chapter 2 (very abstract)

⇒ Decision theory (Wald)

Formulation :

1) parameter space (e.g, mean, variance)

$\theta \in \Theta$ (all the possible values that θ can take)

2) A sample space $\mathcal{X}$, the space that the data X lie

example : ∵ $X = (x_1, x_2, \cdots, x_n)$ $x_i \in R$

∴ $X \in R^n$

For sample : two dimension space X_1, X_2

3) A family of probability distribution

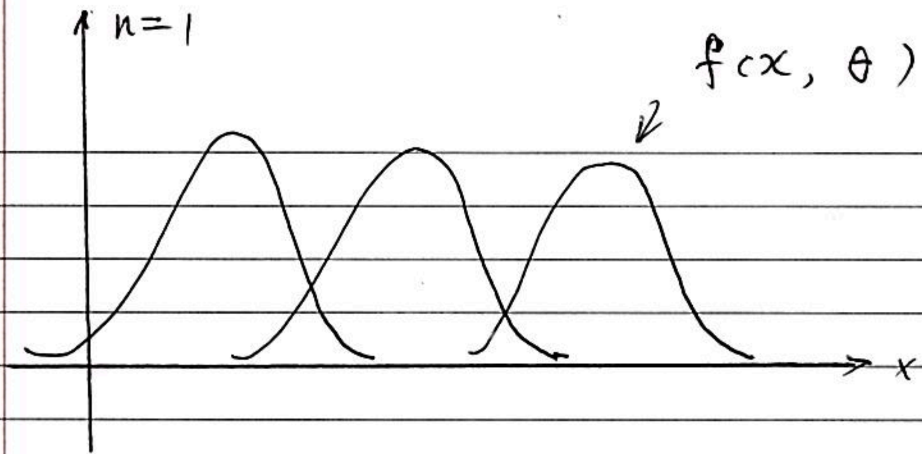
$\{ P_\theta(x) \mid \theta \in \Theta \}$ how likely we see the data given a θ

if X is continuous $P_\theta(x) = f(x, \theta)$ ⇐ density

if X is discrete $P_\theta(x) = f(x, \theta)$ (same notation)

∴ $P_\theta(x) = f(x, \theta)$ → probability mass function if x is discrete.

density if x is continuous



4) An action space

$\mathcal{A}$: all the actions / decisions available to experimentors

example :

(a) hypothesis testing

H_0 Vs H_1

$$\mathcal{A} : a = \begin{cases} 0 & \text{accept } H_0 \\ 1 & \text{reject } H_0 \end{cases}$$

two possible actions, represented by two numbers

(b) In point estimation

$A = \Theta$

an action is an estimate of θ

(c) In predictive inference

we want to predict y_{n+1}

$\mathcal{A} = \mathbb{R}$

(d) In interval estimation, an action is an interval

$A = \{ (l, u) \mid l \in R, u \in R, l \leq u \}$

(5) Loss function

$L : \Theta \times A \rightarrow \mathbb{R}$

$L(\theta, a)$ specifies the loss we may incur if the true parameter is θ, and we take action a.

Generally, $L(\theta, a)$ can be positive and negative

Example:

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

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

3) In hypothesis testing

$$L(\theta, a) = \begin{cases} 1 & \theta \in H_0 \quad a = 1 \\ 1 & \theta \in H_1 \quad a = 0 \\ 0 & \\ 0 & \end{cases} \Big\} \text{ otherwise}$$

$a = 0$ accept H_0

$a = 1$ reject H_0

Note that : loss function means if do something wrong what we will loss.

$\Rightarrow$ decision rule :

An element $d : X \to A$

$d(x)$: the action we take when we gather X

Example :

1) In point estimation

$$d(x) = \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$d(x) = s^2$$

$$d(x) = \text{median of } (x_1, \cdots, x_n)$$

2) In interval estimation

$$d(x) = (l(x), u(x))$$

3) In hypothesis testing

$$d(x) = \begin{cases} 1 & \text{if } \left| \frac{\bar{x} - 0}{s/\sqrt{n}} \right| > t_{n-1, \frac{\alpha}{2}} \\ 0 & \text{otherwise} \end{cases}$$

We want to identify $d(x)$

Risk function :

$$R(\theta, d) = E_\theta[L(\theta, \underbrace{d(x)}_{\text{Average for different possible sample.}})]$$

d : decision by data

therefore ;

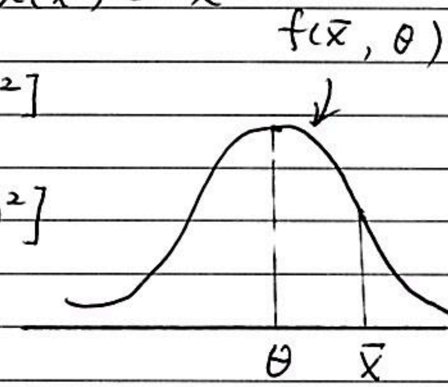
$$R(\theta, d) = \begin{cases} \int_X L(\theta, d(x)) f(x, \theta)\, dx & \Rightarrow \text{continuous} \\ \sum_{x \in X} L(x, d(x)) f(x, \theta) & \Rightarrow \text{discrete} \end{cases}$$

Example :

1) $L(\theta, a) = (\theta - a)^2 \qquad d(x) = \bar{x}$

$$R(\theta, d) = E_\theta[(x - d(\theta))^2]$$

$$= E_\theta[(x - \bar{x})^2]$$



($\Rightarrow$ Mean squre Error)

$$= \int_R (\theta - \bar{x})^2 f(\bar{x}, \theta)\, d\bar{x}$$

2) In hypothesis testing

$$L(\theta, a) = \begin{cases} 1 & \theta \in H_0 \quad a = 1 \\ 1. & \theta \in H_1 \quad a = 0 \\ 0 & o/w \end{cases}$$

$$d(x) = \begin{cases} 1. & \text{if we reject } H_0 \\ 0 & \text{if accept. } H_0 \end{cases}$$

Note that: We have not specified how to reject / accept.

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

false positive Rate

$$= \begin{cases} 1 * P_x(d(x)=1) & \text{if } \theta \in H_0 \\ 1 * P_x(d(x)=0), & \text{if } \theta \in H_1 \end{cases}$$

false negative rate.

<< January 09. 2014 >> 11:30 AM - 12:20 PM

Risk function : $R(\theta, d)$

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

average X a way.

example:

1) $d(x) = \bar{x}$

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

$$R(\theta, d) = R(\theta, \bar{x}) = E_\theta[(\theta - \bar{x})^2]$$

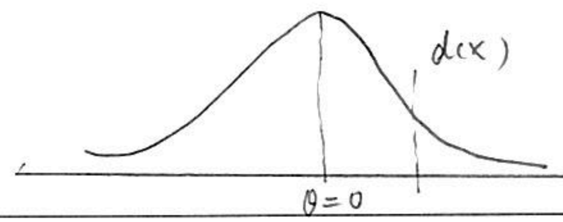
2) In Hypothesis testing

$$a = \begin{cases} 0 & \text{accept } H_0 \\ 1 & \text{reject } H_0 \end{cases}$$

	Accept H_0	reject H_0
H_0	OK	I
H_1	II	OK

$$d(x) = \begin{cases} 0 & \text{if we accept } H_0 \\ 1 & \text{if we reject } H_0 \end{cases}$$

