NAME: Vandana Chudasama

STD: MSc DIV: Previous ROLL NO.:

SUBJECT: Decision Theory

YOUVA

INDEX

SR. NO.	DATE	TITLE	PAGE NO.	TEACHER'S SIGN
		Unit 1.		
	①	Decision problem and 2-person game		
	②	Utility theory		
	③	Loss function		
	④	Expected loss.		
	⑤	decision rule (Non-Randomized and randomized		
	⑥	Optimal Research rule -		
	⑦	Unbiasedness, Invariance		
		Bayes principel.		
		Minimax principle		
		decision principle		
		Inference problem as decision problem		
		geometric interpretation of dec pr.		
		Unit-2		
	①	Concepts of admissibitity and Completeness		
	②	Admissibility of Bayes rule		
	③	Existence of Bayes rule		
	④	Existence of minimal Complete class		
	⑤	Essential Completeness of the class of non randomized decision rule.		
	⑥	Minimax theorem. for finite parameter Space		
	⑦	Complete class theorem		
	⑧	Supporting and Separating hyperplane theorem.		

SR. NO.	DATE	TITLE	PAGE NO.	TEACHER'S SIGN
		Unit 3		
		(1) Review of sufficient statistic		
		(2) Essentially complete class of rules based on sufficient statistics		
		(3) Concept of invariance.		
		(4) Invarient decision rule		
		(5) Admissibility of invariant of decision rule		
		(6) Minimax ~~dot~~ Invariant decision rules.		

Decision Theory

Page No.: 1

Date:

Example Odd or Even game

Player 1 – Nature, Player 2 – Statistician.

Parameter space $\Theta = \{\theta_1 = 1, \theta_2 = 2\}$

Action space $\mathcal{a} = \{a_1 = 1, a_2 = 2\}$

Nature wins if sum of fingers shown up is odd

Statistician wins if sum of fingers shown up is even

The amount won/lost = Sum of the finger shown up

LOSS TABLE FOR STATISTICIAN (negative loss = gain)

$\Theta \setminus \mathcal{a}$	$a_1 = 1$	$a_2 = 2$
$\theta_1 = 1$	-2	3
$\theta_2 = 2$	3	-4

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

L is a mapping / function $L : \Theta \times \mathcal{a}$

Suppose a random observable X is involved in the experiment whose distribution depends on the state of nature $\theta \in \Theta$ denoted as P_θ on the basis of outcome $X = x$, the statistician chooses an action $d(x) \in \mathcal{a}$.

Such a function d which maps $\mathcal{X}$ into $\mathcal{a}$ is an elementary strategy for the statistician.

Thus, $d : \mathcal{X} \to \mathcal{a}$

Decision Rule.

Any function $d(x)$ that maps Sample space $\mathcal{X}$ into $\mathcal{A}$ is called a non-randomised decision rule or a decision formula function provided the loss (or risk) exists and finite $\forall\ \theta \in \Theta$

The class of all non-randomised decision rule is denoted by D.

Example (continue) odd or even :-

Suppose before the game is placed the statistician is allowed to ask nature that how many fingers ~~she~~ he intends to put up

Suppose the nature answers truthfully with probability $3/4$ and untruthfully with prob $1/4$. Then Prob distⁿ of X = answer given by nature

$X = x$	$P(X=x \mid \theta_1)$	$P(X=x \mid \theta_2)$
$x_1 = 1$	$3/4 = P_{\theta_1}(X=1)$	$1/4 = P_{\theta_2}(X=1)$
$x_2 = 2$	$1/4 = P_{\theta_1}(X=2)$	$3/4 = P_{\theta_2}(X=2)$
	1	1

Thus, $\exists$ 4 possible non-randomized decision rule from $d: \mathcal{X} \to \mathcal{A}$ are as follow

$\mathcal{X} \backslash \mathcal{A}$	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

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

No. of decision rules $= m^n$ if
$\quad \mathcal{a} = \{a_1, a_2, \ldots a_m\}$
$\quad \mathcal{X} = \{x_1, x_2, \ldots x_n\}$

In general decision function can be denoted as $d(x)$, That is,

$d_1(x=1) = a_1 = 1$, $d_1(x=2) = a_1 = 1$
$d_2(x=1) = a_1 = 1$, $d_2(x=2) = a_2 = 2$
$d_3(x=1) = a_2 = 2$, $d_3(x=2) = a_1 = 1$
$d_4(x=1) = a_2 = 2$, $d_4(x=2) = a_2 = 2$

<u>Loss</u> <u>function</u> :- If 'a' is the action choosen by statistician and θ is the true value of the parameter (action choosen by nature) then it may result in a loss say $L(\theta, a)$, which is a real-valued function defined for every pair $(a, \theta) \in \mathcal{a} \times \Theta$

LOSS TABLE : $L(\theta, a)$

$\Theta \backslash a$	$a_1 = 1$	$a_2 = 2$
$\theta_1 = 1$	-2	3
$\theta_2 = 2$	3	-4

Note that, if 'a' is the correct action for some value of θ then $L(\theta, a) = 0$, OW: $L(\theta, a) \neq 0$

$L(\theta, d(x))$

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

* Risk function :-

The long term average loss that would result by using a particular decision rule 'd' can be denoted as $R(\theta, d(x))$ and is defined as

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

$$= \int_x L(\theta, d(x)) \cdot f(x,\theta)\, dx$$

$$= \int_x L(\theta,d(x))\, dF_X(x;\theta) \quad (\text{if } x \text{ is cont. r.v})$$

and

$$\sum_x L(\theta,d(x))\, P(X=x) \quad ; (\text{if } x \text{ is discrete r.v})$$

$R_{\theta_1}(\theta, d_1) = (-2)(3/4) + (-2)(1/4) = -2$

$R_{\theta_2}(\theta, d_1) = (3)(1/4) + (3)(3/4) = 3$

$R_{\theta_1}(\theta, d_2) = (-2)(3/4) + (3)(1/4) = -3/4$

$R_{\theta_2}(\theta; d_2) = (3)(1/4) + (-4)(3/4) = -9/4$

$R_{\theta_1}(\theta, d_3) = (3)(3/4) + (-2)(1/4) = 7/4$

$R_{\theta_2}(\theta, d_3) = (-4)(1/4) + (3)(3/4) = 5/4$

$R_{\theta_1}(\theta, d_4) = (3)(3/4) + (3)(1/4) = 3$

$R_{\theta_2}(\theta, d_4) = (-4)(1/4) + (-4)(3/4) = -4$

Risk Table

$\Theta \setminus D$	d_1	d_2	d_3	d_4
θ_1	-2	-0.75	1.75	3
θ_2	3	-2.25	1.25	-4

Examples

(1) Consider a decision problem with $\mathcal{X} = \{x_1, x_2\}$ $\mathcal{A} = \{a_1, a_2, a_3\}$, list out all possible non-randomised decision rules.

no. of decision rules $= m^n = 3^2 = 9$

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
x_1	a_1	a_1	a_1	a_2	a_2	a_3	a_3	a_3	a_3
x_2	a_1	a_2	a_3	a_1	a_2	a_3	a_1	a_2	a_3

(2) $\mathcal{X} = \{x_1, x_2, x_3\}$, $\mathcal{A} = \{a_1, a_2\}$

no. of decision rule $= m^n = 2^3 = 8$

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
x_1	a_1	a_1	a_1	a_1	a_2	a_2	a_2	a_2
x_2	a_1	a_1	a_2	a_2	a_1	a_1	a_2	a_2
x_3	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2

(3) $\mathcal{X} = \{x_1, x_2\}$ $\mathcal{A} = \{a_1, a_2\}$, $\Theta = \{\theta_1, \theta_2\}$

Suppose $P(X = x_1 | \theta_1) = 1/3 = P_{\theta_1}(x_1)$

$P_{\theta_2}(x_1) = 1/4$

$\Theta \setminus \mathcal{A}$	a_1	a_2
θ_1	1	2
θ_2	2	1

$P_\theta(x)$

	θ_1	θ_2
x_1	1/3	1/4
x_2	2/3	3/4

(2) LOSS TABLE :

	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_1	1	2	1	2	2	1	2	1
x_2	1	2	2	1	1	2	2	1
	1	2	1.6	1.25	1.33	1.75	2	1

(1) decision table

	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

(3) Risk table $R(\theta, d(x))$

	d_1	d_2	d_3	d_4
θ_1	1	1.6	1.33	2
θ_2	2	1.25	1.75	1

★ Decision theory problem in terms of classical mathematical statistics.

(1) Suppose $\mathcal{a} = \{a_1, a_2\}$, $\Theta = R$
and loss function is $L(\theta, a_1) = \begin{cases} l_1 & \text{if } \theta > \theta_0 \\ 0 & \text{if } \theta \le \theta_0 \end{cases}$

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

Assume that $l_1, l_2 > 0$
Obviously statistician should take action a_1 if $\theta \le \theta_0$ and a_2 if $\theta > \theta_0$

Let D = set of decision rules where $d : \mathcal{X} \to \mathcal{a}$ with prob that $P_\theta\{d(x) = a_i\}$ is well defined for all values of $\theta \in \Theta$

LOSS TABLE $= L(\theta, a)$

$\Theta \backslash a$	a_1	a_2
$\theta \le \theta_0$	0	$l_2 > 0$
$\theta > \theta_0$	$l_1 > 0$	0

Then in this case, risk function will be

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

Let $H_0 : \theta > \theta_0$ ag $H_1 : \theta \le \theta_0$
and if $a_1 =$ Reject H_0
$a_2 =$ Accept H_0

True state of Nature Θ \ a	Reject H_0	Accept H_0
$H_1 : \theta \le \theta_0$ (H_1 is true)	Reject H_0 when H_1 is true (Correct decision)	Accept H_0 when H_1 is true / H_0 is false (Type II Error)
$H_0 : \theta > \theta_0$ (H_0 is true)	Reject H_0 when H_0 is true (Type I Error)	Accept H_0 when H_0 is true (Correct decision)

$$P_{\theta > \theta_0}\{d(x) = a_1\} = P\{\text{Reject } H_0 \mid H_0 \text{ is true}\}$$
$$= P\{\text{Type I error}\} = \alpha$$

$$P_{\theta \le \theta_0}\{d(x) = a_2\} = P\{\text{Accept } H_0 \mid H_1 \text{ is true}\}$$
$$= P\{\text{type II error}\} = \beta$$

(ii) when $\mathcal{A} = \{a_1, a_2, \ldots a_k\}$, $k \geq 3$
then the decision theoretic problem are called multiple decision problem.
(eg: double/ Multiple Sampling plans, SPRT etc)

(iii) when $\mathcal{A} = (-\infty, \infty) = R$
Consider $\Theta = R$ Suppose the
Loss $f^n = L(\theta, a) = c(\theta - a)^2$, $c > 0$

Thus a decision function can be defined as a real valued function defined on x which may be considered as an "estimate" of the true state of nature θ

The statistician would desire to choose the function d to minimize the risk function

Risk f^n : $R(\theta, d) = E_\theta[L(\theta, d)]$
$= E_\theta[c(\theta - d(x))^2]$
$= c\, E_\theta[(\theta - d(x))^2]$
$= c$ [Mean Squared Error of estimate $d(x)$]

* Other type of loss function:-

$L(\theta, a) = c|\theta - a|$

* Randomization:- Consider a game $(\Theta, \mathcal{A}, L)$ with

$\Theta = \{\theta_1, \theta_2\}$
$\mathcal{A} = \{a_1, a_2, a_3\}$

LOSS table $L(\theta, a)$

$\Theta \backslash a$	a_1	a_2	a_3
θ_1	4	1	3
θ_2	1	4	3

Que Should the statistician ~~ever~~ take the action a_3?

Suppose the statistician tosses a fair coin to choose between a_1 & a_2;
He decides to choose a_1 if heads turns up
He decides to choose a_2 if tails turns up.

Then now, such a decision, denoted by δ, is a randomized decision; such decision allow the actual choice of the action a to be left to a random mechanism and the statistician chooses the prob. of various outcomes.

* Randomization

Consider a game $\Theta = \{\theta_1, \theta_2\}$, $\mathcal{A} = \{a_1, a_2, a_3\}$

LOSS Table

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

$$P_1 = \left(\underset{\substack{\downarrow \\ a_1}}{\tfrac{1}{2}}, \underset{\substack{\downarrow \\ a_2}}{\tfrac{1}{2}}, \underset{\substack{\downarrow \\ a_3}}{0}\right)$$

Expected loss in using P_1

$$L(\theta_1, P_1) = \frac{1}{2} L(\theta_1, a_1) + \frac{1}{2} L(\theta_1, a_2) + 0\, L(\theta_1, a_3)$$

$$= \frac{1}{2} \times 4 + \frac{1}{2} \times 1 = 2 + \frac{1}{2} = \frac{5}{2}$$

$L(\theta_2, P_1) = \frac{1}{2} L(\theta_2, a_1) + \frac{1}{2} L(\theta_2, a_2) + 0 L(\theta_2$

$= \frac{1}{2} \times 1 + \frac{1}{2} \times 4 = \frac{5}{2}$

Put $P_2 = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$

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

$= 1 + \frac{1}{2} + \frac{3}{4} = \frac{9}{4}$

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

$= 3$

Put $P_3 = \left(\frac{3}{8}, \frac{5}{8}, 0\right)$

$L(\theta_1, P_3) = \frac{3}{8} \times 4 + \frac{5}{8} \times 1 = \frac{12}{8} + \frac{5}{8} = \frac{17}{8}$

$L(\theta_2, P_3) = \frac{3}{8} \times 1 + \frac{5}{8} \times 4 = \frac{23}{8} = 2.85$

In general a randomization decision for the statistician in a game $(\Theta, \mathcal{A}, L)$ is a prob dist^n over $\mathcal{A}$.

If p is the probability dist^n over $\mathcal{A}$ and z is r.v taking values in $\mathcal{A}$, whose dist^n is given by P the expected / average loss in the use of the randomized decision P is

$$L(\theta, p) = E[L(\theta, z)] \quad —(1)$$

Provided it exists
The space of randomized decisions p, for which $L(\theta; p)$ exists and is finite for $\forall \theta \in \Theta$ is denoted by D^*

Note:
$D \rightarrow$ Set of non-randomized decision rules
$D^* \rightarrow$ set of randomized decision rules.

Thus the game (Θ, a^*, L) can be considered as the game.
where a^* = Statistician choose action based on probabilities.
(Θ, a, L) in which the statistician is allowed randomization.

