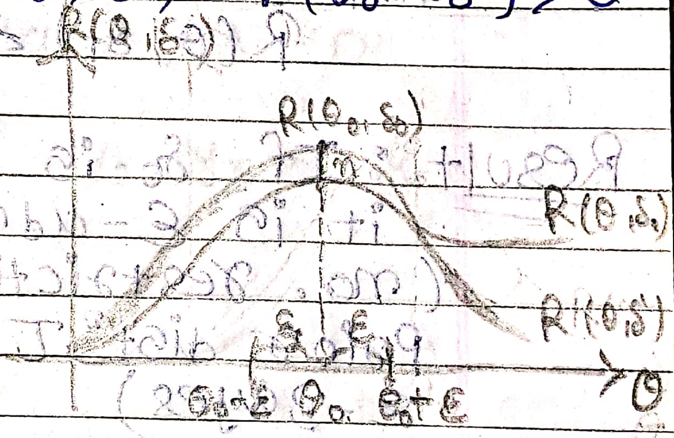
$$R(\theta, \delta') \leq R(\theta, \delta_0) - \frac{\eta}{2} \quad \text{whenever } |\theta_0 - \theta| < \varepsilon$$

where $\eta = R(\theta_0, \delta_0) - R(\theta_0, \delta') > 0$

Let T be a r.v whose dist^n is τ

Then, $r(\tau, \delta_0) - r(\tau, \delta')$

= difference bet^n Bayes risks w.r.to prior τ.



$$= E[R(T, \delta_0) - R(T, \delta')]$$

$$\geq \frac{\eta}{2} \, \tau(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \geq 0$$

(Since whole real line is support of τ $\Rightarrow \theta_0$ is also in support of τ)

~~Since~~
$\Rightarrow r(\tau, \delta_0) > r(\tau, \delta')$ which contradicts the statement that δ_0 is Bayes rule w.r.to τ

Hence δ_0 has to be admissible and hence the proof.

Def^n:- ε-admissible rule:-

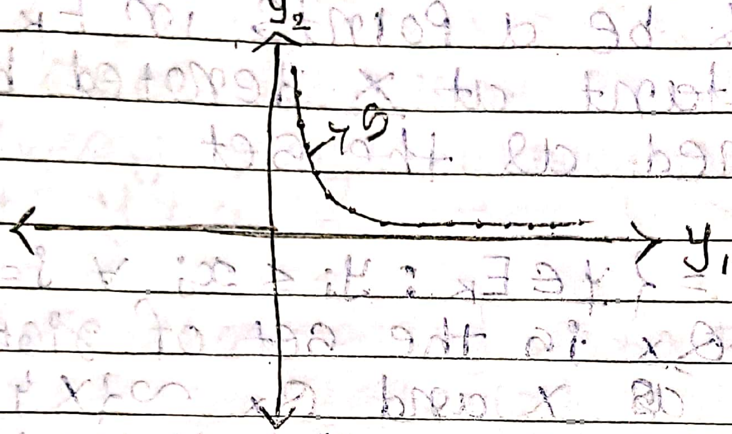
— A dec rule δ_0 is said to be ε-admissible dec rule, if there does not exist any dec rule, say δ_1 $\ni$

$$R(\theta, \delta_1) < R(\theta, \delta_0) - \varepsilon \quad \forall\ \theta \in \Theta$$

Result : If δ_0 is ε-Bayes with $\varepsilon > 0$, then it is ε-admissible
(no. restriction on Θ or on the prior dist^n τ w.r.to which δ_0 is ε-Bayes)

Def^n (1) :- A set S in K-dimensional space E_K is said to be ~~bainded~~ bounded from below if $\exists$ a finite no. say M s.t for every $y = (y_1, y_2 -- y_K) \in S$
$$y_j > -M \text{ for } j = 1, 2 - K$$

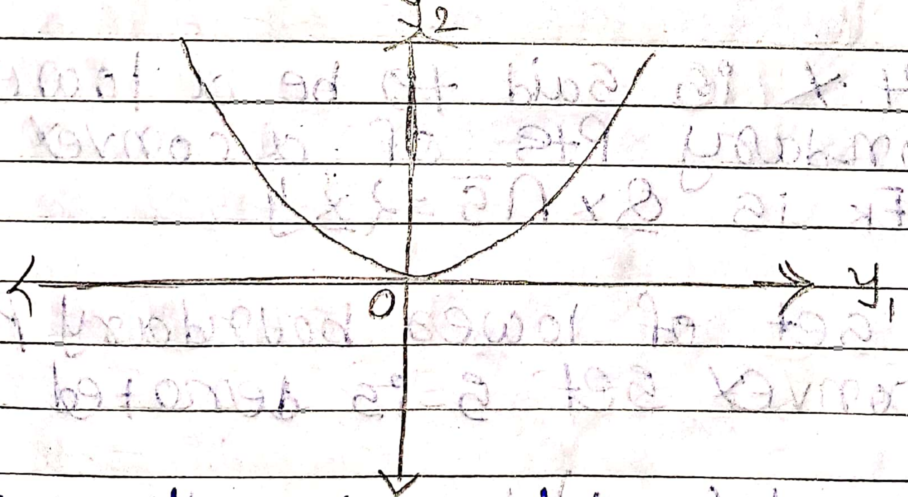
e.g. Consider E_2 and Let $S \subset E_2$ defined as
$S = \{(y_1, y_2),\ y_1 y_2 = 1,\ y_2 > 0\}$



Bounded Below from 0

This set S will be bounded for below and $y_1 > 0$
That is, here $y_1, y_2 > 0$
hence we can take $M = 0$

e.g. Let $S = \{(y_1, y_2),\ y_1^2 = y_2,\ y_2 > 0\}$
consider E_2.



Not set Bounded below

Here $y_2 > 0$ but there is no lower limit on y_1 ##
$\Rightarrow$ The set S is not bounded from below

15/9 $\bar{S} = S$ or limit ke sare closer Point

Def-2 Lower Quantant

Let x be a points in E_K. The Lower Quantant at x denoted by Q_x is defined as the set

$$Q_x = \{ y \in E_K : y_j \leq x_j \ \forall \ j = 1, 2 \ldots K \}$$

Thus, Q_x is the set of risk Pts as good as x and $Q_x \sim \{x\}$ is the set of risk Pts better than x

Def-3 Let $\bar{S}$ denote the closure of the set S, so that $\bar{S}$ is the union of S and the set of all limit Pts of S or alternative $\bar{S}$ is the smallest closed set containg S.

A Pt x is said to be a lower boundary Pts of a convex set $S \subset E_K$ is $Q_x \cap \bar{S} = \{x\}$

The set of lower boundary Pts of a convex set S is denoted by $\lambda(S)$

(Recall: Let (X, d) be a metric space. Let $E \subset X$ and Let d be a distance measure.

i) A neighbourhood of a pt $x \in X$ is a set
$N_r(x) = \{ y \in X : d(x,y) < r \}$

ii) A pt $x \in X$ is a limit pt of set E if every neighbourhood of x contains a pt $y \neq x$ s.t. $y \in E$

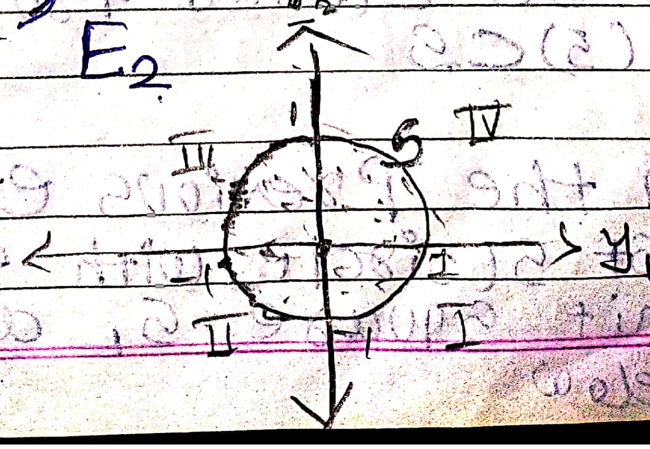
That is for every $r > 0$ $\exists$ a pt $y \in E$ $y \neq x$ st $y \in N_r(x)$

iii) Set $E \subset X$ is closed if every limit pt of E is a pt of E.

iv) E is bounded if $\exists$ a real no. M and a pt $y \in X$ s.t. $d(x,y) < M$ $\forall$ $x \in E$

v) Let E' be the set of all limit pts of E in X
Then $\bar{E} = E \cup E'$ is called the closure of set E

e.g. Consider E_2



Let $S = \{(y_1, y_2), y_1^2 + y_2^2 \le 1\}, S \subset E_2$

$\lambda(S) = \{(y_1, y_2) : y_1^2 + y_2^2 = 1, y_1 \le 0, y_2 \le 0\}$

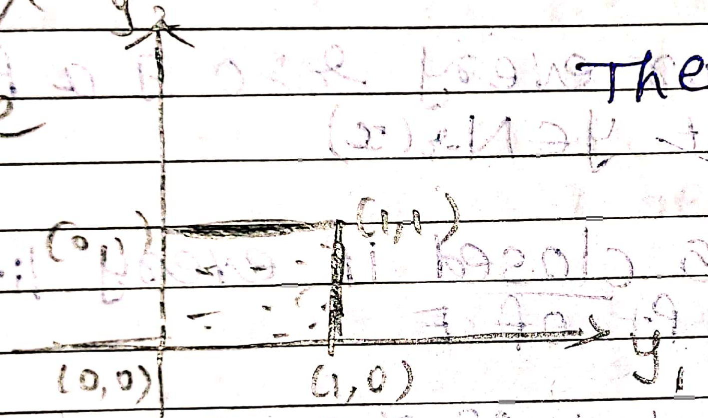
= set of boundary Pts in second ~~quadrant~~ quadrant

e.g E_2 and $S_1 = \{(y_1, y_2) : 0 \le y_1 \le 1, 0 \le y_2 \le 1\}$

; $S \subset E_2$

The $\lambda(S_1) = \{(0, 0)\}$

Unit square



* **Importance of set $\lambda(S)$:-**

- The elements of $\lambda(S)$ lead to the admissible dec rule.