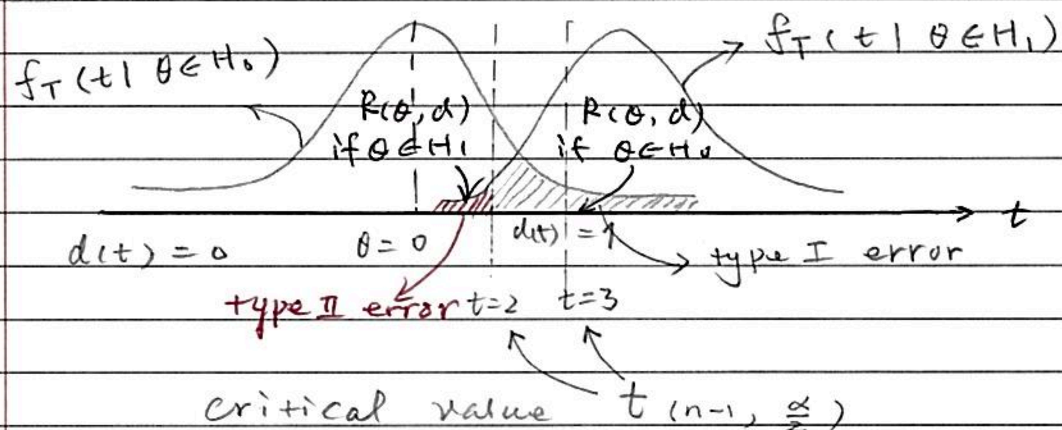
$$L(\theta, a) = \begin{cases} 0 & \text{if } \theta \in H_0 \quad a = 0 \\ 0 & \text{if } \theta \in H_1, \quad a = 1 \\ 1 & \text{if } \theta \in H_0 \quad a = 1 \\ & \quad \text{this is a type I error} \\ 1 & \text{if } \theta \in H_1 \quad a = 0 \\ & \quad \text{this is a type II error} \end{cases}$$



$$R(\theta, d) = E_\theta[L(\theta, d(x))], \quad \theta \in H_0$$

$$= \begin{cases} 0 \times P_r(d(x)=0) + 1 \times P_r(d(x)=1), & \text{for } \theta \in H_0 \\ 0 \times P_r(d(x)=1) + 1 \times P_r(d(x)=0), & \text{for } \theta \in H_1 \end{cases}$$

example: $H_0: \mu \le 0$ Vs $H_1: \mu > 0$



Remark: typically we don't have a single d, such that

$R(\theta, d) \le R(\theta, d')$ for all $\theta \in \Theta$

example 1: $d(t)$: $d=1$ if $t>2$

$d=0$ o/w

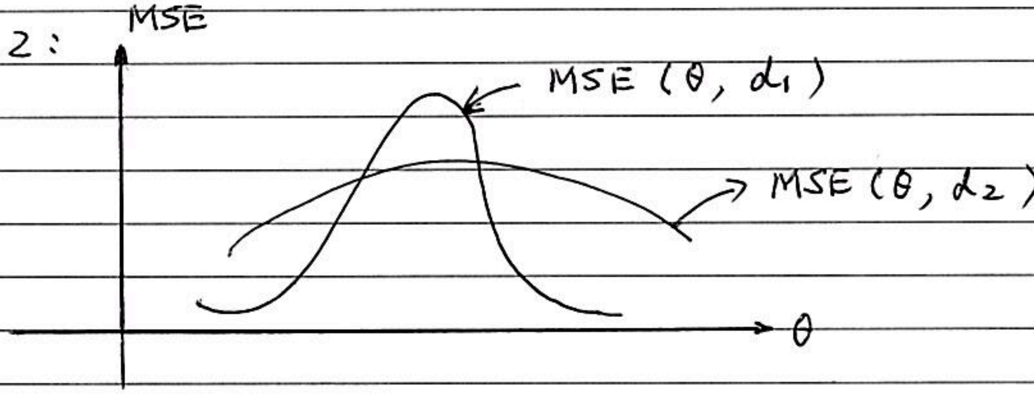
$d'(t)$: $d'=1$, if $t>3$

$d'=0$, o/w

$R(\theta, d) > R(\theta, d')$ if $\theta \in H_0$

But $R(\theta, d) < R(\theta, d')$ if $\theta \in H_1$

example 2:



$\Rightarrow$ Criteria for choosing decision rules

Admissibility:

def: we say d strictly dominate d'

if $R(\theta, d) \leq R(\theta, d')$ for all $\theta \in \Theta$

and $\exists\ \theta_0$ such that $R(\theta_0, d) < R(\theta_0, d')$

Admissible:

if a decision d isn't dominated by any d', then d is admissible; if a decision d is dominated by another decision d', then d is inadmissible.

⇒ Remarks:

1) d is very poor if d is inadmissible;

2) Many daily used statistics are inadmissibles.

examples:

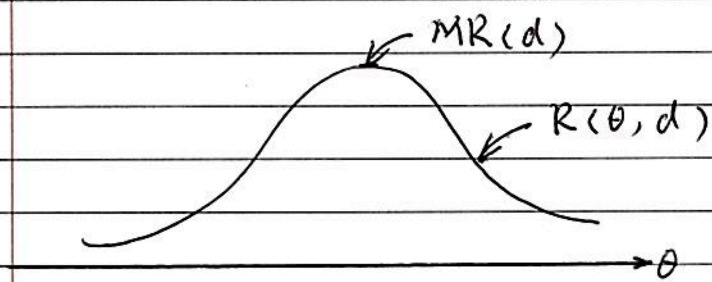
1) $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)$

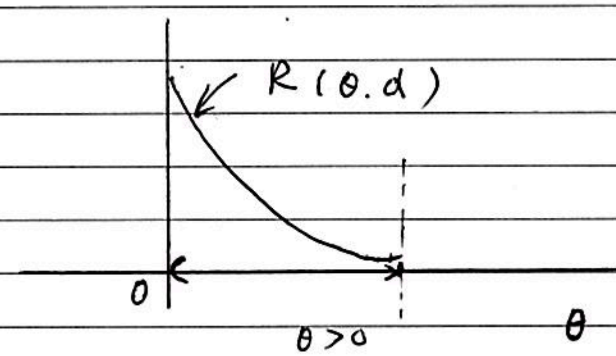
2) For $d \geq 3$ $\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_d)$ is inadmissible

斯特恩的悖论 Stern's paradox

⇒ minimax decision Rules: supremum 上确界 最小上界

Maximum Risk: $MR(d) = \sup_{\theta \in \Theta} R(\theta, d)$



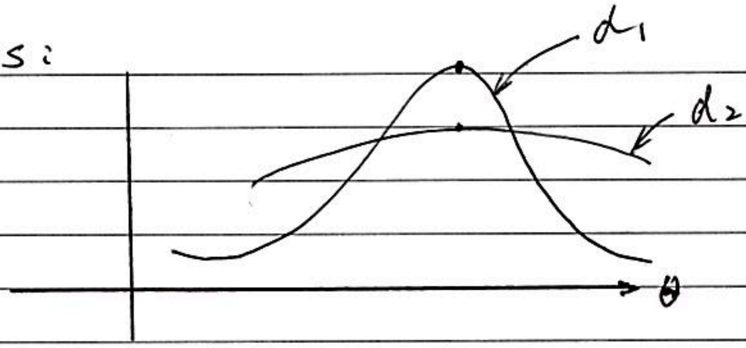


Def for minimax:

A decision d is minimax if $MR(d) \leq MR(d')$ for all $d' \in D$

⇒Remarks:

1)