Defn :- Randomized Action :-

A prob distn p defined over action space A is called randomized action provided
$L(\theta, p) = E[L(\theta, z)]$
where z is r.v taking values provide a with probability distribution exists & finite

$$L(\theta, p) = \sum_{j=1}^{M} L(\theta, a_j) P_j$$

Set of all randomized action is denoted by a^*

Defn :- Randomized Decision Rule

Any prob distn δ on the space of non-randomized decision function D, is called Randomized decision function

Or 'Randomized Decision Rule' provide the risk for $R(\theta, \delta) = E[R(\theta, z)]$ exists and finite $\forall \theta \in \Theta$, where z is a r.v taking value in D with prob dist δ

The space of all randomized dec rules is denoted by D^*

$\theta \backslash D$	d_1	d_2	...	d_{m^n}	Prob
θ_1					
θ_2					
⋮					
θ_n					

x	$P_\theta(x)$: $\theta_1 \ldots \theta_n$	d_1: $\theta_1 \ldots \theta_n$	d_2: $\theta_1 \ldots \theta_n$	... d_{m^n}: $\theta_1 \ldots \theta_n$
x_1	P_1			
x_2	⋮			
⋮	⋮			
x_k	P_k			
		$R(\theta, d_1)$	$R(\theta, d_2)$	$R(\theta, d_{m^n})$

Θ	$d_1 \ldots d_n$
θ_1	
⋮	
θ_n	

Note:- $\delta = (p_1, \ldots, p_m)$ $p_j \geq 0$ $\sum_{j=1}^{m} p_j = 1$

$$D^* = \text{Set of randomized decision} = \{(p_1, \ldots, p_m) \; p_j \geq 0 \; \sum_{j=1}^{m} p_j = 1\}$$

Result:- Let P_1 be the randomized action and P_2 be another randomized action. Then any convex combination of P_1 & P_2 will ~~be~~ also be a randomized action.

That is if $p = \alpha P_1 + (1-\alpha) P_2$
the p will also be randomized action

<u>Result</u> :- $L(\theta, p) = \alpha L(\theta, p_1) + (1-\alpha) L(\theta, p_2)$
$0 < \alpha \leq 1$

$R(\theta, \delta) = \alpha R(\theta, \delta_1) + (1-\alpha) R(\theta, \delta_2)$
if $\delta_1, \delta_2 \in D^*$
$0 \leq \alpha \leq 1$
Then $\delta \in D^*$

<u>Ex</u> $\mathcal{X} = \{x_1, x_2\}$ $\mathcal{A} = \{a_1, a_2\}$
$\Theta = \{\theta_1, \theta_2\}$ $D = \{d_1, d_2, d_3, d_4\}$

LOSS table

$\Theta \backslash \mathcal{A}$	a_1	a_2
θ_1	1	2
θ_2	2	1

$\mathcal{X} \backslash D$	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

$P_{\theta_1}(x_1) = \frac{1}{3}$ $P_{\theta_2}(x_1) = \frac{1}{4}$

$\Theta \backslash D$	d_1	d_2	d_3	d_4
θ_1	1	5/3	4/3	2
θ_2	2	7/4	5/4	1

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

$R(\theta_1, \delta_1) = E[R(\theta_1, d)]$
$= E[R(\theta_1, z)]$, $z \in D$
$D = \{d_1, \ldots d_4\}$

$$R(\theta, \delta_1) = R(\theta_1, d_1)P_1 + \dots + R(\theta_1, d_4)P_4$$

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

$$= \frac{1}{2} + \frac{12}{30} + \frac{5}{30} + \frac{2}{10}$$

$$= \frac{19}{15}$$

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

$$= R(\theta_2, d_1)p_1 + \dots + R(\theta_2, d_4)P_4$$

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

$$= \frac{1}{2} + \frac{21}{40} + \frac{5}{40} + \frac{1}{10}$$

$$= \frac{7}{4}$$

	Behavioral Dec Rule	Randomized Dec Rule
(i)	Tells the statistician how to randomize ~~offer~~ After of the Experiment	(i) Choose of random a dec rule / fn that tells the statistician before observing the outcome of the Expt Exactly what action to take as a result of the Expt
(ii)	Chooses, for each $x \in \mathcal{X}$ a random point $Y_x \in \mathcal{A}$	(ii) Chooses at random, a fn Y_x which is a funᶯ from $\mathcal{X}$ to $\mathcal{A}$
(iii)	specifies the dist of Y_x for each $x \in \mathcal{X}$	(iii) Specifies the distⁿ of the random fn Y_x
(iv)	specifies only the marginal distⁿ of Y_x for each x	(iv) Specifies the entire distribution of Y_x

* Behavioural Decision Rule / function

A function $\delta(x)$, from $\mathcal{X}$ into $\mathcal{A}^*$ is called a behavioural decision function / Rule provided

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

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

Game $(\Theta, \mathcal{A}, L)$ $\xrightarrow{\text{utilize } x \text{ with } P_\theta(x)}$ Decision Rule (Θ, D, R) $\xrightarrow[\text{action}]{\text{Randomized}}$ Randomized decision Rule $(\Theta, \mathcal{D}^*, R)$

Game $(\Theta, \mathcal{A}, L)$ $\xrightarrow{\text{Randomized action}}$ Probability dist associated with action $(\Theta, \mathcal{A}^*, L)$ $\xrightarrow[\text{from of data } x]{\text{Information about } \theta, \text{ in}}$ $(\Theta, \mathcal{D}, \hat{R})$

Ex Let $\mathcal{A} = \{a_1, a_2\}$ $\quad \mathcal{X} = \{x_1, x_2\}$

$\mathcal{A}^*$ = Set of Probability distribution on the Sample Space $\mathcal{A}$ on $[0,1]$, Such that $\pi \in \mathcal{A}^*$ represents probability of taking action a_1 whereas $(1-\pi) \in \mathcal{A}^*$ represents probability of taking action a_2

Suppose the D = decision set = $\{d_1, d_2, d_3, d_4\}$

$\delta \rightarrow$	P_1	P_2	P_3	P_4
$\mathcal{X} \setminus D$	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

Take $D^* =$ Set of all randomized decision rules
$$= \{(P_1, P_2, P_3, P_4), \; P_i \geq 0, \; \sum_{i=1}^{4} P_i = 1\}$$

whereas $\mathcal{D} =$ Set of all behavioural dec rule
$$= \{(\pi_1, \pi_2), \; 0 \leq \pi_1 \leq 1, \; 0 \leq \pi_2 \leq 1\}$$

which maps $\mathcal{X}$ into $\mathcal{a}$.
if x_1 is observed then $\pi_1 \in \mathcal{a}^*$ is used
if x_2 is observed then $\pi_2 \in \mathcal{a}^*$ is used

Let $\delta = \left(P_1 = \frac{1}{2}, \; P_2 = \frac{3}{10}, \; P_3 = \frac{1}{10}, \; P_4 = \frac{1}{10}\right)$

$$P = \left(P(a_1) = \pi_1, \; P(a_2) = 1 - \pi_1\right)$$

$\Rightarrow \hat{\delta} = (\pi_1, \pi_2) \; ; \; 0 \leq \pi_1, \pi_2 \leq 1$

where $\pi_1 = P$ (Choosing a_1 when $x = x_1$)
$= P_{x_1}(a_1)$

$$1 - \pi_1 = P_{x_1}(a_2)$$
$$\pi_2 = P_{x_2}(a_1)$$
$$1 - \pi_2 = P_{x_2}(a_2)$$

With respect to δ :-

$P_{x_1}(a_1) = P_1 + P_2 = \frac{1}{2} + \frac{3}{10} = \frac{8}{10} = \frac{4}{5} = \pi_1$

$P_{x_1}(a_2) = P_3 + P_4 = \frac{1}{10} + \frac{1}{10} = \frac{1}{5} = 1 - \pi_1$

$P_{x_2}(a_1) = P_1 + P_3 = \frac{1}{2} + \frac{1}{10} = \frac{6}{10} = \frac{3}{5} = \pi_2$

$P_{x_2}(a_2) = P_2 + P_4 = \frac{3}{10} + \frac{1}{10} = \frac{4}{10} = \frac{2}{5} = 1 - \pi_2$

Thus, behavioural decision Rule w.r. to δ
$\hat{\delta} = (\pi_1, \pi_2) = \left(\frac{4}{5}, \frac{3}{5}\right)$

Optimal / Best decision Rule

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

Method 1: Restricting the available rules OR Reducing of size of D^* according to some criteria

(i) Unbiasedness :- A decision rule δ is said to be unbiased if
$E_\theta[L(\theta, \delta(x))] \leq E_\theta[L(\theta', \delta(x))]$

(ii) In variance

Method 2: Ordering the decision rule

(i) The Bayes Principle :- the probability distⁿ defined over Θ is called a prior distribution. Suppose we denote the prior dist τ then

$r(\tau, \delta)$ = Bayes Risk of a dec rule δ w.r.t prior dist τ

Define Joint dist of T and X, and conditional dist of Parameter given the obs. which is known as posterior distⁿ of the Parameter given the obs.

$f_1(\theta) \rightarrow$ prior distribution

$f(\theta, x) \rightarrow$ joint distribution of T & X

$g(\theta/x) \rightarrow$ posterior dist

Def: Bayes Rule :- A dec Rule δ_0 is said to be Bayes w.r.t the prior distⁿ $\tau \in \Theta^*$ if

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

Minimum Bayes Risk w.r.t prior dist τ

Def: ε- Bayes Rule :- let $\varepsilon > 0$. A decision rule δ_0 is said to be ε-Bayes w.r to prior dist $\tau \in \Theta$ if

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

(ii) The Minimax Principle

According to this principle the statistician would arrange decision rule in order of the maximum risk involved.

Def A decision rule δ_0 is said to be ~~minimax~~ if Minimax if

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

$\Downarrow$

minimax value

Def Let $\varepsilon > 0$. A decision rule δ_0 is said to be ε-minimax if

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

OR

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

— x —

Def^n: A prior dist^n $\tau_0 \in \Theta^*$ is said to be the least favourable if

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

Maximin → Lower value of the game

Result: A decision rule δ_0 is minimax if and only if

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

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

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

Geometric Interpretation for finite Θ

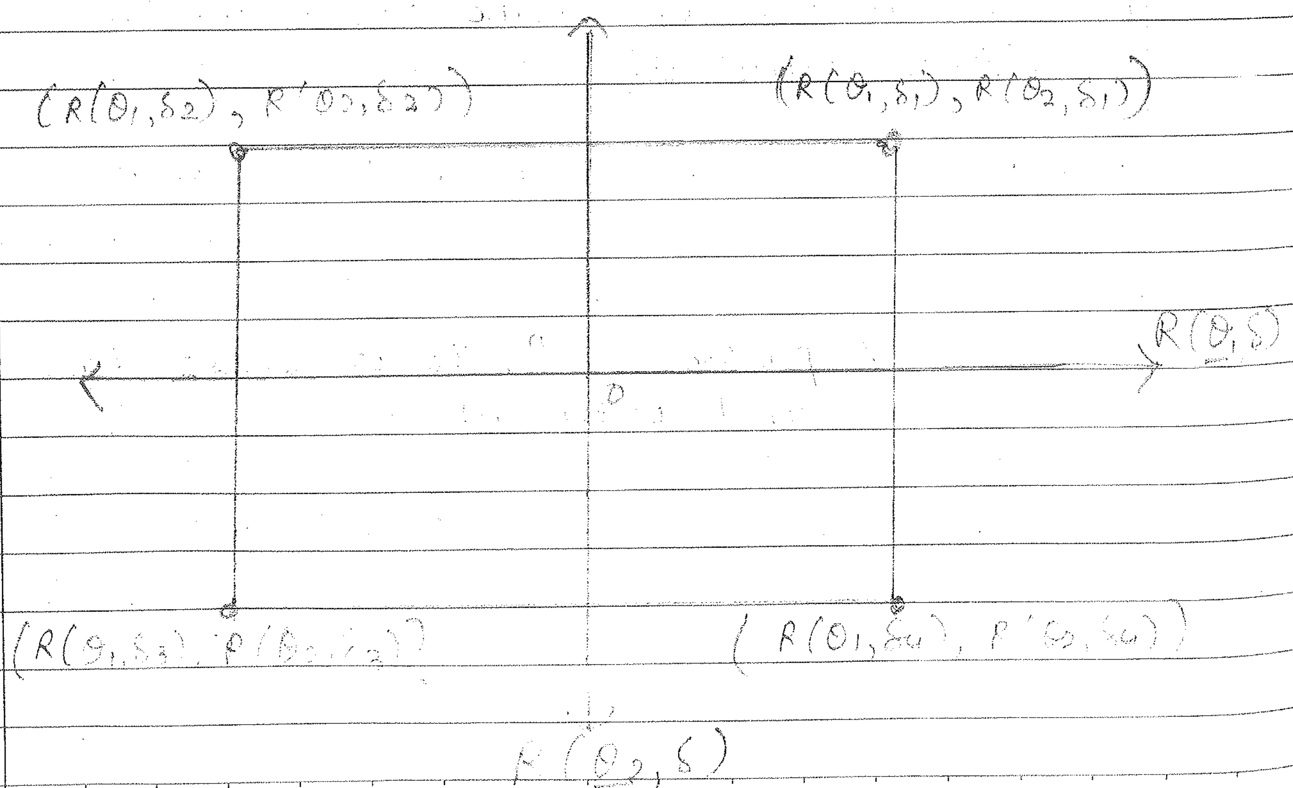
Suppose $\Theta = \{\theta_1, \ldots, \theta_k\}$ for $\delta \in D^*$

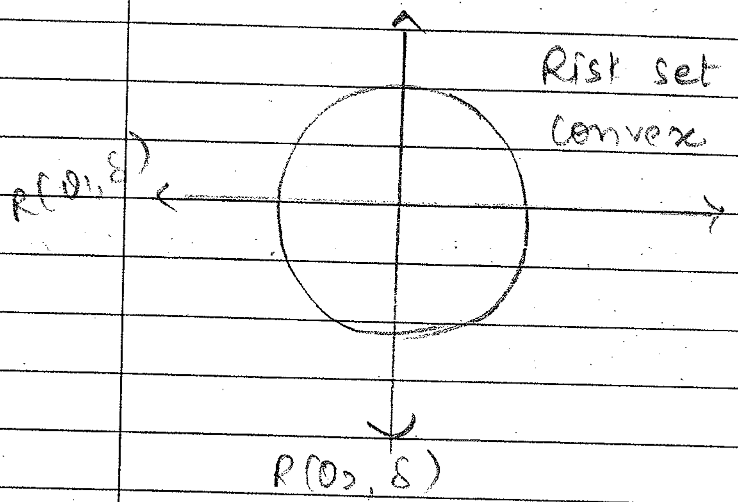
let $y_j = R(\theta_j, \delta)$; $j = 1, \ldots k$
Then $= (y_1, \ldots, y_k) \in E_k$ (is some point in Euclidian k-space)