Def-4 A convex set $S \subset E_K$ is said to be closed from below if $\lambda(S) \subset S$

e.g In the previous examples the unit dist S (circle with radius 1) and unit square S_1 are closed from below

Note : Any closed convex set is closed from below.

Thm: 1 :- If $\underline{x}(R(\theta, \delta_0) ---, R(\theta_K, \delta_0))$ is in $\lambda(S)$ then δ_0 is admissible.

Proof:- Let $\underline{x}$ be a risk Pt ~~correspor~~ corresponding to decision rule δ_0 and $\underline{x} \in \lambda(S)$

Suppose δ_0 is inadmissible

$\Rightarrow \exists$ a decision rule say δ_1 s.t

$$\left.\begin{array}{l} R(\theta_j, \delta_1) \leq R(\theta_j, \delta_0) \;\; \forall \; j = 1, 2 -- K \\ \text{and } R(\theta_j, \delta_1) < R(\theta_j, \delta_0) \text{ for some } j \end{array}\right\} \text{(1)}$$

Let $y_j = R(\theta_j, \delta_1)$, $j = 1, 2 -- K$

Then $\underline{y} = (y_1, y_2 -- y_K) \in S$, where S = Risk set

$\Rightarrow \underline{y} \in \bar{S}$

Also,

(1) $\Rightarrow y_j \leq x_j \;\; \forall \; j = 1, 2 -- K$

$y_j < x_j$ for some j

$\Rightarrow \underline{y} \in Q_x$ and $\underline{y} \neq \underline{x}$

Thus

$Q_x \cap \bar{S} = \{\underline{y}\}$

$\Rightarrow \underline{y}$ is a lower boundary Point of S

$\Rightarrow \underline{x}$ is not a lower boundary Pt of S

$\Rightarrow \underline{x}$ and $\lambda(S)$, which is a contradiction

Hence δ_0 must be admissible.

* Partial converse :-

If δ_0 is admissible and $\underline{x} = (R(\theta_1, \delta_0), \ldots, R(\theta_k, \delta_0))$, then $\underline{x} \in \lambda(S)$ if S is closed

Proof :- suppose $\underline{x} \notin \lambda(S)$, S is closed

$\Rightarrow \underline{x}$ is not a lower boundary pt of S

$\Rightarrow \exists\ \underline{y} \neq \underline{x}$ and $\underline{y} \in Q_x \cap \bar{S}$ (i.e $\underline{y}$ is a lower boundary pt)

$\Rightarrow \underline{y} \in \bar{S}$ and $\underline{y} \in Q_x$

$\Rightarrow y_j \leq x_j\ \forall\ j = 1, 2, \ldots k$ and $y_j < x_j$ for some j

If $\exists$ some decision rule δ_1 for which $R(\theta_j, \delta_1) = y_j$, then δ_1 is better than δ_0

$\Rightarrow \delta_0$ is inadmissible, which is a contradiction

Hence $\underline{x} \in \lambda(S)$.

Date : / /

Page No. :

* Existence of Bayes rule in this convex context

Assume that Θ is finite,

$\Theta = \{\theta_1, \theta_2 --- \theta_K\}$

Thm: 1 : Suppose $\Theta = \{\theta_1, \theta_2 -- \theta_K\}$ and the risk set S is bounded from below and closed from below. Then for every prior dist $(P_1, P_2 -- P_K)$ for which $P_j > 0 \; \forall \; j = 1, 2 -- K$ a Bayes rule w.r.to $(P_1, P_2 -- P_K)$ exists.

Proof :- Let $(P_1 -- P_K)$ be a prior dist over Θ for which $P_j > 0 \; \forall \; j$ Let

b = Bayes risk of a decision rule whose risk pt y, $y \in S$, w.r.to prior distⁿ $(P_1, P_2 -- P_K)$

$$= \sum_{j=1}^{K} P_j y_j$$

and

$B = \{b ; \; b = \sum_{j=1}^{K} P_j y_j , \; y \in S\}$ = set of all bayes risks

Since S is bounded from below B is bounded from below

Let b_0 = glb of $\mathcal{B}$

Consider a seq of pts $\underline{y}^{(n)} \in S$ s.t

$$\sum_{j=1}^{k} p_j y_j^{(n)} \longrightarrow b_0$$

Since $p_j > 0 \ \forall j$, the seq $\underline{y}^{(n)}$ is bounded above also.

Thus seq $\underline{y}^{(n)}$ is bounded

($\because$ bounded above and below both)

seq $\underline{y}^{(n)}$ has limit pt.

Let $\underline{y}^{(0)}$ be the limit pt of seq $\underline{y}^{(n)}$

Then Bayes risk corres to $\underline{y}^{(0)}$

$$= b_0 = \sum_{j=1}^{n} p_j y_j^{(0)}$$

Now, $\underline{y}^{(0)} \in Q_{y^{(0)}}$ and $\underline{y}^{(0)} \in \bar{S}$

$$\Rightarrow \{\underline{y}^{(0)}\} \subset Q_{\underline{y}^{(0)}} \cap \bar{S}$$

If

$$\underline{y}' (\neq \underline{y}^{(0)}) \in Q_{\underline{y}^{(0)}} \cap \bar{S},$$

then $\underline{y}' \in Q_{\underline{y}^{(0)}}$ and $\underline{y}' \in \bar{S}$

$$\Rightarrow y_j' \leq y_j^{(0)} \ \forall j$$

and $y_j' < y_j^{(0)}$ for some j

$\Rightarrow \underline{y}^{(0)}$ is not a limit pt of seq $\underline{y}^{(n)}$, which is contradiction.

$\therefore \{\underline{y}^{(0)}\}$ is the only pt in $Q_{y^{(0)}} \cap \bar{S}$

$\Rightarrow Q_{\underline{y}^{(0)}} \cap \bar{S} = \{\underline{y}^{(0)}\}$

$\Rightarrow \underline{y}^{(0)} \in \lambda(S)$

$\Rightarrow \underline{y}^{(n)} \in \bar{S}$ and since S is closed from below, $\exists$ δ_0 whose risk pt in $y^{(0)}$, and $b_0 = \sum_{j=1}^{k} p_j y_j^{(0)}$

$\Rightarrow \delta_0$ is Bayes rule w.r.to given prior dist $(P_1, P_2 \ldots P_K)$, where $\boxed{P_j > 0 \; \forall j.}$

Result:- If the risk set S is closed from below and bounded (above as well as below), then Bayes rule w.r.to all prior dist exists.

Lemma:- If a non-empty convex set S is bounded from below, then $\lambda(S)$ is not empty.

18/9

$\lambda(S)$ = lower counter or closure bounded

* Existence of a minimal complete class.