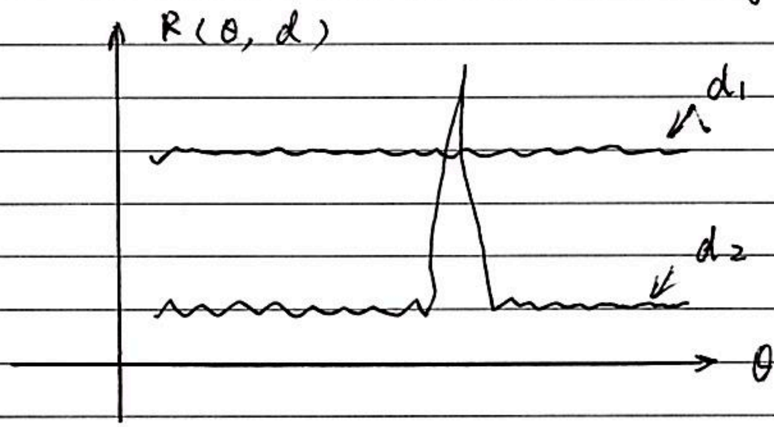


$D = \{d_1, d_2\}$

d_2 is minimax

minimizing the Risk in the worst case

2). minimax rules may not good



$MR(d_1) \leq MR(d_2)$

But $R(\theta\ d_1) > R(\theta, d_2)$ for most θ

<< STAT 846, January 12, 2014 >>

⇒ unbiaseness:

Defination: A decision rule d is unbiased

if $E_\theta\{L(\theta', d(x))\} \geq E_\theta\{\underbrace{L(\theta, d(x))}_{\to R(\theta, d)}\}$ —— (1)

for all $\theta, \theta' \in \Theta$

This is a generalization of unbiaseness for point estimate

$E_\theta(d(x)) = \theta$ ⟵ unbiased —— (2)

~~practice~~

Show that: (1) = (2) under squre loss

proof: suppose the loss function is the squred error loss $L(\theta, d) = (\theta - d)^2$

Fix θ and Let $E_\theta(d(x)) = \phi$, then, for d to be an unbiased decision rule, we require that, for all θ',

$$0 \leq E_\theta\{L(\theta', d(x))\} - E_\theta\{L(\theta, d(x))\}$$

$$= E_\theta\{(\theta' - d(x))^2\} - E_\theta\{(\theta - d(x))^2\}$$

$$= (\theta')^2 - 2\theta'\phi + E_\theta(d^2(x)) - \theta^2 + 2\theta\phi - E_\theta(d^2(x))$$

$$= (\theta' - \phi)^2 - (\theta - \phi)^2$$

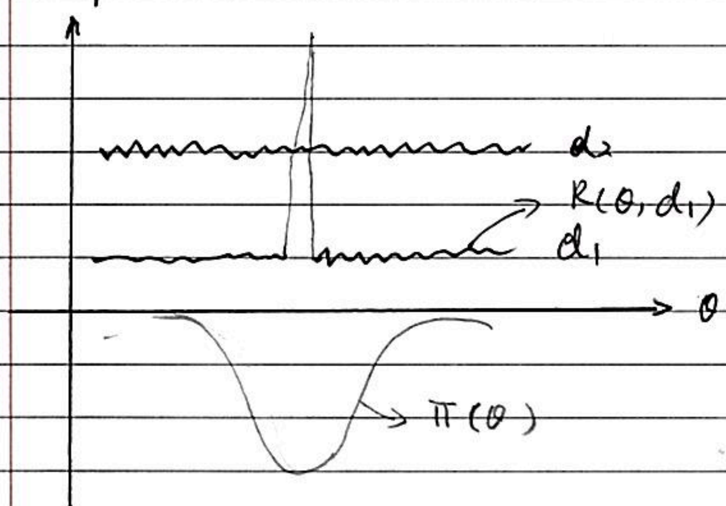
if $\phi = \theta$, then this statement is obviously true

If $\phi \neq \theta$, then set $\theta' = \phi$ to obtain a contradiction.

therefore, $E_\theta(d(x)) = \phi = \theta$ for all θ

$\Rightarrow$ $d(x)$ is said to be unbiased.

$\Rightarrow$ Bayes decision Rules:



We specify a prior $\pi(\theta)$ for θ

Definition:

1) A Bayes Risk (not a function of θ) of a decision Rule is $r(\pi, d) = \int_\theta R(\theta, d)\pi(\theta)\,d\theta$

2) A decision Rule d is said to be a bayes Rule, with respect to a given prior $\pi(\cdot)$, if it minimises the Bayes risk, so that

$$r(\pi, d) = \inf_{d' \in D} r(\pi, d') = m_\pi$$

[less formaly $r(\pi, d) \leq r(\pi, d')$, for all $d' \in D$]

3) ε - Bayes

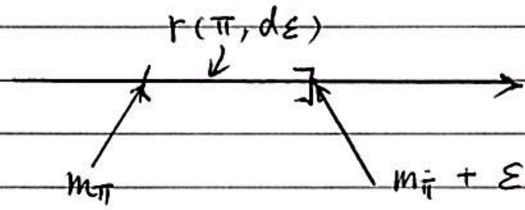
Let $m_\pi = \inf_{d' \in D} r(\pi, d')$, for any $\varepsilon > 0$

We can find a decision rule d_ε for which

$r(\pi, d_\varepsilon) < m_\pi + \varepsilon$

and in this case d_ε is said to be ε- Bayes

with respect to the prior distribution $\pi(\cdot)$.



=> Randomized decision Rules

$d(x) : \underset{\substack{\text{sample space} \\ \text{(data)}}}{X} \rightarrow \underset{\text{action space}}{A}$

Given a collection of decision rules

$d_1, d_2, \cdots d_I$, a set of probability weights

$p_1, p_2, \cdots, p_I$ $\left(\sum_{i=1}^{I} p_i = 1 \right)$

Defination the decision Rule $d^* = \sum_{i=1}^{I} p_i d_i$ to be

The rule 'selected d_i with probability p_i', then

d^* is a randomised decision rule.

example :

$$d_1 = \bar{x} \qquad d_2 = \text{median}(\bar{x})$$

$$P_1 = 1/4 \qquad P_2 = 3/4$$

$$X \text{ (denote toss the two coins)} \begin{cases} \xrightarrow{\frac{1}{4}\ (HH)} d_1 \\ \xrightarrow[(o/w)]{3/4} d_2 \end{cases}$$

$\Rightarrow$ Risk function of d^*

$$R(\theta, d^*) = \sum_{i=1}^{I} P_i R(\theta, d_i)$$

$\Rightarrow$ Finite decision problems

(for illustrating concepts in decision theory)

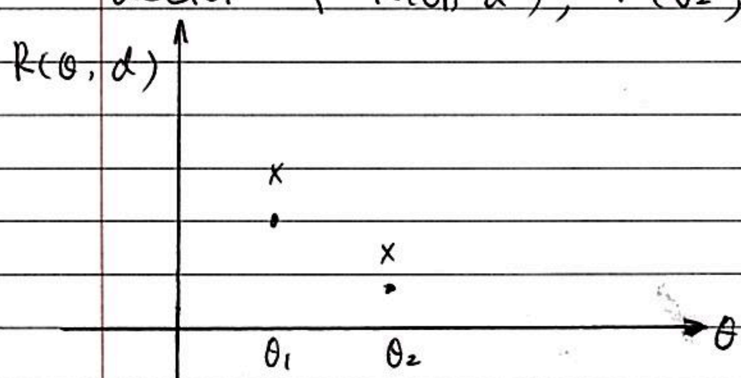
parameter space is a finite set :

$$\Theta = \{\theta_1, \theta_2, \cdots, \theta_t\}$$

$\Rightarrow$ Risk set

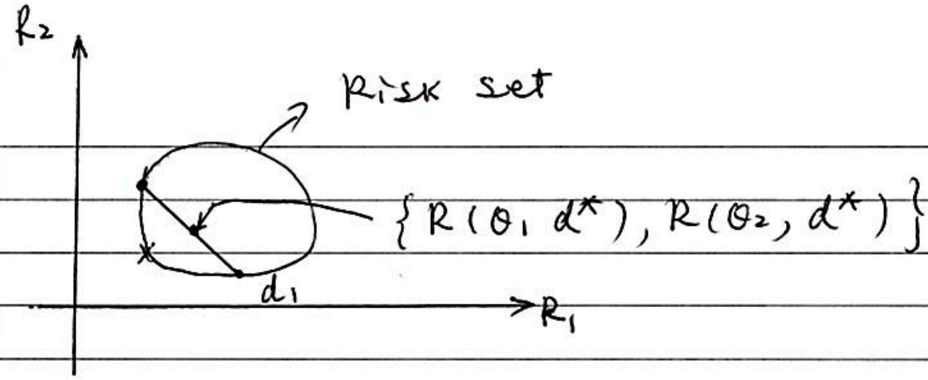
a subst in $\mathbb{R}^+$ containning all possible