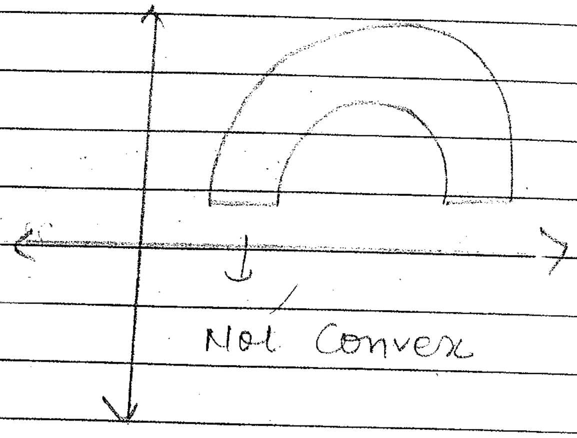
let S = Risk Set
$= \{(y_1, \ldots, y_k) \mid y_j = R(\theta_j, \delta) ; j = 1, 2, \ldots k, \delta \in D^*\}$

Then $S \subset E_k$
If $k = 2$ $\Theta = \{\theta_1, \theta_2\}$ let $D^* = \{\delta_1, \delta_2, \delta_3, \delta_4\}$

For δ_1 (under $\theta = \theta_1$) $\to y_1 = R(\theta_1, \delta)$
δ_1 (under $\theta = \theta_2$) $\to y_2 = R(\theta_2, \delta)$
$\vdots$
For δ_4 (under $\theta = \theta_1$) $\to y_1 = R(\theta_1, \delta)$
δ_4 (under $\theta = \theta_2$) $y_2 = R(\theta_2, \delta)$







lemma: The Risk set S is a convex subset of E_k

$$S = \{ (y_1, \ldots y_k) \mid y_j = R(\theta_j, d) \}$$

Convex Hull :- Let $S_0 = \{ (y_1, \ldots y_k) \mid y_j = R(\theta_j, d) ; j = 1 \ldots k, d \in D \}$ be the set of risk points corresponding to the non-randomized decision rule d. Then the convex hull of a set S_0 is defined as the smallest set containing S_0 or it is the intersection of all convex sets containing S_0

Note :- Baye's Rule in Estimation

In a problem of Estimation if $\exists$ a randomized Bayes rule w.r.to a prior distⁿ τ, then there will be always a non-randomized Bayes rule with respect to the same prior distⁿ x

Obtaining the Bayes Estimate :-

We know $r(\tau, \delta) = E[R(z, \delta)]$

where z is a r.v taking values in Θ with prior distⁿ τ, δ_0 is Bayes rule w.r. to τ is

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

Also,

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

If d_0 is a non-randomized Bayes rule w.r.to a prior dist τ then

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

where

$$r(\tau, d) = \int_\theta R(\theta, d)\, d\tau(\theta)$$

$$R(\theta, d) = E_\theta[L(\theta, d(x))] = \int_x L(\theta, d(x))\, dF(x|\theta)$$

where $F(x|\theta)$ is a conditional dist function of X given θ

$$r(\tau, d) = \int_\theta \int_x L(\theta, d(x))\, dF(x|\theta) \cdot d\tau(\theta)$$

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

$$\left(\because F(x, \theta) = F(x|\theta) \cdot \tau(\theta) = \tau(\theta|x) F(x) \right)$$

To find Bayes Estimate, we need to find d which minimizes $r(\tau, d)$.

That is to minimize $\int_\theta L(\theta, d(x)) \cdot d\tau(\theta|x)$, for each given x

That is to minimize conditional expected Loss

$$\int_\theta L(\theta, d(x)) \cdot d\tau(\theta|x), \text{ for each given } x$$

Ex Let $\Theta = \mathcal{A} = (0, \infty)$ and $L(\theta, a) = c(\theta - a)^2$ and Suppose $(X|\theta) \sim U(0, \theta)$, that is

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

Let prior distⁿ of θ be:

$$\tau(\theta) = \begin{cases} \theta e^{-\theta}, & \theta > 0 \\ 0, & o.w \end{cases} \qquad (\text{i.e } \theta \sim G(2))$$

To find Bayes estimate we have to minimize conditional expected Loss given $X = x$

The joint pdf of X and $\theta = f_1(x, \theta)$

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

$$= \begin{cases} 1/\theta \cdot \theta e^{-\theta}; & 0 < x < \theta \\ & \theta > 0 \\ 0, & o.w \end{cases}$$

$$f_1(x, \theta) = \begin{cases} e^{-\theta}; & 0 < x < \theta,\ \theta > 0 \\ 0; & o.w \end{cases}$$

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 $x > 0$
$\Rightarrow \theta > x$

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

$$= e^{-x} - e^{-\infty} = e^{-x} \Rightarrow h(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & o.w \end{cases}$$

Conditional distribution of $(\theta / X = x)$

$$= g(\theta/x) = \frac{f_1(x,\theta)}{h(x)} = \frac{e^{-\theta}}{e^{-x}} = e^{x-\theta}$$

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

Conditional Expected Loss $= S$

$$= \int L(\theta, d(x)) \cdot d\tau(\theta/x)$$

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

To minimise S,

$$\frac{\partial S}{\partial a} = 0 \Rightarrow C \int_{x}^{\infty} 2(\theta-a)(-1) \cdot e^{x-\theta} \, d\theta = 0$$

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

$$\Rightarrow \int_{x}^{\infty} \theta e^{x-\theta} \, d\theta = a \int_{x}^{\infty} e^{x-\theta} \, d\theta \qquad \left[\begin{array}{l} \because e^{x-\theta} = \tau(\theta/x) \\ \int \tau(\theta/x) = 1 \end{array} \right]$$

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

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

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

$\Rightarrow e^{x}[x e^{-x} - (\cancel{\;} - e^{-x})] = a$

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

$\Rightarrow a = x+1$

$\Rightarrow d(x) = x+1$ = Bayes Estimate of θ w.r.to prior distribution gamma with parameter 2.

Note:- The problem of estimation of a real parameter using quadratic loss, occurs frequently

The posterior loss, given $X=x$ for a quadratic loss function at action 'a' is

$E[L(\theta,a)\,|\,X=x] = c\int(\theta-a)^2\, d\tau(\theta|x)$
= second raw moment about 'a' of posterior $dist^n$ of a given x

If $a = E(x)$, then $E[L(\theta,a)|X=x]$ is min
Hence Bayes decision rule is simply
$\boxed{d(x) = E[\theta|X=x]}$

Rule 1: In problem of estimating a real para θ, with loss proportional to squared error, a Bayes decision rule with respect to a given prior dist $\tau(\theta)$ is, to estimate θ as the mean of the posterior $dist^n$ of θ, given the observation x that is Bayes ~~de~~ Estimate $d(x) = E[\theta|X=x]$
$= \int \theta\, d\tau(\theta|X=x)$

$d(x)$ = median of conditional dist of $\theta|X=x$
(In this case S = Conditional Expected loss
$= C\int|\theta-a|\,\tau(\theta|x)\,dx$)

Rule 2 In problem of estimating a real para θ, with loss proportional to absolute error, a Bayes decision rule w.r. to a given prior dist $\tau(\theta)$ is to estimate θ as the median of the posterior dist of 'θ', given the Observation x ~~of the posterior of θ, given the observation x~~ that is, Bayes estimate

$d(x)$ = median of conditional dist of $\theta | \cancel{X} = x$

(In this case, S = Conditional Expected loss)
$$= C\int |\theta - a| \, \tau(\theta|x) \, d\theta$$

* Generalization of Squared error loss, weighted Squared error loss:-

let $L(\theta, a)$ = weighted Squared error loss fn
$$= w(\theta)(\theta - a)^2$$

Then to find Bayes estimate w.r.t $\tau(\theta)$, we need to minimize

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

$$= \int w(\theta)(\theta - a)^2 \, d\tau(\theta | X = x)$$

To minimize S, $\frac{\partial S}{\partial a} = 0$

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

$$\Rightarrow d(x) = a = \frac{\int \theta\, w(\theta) \, d\tau(\theta|x)}{\int w(\theta) \, d\tau(\theta|x)}$$

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

Ex Let $\Theta = R = \mathcal{a}$, $L(\theta, a) = (\theta - a)^2$ & let $X \sim N(\theta, 1)$
Find Bayes Estimate of θ w.r.to a prior dist of $\theta \sim N(0, \sigma^2) = \tau_{\sigma^2}$ or τ_σ

Given $f(x|\theta) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$; $-\infty < x < \infty$, $-\infty < \theta < \infty$

Now, $\theta \sim N(0, \sigma^2)$

$\Rightarrow \tau_{\sigma^2}(\theta) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{\theta^2}{\sigma^2}}$, $-\infty < \theta < \infty$, $\sigma > 0$

Jt pdf X & $\theta = f_1(x, \theta)$

$= f(x|\theta)\, \tau_{\sigma^2}(\theta)$

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

To obtain marginal distn of x :-

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

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

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

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

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

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

$$f_1(x,\theta) = \frac{1}{2\pi\sigma} e^{-\frac{1}{2}\frac{1+\sigma^2}{\sigma^2}\left(\theta - \frac{x\sigma^2}{1+\sigma^2}\right)^2} e^{-\frac{1}{2}\frac{x^2}{1+\sigma^2}}, \quad -\infty < x < \infty, \ -\infty < \theta < \infty$$

Marginal pdf of $x = f_2(x) = f_1(x,\theta)\, d\theta$

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

$$= \frac{\sigma}{\sqrt{1+\sigma^2}}\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{x^2}{1+\sigma^2}} \underbrace{\int_{\theta=\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{1+\sigma^2}}} e^{-\frac{1}{2}\frac{\left(\theta - \frac{x\sigma^2}{1+\sigma^2}\right)^2}{\sigma^2/1+\sigma^2}} d\theta}_{\int \text{pdf of } N\left(\frac{x\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right) = 1}$$

$$\Rightarrow \cancel{\frac{1}{\sqrt{2\pi}\,\sigma} \times \frac{\sigma}{\sqrt{1}}}$$

$\Rightarrow$ posterior distn of $\theta | x = x$ = conditional dist of $\theta | x = x$

$$= \tau(\theta | x = x)$$

$$= \frac{f_1(x,\theta)}{f_2(x)}$$

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

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

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

8\. The Bayes rule or Bayes estimate w.r.to prior dist τ_σ

$$= d_\sigma(x) = E[\theta / X = x] = \frac{x\sigma^2}{1+\sigma^2}$$

(using Rule 1)

which has the the Bayes Risk

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

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

$$= \text{variance of } \tau(\theta / X = x)$$

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

Note that :, if we take some other dec rule, say $d(x) = x$

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

As $\sigma \to \infty$ then $d_\sigma(x) \to d(x)$ in distribution

★ Useful Extensions in the definition of Bayes Rule

Defn 1:- A rule δ is said to be a limit of Bayes rule δ_n if for almost all x, $\delta_n(x) \to \delta(x)$ in distribution

Remark:- for non-randomized dec rules, this defn becomes $d_n \to d$ if $d_n(x) \to d(x)$ for almost all x in distribution

Ex In the previous Ex, we have seen that ~~$d_\sigma(x)$~~
$d_\sigma(x) \to d(x)$, as $\sigma \to \infty$
$\Rightarrow d(x)$ is a limit of Bayes rule d_σ

Def 2 :- Generalized Bayes Rule

A rule δ_0 is said to be a "generalized Bayes rule" if $\exists$ a measure τ on Θ (or a nondecreasing fn on Θ if Θ is real) 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; here $\tau(\theta)$ need not be a density fn, and hence not necessarily a prior distn of θ

Ex consider prev Ex, where $X \sim N(\theta, 1)$. Let $d(x) = x$ be a generalized Bayes rule when $d\tau(\theta) = d\theta$, i.e $\tau(\theta) = \theta$

To verify $d(x) = x$ is a generalized Bayes rule
let

$$S = \int L(\theta, d) f_X(x|\theta)\, d\tau(\theta)$$

Minimise

$$S = \int (\theta - a)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-a)^2} d\theta$$

$$\frac{dS}{da} = 0 \Rightarrow -\frac{2}{\sqrt{2\pi}} \int (\theta) e^{-\frac{1}{2}(x-\theta)^2} d\theta \quad \text{LHS}$$

$$= \frac{a}{\sqrt{2\pi}} \int_{\theta=-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} d\theta \quad \text{RHS}$$

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

(i.e if we know $\theta \sim N(x,1)$)

$\Rightarrow d(x)$ is a generalized Bayes rule

Def: 3 Extended Bayes Rule

A rule δ_0 is said to be Extended Bayes if δ_0 is ε-Bayes for every $\varepsilon > 0$, that is $\exists$ a prior τ such that δ_0 is ε-Bayes w.r.to τ that is, $r(\tau, \delta_0) \leq \inf_{\delta} r(\tau, \delta) + \varepsilon$

Ex Prev Ex $X \sim N(0,1)$, To show that dec rule $d(x) = x$ is Extended Bayes rule compute

$$r(\tau_\sigma, d) = E\left[E\{(\theta - x)^2 \mid \theta\}\right]$$
$$= E\left[E\{(x-\theta)^2 \mid \theta\}\right]$$
$$= V(x \mid \theta)$$
$$= 1$$

$\left(\because x \sim N(0,1) \Rightarrow E(x) = \theta,\ V(x) = 1 \right)$

But Because $\inf_{\delta} r(\tau_\sigma, \delta) = r(\tau_\sigma, d_\sigma)$

($\because d_\sigma$ is Bayes rule $\Rightarrow r(\tau_\sigma, d_\sigma)$ will be min)

$$= \frac{\sigma^2}{1+\sigma^2} = \text{Bayes risk}$$

$\Rightarrow r(\tau_\sigma, d) = \inf_{\delta} r(\tau_\sigma, d_0) + \varepsilon$ $\left(\text{for } \varepsilon = \frac{1}{1+\sigma^2} \right)$

$\left(\text{Because } \inf_{\delta} r(\tau_\sigma, d_0) + \varepsilon = \frac{\sigma^2}{1+\sigma^2} + \frac{1}{1+\sigma^2} = 1 = r(\tau_\sigma, d_0) \right)$

$\Rightarrow d(x) = d$ is ε-Bayes for every $\varepsilon > 0$

Note:- When a notion of an "almost" Bayes rule is required, the Extended Bayes rule of def 3 will be choosen.

UNIT - II

* Main theorems of Decision theory

* Admissibility and completeness

Def 1: Natural Ordering :- A decision rule δ_1 is said to be as good as a rule δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2) \ \forall \ \theta \in \Theta$

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

Note: (i) The natural ordering gives a partial ordering of the space D^* (or $\mathcal{D}$) of dec rules

(ii) If δ_1 is as good as δ_2 then δ_2 is not to be performed over δ_1
Both the minimax and Bayes ordering satisfy their requirement.