Thm :- Suppose $\Theta = \{\theta_1, \theta_2 \ldots \theta_k\}$, and the risk set S is bounded from below a closed from below. Then the class of dec rules $D_0 = \{$ $\ni$

$$D_0 = \{\delta \in D^* : (R(\theta_1, \delta), \ldots R(\theta_k, \delta)) \in \lambda(S)\}$$

Proof :- First we shall P.T D_0 is complete class

Let δ be any decision rule s.t. $\delta \notin D_0$ but $\delta \in D^*$

Let S = set of risk pts corresponding to decision rules $\delta \notin D_0$

That is $\underline{x} = (R(\theta_1, \delta), \ldots R(\theta_k, \delta)) \in S$ but $\underline{x} \notin \lambda(S)$ since $\delta \notin D_0$

Let $S_1 = Q_x \cap \bar{S}$ then S_1 is non-empty convex set (because closure of a convex set is convex and the intersection of two convex sets is also convex), and bounded from below then using previous lemma, that is if a non-empty convex set S is bounded from below then $\lambda(S)$ is non-empty, $\lambda(S_1)$ is not empty

$\underline{Q_x} \Rightarrow y_i \leq x_i \rightarrow \delta_0$
δ (as good as)
Risk point ~~these~~ point
lower bounded hote he

$\overline{A \cap B}$
$= \overline{A} \cap \overline{B}$

Let $\underline{y} \in \lambda(S_1)$, then $\{\underline{y}\} = Q_y \cap \bar{S}_1$
(by def^n of $\lambda(S_1)$)
Further $\underline{y} \in Q_x$ because $\underline{y} \in \bar{S}_1$
$\Rightarrow \underline{y} \in Q_x \cap \bar{S}$
$\Rightarrow \underline{y} \in \bar{Q}_x \cap S$ $\left[\begin{array}{l} \because S_1 = Q_x \cap \bar{S} \\ Q_x \cap \bar{S} = \bar{Q}_x \cap \bar{S} \\ = \bar{Q}_x \cap S \end{array} \right]$
$\Rightarrow \underline{y} \in \bar{Q}_x$
$\Rightarrow \underline{y} \in Q_x$
Also, $\underline{y} \in \lambda(S)$ because $\{\underline{y}\} = Q_y \cap \bar{S}$
$= Q_y \cap S$
Thus because S is closed from below $\exists$ a dec. rule $\delta_0 \in D_0$ for which $\underline{y} = (R(\theta_1, \delta_0), \ldots (\theta_K, \delta_0))$ and it is better than δ since $\underline{y} \in Q_x \sim \{\underline{x}\}$ better to $\lambda(S)$
This means D_0 is complete
Now, using thm if $\underline{x} = (R(\theta_1, \delta_0) \ldots R(\theta_K, \delta_0)) \in \lambda(S)$,
Then δ_0 is admissible, therefore δ_0 is admissible $(\because y \in \lambda(S))$
$\Rightarrow$ Every rule in D_0 is admissible
Hence no proper subset of D_0 can be complete because every complete class must contain all admissible rules.

$\Rightarrow D_0$ is minimal complete class Hence the proof.

* Separating Hyperplane Theorem

Lemma 1:- If S is closed convex subset of E_K and $0 \notin S$, then $\exists$ a vector $\underline{p} \in E_K$ s.t $\underline{p}'\underline{x} > 0$ for all $\underline{x} \in S$.

Lemma-2:- If S is convex subset of E_K, A is an open subset of E_K and $A \subset \bar{S}$ then $A \subset S$.

Thm-1:- If S is a convex subset of E_K and $\underline{x}_0$ is not an interior pt of S (that is either $\underline{x}_0 \notin S$ or $\underline{x}_0$ is boundary pt of S), then $\exists$ a vector $\underline{p} \in E_K$, $\underline{p} \neq 0$, such that $\underline{p}'\underline{x} \geq \underline{p}'\underline{x}_0$ for all $\underline{x} \in S$.

Thm 2: Separating Hyperplane theorem

Let S_1 and S_2 be disjoint convex subsets of E_K then $\exists$ a vector $\underline{p} \neq 0$ s.t $\underline{p}'\underline{y} \leq \underline{p}'\underline{x}$ for all $\underline{x} \in S_1$ and all $\underline{y} \in S_2$

22/9

Date : / /
Page No. :

Result: For any decision problem (Θ, D, R), $\underline{V} \le \overline{V}$

Proof:- $r(\tau, \delta) \le \sup_{\tau' \in \Theta^*} r(\tau', \delta) \quad \forall\ \delta \in D^*$ and $\forall\ \tau \in \Theta^*$ — (1)

Also

$$\inf_{\delta' \in D^*} r(\tau, \delta') \le r(\tau, \delta) \quad \forall\ \tau \in \Theta^* \text{ and } \delta \in D^* \text{ — (2)}$$

$\because$ (1) and (2) $\Rightarrow \inf_{\delta' \in D^*} r(\tau, \delta') \le \sup_{\tau' \in \Theta^*} r(\tau', \delta)$

$\forall\ \delta \in D^*$ and $\forall\ \tau \in \Theta^*$

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

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

$\Rightarrow \underline{V} \le \overline{V}$

(maximin) $\le$ (minimax)

* Minimax Theorem :-

Statement:- If for a given decision problem (Θ, D, R) with finite $\Theta = \{\theta_1, \theta_2 \ldots \theta_k\}$ the risk set S is bounded below then

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

That is $\underline{V} = \bar{V} = V$

$$\left(\begin{array}{ccc} \text{maximin} & = & \text{minimax} \\ \text{value} & & \text{value} \end{array} \right)$$

and $\exists$ a least favourable prior dist^n τ_0

(Additionally) Moreover, if S is closed from below, then $\exists$ an admissible minimax decision rule δ_0 which is Bayes w.r.to prior τ_0

Example:- This example demonstrates that if Θ is not finite, then minimax theorem does not necessarily hold.

Let $\Theta = \mathcal{A} = \{1, 2, 3, \ldots\} \rightarrow$ not finite

and $L(\theta, a) = \begin{cases} 1 & \text{if } a < \theta \\ 0 & \text{if } a = \theta \\ -1 & \text{if } a > \theta \end{cases}$

Then

Loss Table:-

$\Theta \backslash \mathcal{A}$	a_1 1	a_2 2	a_3 3	a_4 4	$\inf r(\tau, \delta)$
$\theta_1 = 1$	0	-1	-1	-1	-1
$\theta_2 = 2$	1	0	-1	-1	-1
$\theta_3 = 3$	1	1	0	-1	-1
$\theta_4 = 4$	1	1	1	0	-1
$\vdots$					
$\sup r(\tau, \delta)$	1	1	1	1 ...	

$\rightarrow$ sup

$\hookrightarrow$ inf

Suppose observable r.v X is degenerate at $X=0$. that is, $P_\theta(X=0)=1 \ \forall \ \theta \in \Theta$
Then $\mathcal{X}=\{0\}$ = sample space.

$\mathcal{D}$ = set of non-randomized decision rules
$=\{d_1, d_2, d_3 \ldots\}$
($\because$ no. of non-randomized decision rules $=(\text{no. in } a)^{(\text{no. in } \mathcal{X})}=\infty^1=\infty$)
that is $\mathcal{D}=\{d_i(0)=i, \ i=1,2,\ldots\}$
Then $R(\theta_i, d_i)=E_\theta[L(\theta_i, d_i)]$
$$=\sum_{i=1}^{\infty} L(\theta_i, d_i) \cdot p_i$$
$$=L(\theta_i, d_i)$$
($\because p_i=1$ as X is degenerate)

That is the risk table will be same as the loss table

$\therefore$ From the table $\sup_\tau r(\tau,\delta)=1 \ \forall \ \delta$ and

$\inf_\delta r(\tau,\delta)=-1 \ \forall \ \tau$

$\therefore \underline{V}=\sup_\tau \inf_\delta r(\tau,\delta)=-1$

and $\overline{V}=\inf_\delta \sup_\tau r(\tau,\delta)=1$

Thus $\underline{V} \neq \overline{V}$

Theorem 1:- If δ is admissible and Θ is finite, then δ is Bayes rules w.r.to some prior distribution.

Theorem 2:- complete class theorem:-

If, for a decision problem $(\Theta, \mathcal{D}, R)$ with finite Θ, the risk set S is bounded from below and closed from below, then the class of all Bayes rules is complete and the admissible Bayes rules from a minimal complete class

example:- Suppose $\Theta = \left\{\frac{1}{3}, \frac{2}{3}\right\}$, $\mathcal{A} = R$,

$L(\theta, a) = (\theta - a)^2$ suppose a coin is tossed with prob. θ of getting a head. Find a Bayes rule w.r. to prior $dist^n$ τ which assigns prob π to $\theta = \frac{1}{3}$ and $(1-\pi)$ to $\theta = \frac{2}{3}$

$(0 \leq \pi \leq 1)$

Sol^n: Here, $\mathcal{X} = \{H, T\}$

$\therefore$ No. of non-randomized decision rules $= \infty$

Date :

Page No. :

Let us define a general non-randomized decision rule, $d(H) = x$, $d(T) = y$.

$(x, y) = R^2$ $\quad x, y \in \mathcal{A}$ (where $\mathcal{A} = R$)

$\hookrightarrow$ combination of x and y

$\therefore D = \{(x, y), x, y \in R, d(H) = x, d(T) = y\}$
= set of all non-randomized dec. rules.