Vector $(R(\theta_1, d), R(\theta_2, d), \cdots, R(\theta_t, d))$ for all $d \in D$



a vector $\{R(\theta_1, d), R(\theta_2, d)\}$

$d_1 \Rightarrow \bullet \qquad d_2 \Rightarrow \times$

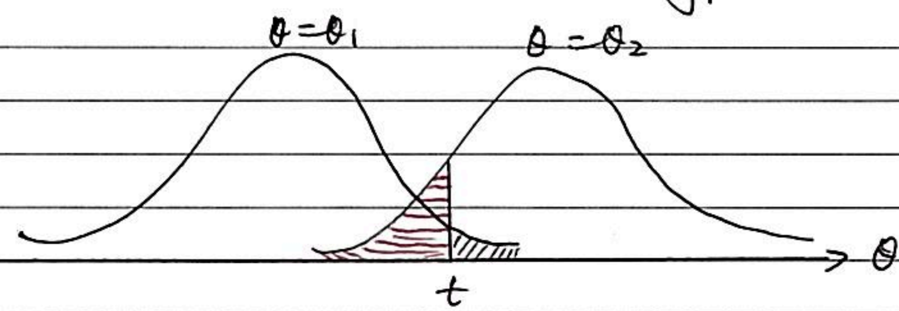


while $d^* = P_1 d_1 + P_3 d_3$

because $R(\theta, d^*) = P_1 R(\theta, d_1) + P_3 R(\theta, d_3)$

example: $\theta = (\theta_1, \theta_2)$, θ_1 is null hypothesis

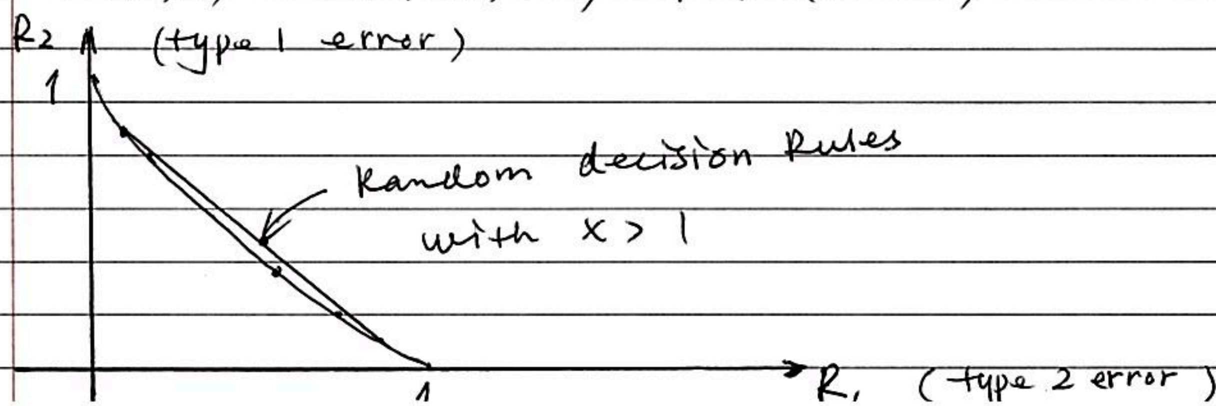
θ_2 is alternative hypothesis



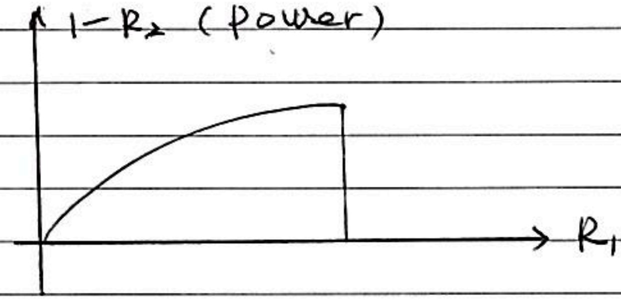
$$R(\theta, d) = \begin{cases} Pr_{\theta_1}(d(x) = 1) & \text{if } \theta = \theta_1 \\ Pr_{\theta_2}(d(x) = 0) & \text{if } \theta = \theta_2 \end{cases} \quad ; \quad d(x) = \begin{cases} 1 & x > t \\ 0 & x < t \end{cases}$$

$$R_1 = R(\theta_1, d) = Pr_{\theta_1}(d(x) = 1) = Pr_{\theta_1}(x > t)$$

$$R_2 = R(\theta_2, d) = Pr_{\theta_2}(d(x) = 0) = Pr_{\theta_2}(x < t)$$



$\Rightarrow$ ROC curve is an example of risk set



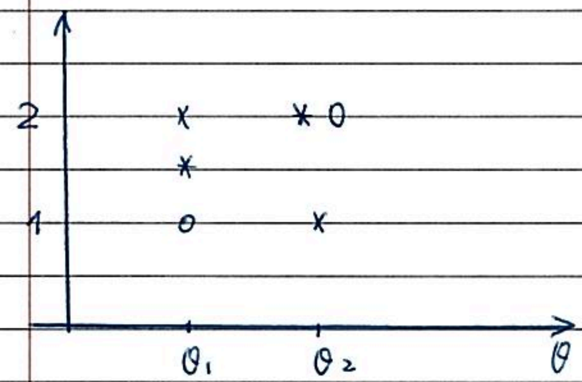
<< January 14. 2015 >>

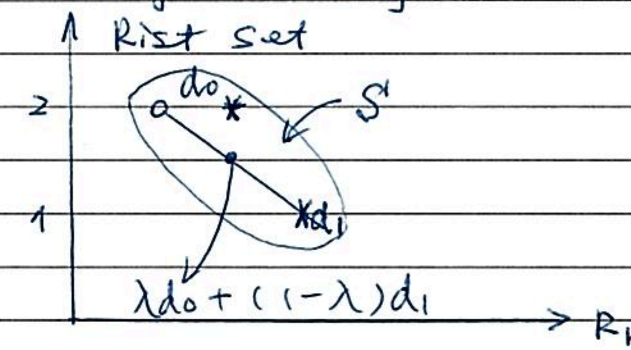
Risk set $\theta = \{\theta_1, \theta_2, \cdots, \theta_t\}$

$$R_i = R(\theta_i, d) \quad \text{for } i = 1, \cdots, t$$

Risk set : $R = \{R(\theta_1, d), R(\theta_2, d), \cdots, R(\theta_t, d) \mid \text{for any } d \in D\}$

R(θ



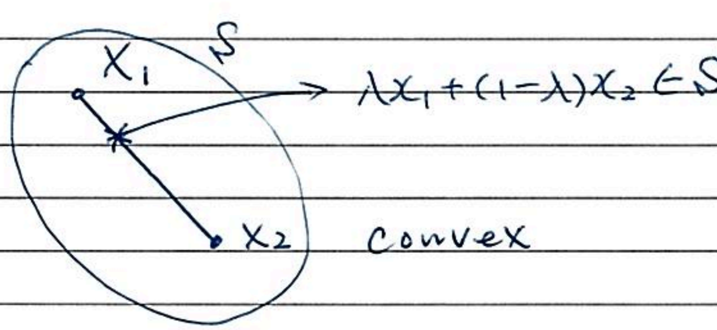


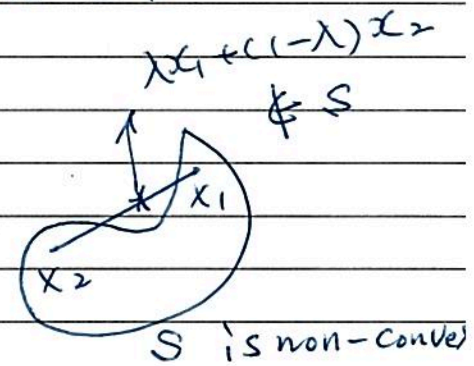
o: do   *: d1

$\Rightarrow$ Lemma 2.1 : A Risk set S is convex.