Def 2: Inadmissible :- 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.

Def 3 A class C of decision rule $C \subset (D^*)$ is said to be complete if given any rule $\delta \in D^*$ not in C, $\exists$ a rule $\delta_0 \in C$ that is better than δ.

Essentially Complete:- A class C of dec rules is said to be "essentially complete" if given any rule $\delta \notin C$ $\exists$ a rule $\delta_0 \in C$ that is as good as δ.

* The full lemma will be useful to understand the difference between complete and essentially complete classes of decision rules.

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

Proof:- A is set of all admissible decision rule
Let δ be any admissible decision rule
$\Rightarrow \delta \in A$ Suppose $\delta \notin C$

Since C is a complete class, $\exists$ a decision say $\delta' \in C$ s.t δ' is better than δ
$\Rightarrow \delta$ is inadmissible dec rule
therefore $\delta \notin A$, which is a contradiction
$\therefore \delta \in C$ Thus $\delta \in A \Rightarrow \delta \in C \Rightarrow A \subset C$

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

proof: C is essentially complete classes, δ is admissible and $\delta \notin C$.
$\therefore \exists$ $\delta' \in C$ s.t δ' is as good as δ
(by defn of essentially complete class)

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

If $R(\theta, \delta') < R(\theta, \delta)$ for some $\theta \in \Theta$.
then since $R(\theta, \delta') \leq R(\theta, \delta) \quad \forall\ \theta \in \Theta$
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
$\therefore$ for no. $\theta \in \Theta$, $R(\theta, \delta') \leq R(\theta, \delta)$
$\Rightarrow R(\theta, \delta') = R(\theta, \delta) \quad \forall\ \theta \in \Theta$
$\Rightarrow \delta'$ is equivalent to δ.

<u>Def</u>ⁿ <u>Minimal Complete</u>:- A class C of decision rule is said to be "<u>minimal complete</u>" if C is complete and if no proper subsets of C is complete.

A class C of decision rules is said to be "minimal essentially complete" if C is essentially complete and if no proper subclass of C is essentially complete.

<u>Note</u> ~~It~~ is not necessary that minimal complete or minimal essentially complete classes exists.

<u>Remark</u>:- If the statistician knows the complete (or essentially complete) class of decision rule, then he can find a decision rule just well This class only.
Thus the statistician can simplify his task by finding a small complete (or essentially complete) class from which he can make

his choice. The smallest such class may not exists but if it exists it is called minimal complete (or essentially complete) class, and it gives the maximum reduction of this problem

Thm 1 If a minimal complete class exists it consists of exactly the admissible dec rules.

Proof Let C denote a minimal complete class and let A denote the class of all admissible rule

To s.t $C = A$

lemma 1 $\Rightarrow$ $A \subset C$, because a minimal complete class is complete.

So now it remains to show that $C \subset A$

Assume that $C \not\subset A$

Let $\delta_0 \in C$ and suppose $\delta_0 \notin A$

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

$\Rightarrow$ δ_0 is inadmissible as $\exists$ $\delta_1 \in C$ which is better than δ_0

Because δ_0 is inadmissible, $\exists$ a rule δ 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, hence better than δ_0

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

Now, let $C_1 = C \sim \{\delta_0\}$
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

If $\delta \neq \delta_0$, $\exists\ \delta' \in C$ which is better than δ
if $\delta' = \delta_0$, then $\delta_1 \in C_1$ is better than δ,
if $\delta' \neq \delta_0$, then $\delta' \in C_1$ is better then δ

In any case $\exists$ an element (decision rule) of C_1 better than δ, which means $C_1 \subset C$ is complete

$\Rightarrow$ C is not minimal complete our Assumption that $C \not\subset A$ is wrong
$\Rightarrow C \subset A$
Thus, $A \subset C$ and $C \subset A$ Hence $C = A$

(Converse of the Peev theorem)
If the class of admissible rule is complete then it is minimal complete

Proof: Let A be the class of admissible decision rule which is complete.
Suppose if possible A is not minimal complete
Since, minimal complete class is intersection of all complete classes.

let A 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 $\delta_1 \in A_1$ & δ_1 is better than δ_0

$\Rightarrow \delta_0$ is inadmissible, which is a contradiction, bc A is a class of admissible decision rule and $\delta_0 \in A$

$\Rightarrow A$ has to be a minimal complete class

Result :- ~~Such~~ Show that $\mathcal{T}$ is complete and contains no proper essentially complete subclass, then $\mathcal{T}$ is minimal complete and minimal Essentially complete.

Proof Given $\mathcal{T}$ is complete and if $\mathcal{T}_1 \subset \mathcal{T}$, then $\mathcal{T}_1$ is not essentially complete

(i) To Prove that $\mathcal{T}$ is minimal complete
Given $\mathcal{T}$ is complete $\Rightarrow$ To P.T $\mathcal{T}_1 \subset \mathcal{T}$ is not complete

Suppose $\mathcal{T}_1$ is complete
$\therefore$ If $\delta_0 \notin \mathcal{T}_1$, then $\exists \delta \in \mathcal{T}_1$ which is better than δ_0 i.e

$$\left.\begin{aligned} &R(\theta, \delta) \leq R(\theta, \delta_0) \quad \forall \theta \in \Theta \\ \text{and } &R(\theta, \delta) < R(\theta, \delta_0) \text{ for Some } \theta \in \Theta \end{aligned}\right\} - (1)$$

But $\mathcal{T}_1$ is not essentially complete
$\therefore$ For $\delta_0 \notin \mathcal{T}_1$, $\exists \delta \in \mathcal{T}_1$ which is as good as δ_0

i.e $R(\theta, \delta) \leq R(\theta, \delta_0) \quad \forall \theta \in \Theta$ — (2)

Thus (1) & (2) contradicts each other

$\therefore \tau_1$ is not complete and $\tau_1 \subset \tau$

$\Rightarrow \tau$ is minimal complete class

(ii) To P.t τ is Minimal essentially complete first we will prove τ is essentially complete

Suppose τ is not essentially complete.

$\therefore$ for $\delta_0 \notin \tau$, $\exists\ \delta_1 \in \tau$ such that δ_1 is as good as δ_0 i.e

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

If $R(\theta, \delta_1) \nless R(\theta, \delta_0)$ for some θ, the δ_1 is not better than δ_0

$\therefore \delta_1 \in \tau$ which is better than δ_0

$\therefore \delta_0$ is admissible and $\delta_0 \notin \tau$, but τ is minimal complete ($\because$ Part i). So $\delta_0 \notin \tau$ is a contradiction $\Rightarrow \delta_0 \in \tau$

Also if $R(\theta, \delta_1) \neq R(\theta, \delta_0)$ $\forall\ \theta \in \Theta$ then δ_0 and δ_1 cannot be equivalent

Thus τ is Essentially complete

Let $\tau_1 \subset \tau$. But as per the statement τ_1 is not essentially complete

$\Rightarrow \tau$ is minimal essentially complete

Hence the proof.

Ex let $\Theta = \{\theta_1, \theta_2, \theta_3\}$ $\mathcal{A} = \{a_1, a_2, a_3\}$ $\mathcal{X} = \{x_1, x_2\}$

Risk table

Θ	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
θ_1	0	$1/64$ $=0.0156$	$4/64$ 0.0625	$3/64$ 0.0468	$4/64$	$7/64$	$12/64$	$16/64$	$16/64$
θ_2	$4/64$	$2/64$ 0.0312	$4/64$	$2/64$	0	$2/64$	$4/64$	$2/64$	$4/64$
θ_3	$1/4 = \frac{16}{64}$	$7/64$	$4/64$	$13/64$	$4/64$	$1/64$	$12/64$	$3/64$	0
	✓	✓	×	×	✓	✓	×	×	✓

No dec rule is better than d_1

$\Rightarrow$ d_1 is admissible decision rule d_2 is better than d_4.

$\Rightarrow$ d_4 is inadmissible, d_2 is better than d_7

$\Rightarrow$ d_7 is inadmissible, $\nexists$ any decision rule better than d_2

$\Rightarrow$ d_2 is admissible, d_5 is better than d_3

$\Rightarrow$ d_3 is inadmissible, $\nexists$ any decision rule better than d_5

$\Rightarrow$ d_5 is admissible, d_6 is better than d_8

$\Rightarrow$ d_8 is indmissible, $\nexists$ any decision rule better than d_6 as well as d_9

$\Rightarrow$ d_8 and d_9 are admissible

$\Rightarrow$ Set of admissible decision rules = $\{d_1, d_2, d_5, d_6, d_9\}$
and set of inadm dec rule = $\{d_3, d_4, d_7, d_8\}$

* Complete class which is not minimal complete
= all admissible decision rule +
atleast one inadmissible decision rule

$\Rightarrow$ $C_1 = \{d_1, d_2, d_5, d_6, d_9, d_3\}$ or
$C_2 = \{d_1, d_2, d_5, d_6, d_9, d_4\}$ or
$C_3 = \{d_1, d_2, d_5, d_6, d_9, d_7\}$ or
$C_4 = \{d_1, d_2, d_5, d_6, d_9, d_8\}$

* Minimal complete class = set of admission decision rule
$= \{d_1, d_2, d_5, d_6, d_9\}$

Suppose the prior distribution is
$P(\theta_1) = P(\theta_2) = P(\theta_3) = 1/3$

θ :	θ_1	θ_2	θ_3
$P(\theta)$:	$1/3$	$1/3$	$1/3$

Bayes decision rule = min $r(\tau, d)$
where $r(\tau, d) = E[R(\theta, d)]$
$$= \sum_{i=1}^{3} R(\theta, d_i) \cdot P(\theta_i)$$

$\therefore r(\tau, d_1) = E[R(\theta, d_1)]$
$$= \sum_{i=1}^{3} R(\theta_i, d_1)\, P(\theta_i)$$
$$= (0)\frac{1}{3} + \frac{1}{16} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{20}{192}$$

$r(\tau, d_2) = \frac{10}{192}$, $r(\tau, d_3) = \frac{12}{192}$

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

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

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

$\Rightarrow d_5$ is Bayes dec rule

Important Theorems on
* Decision theory and Game theory

Recall Θ^* = Set of all prob. dist defined over Θ or
Set of all prior distⁿ

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

→ Maximin value $= \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta)$
$= \underline{v}$ = lower value of game — (1)

(1) Represents the 'maximin' or 'lower value of the game'. If 'nature' is an opponent who is playing to ruin the "statistician" then it could use a least favourable distribution if it existed which guarantees it that the statistician expected loss would be at least equal to $\underline{v}$ no matter what decision rule he might use

→ Minimax value $= \inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta)$
= Upper value of the game
$= \bar{v}$ — (2)

(2) Represents the 'minimax' value or "upper value of the game" since for every $\delta \in D^*$

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

(we shall prove (3) later.)
Thus, the statistician has a rule which ensures him that his expected loss will not be greater than any preassigned number larger than the upper value, no matter what prior distn nature decides to use

Recall:- A decision rule δ_0 is said to be minimax decision rule if

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

$(= \bar{v} =$ minimax value$)$

Thus, δ_0 is minimax $\Rightarrow$ upper bound on risk $= \max R(\theta, \delta_0) \Rightarrow$ no matter what strategy /decision, the nature chooses the statistician's risk will never be more then $\bar{v}$.

Similarly, τ_0 is least favourable prior distribution for nature if

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

$=$ maximin value

which means if the nature chooses τ_0, it is ensured that whatever may be the decision rule choosen by the statistician his loss would be atleast $\underline{v}$.

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

Consider a game and assume that the risk set S is bounded from below. Then the game has value.

$$\underline{V} = \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta)$$
$$= \inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta) = \bar{V} \quad \text{--- (4)}$$

(we can also denote $\underline{V} = \bar{V} = V$)

Moreover if S is closed from below then a minimax strategy (i.e a minimax decision rule δ) exists, and
$$r(\tau_0, \delta_0) = V$$

(assuming τ_0 is least favourable prior distribution and δ_0 is minimax decision rule)

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

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

③ In decision theory, the minimax theorem is helpful in helping the statistician to find a minimax decision rule

④ Also the minimax theorem is useful in answering the question when are the degenerated distⁿ = have 1 value at some point

minimax rules also Bayes rule w.r. to some prior distⁿ?

The answer is :- If the minimax theorem holds and if a least favourable prior distⁿ τ_0 exists then any minimax rule so is Bayes with respect to τ_0

<u>Result</u> :- Prove that $\sup_{\tau \in \Theta^*} r(\tau, \delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$

<u>Proof</u> Let Θ^* = Set of all Prob. dist defined over Θ

ω = set of all degenerate distribution defined over Θ

Then $\omega \subset \Theta^*$

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

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

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

Thus (1) & (2) $\Rightarrow \sup_{\theta \in \Theta} R(\theta, \delta) \leq \sup_{\tau \in \Theta^*} r(\tau, \delta) \quad -(3)$

Now consider

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

$$r(\tau,\delta) = E[R(\theta',\delta)]$$

Taking Expectation on both the sides with respect to prior dist τ.

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

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

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

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

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

Theorem :- If minimax thm holds and if there exists least favourable prior dist τ_0, then any minimax decision rule δ_0 will be Bayes with respect to prior distⁿ τ_0.

Proof: Since minimax theorem holds

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

Suppose τ_0 is least favourable dist

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

Let δ_0 be a minimax dec rule

$$\Rightarrow \sup_{\theta \in \Theta} R(\theta, \delta_0) = \sup_{\tau \in \Theta^*} r(\tau, \delta_0) \quad \text{--- (3)}$$

$$= \inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta) \quad \text{--- (4)}$$

($\because$ def^n of minimax rule)

To prove that δ_0 is Bayes with respect to τ_0 we can P.t

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

Consider

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

$$= \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta) \quad (\because (4))$$

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

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

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

Hence equality holds at all places
($\because$ $a > a$ is not possible)

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

Hence δ_0 is Bayes rule.

Complete class theorem :-
(The fundamental theorem of decision theory)

Ques When does the collection of Bayes rules (or extended Bayes rule) form a complete class? What subset of the class of Bayes rules forms a minimal complete class?

Discussion:- An advantage that the Bayes principle has over most other principles leading to optimal rules (including the minimax principle). is that often it leads to rules that are relatively easy to compute. Thus it is essentially important to know whether the statistician can restrict his attention to the class of Bayes rules if it forms a complete class or an essentially complete class.

If the Bayes rules form a complete class, and a minimal complete class exists, it is then of interest to characterize the prior distribution where corresponding Bayes rules form the minimal complete class.

