<< STAT 846 Notes >>

"Statistics Inference"

January 05 2015 -

Aprial 10 2015

<< STAT 846 Notes >>

"Statistics Inference"

January 05 2015 –

April 10 2015

January 05, 2014.

=> Theoretical statistics

Compare different statistical methods.

=> Buy the textbook : Essential of statistical Inference
chapter 1 to chapter 8

=> Lab : 3:30 PM to 4:50 PM

One week one student

=> Lab question -> term-test -> Final exam

plus presentation skill

=> two term test (subjective) + (attendance)

Consider only the higher mark

Final time : April

=> what is statistical inference?

=> observations on n units

$x_1, x_2, \cdots, x_n$ (x_i may be vector)

=> population Distribution

regard $x_1, x_2, \cdots, x_n$ as a realization

of random variables $X_1, \cdots, X_n$.

$\Rightarrow$ For simplicity we assume that

$x_1, x_2, \cdots x_n \overset{iid}{\sim} f(x)$

$f(x)$ is called a population distribution

Non iid example: spatial data (spatial correlation)
time series.

$\Rightarrow$ Furthermore we assume $f(x)$ is of known analytic form, but involves unknow parameters

Example $f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

parametric statistics

$\Rightarrow$ parameter space

$\theta \in \Theta$

Θ: all possible values of θ

Example: $x_1, \cdots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\sigma \in [0, +\infty)$

$\Rightarrow$ statistical inference:

We want to learn some aspects of unknow θ from the observations $x_1, \cdots x_n$

(from sample to population)

probability:

θ is known $\longrightarrow f(x_1, \cdots, x_n \mid \theta)$

Statistics :

θ is unknow $\longleftarrow x_1, \cdots, x_n$

$\Rightarrow$ Different types of inference :

(1) point estimation $\hat{\theta}(x_1, \cdots, x_n)$

$\hat{\theta}$ is a single number to capture θ [illegible]

(2) interval estimation

$\theta \in (L(x_1, \cdots, x_n), \; \mu(x_1, \cdots, x_n))$

The true parameter is within this interval

(3) Hypothesis testing

$\theta = \theta_0$ Vs $\theta \neq \theta_0$

$\theta \in \Theta_0$ Vs $\theta \in \Theta_1$

$\theta < \theta_0$ Vs $\theta > \theta_0$

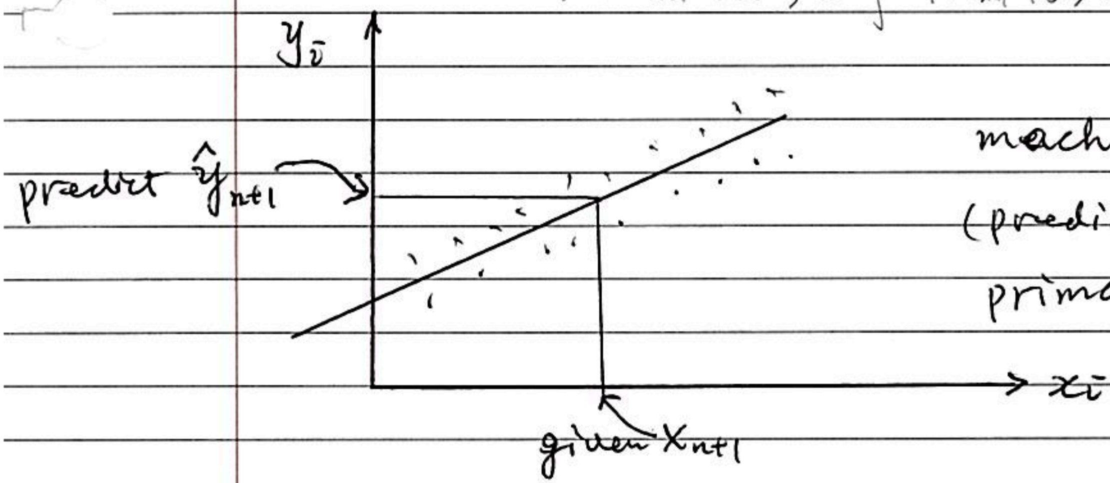
(4) Predictive inference $x_1, x_2, \cdots, x_n \xrightarrow{\text{predict}} x_{n+1}$

Example : $f(y_i \mid x_i, \hat{\theta})$

from $(x_1, y_1), \cdots, (x_n, y_n)$

We want to predict y_{n+1} given x_{n+1}

$$f(x_{n+1} \mid X) \overset{?}{=} \int f(x_{n+1} \mid \theta)\, \pi(\theta \mid X)\, dx$$



machine learning
(prediction is the primary goal)

⇒ how to do : standard paradigms for stat inference :

- Bayesian inference

  prior distribution $\theta \sim \pi(\theta)$

  data model $D(\theta \sim f(d \mid \theta))$

  posterior : $f(\theta \mid D=d) = \dfrac{\pi(\theta) f(d \mid \theta)}{\int \pi(\theta) f(d \mid \theta)\, d\theta}$

- In 1920, fisher devises a system of inference approaches that we indep of π

  repeated sampling principle :

  $\hat{\theta}(x_1, \cdots, x_n)$ has a sampling distribution

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

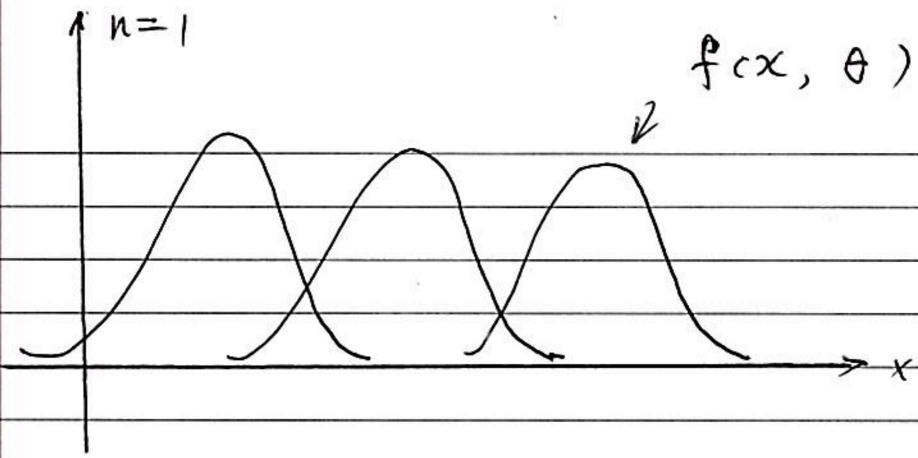
$\{ P_\theta(x) \mid \theta \in \Theta \}$ how likely we see the data given a θ

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

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

$L(\theta, a)$ 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 : Average for different possible sample.

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

decision by data

therefore;

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

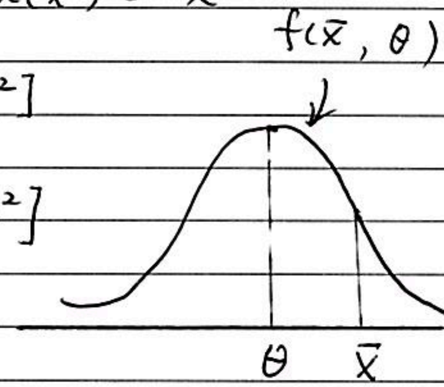
Example :

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

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

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

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

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

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

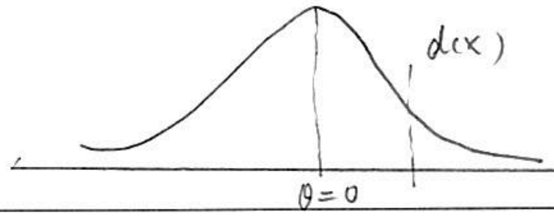
2) In Hypothesis testing

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

	Accept H_0	reject H_0
H_0	OK	I
H_1	II	OK

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

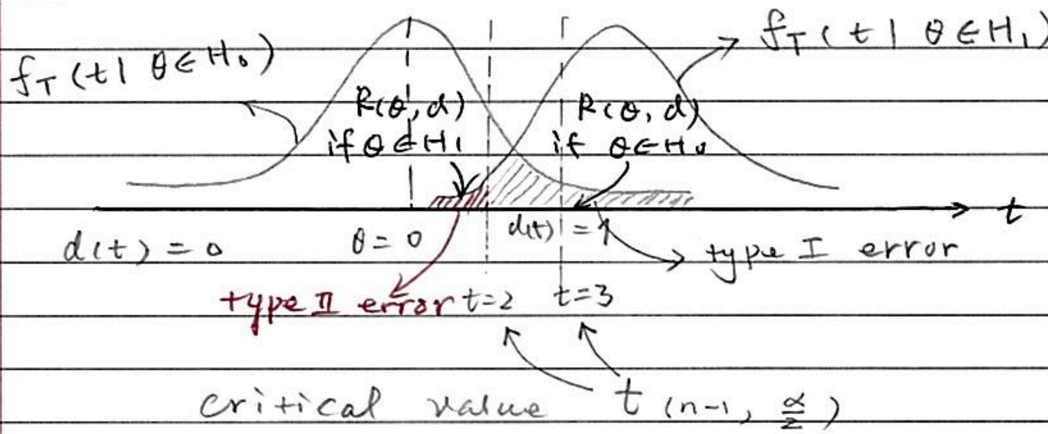
$$L(\theta, a) = \begin{cases} 0 & \text{if } \theta \in H_0 \quad a = 0 \\ 0 & \text{if } \theta \in H_1, \quad a = 1 \\ 1 & \text{if } \theta \in H_0 \quad a = 1 \\ & \quad \text{this is a type I error} \\ 1 & \text{if } \theta \in H_1 \quad a = 0 \\ & \quad \text{this is a type II error} \end{cases}$$



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

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

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

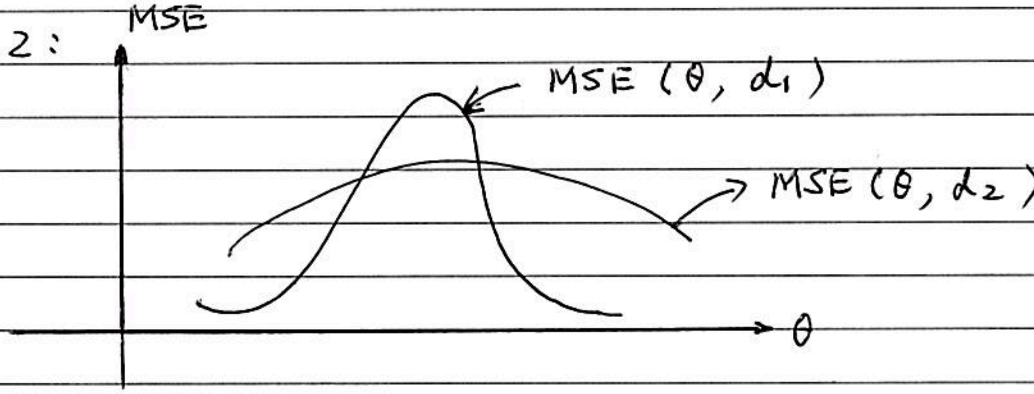
$d'(t)$: $d'=1$, if $t>3$

$d'=0$, 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.

$\Rightarrow$ Remarks:

1) d is very poor if d is inadmissible;

2) Many daily used statistics are inadmissibles.

examples:

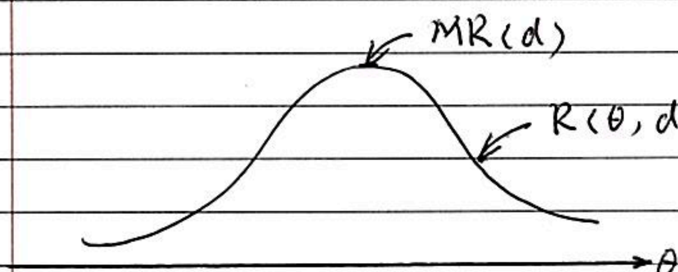
1) $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)$

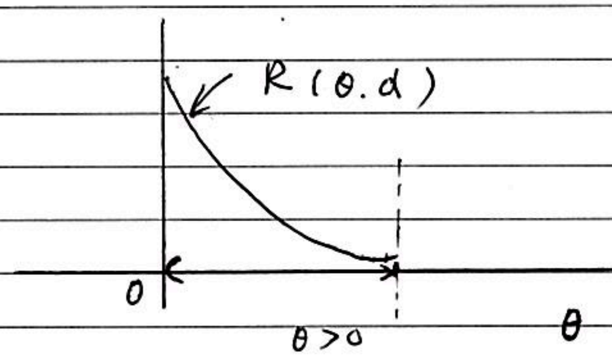
2) For $d \geq 3$ $\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_d)$ is inadmissible

斯特恩的悖论 stern's paradox

$\Rightarrow$ minimax decision Rules: supremum 上确界 最小上界

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



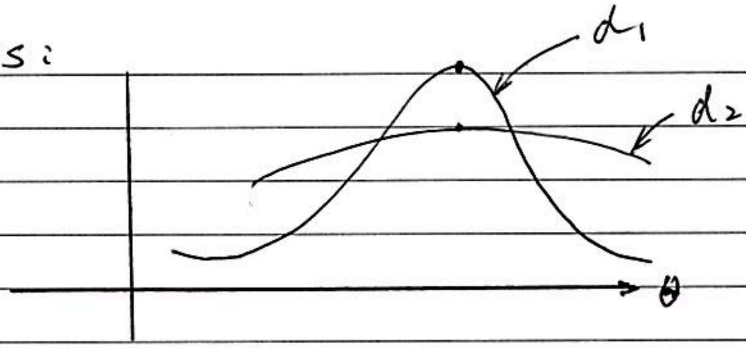


Def for minimax:

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

⇒Remarks:

1)

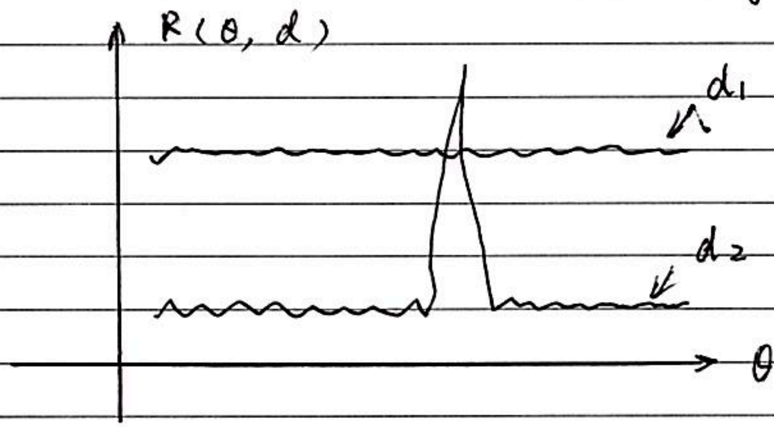


$D = \{d_1, d_2\}$

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',

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

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

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

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

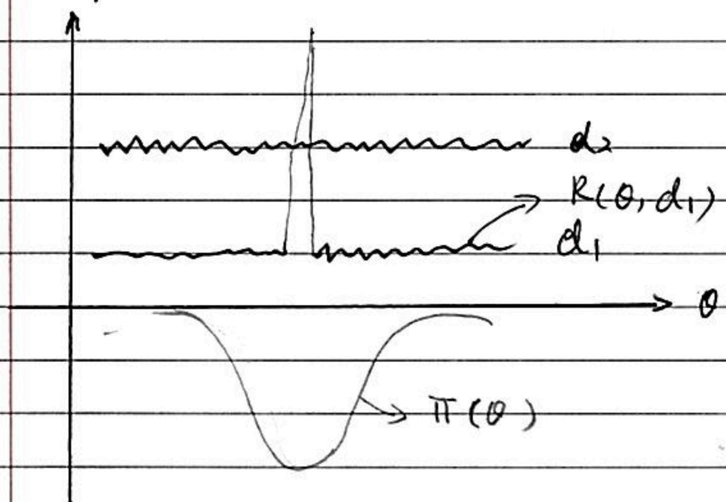
if $\phi = \theta$, then this statement is obviously true

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

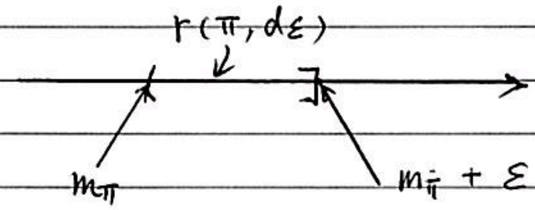
Let $m_\pi = \inf_{d' \in D} r(\pi, d')$, for any $\varepsilon > 0$

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

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

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 (denote toss the two coins) $\xrightarrow{\frac{1}{4}\ (HH)} d_1$

X (denote toss the two coins) $\xrightarrow{3/4\ (o/w)} d_2$

$\Rightarrow$ 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)

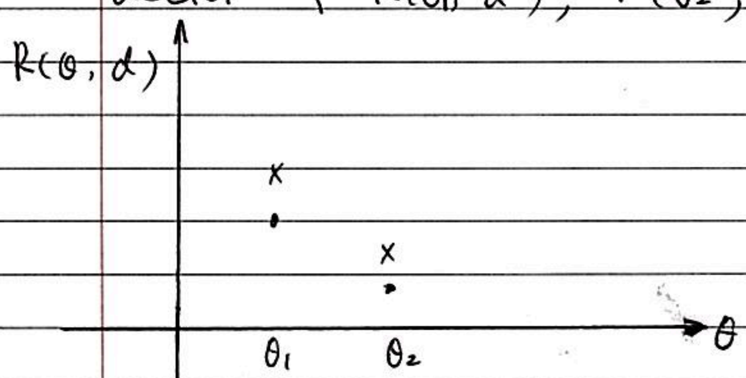
parameter space is a finite set:

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

$\Rightarrow$ Risk set

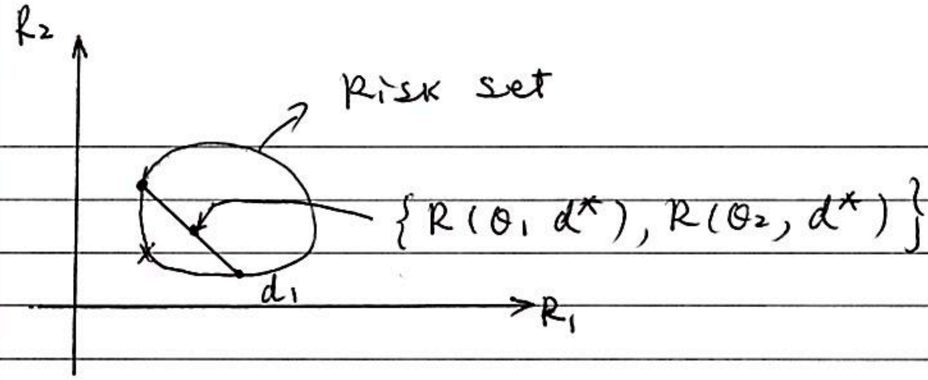
a subst in $\mathbb{R}^t$ containing all possible

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



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

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

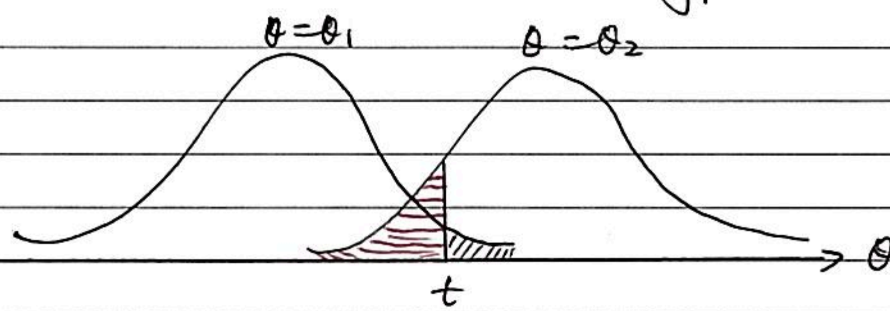


while $d^* = P_1 d_1 + P_3 d_3$

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

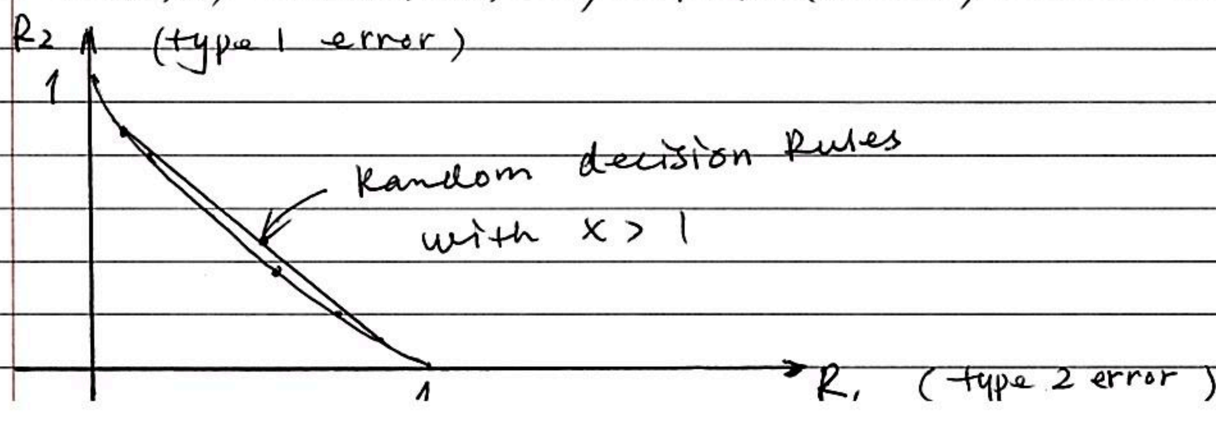
θ_2 is alternative hypothesis



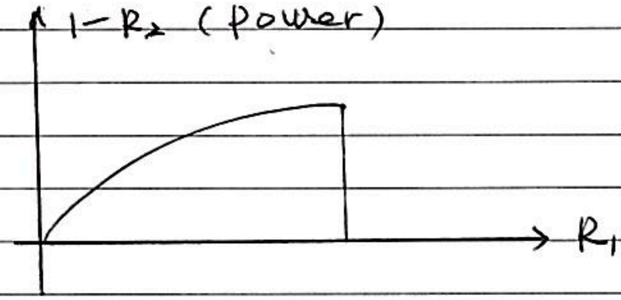
$$R(\theta, d) = \begin{cases} Pr_{\theta_1}(d(x)=1) & \text{if } \theta=\theta_1 \\ Pr_{\theta_2}(d(x)=0) & \text{if } \theta=\theta_2 \end{cases} ; \quad d(x) = \begin{cases} 1 & x>t \\ 0 & x<t \end{cases}$$