Def: A S is convex, if for any $x_1, x_2 \in S$

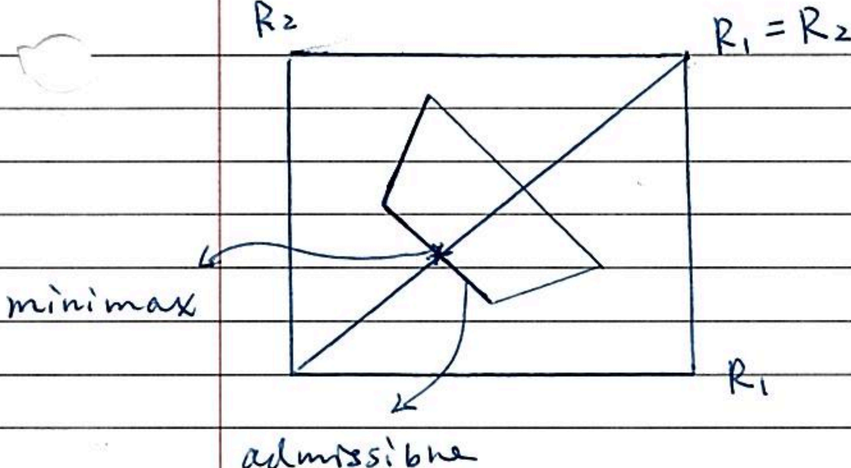
$$\lambda \in [0, 1], \quad \lambda x_1 + (1-\lambda) x_2 \in S$$





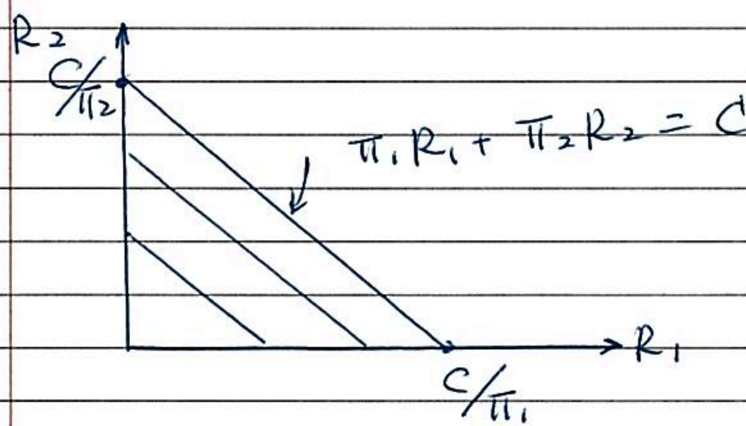
Proof of Lemma 2.1

Suppose that $X_1 = \{R(\theta_1, d_1) \cdots R(\theta_t, d_1)\}$

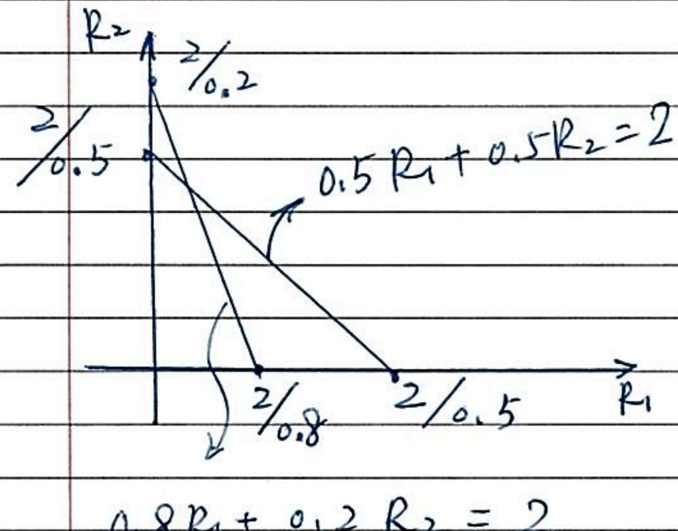


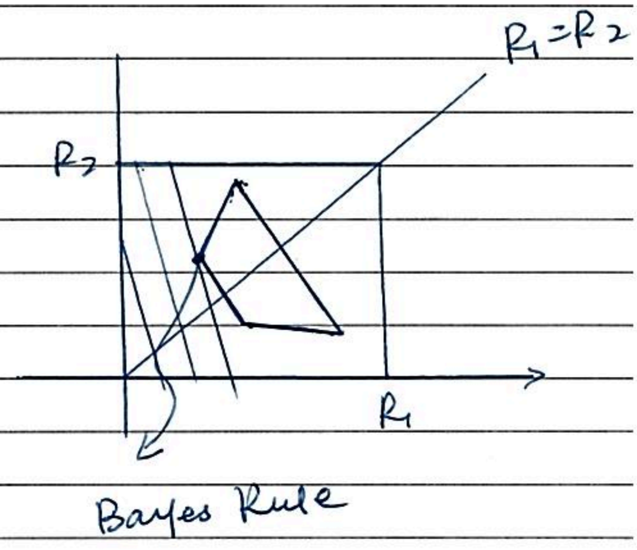
$\Rightarrow$ Bayes Risk : Given $\pi = (\pi_1, \pi_2)$

$B(R_1, R_2) = \pi_1 R_1 + \pi_2 R_2$ is Bayes risk



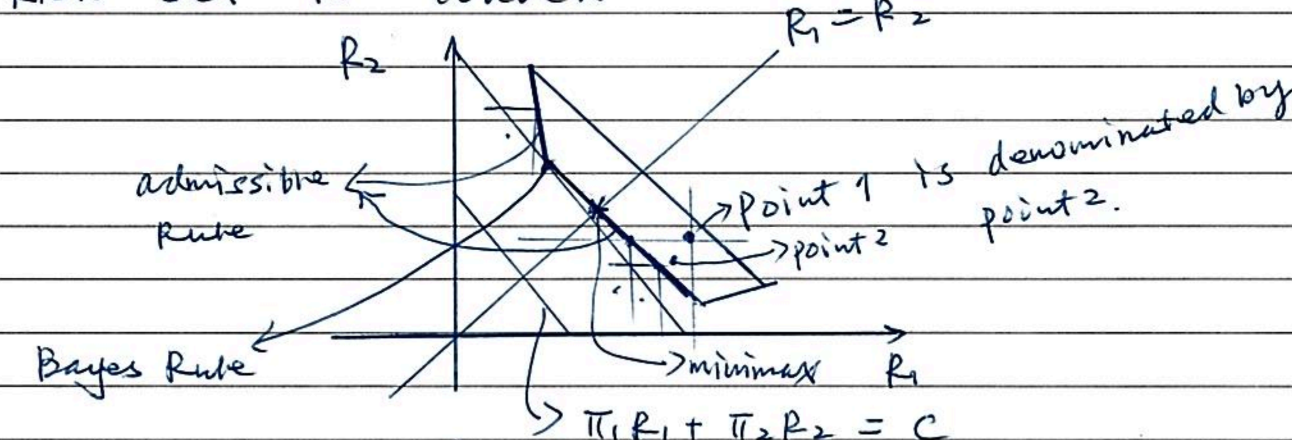
examples:

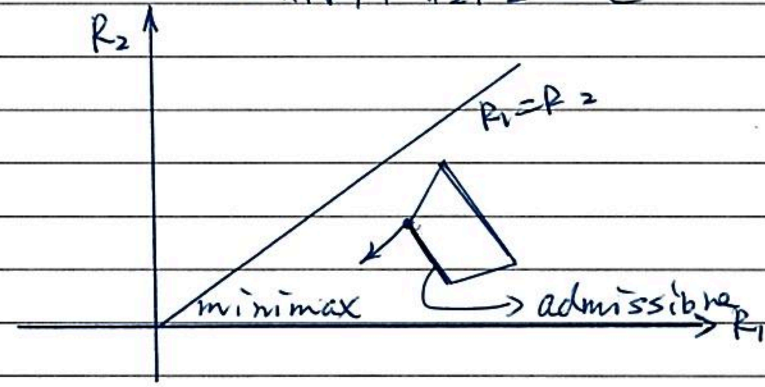




<< January 16, 2014 >> STAT : 846

⇒ Risk set is convex





Example: box 1   box 2

A real necklace in box; an imitation necklace in another box.