- Thus, other questions of interests are
  - when are Bayes rules admissible?
  - when do Bayes rules exists?
  - Are minimax rules also Bayes rules?
  - Are all admissible rules also Bayes rule?
  - when does a minimal complete class exist?

* There exists relatively complete answer to these questions when (H) is finite for which geometrical interpretation is handy

Partial answer to some of these question exists in case when (H) is infinite

* Use of randomized decision rules has advantage in answering above question, but at the same time it has disadvantage of greatly complicated nature of decision space from which we must choose the decision rule

* Therefore it becomes important to know when it is possible to restrict attention the class D of non-randomized decision rules

* Thus, another important question when is the class D of non-randomized decision rules essentially complete?

* Another most general problem is concerned with distribution of the random variable X observed by the statistician.

These problems are specific to decision theory and do not arise in game theory

For these problem, the choice of decision rules will be based on sufficient statistics

Ques When are Bayes rules admissible?

Thm 1 If for a given prior distⁿ τ, a Bayes rule w.r. to τ is unique upto equivalence, then Bayes rule is admissible.

(Recall :- δ_1 & δ_2 are said to equivalent dec rules if $R(\theta, \delta_1) = R(\theta, \delta_2) \ \forall \theta \in \Theta$)

Proof Suppose δ_0 is Bayes w.r. to prior dist τ and is unique upto equivalent.
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

$$\left.\begin{array}{l} R(\theta, \delta_1) \leq R(\theta, \delta_0) \quad \forall \ \theta \in \Theta \\ R(\theta, \delta_1) < R(\theta, \delta_0) \ \text{for atleast one } \theta \in \Theta \end{array}\right\} \quad (1)$$

let $\Theta = \{\theta_1, \ldots \theta_k\}$ (finite)
$$\therefore r(\tau, \delta_1) \leq r(\tau, \delta_0) \quad \text{--- } (2)$$

Inequality hold in (2) if τ attached positive prob. for these θ for which inequality hold 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 contradiction in either case
Hence δ_0 must be admissible

Thm 2 Assume that $\Theta = \{\theta_1, \dots \theta_k\}$ and a Bayes rule δ_0 w.r.to the prior dist $= (p_1, \dots p_k)$ exists. If $p_j > 0 \ \forall \ j = 1, \dots k$, then δ_0 is admissible.

Proof Suppose δ_0 is inadmissible. Then $\exists$ a decision rule $\delta' \in D^*$ which is better than δ_0 that is,

$$R(\theta_j, \delta') \leq R(\theta_j, \delta_0) \quad \forall j$$
$$\& \quad R(\theta_j, \delta') < R(\theta_j, \delta_0) \quad \text{for same } j$$

Since all p_j are positive.

$$\sum_{j=1}^{k} p_j R(\theta_j, \delta') < \sum_{j=1}^{k} p_j R(\theta_j, \delta_0)$$

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

$\Rightarrow$ δ_0 is not Bayes w.r. to prior dist $\tau = (p_1, \dots p_k)$

($\because$ Recall: A decision rule δ_0 is said to be Bayes w.r. to Prior dist τ if
$$r(\tau, \delta_0) = \inf_{\delta \in D^*} r(\tau, \delta) \text{)}$$

This is a contradiction to the statement. Hence, δ_0 is admissible

Ex To demonstrate that if $P_j > 0, \forall j = 1, \dots k$ is violated in above thm-2 So is not necessarily admissible

Let $\Theta = \{\theta_1, \theta_2\}$, $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$

Suppose dist of r.v x is degenerate at zero $\forall \theta$ $\Rightarrow \mathcal{X} = \{x_1 = 0\}$

$(\because P(x=0) = P(x = x_1) = 1)$ & $P(x \neq 0) = 0)$

Suppose the loss table is

$\Theta \setminus \mathcal{A}$	a_1	a_2	a_3	a_4
θ_1	1	1	2	2
θ_2	0	1	0	1

No of non-randomized decision rules

$= (\text{No. of elements in } \mathcal{A})^{(\text{No. of elements in } \mathcal{X})}$

$= 4^1 = 4$

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

Decision table

$\mathcal{X} \setminus D$	d_1	d_2	d_3	d_4
$x_1 = 0$	a_1	a_2	a_3	a_4

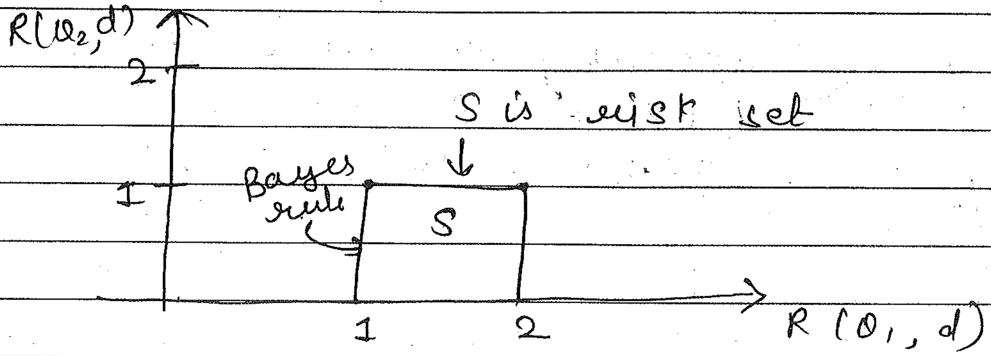
$d_1(x_1) = a_1$, $d_2(x_1) = a_2$, $d_3(x_1) = a_3$, $d_4(x_1) = a_4$

$P_\theta(x)$	d_1		d_2		d_3		d_4	
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
$P(x = x_1)$ $P(x=0)$	1	0	1	1	2	0	2	1
$R(\theta, d)$	1	0	1	1	2	0	2	1

Risk table $R(\theta, d)$

	d_1	d_2	d_3	d_4
θ_1	1	1	2	2
θ_2	0	1	0	1

$\Rightarrow$ d_1 is Bayes rule as $r(\theta, d_1)$ is min for θ_1 & θ_2



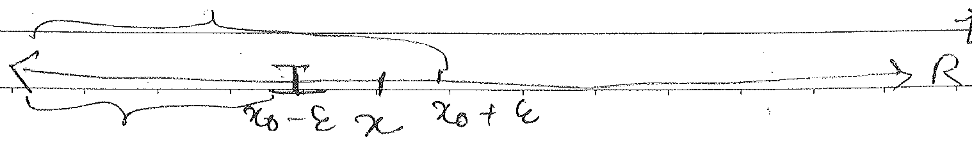
* **Support of a dist^n:-**

(1) Let $\Theta = E_1$ = Euclidian 1-dim space = Real line Θ is infinite.
A point $\theta_0 \in E_1$ is said to be in the support of a dist τ on the real line if for every $\varepsilon > 0$ the interval $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ has positive prob $\tau(\theta_0 - \varepsilon, \theta_0 + \varepsilon) > 0$, that is, if $P(\theta_0 - \varepsilon < \theta < \theta_0 + \varepsilon) > 0$

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

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

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



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

Ex (1) let $X \sim B(n, \theta)$ $\quad 0 < \theta < 1$

Then for $x_0 = 0$, $P(0 - \varepsilon < x \leq 0 + \varepsilon) = P(x = 0) > 0$

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

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

But for $x_0 = 1.5$, $P(1.5 - \varepsilon < x \leq 1.5 + \varepsilon) = 0$

In general for any x other than integers $0, 1, \ldots n$ $\quad P(x - \varepsilon < x \leq x + \varepsilon) = 0$

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

Ex (2) let $X \sim U[0, 1]$

Then $\forall\, x_0 \in [0, 1]$

$$P(x_0 - \varepsilon < x < x_0 + \varepsilon) = \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} 1 . dx = 2\varepsilon > 0$$

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

Thm 3 Let $\Theta = E_1$ 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 probability 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 decision rule $\delta' \in D^*$ which is better than δ_0 that is,

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

$$\& \quad R(\theta, \delta') < R(\theta, \delta_0) \text{ for some } \theta \in \Theta$$

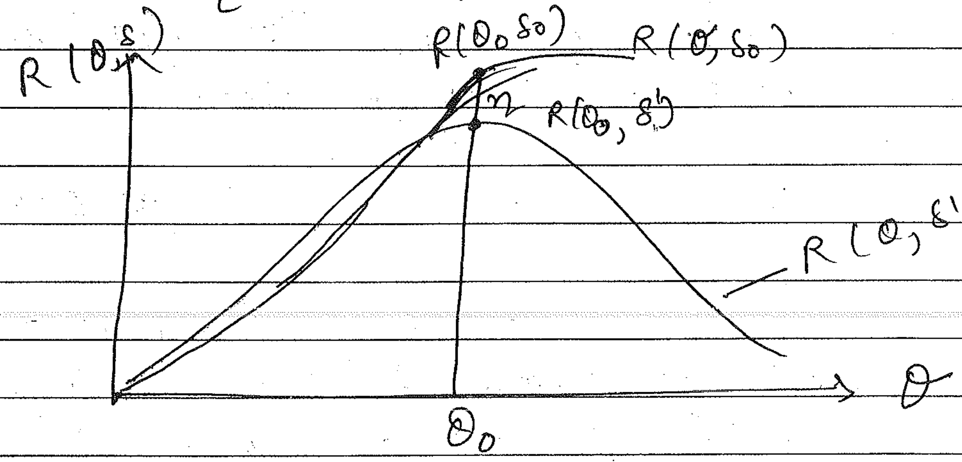
Suppose the ~~second~~ second condition hold for some $\theta_0 \in \Theta$ that is

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

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

$$R(\theta, \delta') \leq R(\theta, \delta_0) = \frac{\eta}{2}$$

whenever $|\theta_0 - \theta| < \varepsilon$
where $\eta = R(\theta_0, \delta_0) - R(\theta_0, \delta') > 0$



Because $R(\theta, \delta)$ is continuous

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

let T be a r.v whose distn is τ
Then

$r(\tau, \delta_0) - r(\tau, \delta')$ = Difference between Bayes risk with respect to prior dist τ

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

$$\geq \frac{\eta}{2}\, \underbrace{\tau(\theta_0 - \varepsilon, \theta_0 + \varepsilon)}_{\text{Support = Prob} > 0} > 0$$

(Since θ_0 is in the support of τ)

$\Rightarrow r(\tau, \delta_0) > r(\tau, \delta')$ which

contradicts the assumption that δ_0 is a Bayes rule w.r. to τ
Hence δ_0 is admissible

Def **ε- admissible decision rule:-**

A decision rule δ_0 is said to be ε-admissible decision rule if there does not exist any rule say δ_1 such that

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

Result:- If δ_0 is ε - Bayes with $\varepsilon > 0$ then it is ε- admissible (no restriction on Θ or on the distribution τ with respect to which δ_0 is ε- Bayes.

* Assumptions for the problem raised:-

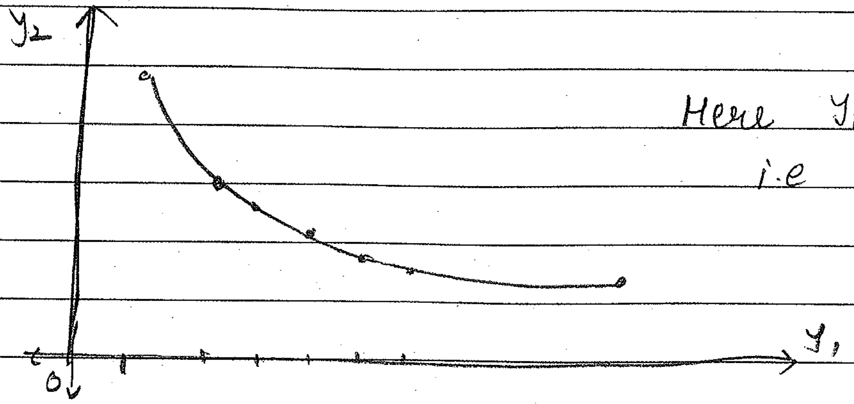
Bounded from ~~and~~ ~~Bounded~~ below:-
A set S in k-dimensional Euclidian space E_k, is said to be bounded from if $\exists$ a finite no. m, such that for every $\underline{y} = (y_1, \dots y_k) \in S$

$$y_j > -M \quad \text{for } j = 1 \dots k.$$

Ex Consider a two-dimensional space E_2

Let $S = \{(y_1, y_2) ;\ y_1 y_2 = 1,\ y_2 > 0\}$

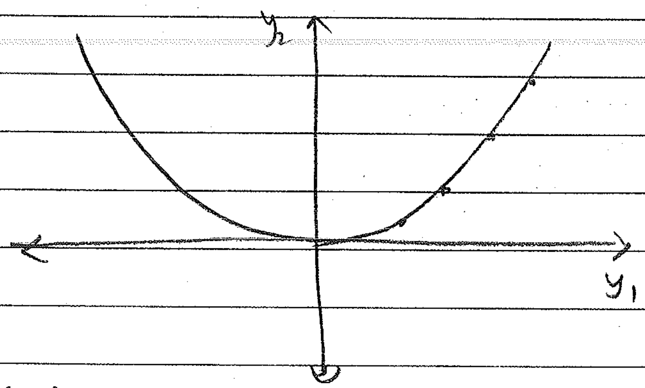
Then S will be bounded from below and $y_1 > 0$



Here $y_1, y_2 > 0$
i.e $m = 0$

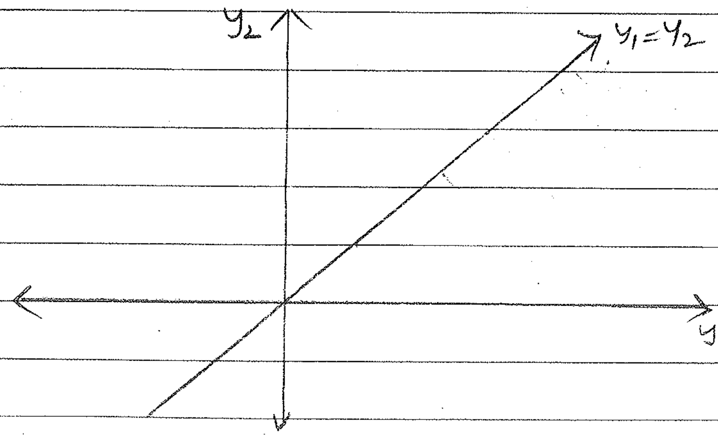
Note:- If condition $y_2 > 0$ is not there then S would not be bounded from below.

Ex (2) let $S = \{ (y_1, y_2) , y_1^2 = y_2, \ y_2 > 0 \}$



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

Ex (3) let $S = \{ (y_1, y_2) , y_1 + y_2 < 1 \}$



This set S is not bounded from below.