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

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



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

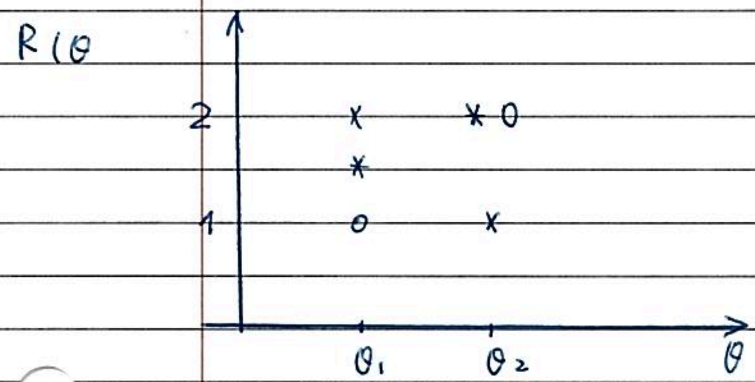


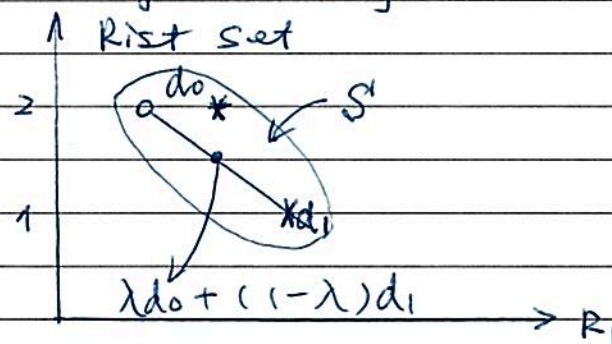
<< January 14. 2015 >>

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



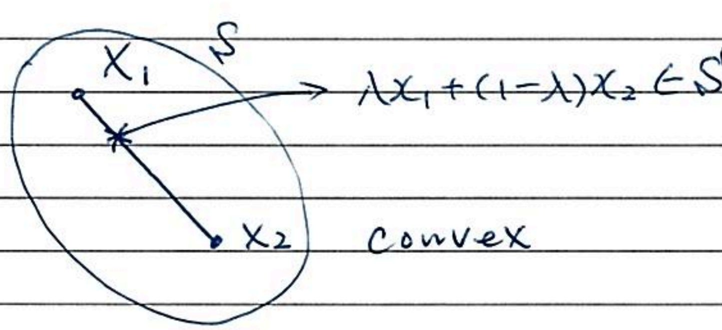


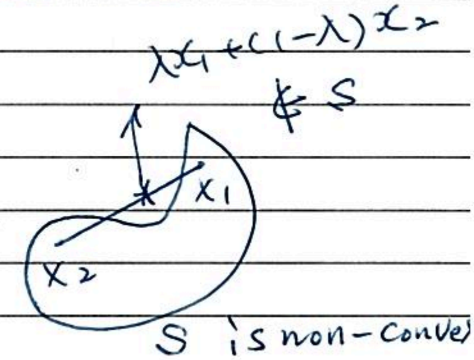
o : d_0 $\quad$ * : d_1

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

$x_2 = \{ R(\theta_1, d_2), \dots \; R(\theta_1, d_2) \}$

$x_1, \; x_2 \in S'$, thus for $\lambda \in [0, 1]$

$\lambda x_1 + (1-\lambda) x_2 = \{ \lambda R(\theta_1, d_1) + (1-\lambda) R(\theta_t, d_2), \dots$

$\dots \lambda R(\theta_t, d_1) + (1-\lambda) R(\theta_t, d_2) \}$

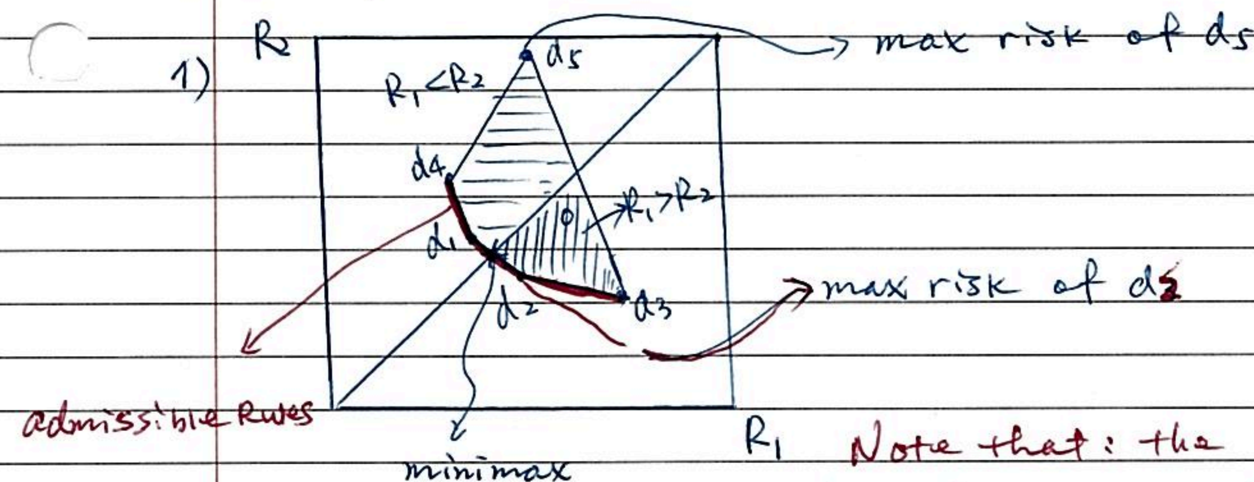
we see that $\lambda x_1 + (1-\lambda) x_2$ is the risk

function (vector) of $\lambda d_1 + (1-\lambda) d_2$

So, $\lambda x_1 + (1-\lambda) x_2 \in S'$

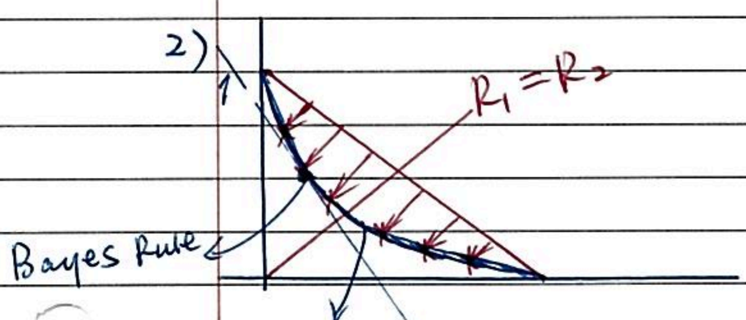
$\Rightarrow$ Examples of Risk set

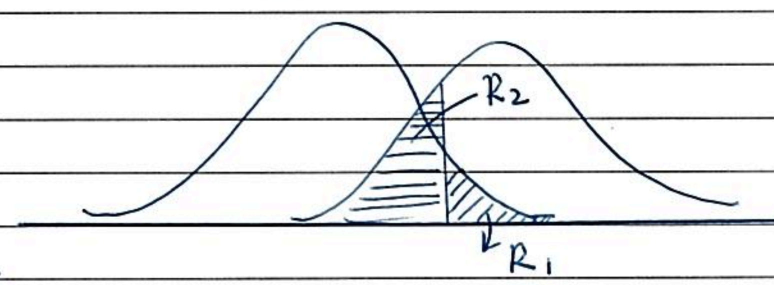
1)



Note that: the lef-hand boundary are admissible rules.

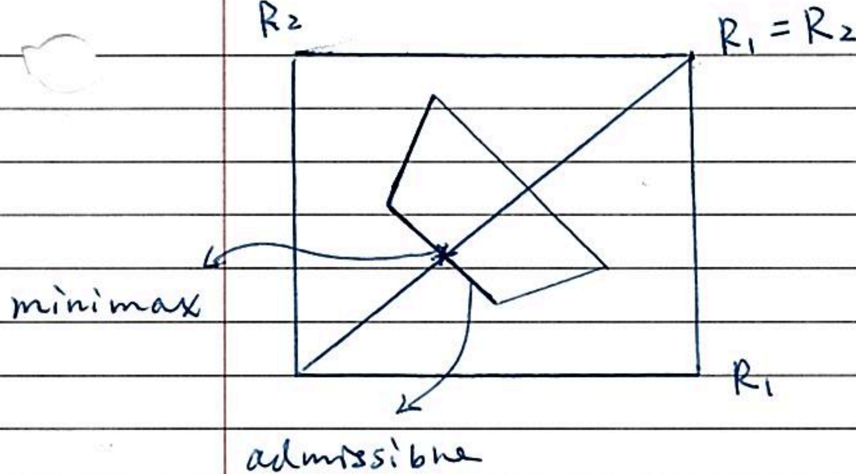
2)





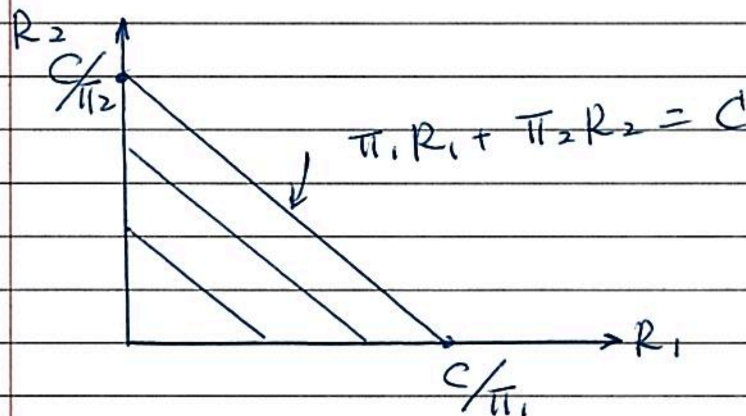
the t, s.t, $R_1 = R_2$ is minimax

* Any randomised decision rules are inadmissible.

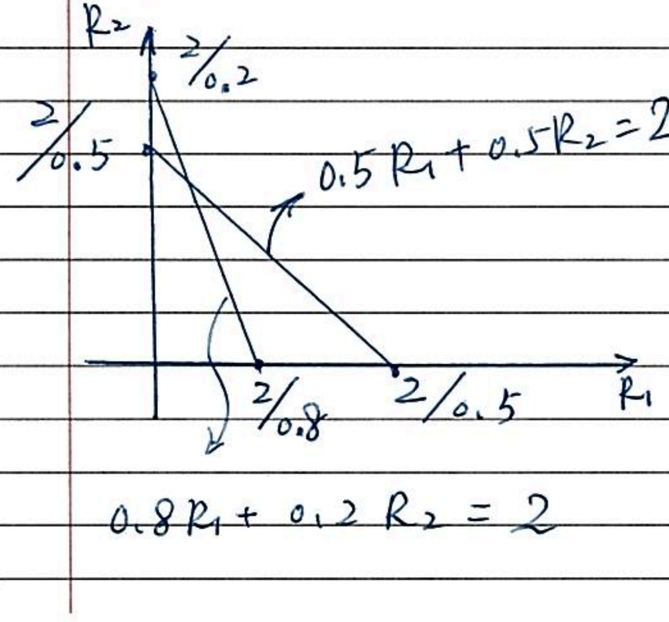


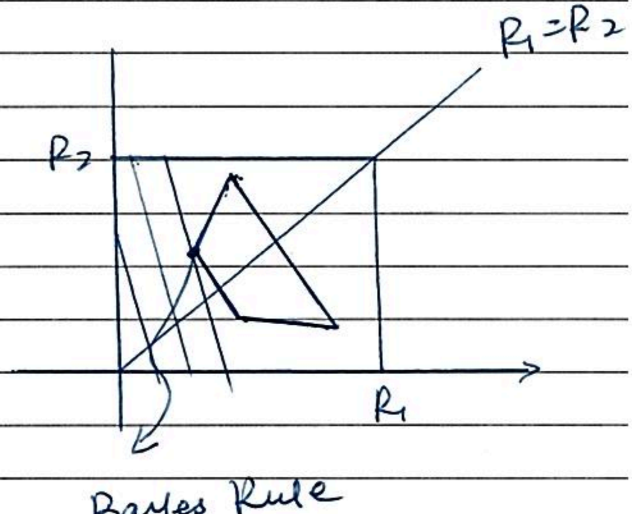
$\Rightarrow$ Bayes Risk : Given $\pi = (\pi_1, \pi_2)$

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



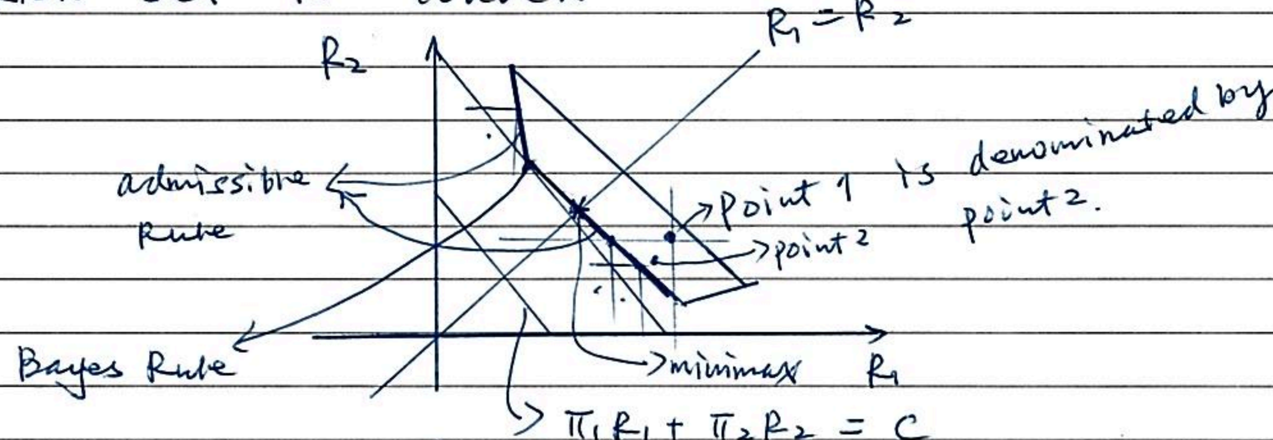
examples:

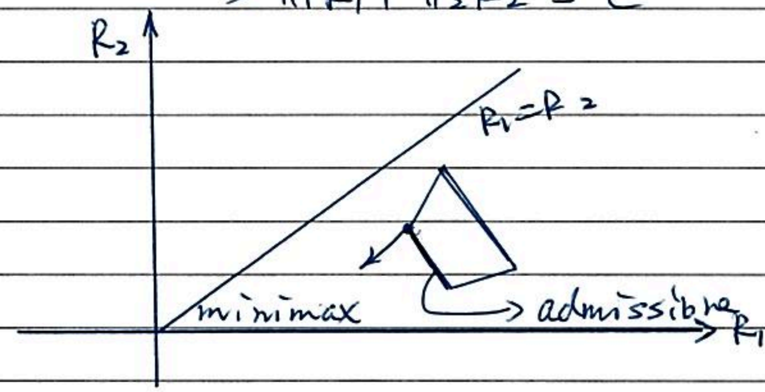




<< January 16, 2014 >> STAT : 846

⇒ 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

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

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

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

$\Rightarrow$ 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

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

$\Rightarrow$ Risk functions :

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

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

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

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

$$R(2, d_3(x)) = L(2\ d_3(x=1))\times\tfrac{1}{2} + L(2, d_3(x=2))\times\tfrac{1}{2}$$
$$= L(2,1)\times\tfrac{1}{2} + L(2,2)\times\tfrac{1}{2}$$
$$= 1\times\tfrac{1}{2} + 0\times\tfrac{1}{2}$$
$$= \tfrac{1}{2}$$

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

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

$$= 1\times 1 = 1$$

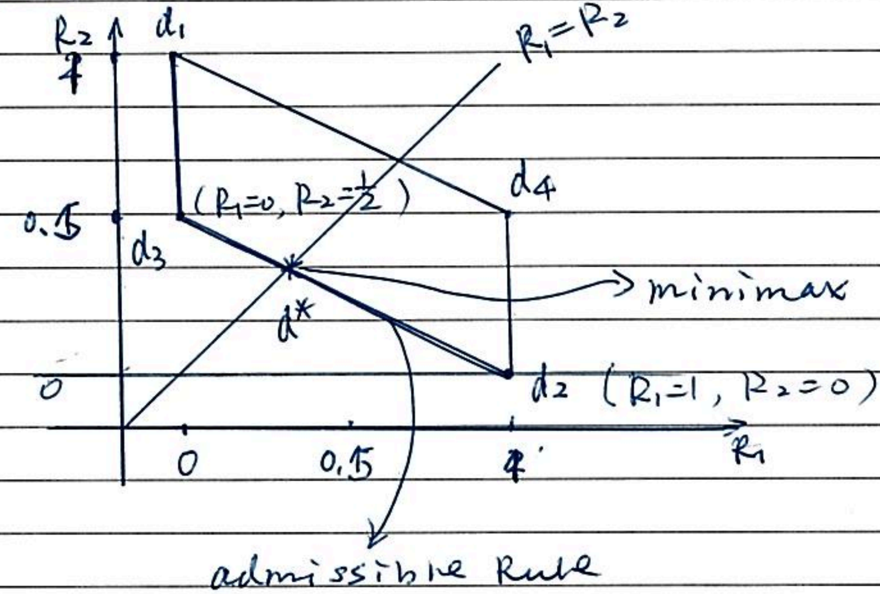
$$R(2, d_4(x)) = L(2, d_4(x=1)) \times \tfrac{1}{2} + L(2, d_4(x=2)) \times \tfrac{1}{2}$$

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

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

$$= 1/2$$

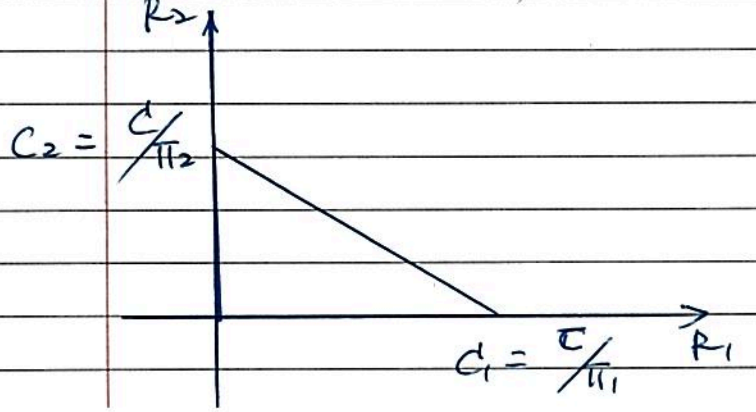
$\Rightarrow$ Risk set :



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

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

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



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

$$C_2 R_1 + C_1 R_2 = C$$

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

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

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

$\Rightarrow \lambda = 2/3$

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

$\Rightarrow$ 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.

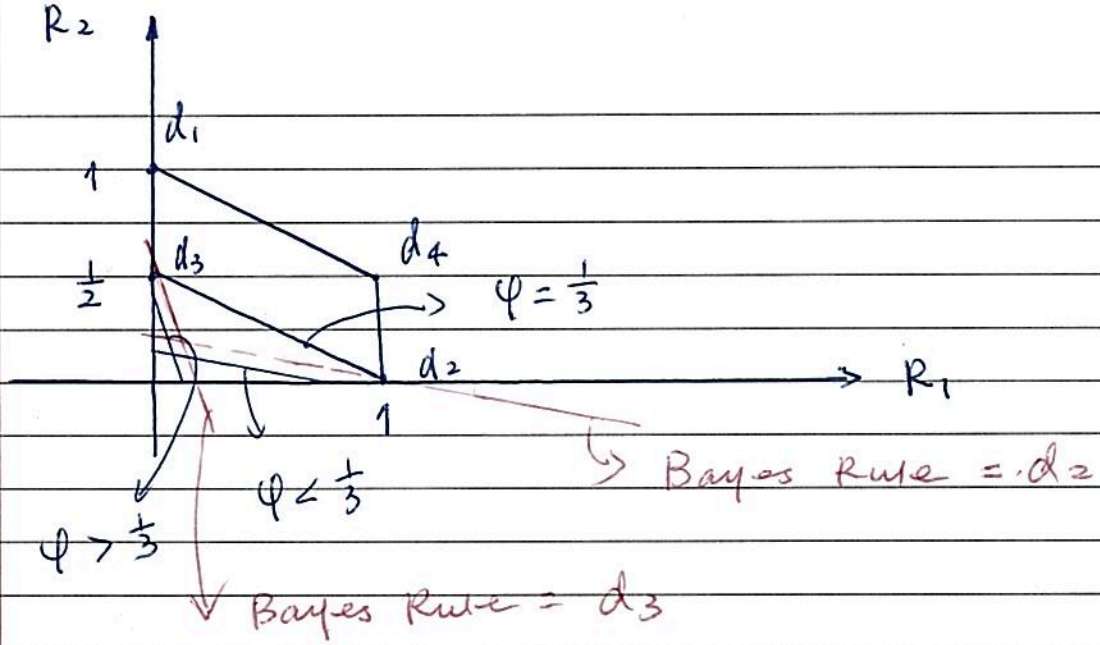
$\pi(\theta=1) = \varphi.$ (in our case $\varphi = 1/3$

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

$\Rightarrow$ 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 $= \lambda d_3 + (1-\lambda) d_2$

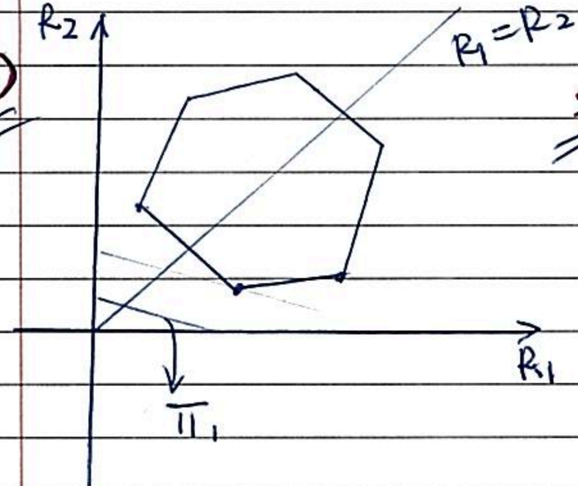
(3) $\varphi > 1/3$, Bayes Rule $= d_3$

<< January 19, 2014 >> STAT 846

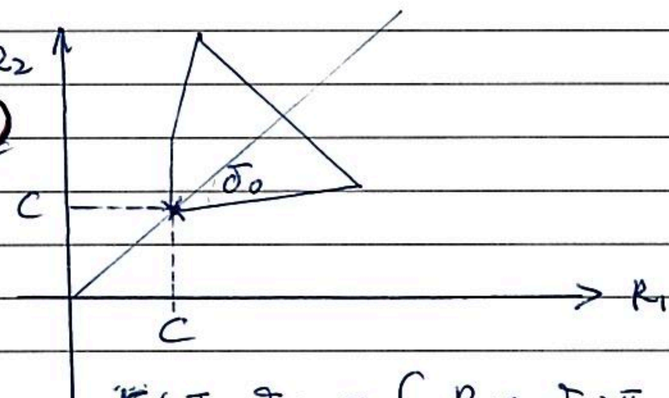
⇒ In chapter 3, we will see it is easy to find a Bayes Rule with a prior π, How to find minimax from Bayes Rules.