Then

$$R(\theta, (x, y)) = E[L(\theta, (x, y))]$$

or $E_\theta[L(\theta, d)]$

$$= \sum_{a \in (x,y)} (\theta - a)^2 P_\theta(d)$$

$$= (\theta - x)^2 P_\theta(H) + (\theta - y)^2 P_\theta(T)$$

$$= (\theta - x)^2 \theta + (\theta - y)^2 (1 - \theta)$$

(assuming $P_\theta(H) = \theta$
$\Rightarrow P_\theta(T) = 1 - \theta$)

$$\therefore R\left(\theta = \frac{1}{3}, (x, y)\right) = \left(\frac{1}{3} - x\right)^2 \left(\frac{1}{3}\right) + \left(\frac{1}{3} - y\right)^2 \left(\frac{2}{3}\right)$$

and $$R\left(\theta = \frac{2}{3}, (x, y)\right) = \left(\frac{2}{3} - x\right)^2 \left(\frac{2}{3}\right) + \left(\frac{2}{3} - y\right)^2 \left(\frac{1}{3}\right)$$

Now $r(\tau, (x, y)) = E_\theta[R(\theta, z)]$

$$= \pi \cdot R\left(\theta = \frac{1}{3}, (x, y)\right) + (1 - \pi) \cdot R\left(\theta = \frac{2}{3}, (x, y)\right)$$

$$= \pi\left[\frac{1}{3}\left(\frac{1}{3}-x\right)^2+\left(\frac{2}{3}\right)\left(\frac{1}{3}-y\right)^2\right]$$
$$+(1-\pi)\left[\frac{2}{3}\left(\frac{2}{3}-x\right)^2+\frac{1}{3}\left(\frac{2}{3}-y\right)^2\right]$$

we want to find Bayes rules that is a rule for which Bayes risk is minimum.

Thus, for $r(\tau,(x,y))$ to be minimum, we want to determine x and y which minimizes $r(\tau,(x,y))$.

$$\therefore \frac{\partial}{\partial x} r(\tau,(x,y)) = 0 \text{ and } \frac{\partial}{\partial y} r(\tau,(x,y)) = 0$$

$$\Rightarrow \pi\left(\frac{1}{3}\cdot 2\left(\frac{1}{3}-x\right)(-1)\right)+(1-\pi)\left[\frac{2}{3}\cdot 2\left(\frac{2}{3}-x\right)(-1)\right]=0$$

$$\Rightarrow -\frac{2}{3}\left(\frac{1}{3}-x\right)\pi-\frac{4}{3}\left(\frac{2}{3}-x\right)(1-\pi)=0$$

$$\Rightarrow -\frac{2}{9}\pi+\frac{2}{3}x\pi-\frac{8}{9}+\frac{4}{3}x+\frac{8}{9}\pi-\frac{4}{3}x\pi=0$$

$$\Rightarrow -2\pi-6x\pi+8-12x+8\pi+12\pi x=0$$

(Lcm by -9)

$$\Rightarrow -6\pi+6x\pi-12x+8=0$$

$$\Rightarrow 6x(\pi-2)=6\pi-8$$

Date :

Page No. :

$\Rightarrow 3x(\pi-2) = 3\pi-4$ ($\because$ Divide by 2)

$\Rightarrow x = \frac{3\pi-4}{3(\pi-2)} = \boxed{\frac{4-3\pi}{6-3\pi}}$ ($\because$ N and P multiply by -1)

$$\frac{\partial}{\partial y} r(\tau,(x,y)) = 0$$

$$\Rightarrow \pi\left[\frac{2}{3}\cdot 2\left(\frac{1}{3}-y\right)(-1)\right] + (1-\pi)\left[\frac{1}{3}\cdot 2\left(\frac{2}{3}-y\right)(-1)\right] = 0$$

$$\boxed{y = \frac{2}{3(1+\pi)}}$$

5|9

e.g A coin with unknown prob. of head θ is tossed once to estimate θ. If the loss f^n is $L(\theta,a) = (\theta-a)^2$ find Bayes rule w.r.to a prior dist^n τ. Show that the Bayes rule w.r.to prior τ' having same first two moments as τ are same

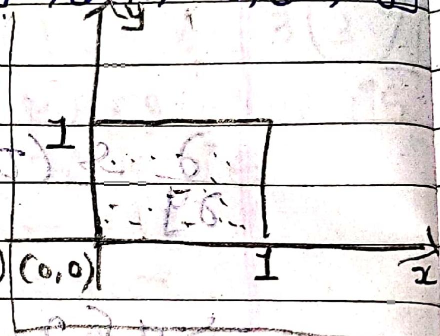
Sol^n :- Here $\Theta = \mathcal{A} = [0,1]$, $L(\theta,a) = (\theta-a)^2$, $\mathcal{X} = \{H,T\}$

general non-randomized decision rule

can be defined as

$$d(H) = x,\ d(T) = y \quad \therefore\ 0 \le x, y \le 1$$

and $D = \{(x,y) : x, y \in [0,1],\ d(H) = x,\ d(T) = y\}$



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

$$= \sum_{a} (\theta-a)^2 \times prob.$$

(Here $a = x$ or y)

$$= (\theta-x)^2 P(H) + (\theta-y)^2 P(T)$$

$$= (\theta-x)^2\,\theta + (\theta-y)^2(1-\theta)$$

Bayes risks w.r.to Prior τ

$$\Rightarrow r(\tau,(x,y)) = E[R(\theta,(x,y))]$$

$$= E[(\theta-x)^2\theta + (\theta-y)^2(1-\theta)]$$

$$= E_\theta[\theta^3 - 2\theta^2 x + x^2\theta + \theta^2 - 2\theta y + y^2 - \theta^3 + 2\theta^2 y - \theta y^2]$$

$$\cancel{= E_\theta[(1-2x+2y)]\theta^2}$$

$$= E[(1-2x+2y)\theta^2 + (x^2-y^2-2y)\theta + y^2]$$

$$= (1-2x+2y)E(\theta^2) + (x^2-y^2-2y)E(\theta) + y^2$$

since τ is a prior $dist^n$ of θ.

Let m_1' = first row moment of $\tau = E(\theta)$

and m_2' = second row moment of $\tau = E(\theta^2)$

$\therefore \boxed{r(\tau,(x,y)) = (1-2x+2y)m_2' + (x^2-y^2-2y)m_1' + y^2}$

Therefore, (different)

the $diff^n$ priors say τ and τ', whose first two moments (m_1' and m_2') are same will have the same Bayes rules as $r(\tau,(x,y))$ only involves m_1' and m_2' in term of θ.

To find the Bayes rule, means to find value of x and y s.t. $r(\tau,(x,y))$ is minimum.

$\therefore \frac{\partial}{\partial x} r(\tau,(x,y)) = 0$ and

$$\frac{\partial}{\partial y} r(\tau,(x,y)) = 0$$

$\Rightarrow -2m_2' + 2xm_1' = 0$ and $2m_2' - 2ym_1' - 2m_1' + 2y = 0$

$\Rightarrow \boxed{x = \frac{m_2'}{m_1'}}$ and ($\therefore m_2' + m_1' = 2ym_1' - 2y$

$m_2' - m_1' = y(m_1' - 1)$)

$\boxed{y = \frac{m_2' - m_1'}{m_1' - 1}}$ $\cancel{y = \frac{(m_1' - 1)}{m_2' - 1}}$

* **Solving for minimax Rule :-**

Method (1) :-

Thm :- If δ_0 is Bayes rule w.r.to τ_0 and for all $\theta \in \Theta$,
$$R(\theta, \delta_0) \leq r(\tau_0, \delta_0) \longrightarrow (1)$$
then the game has a value, δ_0 is minimax rule and τ_0 is least favourable prior distn.

NOTE :- This thm means that, guess a least favourable prior distn τ_0 find a Bayes rule δ_0 w.r.to τ_0 if this δ_0 is equilizer rule, then it is minimax.

Thm :- If δ_n is Bayes w.r.to τ_n, $r(\tau_n, \delta_n) \rightarrow c$ and $R(\theta, \delta_n) \leq c$ $\forall\, \theta \in \Theta$ then the game has a value and δ_0 is minimax.

Method (2) :-

Lemma :- Suppose the game (Θ, D, R) with finite Θ has a value V and that a minimax rule δ_0 exist, then for any $\theta \in \Theta$ that ~~res~~ receive positive weight from any least favourable prior dist^n

$$R(\theta, \delta_0) = V = \text{value of the game}$$

$\Theta = \{\theta_1, \theta_2 \dots \theta_K\}$
$I \rightarrow p_1 \quad p_2 \dots p_K$
$\geq 0 \quad \geq 0 \quad \geq 0$

Thm:- If an equilizer rule is extended Bayes then it is a minimax rule.

Note: Thus, in second method we search for an equilizer rule in order to find a minimax rule

e.g Let $\Theta = \left\{\frac{1}{3}, \frac{2}{3}\right\}$, $\mathcal{A} = R$, $L(\theta, a) = (\theta - a)^2$

Suppose a coin is tossed once, for which prob. of head is θ, then find a minimax rule.