Def^2 : let $\underline{x}$ be a point in E_k, The lower quadrant at x, denoted by Q_x is defined as the set.

$$Q_x = \{ \underline{y} \in E_k : y_j \leq x_j \text{ for } j = 1, \dots k \}$$

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

Def^n 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 points of S or, alternatively, $\bar{S}$ is the smallest closed set containing S.

A point $\underline{x}$ is said to be a lower boundary point of a convex set $S \subset E_k$, if

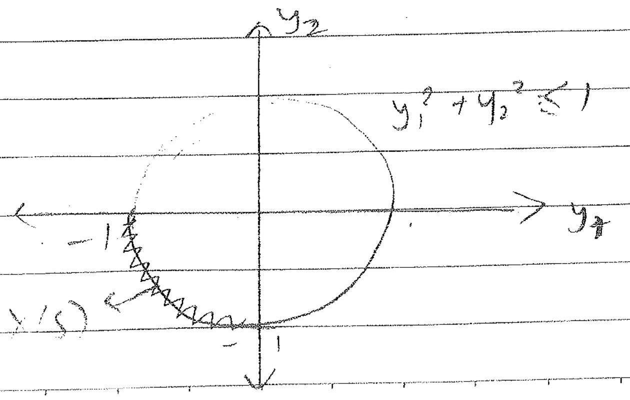
$$Q_{\underline{x}} \cap \bar{S} = \{ \underline{x} \}$$

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

Ex(1) Consider E_2 and let

$$S = \{ (y_1, y_2), y_1^2 + y_2^2 \leq 1 \}$$

Circle with randius 1 / unit disk

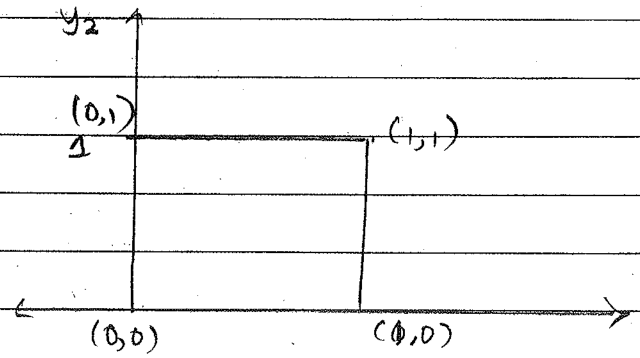


$\lambda(S) = \{(y_1, y_2) : y_1^2 + y_2^2 = 1, \ y_1 \leq 0, \ y_2 \leq 0\}$

only boundary points

Ex(2) let $S_1 = \{(y_1, y_2) ; 0 \leq y_1 \leq 1, 0 < y_2 \leq 1\}$

unit Square



$\Rightarrow$ $\lambda(S_1)$ = set of lower boundary points $= \{0, 0\}$

★ The important of the lower boundary of the convex set S is that the elements of $\lambda(S)$ lead to the admissible decision rules.

Defn 4: A convex set $S \subset E_x$ is said to be Closed from below if $\lambda(S) \subset S$

Example(1) unit disk :- S is closed from below
Example(2) unit Square :- S_1 is closed from below

Note Any closed convex set is closed from below.

— x —

Recall :- let (X, d) be a metric space & let $E \subset X$ and let 'd' be a distance measure.

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

(2) A point $x \in X$ is a limit point of set E if every neighbourhood of x contains a points $y \neq x$ such that $y \in E$ that is for every $r > 0$, $\exists$ a point $y \in E$, $y \neq x$, st $y \in N_r(x)$

(3) Set $E \subset X$ is closed if every limit point of E is a point of E (i.e limit pt of E may or may not belong to E)

(4) E is bounded if $\exists$ a real no. M and a point $y \in X$ such that $d(x,y) < M$ for all $x \in E$.

(5) let E' be the set of all limit points of E in X Then the closure of E is the set

$$\bar{E} = E \cup E'$$

* Thm 1: If $\underline{x} = (R(\theta_1, \delta_0), \dots R(\theta_k, \delta_0))$ is in $\lambda(S)$, then δ_0 is admissible

Proof let $\underline{x}$ be the risk point corresponding to dec rule δ_0 and $\underline{x} \in \lambda(S)$.
Suppose δ_0 is inadmissible

$\Rightarrow \exists$ a decision rule say, δ_1, such that

$$\left.\begin{aligned} &R(\theta_j, \delta_1) \leq R(\theta_j, \delta_0) \quad \forall\ j = 1, 2, \dots k \\ &\& \\ &R(\theta_j, \delta_1) < R(\theta_j, \delta_0) \quad \text{for some } j \end{aligned}\right\} \quad (1)$$

let $y_j = R(\theta_j, \delta_1)$; $j = 1, 2, \dots k$
Then $\underline{y} = (y_1, \dots y_k) \in S$

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

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

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

Thus $Q_x \cap \bar{S} = \{ y \}$ $\Rightarrow$ $x \notin \lambda(S)$

i.e $\underline{x}$ is not a lower boundary point of 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 $x \notin \lambda(S)$ $\Rightarrow$ $\underline{x}$ is not a lower boundary point

$\Rightarrow \exists$ $\underline{y} \neq \underline{x}$ and $y \in Q_x \cap \bar{S}$
$\Rightarrow$ $\underline{y} \in \bar{S}$ and $y \in Q_x$
$\Rightarrow$ $y_j < x_j$ $\forall$ $j = 1, \ldots k$
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
$\Rightarrow$ δ_0 is inadmissible, which is a contradiction

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

Existence of Bayes Rule

Assume that Θ is finite and let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$

Thm 1 Suppose $\Theta = \{\theta_1, \ldots \theta_k\}$ and the risk set S is bounded from below and closed from below. Then, for every prior dist $(P_1, \ldots P_k)$ for which $P_j > 0$ $j = 1, \ldots k$ a Bayes rule w.r.to $(P_1, \ldots P_k)$ exists

Proof let $(P_1, \ldots P_k)$ be a prior dist over Θ for which $P_j > 0$ $\forall j$

let b = Bayes risk of a decision rule whose point is $\underline{y}$, $\underline{y} \in S$

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

& $B = \{b ;\ b = \sum_{j=1}^{k} P_j y_j,\ \underline{y} \in S\}$

= Set of all Bayes risks

Since S is bounded from below, B is bounded from below.
let b_0 = glb of B

Consider a seq of points $\underline{y}^{(n)} \in S$ such that

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

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

Thus seq $\underline{y}^{(n)}$ is bounded below and bounded above both $\Rightarrow$ it is bounded

$\therefore$ sequence $\underline{y}^{(n)}$ has limit point
let $\underline{y}^{(0)}$ be the limit point of sequence $\underline{y}^{(n)}$

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

Now $\underline{y}^{(0)} \in Q_{\underline{y}^{(0)}}$ & $\underline{y}^{(0)} \in \bar{S}$
($\because \underline{y}^{(0)}$ is a limit point of $\underline{y}^{(n)}$)

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

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

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

$\Rightarrow$ $y_j' \leq y_j^{(0)}$ $\forall j$
and
$y_j' < y_j^{(0)}$ for some j

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

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

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

$\Rightarrow$ $\underline{y}^{(0)} \in \bar{S}$ and since S is closed from below,
$\exists$ δ_0 whose risk point is $\underline{y}^{(0)}$, and

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

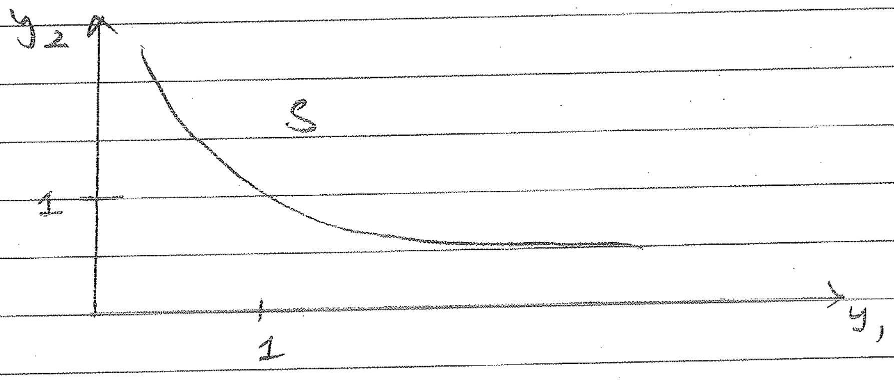
$\Rightarrow$ δ_0 is Bayes rule w.r.t. to τ_0.

Note The condition $p_j > 0$ is necessary in the above thm.

Ex Take $k=2$ and suppose

$S = \{(y_1, y_2) ; y_1 y_2 \geq 1, y_1 > 0\}$

S is convex, bounded from below, and closed from below.



All points on the boundary of S are lower boundary points

$\lambda(S) = \{(y_1, y_2) ; y_1 y_2 = 1, y_1 > 0\}$

$\Rightarrow \lambda(S) \subset S$

Consider a prior dist $\tau = (1, 0)$

for $y \in S$

$$\sum_{j=1}^{k} p_j y_j = p_1 y_1 + p_2 y_2$$

$$= (1) y_1 + (0) y_2 = y_1$$

Now $\inf_{\delta} r(\tau, \delta) = 0$

But there is no decision rule where Bayes risk is zero, for this prior dist τ

$\Rightarrow p_j > 0 \ \forall j$ is necessary

* Result: If the risk set S is closed from below & bounded (above as well as below), the Bayes solutions w.r. to all prior distn exist.

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

* Existence of a Minimal Complete class

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

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

is minimal complete class.

Proof: First we shall prove that D_0 is complete class. Let δ be any decision rule such that

$\delta \notin D_0$ but $\delta \in D^*$

Let S = set of risk points corresponding to decision rule $\delta \notin D_0$

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 $S_1 = Q_{\underline{x}} \cap \bar{S}$ $\therefore$ S_1 is non empty convex

($\because$ closure of a convex set is convex $\}$

and the intersection of two convex sets is convex) and ~~the~~ bounded from below. Then using previous lemma that ; if a non-empty convex set S is bounded from below, then $\lambda(S)$ is not empty, $\lambda(S_1)$ is not empty.

let $\underline{y} \in \lambda(S_1)$, then $\{y\} = Q_y \cap \bar{S}_1$
(by def^n of $\lambda(S_1)$)

Further $\underline{y} \in Q_x$;

Because $\underline{y} \in \bar{S}_1$, $(\bar{S}_1 = \overline{Q_x \cap S})$

$\Rightarrow \underline{y} \in \overline{Q_x \cap \bar{S}}$

$\Rightarrow \underline{y} \in \bar{Q}_x \cap \bar{S}$

$\Rightarrow \underline{y} \in \bar{Q}_x$

$\Rightarrow \underline{y} \in Q_x$

$$\overline{Q_x \cap \bar{S}} = \bar{Q}_x \cap \bar{S} = \bar{Q}_x \cap S$$

Also we know $A \cap B \subset A$ and $A \cap B \subset B$

So, $\bar{Q}_x \cap S \subset \bar{Q}_x$

$(\because y_j \leq x_j)$

Also, $\underline{y} \in \lambda(S)$ because

$$\{y\} = Q_{\underline{y}} \cap \bar{S}_1 = Q_{\underline{y}} \cap \overline{Q_{\underline{x}} \cap \bar{S}}$$

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

$(\because (Q_{\underline{x}} \cap \bar{S}) \subset \bar{S})$

Thus, because S is closed from below, $\exists$ a decision rule $\delta_0 \in D_0$ for which

$$\underline{y} = (R(\theta_1, \delta_0), \ldots R(\theta_k, \delta_0))$$

and it is better than δ since $\underline{y} \in Q_{\underline{x}} \sim \{x\}$

This means D_0 is complete

Now, using thm; if $\underline{x} = (R(\theta_1, \delta_0), \ldots R(\theta_k, \delta_0))$ is in $\lambda(S)$, then δ_0 is admissible

therefore δ_0 is admissible $(\because \underline{y} \in \lambda(S))$

$\Rightarrow$ 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
Hence the proof

* Separating Hyperplane Theorem:-

Roughly - this theorem states that any two disjoint convex sets can be separated by a plane

lemma:- If S is closed convex subset of E_k and $\underline{0} \notin S$ then $\exists$ a better vector $\underline{P} \in E_k$ such that $\underline{P}'\underline{X} > 0$ for all $\underline{X} \in S$

lemma 2: If S is a 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 point of S (that is either $\underline{x}_0 \notin S$ or x_0 is a boundary point of S) then $\exists$ a vector $P \in E_k$, $\underline{P} \neq \underline{0}$ such that

$$\underline{P}'\underline{x} \geq \underline{P}'\underline{x}_0 \quad \text{for all } \underline{x} \in S$$

Thm 2: Separating Hyperplane Thm:-
Let S_1 and S_2 be disjoint convex subsets of E_k. Then $\exists$ a vector $\underline{P} \neq \underline{0}$ such that

$$\underline{P}'\underline{y} \leq \underline{P}'\underline{x} \quad \text{for all } x \in S_1 \text{ and all } \underline{y} \in S_2$$

* Minimax Theorem (Statement) :-

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

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

$\underline{V}$ Maximum value $\qquad \overline{V}$ = Minimax value

and $\exists$ a least favourable prior dist τ_0. moreover if S is closed from below, then $\exists$ an admissible minimax dec rule δ_0 which is Bayes w.r.to τ_0

Result: For any decision problem (Θ, D, R)

$$\underline{V} \leq \overline{V}$$

Recall :-

(1) A point $x \in X$ is an 'interior' point of $E \subset X$ if there is a neighborhood N of x such that $N \subset E$ that is, if $\exists$ a radius $r > 0$ such that $N_r(x) \subset E$

(2) A set E is said to be 'open' if every point of E is an interior point of E

Ex Let $\Theta = \mathcal{A} = \{1, 2, 3, \ldots\}$

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

Thus

$\Theta \backslash \mathcal{A}$	1	2	3	4	...	Inf $r(\tau, \delta)$
1	0	-1	-1	-1	...	-1
2	1	0	-1	-1	...	-1
3	1	1	0	-1	...	-1
4	1	1	1	0	...	-1
⋮	⋮	⋮	⋮	⋮		⋮
Sup $r(\tau, \delta)$	1	1	1	1	...	