⇒ Some Toy examphe

1)

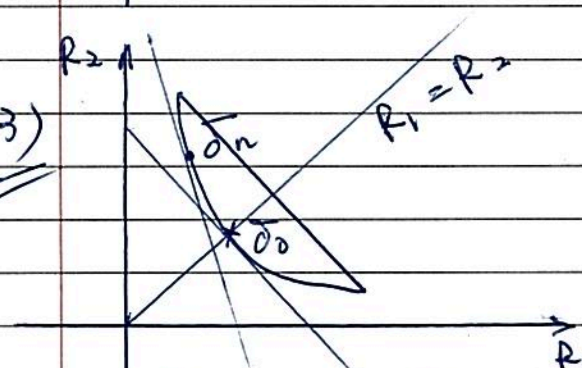


2)



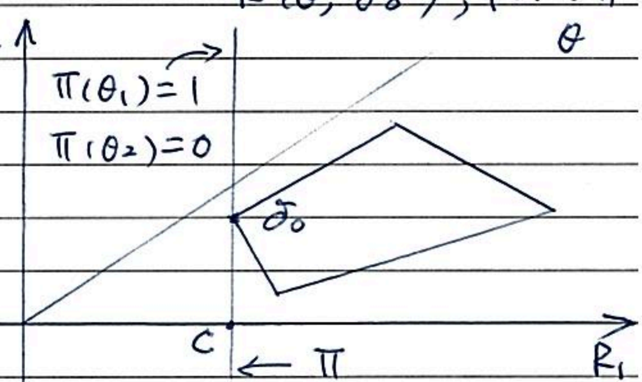
$$r(\pi, \delta_0) = \int_\Theta R(\theta, \delta_0)\pi(\theta)\,d\theta = R(\theta, \delta_0), \text{ for all } \theta$$

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

$\Rightarrow$ 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 $m_\pi = \inf_{\delta} r(\pi, \delta)$)

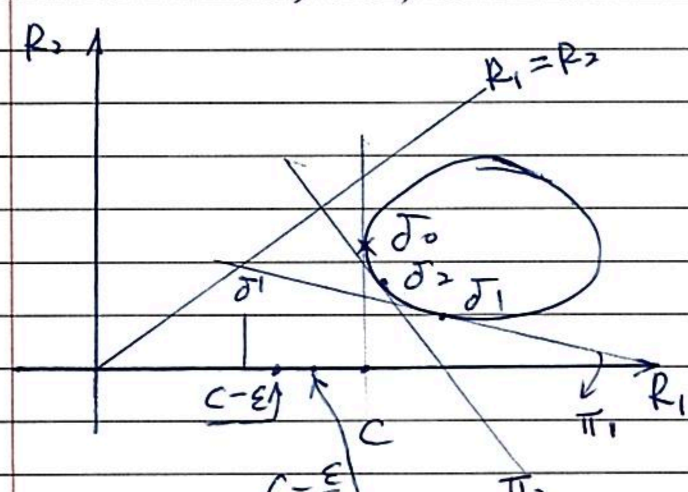
$$r(\pi, d) < m_\pi + \varepsilon$$

m_π $\qquad$ $m_\pi + \varepsilon$

$\Rightarrow$ 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 $R(\theta, \delta_0) \leq C$ for all $\theta \in \Theta$



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

$\pi = (1, 0)$

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

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

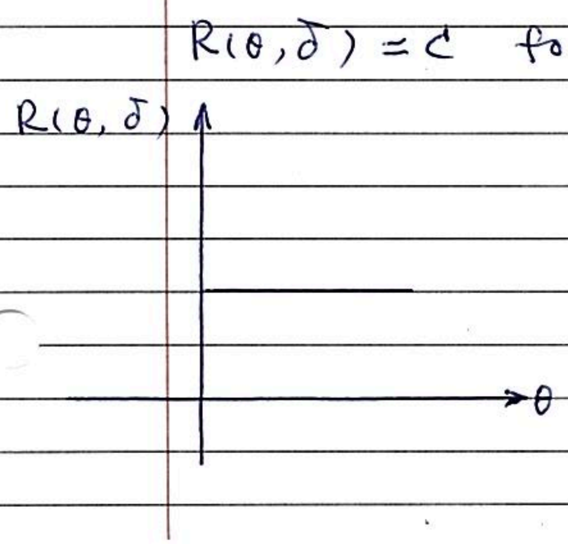
~~for all $\theta \in \Theta$~~

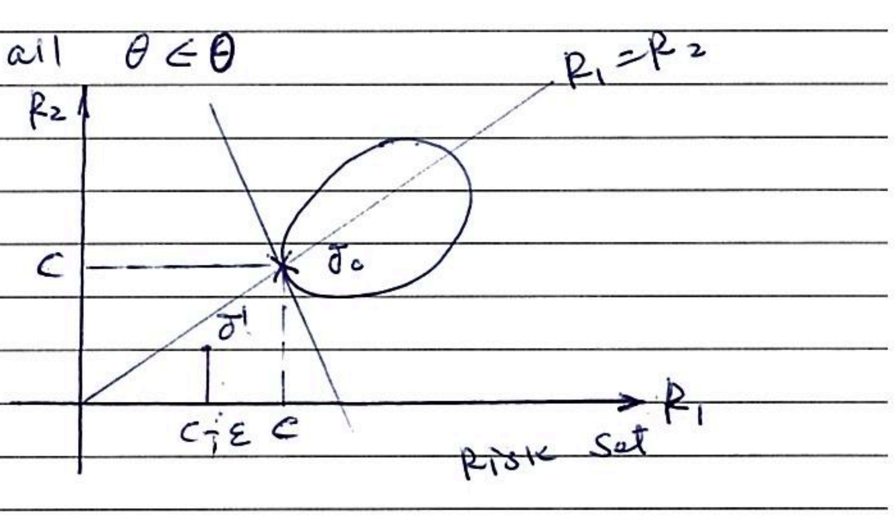
⇒ proof theorem 2.1

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 $r(\pi_n, \delta_n) \to C$, we can find an n for which $r(\pi_n, \delta_n) > C - \varepsilon/2$. But $r(\pi_n, \delta') \le C - \varepsilon$ therefore, δ_n cannot be the Bayes rule with respect to π_n. This creates a contradiction, and hence proves the theorem.

⇒ 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.

<< January 21, 2015 >> STAT 846

=> Admissibility, Bayes Rules

=> Theorem 2.3 : Assume that $\Theta = \{\theta_1, \cdots, \theta_t\}$ is finite, and that the prior $\pi(\cdot)$ gives positive probability to each θ_i. Then a Bayes rule with respect to $\pi(\cdot)$ is admissible.

=> Theorem 2.4 : if a Bayes Rule is unique, it is admissible.

=> theorem 2.5

(1) Let Θ be a subset of the real line;

(2) Assume that the risk functions $R(\theta, d)$ are continuous in θ for all decision rules d;

(3) Suppose that for any $\varepsilon > 0$ and any θ the interval $(\theta - \varepsilon, \theta + \varepsilon)$ has positive probability under the prior $\pi(\cdot)$.

Then, a Bayes rule with respect to $\pi(\cdot)$ is admissible.

Chapter 3: Bayesian methods

3.1 Fundamental elements of Bayesian inferrence

Bayes theorem (Law)

suppose $\theta \sim \pi(\theta)$; $x | \theta \sim f(x; \theta)$

then the posterior density of θ, given x

$$\pi(\theta | x) = \pi(\theta) \cdot f(x; \theta) \Big/ \int_\Theta \pi(\theta) \cdot f(x; \theta)\, d\theta$$

$$\pi(\theta | x) \propto \pi(\theta) \cdot \text{likelihood}$$

⇒ examples :

$$X | \theta \sim Bin(n, \theta) \Rightarrow \pi(x | \theta) = \binom{n}{x} \cdot \theta^{x} (1-\theta)^{n-x}$$

$$\theta \sim Beta(a, b) \Rightarrow \pi(\theta) = \frac{\theta^{a-1} \cdot (1-\theta)^{b-1}}{B(a, b)}$$

$$B(a, b) = \int_0^1 \theta^{a-1} \cdot (1-\theta)^{b-1}\, d\theta$$

⇒ A note :

$$\pi(\theta | X) = \frac{\pi(\theta) f(x; \theta)}{\int_\Theta \pi(\theta) f(x; \theta)\, d\theta} = \frac{C\, \pi(\theta) \cdot f(x; \theta)}{\int C\, \pi(\theta) \cdot f(x; \theta)\, d\theta}$$

C is free of θ

⇒ how to find the posterior $\pi(\theta | X)$

example 1 : $\pi(\theta) \cdot f(x; \theta) \propto \theta^{a-1} \cdot (1-\theta)^{b-1} \cdot \theta^{x} \cdot (1-\theta)^{n-x}$

$$= \theta^{a+x-1} \cdot (1-\theta)^{n+b-x-1}$$

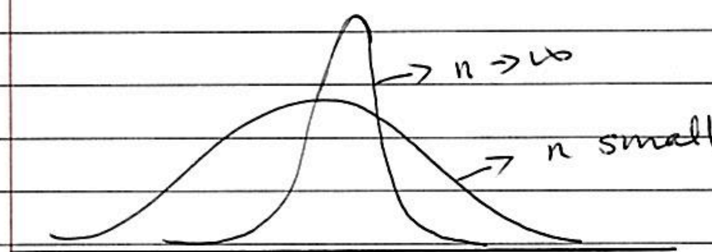
So, we can conclude that:

$$\theta \mid x \sim B(a+x,\ b+n-x)$$

$$E(\theta \mid x) = \frac{a+x}{a+b+n} \qquad \text{as } n \to \infty \quad E(\theta \mid x) \approx \frac{x}{n}$$

$$var(\theta \mid x) = \frac{(a+x)(n+b-x)}{(a+b+n)^2(a+b+n+1)} \approx \frac{x(n-x)}{n^3}$$

$$= \frac{x}{n} \cdot \left(1 - \frac{x}{n}\right) \cdot \frac{1}{n} = \hat{p}(1-\hat{p}) \cdot \frac{1}{n}$$



example 2:

$x_1, x_2, \cdots, x_n \mid \mu \qquad x_i \sim N(\mu, \alpha^2)$

assume α^2 is know, $\mu \sim N(\mu_0, \sigma_0^2)$

To find $\pi(\mu \mid x_1, \cdots, x_n) \quad \underset{\sim}{x} = (x_1, \cdots, x_n)$

$$\pi(\mu \mid x_1, \cdots, x_n) \propto \pi(\mu) \cdot f(\underset{\sim}{x}; \theta)$$

$$= e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \cdot e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\alpha^2}}$$

$$\propto e^{-\left\{\left(\frac{1}{2\sigma_0^2} + \frac{n}{2\alpha^2}\right)\mu^2 + \left(-\frac{\mu_0}{\sigma_0^2} - \frac{\Sigma x_i}{\sigma^2}\right)\mu\right\}}$$

We see $\mu \mid \underset{\sim}{x}$ is a normal, suppose $N(\mu_1, \sigma_1^2)$

We will find μ_1 and σ_1^2

$$\pi(\mu \mid \underset{\sim}{x}) \propto e^{-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}} \propto e^{-\left(\frac{1}{2\sigma_1^2}\mu^2 - \frac{\mu_1}{\sigma_1^2}\mu\right)}$$

So, $$\frac{1}{2\sigma_1^2} = \frac{1}{2\sigma_0^2} + \frac{n}{2\sigma^2} \iff$$

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

Let $\tau_0 = \frac{1}{\sigma_0^2}$; $\tau_1 = \frac{1}{\sigma_1^2}$; $\tau = \frac{1}{\sigma^2}$

$$\Rightarrow \tau_1 = \tau_0 + n \cdot \tau$$

$\iff$ posterior precision = prior precision + $\sum_{i=1}^{n}$ precision of x

So, $$\frac{\mu_1}{\sigma_1^2} = \frac{\mu_0}{\sigma_0^2} + \frac{\Sigma x_i}{\sigma^2} \iff$$

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\Sigma x_i}{\sigma^2}}{\frac{1}{\sigma_1^2}} = \frac{\mu_0\tau_0 + \bar{x} \cdot n \cdot \tau}{\tau_1}$$

$$= \frac{\mu_0\tau_0 + \bar{x} \cdot n\tau}{\tau_0 + n\tau}$$

So, $$\mu \mid x_1, \dots, x_n \sim N\left(\frac{\mu_0\tau_0 + \bar{x} \cdot n\tau}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau}\right)$$

<< January 23, 2015 >> STAT 846

Examples:

1) parameter (prior Dis. $\pi(\theta)$)

$\theta \sim$	1	2	3
$\pi(\theta)$	$1/3$	$1/3$	$1/3$

Data Distribution

$X \mid \theta \sim$	$\theta-1$	θ	$\theta+1$
$\pi(X\mid\theta)$	$1/4$	$1/2$	$1/4$

$$\pi(\theta \mid X) \propto \pi(\theta)\cdot\pi(X\mid\theta)$$

$\pi(\theta)\cdot\pi(X\mid\theta)$

$\theta \backslash X$	0	1	2	3	4
1	$\frac{1}{3}\times\frac{1}{4}$	$\frac{1}{3}\times\frac{1}{2}$	$\frac{1}{3}\times\frac{1}{4}$	0	0
2	$\frac{1}{3}\times 0$	$\frac{1}{3}\times\frac{1}{4}$	$\frac{1}{3}\times\frac{1}{2}$	$\frac{1}{3}\times\frac{1}{4}$	0
3	$\frac{1}{3}\times 0$	$\frac{1}{3}\times 0$	$\frac{1}{3}\times\frac{1}{4}$	$\frac{1}{3}\times\frac{1}{2}$	$\frac{1}{3}\times\frac{1}{4}$

$\pi(\theta \mid X=0) \sim$

θ	1	2	3
$\pi(\cdot)$	1	0	0

$$\pi(\theta\mid X) = \frac{\pi(\theta)\cdot\pi(X\mid\theta)}{\sum_{\theta=1}^{3}\pi(\theta)\cdot\pi(X\mid\theta)}$$

$\pi(\theta \mid X=1) \sim$

θ	1	2	3
$\pi(\cdot)$	$\frac{2}{3}$	$\frac{1}{3}$	0

$\pi(\theta \mid X=2) \sim$

θ			
$\pi(\cdot)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

2) $X_1, \cdots, X_n \mid (\mu, \tau) \sim N(\mu, \frac{1}{\tau})$

$\tau \sim \text{Gamma}(\alpha, \beta)$

$$\pi(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \cdot e^{-\beta\tau}$$

$$\mu \mid \tau \sim N(\nu, \frac{1}{k\tau})$$

$$\pi(\mu \mid \tau) = \frac{1}{\sqrt{2\pi}} (k\tau)^{\frac{1}{2}} e^{-\frac{k\tau}{2}(\mu-\nu)^2}$$

$$\pi(\tau, \mu) \propto \tau^{\alpha-\frac{1}{2}} e^{-\tau(\beta + \frac{k}{2}(\mu-\nu)^2)}$$

$$\pi(\tau, \mu \mid X_1, X_2, \cdots, X_n) \propto \pi(\tau, \mu) \cdot f(X_1, \cdots X_n; \tau, \mu)$$

$$\propto \tau^{\alpha-\frac{1}{2}} \cdot e^{-\tau(\beta + \frac{k}{2}(\mu-\nu)^2)} \prod_{i=1}^{n} \tau^{\frac{1}{2}} \cdot e^{-\frac{\tau}{2}(X_i-\mu)^2}$$

$$\propto \tau^{\alpha' - \frac{1}{2}} \cdot e^{-\tau(\beta' + \frac{k'}{2}(\mu-\nu')^2)}$$

where,
$$\alpha' = \alpha + n/2$$

$$\beta' = \beta + \frac{1}{2} \cdot \frac{nk}{n+k} (\bar{X} - \nu)^2 + \frac{1}{2} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$k' = k + n$$

$$\nu' = \frac{k\nu + n\bar{X}}{k+n}$$

$\Rightarrow$ Conjugate prior : if the parametric form of the prior and the posterior is the same, then we say this prior is conjugate to this problem.

⇒ General form of Bayes rule, give a fixed prior.

⇒ Risk function:

$$R(\theta, d) = \int_X L(\theta, d(x))\, f(x,\theta)\, dx$$

Bayes Risk:

$$r(\pi, d(x)) = \int_\theta \pi(\theta) \cdot R(\theta, d(x))\, d\theta$$

$$= \int_\theta \pi(\theta) \left[\int_X L(\theta, d(x)) f(x,\theta)\, dx \right] d\theta$$

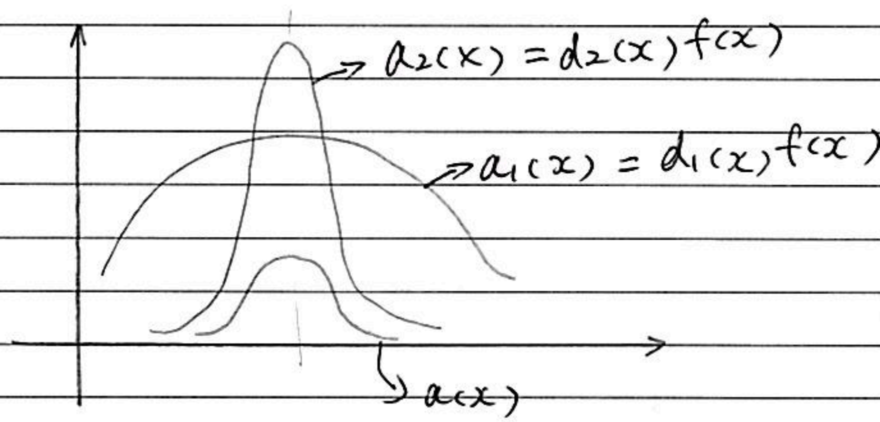
$$= \int_X f(x) \left\{ \int_\theta L(\theta, d(x)) \cdot \frac{\pi(\theta) \cdot f(x,\theta)}{f(x)}\, d\theta \right\} dx$$

where, $f(x) = \int_\theta \pi(\theta) \cdot f(x,\theta)\, d\theta$

$$= \int_X f(x) \cdot \left[\int_\theta L(\theta, d(x))\, \pi(\theta|x)\, d\theta \right] dx$$

To find Bayes Rule $d(x)$ to minimize $r(\pi, d(x))$ is equivalent to that for each x, we find $d(x)$ to minimize $\int_\theta \underbrace{L(\theta, d(x))\, \pi(\theta|x)}_{a(x)}\, d\theta = E_{\theta|x}\{L(\theta, d(x))\}$

Called expected posterior loss



⇒ How to find the Bayes Rule :

examples:

1) Point estimate:

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

$$E_{\theta|x}(L(\theta, d(x))) = E_{\theta|x}((\theta - d(x))^2) = A(d(x))$$

To find a value $d(x)$ to minimize $A(d(x))$,

Solving $\frac{\partial}{\partial d}(A(d(x))) = E_{\theta|x}\left(\frac{\partial}{\partial d}(\theta - d)^2\right) = E_{\theta|x}(2(\theta - d)(-1))$

$$= 0$$

$\Rightarrow E_{\theta|x}(\theta) - d(x) = 0 \quad \Rightarrow d(x) = E_{\theta|x}(\theta)$

So, the Bayes Rule is the mean of posterior θ.

2) Point estimate:

Loss function : $L(\theta, d) = |\theta - d|$

Solution: $E_{\theta|x}(L(\theta, d)) = E_{\theta|x}(|\theta - d|)$

$$= \int_{-\infty}^{\infty} |\theta - d| \, \pi(\theta|x) \, d\theta$$

$$= \int_{-\infty}^{d} (d - \theta) \pi(\theta|x) d\theta + \int_{d}^{\infty} (\theta - d) \pi(\theta|x) d\theta$$

Differential : $\frac{\partial}{\partial d}\{E_{\theta|x}(L(\theta, d))\} = 0$

$$\Rightarrow \int_{-\infty}^{d} \pi(\theta|x) d\theta - \int_{d}^{\infty} \pi(\theta|x) d\theta = 0$$

$\because \int_{-\infty}^{\infty} \pi(\theta|x) d\theta = 1$

$$\Rightarrow \int_{-\infty}^{d} \pi(\theta|x) d\theta = \int_{d}^{\infty} \pi(\theta|x) d\theta = \frac{1}{2}$$

« January 26, 2015 » STAT 846

⇒ Examples of Bayes Rules

1) point estimates:

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

$$E_{\theta|x}(|\theta - d(x)|) = \int_{-\infty}^{\infty} |\theta - d(x)| \, \pi(\theta|x) \, d\theta$$

$$= r(d(x)|x)$$

$$\frac{\partial r(d(x)|x)}{\partial d(x)} = \int_{-\infty}^{\infty} [1 \times I(\theta < d) + (-1) I(\theta > d)] \pi(\theta|x) \, d\theta$$

$$= \int_{-\infty}^{d} \pi(\theta|x) \, d\theta - \int_{d}^{\infty} \pi(\theta|x) \, d\theta = 0$$

$$\Rightarrow \Pr(\theta < d|x) = \Pr(\theta > d|x)$$

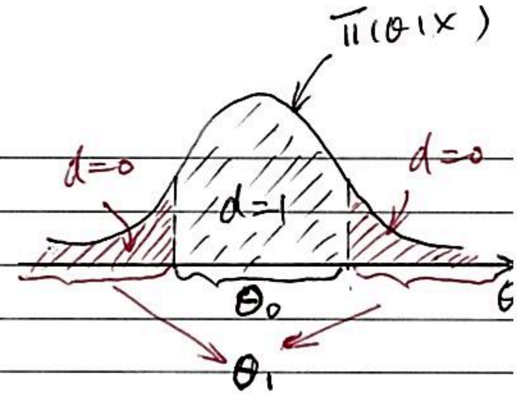
$d(x)$ = median of $\pi(\theta|x)$

2) In hypothesis testing

$$L(\theta, a) = \begin{cases} 1 & \text{if } \theta \in \Theta_0, \ a = 1 \\ 1 & \text{if } \theta \in \Theta_1, \ a = 0 \\ 0 & \text{o/w} \end{cases}$$

$$r(d|x) = \begin{cases} E_{\theta|x}\{L(\theta, d=1)\}, & \text{if } d = 1 \\ E_{\theta|x}\{L(\theta, d=0)\} & \text{if } d = 0 \end{cases}$$

$$= \begin{cases} Pr(\theta \in \Theta_0 | x) & \text{if } d=1 \\ Pr(\theta \in \Theta_1 | x) & \text{if } d=0 \end{cases}$$



Bayes Rule: $d(x)$:

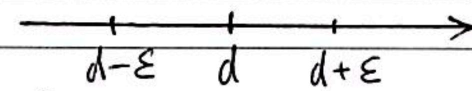
$$d(x) = \begin{cases} 1 & \text{if } Pr(\theta \in \Theta_1 | x) \geq Pr(\theta \in \Theta_0 | x) \\ 0 & \text{if } Pr(\theta \in \Theta_0 | x) \geq Pr(\theta \in \Theta_1 | x) \end{cases}$$

We will come back in chapter 4.