Sol^n: $D = \{(x, y) : x, y \in R, d(H) = x, d(T) = y\}$ decision set

$$R(\theta, (x, y)) = \sum_a L(\theta, a)$$

$$= (\theta - x)^2 P_\theta(H) + (\theta - y)^2 P(T)$$

Date :

Page No. :

$$= (\theta - x)^2 \theta + (\theta - y)^2 (1-\theta)$$
$$= (-2x+2y+1)\theta^2 + (x^2 - 2y - y^2)\theta + y^2$$

(same as last ex)

To find an equilizer rule
Consider a prior dist^n say $\tau_0 = \left(\frac{1}{2}, \frac{1}{2}\right)$.

$$r(\tau_0, (x,y)) = E[R(\theta, (x,y))].$$

$$= R(\theta_1, (x,y)) P_{\tau_0}(\theta_1) + R(\theta_2, (x,y)) P_{\tau_0}(\theta_2)$$

$$= R\left(\frac{1}{3}, (x,y)\right) \cdot \frac{1}{2} + R\left(\frac{2}{3}, (x,y)\right) \frac{1}{2}$$

$$= \frac{1}{2}\left[\frac{1}{9}(-2x+2y+1) + \frac{1}{3}(x^2-2y-y^2) + y^2 + \frac{4}{9}(-2x+2y+1) + \frac{2}{3}(x^2-2y-y^2) + y^2\right]$$

$$\cancel{= \frac{1}{2}\left[\frac{5}{9}(-2x+1\right.}$$

find the Bayes rule we need to minimize $r(\tau_0, (x,y))$ w.r.to x and y

$$\therefore \frac{\partial}{\partial x} r(\tau_0, (x,y)) = 0$$

$$\Rightarrow \frac{1}{2}\left[\frac{1}{9}(-2) + \frac{1}{3}(2x) + \frac{4}{9}(-2) + \frac{2}{3}(2x)\right] = 0$$

$\Rightarrow -2+6x-8+12x=0$

$\Rightarrow 18x = 10 \Rightarrow x = \frac{10}{18} = \boxed{\frac{5}{9}}$

$\frac{\partial}{\partial y} r(\tau_0; (x,y)) = 0$

$$\Rightarrow \frac{1}{2}\left[\frac{1}{9}(2) + \frac{1}{3}(-2-2y) + 2y + \frac{y}{9}(2) + \frac{2}{3}(-2-2y) + 2y\right] = 0$$

$\Rightarrow 2 - 6 - 6y + 18y + 8 - 12 - 12y + 18y = 0$

$\Rightarrow 18y = 8 \Rightarrow y = \frac{8}{18} = \boxed{\frac{4}{9}}$

The Bayes rule w.r. to Prior $\tau_0 = \left(\frac{1}{2}, \frac{1}{2}\right)$ is $\delta_0 = \left(\frac{5}{9}, \frac{4}{9}\right)$, that is $d(H) = \frac{5}{9}$, $d(T) = \frac{4}{9}$

Also, Bayes risk $= r(\tau_0; \delta_0) = E[R(\theta,(x,y))]$ (using (1))

$$= \frac{1}{2}\left[\frac{1}{9}\left(-2\left(\frac{5}{9}\right) + 2\cdot\frac{4}{9} + 1\right) + \frac{1}{3}\left(\frac{25}{81} - 2\left(\frac{4}{9}\right) - \left(\frac{16}{81}\right)\right) + \frac{16}{81} + \frac{4}{9}\left(-2\left(\frac{5}{9}\right) + 2\cdot\frac{4}{8} + 1\right) + \frac{2}{3}\left(\frac{25}{81} - \frac{8}{9} - \frac{16}{81}\right) + \frac{16}{81}\right] = \boxed{V = \frac{2}{81}}$$

Also $R(\theta_1, \delta_0) = R\left(\frac{1}{3}, \delta_0\right)$

$$= (\theta - x)^2 \theta + (\theta - y)^2 (1-\theta)$$

$$= \left(\frac{1}{3} - \frac{5}{9}\right)^2 \cdot \frac{1}{3} + \left(\frac{1}{3} - \frac{4}{9}\right)^2 \frac{2}{3}$$

$$= \boxed{\frac{2}{81}}$$

$$R(\theta_2, \delta_0) = R\left(\frac{2}{3}, \delta_0\right)$$

$$= \left(\frac{2}{3} - \frac{5}{9}\right)^2 \cdot \frac{2}{3} + \left(\frac{2}{3} - \frac{4}{9}\right) \frac{2}{3}$$

$$= \boxed{\frac{2}{81}}$$

e.g. $\Theta = \mathcal{A} = [0, 1]$, $L(\theta, a) = (\theta - a)^2$, $x = \{H, T\}$ (Tossing a coin once)

$\mathcal{D} = \{(x, y) ; 0 \le x, y \le 1, d(H) = x ; d(T) = y\}$

From prev ex, we known,

$$R(\theta ; (x, y)) = \theta(\theta - x)^2 + (1-\theta)(\theta - y)^2$$
$$= \theta^2(1 + 2y - 2x) + \theta(x^2 - y^2 - 2y) + y^2$$

Since $\Theta = [0, 1]$ is continuous it is ~~differ~~ difficult to guess a least favourable prior so we will try to find an equilizer rule.

For equilizer rule, $R(\theta,(x,y))$ must be indep^t of parameter θ so equate coefficients of θ^2 and θ is $R(\theta,(x,y))$ to zero.

$\Rightarrow 1+2y-2x=0$ and $x^2-y^2-2y=0$

$\Rightarrow 2x = 1+2y$

$\Rightarrow x = \frac{1+2y}{2}$

$\Rightarrow \left(\frac{1+2y}{2}\right)^2 - y^2 - 2y = 0$

$\Rightarrow 1+4y+4y^2-4y^2-8y=0$

$\Rightarrow 1-4y=0$

$\Rightarrow 1 = 4y$

$\Rightarrow \boxed{y = \frac{1}{4}}$

$x = \frac{1+2\left(\frac{1}{4}\right)}{2}$

$x = \frac{1+\frac{1}{2}}{2}$

$x = \frac{2+1}{4} = \boxed{\frac{3}{4}}$

equilizer rule :- $d(H) = \frac{3}{4}$, $d(T) = \frac{1}{4}$

and risk of equilizer rule $= R(\theta,(x,y))$

$= y^2$

$= \boxed{\frac{1}{16}}$

Now, suppose $x = \frac{m_0'}{m_1'}$ and $y = \frac{m_1'-m_0'}{1-m_1'}$

Scanned with OKEN Scanner

Date : / /
Page No. :

Que will the decision rule, $d(H) = x = \frac{3}{4}$ and $d(T) = y = \frac{1}{4}$ be a Bayes rule?

So, we have $x = \frac{3}{4} = \frac{m_2'}{m_1'}$

$$\Rightarrow m_1' = \frac{4m_2'}{3}$$

$$y = \frac{1}{4} = \frac{m_1' - m_2'}{1 - m_1'} \Rightarrow \frac{\frac{4m_2'}{3} - m_2'}{1 - \frac{4m_2'}{3}}$$

$$\Rightarrow \frac{4m_2' - 3m_2'}{3 - 4m_2'}$$

$$\Rightarrow \frac{m_2'}{3 - 4m_2'}$$

$$\Rightarrow 3 - 4m_2' = 4m_2'$$

$$\Rightarrow 3 = 8m_2'$$

$$\Rightarrow \boxed{m_2' = \frac{3}{8}}$$

$$m_1' = \frac{4\left(\frac{3}{8_2}\right)}{3} = \frac{3}{6} = \boxed{\frac{1}{2}} = m_1'$$

$Bin \rightarrow E(X) = np = m_1'$

mean variance

$\boxed{np > npq}$

$V(x) = npq = E(x^2) - (E(x))^2 = npq + n^2p^2$

$\Rightarrow V(X) = m_2 = m_2' - m_1'^2 = \frac{3}{8} - \frac{1}{4} = \boxed{\frac{1}{8}}$

$E(X) = m_1' = \boxed{\frac{1}{2}}$

From the values of m_1' and m_2' if we guess a prior dist on [0, 1] say for ex, $U[0,1]$ then $m_1' = \frac{1}{2}$ but $m_2' = \frac{1}{3} \neq \frac{3}{8}$

Suppose we assume prior dist^n $\beta_I(\alpha, \beta)$, with

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

Then

$$E(x^r) = \int_0^1 x^r \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\, dx$$

$$= \frac{B(r+\alpha, \beta)}{B(\alpha,\beta)}, \quad r = 1, 2 \ldots$$

$$E(X) = m_1' = \frac{B(\alpha+1, \beta)}{B(\alpha,\beta)} \quad (r=1)$$

$$= \frac{\alpha}{\alpha+\beta}$$

$$E(x^2) = m_2' = \frac{B(\alpha+2, \beta)}{B(\alpha,\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

Date : / /
Page No. :

Taking $\alpha = \beta = 1$ then $m_1' = \frac{1}{2}$, $m_2 = \frac{1}{3} \neq \frac{3}{8}$

Taking $\alpha = \beta = \frac{1}{2}$ then $m_1' = \frac{1}{2}$, $m_2' = \frac{3}{8}$

Thus decision rule $d(H) = x = \frac{3}{4}$ and

$d(T) = y = \frac{1}{4}$ is a

Bayes rules w.r. to Prior $B\left(\frac{1}{2}, \frac{1}{2}\right)$

29/9

e.g Let $X \sim B(n, \theta)$, $\Theta = [0, 1] = \mathcal{A}$,

$L(\theta, a) = (\theta - a)^2$ Let the prior dist^n

of θ be $g(\theta) = \dfrac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$, $0 < \theta < 1$

i) Find Bayes estimator of θ

ii) Does an equilizer rule exist?