⇒ parameter space Θ:

$$\theta = \begin{cases} 1 & \text{if real necklace in box 1} \\ 2 & \text{if real necklace in box 2} \end{cases}$$

$$\Theta = \{1, 2\}$$

$\Rightarrow$ action space $\mathcal{A}$:

$$a = \begin{cases} 1 & \text{if choose box 1} \\ 2 & \text{if choose box 2.} \end{cases}$$

$$\mathcal{A} = \{1, 2\}$$

$\Rightarrow$ Loss function $L(\theta, a)$:

$$L(\theta, a) = \begin{cases} 0 & \text{if } \theta = 1 \quad a = 1 \\ 0 & \text{if } \theta = 2 \quad a = 2 \\ 1 & \text{if } \theta = 1 \quad a = 2 \\ 1 & \text{if } \theta = 2 \quad a = 1 \end{cases}$$

$\Rightarrow$ sample space (data X): example see textbook page 15

data X: the judgement of Great Aunt

$$X = \begin{cases} 1 & \text{if Great Aunt choose box 1} \\ 2 & \text{if Great Aunt choose box 2} \end{cases}$$

$\theta_1 = 1$: $P_{\theta=1}(X=1) = 1$, $P_{\theta=1}(X=2) = 0$

$$\theta_2 = 2 : \begin{cases} P_\theta(X=1) = 1/2 \\ P_\theta(X=2) = 1/2 \end{cases}$$

$\Rightarrow$ Decision space:

$d_1(x)$: choose box 1 ignnoring Great Aunt's judgement

$d_2(x)$: choose box 2 ignoring Great Aunt's judgement.

$d_3(x)=x$: follow Great Aunt's judgement

$d_4(x)$: follow the reverse of Great Aunt's judgement

$$d_4(x) = 3 - x = \begin{cases} 2 & \text{if } x=1 \\ 1 & \text{if } x=2 \end{cases}$$

=> Risk functions :

x θ	$d_1(x)$	$d_2(x)$	$d_3(x)$	$d_4(x)$
1	0	1	0	1
2	1	0	½	½

The Risk function of the decision Rule is

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

$$R(1, d_3(x)) = L(1, d_3(x=1)) P_\theta(x=1) + L(1, d_3(x=2)) \times P_\theta(x=2)$$

$$= L(1, 1) \times 1 + L(1, 2) \times 0$$

$$= L(1, 1)$$

$$= 0$$

$$R(2, d_3(x)) = L(2 \; d_3(x=1)) \times \tfrac{1}{2} + L(2, d_3(x=2)) \times \tfrac{1}{2}$$

$$= L(2, 1) \times \tfrac{1}{2} + L(2, 2) \times \tfrac{1}{2}$$

$$= 1 \times \tfrac{1}{2} + 0 \times \tfrac{1}{2}$$

$$= \tfrac{1}{2}$$

$$R(1, d_4(x)) = L(1, d_4(x=1)) \times 1 + L(1, d_4(x=2)) \times 0$$

$$= L(1, 2) \times 1 + L(1, 1) \times 0$$

$$= 1 \times 1 = 1$$

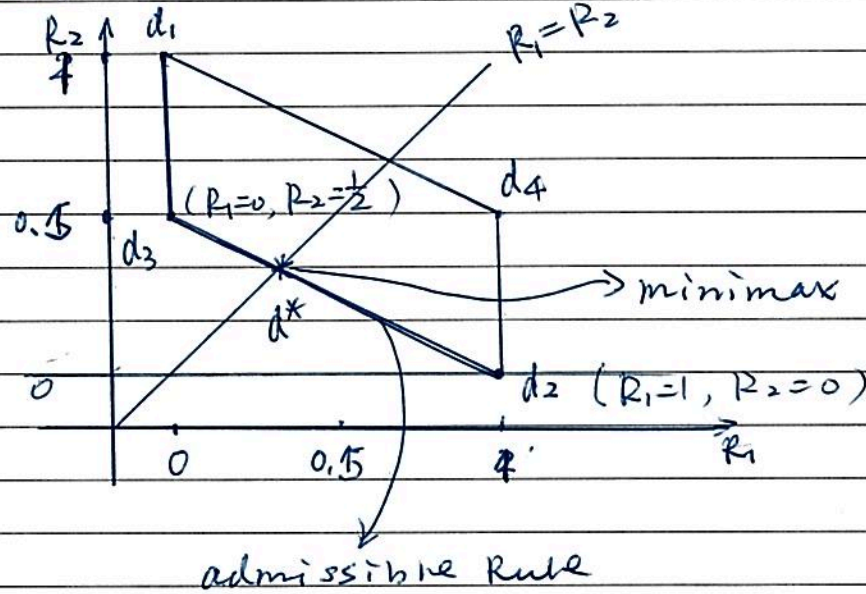
$$R(2, d_4(x)) = L(2, d_4(x=1)) \times \tfrac{1}{2} + L(2, d_4(x=2)) \times \tfrac{1}{2}$$

$$= L(2, 2) \times \tfrac{1}{2} + L(2, d_1) \times \tfrac{1}{2}$$

$$= 0 \times 1/2 + 1 \times 1/2$$

$$= 1/2$$

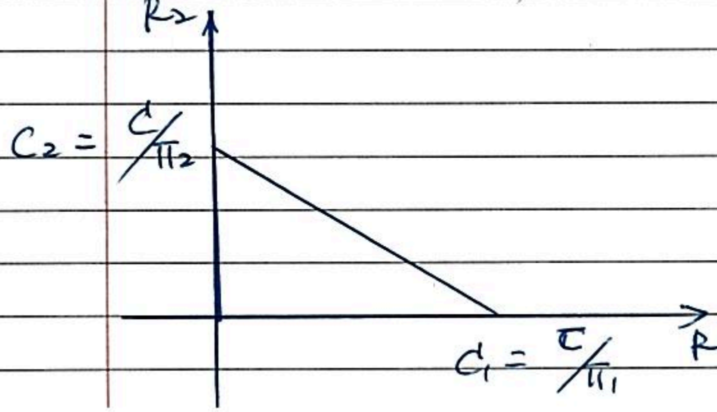
$\Rightarrow$ Risk set :



$$\text{minimax}: \begin{cases} R_1 = R_2 \\ \dfrac{R_2 - \frac{1}{2}}{R_1 - 0} = \dfrac{0 - \frac{1}{2}}{1 - 0} = -\dfrac{1}{2} \end{cases}$$

$$\Rightarrow R_2 - \tfrac{1}{2} = -\tfrac{1}{2} R_1$$

$$\Rightarrow 1/2\, R_1 + R_2 = 1/2$$



$$\pi_1/\pi_2 = \frac{C/\pi_2}{C/\pi_1}$$

$$C_2 R_1 + C_1 R_2 = C$$

$$\begin{cases} R_1 = R_2 \\ \frac{1}{2}R_1 + R_2 = \frac{1}{2} \end{cases} \iff \begin{cases} R_1 = \dfrac{\frac{1}{2}}{1+\frac{1}{2}} = \frac{1}{3} \\ R_2 = \frac{1}{3} \end{cases}$$

$\because$ $R_1 = R(\theta=1, d^*) = R(\theta=2, d^*) = R_2$

$$d^* = \lambda d_3 + (1-\lambda) d_2$$

$$\Rightarrow \lambda = 2/3$$

$$\Rightarrow d^* = \frac{2}{3} d_3 + \frac{1}{3} d_2$$

$\Rightarrow$ Admissible rule :

look at risk set d_4 is inadmissible

$\Rightarrow$ Bayes Rule :

suppose that jewel cleaner will have placed the true necklace in the Box 1 with probability φ.