=> Lindley paradox

3) Interval estimate

$$\mathcal{A}_1 = \{d\} = \theta$$



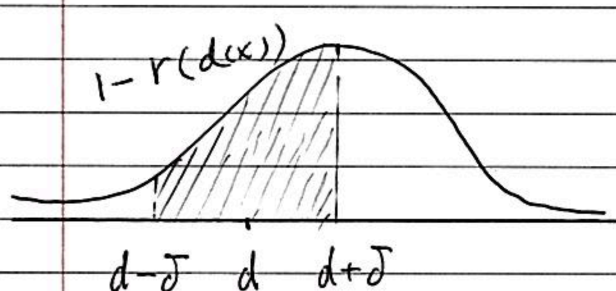
$$\mathcal{A}_1 = \{(d-\delta, d+\delta) \mid d \in \theta\}$$

Prescribe the length of in interval δ

$$L(\theta, d) = \begin{cases} 0 & \theta \in (d-\delta, d+\varepsilon) \\ 1 & \theta \notin (d-\delta \quad d+\varepsilon) \end{cases}$$

$$r(d|x) = E_{\theta|x}\{L(\theta, d)\} = Pr\{\theta \notin (d-\delta, d+\delta)|x\}$$

the Bayes Rule $d(x)$ is



$$d(x) = \underset{d}{\arg\min}\, Pr(d-\delta < \theta < d+\delta \mid x)$$

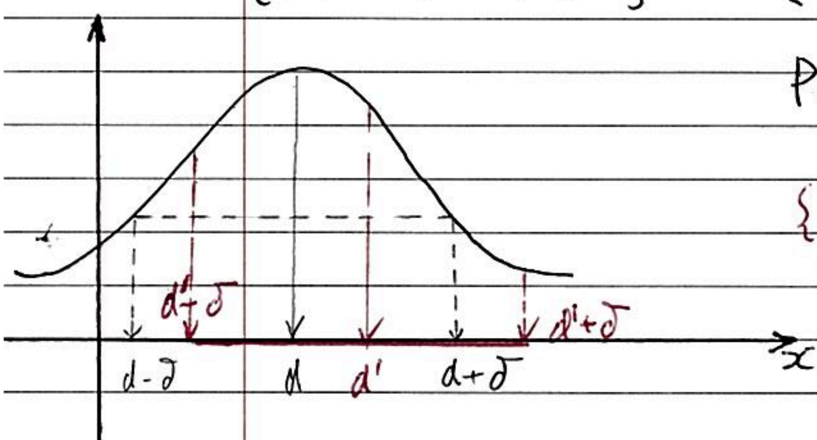
= highest Posterior Density

= HPD

⇒ A special case

If $\pi(\theta|x)$ is unimodel, the Bayes Rule interval has the form

$\{\theta \mid \pi(\theta|x) \geq c\}$ OR $\{d \mid \pi(d+\delta|x) = \pi(d-\delta|x)\}$



$P(\theta \in d' \pm \delta|x) > Pr(\theta \in d \pm \delta|x)$

{But it is higher d to evaluate $\pi(\theta|x)$, when $\pi(\theta)$ OR $P(x|\theta)$ does't have close form}

⇒ in practice, we prescribe $Pr(\theta \in I|x) = 1-\alpha$

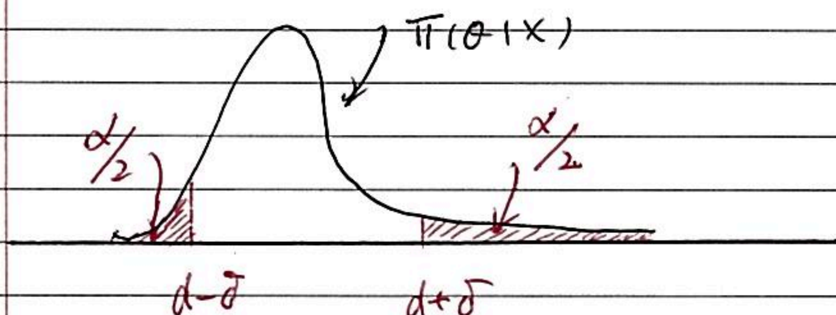
then we search I such that length(I) is minimized.

⇒ In practice, two alternatives:

a) $d(x)$ is set such that

$$Pr(\theta > d(x)|x) = Pr(\theta < d(x)|x)$$

Equil tailed interval



2) Normal interval $(\mu - Z_{\frac{\alpha}{2}} \cdot \sigma \quad \mu + Z_{\frac{\alpha}{2}} \cdot \sigma)$

where, $\mu = E(\theta | X)$ $\quad \sigma^2 = Var(\theta | x)$

Based on Bayesian Asymptotic

$\theta | x \sim N(\mu, \sigma^2)$

$\Rightarrow$ point estimate : how to Find the minimax Rule

$X | \theta \sim Bin.l(n, \theta)$

$\theta \sim Beta(a, b)$

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

$\theta | x \sim Beta(a+x, b+n-x)$

the Bayes Rule is

$$d(x) = E(\theta | x) = \frac{a+x}{a+b+n}$$

calculated the Risk function

If $R(\theta, d(x)) =$ constant for all $\theta \in (0, 1)$

By theorem 2.2, this $d(x)$ is minimax.

Let $c = a+b+n$

$$R(\theta, d) = E_X\{(\theta - d(x))^2\} = E_X\{(\theta - \frac{a+x}{a+b+n})^2\}$$

$$= E_X\{(\theta - \frac{a+x}{c})^2\} = \frac{1}{c^2} E_X\{(c\theta - a - x)^2\}$$

$$= \frac{1}{c^2} E_X\{(c\theta - a)^2 - 2(c\theta - a) \cdot x + x^2\}$$

$$= \frac{1}{c^2} \cdot (c\theta - a)^2 - \frac{2}{c^2}(c\theta - a) E_X(x) + \frac{1}{c^2} E_X(x^2)$$

Note that : $E(X) = n\theta$, $E(X^2) = (n\theta)^2 + n\theta(1-\theta)$

$= \frac{1}{c^2}\{(c\theta - a)^2 - 2(c\theta - a)\cdot n\theta + n^2\theta^2 + n\theta(1-\theta)\}$

To search a, b such that

$R(\theta, d) = \text{constant}$ for all θ

We write that

$$\begin{cases} n + 2na - 2ac = 0 & \text{coefficent of } \theta \\ -n + n^2 - 2nc + c^2 = 0 & \text{coefficent of } \theta^2 \end{cases}$$

$$\Rightarrow \begin{cases} a = \sqrt{n}/2 \\ b = \sqrt{n}/2 \end{cases}$$

the minimax rule is $d(x) = \dfrac{\sqrt{n}/2 + x}{n + \sqrt{n}}$

Note that : $d(x) \xrightarrow{\text{as } n \to \infty} \dfrac{\sqrt{n}/2 + n\theta}{n +} \to \theta$

« January 28, 2015 » STAT 846

$X \sim Bin(n, p) \Rightarrow f(x, p) = \binom{n}{x} p^x (1-p)^{n-x}$

$p \sim Beta(a, b)$

Bayes estimator: $\hat{p}_{BS} = \frac{x+a}{a+b+n}$ $\quad \hat{p} = \frac{X}{n}$.

So $\hat{p}_{BS}$ shrinks $\hat{p}$ to $\frac{a}{a+b+n}$

$\Rightarrow$ Shrinkage and Jame-stein estimation.

$X_i \mid \mu_i \sim N(\mu_i, 1)$ for $i = 1, 2, \cdots, p$

We want to estimate $\mu_1, \mu_2, \cdots, \mu_p$

A decision Rule $d = (d_1, d_2, \cdots d_p)$

Loss function:

$L(\mu, d) = \sum_{i=1}^{p} (\mu_i - d_i)^2 = \| \mu - d \|^2$

Let $X = (X_1, X_2, \cdots, X_p)^T$, $d = (d_1, d_2, \cdots, d_p)^T$

A class of d^a: $d^a(x) = \left(1 - \frac{a}{\|X\|^2}\right) \cdot X$

$\Rightarrow$ A special case $a = 0 \Rightarrow d^0(x) = X$

$\Rightarrow$ When $a > 0$, $d^0(x)$ closer to $(0, 0, \cdots, 0)^T$

$\Rightarrow$ We will show that, when $p \geq 3$, d^a dominates d^0.

~~the Risk function:~~ $R(\mu, d^0(x)) = E_X\{\| X - \mu \|^2\}$

$$= \sum_{i=1}^{p} E_{x_i}\{(x_i-\mu_i)^2\} = \sum_{i=1}^{p} 1 = p$$

$\Rightarrow$ Stein Lemma: if $x_i \sim N(\mu_i\ 1)$, then

$$E_x\{(x_i-\mu_i)h(x)\} = E_x\left(\frac{\partial h(x)}{\partial x_i}\right)$$

proof: suppose that $p=1$, Let φ be pdf of $N(0,1)$

$$\frac{\partial \varphi(x-\mu)}{\partial x} = -\varphi(x-\mu)\cdot(x-\mu)$$

$\left\{\text{Note that: since } \varphi(x-\mu) = \frac{1}{\sqrt{2\pi}}\cdot \exp^{-\frac{(x-\mu)^2}{2}}\right\}$

$$E_x\{(x-\mu)h(x)\} = \int (x-\mu)h(x)\cdot\varphi(x-\mu)\,dx$$

$$= -\int_{-\infty}^{\infty} h(x)\, d(\varphi(x-\mu))$$

$$= -\left\{h(x)\varphi(x-\mu)\Big|_{-\infty}^{\infty} - \int \varphi(x-\mu)\, dh(x)\right\}$$

$$= -h(x)\varphi(x-\mu)\Big|_{-\infty}^{\infty} + \int \varphi(x-\mu)\, dh(x)$$

$$= \int \varphi(x-\mu)\cdot\frac{\partial h(x)}{\partial x}\cdot dx$$

$$= E_x\left\{\frac{\partial h(x)}{\partial x}\right\}$$

therefore, $E_x\{(x_i-\mu_i)h(x)\} = E_x\left(\frac{\partial h(x)}{\partial x}\right)$

$\Rightarrow$ Risk function $R(\mu, d^a) = E_x(\|\mu - d^a(x)\|^2)$

$$= E_x\left\{\left(\left\|\mu - \left(1-\frac{a}{\|x\|^2}\right)\cdot x\right\|\right)^2\right\}$$

$$= E_x\left\{\left(\left\|(\mu-x) + \frac{a}{\|x\|^2}\right\|\right)^2\right\}$$

$$= E_X\{\|\mu - X\|^2) - 2a \cdot E_X\left\{\frac{X^T(X-\mu)}{\|X\|}\right\} + a^2 \cdot E_X\left\{\left(\frac{1}{\|X\|^2}\right)\right\}$$

$$\Bigg\{ \text{where } E_X\left(\frac{X^T(X-\mu)}{\|X\|^2}\right)$$

$$= \sum_{i=1}^{p} E_X\left(\frac{X_i(X_i-\mu_i)}{\sum_{j=1}^{p} X_j^2}\right) \quad . \quad h(X_i) = \frac{X_i}{\sum_{j=1}^{p} X_j^2}$$

$$= \sum_{i=1}^{p} E_X\left\{\frac{\partial}{\partial X_i} \cdot \left(\frac{X_i}{\sum_{j=1}^{p} X_j^2}\right)\right\} = \sum_{i=1}^{p} E_X\left\{\frac{\sum_{j=1}^{p} X_j^2 - 2X_i^2}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right\}$$

Note(1)

$$= (p-2) E_X\left(\frac{1}{\|X\|^2}\right) \Bigg\}$$

$$= p - 2a \cdot (p-2) E_X\left(\frac{1}{\|X\|^2}\right) + a^2 E_X\left(\frac{1}{\|X\|^2}\right)$$

$$= p - (2a(p-2) - a^2) E_X\left(\frac{1}{\|X\|^2}\right)$$

so, $R(\mu, d^a) < R(\mu, d^0) = p$

$\Leftrightarrow \quad 2a(p-2) - a^2 > 0$

$\Leftrightarrow \quad 2(p-2) > a$

when $p \geq 3$ there exist $a > 0$ such that

$a < 2(p-2)$

$\Leftrightarrow$ therefore d^0 is inadmissible

Note 1:

$$\sum_{i=1}^{p} E_x \left\{ \frac{\sum_{j=1}^{p} x_j^2 - 2x_i^2}{\left(\sum_{j=1}^{p} x_j^2\right)^2} \right\}$$

$$= E_x \left\{ \sum_{i=1}^{p} \frac{\|x\|^2 - 2x_i^2}{\|x\|^4} \right\}$$

$$= E_x \left\{ \frac{1}{\|x\|^2} \cdot \sum_{i=1}^{p} 1 - \frac{2}{\|x^4\|} \cdot \sum_{i=1}^{p} x_i^2 \right\}$$

$$= E_x \left(\frac{p}{\|x\|^2} - \frac{2\|x\|^2}{\|x\|^4} \right)$$

$$= E_x \left(\frac{p-2}{\|x\|^2} \right)$$

$$= (p-2) E_x \left(\frac{1}{\|x\|^2} \right)$$

《 January 30, 2015 》 STAT 846

=> Home Run Examples

	pre-season			Regular season						
	Y_i	n_i	X_i	N_i (AB)	$\hat{\mu}_i^{JS}$	μ_i	$\hat{\mu}_i = X_i$	HR_i	$\hat{HR}_i^{JS}$	$\hat{HR}_i$
Sosa	7	58	-6.56	509	-7.12	-6.18	-6.56	70	50	61
M. G	9	59	-5.90	643	-6.71	-7.06	-5.90	66	75	98

Note that : $\hat{HR}_i = (Y_i / n_i) \times N_i$

$61 = \frac{7}{58} \times 509 \qquad 98 = \frac{9}{59} \times 643$

$Y_i \mid n_i, P_i \sim Bin(n_i, P_i)$

$x_i = f_{n_i}\left(\frac{y_i}{n_i}\right)$, when $n_i = n$ $f_n\left(\frac{y}{n}\right) = \sqrt{n} \cdot \sin^{-1}\left(2\frac{y}{n} - 1\right)$

$SD(y/n) = \sqrt{P(1-P)/n}$, $SD(X_i) = \sigma$

New model : $x_i \mid \mu_i \sim N(\mu_i, \sigma^2)$; $\mu = f_{n_i}(P_i)$

We will then apply J-S to $X_1, \ldots, X_p$ to estimate $\mu_1, \ldots, \mu_p$, then we can estimate

$P_i = f_{n_i}^{-1}(\mu_i)$

$d^{p-2}(x) = \bar{X} + \left(1 - \frac{p-2}{V}\right)(X - \bar{X})$

$V = \sum_{i=1}^{p} (X_i - \bar{X})^2$, where, $X = (X_1, \cdots X_p)^T$

$\sum_{i=1}^{p} (\hat{\mu}_i - \mu_i)^2 = 19.68 \qquad \sum_{i=1}^{p} (\hat{\mu}^{JS} - \mu_i)^2 = 8.07$

$\hat{HR}_i^{JS} = f_{N_i}^{-1}(\hat{\mu}_i^{JS}) \cdot N_i \quad \approx \hat{P}^{JS}$

$\hat{HR}_i = \frac{Y_i}{n_i} \times N_i$

=> Remarks:

when $p=1$, and 2, $\bar{X}$ is admissible under squre Loss.

(*) => Review: $X \mid \mu \sim N(\mu, \sigma^2)$

$\mu \sim N(\mu_0\ \sigma_0^2)$

What is the marginal distribution of X

Method 1 => $f(x) = \int_{-\infty}^{\infty} f(x \mid \mu) \cdot \pi(\bar{u})\, d\mu$

Method 2 => $x = \mu + z$ where $\mu \sim N(\mu_0, \sigma_0^2)$

$z \sim N(0, \sigma^2)$

(*) => A general result: $x_i \overset{iid}{\sim} N(\mu_i\ \sigma_i^2)$,

then $\sum_{i=1}^{n} a_i x_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$, the

distribution of X is $N(\mu_0, \sigma_0^2 + \sigma^2)$

=> predictive distribution

Given $X_1, \cdots X_n$ we want to estimate x^* by

finding $f(x^* \mid x_1, \cdots, x_n) = \int f(x^* \mid \theta, x_1 \cdots x_n) f(\theta \mid x_1 \cdots x_n) d\theta$

This is a generalization of $f(x) = \int f(x \mid \theta) f(\theta) d\theta$

=> A special case $X_1, \cdots, X_n, X^* \mid \theta \overset{iid}{\sim} f(x \mid \theta)$

$f(x^* \mid x_1, \cdots, x_n) = \int f(x^* \mid \theta) \pi(\theta \mid x_1, \cdots, x_n)\, d\theta$

Example: $X_1, \cdots, X_n, X^* \mid \mu \overset{iid}{\sim} N(\mu, \sigma^2)$ $\quad \sigma^2$ known

$\mu \sim N(\mu_0, \sigma_0^2)$

(*) Question is how to find the predictive distributio

$f(X^* \mid X_1, \cdots, X_n)$

Sol: $\mu \mid X_1, \cdots, X_n \sim N(\mu_1, \sigma_1^2)$

where, $\sigma_1^2 = \left[\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right]^{-1}$ $\quad \mu_1 = \left[\frac{\mu_0}{\sigma_0^2} + \frac{n}{\sigma^2} \cdot \bar{X}\right] \times \sigma_1^2$

$f(x^* \mid x_1, \cdots, x_n) = \int f(x^* \mid \mu) \pi(\mu \mid x_1, \cdots, x_n) d\mu$

$$\left.\begin{array}{l} x^* \mid \mu \sim N(\mu, \sigma^2) \\ \mu \mid x_1, \cdots x_n \sim N(\mu_1\ \sigma_1^2) \end{array}\right\} \Rightarrow x^* \mid x_1, \cdots x_n \sim N(\mu_1, \sigma_1^2 + \sigma^2)$$

$\Rightarrow$ Emprical Bayes:

$\Rightarrow$ Review: 1) $x_i \mid \mu_i \sim N(\mu_i, 1)$

$\mu_i \sim N(0\ \ \sigma^2)$

the marginal distribution of x_i

$x_i \sim N(\mu_i\ \ 1 + \sigma^2)$

2) $X \sim \text{Gamma}(\alpha, \lambda) \Rightarrow E(X^k) = ?$

proof: $f(x) = \frac{1}{\Gamma(\alpha)} \cdot \lambda^\alpha \cdot x^{\alpha-1} e^{-\lambda x}$ for $x > 0$

Let $x = z/\lambda$ where, $Z \sim \text{Gamma}(\alpha, 1)$

$Z \sim f(z) = \frac{1}{\Gamma(\alpha)} z^{\alpha-1} \cdot e^{-z}$ for $z > 0$

$$E(Z^k) = \int_0^\infty Z^k \cdot \frac{1}{\Gamma(\alpha)} Z^{\alpha-1} \cdot e^{-z} \, dz$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty z^{\alpha+k-1} \cdot e^{-z} \, dz$$

$$= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}$$

« Feb, 02, 2015 » STAT 846

$x \sim \text{Gamma}(\alpha, \lambda)$; λ is rate.

$$E(x^k) = \frac{\Gamma(\alpha+k)}{\lambda^k \Gamma(\alpha)}$$

$$x \sim \chi_p^2 = \text{Gamma}\left(\frac{p}{2}, \lambda = \frac{1}{2}\right)$$

χ_p^2 is distribution of $x = \sum_{i=1}^{p} z_i^2$

where $z_i \sim N(0, 1)$

$$E(x^k) = \frac{\Gamma(\frac{p}{2}+k)}{(\frac{1}{2})^k \Gamma(\frac{p}{2})} = \frac{\Gamma(\frac{p}{2}+k)}{\Gamma(\frac{p}{2})} \cdot 2^k$$

$$\Rightarrow E(x^{-1}) = \left\{\Gamma\left(\frac{p}{2}-1\right) / \Gamma\left(\frac{p}{2}\right)\right\} \cdot \frac{1}{2}$$

$$= \frac{\Gamma(\frac{p}{2}-1)}{(\frac{p}{2}-1)\Gamma(\frac{p}{2}-1)} \cdot \frac{1}{2} = \frac{1}{p-2} \quad (\text{for } p > 2)$$

$\Rightarrow$ Emparical Bayes:

$$\text{Model: } \begin{cases} x_i \mid \mu_i \sim N(\mu_i, 1) \\ \mu_i \sim N(0, \tau^2) \end{cases} \quad \text{for } i = 1, 2, \cdots, p$$

Suppose we know τ^2, Bayes estimator for

$\mu = (\mu_1, \cdots, \mu_p)^T$, $\delta^\tau(x) = E_{\mu|x}(\mu \mid x)$

$$\mu_i \mid x_i \sim N\left(\frac{\tau^2}{1+\tau^2} x_i, \frac{\tau^2}{1+\tau^2}\right)$$

$$\delta^\tau(x) = E_{\mu|x}(\mu \mid x) = \frac{\tau^2}{1+\tau^2} \cdot x$$

$\Rightarrow$ Bayes Risk: $r(\pi^\tau, \delta^\tau) = E_{\mu \sim \pi^\tau}\{R(\mu, \delta^\tau)\}$

$= E_\mu E_{x|\mu}\{\|\mu - \delta^\tau(x)\|^2 \mid \mu\}$

$= E_x E_{\mu|x}\{\|\mu - \delta^\tau(x)\|^2 \mid x\}$

$= E_x E_{\mu|x}\{\sum_{i=1}^{p}(\mu_i - \delta^\tau(x)_i)^2 \mid x\}$

$= E_x\{\sum_{i=1}^{p} E_{\mu|x}(\mu_i - \delta^\tau(x_i))^2 \mid x\}$

$= E_x\{\sum_{i=1}^{p} Var(\mu_i \mid x_i)\} = E_x\{\sum_{i=1}^{p} \frac{\tau^2}{1+\tau^2}\} = p \cdot \frac{\tau^2}{1+\tau^2}$

$\Rightarrow$ Emprical Bayes estimate:

We want to replace τ^2 with a statistic (of X) using marginal distribution of X

$\left.\begin{array}{l} x_i \mid \mu_i \sim N(\mu_i, 1) \\ \mu_i \sim N(0, \tau^2) \end{array}\right\} \Rightarrow x_i \sim N(0, 1+\tau^2)$

If $\frac{x_i}{\sqrt{1+\tau^2}} \sim N(0,1) \Rightarrow \sum_{i=1}^{p} \frac{x_i^2}{1+\tau^2} \sim \chi_p^2$

$E_x\left(\frac{1+\tau^2}{\|x\|^2}\right) = \frac{1}{p-2}$

$\Rightarrow E_x\left\{\frac{1}{\|x\|^2}\right\} = \frac{1}{(p-2)(1+\tau^2)}$

$E_x\left(1 - \frac{p-2}{\|x\|^2}\right) = 1 - \frac{1}{1+\tau^2} = \frac{\tau^2}{1+\tau^2}$

We can replace $\frac{\tau^2}{1+\tau^2}$ by $\left(1 - \frac{p-2}{\|x\|^2}\right)$

A note :

Another approach to estimate σ^2

$$x_1, \dots, x_p \overset{iid}{\sim} N(0, 1+\sigma^2)$$

$$\hat{\sigma}^2 = \frac{\|X\|^2}{p-2} - 1, \quad \widehat{1+\sigma^2} = \sum_{i=1}^{p} X_i^2 \Big/ p-2$$

$$\Rightarrow \frac{\hat{\sigma}^2}{1+\hat{\sigma}^2} = 1 - \frac{p-2}{\|X\|}$$

$\Rightarrow$ An emprical Bayes estimate replace $\frac{\sigma^2}{1+\sigma^2}$ in

$$\delta^\sigma(x) = \frac{\sigma^2}{1+\sigma^2} X \quad \text{by} \quad 1 - \frac{p-2}{\|X\|^2}$$

$$d^{p-2}(x) = \left(1 - \frac{p-2}{\|X\|^2}\right) X$$

to find the Bayes Risk of d^{p-2}

$$r(\pi^\sigma, d^{p-2}(x)) = E_\mu E_{x|\mu}\left(\|\mu - d^{p-2}(x)\|^2 \mid x\right)$$

$$= E_x E_{\mu|x}\left(\left\|\mu - \left(1 - \frac{p-2}{\|X\|^2}\right) X\right\|^2 \mid x\right)$$

$$= E_x \left\{ \left(p - \frac{p-2}{\|X\|^2}\right) \right\}$$

$$= p - (p-2)^2 E_x\left(\frac{1}{\|X\|^2}\right)$$

$$= p - (p-2) \times \frac{1}{1+\sigma^2}$$

$$= p - \frac{p-2}{1+\sigma^2} = r(\pi^\sigma, \delta^\sigma(x)) + \frac{2}{1+\sigma^2}$$

$$r(\pi^\sigma, \delta^\sigma(x)) = p\sigma^2 \Big/ 1+\sigma^2$$

« Feb. 04. 2015 » STAT 846

⇒ choice of prior {how to choice the prior}

Methods: 1) Empiracal Bayes :

$X|\theta \sim f(X|\theta)$

$\theta|\tau \sim f(\theta|\tau)$

$f(x|\tau) = \int f(x|\theta) f(\theta|\tau) d\theta$

we want to find a $\hat{\tau}$ from $f(x|\tau)$

2) physical method by Bayes:

$x_1, \cdots, x_n | \theta \sim Bin(n; \theta)$ $\theta \sim unif([0,1])$

An example : θ : recombination Rate

3) Non-informative prior by Jeffrey and Laplace

Roughly, Bayes inference = MLE

Example: $x_1, \cdots, x_n | \mu \sim N(\mu, \sigma^2)$ $\mu \sim N(\mu_0, \sigma_0^2)$

$$\mu | x_i \sim N\left\{\left(\frac{\mu_0}{\sigma_0^2} + \frac{n}{\sigma^2}\cdot\bar{X}\right) \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right), *\right\}$$

if we set $\sigma_0^2 = +\infty$ $E(\mu|X) = \bar{X}$

Jeffrey prior : $\pi(\theta) \propto \sqrt{I(\theta)}$

$$I(\theta) = E_X\left\{\frac{\partial^2}{\partial\theta^2}(\log f(x;\theta))\right\}$$

4) personal probability (Subjective)

- $\pi(\theta)$ reflects a person's judgement on θ

θ = Average heights of all U of S students.