(i) $\Rightarrow$ we known that ~~$d_{\alpha,\beta}$~~

$d_{\alpha,\beta}(X) = E[\theta | X = x]$ and the

Posterior dist^n $(\theta | X = x) \sim \beta_1(x+\alpha, n+\beta-x)$

$$d_{\alpha,\beta}(X) = E[\theta | X = x]$$

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

= conditional mean of Posterior dist^n

($\because$ If $X \sim \beta_1(\alpha, \beta)$

$$E(X) = \frac{\alpha}{\alpha+\beta})$$

= Bayes estimate of θ.

ii) A decision rule is equilizer rule if it is indep^t of θ

$\therefore$ Consider $R(\theta, d_{\alpha,\beta}(x)) = E_\theta\left[\frac{x+\alpha}{n+\alpha+\beta} - \theta\right]^2$, $\alpha, \beta > 0$

$$= \frac{1}{(n+\alpha+\beta)^2} E_\theta[x+\alpha - \theta(n+\alpha+\beta)]^2$$

$$= \frac{1}{(n+\alpha+\beta)^2} E_\theta\left[(x+\alpha)^2 - 2\theta(x+\alpha)(n+\alpha+\beta) + \theta^2(n+\alpha+\beta)^2\right]$$

$$\Rightarrow R(\theta, d(x)) = \frac{1}{(n+\alpha+\beta)^2} E_\theta\left[x^2+2x\alpha+\alpha^2 - 2\theta x n - 2\theta x\alpha - 2\theta x\beta - 2\theta\alpha(n+\alpha+\beta) + \theta^2(n+\alpha+\beta)^2\right]$$

$$= \frac{1}{(n+\alpha+\beta)^2} E_\theta\left[x^2+2x\alpha+\alpha^2 - 2\theta(n+\alpha+\beta)x - 2\theta(n+\alpha+\beta)\alpha + \theta^2(n+\alpha+\beta)^2\right]$$

$$= \frac{1}{(n+\alpha+\beta)^2}\left[E_\theta(x^2) + 2\alpha E_\theta(x) + \alpha^2 - 2\theta(n+\alpha+\beta)E_\theta(x) - 2\theta\alpha(n+\alpha+\beta) + \theta^2(n+\alpha+\beta)^2\right]$$

$\left(\because X \sim B(n, \theta) \right.$
$E(x) = n\theta, \; V(x) = n\theta(1-\theta)$
$\left. E(x^2) = n\theta - n\theta^2 + n^2\theta^2 \right)$

$$= \frac{1}{(n+\alpha+\beta)^2}\left[n\theta - n\theta^2 + n^2\theta^2 + 2\alpha n\theta - 2n\theta^2(n+\alpha+\beta) - 2\theta\alpha(n+\alpha+\beta) + \theta^2(n+\alpha+\beta)^2\right]$$

$$= \frac{1}{(n+\alpha+\beta)^2}\left[\theta^2(-n+n^2-2n^2-2n\alpha-2n\beta+n^2+\alpha^2+\beta^2+2n\alpha+2n\beta+2\alpha\beta) + \theta(n+2\alpha n - 2n\alpha - 2\alpha^2 - 2\alpha\beta) + \alpha^2\right]$$

$$= \frac{1}{(n+\alpha+\beta)^2}\left[\theta^2(\alpha^2+\beta^2+2\alpha\beta-n) + \theta(n-2\alpha^2-2\alpha\beta) + \alpha^2\right]$$

For equilizer rule coeff of θ^2 and θ should be zero

$\Rightarrow \alpha^2+\beta^2+2\alpha\beta-n=0$ and $n-2\alpha^2-2\alpha\beta=0$

$\Rightarrow n=\alpha^2+\beta^2+2\alpha\beta$ and $n=2\alpha^2+2\alpha\beta$

$\Rightarrow \cancel{\alpha^2}+\beta^2+\cancel{2\alpha\beta} = 2\alpha^2+\cancel{2\alpha\beta}$

$\Rightarrow \alpha^2 - \beta^2 = 0$

$\Rightarrow (\alpha-\beta)(\alpha+\beta) = 0$

eg. take $\alpha = \beta = \frac{n}{2}$ then e^n $n=\alpha^2+\beta^2+2\alpha\beta$ does not satisfies.

take $\alpha = \beta = \frac{\sqrt{n}}{2}$ the both $n=\alpha^2+\beta^2+2\alpha\beta$ and $n=2\alpha^2+2\alpha\beta$ are satisfies

$\therefore$ Bayes rule $d(x)$ with $\alpha = \beta = \frac{\sqrt{n}}{2}$ that

is $d(x) = \frac{x+\alpha}{n+\alpha+\beta} = \frac{x+\sqrt{n/2}}{n+\sqrt{\frac{n}{2}}+\sqrt{\frac{n}{2}}}$

$= \frac{x+\sqrt{\frac{n}{2}}}{n+\sqrt{n}}$ is an

equilizer rule.

Also it is minimax rule.

e.g. Let $X \sim B(n,\theta)$, $L(\theta,a)$ = weight sq error

$= \frac{(\theta-a)^2}{\theta(1-\theta)}$

$\therefore$ that is $\omega(\theta) = \frac{1}{\theta(1-\theta)}$

S.T $d(x) = \frac{x}{n}$ is minimax rule.

$\rightarrow R(\theta,d) = E_\theta\left[\frac{(\theta-d)^2}{\theta(1-\theta)}\right] = \frac{1}{\theta(1-\theta)} E\left[\frac{X}{n}-\theta\right]^2$

$= \frac{1}{n^2\theta(1-\theta)} E[X-n\theta]^2$

Date :

Page No. :

$$= \frac{1}{n^2\theta(1-\theta)} \underbrace{E[X-E(X)]^2}_{V(X)} \quad \left(\because X \sim B(n,\theta) \quad E(X) = n\theta \right)$$

$$= \frac{1}{n^2\theta(1-\theta)} \cancel{n\theta(1-\theta)} = \frac{1}{n} \quad \text{which is indep of } \theta$$

$d(X) = \frac{X}{n}$ is an equilizer rule and hence $d(x) = X/n$ is a minimax rule

* Suppose we take prior of θ as $U(0,1)$
posterior dist $(\theta \mid X=x) \sim \beta_I(x+1, n-x+1)$

$\Rightarrow d(x) =$ Bayes rule or Bayes estimate of θ

$$= \frac{E[\theta \,.\, w(\theta) \mid X=x]}{E[w(\theta) \mid X=x]}$$

$$= \frac{n(n+1)(n-x)}{n(n+1)|x(n-x)} = \boxed{\frac{X}{n}}$$

(invariance decision Problem)

Unit - 3

* Sufficient statistics :-

- Let X denote a r.v. whose $dist^n$ depends on parameter $\theta \in \Theta$. A real value f^n T of X, i.e $T(X)$, is said to be 'sufficient' for θ if the conditional $dist^n$ of X given $T = t$, is $indep^t$ of θ.

e.x : If $X_1 \ldots X_n$ are i.i.d $P(\lambda)$ r.v.s then

$$P(X_j = x) = \frac{e^{-\lambda}\lambda^{x_j}}{x_j!}, \quad x_j = 0, 1, 2 \ldots$$

Let $T(X) = T = \sum_{i=1}^{n} X_i$

Then $f(x_1, \ldots x_n | T = t)$

$$= \frac{P_\lambda(X_1 = x_1, X_2 = x_2 \ldots, X_n = x_n, T = t)}{P_\lambda(T = t)} \quad \left(\because T = \sum_{i=1}^{n} x_i\right)$$

$$= \frac{P_\lambda(X_1 = x_1, \ldots X_n = x_n)}{P_\lambda(T = t)}$$

$$= \frac{\frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}}{\frac{e^{-n\lambda}\lambda^{t}}{(\sum x_i)!}} \quad \left(T = \sum x_i \sim P(n\lambda)\right)$$

$= indep^t$ of λ

$\Rightarrow T(x) = \sum_{i=1}^{n} x_i$ is independent of λ and hence sufficient statistic.