Suppose observable r.v. x is degenerate at zero
$\Rightarrow P_\theta(X=0) = 1 \quad \forall \ \theta \in \Theta$

Thus $\mathcal{X} = \{0\}$

D = set of Non-randomised dec rule

No of non - ~~no~~ randomized dec rule $= \infty^\infty = \infty$

Let $D = \{ d_i(\theta) = i \ , \ i = 1, 2, \ldots \}$

let δ be a randomized decision rule

Then $R(\theta_i, d_i) = E[L(\theta_i, d_i)]$

$$= \sum_{i=1}^{\infty} L(\theta_i, d_i) P_i$$

$$= \sum_{i=1}^{\infty} L(\theta_i, d_i) \quad (\because P_i = 1 \text{ as r.v is degenerate})$$

$\therefore$ Risk table is same as above loss ~~table~~

$\therefore \sup\limits_{\tau} r(\tau, \delta) = 1 \quad \forall \delta$

and $\inf\limits_{\delta} r(\tau, \delta) = -1 \quad \forall \tau$

So

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

and

$$\bar{v} = \inf_{\delta} \sup_{\tau} r(\tau, \delta) = 1$$

Thus $\underline{v} \neq \bar{v}$, hence the game does not have a value

* Thm 1 :-

If δ is admissible and Θ is finite, then δ is Bayes rule w.r.to some prior distribution

Thm 2: Complete Class Theorem

If, for a given decision problem (Θ, 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 form a minimal complete class.

Ex Suppose $\Theta = \{\frac{1}{3}, \frac{2}{3}\}$ $\mathcal{a} = R$

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

Suppose a coin is tossed with prob θ of getting a head

Find Bayes rule w.r.to prior distⁿ τ which assign prob π to $\theta = \frac{1}{3}$ and $1-\pi$ to $\theta = \frac{2}{3}$
$(0 \leq \pi \leq 1)$

Ans Here $\mathcal{X} = (H, T)$

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

Let us define a general non-randomized decision rule

$$d(H) = x, \quad d(T) = y, \quad x, y \in R$$

$$\therefore (x, y) \in R^2$$

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

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

$$= \sum_{a} (\theta - a)^2 \times \text{Prob}$$

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

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

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

and

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

Now $r(\tau, (x,y)) = E[R(\tau, d)] =$ Bayes Risk

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

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

we want to find Bayes rule, that is a rule for which Bayes risk is min

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

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

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

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

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

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

$$\Rightarrow (6-3\pi)x = 4 - 3\cdot\pi$$

$$\Rightarrow x = \frac{4-3\pi}{6-3\pi} = \frac{8-3}{12-3} = \frac{5}{9}$$

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

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

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

$$\Rightarrow \frac{4}{9} + \left(\cancel{\frac{4}{9}} - \cancel{\frac{4}{9}}\right)\pi + \underbrace{\left(\frac{2}{3}-\frac{4}{3}\right)}_{=(-2/3)}\pi y = \frac{2}{3}y$$

$$\Rightarrow \frac{4}{9} = \frac{2}{3}y + \frac{2}{3}\pi y = \left(\frac{2}{3}+\frac{2}{3}\pi\right)y$$

$$\Rightarrow y = \frac{\frac{4}{9}}{\frac{2}{3}(1+\pi)} = \frac{2}{3(1+\pi)} = \frac{2}{3\left(\frac{2+1}{2}\right)} = \frac{4}{9}$$

Ex 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 τ.
Show that the Bayes rule w.r. to prior dist τ' having same first moments as τ are same.

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

As earlier Ex. general non-randomised dec rule can be defined as

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

and

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

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

$$= \sum_a (\theta - a)^2 \cdot \text{Prob}$$

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

and

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

$$= E[\cancel{\theta^3} - 2\theta^2 x + x^2\theta + \theta^2 - 2\theta y + y^2 - \cancel{\theta^3} + 2\theta^2 y - \theta y^2]$$

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

$$\theta^3 - 2x\theta^2 + x^2\theta + (\theta^2 - 2y\theta + y^2)(1-\theta)$$
$$+ \theta^2 - 2y\theta + y^2 - \theta^3 + 2y\theta^2 - y^2\theta$$

Since τ is a prior distribution of θ, let m_1' and m_2' be the first two raw moments of distribution τ

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

Also since first two moments are same for all prior dist, the Bayes rule will be the one for which this Bayes risk is minimum.

$\therefore$ To find the Bayes rule means to find x & y such that $r(\tau,(x,y))$ is min.

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

$\Rightarrow$ $-2m_2' + 2x\,m_1' = 0 \Rightarrow x = \frac{m_2'}{m_1'}$

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

$\Rightarrow$ $2m_2' - 2y\,m_1' + 2y - 2m_1' = 0$

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

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

* Solving for minimax rule

Method (1)

Thm If δ_0 is Bayes rule w.r.to τ_0 and for all $\theta \in \Theta$, $R(\theta, \delta_0) \leq r(\tau_0, \delta_0)$ — (1)

Then the game has a value, δ_0 is minimax rule and τ_0 is least favourable Prior dist
(This thm means that, guess a least favourable Prior dist. 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,
$r(\tau_n, \delta_n) \to c$ and $R(\theta, \delta_n) \leq c \quad \forall \theta \in \Theta$
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 receive possible weight from any least favourable dist

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

Thm If An equilizer rule is extended Bayes it is a minimax rule.
(Thus, in second method we search for an equilizer rule in Order to find a minimax rule)

Ex let $\Theta = \{\frac{1}{3}, \frac{2}{3}\}$, $\mathcal{A} = R$, $L(\theta, a) = (\theta - a)^2$

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

Sol $D = \{(x,y),\ x,y \in R,\ d(H) = x,\ d(T) = y\}$

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

To find an equilizer rule
consider a prior dist $\tau_0 = \left(\frac{1}{2}, \frac{1}{2}\right)$

$$\therefore\ r(\tau_0,(x,y)) = \frac{1}{2} R\left(\frac{1}{3},(x,y)\right) + \frac{1}{2} R\left(\frac{2}{3},(x,y)\right)$$

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

$$= \frac{1}{2\cdot 27}\left[(1-3x)^2 + 2(1-3y)^2 + 2(2-3x)^2 + (1-3y)^2\right]$$

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

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

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

$\Rightarrow (1-3x) + 4 - 6x = 0 \Rightarrow 9x = 5$

$\Rightarrow x = \frac{5}{9}$

and $\frac{d}{dy} r(\tau_0, (x,y)) = 0$

$\Rightarrow 4(1-3y)(-3) + 2(2-3y)(-3) = 0$

$\Rightarrow 2-6y+2-3y=0 \Rightarrow 4-9y=0$

$\Rightarrow y = 4/9$

$\therefore$ The Bayes rule w.r. to prior dist $\tau = \left(\frac{1}{2}, \frac{1}{2}\right)$ is $\delta_0 = \left(\frac{5}{9}, \frac{4}{9}\right)$

where $d(H) = \frac{5}{9}$, $d(T) = \frac{4}{9}$

$$\therefore r(\tau_0, \delta_0) = \frac{1}{54}\left[\left(1-3\cdot\frac{5}{9}\right)^2 + 2\left(1-3\cdot\frac{4}{9}\right)^2 + 2\left(1-3\cdot\frac{5}{9}\right)^2 + \left(2-3\cdot\frac{4}{9}\right)^2\right]$$

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

Also

$$R\left(\frac{1}{3}, \delta_0\right) = \frac{1}{3}\left(\frac{1}{3}-\frac{5}{9}\right)^2 + \frac{2}{3}\left(\frac{1}{3}-\frac{4}{9}\right)^2 = \frac{2}{81}$$

and $$R\left(\frac{2}{3}, \delta_0\right) = \frac{2}{3}\left(\frac{2}{3}-\frac{5}{9}\right)^2 + \frac{1}{3}\left(\frac{2}{3}-\frac{4}{9}\right)^2 = \frac{2}{81}$$

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

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

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

Since $\Theta = [0,1]$ it is difficult to guess a least favourable prior dist.

So, we will try to find an equilizer rule for, equilizer rule $R(\theta,(x,y))$ must be indept of parameter θ.

So equate coeff of θ^2 & θ in $R(\theta,(x,y))$ to zero

$\Rightarrow$ $1+2y-2x = 0 \Rightarrow 2x = 1+2y \Rightarrow x = \frac{1+2y}{2}$

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

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

$\Rightarrow 1-4y = 0 \Rightarrow y = \frac{1}{4}$

$\Rightarrow x = \frac{1+2(1/4)}{2} = \frac{3}{4}$

$\therefore$ Equilizer rule is : $d(H) = \frac{3}{4}$, $d(T) = \frac{1}{4}$

$R(\theta,(x,y)) = 0 + 0 + y^2 = \frac{1}{16}$

= Risk of an equilizer rule

* Suppose $x = \frac{m_2'}{m_1'}$, $y = \frac{m_1'-m_2'}{1-m_1'}$

Is this a Bayes rule?

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

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

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

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

$$\Rightarrow 3 - 4m_2' = 4m_2' \Rightarrow 8m_2' = 3$$

$$\Rightarrow m_2' = \frac{3}{8} \quad \text{and} \quad m_1' = \frac{1}{2}$$

From these two moments if we guess any prior dist^n bet^n $[0,1]$, say for Ex,

$\beta_1(\alpha, \beta)$ with $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$, $0 < x < 1$, $\alpha, \beta > 0$

Then

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

we know

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

Take $r = 1 \Rightarrow E(x) = m_1' = \frac{\beta(1+\alpha, \beta)}{\beta(\alpha,\beta)} = \frac{\alpha}{\alpha+\beta}$

& $r=2 \Rightarrow E(X^2) = m_2' = \dfrac{\beta(2+\alpha, \beta)}{\beta(\alpha, \beta)}$

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

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

$\Rightarrow$ Dec rule $d(H) = x = \frac{3}{4}$, $d(T) = y = \frac{1}{4}$ is Bayes rule

Ex Let $X \sim B(n, \theta)$, $\Theta = [0,1] = \mathcal{A}$
$L(\theta, a) = (\theta - a)^2$

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

(i) Find Bayes estimate of θ

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

posterior dist $(\theta / X = x) \sim \beta_1(x+\alpha, n+\beta-x)$
and
$d_{\alpha,\beta}(x) = \dfrac{x+\alpha}{n+\alpha+\beta}$ = Bayes Estimate of θ

(ii) Is there an equilizer rule?

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

A dec rule is equilizer rule if it is indept of θ

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

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

$\Rightarrow \frac{\alpha^2+\beta^2+2\alpha\beta-n}{(n+\alpha+\beta)^2}=0$ & $\frac{n-2\alpha^2-2\alpha\beta}{(n+\alpha+\beta)^2}=0$

($\because$ Coeff of θ^2 & θ should be zero)

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

$\alpha^2+\beta^2+2\alpha\beta-n=0 \Rightarrow \alpha^2+\beta^2+\cancel{2\alpha\beta}=2\alpha^2+\cancel{2\alpha\beta}$

$\Rightarrow \alpha^2-\beta^2=0 \Rightarrow (\alpha-\beta)(\alpha+\beta)=0$

and if $g(\theta) \sim B_1\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$

$\therefore \alpha^2-\beta^2=0 \Rightarrow \alpha=\beta=\frac{\sqrt{n}}{2}$

$$\therefore d(x) = \frac{x+\frac{\sqrt{n}}{2}}{n+\frac{\sqrt{n}}{2}+\frac{\sqrt{n}}{2}} = \frac{x+\frac{\sqrt{n}}{2}}{n+\sqrt{n}}$$

is a minimax dec rule

Ex let $X \sim B(n, \theta)$ & $L(\theta, a)$ = weighted sq error

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

Show that $d(x) = \frac{x}{n}$ is minimax rule

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

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

$$= \frac{1}{n^2 \theta(1-\theta)} n\theta(1-\theta) = \frac{1}{n}$$ indep of θ

$\therefore$ $d(x) = \frac{X}{n}$ is an equilizer rule

Suppose prior distn of θ is $U(0,1)$

$$g(\theta) = 1, \quad 0<\theta<1$$
$$= 0 \quad \text{o.w}$$

$\therefore$ $f(x, \theta)$ = dist of x & θ

$$= \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad ; \quad x = 0, \ldots n, \quad 0<\theta<1$$

$h(x)$ = marginal dist of x

$$= \int_{\theta=0}^{1} f(x, \theta)\, d\theta$$

$$= \binom{n}{x} \beta(x+1, n-x+1)$$

$\tau(\theta | X=x)$ = posterior dist of θ given $X=x$

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

$$= \frac{\theta^{x}(1-\theta)^{n-x}}{\beta(x+1, n-x+1)}$$

To find Bayes rule :-

$d(x)$ = Bayes Estimate of θ

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

$$E[\theta\,\omega(\theta) | X=x] = \frac{n+1}{n-x}$$

$$E[\omega(\theta) | X=x] = \frac{n(n+1)}{x(n-x)}$$

$\therefore$ $d(x) = \frac{X}{n}$ & is minimax ~~bec~~

bec $R(\theta, d)$ is indept of θ

UNIT - 3

[Recall]

Sufficient statistics:-

Let X denote a R.V whose distribution depends on a parameter $\theta \in \Theta$. A real-valued function T of X (i.e $T(X)$) is said to be 'sufficient' for θ if the conditional dist of X, given $T=t$, is independent of θ

Eg If $X_1, \ldots X_n$ are iid $B(1,P)$ r.s then

$$P(X_j = x) = p^x (1-p)^{1-x} \quad ; x = 0, 1$$

let $T(x) = T = \sum_{i=1}^{n} x_i$

Then $$f(x_1, \ldots x_n \mid T=t) = \frac{P(X_1 = x_1, \ldots X_n = x_n \mid T=t \mid P)}{P(T=t \mid P)}$$

$$= \frac{P(X_1 = x_1, \ldots X_n = x_n \mid P)}{\binom{n}{t} p^t (1-p)^{n-t}} \rightarrow (\because \Sigma x_i = T)$$

$$(\because T = \sum_{i=1}^{n} x_i \sim B(n, p))$$

$$= \frac{p^{x_1}(1-p)^{1-x_1} \ldots p^{x_n}(1-p)^{x_n}}{\binom{n}{t} p^t (1-p)^{n-t}}$$

$$= \frac{p^{\Sigma x_i} (1-p)^{n - \Sigma x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{1}{\binom{n}{t}}$$

which is independent of p

$\Rightarrow$ $T(x) = \Sigma x_i$ is sufficient for p

Thm: <u>The factorization Theorem</u>:-

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

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

* <u>Essentially complete class of decision rule based on sufficient statistics</u> :-

The notion that a sufficient statistics carries all information the ~~true~~ sample has to give about the true value of the parameter can be formalize for decision rules based on a sufficient statistics forms an essentially complete class.

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

Thm **Rao - Blackwell Thm:-**

Let $\mathcal{a}$ be a convex subset of E_k, 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 statistic for θ. If $d(x)$ is a non-randomized decision rule then the non-randomized decision rule based on T

$$\hat{\theta}(t) = E(d(x) \mid T = t)$$

provided this expectation exists, is as good as d.

Note:- Rao - Blackwell thm gives the explicit formula by which a non randomised dec rule may be improved by a non-randomized rule based on a sufficient statistic.