θ is unknow, can not be replicated.

θ is a R.V only because θ is unknow.

$\theta \sim N(170, ...^2)$

- $\pi(\theta)$ is information, external to data subjective to persons (not all we the same smart)

- inference results can be still judged with frequentist criterion

5) choose convenient prior, such as conditional conjuge prior.

6) hierachical Modelling

$$X|\theta \sim f(x|\theta), \quad \theta_1, \cdots, \theta_n | \tau \sim f(\theta|\tau)$$

$$\tau \sim \pi(\tau)$$

Hierachical Modelling example:

$$x_i | \mu_i \sim N(\mu_i, 1) \quad \left\{ Y_{i1}, \cdots Y_{in} | \mu_i \sim N(\mu_i, \sigma_i^2) \quad x_i = \bar{Y}_i \right\}$$

$$\mu_1, \cdots \mu_n | \tau^2 \sim N(\theta, \tau^2)$$

$$\tau^2 \sim \text{inv-Gamma}(\alpha^*, \beta^*) \text{ OR}$$

$$f(\tau^2) d\tau^2 \propto (\tau^2)^{-(\alpha^*+1)} e^{-\frac{\beta^*}{\tau^2}} d\tau^2$$

$$\theta \sim \text{Unif}(-\infty, +\infty) \longrightarrow \text{Non-informative}$$

$$x \sim \text{Gamma}(\alpha^*, \beta^*)$$

$$f_x(x) dx \propto x^{\alpha^*-1} \cdot e^{-\beta^* \cdot x} dx$$

Let $\tau^2 = 1/x \sim \text{inverse-Gamma}(\alpha^*, \beta^*)$

$$f_{\tau^2}(\tau^2) d\tau^2 \propto f_x\left(\frac{1}{\tau^2}\right) d\frac{1}{\tau^2}$$

$$\propto (\tau^2)^{-(\alpha^*-1)} e^{-\frac{\beta^*}{\tau^2}} (\tau^2)^{-2} d\tau^2$$

$$= (\tau^2)^{-(\alpha^*+1)} e^{-\frac{\beta^*}{\tau^2}} d\tau^2$$

Textbook : $\tau = \sqrt{\tau^2}$ $\quad f_\tau(\tau)\, d\tau = f_{\tau^2}(\tau^2)\, d\tau^2$

$$= f_{\tau^2}(\tau^2) \cdot 2\tau\, d\tau = (\tau^2)^{-(\alpha^*+1)} \tau \cdot e^{-\frac{\beta^*}{\tau^2}} d\tau$$

$$= (\tau^2)^{-(\alpha^*+1)} \cdot (\tau^2)^{\frac{1}{2}} e^{-\frac{\beta^*}{\tau^2}} d\tau$$

$$= (\tau^2)^{-\alpha^*-\frac{1}{2}} e^{-\frac{\beta^*}{\tau^2}} d\tau = (\tau^2)^{-(\alpha^*+\frac{1}{2})} e^{-\frac{\beta^*}{\tau^2}} d\tau$$

$\Rightarrow$ Joint posterior of μ, θ, τ^2

$$f(\mu_1, \cdots, \mu_n, \theta, \tau^2 \mid x_1, x_2, \cdots, x_n)$$

$$\propto \prod_{i=1}^n f(x_i \mid \mu_i) \cdot \prod_{i=1} f(\mu_i \mid \theta, \tau^2)$$

$$\cancel{(\pi_\theta(\theta) \cdot \pi_{\tau^2}(\tau^2) \times d\theta\, d\mu_i\, d\tau^2)}$$

$$\propto e^{-\sum_{i=1}^n \frac{(x_i-\mu_i)^2}{2}} \times (\tau^2)^{-\frac{n}{2}} \cdot e^{-\sum_{i=1}^n \frac{(\mu_i-\theta)^2}{2\tau^2}}$$

$$(\tau^2)^{-(\alpha^*+1)} \cdot e^{-\frac{\beta^*}{\tau^2}} d\theta\, d\mu_i\, d\tau^2$$

We will derive full conditionals :

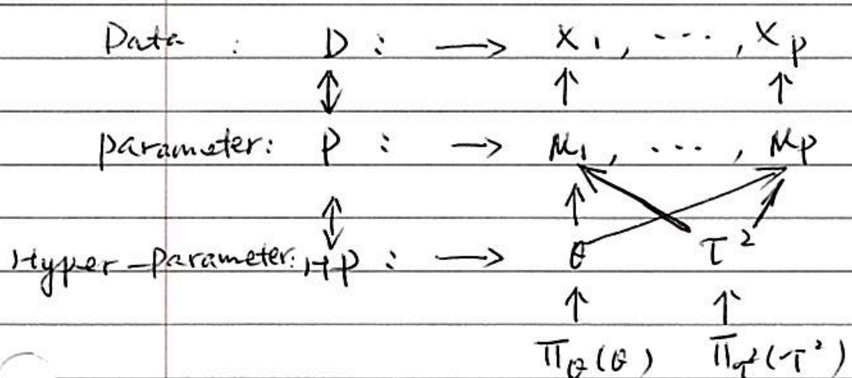
$$\begin{cases} f(\mu_i \mid \mu_{-i};\ \theta, \tau^2, x_1, \cdots, x_n) \\ f(\theta \mid \mu_{1:n}, \tau^2, x_1, \cdots, x_n) \\ f(\tau^2 \mid \mu_{1:n}, \theta, x_1, \cdots, x_n) \end{cases}$$

« Feb. 06, 2015 » STAT 846

⇒ Example :

$$\begin{aligned} x_i \mid \mu_i &\sim N(\mu_i, 1) \quad \text{for } i = 1, \cdots, p \\ \mu_i \mid \theta &\sim N(\theta, \tau^2) \\ \theta &\sim N(\theta_0, \sigma_0^2) \\ \tau^2 &\sim \text{inv-Gamma}(\alpha^*, \beta^*) \end{aligned}$$

$$f(\mu_1, \cdots, \mu_p, \theta, \tau^2) \propto \prod_{i=1}^{p} f(x_i \mid \mu_i) \cdot \prod_{i=1}^{p} f(\mu_i \mid \theta, \tau^2) \cdot \pi_\theta(\theta) \cdot \pi_{\tau^2}(\tau^2)$$



⇒ Full Conditionals

1) $\mu_i \mid x_i, \theta, \tau^2$

$$\begin{cases} x_i \mid \mu_i \sim N(\mu_i, 1) \\ \mu_i \mid \theta, \tau^2 \sim N(\theta, \tau^2) \end{cases} \Rightarrow$$

$$\mu_i \mid x_i \sim N\left(\frac{\frac{\theta}{\tau^2} + \frac{x_i}{1}}{\frac{1}{\tau^2} + \frac{1}{1}}, \frac{1}{\frac{1}{\tau^2} + 1} \right) = N\left(\theta + \tau_2^2 (x_i - \theta), \tau_2^2\right)$$

where, $\tau_2^2 = \frac{\tau^2}{1+\tau^2}$

2) $\theta \mid \tau^2, \mu_i, x_i$

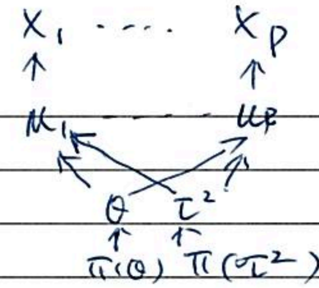
$$\mu_1, \cdots, \mu_p \mid \theta, \tau^2 \sim N(\theta, \tau^2)$$

$$\theta \sim N(\theta_0, \sigma_0^2)$$

$$\theta \mid \tau^2, \mu_i, x_i \sim N\left(\frac{\frac{\theta_0}{\sigma_0^2} + \frac{p}{\tau^2} \cdot \bar{\mu}}{\frac{1}{\sigma_0^2} + \frac{p}{\tau^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{p}{\tau^2}} \right) = N\left(\bar{\mu}, \frac{\tau^2}{p}\right)$$

[illegible]

<< Feb. 09. 2015 >> STAT 846



⇒ Baseball example:

Let $X = (x_1, \cdots, x_p)$

$$f(\mu_1, \cdots \mu_p, \theta, \tau^2 \mid x) \propto \prod_{i=1}^{p} f(x_i \mid \mu_i) \cdot \prod_{i=1}^{p} f(\mu_i \mid \theta, \tau^2) \cdot \pi_0(\theta) \cdot \pi_{\tau^2}(\tau^2)$$

After we have samples:

$$\{ \mu_1^{(i)}, \cdots \mu_p^{(i)}, (\tau^2)^{(i)}, \theta^{(i)} \mid i = 1, 2, \cdots, N \}$$

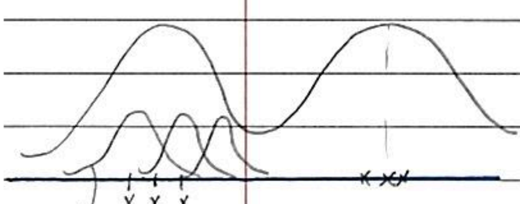
To look at $f(\mu_i \mid X)$ marginal posterior of μ_i

1) look at density $\hat{f}(\mu_j)$ of $\{ \mu_j^{(i)} \mid i = 1, 2, \cdots, N \}$

2) $f(\mu_j \mid X) = \int_N f(\mu_j \mid X, \theta\ \tau^2) f(\theta, \tau^2)\, d\theta\, d\tau^2$

(where $f(\mu_i \mid X, \theta, \tau^2) \sim N(\theta + \tau_2(x_i - \theta), \tau_2)$

$$\approx \sum_{i=1}^{N} f(\mu_j \mid X, \theta^{(i)}, (\tau^2)^{(i)})$$



Kernal Density

$\varphi\left(\frac{x_i - \bar{x}}{h}\right) \cdot \frac{1}{h}$

$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \varphi\left(\frac{x_i - \bar{x}}{h}\right) \cdot \frac{1}{h}$

This is an application of Rao-Blackwell formula

$$E\{ t(\mu, \theta) \} = E_\theta \{ E_\mu (t(\mu, \theta) \mid \theta) \}$$

make a Note in here:

$$f(\mu_j \mid X, \theta, \tau^2) = \int f(\mu_j\ \mu_{-j} \mid X, \theta, \tau^2)\, d\mu$$

⇒ Empirical Bayes estimator of $f(\mu_j \mid X)$

Suppose, we have an estimator of θ, τ^2 by

looking $f(x|\theta, \tau^2)$, denoted by $\hat\theta$, $\hat\tau_2 \left(= \frac{\hat\tau^2}{1+\hat\tau^2}\right)$

$f(\mu_j|x) \approx f(\mu_j|x, \hat\theta, \hat\tau^2) = N(\hat\theta + \hat\tau_2(\hat\theta - x_i), \hat\tau_2)$

in particular : J-S : $\hat\theta = \bar x$ $\quad \hat\tau_2 = 1 - \frac{p-3}{V}$

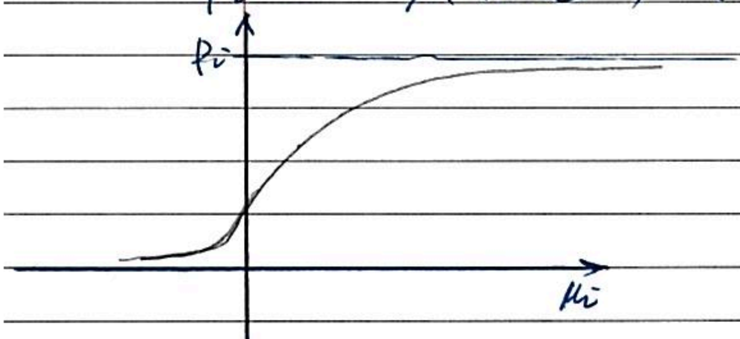
where, $V = \sum_{i=1}^{p}(x_i - \bar x)^2$

$\Rightarrow$ predictive distribution (Another Method.)

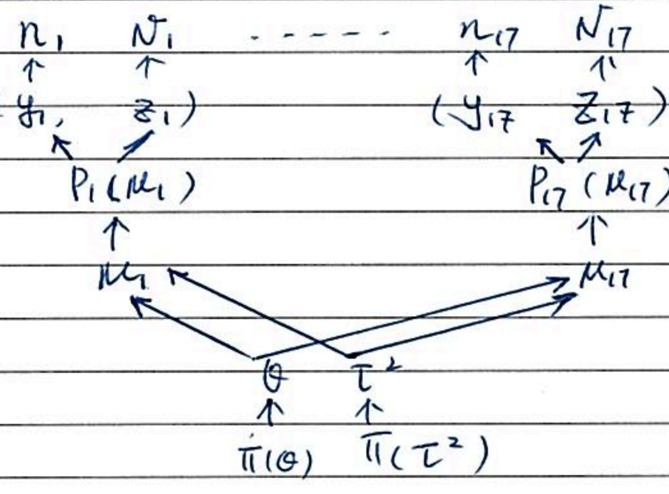
Example : $y_i | n_i \;\; p_i \sim Bin(n_i, p_i)$
$\uparrow$
pre-season

$z_i | N_i, p_i \sim Bin(N_i, p_i)$
$\uparrow$
~~full season~~

$p_i = e^{\mu_i}/(1+e^{\mu_i})$ (inv-logistic transformation)



$\mu_1, \dots, \mu_p | \theta, \tau^2 \sim N(\theta, \tau^2)$
$\tau^2 \sim$ inv-Gamma (α^*, β^*)
$\theta \sim N(0, \sigma_0^2)$



Posterior of μ_i, θ, τ^2

$f(\mu_1, \cdots \mu_{17}, \theta, \tau^2) \propto \prod_{i=1}^{17} f(y_i|\mu_i) \cdot \prod_{i=1}^{17} f(\mu_i|\theta, \tau^2) \cdot \pi(\theta), \pi(\tau^2)$

$$f(y_i \mid \mu_i) = \binom{n_i}{y_i}\left(\frac{e^{\mu_i}}{1+e^{\mu_i}}\right)^{y_i} \cdot \left(1 - \frac{e^{\mu_i}}{1+e^{\mu_i}}\right)^{n_i - y_i}$$

We don't have close form for $f(\mu_i \mid y_i, \theta, \tau^2)$

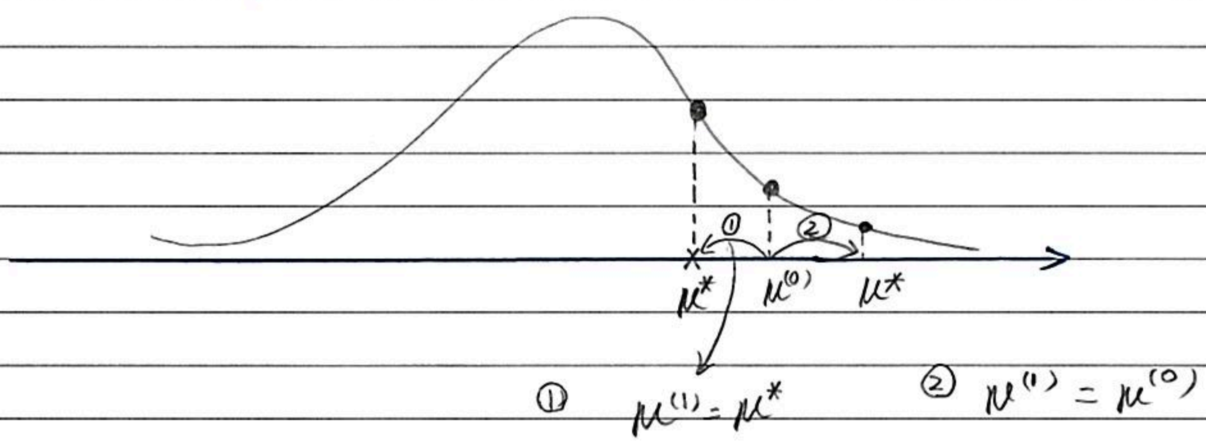
$\Rightarrow$ Gibbs sampling

$$\begin{cases} 1)\ \mu_i \mid y_i, \theta, \tau^2 \\ 2)\ \theta \mid \mu_1, \cdots \mu_{17}, \tau^2, y_1, \cdots y_{17} \\ 3)\ \tau^2 \mid \theta, \mu_1, \cdots \mu_{17}, y_1, \cdots y_{17} \end{cases}$$

$\Rightarrow$ Metropolis – Hasting

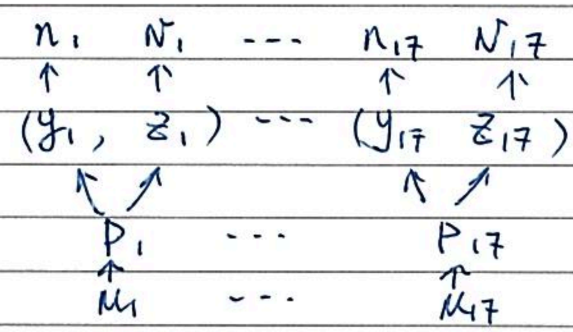
$\mu \sim \pi(\mu)$, repeat N times

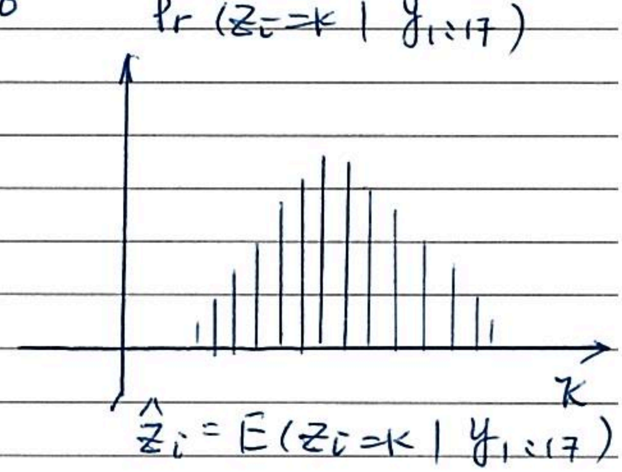
$$\begin{cases} \text{Starting from } \mu^{(0)} \\ 1)\ \text{propose } \mu^* \text{ from } P \\ 2)\ \text{Draw } u \sim \text{unif}([0,1]) \\ 3)\ \text{if } u < \min\left\{1,\ \pi(\mu^*) P(\mu^* \mid \mu^{(0)}) / \pi(\mu^{(0)}) \pi(\mu^{(0)} \mid \mu^*)\right. \\ \quad \text{Accept. } \mu^{(1)} = \mu^* \\ \quad \text{o/w } \mu^{(1)} = \mu^{(0)} \end{cases}$$



<< Feb. 11, 2015 >> STAT 846

Example:





$y_i \mid \mu_i \sim Bin(n_i, p_i)$

$z_i \mid \mu_i \sim Bin(N_i, p_i)$

n_i, N_i are co-variates

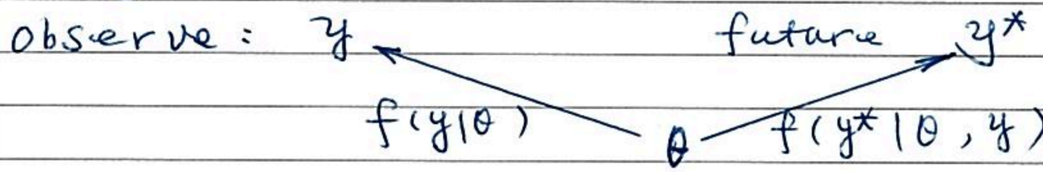
$\mu_1, \cdots, \mu_{17} \mid \theta, \tau^2 \sim N(\theta, \tau^2)$

$\theta \sim \pi_\theta$, $\tau^2 \sim \pi_{\tau^2}$

We want to find $\hat{Pr}(z_i = k \mid y_1, \cdots y_{17})$ for $i = 1, \cdots 17$,

prediction PMF of $z_i \mid y_{1:17}$,

=> in general term:



=> Joint Model for (θ, y, y^*):

$$f(\theta, y, y^*) = \pi(\theta) f(y \mid \theta) \cdot f(y^* \mid \theta, y)$$

"Our Goal to find the $f(y^* \mid y)$"

=> Joint posterior of $\theta, y^* \mid y$

$$f(\theta, y^* \mid y) = \frac{f(\theta, y, y^*)}{\iint f(\theta, y, y^*)\, dy^*\, d\theta}$$

$$= \frac{\pi(\theta)\, f(y|\theta)\, f(y^*|\theta, y)}{\left(\int f(y^*|y,\theta)\, dy^*\right)\left(\int \pi(\theta)\, f(y|\theta)\, d\theta\right)}$$

(Note: $\int f(y^* | y, \theta)\, dy^* = 1$)

$$= \frac{\pi(\theta) \cdot f(y|\theta)}{\int \pi(\theta) f(y|\theta)\, d\theta} \cdot f(y^* | \theta, y)$$

$$= f(\theta | y) \cdot f(y^* | \theta, y)$$

(Note that: if omitting $|y$, $f(\theta, y^*) = f(\theta) \cdot f(y^*|\theta)$)

$$\Rightarrow f(y^* | y) = \int f(\theta, y^* | y)\, d\theta$$

$$= \int f(\theta|y) \cdot f(y^* | \theta, y)\, d\theta$$

$$= E_{\theta|y}\{ f(y^* | \theta, y) \}$$

$\Rightarrow$ Two approaches to finding $f(y^* | y)$

Method 1:

$$f(y^* = k | y) = E_{\theta|y}\{ f(y^* = k | \theta, y) \}$$

Given samples of $\theta^{(i)} \sim f(\theta | y)$

$$\hat{f}(y^* = k | y) = \sum_{i=1}^{N} f(y^* = k | \theta^{(i)}, y) / N$$

Method 2: $f(y^* | y)$ is marginal of $f(\theta, y^* | y)$

for discrete y^*.

$$f(y^* = k | y) = E_{y^*, \theta | y}\{ I(y^* = k) \}$$

Given samples of $(y^{*,(i)}, \theta^{(i)}) \sim f(\theta, y^* | y)$

$$\hat{f}(y^* = k \mid y) = \sum_{i=1}^{N} I(y^{*,i} = k) \,/\, N$$

If y^* is continuous density, appy kernal estimate to $y^{*,1}, \dots, y^{*,N}$

⇒ Re-Mark:

Method 1 : is Rao-Blackwellization of Method 2.

$$E_X(t(x)) = E_Y(\underbrace{E_X(t(x) \mid Y)}_{\tilde{t}(y)}) = E_Y(\tilde{t}(Y))$$

⇒ Note that $Var(t(x)) > Var(\tilde{t}(Y))$

⇒ Back to example:

Method 1 : Draw $\theta^{(i)} \sim f(\theta \mid y_1, \dots, y_{17})$

OR. Draw $\theta^{(i)}, z_j^{(i)} \sim f(\theta, z_j \mid y_{1:17})$

then discard $z_j^{(i)}$

$$\hat{f}(z_j = k \mid y_{1:17}) \sim \frac{\sum_{i=1}^{N} dbino(k; N_j, P_j^{(i)})}{N}$$