Thm: Factorization Thm:-

Let X be a discrete random quantity whose pmf $f(x|\theta)$ depends on a parameter $\theta \in \Theta$. A f^n $T = t(x)$ is sufficient for θ if and only if the pmf factors into a product of a f^n of $t(x)$ and θ and a f^n of X alone that is, ~~$f(x|\theta)$~~.

$$f(x|\theta) = g(t(x), \theta) . h(x)$$

*** Essentially complete class of decision rules based on sufficient statistices:-**

- The nation that a sufficient statistics carries ~~all~~ all informations the samples has to give above ~~true~~ true value of the parameter can be formalized for dec theory as the class of dec rules based as a sufficient statistic terms an Essentially complete class.

Thm :- Consider the game $(\Theta, \mathcal{a}, L)$ where the statistician observes a random vector $\underline{X}$ whose dist. depends on θ. If T is a sufficient statistic for θ, then the set $\mathcal{D}_0$ of decision rules in $\mathcal{D}$ $(\mathcal{D}_0 \subset \mathcal{D})$ which are based on T, forms an essentially complete class in the game $(\Theta, \mathcal{D}, \hat{R})$.

Thm :- Rao-Blackwell thm :-

- This thm gives the explicit formula by which a non-randomized dec. rules may be improved by a non-randomized rule based on a sufficient statistic.

Statement :- Let $\mathcal{a}$ be a convex subset of E_K (m-d parameter), Let $L(\theta, a)$ be a convex fn of $a \in \mathcal{a}$ for each $\theta \in \Theta$, and suppose that T is a sufficient stat for θ. If $d(X)$ is a non-randomized decision rule, then the non-randomized rule based on T, say

$$\hat{\theta}(t) = E[d(X) \mid T = t],$$

Provided this expectation exists is as good as d.

ex: Let $X_i \sim N(\theta, \sigma^2)$ where $\sigma^2 = 1$, $i = 1, 2 \ldots n$. Let $L(\theta, a) = (\theta - a)^2$, Also we known that $T = \sum_{i=1}^{n} x_i$ is a sufficient statistics for θ. A ~~reasonable~~ reasonable estimate of θ when the assumption of normality is doubtful but when symmetry seems reasonable, is the median of X.

~~If the assumpt~~

If the assumption of normality is exactly satisfied, an improved estimate is,

$$\hat{\theta}(T) = E[\text{median of } x_i | T] = \bar{x}$$

Now,

$$E(T) = E\left(\sum_{i=1}^{n} x_i\right) = n\theta$$

$$E\left(\frac{T}{n}\right) = \frac{1}{n} E(T) = \frac{n\theta}{n} = \theta$$

$\Rightarrow \frac{T}{n}$ is unbiased estimator of θ

and $$V\left(\frac{T}{n}\right) = \frac{1}{n^2} V(T) = \frac{1}{n} V(\Sigma x_i)$$

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

$$= \frac{1}{n^2}(n) = \boxed{\frac{1}{n}}$$

$(\because V(x_i) = 1 \; \forall i)$

whereas under the assumption of Symmetry (but not normality), then median of X_i is unbiased estimator of θ, and

$$V(\text{median of } X_i) = \frac{\pi}{2n} \text{ ; for large } n$$

Here, $\bar{X}$ is as good as median of X_i for loss f^n $L(\theta, a) = |\theta - a|$ or $(\theta - a)^2$

$\boxed{6|10}$

Def: A family of distribution on the real line with pdf/pmf $f(x|\theta)$, $\theta \in \Theta$ is said to be an exponential family of distribution is $f(x|\theta)$ is of the form

$$f(x|\theta) = c(\theta)\, h(x)\, \exp\left[\sum_{i=1}^{k} \Pi_i(\theta)\, t_i(x)\right]$$

Result:- If $X_1, X_2 \ldots X_n$ is a random sample of size n from exponential family of distributions, then

$$T = (T_1, T_2, \ldots, T_k) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j) \ldots \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient (or jointly sufficient) statistic

Lemma :- Let $X_1, X_2 \ldots X_n$ be a sample from the exponential family of distribution. Then the dist^n of the sufficient statistic T has prob. f^n of the form

$$f_T(t|\theta) = C_0(\theta) h_0(t) \exp\left[\sum_{i=1}^{k} \pi_i(\theta) t_i(x)\right]$$

* Complete sufficient statistics :-

— We have seen that a great reduction in the complexity of the data may be achieved by means of sufficient statistics and thus it is also important to know how for such a reduction can be carried for a given problem.

The smallest amount of data that is still sufficient for the parameter is called a minimal sufficient statistics

The ~~Poperty~~ Property of the completeness is some what stronger than the property of minimal sufficient Statistic.

Defn: A sufficient statistic T for a parameter $\theta \in \Theta$ is said to be complete if for every real-valued fn g,

$$E_\theta[g(T)] = 0$$

$$\Rightarrow P_\theta[g(T) = 0] = 1$$

Defn: A sufficient statistic T is said to be boundedly complete if for every real-valued bounded fn g,

$$\left[E_\theta[g(T)] = 0\right] \Rightarrow P_\theta[g(T) = 0] = 1$$

Note:-

i) completeness $\Rightarrow$ bounded completeness but converse is not necessarily true.

Date :
Page No. :

ii) T is boundedly complete sufficient statistices if it is sufficient and if g is a bounded f^n for which $E_\theta[g(T)]$ exists and is equal to zero for all $\theta \in \Theta$ we have

$\therefore g(t) = 0 \quad \forall\, t$

eg Let $X_1, \ldots, X_n$ be a r.s from $B(m, \theta)$ then $T = \sum_{j=1}^{n} X_j$ is sufficient for θ

and $T \in B(N, \theta)$ where ~~N = min~~ $N = nm$

$\Rightarrow$ If $E_\theta[g(T)] = 0 \;\forall\, \theta \in [0,1]$, that is if

$$\sum_{t=0}^{N} g(t) \cdot p(t) = 0$$

$$\Rightarrow \sum_{t=0}^{N} g(t) \binom{N}{t} \theta^t (1-\theta)^{N-t} = 0 \quad \forall\, \theta \in [0,1]$$

$$\Rightarrow \sum_{t=0}^{N} g(t) \binom{N}{t} \left(\frac{\theta}{1-\theta}\right)^t = 0 \quad \forall\, \theta \in [0,1] \quad \text{(1)}$$

(1) $(\because (1-\theta)^N$ is constant)

If a convergent power series $\sum a_n z^n$ is zero for z in some open interval then each of the coefficient a_n must be zero.

The expression ① is a polynomial of degree N in $\left(\frac{\theta}{1-\theta}\right)$, or is a power series which converges to zero, so that $g(t)\binom{N}{t} = 0 \ \forall \ t = 0, 1, \ldots N$

that is $g(t) = 0 \ \forall \ t = 0, 1, \ldots N$

$\therefore P_\theta[g(T) = 0] = 1 \ \forall \ \theta$

Hence, $T = \sum_{j=1}^{n} x_j$ is a complete sufficient statistic.

Ex: Let $x_1, x_2 \ldots x_n$ be a r-s from $U(0, \theta)$, $\theta > 0$ then $T = \max x_i$ is sufficient for θ.

The density f^n of $T = \max x_i$ is

$$f_T(t|\theta) = n\left(\frac{t}{\theta}\right)^{n-1} \frac{1}{\theta} = \frac{n}{\theta^n} t^{n-1} I_{(0,\theta)}(t)$$

where $I_{(0,\theta)}(t) = \begin{cases} 1 & \text{if } 0 < t < \theta \\ 0 & \text{o.w} \end{cases}$

(∵ Recall: If $X_{(1)}, \ldots X_{(n)}$ is order Statistics some distribution of $Y = X_{(n)}$ = highest / maximum order Statistics is

$$f_1(y) = n[F(y)]^{n-1} f(y)$$

If $X \sim U(0,\theta) \Rightarrow f(x) = \frac{1}{\theta} \quad 0 < x < \theta$

$$0 \qquad 0.\omega.$$

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ x/\theta & \text{if } 0 < x < \theta \\ 1 & \text{if } x \ge \theta \end{cases}$$

So Pdf of $(T = X_{(n)}) = f_1(t)$

$$\Rightarrow f_1(t) = n\left(\frac{t}{\theta}\right)^{n-1} \frac{1}{\theta} \quad ; \; 0 < t < \theta$$

Now If $E_\theta[g(T)] = 0$

$$\Rightarrow \int_0^\theta g(t)\, f(t)\, dt = 0$$

$$\Rightarrow \int_{t=0}^{\theta} g(t)\, \frac{n}{\theta^n} t^{n-1}\, dt = 0 \quad (\because I_{(0,\theta)} = 1 \text{ if } 0 < t < \theta)$$

$\cancel{\Rightarrow \int g(t)\, t^{n-1}\, dt = 0}$

$$\Rightarrow \int_{t=0}^{\theta} g(t)\, t^{n-1}\, dt = 0 \quad \text{for } 0<t<\theta$$

$$\Rightarrow \quad g(t) = 0 \quad \forall \; t \in (0,\theta)$$

Hence $P_\theta[g(T)=0] = 1 \quad \forall \; \theta > 0$

$\Rightarrow T = \max X_i$ is complete sufficient

$\cancel{\Rightarrow \int g(t)\, t^{n-1}\, dt = 0}$

$$\Rightarrow \int_{t=0}^{\theta} g(t)\, t^{n-1}\, dt = 0 \quad \text{for } 0 < t < \theta$$

$$\Rightarrow g(t) = 0 \quad \forall\ t \in (0, \theta)$$

Hence $P_\theta[g(T) = 0] = 1 \quad \forall\ \theta > 0$

$\Rightarrow T = \max X_i$ is complete sufficient

10/10/25

* <u>Invariance</u> :-

- The invariance principle involves groups of transformations over the 3 spaces :- $\mathcal{X}$, $\mathcal{A}$, Θ associated with any decision problem.