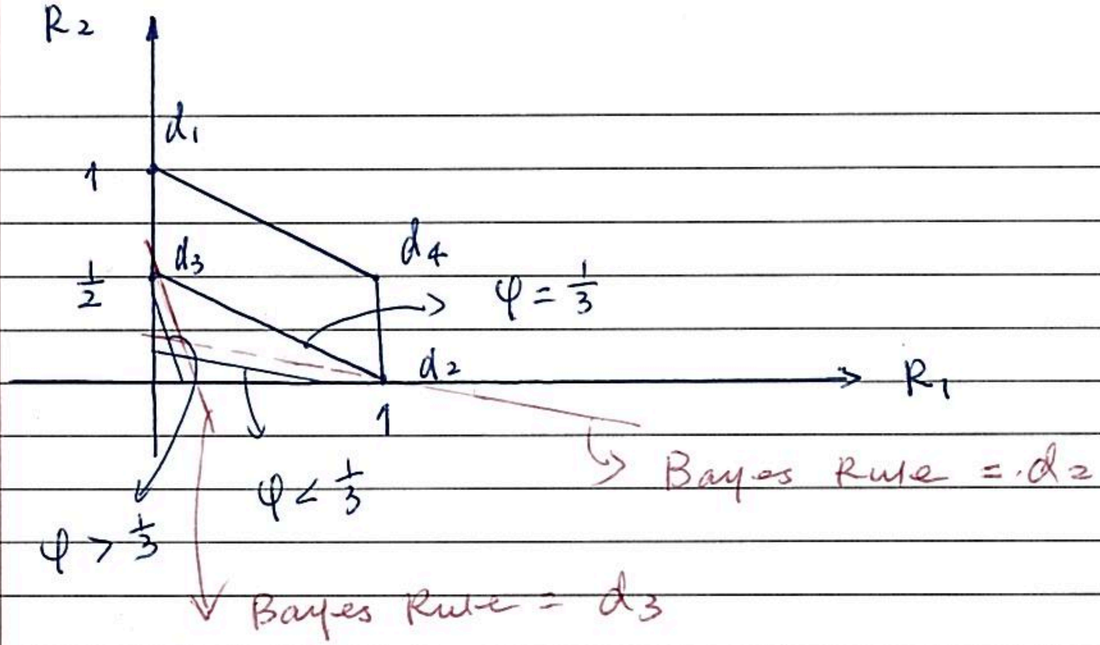
$\pi(\theta=1) = \varphi$. (in our case $\varphi = 1/3$

$\pi(\theta=2) = 1-\varphi$ ($1-\varphi = 2/3$)

$\Rightarrow$ Bayes Risk :

The Bayes risk of a rule d is then

$$r(\pi, d) = \varphi\,(R\,\theta=1, d) + (1-\varphi)\,R(\theta=2, d)$$



(1) $\varphi < \frac{1}{3}$, Bayes Rule $= d_2$

(2) $\varphi = 1/3$, Bayes Rule $= \lambda d_3 + (1-\lambda) d_2$

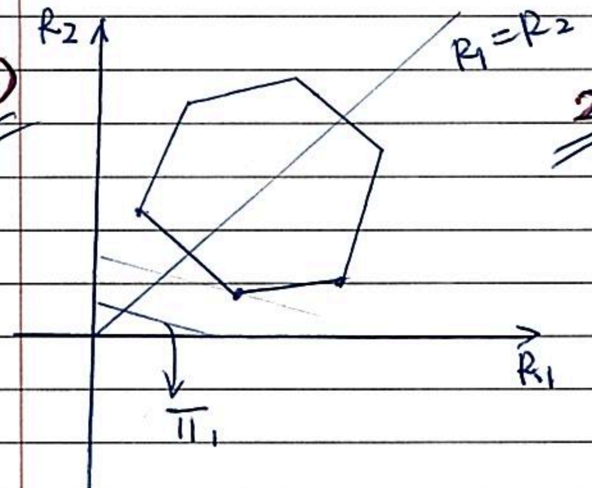
(3) $\varphi > 1/3$, Bayes Rule $= d_3$

<< January 19, 2014 >> STAT 846

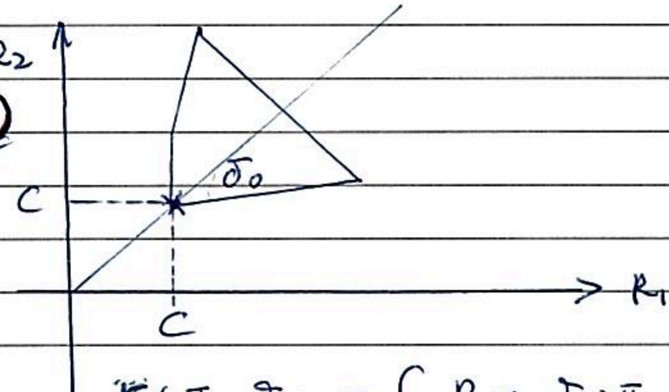
=> In chapter 3, we will see it is easy to find a Bayes Rule with a prior π, How to find minimax from Bayes Rules.

=> Some Toy example

1)

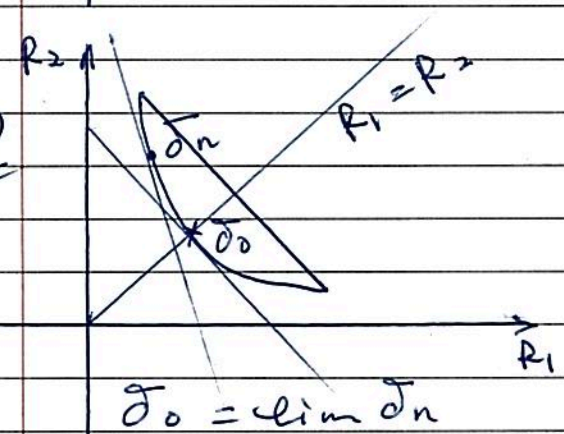


2)



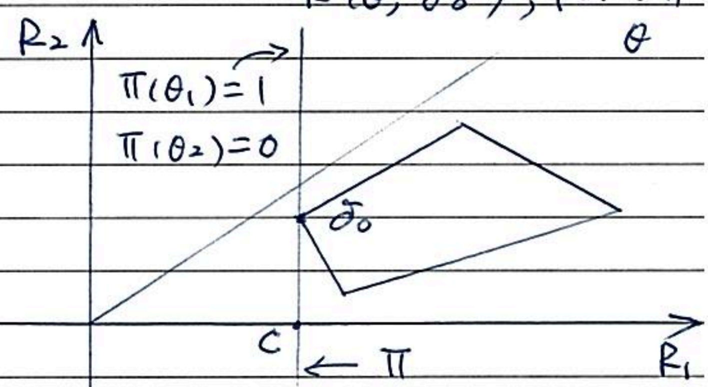
$$r(\pi, \delta_0) = \int_\theta R(\theta, \delta_0)\pi(\theta)\,d\theta = R(\theta, \delta_0), \text{ for all } \theta$$