Method 2 : Draw $\theta^{(i)}, z_j^{(i)} \sim f(\theta, z_j \mid y_{1:17})$

using $z_j^{(1)} \cdots z_j^{(N)}$ only $(z_j^{(i)} \sim f(z_j \mid y_{1:17}))$

$$\hat{f}(z_j = k \mid y_{1:17}) = \frac{\sum_{i=1}^{N} I(z_j^{(i)} = k)}{N}$$

$\Rightarrow$ Method 1 : $\displaystyle Var(\bar{X}) = \frac{Var(x_i)}{N}$

Method 2 : $\displaystyle Var(\bar{Y}) = \frac{Var(y_i)}{N}$

<< Feb 13, 2015 >> STAT 846

Chapter 4: Hypothesis Testing

=> Hypothesis testing (Model comparison)

=> Formulation : We want to know that

$$H_0 : \theta \in \Theta_0 \quad Vs \quad H_1 : \theta \in \Theta_1$$

H_0 ↓ (Null) $\qquad$ H_1 ↓ (alternative)

=> Simple hypothesis

θ_0 has a single value

=> Composite hypothesis :

θ_0 has > 1 value

=> example :

$X_1, \cdots, X_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

If σ^2 is known, $H_0 : \mu = \mu_0$ is simple

If σ^2 is un-known, $H_0 : \mu = \mu_0$ is composite

=> size of a test

$Pr(\text{"Reject } H_0\text{"}(X) \mid \theta) \leq \alpha$ for all $\theta \in \Theta_0$

then α is called the size of "Reject H_0" (X)

$$\alpha = \sup_{\theta} Pr(\text{"Reject } H_0 \mid \theta\text{"})$$

$\Rightarrow$ Critical Region based on $t(x)$

Rejected H_0 : if $t(x) \in C_\alpha$

C_α is called Critical Region

$\Rightarrow$ Test function:

$$\phi(x) = \begin{cases} 1 & \text{(Rejected) if } t(x) \in C_\alpha \\ 0 & \text{o/w} \end{cases}$$

$\Rightarrow$ example:

$$X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

$H_0 : \mu = \mu_0$ VS H_1 $\mu > \mu_0$

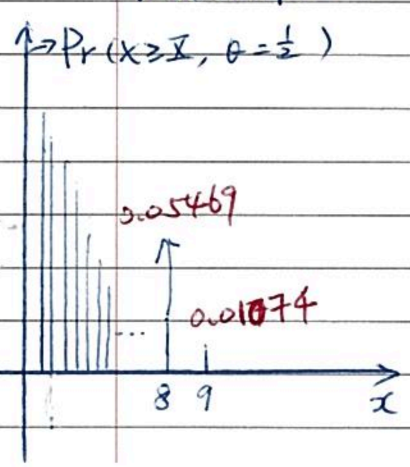
$$t(x) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

test. function in case:

$$\phi(x) = \begin{cases} 1, & \text{if } t(x) > t_{n-1, \alpha} \\ 0 & \text{o/w} \end{cases}$$

$\Rightarrow$ Randomised test function:

Example. $X \sim Bin(n=10, \theta)$



$H_0 : \theta = \frac{1}{2}$; VS, $H_1 : \theta > \frac{1}{2}$

$t(x) = X$

Critical Region $x \geq K_\alpha$

Given $\alpha = 5\%$ we can't find

the integer K_α, s.t $P(X \geq K_\alpha \mid \theta = \frac{1}{2}) = 0.05$

Randomised test function:

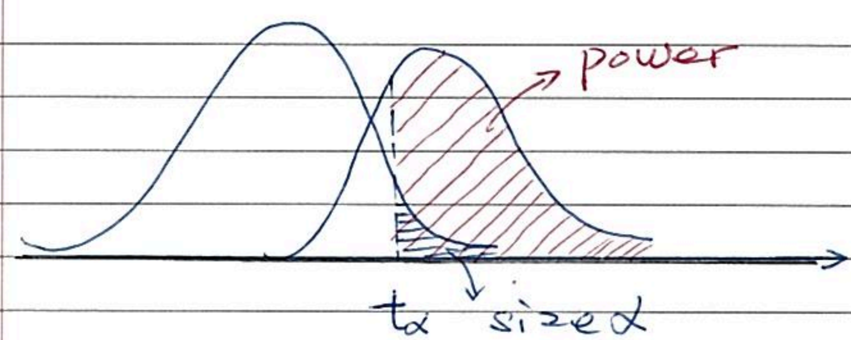
$$\phi(x) = \begin{cases} 1 & \text{if } x \geq 9 \\ \frac{67}{75} & \text{if } x = 8 \text{ (Reject } H_0 \text{ with Pr.} = \frac{67}{75}) \\ 0 & \text{if } x \leq 7 \end{cases}$$

$E(\phi(x)) = P(\text{Reject } H_0)$

$= Pr(X \geq 9) \times 1 + Pr(X=8) \times Pr(\text{Reject } H_0, X=8)$

$= 0.01074 \times 1 + (0.05469 - 0.01074) \times \frac{67}{75} = 5\%$

$\Rightarrow$ Power function:



$W(\theta) = Pr(\text{Reject } H_0 \mid \theta)$ for $\theta \in H_1$

$= E(\phi(x) \mid \theta)$, for $\theta \in H_1$

We have many way tests but we need to the best way

Ideally, a test function $\phi(x)$ is good if

(1) $E\{\phi_0(x) \mid \theta\} \leq \alpha$ for $\theta \in H_0$

(2) $E\{\phi_0(x) \mid \theta\} \geq E\{\phi(x) \mid \theta\}$

For each $\theta \in H_1$ and for all $\phi_{(x)}$ with size α

Within this framework, we have

(1) Simple H_0 $\leftrightarrow$ simple H_1

Neymon-Pearson formed optimal $\phi(x)$:

$$\phi(x) = \begin{cases} 1 & \text{if } \Lambda(x) = f(x \mid \theta \in H_1)/f(x \mid \theta \in H_0) > k_\alpha \\ r & \text{if } \Lambda(x) = k_\alpha \\ 0 & \text{if } \Lambda(x) < k_\alpha \end{cases}$$

(2) simple H_0 VS composite H_1

but $\Lambda(x)$ is monotone, we can have

optimal $\phi(x)$ for all $\theta \in H_1$

uniformly most powerful test

(3) Composite H_0 VS Composite H_1

we don't have optimal test

1.

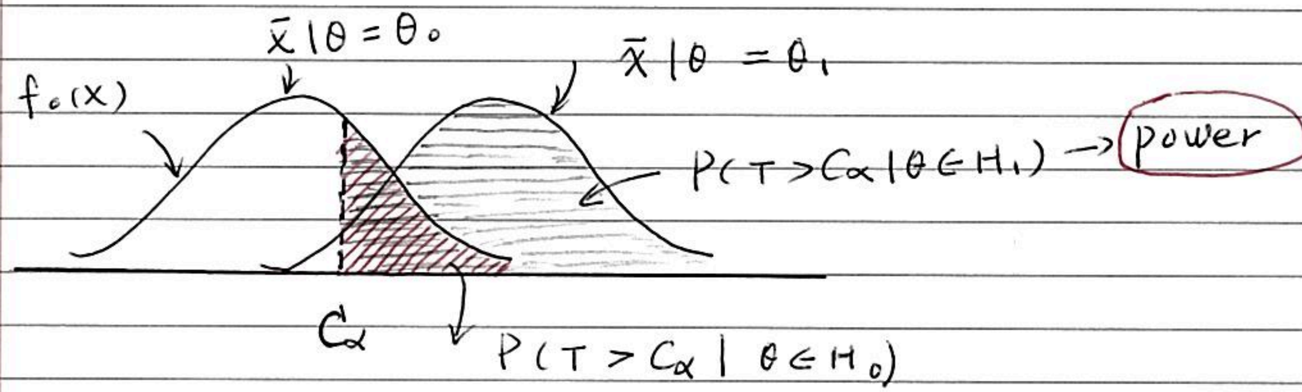
<< Feb. 23, 2015 >> STAT: 846

⇒ Test Function:

$$\phi(x) = \begin{cases} 1 & \text{if } x \in C_\alpha \\ \lambda(x) & \text{if } x = x_0 \\ 0 & \text{if } x \notin C_\alpha \cup \{x = x_0\} \end{cases}$$

⇒ size: $\sup_{\theta \in \Theta_0} E(\phi(x) \mid \theta)$

⇒ power: $W(\theta) = Pr(\phi(x) = 1 \mid \theta) = E(\phi(x) \mid \theta)$ for $\theta \in \Theta_1$



⇒ Neyman – Pearson theorem

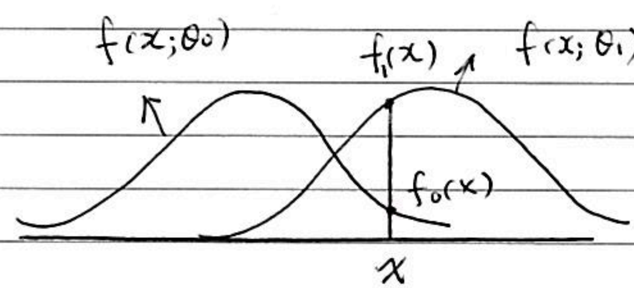
$H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta_1$

(LR) ⇒ Likelihood Ratio: Define the likelihood Ratio $\Lambda(x)$ by

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)}$$

pdf of X: $f_1(x) = f(x; \theta_1)$

$f_0(x) = f(x; \theta_0)$



$\Rightarrow$ Likelihood Ratio Test (LRT)

The (Randomised) test with the test function

$\phi_0(x)$ is said to be a likelihood ratio test

if it is of the form

$$\phi_0(x) = \begin{cases} 1 & \text{if } \Lambda(x) = \frac{f_1(x)}{f_0(x)} > K \\ \gamma(x) & \text{if } \Lambda(x) = f_1(x)/f_0(x) = K. \\ 0 & \text{if } \Lambda(x) = f_1(x)/f_0(x) < K. \end{cases}$$

$\Rightarrow$ Theorem 4.1 (Neyman – Pearson)

(最优性)

(a) (optimality). For any K and $\gamma(x)$, the test ϕ_0 has maximum power among all tests whose sizes are no greater than the size of $\phi_0(x)$

proof:

Let ϕ be any test for which $E_{\theta_0}\phi(x) \leq E_{\theta_0}\phi_0(x)$

Define $U(x) = \{\phi_0(x) - \phi(x)\}\{f_1(x) - K \cdot f_0(x)\}$,

When, $f_1(x) - K \cdot f_0(x) > 0$, we have $\phi_0(x) = 1$

So, $U(x) \geq 0$.

when $f_1(x) - K f_0(x) < 0$, we have $\phi_0(x) = 0$

So, $U(x) \geq 0$

When $f_1(x) - K f_0(x) = 0$, of course $U(x) = 0$

Thus $U(x) \geq 0$ for all x. Hence

$0 \le \int V(x)\,dx$

$= \int \{\phi_0(x) - \phi(x)\}\{f_1(x) - K f_0(x)\}\,dx$

$= \int \phi_0(x) \cdot f_1(x)\,dx - \int \phi(x) f_1(x)\,dx$

$+ K \{ \int \phi(x) f_0(x)\,dx - \int \phi_0(x) f_0(x)\,dx \}$

$= E_{\theta_1}\{\phi_0(x)\} - E_{\theta_1}\{\phi(x)\} + K\{E_{\theta_0}\{\phi(x)\} - E_{\theta_0}\{\phi_0(x)\}\}$

$\because E_{\theta_0}\{\phi(x)\} - E_{\theta_0}(\phi_0(x)) \le 0$, because of the assumption that the size of ϕ is no greater than the size of ϕ_0, therefore,

$\int \phi_0(x) f_1(x)\,dx - \int \phi(x) f_1(x)\,dx \ge 0$

which establishes that the power of ϕ cannot be greater than the power of $\phi_0(x)$.

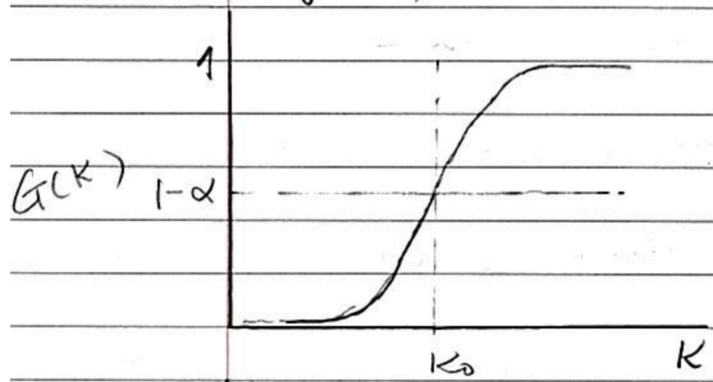
(b) (Existence) Given $\alpha \in (0, 1)$, there exist constants K and γ_0 such that the LRT defined by this K and $\gamma(x) = \gamma_0$ for all x has size exactly α.

Proof: The probability distribution function

$G(K) = Pr_{\theta_0}\{\Lambda(x) \le K\}$ is non-Decreasing as K

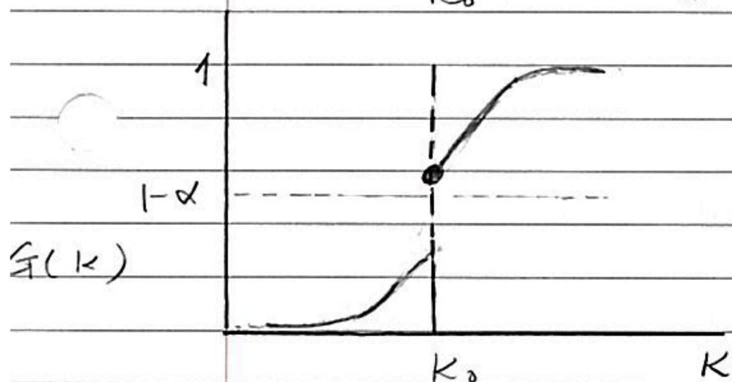
increases; it is also right-continuous.

Try to find a value K_0 for which $G(K_0) = 1-\alpha$. As can be seen from the below figure, There are two possibilities:



(1) Such K_0 exist or

(2) we cannot exactly solve the equation $G(K_0) = 1-\alpha$ but we can find a K_0 for which $G_-(K_0) = Pr_{\theta_0}\{\Lambda(X) < K_0\} \leq 1-\alpha \leq G(K_0)$



in (1), we are done (set $\gamma_0 = 0$)

in case (2), set $\gamma_0 = \dfrac{G(K_0) - (1-\alpha)}{G(K_0) - G_-(K_0)}$

Then it is an easy exercise to demonstrate that this test has size exactly α.

(C) (uniqueness). if the test ϕ has size α, and is of maximum power amongst all possible tests of size α, then ϕ is necessarily a likelihood ratio test, except possibly on a set of values of x which has probability 0 under both H_0 and H_1.

proof:

Let ϕ_0 be the LRT defined by the constant K and function $r(x)$, and suppose ϕ is another test of the same size α and the same power as ϕ_0. Define $U(x)$ as in (a), then $U(x) \geq 0$ for all x, but, because ϕ and ϕ_0 have the same size and power, $\int U(x)\,dx = 0$. So the function $U(x)$ is non-negative and integrates to 0. Hence $U(x) = 0$ for all x, except possibly on a set, S say, of values of x, which has probability zero under both H_0 and H_1. This in turn means that, except on the set S $\phi(x) = \phi_0(x)$ or $f_1(x) = K f_0(x)$, so that $\phi(x)$ has the form of a LRT. This establishes the uniqueness result, and so completes the proof of the theorem.

<< Feb. 27. 2015 >> STAT 846

⇒ Neyman - Pearson Lemma :

$$\phi_0(x) = \begin{cases} 1 & \text{if } \Lambda(x) = f_1(x)/f_0(x) > K \\ \cancel{\gamma(x)} & \cancel{\text{if } \Lambda(x) = f_1(x)/f_0(x) = K} \\ 0 & \text{if } \Lambda(x) = f_1(x)/f_0(x) < K \end{cases}$$

LRT

⇒ **b** Given $\alpha \in (0, 1)$, we can find K and $\gamma(x)$ such that: $E\{\phi(x) \mid \theta_0\} = \alpha$

a

⇒ $\phi_0(x)$ is specified with K and $\gamma(x)$ given in (b) so $E(\phi_0(x) \mid \theta_0) = \alpha$ For any test $\phi(x)$ with $E(\phi(x) \mid \theta_0) \le \alpha$

$E(\phi(x) \mid \theta_1) \le E(\phi_0(x) \mid \theta_1)$

⇒ proof :

suppose that : $V(x) = (\phi_0(x) - \phi(x))[f_1(x) - Kf_0(x)]$

We see that $V(x) \geq 0$

$$\propto \int V(x)\,dx = \int \{\phi_0(x) - \phi(x)\}[f_1(x) - Kf_0(x)]\,dx$$

$$= E_{\theta_1}(\phi_0(x) \mid \theta_1) - K E_{\theta_0}(\phi_0(x) \mid \theta_0)$$

$$- E_{\theta_1}(\phi(x) \mid \theta_1) + K E_{\theta_0}(\phi(x) \mid \theta_0)$$

$$= \{E_{\theta_1}(\phi_0(x)) - E_{\theta_1}(\phi(x))\} + K\{E_{\theta_0}(\phi(x)) - E_{\theta_0}(\phi_0(x))\}$$

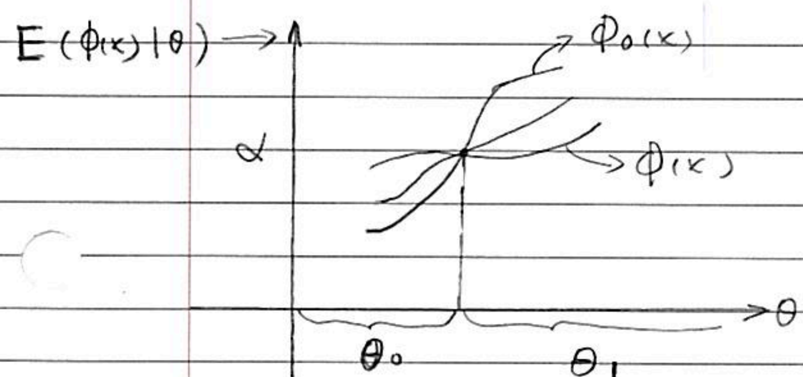
$$\Rightarrow E_{\theta_1}(\phi_0(x)) - E_{\theta_1}(\phi(x)) \geq 0$$

⇒ Definition of uniformly most powerful test of size α is a test $\phi_0(x)$ such that

(1) $E_\theta\{\phi_0(x)\} \le \alpha$ for all $\theta \in \Theta_0$

(2) given any other test $\phi(x)$ for which $E_\theta\{\phi(x)\} \le \alpha$ for all $\theta \in \Theta_0$, we have

$E_\theta\{\phi_0(x)\} \ge E_\theta\{\phi(x)\}$ for all $\theta \in \Theta_1$



⇒ Review of Gamma (α, θ)

(1) θ is a scale

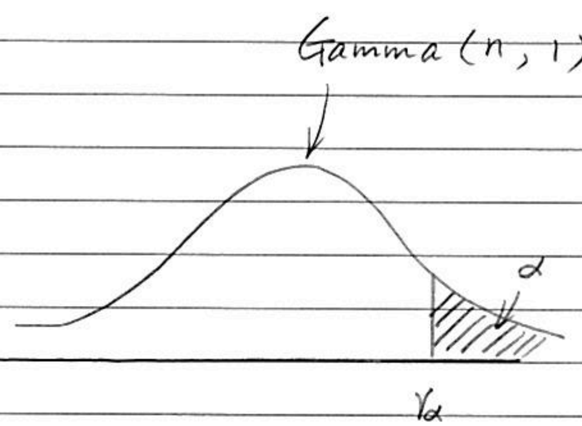
$$f(x|\theta) = \frac{1}{\Gamma(\alpha)} \cdot \left(\frac{1}{\theta}\right)^{\alpha} \cdot e^{-\frac{x}{\theta}} \quad \text{for } x>0$$

(2) Gamma $(\alpha=1, \theta) = \exp(\theta)$

$$f(x|\theta) = \frac{1}{\theta} \exp^{-\frac{x}{\theta}}$$

$x_1, \dots, x_n \overset{iid}{\sim} \exp(\theta)$

$\Sigma x_i \sim$ Gamma (n, θ)



(3) $\Sigma x_i / n \sim$ Gamma $(n. \ 1)$

Let γ_α be the upper α quantile of Gamma $(n, 1)$

$E(\Sigma x_i / n) = \frac{1}{n} \Sigma E(x_i) = \frac{1}{n} \cdot n\theta = \theta$

$\Rightarrow$ example 4.2.