<u>Ex</u> :- Let $X \sim f(x|\theta)$ where $f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}$; $x > 0,\ \theta > 0$

$= 0$ o.w

Generally, for this $f(x|\theta)$, which is exponential with mean θ, X denotes the lifetime. Suppose our interest is to estimate θ based on single obs^n $X = x$ under squared error loss func^n

$$L(\theta, a) = \left(1 - \frac{a}{\theta}\right)^2 = \left(\frac{a}{\theta} - 1\right)^2$$

Suppose X is "measured in terms of seconds"

Date :

Page No. :

and decision rule $\delta_0(x)$ is proposed as an estimate of θ.

Now, suppose X is to be measured in terms of "minutes" instead of "seconds". Then, define obs^n $Y = \frac{X}{60}$ (X : in seconds, Y : in minutes)

Define $\eta = \frac{\theta}{60}$

Then pdf of $Y = \frac{X}{60}$ will be

$$f(y/\eta) = \begin{cases} \frac{1}{\eta} e^{-y/\eta} & ; y>0, \eta>0 \\ 0 & o.w. \end{cases}$$

If the actions, say a^* in this problem are expressed in terms of minutes, so that $a^* = \frac{a}{60}$; then the loss funct^n would be

$$L(\eta, a^*) = \left(\frac{a^*}{\eta} - 1\right)^2 = \left(\frac{a/60}{\theta/60} - 1\right)^2$$

$$= \left(\frac{a}{\theta} - 1\right)^2 = L(\theta, a)$$

Thus, we can observe that the formal structure of the problem in terms of minutes is exactly the same as the structure of the problem in terms of seconds."

Consequently, we may use the same dec. rule for both the formulations.

Let $\delta_0^*(y)$ denote the proposed dec. rule for the transform problem ~~and~~ when this means that $\delta_0^*(y)$ should be equal to $\delta_0(y)$.

In other words, it is reasonable to expect that the same decision can be made, no matter what is the unit of measurements.
This implies that

$$\delta_0^*(y) = \frac{\delta_0(x)}{60}$$

Thus,
$$\delta_0(x) = 60\,\delta_0^*(y)$$
$$= 60\,\delta_0(y) = 60\,\delta_0\left(\frac{x}{60}\right)$$

Thus, in general, for any constant $c > 0$ and for transformation $y = cx$,

$$\delta_0(x) = \frac{1}{c}\cdot\delta_0(cx) \quad \text{——— (1)}$$

Our objective is to determine $\delta_0(x)$, so that eq^n (1) is satisfied, eq^n (1) can be solved easily by taking $c = \frac{1}{x}$ so that

$$(1) \Rightarrow \delta_0(x) = x\cdot\delta_0(1) = kx \quad \text{——— (2)}$$

$\frac{x-a}{b} \rightarrow$ change of origin & scale transformation

$y = ax + b \longrightarrow$ linear transformation (change in origin)

$y = cx \rightarrow$ change in scale

where $k = \delta_0(1) > 0 \Rightarrow$ constant

Such rules are said to be invariant for a given dec. problem. A best invariant rule for this problem thus consist of choosing the constant k which minimizes the risk.

NOTE :- The invariance principle involves groups of transformations over 3 spaces involved in dec. theory : Θ, $\mathcal{A}$ & $\mathcal{X}$

Def^n :- A transformation g from $\mathcal{X}$ into itself is said to be onto $\mathcal{X}$ if the range of g is the whole of $\mathcal{X}$; that is, if for every $x_1 \in \mathcal{X}$, $\exists$ an $x_2 \in \mathcal{X}$ such that $g(x_2) = x_1$ $(g : \mathcal{X} \Rightarrow \mathcal{X})$

A transformation g for $\mathcal{X}$ into itself is said to be one-to-one is $g(x_1) = g(x_2) \Rightarrow x_1 = x_2$

Def^n : The family of distribution P_θ, $\theta \in \Theta$ is said to be invariant under the group of transformation $\mathcal{g}$ if for every $g \in \mathcal{g}$ and every $\theta \in \Theta$, $\exists$ a unique $\theta' \in \Theta$ s.t. the dist^n of $g(x)$ is given by $P_{\theta'}$ whenever the dist^n of X is given by P_θ. The θ' uniquely)

$f : A \xrightarrow{\text{into}} B$

$f : A \xrightarrow{\text{onto}} B$

determined by g and θ is denoted by $\bar{g}(\theta)$.

ex If $X \sim N(\mu, \sigma^2)$.

$$P_{\mu,\sigma}(x) = f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < \mu < \infty$$
(θ)

$-\infty < x < \infty$

$\sigma > 0$

$$g(x) = \frac{X-\mu}{\sigma} = Z \sim N(0,1)$$

$$P_{\theta}(z) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$$

↳ (Invariant standarized dist^n)
(normal to normal)

* Remark : The condition that the family of distribution P_θ be invariant under g is that for every measurable set $A \subset \mathcal{X}$, $P_\theta\{g(X) \in A\} = P_{\bar{g}(\theta)}\{x \in A\}$.

In term of expectation this is equivalent to say that for every integrable real-valued f^n ϕ.

$E_\theta[\phi(g(X))] = E_{\bar{g}(\theta)}[\phi(X)]$ when the dist^n of X is P_θ

Lemma:- If a family of distribution P_θ, $\theta \in \Theta$ is invariant under $\mathcal{G}$ then $\bar{\mathcal{G}} = \{\bar{g} : g \in \mathcal{G}\}$ is a group transformation of Θ into itself.

Note: All transformation $\bar{g} \in \bar{\mathcal{G}}$ are one-to-one and onto f^n.

Def^n: Invariance of statistical decision problem:-

— A dec problem, consisting of the game $(\Theta, \mathcal{A}, L)$ and dist^n P_θ over $\mathcal{X}$ is said to be invariant under the group $\mathcal{G}$ if the family of dist^n P_θ, $\theta \in \Theta$ is invariant under $\mathcal{G}$ and if the loss f^n is invariant under $\mathcal{G}$ in the sense that for every $g \in \mathcal{G}$ and $a \in \mathcal{A}$, $\exists$ a unique $a' \in \mathcal{A}$, s.t.

$$L(\theta, a) = L(\bar{g}(\theta), a') \quad \forall \; \theta \in \Theta$$

Lemma:- If a dec. problem is invariant under a group $\mathcal{G}$ then $\tilde{\mathcal{G}} = \{\tilde{g} : g \in \mathcal{G}\}$ is a group of transformation of $\mathcal{A}$ into itself.

ex :- Suppose $X \sim N(\theta, 1)$, $\Theta = \mathcal{A} = \mathbb{R}$,
$L(\theta, a) = (\theta - a)^2$
Consider a group of transformation $\mathcal{G}$, with $g_c(x) = x + c$
then.
$$g_c(X) = X + c \sim N(\theta + c, 1)$$
Thus, the $N(\theta, 1)$ family is invariant under $\mathcal{G}$ and $\bar{g}_c(\theta) = \theta + c$

Further, $L(\theta, a)$ with $\bar{g}_c(\theta) = \theta + c$
and $\tilde{g}_c(a) = a + c$
then,
$$L(\bar{g}_c(\theta), \tilde{g}_c(a))$$
$$= (\bar{g}_c(\theta) - \tilde{g}_c(a))^2$$
$$= (\theta + c - a - c)^2$$

$L(\theta, a) = (\theta - a)^2$
Thus, $L(\theta, a) = L(\bar{g}_c(\theta), \tilde{g}_c(a)) \ \forall \theta$
Thus, Loss is invariant under group of transformation $\mathcal{G}$ hence the decision problem is invariant under $\mathcal{G}$.

~~ex Let $X \sim B(n, \theta)$, n is known and $\Theta = [0, 1]$~~
~~Also Let $\mathcal{A} = [0, 1]$ and $L(\theta, a) = W(\theta - a)$, some~~
~~even fn of $(\theta - a)$~~