Ex. Let $X_i \sim N(\theta, \sigma^2)$ where $\sigma^2 = 1$ $i = 1, 2, \ldots n$
The problem is to estimate θ, using squared error loss f^n $L(\theta, a) = (\theta - a)^2$

It is known that $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ in this case

A reasonable estimate of a, when the assumption of normality is doubtful but when symmetry seems reasonable, is the median of x.

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

$$\hat{\theta}(T) = E[\text{median } x_i \mid T]$$
$$= \left(\frac{1}{n}\right) T = \bar{x} \quad (\because \text{ Symmetry})$$

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

$$\Rightarrow E(T/n) = \theta$$

T/n is unbiased Estimator of θ

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

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

whereas under the assumption of symmetry (but not normality), then median x is unbiased estimator of θ but

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

Here $\bar{x}$ is as good as median of x_i for loss fn $L(\theta, a) = |\theta - a|$ or
$L(\theta, a) = (\theta - a)^2$
with respect to $L(\theta, a) = (\theta - a)^2$
neither of these estimates is very good

Def **Exponential family of distributions :-**

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

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

Result :- If $x_1, \dots x_n$ is a sample of size n from exponential family of distribution then

$$T = (T_1, \dots T_k) = \left(\sum_{j=1}^{n} t_1(x_j) \dots \sum_{j=1}^{k} t_k(x_j)\right)$$

is a sufficient statistic

Lemma :- Let $x_1, \dots x_n$ be a sample from the Exponential family of distribution then the distribution of the sufficient statistic T has prob f^n of the form

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

* **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 far such a reduction can be carried for a given problem.

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

The property of completeness is somewhat stronger then the property of minimal sufficient statistics

Defⁿ:- A sufficient statistics 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.

$$E_\theta(g(T)) = 0 \Rightarrow P_\theta\{g(T) = 0\} = 1$$

Note (i) Completeness $\Rightarrow$ bounded completeness but converse is not necessarily true

(ii) T is boundedly complete sufficient stat if it is sufficient, and if g is a bounded fⁿ for which $E_\theta[g(T)]$ exists and is equal to zero for all $\theta \in \Theta$ we have $g(t) = 0 \quad \forall t$

Ex let $x_1, \ldots x_n$ be a sample from $B(m, \theta)$. Then $T = \sum_{j=1}^{n} x_j$ is sufficient for θ and

$T \in B(N, \theta)$, where $N = nm$.

If $E_\theta[g(T)] = 0 \ \forall \ 0 < \theta < 1$ is

$$\text{if } \sum_{t=0}^{N} g(t) \binom{N}{t} \theta^t (1-\theta)^{N-t} = 0 \quad \forall \ 0 < \theta < 1$$

$$\Rightarrow \sum_{t=0}^{N} g(t) \binom{N}{t} \left(\frac{\theta}{1-\theta}\right)^t = 0 \quad \forall \ 0 < \theta < 1$$

$$\text{--- (1)} \quad (\because (1-\theta)^N \text{ is const})$$

If a convergent Power series $\sum a_n z^n$ is zero for z in some open interval, then each of the coefficients 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 is a ~~power~~ & identically equal to zero, So that $g(t) \binom{N}{t} = 0, \ \forall \ t = 0, 1, \ldots N$

i.e $g(t) = 0$, for $t = 0, 1, \ldots N$

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

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

Ex let $x_1, \ldots x_n$ be a sample from $U(0,\theta)$ $\theta > 0$
Then $T = \max x_i$ is sufficient for θ.

The density fn of T is

$$f_T(t|\theta) = \frac{n}{\theta^n} t^{n-1} I_{(0,\theta)}(t) \quad \text{--- (i)}$$

[Recall :- If $X_{(1)}, \ldots, X_{(n)}$ is order stat,
then dist of $Y = X_{(n)}$ = highest order stat

is $f(y) = n[F(y)]^{n-1} \cdot f(y)$

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

$= 0$ o.w

and $F(x) = 0$, if $x \leq 0$

$= \frac{x}{\theta}$, if $0 < x < \theta$

$= 1$, if $x \geq \theta$

So, for $T = \max x_i$

$$f_T(t|\theta) = n[F(t)]^{n-1} \cdot f(t)$$

$$= n\left(\frac{t}{\theta}\right)^{n-1} \frac{1}{\theta}$$

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

Recall : $I_{(0,\theta)}(t) = 1$, if $0 < t < \theta$,
$= 0$, o.w

Thus from (1) , if $E_\theta[g(t)] = 0$
that is, if $\int_0^\theta g(t) \cdot f(t)\, dt = 0$

that is, if $\frac{n}{\theta^n} \int_0^\theta g(t)\, t^{n-1}\, dt = 0$

Then it implies $\int_0^\theta g(t) \cdot t^{n-1}\, dt = 0$ for $0 < t < \theta$

$\Rightarrow g(t) = 0 \quad \forall\ t > 0$

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

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

Invariant statistical decision problems:-

The invariate principle involves groups of transformation over the 3 spaces $(\mathcal{X}, \mathcal{a}, \Theta)$ associated with any decision problem.

Ex-1 : 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 observation $X = x$ under squared error loss f^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" and a dec 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 the observation would be $Y = \frac{X}{60}$

Define $\eta = \frac{\theta}{60}$, Then Pdf of $Y = \frac{X}{60}$ will be

$$F\left(\frac{y}{\eta}\right) = \frac{1}{\eta} e^{-y/\eta} \quad ; y > 0 \; ; \eta > 0$$

$$= 0 \qquad \text{o.w}$$

If the action say a^*, in this problem are also expressed in terms of minutes so that $a^* = \frac{a}{60}$; then the loss

fn will be of the form

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

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

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

Thus, we can observe that the formed 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 decision rule for both the formulations

Let $\delta_0^*(y)$ denote the proposed dec rule for the transformed problem. Then 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 be made no matter what is the unit of measurement

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 $c>0$ and for transformation of form $y = cx$

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

Our objective is to determine $\delta_0(x)$ so that (1) is satisfied.

The functional eq^n (1) can be solved easily by taking $c = 1/x$, so that

$$(1) \Rightarrow \delta_0(x) = x\,\delta_0(1) = kx \quad -(2)$$

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

Such rules are said to be "invariant" for a given decision problem.

A best invariant rule for this problem, then consists in choosing the constant 'k' which minimizes the risk.

Hence invariance principle completely determines the decision rule to be used

* Invariance

The invariance principle involves groups of transformation over 3 spaces involved in decision theory:

the parameter space Θ

the action space $\mathcal{A}$ and

the sample space

and assumed to be a subset of E_n.

Defn:- 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 $g(x_2) = x_1$

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

Def^n The family of distribution P_θ $\theta \in \Theta$ is said to be invariant under the group $\mathcal{G}$ if for every $g \in \mathcal{G}$ and every $\theta \in \Theta$ $\exists$ a unique $\theta' \in \Theta$ such that the dist of $g(x)$ is given by $P_{\theta'}$ whenever the dist of X is given by P_θ

The θ' uniquely determined by g & θ is denoted by $\bar{g}(\theta)$

Remark :- The condition that the family of dist^n P_θ be invariant under $\mathcal{G}$ is that for every measurable set $A \subset \mathcal{X}$

$$P_\theta\{g(x) \in A\} = P_{\bar{g}(\theta)}\{x \in A\}$$

In terms of expectation this is equivalent to say that for every integrable real-valued fn ϕ

$$E_\theta[\phi(g(x))] = E_{\bar{g}(\theta)}[\phi(x)]$$

when the dist of X is P_θ

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

Note All transformations $\bar{g} \in \bar{\mathcal{G}}$ are one to one and onto

Def^n <u>Invariance of statistical Decision Problem</u>:-

A decision problem, consisting of the game $(\Theta, \mathcal{a}, L)$ and distribution P_θ over $\mathcal{X}$ is said to be invariant under the group $\mathcal{G}$ if the family of distribution

P_θ $\theta \in \Theta$, is invariant under $\mathcal{G}$ and if the loss function is invariant under $\mathcal{G}$ in the sense that for every $g \in \mathcal{G}$ and $a \in \mathcal{a}$, $\exists$ a unique $a' \in \mathcal{a}$ such that

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

Lemma:- If a decision 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 $X \sim N(\theta, 1)$ $\Theta = \mathcal{a} = R$

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

Consider group of transformation $\mathcal{G}$ with $g_c(x) = x + c$

Then $g_c(x) = x + c \sim N(\theta + c, 1)$

Thus, ~~$g_c(x) = x +$~~

The $N(\theta, 1)$ is invariant under $\mathcal{G}$ and $\bar{g}_c(\theta) = \theta + c$

further $L(\theta, a) = L(\bar{g}_c(\theta), \tilde{g}_c(a))$ is

Satisfied for all θ, when $\tilde{g}_c(a) = a+c$

$\left(\because \bar{g}_c(\theta) = \theta + c \ \& \ \tilde{g}_c(a) = a+c \right.$

$\Rightarrow L(\bar{g}_c(\theta), \tilde{g}_c(a)) = (\bar{g}_c(\theta) - \tilde{g}_c(a))^2$

$\left. = (\theta + \cancel{c} - a - \cancel{c})^2 = (\theta - a)^2 = L(\theta, a) \right)$

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 Suppose $\mathcal{a} = [0,1]$
and $L(\theta, a) = w(\theta - a)$, some even f^n of $(\theta - a)$

Let $\mathcal{g}$ be the group of transformation considering e and $g(x) = n - x$
We know, if $X \sim B(n, \theta)$ then $Y = n - x \sim B(n, 1-\theta)$
Thus, the dist of $g(x) = n - x$ is $B(n, 1-\theta)$
so that the $B(n, \theta)$ dist is invariant under g and $\bar{g}_c(\theta) = 1-\theta$ with $\tilde{g}(a) = 1-a$

$$L(\bar{g}_c(\theta), \tilde{g}_c(a)) = w(\bar{g}_c(\theta) - \tilde{g}_c(a))$$

$$= w(1-\theta, -1+a) = w(a-\theta)$$

$\therefore$ the loss is invariant & hence the decision problem is invariant

* Invariant Decision Rules

Consider a decision problem $(\Theta, \mathcal{A}, L)$ with an observable random quantity X whose dist P_θ depends on θ and assume that the dec problem is invariant under a group $\mathcal{G}$ of transformation on the sample space $\mathcal{X}$.

The decision problem in which the statistician observes a random quantity $Y = g(x)$ whose dist is P_ϕ, $\phi \in \Theta$ [where $\phi = \bar{g}(\theta)$], and must choose a point $b \in \mathcal{A}$ with loss fn $L(\phi, b)$ is exactly the same problem as stated above since for any $a \in \mathcal{A}$ that he may choose in the original problem the choice $b = \tilde{g}(a)$ yields the same loss in the new problem

$$(L(\theta, a) = L(\bar{g}(\theta), \tilde{g}(a)) = L(\phi, b))$$

Thus the decision to take action a_0 if $X = x_0$ in the first problem is equivalent to the decision to ~~the~~ take action $\tilde{g}(a_0)$ if $Y = g(x_0)$ in the second

Because the two problems are identical hoe we should be willing to take action $\tilde{g}(a_0)$ if $X = g(x_0)$ in the original problem

Thus there should be $d(g(x_0)) = \tilde{g}(d(x_0))$ for all $g \in \mathcal{G}$

Rules that satisfy this equality are called invariant non-randomized decision rule

Def^n Given an invariant decision problem a non-randomized dec rule $d \in D$ is said to be invariant under $\mathcal{G}$ if for all $x \in \mathcal{X}$ and all $g \in \mathcal{G}$

$$d(g(x)) = \tilde{g}(d(x))$$

A randomized dec rule $\delta \in D^*$ is said to be invariant if δ, as a prob dist over D, gives all its mass to the subset of invariant non-randomized decision rule

A behavioral dec rule $\delta \in D$ is said to be invariant if for all $x \in \mathcal{X}$ and all $g \in \mathcal{G}$
$\delta(g(x)) = \tilde{g}(\delta(x))$
where by $\tilde{g}(\delta)$ we mean the dist of $\tilde{g}(z)$ when z has dist δ.

* Two points $\theta_1, \theta_2 \in \Theta$ are said to be equivalent if there exists a $\bar{g} \in \bar{\mathcal{G}}$ such that $\theta_1 = \bar{g}(\theta_2)$. This is an equivalence relation, which breaks Θ into equivalence classes

An important property of an invariant dec rule is that the risk f^n of an invariant dec rule δ is constant that is,

$$R(\theta, \delta) = R(\bar{g}(\theta), \delta) \quad \forall \theta \in \Theta$$
& all $\bar{g} \in \bar{\mathcal{G}}$

For a non randomized dec rule d there will be

$$R(\theta, d) = R(\bar{g}(\theta), d)$$

lemma:- If a dec problem is invariant under $\mathcal{G}$, then for any behavioral dec rule $\delta \in \mathcal{D}$ and $g \in \mathcal{G}$

$$\hat{R}(\theta, \delta^g) = \hat{R}(\bar{g}(\theta), \delta)$$

Thm If a decision problem is invariant under $\mathcal{G}$, then for any invariant dec rule $\delta \in \mathcal{D}$, $\hat{R}(\theta, \delta) = \hat{R}(\bar{g}(\theta), \delta)$ for all $\theta \in \Theta$ and all $\bar{g} \in \bar{\mathcal{G}}$

* Admissible & Minimax Invariant Rules:-

Thm 1 Suppose that a given dec problem is invariant under a finite group $\mathcal{G}$. Then if there exists a minimax rule $\exists$ a minimax rule which is (behavioural) invariant. If a rule is minimax within the class of behavioural invariant rules it is minimax.

Thm 2 Suppose that a given decision problem is invariant under a finite group $\mathcal{G}$. If an invariant rule $\delta_0 \in \mathcal{D}$ is admissible within the class of all invariant rules, it is admissible.

* Invariant Prior dist^n

A prior dist. τ on Θ is said to be invariant under $\bar{\mathcal{G}}$ if for all $\bar{g} \in \bar{\mathcal{G}}$, $\bar{g}\tau = \tau$ where by $\bar{g}\tau$ we mean the dist of $\bar{g}T$ when T has dist τ.

Thm 3 Suppose that a given dec problem is invariant under a finite group $\mathcal{G}$

(a) If δ_0 is Bayes rule w.r.to prior dist^n τ and δ_0 is invariant, $\exists$ an invariant prior dist τ_0 w.r.to which δ_0 is Bayes

(b) If δ is Bayes rule w.r.to τ_0 and τ_0 is invariant, $\exists$ an invariant rule δ_0 which is Bayes w.r.to τ_0

(c) If $\exists$ a least favourable prior dist τ $\exists$ an invariant least favourable dist τ_0.