$x_1, x_2, \cdots, x_n \overset{iid}{\sim} \exp(\theta)$

Let $x = (x_1, x_2, \cdots, x_n)$

$f(x|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp^{-x_i/\theta} = \frac{1}{\theta^n} \exp^{-\frac{\Sigma x_i}{\theta}}$

We want to test

Test$_0$: $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ $(\theta_1 > \theta_0)$

By N-P Lemma, the UMP test $\phi_0(x)$ is given

$$\phi_0(x) = \begin{cases} 1, & \text{if } f(x|\theta_1)/f(x|\theta_0) > K \\ 0 & \text{if } f(x|\theta_1)/f(x|\theta_0) < K \end{cases}$$

$$\frac{f(x|\theta_1)}{f(x|\theta_0)} = \frac{\theta_0^n}{\theta_1^n} e^{-\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\Sigma x_i}$$

$$\phi_0(x) = \begin{cases} 1 & \text{if } \Sigma x_i > t \\ 0 & \text{if } \Sigma x_i < t \end{cases}$$

what is t, we should determine t, given

α. $E(\phi_0(x) | \theta_0) = \alpha$

$Pr(\Sigma x_i > t \mid \theta = \theta_0) = \alpha \Rightarrow Pr\left(\frac{\Sigma x_i}{\theta} > \frac{t}{\theta} \mid \theta = \theta_0\right) = \alpha$

So, $\frac{t}{\theta_0} = r_{\alpha,n}$

the UMP is $\quad \phi_0(x) = \begin{cases} 1 & \text{if } \sum x_i > \theta_0 r_{\alpha,n} \\ 0 & \text{if } \sum x_i \le \theta_0 r_{\alpha,n} \end{cases}$

By N-P Lemma

Note that:

(1) $E(\phi_0(x) \mid \theta = \theta_0) = \alpha$

(2) $E(\phi_0(x) \mid \theta = \theta_1) \geq E(\phi(x) \mid \theta = \theta_1)$ for all $\theta_1 > \theta_0$

and all $\phi(x)$ with $E(\phi(x) \mid \theta = \theta_0) \le \alpha$

(3) $\phi_0(x)$ doesn't depend on θ_1. $\Lambda(x)$ is monotone with respect to $\sum x_i$

(4) $E(\phi_0(x) \mid \theta) \le \alpha$ for all $\theta \le \theta_0$

$\Rightarrow$ proof: For $\theta \le \theta_0$ $\quad E(\phi_0(x) \mid \theta)$

$= Pr\left(\sum_{i=1}^{n} x_i > \theta_0 \cdot r_{\alpha,n} \mid \theta\right) \times \underset{1}{\underset{\shortparallel}{\phi_0(x)}} + Pr\left(\sum x_i < \theta_0 \cdot r_{\alpha,n} \mid \theta\right) \cdot \underset{0}{\underset{\shortparallel}{\phi_0(x)}}$

$= Pr\left(\frac{\sum x_i}{\theta} > \frac{\theta_0}{\theta} \cdot r_{\alpha,n} \mid \theta\right)$ $\quad$ Let $Y = \frac{\sum x_i}{\theta}$

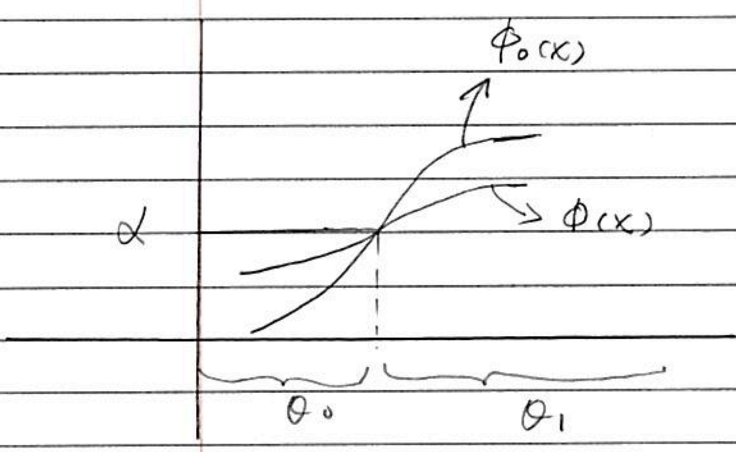
$= Pr\left(Y > \frac{\theta_0}{\theta} \cdot r_{\alpha,n} \mid \theta\right)$

where $Y \sim \text{Gamma}(n, 1)$

We can combining (1), (2), (3), (4) we see that

$\phi_0(x)$ is an UMP test with size α for

Test 1 $\quad \theta \le \theta_0 \iff \theta > \theta_0$


$\phi_0(x)$
$\phi(x)$
α
θ_0
θ_1

<< Mar 02 2015 >> STAT 846

⇒ Definition

A family of densities $\{ f(x|\theta) \mid \theta \in \Theta \}$ with scalar Q

This family has monotone likelihood ratio (MLR)

if $\Lambda(x) = f(x|\theta_1) / f(x|\theta_0)$ is an non-decreasing

function of a statistic $t(x)$ whenever $\theta_0 < \theta_1$

⇒ examples 1:

1) $X_1, \ldots X_n \overset{iid}{\sim} \exp(\lambda = \frac{1}{\theta})$; Let $X = (X_1, \ldots, X_n)$

$f(x|\theta) = \theta^{-n} \cdot e^{-\frac{1}{\theta}\Sigma x_i}$ For $\theta_0 < \theta_1$

$\Lambda(x) = \left(\frac{\theta_0}{\theta_1}\right)^n e^{-\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\Sigma x_i}$

Let $t(x) = \Sigma x_i$

$\Lambda(x)$ is ↗ with respect to $t(x)$

That is : $\Lambda(x) > K \iff \Sigma x_i > t_0$ (determined by θ_0)

⇒ example 2:

$X_1, \ldots X_n \overset{iid}{\sim} f(X|\theta) = C(\theta) \cdot h(x) \cdot e^{\theta T(x)}$

exponential family For $\theta_0 < \theta_1$

$$\Lambda(x) = \frac{f(x|\theta_1)}{f(x|\theta_0)} = \frac{C(\theta_1) \cdot h(x) e^{\theta_1 T(x)}}{C(\theta_0) h(x) e^{\theta_0 T(x)}}$$

$$= \frac{C(\theta_1)}{C(\theta_0)} \cdot e^{(\theta_1 - \theta_0) T(x)}$$

Let $t(x) = T(x)$, $\Lambda(x)$ ↗ function W.R.T $T(x)$

$\Rightarrow$ Theorem 4.2: suppose X has a distribution from a family with MLR with respect to $t(x)$; also the distribution of $t(x)$ is continuous

ⓐ The test:
$$\phi_0(x) = \begin{cases} 1 & \text{if } t(x) > t_0 \\ 0 & \text{if } t(x) \le t_0 \end{cases}$$

where t_0 is a value such that (s.t)
$$E(\phi_0(x) \mid \theta_0) = \alpha$$

[i.e, t_0 is determined by θ_0 and α]

$\phi_0(x)$ is the UMP test for

$$H_0: \theta \le \theta_0 \quad \text{Vs} \quad H_1: \theta > \theta_0$$

$\Rightarrow$ proof: suppose we want to test

$$H_0^*: \theta = \theta_0 \quad \text{Vs} \quad H_1^*: \theta = \theta_1 \quad (\theta_1 > \theta_0)$$

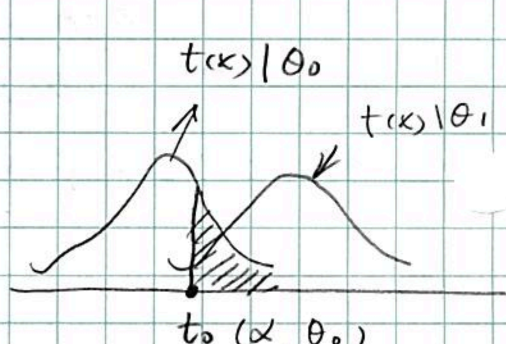
Let $\phi_0(x) = \begin{cases} 1 & \text{if } \Lambda(x) > K \\ 0 & \text{if } \Lambda(x) \le K \end{cases}$

Because $\Lambda(x)$ is monotone w.R.t $t(x)$

$$\phi_0(x) = \begin{cases} 1 & \text{if } t(x) > t_0 \\ 0 & \text{if } t(x) \le t_0 \end{cases}$$

We can determined t_0 by

$$E(\phi_0(x) \mid \theta_0) = \alpha$$



We note:

t_0 doesn't depend on θ_1

By N-P Lemma

$E\{\phi_0(x) \mid \theta_1\} \geq E\{\phi(x) \mid \theta_1\}$ for all ϕ with $E(\phi(x) \mid \theta_0) = \alpha$

We note that: the above statement is true for all $\theta_1 \in \Theta_1 = (\theta_1, +\infty)$

We remain to show that

$E(\phi_0(x) \mid \theta) \leq \alpha$ for all $\theta \leq \theta_0$

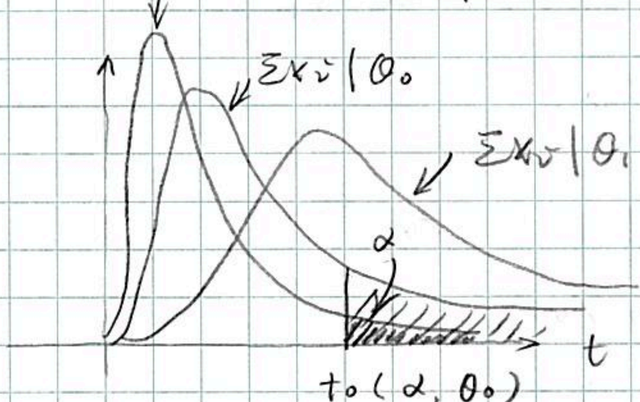
$\Rightarrow$ Example: $X_1, \dots X_n \overset{iid}{\sim} \exp(\lambda = \frac{1}{\theta})$

$t(x) = \Sigma x_i$

scalar

$t(x) \mid \theta \sim \text{Gamma}(n, \beta = \theta)$, θ is a scala

$\Sigma x_i \mid \theta_2$



$$\phi_0(x) = \begin{cases} 1 & \text{if } \Sigma x_i > t_0(\theta_0, \alpha) \\ 0 & \text{if } \Sigma x_i \leq t_0(\theta_0, \alpha) \end{cases}$$

By N-P Lemma: $E(\phi_0(x) \mid \theta_1) \geq E(\phi(x) \mid \theta_1)$

with $E(\phi(x) \mid \theta_0) = \alpha$

also for this example, we can see

$E\{\phi_0(x) \mid \theta_2\} < E(\phi_0(x) \mid \theta_0) = \alpha$

for all $\theta_2 \leq \theta_0$ [this statement is true, general

as long as $\phi(x)$ has MLR; so ϕ is UMP for $\theta \leq \theta_0$ vs

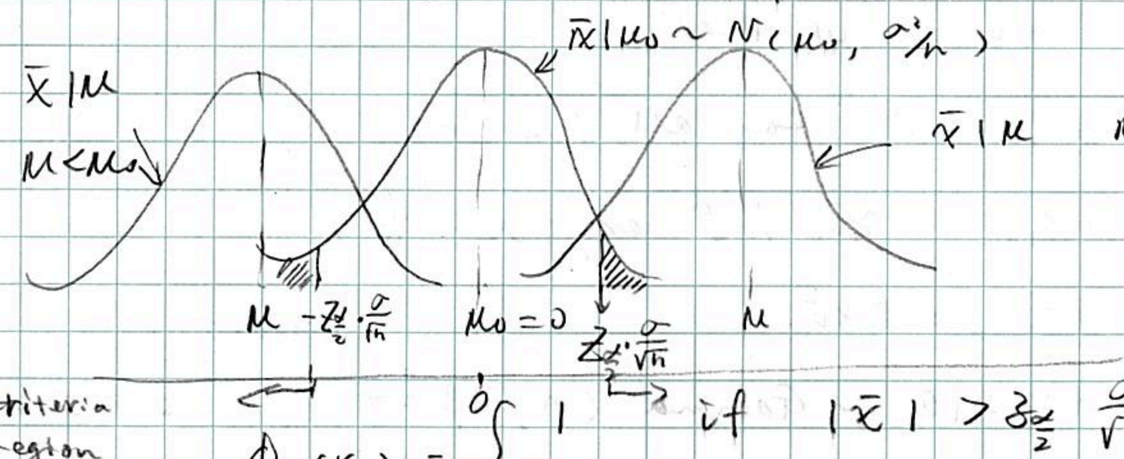
$\Rightarrow$ There is no UMP test for

$$H_0: \theta = \theta_0 \quad vs \quad H_1: \theta \neq \theta_0$$

Example: $X_1, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ σ^2 is known.

$\Lambda(x)$ is monotone W.R.T $t(x) = \Sigma x_i / n$

$H_0: \mu = \mu_0 = 0 \leftrightarrow H_1: \mu \neq \mu_0 = 0$

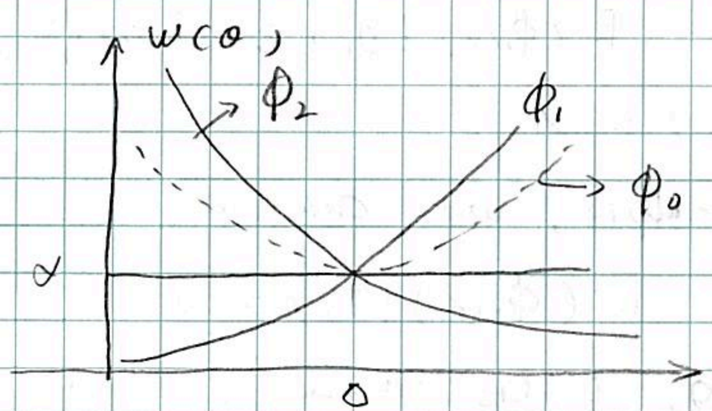


$$\phi_0(x) = \begin{cases} 1 & \text{if } |\bar{x}| > Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ 0 & o/w \end{cases}$$

$$\phi_1(x) = \begin{cases} 1 & \text{if } |\bar{x}| > Z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \\ 0 & o/w \end{cases}$$

$$\phi_2(x) = \begin{cases} 1 & \text{if } \bar{x} < Z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \\ 0 & o/w \end{cases}$$

we can look power function:



when $\mu < 0$ ϕ_2 is Most power
$\mu > 0$ ϕ_1 is Most power

No ϕ_0 s.t $W(\mu)$ is Most power for $\mu \in \Theta_1$

$W(\mu) = E(\phi(x) \mid \mu)$

<< March 04. 2015 >> STAT 846.

⇒ Regular class

Likelihood function:

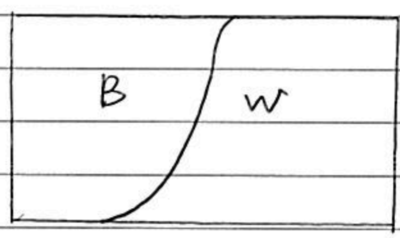
Mode : $x \mid \theta \sim f(x \mid \theta)$

$$\begin{tikzcd}
\theta \arrow[r] \arrow[dr] \arrow[ddr] & 1 \\
 & 2 \\
 & 3
\end{tikzcd}$$

$f(x \mid \theta)$

$L(\theta \mid x) = f(x \mid \theta)$ θ is variable and x is fixed

⇒ Example :



θ = propotion of black balls

$$x_i = \begin{cases} 1 & \text{if } i^{th} \text{ is black} \\ 0 & o/w \end{cases}$$

$x_1, \cdots, x_n \mid \theta \overset{iid}{\sim} \text{Bern}(\theta)$

$f(x_1, \cdots, x_n \mid \theta) = \theta^{n_1} (1-\theta)^{n-n_1}$

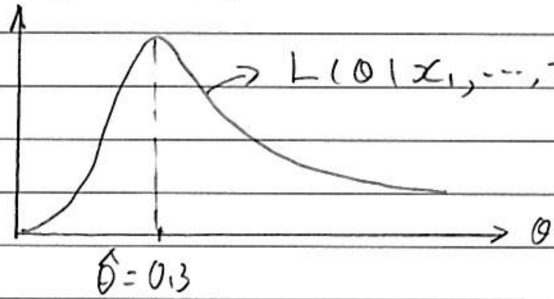
$n_1 = \#$ of black $= \sum_{i=1}^{n} x_i$

$L(\theta \mid x_1, \cdots x_n) = \theta^{n_1} (1-\theta)^{n-n_1}$

Data example $n_1 = 30 = \sum_{i=1}^{n} x_i$, $n = 100$

$L(\theta \mid x_1, \cdots, x_n) = \theta^{30} \cdot (1-\theta)^{70}$

According to graph 0.3 is more likely $\hat{\theta} = 0.3 \rightarrow ML\theta$



$\Rightarrow$ Chapter 5. special Model

Exponential families

$\Rightarrow$ Defination: $f(x|\theta)$ is said to be a model in exponential family if

$$f(x|\theta) = C(\theta)\, h(x) \cdot e^{\sum_{i=1}^{K} \pi_i(\theta) \cdot T_i(x)}$$

Note: $\theta = (\theta_1, \theta_2, \cdots \theta_d)$ d may be $< K$

$\Rightarrow$ Example:

1) $X_1, \cdots, X_n | \theta \overset{iid}{\sim} \exp(\theta)$ θ is scale

$X = (X_1, \cdots, X_n)$

$$f(x|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$$

$\pi_i(\theta) = -\frac{1}{\theta}$ $T_i(x) = \sum_{i=1}^{n} x_i$ $C(\theta) = \theta^{-n}$ $h(x) = 1$

2) $X_1, \cdots x_n | \theta \overset{iid}{\sim}$ Gamma (α, θ)

$$f(x|\theta) = \prod_{i=1}^{n} \frac{x_i^{\alpha-1}}{\Gamma(\alpha) \cdot \theta^{\alpha}} e^{-\frac{x_i}{\theta}}$$

$$= \underbrace{\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \left(\Gamma(\alpha)\right)^{-n}}_{h(x)} \; \underbrace{\theta^{-n\alpha}}_{C(\theta)} \cdot e^{\underbrace{-\frac{1}{\theta}}_{\pi_i(\theta)} \underbrace{\sum_{i=1}^{n} x_i}_{T_i(x)}}$$

3) $X_1, \cdots, X_n \mid \theta \overset{iid}{\sim} \text{Beta}(a, b)$

$$\theta = (a, b), \quad f(x \mid \theta) = \prod_{i=1}^{n} \frac{1}{\text{Beta}(a,b)} x_i^{a-1} \cdot (1 - x_i)^{b-1}$$

$$f(x \mid \theta) = \{\text{Beta}(a, b)\}^{-n} \cdot \left(\prod_{i=1}^{n} x_i\right)^{a-1} \cdot \left(\prod_{i=1}^{n} (1 - x_i)\right)^{b-1}$$

$$= \underbrace{\{\text{Beta}(a, b)\}^{-n}}_{C(\theta)} \cdot e^{\underbrace{(a-1)}_{\pi_1(\theta)} \underbrace{\sum \log x_i}_{T_1(x)} + \underbrace{(b-1)}_{\pi_2(\theta)} \underbrace{\sum \log(1 - x_i)}_{T_2(x)}}$$

$k = d$:
how many terms in the exponential

4) $X_1, \cdots X_n \mid \theta \overset{iid}{\sim} N(\mu, \sigma^2) \qquad \theta = (\mu, \sigma^2)$

$$f(x \mid \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\sum \frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum (x_i^2 - 2\mu x_i + \mu^2)}{2\sigma^2}}$$

$$= \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{n\mu^2}{2\sigma^2}}}_{C(\theta)} \cdot e^{\underbrace{\left(-\frac{1}{2\sigma^2}\right)}_{\pi_1(\theta)} \underbrace{\sum x_i^2}_{T_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{\pi_2(\theta)} \underbrace{\sum x_i}_{T_2(x)}}$$

5) $X_1, \cdots, X_n \mid \theta \sim \text{Bern}(\theta)$

$$f(x \mid \theta) = \prod_{i=1}^{n} \theta^{x_i} \cdot (1 - \theta)^{1 - x_i}$$

$$= e^{\log\left(\prod_{i=1}^{n} \theta^{x_i} \cdot (1 - \theta)^{1 - x_i}\right)}$$

$$= e^{\underbrace{\log\theta}_{\Pi_1(\theta)} \cdot \underbrace{\Sigma x_i}_{T_1(x)} + \underbrace{\log(1-\theta)}_{\Pi_2(\theta)} \cdot \underbrace{(n - \Sigma x_i)}_{T_2(x)}}$$

$d = 1, \quad K = 2$

A model is not in exponential family:

$X \mid \theta \sim \text{Cauchy}(\theta)$

$$f(x \mid \theta) = \frac{1}{\pi} \frac{1}{1 + (x-\theta)^2}$$

$$= \frac{1}{\pi} \cdot e^{-\log(1 + (x-\theta)^2)}$$

cannot be written as $\Sigma\, \Pi_i(\theta)\, T_i(x)$

$\Rightarrow \int f(x \mid \theta)\, dx = 1$

Suppose we can change $\frac{\partial}{\partial\theta}$ and $\int$

[Sometimes we cannot

$X \mid \theta \sim \text{unif}(0, \theta)$

$\int_0^\theta \frac{1}{\theta}\, dx = 1 \Rightarrow \frac{\partial}{\partial\theta} \int_0^\theta \frac{1}{\theta}\, dx \neq \int_0^\theta \frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right) dx$]

$E_{X|\theta}\left(\frac{\partial}{\partial\theta_j} \log f(x \mid \theta) \mid \theta\right) = 0 \quad$ for $j = 1, 2, \cdots d$.

$= E_{X|\theta}\left\{\frac{\partial}{\partial\theta_i \partial\theta_j} \log f(x \mid \theta) \mid \theta\right\} = -E_{X|\theta}\left(\frac{\partial \log f}{\partial\theta_i} \cdot \frac{\partial \log f}{\partial\theta_j}\right)$

proof: Right hand side

$$= E\left(\frac{\partial \log f}{\partial\theta_j} \mid \theta\right) = \int \frac{\partial \log f}{\partial\theta_j} \cdot f(x \mid \theta)\, dx$$

$$= \int \frac{1}{f(x|\theta)} \cdot \frac{\partial f(x|\theta)}{\partial \theta_j} \cdot f(x|\theta)\, dx$$

$$= \int \frac{\partial f(x|\theta)}{\partial \theta_j}\, dx \quad = \frac{\partial}{\partial \theta_j} \int f(x|\theta)\, dx$$

$$= \frac{\partial}{\partial \theta_j}(1) \quad = 0$$

$\Rightarrow$ If $f(x|\theta) = C(\theta) \cdot h(x) \cdot \exp^{\sum \Pi_i(\theta) T_i(x)}$

$$\log f(x|\theta) = \log C(\theta) + \log h(x) + \sum \Pi_i(\theta) \cdot T_i(x)$$

$$\frac{\partial \log f(x|\theta)}{\partial \theta_j} = \frac{1}{C(\theta)} \cdot \frac{\partial C(\theta)}{\partial \theta_j} + \sum \frac{\partial \Pi_i(\theta)}{\partial \theta_j} \cdot T_i(x)$$

So $E\Big(\underbrace{\frac{\partial \log f(x|\theta)}{\partial \theta_j}}_{0}\Big) = \frac{1}{C(\theta)} \cdot \frac{\partial C(\theta)}{\partial \theta_j} + \sum \frac{\partial \Pi_i(\theta)}{\partial \theta_j} \cdot E(T_i(x)) = 0$

for $i = 1, \cdots, k$

we can find $E[T_i(x)]$ by solving these equations.

Example : $f(x| \mu, \sigma^2) = e^{-\frac{1}{2\sigma^2}\sum x_i^2 - \frac{\mu}{\sigma^2}\sum x_i}$

$$E(\sum x_i^2) = 0 \quad , \quad E(\sum x_i) = 0$$

$$E(T_i(x)) = 0$$

<< Mar 06. 2015 >> STAT 846

=> if $f(x|\theta)$ is in the exponential family; then

$$\frac{\partial}{\partial \theta_j} \log c(\theta) + \sum_{i=1}^{k} \frac{\partial \pi_i}{\partial \theta_j} E(T_i(x)) = 0 \text{ for } j = 1, 2, \cdots d$$

=> Def: Natural parameterization

suppose $f(x|\theta) = c(\theta) h(x) e^{\sum_{i=1}^{k} \pi_i(\theta) T_i(x)}$

Natural parameter

let $\boxed{\eta_i = \pi_i(\theta)}$, $\eta = (\eta_1, \cdots \eta_n)$

$f(x|\eta) = c'(\eta) h(x) e^{\sum \eta_i T_i(x)}$

in such case, we say $f(x|\eta)$ is in its natural parameterization.