3)



$$\delta_0 = \lim_{n\to+\infty} \delta_n$$

4)



$$\begin{aligned} r(\pi, \delta_0) &= \pi(\theta_1)\cdot R(\theta_1, \delta_0) + \pi(\theta_2)\cdot R(\theta_2, \delta_0) \\ &= 1 \times R(\theta_1, \delta_0) + 0 \cdot R(\theta_2, \delta_0) \end{aligned}$$

=> General But informal observation a decision Rule δ is a minimax rule if

1) δ is Bayes with respect to (W.R.t) π

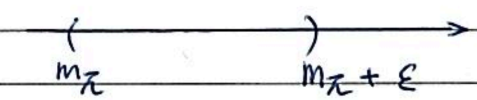
2) $\max_{\theta} R(\theta, \delta) \le r(\pi, d)$

$\Rightarrow$ Examples Bayes:

Def: (informally a δ_0 is extended Bayes if it is a Bayes rule with respect to a certain π)

Formally, A decision Rule δ is extended Bayes if for every $\varepsilon > 0$, δ is ε - Bayes W.R.t a prior π (let $m_\pi = \inf_{\delta} r(\pi, \delta)$)

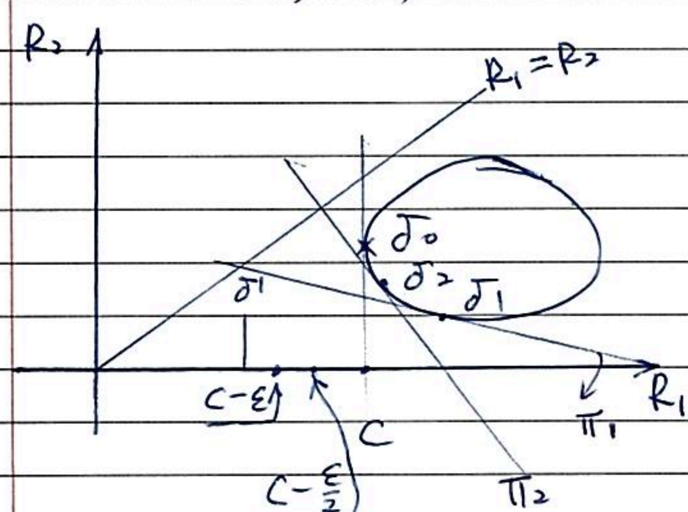
$$r(\pi, d) < m_\pi + \varepsilon$$



$\Rightarrow$ Theorem 2.1:

if δ_n is Bayes with respect to a prior π_n and $r(\pi_n, \delta_n) \to C$ as $n \to +\infty$

and $R(\theta, \delta_0) \le C$ for all $\theta \in \Theta$



$C = R(\theta_1 \delta_0) = \max_{\theta} R(\theta, \delta_0)$

$\pi = (1, 0)$

$r(\pi, \delta_0) = C$

$R(\theta, \delta_0) \le C$

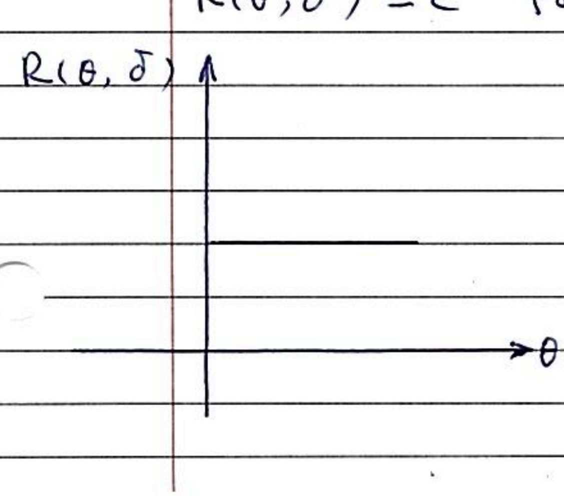
for all $\theta \in \Theta$

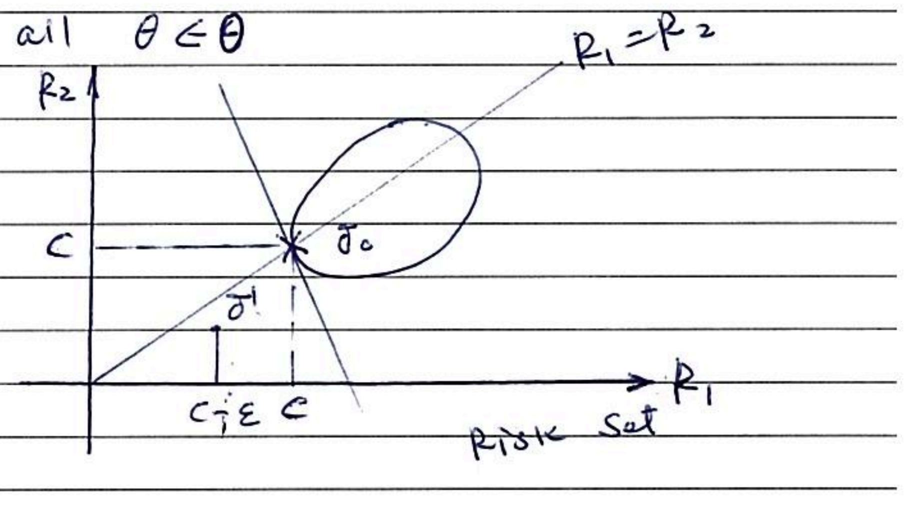
$\Rightarrow$ proof theorem 2.1

Suppose δ_0 satisfied the conditions of the theorem but is not minimax. then there must exist some decision Rule δ' for which $\sup_\theta R(\theta, \delta') < C$, the inequality must be strict, because, if the maximum risk of δ' was the same as that of δ_0, that would not contradict minimaxity of δ_0. So there is an $\varepsilon > 0$ for which $R(\theta, \delta') < C - \varepsilon$ for every θ. Now since $r(\pi_n, \delta_n) - C$, we can find an n for which $r(\pi_n, \delta_n) > C - \varepsilon/2$. But $r(\pi_n, \delta') \leq C - \varepsilon$ therefore, δ_n cannot be the Bayes rule with respect to π_n. This creates a contradiction, and hence proves the theorem.

$\Rightarrow$ theorem 2.2.

Definition: A decision rule δ is equaliser if $R(\theta, \delta) = C$ for all $\theta \in \Theta$





⇒ Theorem 2.2 : A decision Rule δ_0 that is equaliser & extended Bayes must be minimax

proof : 1) Let $c = R(\theta, \delta_0)$ for all $\theta \in \Theta$, we see that $r(\pi, \delta_0) = \int R(\theta, \delta_0)\pi(\theta)d\theta$

$$= \int c \cdot \pi(\theta) d\theta$$

$$= c$$

2) δ_0 is an extended Bayes

Suppose δ_0 is not minimax there must exist a δ' s.t

$$\sup_\theta R(\theta, \delta') < c$$

Let $\varepsilon = c - \sup_\theta R(\theta, \delta')$, then, we write $R(\theta, \delta') \le c - \varepsilon$ for all $\theta \in \Theta$, therefore, $r(\pi, \delta') \le c - \varepsilon$ for all π, then δ_0 is an extended Bayes. By the extended Bayes property of δ_0, we can find a prior π for which

$$r(\pi, \delta_0) = c < \underbrace{\inf_\delta r(\pi, \delta)}_{= m_\pi} + \varepsilon/2$$

that $m_\pi > c - \frac{\varepsilon}{2}$

But $r(\pi, \delta') \le c - \varepsilon$, So this gives another contradiction, and hence proves the theorem.