=> Example: $f(x|\theta) = (2\pi)^{-\frac{n}{2}} \cdot (\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}}$

$$\cdot e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum x_i} \qquad \theta = (\mu, \sigma^2)$$

Let $\eta_1 = -\frac{1}{2\sigma^2}$, $\eta_2 = \frac{\mu}{\sigma^2}$ $\quad (\mu, \sigma^2) \Leftrightarrow (\eta_1, \eta_2)$

$f(x|\eta) = (2\pi)^{-\frac{n}{2}} \hat{c}(\eta) e^{\eta \sum x_i^2 + \eta_2 \sum x_i}$

=> Def: Natural Parameterization space:

$$\Pi = \{ \eta = (\pi_1(\theta), \cdots \pi_k(\theta)) \mid \theta \text{ such that } c(\theta) \text{ is finite} \}$$

k is dimention of Π (Natural parameterization space)

$\theta = (\theta_1, \theta_2, \cdots \theta_d)$ $\quad \pi = (\pi_1(\theta), \cdots \pi_k(\theta))$

$\theta_d \to d$ $\quad \pi_k \to k$

if $d = k$ $\boxed{\text{dimention}(\theta) = \text{dimention}(\pi)}$, then we say $f(x|\theta)$ is in full exponential family

If $d < k$, then we say $f(x|\theta)$ is in curved exponential family. (Reduce)

=> example:

1) $X_1, \cdots X_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$, it is in full exponential family

2) $x_1, \cdots, x_n \mid \theta \sim$ Bernuli (θ)

$$f(x|\theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= e^{\log\theta \sum x_i + \log(1-\theta)\sum(1-x_i)}$$

here, $\pi_1 = \log(\theta)$ $\quad \pi_2 = \log(1-\theta)$

(π_1, π_2) $\quad k = 2$ $\quad d = 1$ (only θ)

therefore, $f(x|\theta)$ is in a curved exponential family

=> A fact: {E, and Cov. for natural exponential family}

If $f(x|\theta)$ is in nat parameterization, i.e.

$$f(x|\theta) = C(\theta)h(x) \cdot e^{\sum_{i=1}^{k} \theta_i T_i(x)}, \text{ where, } \pi_i(\theta) = \theta_i$$

$$\frac{\partial}{\partial\theta_j}\log(C(\theta)) + \sum_{i=1}^{k} \frac{\partial \pi_i}{\partial\theta_j} E(T_i(x)) = 0$$

$$\frac{\partial}{\partial\theta_j}\log C(\theta) + E(T_j(x)) = 0$$

$$E(T_j(x)) = -\frac{\partial}{\partial\theta_j}\log C(\theta)$$

Also, we can show

$$\mathrm{cov}(T_i(x), T_j(x) \mid \theta) = -\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log C(\theta)$$

$\Rightarrow$ Lemma 5.1

Suppose that, $f(x|\theta) = C(\theta) h(x) e^{\sum_{i=1}^{k} \pi_i(\theta) T_i(x)}$

the the joint distribution of $(T_1, \cdots, T_k)$ is in exponential family. natural parameter $(\pi_1, \cdots \pi_k)$

proof: Let x be discrete

$$Pr(T_1 = y_1, \cdots, T_k = y_k \mid \theta) = \sum_{\left\{ x \,\middle|\, \begin{array}{c} T_1(x) = y_1 \\ \vdots \\ T_k(x) = y_k \end{array} \right\}} Pr(X = x \mid \theta)$$

$$= \sum_{\{x \mid T_i(x) = y_i\}} C^n(\theta) h(x) \underbrace{e^{\sum_{i=1}^{k} \pi_i(\theta) T_i(x)}}$$

$\hookrightarrow$ constant over Σ

$$= C^n(\theta) \left(\sum_{\{x \mid T_i(x) = y_i\}} h(x) \right) e^{\sum \pi_i(\theta) \cdot y_i}$$

$$= C^n(\theta) \cdot \tilde{h}(y) e^{\sum \pi_i(\theta) \cdot y_i}$$

$\Rightarrow$ Lemma 5.2

~~suppose that~~ $f(x|\theta) = C(\theta) \cdot h(x) \cdot e^{\sum \pi_i(\theta) T_i(x)}$

Let S be a subset of $\{1, 2, \cdots, k\}$, then the

discrete $T_i \quad i \in S \mid T_j \quad j \notin S$

is of exponential family with natural parameterization

$\{ \pi_i(\theta) \mid i \in S \}$

$\Rightarrow$ proof: Let x be discrete, Let $S = \{1, 2, \cdots \ell\}$

$Pr(T_1 = y_1, \cdots T_k = y_k \mid \theta)$

$= C(\theta) \tilde{h}(y) e^{\sum \pi_i(\theta) y_i}$

$Pr(T_1 = y_1, \cdots T_\ell = y_\ell \mid T_{\ell+1} = y_{\ell+1}, \cdots T_k = y_k, \theta)$

$= Pr(T_1 = y_1, \cdots T_k = y_k) \Big/ \sum_{y_1 \cdots y_\ell} [Pr(T_1 = y_1 \cdots T_k = y_k \mid \theta)]$

$= C(\theta) \tilde{h}(y) \prod e^{\pi_i(\theta) y_i} \Big/ \sum_{y_1 \cdots y_\ell} C(\theta) \tilde{h}(y) \prod e^{\pi_i(\theta) y_i}$

$= \dfrac{\tilde{h}(y)}{\left(\sum_{y_1 \cdots y_\ell} \tilde{h}(y) \cdot e^{\sum_{i=1}^{\ell} \pi_i(\theta) y_i} \right)} \cdot e^{\sum_{i=1}^{\ell} \pi_i(\theta) y_i}$

$\uparrow$

free of $y_1, \cdots y_\ell$, related to θ i.e. can be written as $\tilde{C}(\theta)^{-1}$

<<Chapter 6>> : Sufficiency and Completness

<< Mar 09. 2015 >> STAT 846

⇒ Sufficient statistics

likelihood function: $L(\theta; x) = f(x \mid \theta)$

θ is variable, x is fixed.

⇒ Example: $X_1, \cdots X_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$$x = (x_1, \cdots, x_n), \quad \theta = (\mu, \sigma^2)$$

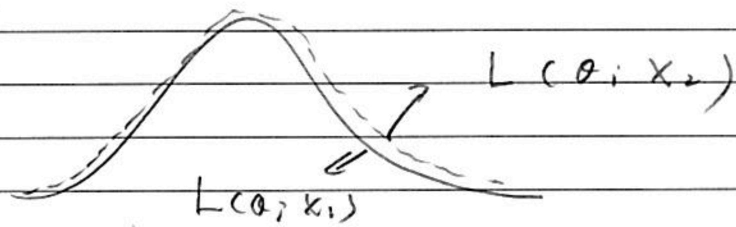
$$L(\theta, x) = f(x \mid \theta) = (2\pi)^{-\frac{n}{2}} \cdot (\sigma^2)^{-\frac{n}{2}} \cdot e^{\frac{-n\mu}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}\Sigma x_i^2 + \frac{\mu}{\sigma^2}\Sigma x_i}$$

$$= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2}[n(\mu-\bar{x})^2 + (n-1)v^2]}$$

⇒ informally: $L(\theta; x)$ is determined by $\bar{x}$ and v^2, it means that: For two data sets x_1 and x_2 of, if $\begin{cases} \bar{x}_1 = \bar{x}_2 \\ V_1^2 = V_2^2 \end{cases}$, then

given.

$L(\theta; x_1) = L(\theta; x_2)$



⇒ Definition:

A stististics $T = T(x)$ is sufficient for θ, if one of the following conditions are true:

1) Factorization:

irrelavant to θ

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

2) likelihood ratio theorem.

For any pair of data sets x and x' such that $t(x) = t(x')$, we have

$$\Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$$

where, $\Lambda_x(\theta_1, \theta_2) = \dfrac{f(\theta_2, x)}{f(\theta_1, x)}$

3) $f(x \mid t(x), \theta)$ is independent of θ

i.e $f(x \mid t(x), \theta) = f(x \mid t(x))$

$\Rightarrow$ Lamma 6.1

Condition 1 $\Longleftrightarrow$ Condition 2

pf: For 1) $\Rightarrow$ 2) suppose $t(x) = t(x')$

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

$$\Lambda_x(\theta_1, \theta_2) = h(x) \cdot g(t(x), \theta_2) \,/\, h(x) \cdot g(t(x), \theta_1)$$

$$= g(t(x), \theta_2) \,/\, g(t(x), \theta_1)$$

$$\Lambda_{x'}(\theta_1, \theta_2) = g(t(x'), \theta_2) \,/\, g(t(x'), \theta_1)$$

$$\Rightarrow \Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$$

For 2) $\Rightarrow$ 1)

pf: Fixed θ_0, for any θ

$\dfrac{f(x;\theta)}{f(x,\theta_0)}$ as a function of θ; is the same

as $\dfrac{f(x',\theta)}{f(x',\theta_0)}$ as long as $t(x)=t(x')$

it implies $\dfrac{f(x,\theta)}{f(x,\theta_0)} = g^*(t(x);\theta,\theta_0)$

$$f(x,\theta) = f(x,\theta_0)\cdot g^*(t(x);\theta,\theta_0)$$

$$= h(x)\cdot g(t(x);\theta)$$

$\Rightarrow$ Condition 3) $\Leftrightarrow$ Condition 1)

from 1) $\Rightarrow$ 3)

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

suppose that x is discrete for all x, t such that $t=t(x)$

$$f(x|t(x),\theta) = f(x,t=t(x)|\theta) \,/\, f(t(x)=t|\theta)$$

$$= f(x|\theta) \Big/ \sum_{t(x)=t} f(x|\theta)$$

$$= h(x)\cdot \underline{g(t(x)\overset{=t}{,}\theta)} \Big/ \sum_{t(x)=t} h(x)\cdot \underbrace{g(t(x)=t,\theta)}_{\text{constant}}$$

$$= \frac{h(x)}{\sum h(x)} \quad \text{free of } \theta$$

example for above pf:

$$P(x_1=1, x_2=2 \mid x_1+x_2=3, \theta) = \frac{P(x_1=1, x_2=2, x_1+x_3=3 \mid \theta)}{P(x_1+x_2=3 \mid \theta)}$$

Form 3) $\Rightarrow$ 1)

$f(x|t(x),\theta)$ is free of θ

$\exists\, h(x)$ such that $f(x|t(x),\theta) = h(x)$

For ~~X~~, t such that $t(x) = t$, then $f(x|\theta) = f(x,\ t(x) \mid \theta)$

$= f(t(x) = t \mid \theta) \cdot f(x \mid t(x) = t, \theta)$

by product rule

$= h(x) \cdot g(t(x) \mid \theta)$

<< Mar 11, 2015 >> STAT 846.

=> **Definition:** T is a sufficient statistic if for any pair x, x' with $T(x) = T(x')$, then.

$\Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$ for all θ_1, θ_2

in other words: $L(\theta, x) = L(x, x') L(\theta, x')$

=> **Definition**: Minimal sufficient statistics (MSS)

A sufficient statistics $T(x)$ is minimal sufficient statistics if $T(x)$ is a function of any other sufficient statistics, i.e. For any sufficient statistic S, where exist a g such that $T(x) = g(S)$

=> Remark:

(1) Minimal sufficient statistic is not unique

example: $X_1, X_2, \dots X_n | \theta \sim \text{poisson}(\theta)$

$T_1(x) = \Sigma X_i$ is minimal sufficient

$S(x) = 2\Sigma X_i$ is minimal sufficient

i.e $\exists$ one-to one g such that $T_1 = g(T_2)$

=> Lemma 6.2:

if T and S are two minimal sufficient statistics, then $\exists$ injective function g_1 and g_2

such that $T = g_1(S)$ $\quad S = g_2(T)$

$\Rightarrow$ theorem 6.1:

$T(x)$ is minimal sufficient statistics

$\Leftrightarrow$ For any pair X and X', $T(x) = T(x')$

$\Leftrightarrow \Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$ for any θ_1 and θ_2

$\Rightarrow$ example 6.1.

$X_1, \cdots X_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$$f(x ; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\Sigma x_i^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\Sigma x_i - \frac{n\mu^2}{2\sigma^2}}$$

$T(x) = (\Sigma x_i^2, \Sigma x_i)$

$$\Lambda_x(\theta_1, \theta_2) = \frac{C(\theta_1)}{C(\theta_2)} \cdot e^{-\frac{\Sigma x_i^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) + \Sigma x_i\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right)}$$

For any pair of data set x and x'

if $T(x) = T(x')$

the $\Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$ for all θ_1 and θ_2

$\Rightarrow T$ is sufficient.

$\Rightarrow$ proof theorem 6.1

For any pair of data sets X and X'

If $\Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$ for some θ_1, θ_2

then we want to say $T(x) = T(x')$

This is equivelent to prove that

if $T(x) \neq T(x')$ then $\lambda_x(\theta_1, \theta_2) \neq \lambda_{x'}(\theta_1, \theta_2)$

for some θ_1 and θ_2.

Minimal: If $L(\theta, x') = c\, L(\theta, x)$, then

$$T(x) = T(x')$$

$\Leftrightarrow$ if $T(x) \neq T(x')$ then $L(\theta, x') \neq c\, L(\theta, x)$

$\Rightarrow$ Example 6.2:

If $f(x_i \mid \theta)$ is in exp family i.e.

$$f(x \mid \theta) = C(\theta)\, h(x) \cdot e^{\sum \lambda_i(\theta) T_i(x)}$$

and Θ contains an open rectangle.

Then $T(x) = (T_1(x), \cdots, T_k(x))$ is minimal suffi statistic Indeed it is complete, too.

$\Rightarrow$ HW:

$$X_1 \cdots X_n \mid \theta \sim \text{poisson}(\theta)$$

to show: $T\left(\sum_{i=1}^{5} x_i, \sum_{i=6}^{n} x_i\right)$ is not minimal sufficient by using the therom 6.1

3) $\tau^2 \mid \theta, \mu_1, \cdots, \mu_p, x_1, \cdots, x_p$

$$f(\tau^2 \mid \cdots\cdots) \propto \prod_{i=1}^{p} f(\mu_i \mid \tau^2, \theta) \cdot \pi_{\tau^2}(\tau^2)$$

$$= e^{\sum_{i=1}^{p} -\frac{(\mu_i-\theta)^2}{2\tau^2}} \cdot (\tau^2)^{-\frac{p}{2}} \cdot (\tau^2)^{-(\alpha^*+1)} \cdot e^{-\frac{\beta^*}{\tau^2}}$$

$$= e^{-\frac{1}{\tau^2}\left(\underline{\sum_{i=1}^{p}\frac{(\mu_i-\theta)^2}{2} + \beta^*}\right)} \cdot (\tau^2)^{-(\underline{\alpha^* + \frac{p}{2}} + 1)}$$

$$\Rightarrow \tau^2 \mid \cdots \sim \text{inv-Gamma}\left(\frac{2\alpha^*+p}{2}, \frac{\sum_{i=1}^{p}(\mu_i-\theta)^2 + 2\beta^*}{2}\right)$$

Let $\alpha^* = \frac{K^*}{2}$, $\beta^* = \frac{\lambda^*}{2}$

$$\Rightarrow \tau^2 \mid \cdots \sim \text{inv-Gamma}\left(\frac{K^*+p}{2}, \frac{\lambda^* + \sum_{i=1}^{p}(\mu_i-\theta)^2}{2}\right)$$

$$E(\tau^2 \mid \cdots) = \frac{\lambda^* + \sum_{i=1}^{p}(\mu_i-\theta)^2}{K^*+p}$$

$\Rightarrow$ Gibbs sampling

$(\theta_1, \cdots, \theta_p) \sim \pi(\theta_1, \cdots, \theta_p)$

starting from $\theta^{(0)} = (\theta_1^{(0)}, \cdots, \theta_p^{(0)})$

reapeat N times of :

(1). Draw $\theta_1^{(1)} \sim \pi(\theta_1 \mid \theta_2^{(0)}, \cdots \theta_p^{(0)})$

(2). Draw $\theta_2^{(1)} \sim \pi(\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)}, \cdots, \theta_p^{(0)})$

⋮

(K). Draw $\theta_K^{(1)} \sim \pi(\theta_K \mid \theta_1^{(1)}, \theta_2^{(1)} \cdots \theta_{K-1}^{(1)}\ \theta_{K+1}^{(0)} \cdots \theta_p^{(0)})$

⋮

(p). Draw $\theta_p^{(1)} \sim \pi(\theta_p \mid \theta_1^{(1)}, \theta_2^{(1)} \cdots \theta_{p-1}^{(1)})$

set $\theta^{(0)} = \theta^{(1)}$

informally, Repeat N times for $i = 1, \cdots, p$

$\theta_i \sim \pi(\theta_i \mid \theta_{-i})$

<< Mar 13, 2015 >> STAT 846

=> Mini sufficient

A statistic T is said to be MSS $\overset{\text{iff}}{\Longleftrightarrow}$

For any pair of data sets x, x'

(1) $T(x) = T(x') \Rightarrow \Lambda_x(\theta_1, \theta_2) = \Lambda_{x'}(\theta_1, \theta_2)$
for all θ_1, θ_2.

(2) $T(x) \neq T'(x) \Rightarrow \Lambda_x(\theta_1, \theta_2) \neq \Lambda_x(\theta_1, \theta_2)$
for some θ_1, θ_2.

=> $\boxed{\text{Complete statistic.}}$

A Statistic T is said to be complete

for any real-value function g, if

$E(g(T) \mid \theta) = 0$ for all θ, then $Pr(g(T) = 0 \mid \theta) = 1$

=> what completeness implies?

if T is complete, then there exists only

one unbiased estimator for θ, Let

$E(g_1(t) \mid \theta) = 0, \quad \forall \theta.$

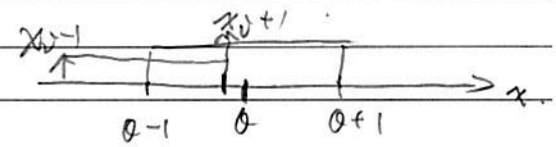
$E(g_2(t) \mid \theta) = 0, \quad \forall \theta$

$E(g_1(T) - g_2(T) \mid \theta) = 0$

if T is not complete for θ, unbiased estimator for θ is not unique.

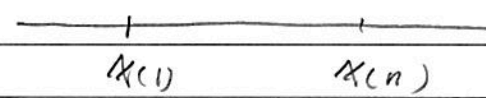
If T is complete, then unbiased estimator is unique.

$\Rightarrow$ example: (not complete)



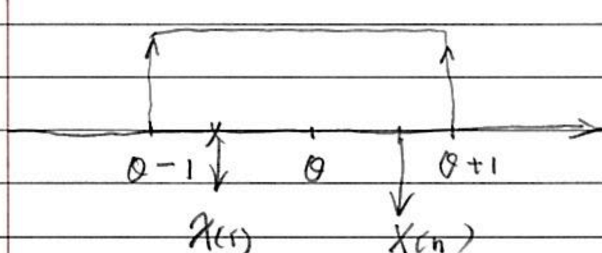
$X_1, \dots X_n \mid \theta \overset{iid}{\sim} \text{unif}(\theta-1, \theta+1)$

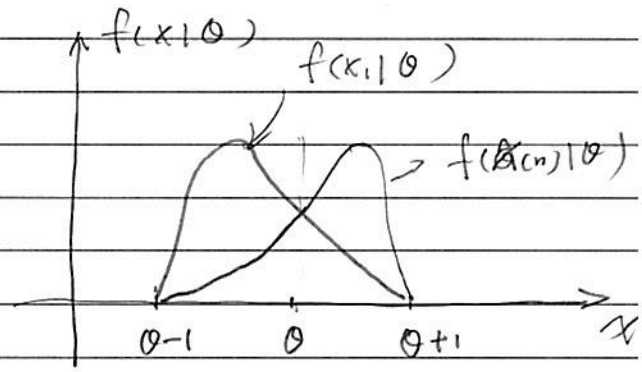
$$f(x) = \prod_{i=1}^{n} I(\theta-1 < x_i < \theta+1)$$
$$= \prod_{i=1}^{n} I(x_i-1 < \theta < x_i+1)$$
$$= I(\theta \in (x_i-1, x_i+1)) \quad \text{for all } i=1,2,\dots n)$$
$$= I(\theta \in (X_{(n)}-1, x_{(1)}+1))$$



$T(X_{(1)}, T_{(n)})$ is minimal sufficient statistic.

but we will see T is not complete





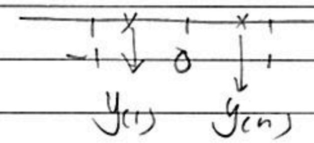
$f(x_{(1)} \mid \theta)$ is symetric about θ with $f(x_{(n)}, \theta)$

$$E\left(\frac{X_{(1)} + X_{(n)}}{2}\right) = \theta$$

The distribution of $X_{(n)} - X_{(1)}$ is independent of θ,

Let $Y_i = X_i - \theta \sim \text{unif}(-1, 1)$

$$Y_{(n)} = X_{(n)} - \theta$$
$$Y_{(1)} = X_{(1)} - \theta$$



$X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$ the distribution of $Y_{(n)} - Y_{(1)}$ doesn't depend on θ so the distribution of $X_{(n)} - X_{(1)}$ doesn't depend on θ too!

$E(X_{(n)} - X_{(1)} \mid \theta) = c$ for all θ, c doesn't depend on θ.

Now we have:

$$E\left(\frac{X_{(1)} + X_{(n)}}{2} \,\Big|\, \theta\right) = \theta \text{ for all } \theta$$

$$E\left(\frac{X_{(1)} + X_{(n)}}{2} - \left((X_{(n)} - X_{(1)}) - c\right) \Big|\, \theta\right) = \theta$$

for all

There are two unbiased estimator which are function $f(X_{(1)}, X_{(n)})$, for θ.

=> Lemma 6.3

If $T = (T_1, \cdots, T_K)$ is the natural statistical of an exp family also Θ contains an open rectangle, then, T is complete.

proof: the pdf / pmf of $(T_1, \cdots, T_K)$ has the form as follows

$$f(t_1, \cdots, t_K | \theta) = C(\theta)\, h(t) \cdot e^{\sum_{i=1}^{K} \theta_i t_i}$$

suppose for a real-value function g such that

$$E(g(T_1, \cdots, T_K) | \theta) = 0 \quad \text{for all } \theta$$

we will show $g(t_1, \cdots, t_K) = 0$ a.s.

Let $t = (t_1, \cdots, t_K)$

$$E(g(T_1, \cdots, T_K) | \theta)$$

$$= \int g(t)\, C(\theta)\, h(t)\, e^{\sum_{i=1}^{K} \theta_i t_i}\, dt$$

$$= C(\theta) \int g(t) \cdot e^{\sum_{i=1}^{K} \theta_i t_i}\, dt$$

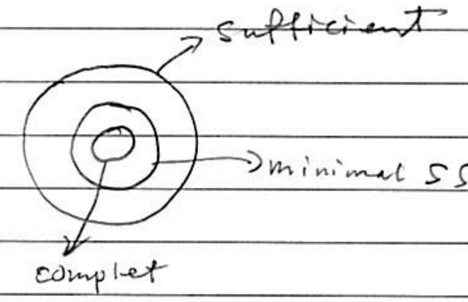
This is the Laplace transformal of $g(t) h(t)$, so integral $= 0$ => $g(t) h(t) = 0$

//

<< Mar 16, 2015 >> STAT 846

=> Lehman - Scheffe

theorem: if T is sufficient and complete, then T is minimal sufficient statistic (MSS)



=> UMVUE: we says a statistic $T(x)$ is an uniformly minimum variance unbiased estimator for θ if

1) $E(T(x) \mid \theta) = \theta$ for all θ

2) $V(T(x) \mid \theta) \le V(d(x) \mid \theta)$, for all θ, ; for all $d(x)$

with $E(d(x) \mid \theta) = \theta$

$V(T(x) \mid \theta) = E((T(x) - \theta)^2 \mid \theta)$

because $E(T \mid \theta) = \theta$

=> theorem 6.3 (Rao - Blackwell Theorem)

Given

$E(d_1(x) \mid \theta) = \theta$

T is a sufficient statistics

$g(T) = E(d_1(x) \mid T, \theta)$

then,

1) $g(T)$ is a statistic i.e free of θ

2) $E(g(T)|\theta) = \theta$

3) $Var(g(T)|\theta) \leq Var(d(x)|\theta)$

proof:

1) T is sufficient for θ

$X|T$ is independent of θ

so $E(d_1(x)|T, \theta)$ is free of θ

2) $E(g(T)|\theta) = E_T(E_X(d_1(x)|T, \theta))$

$= E_X(d_1(x)|T)$

$= \theta$

3)

$E_X[(d_1(x) - \theta)^2]$

$= E_T E_X((d_1(x) - \theta)^2 | T)$

Note that:

$E_X((d_1(x) - \theta)^2 | T) \quad = (g(T) - \theta)^2$

$\geq (E(d_1(x)|T) - \theta)^2$ by Jensen Inequality

Then, we see that

$$E_X[(d_1(x) - \theta)^2] \geq E_T[(g(T) - \theta)^2]$$

Remarks:

1) if T is complete ; $g(T)$ is unique

So, the $g(T)$ is UMVUE

T is complete

$$d_1(x), \quad d_2(x), \quad \cdots\cdots, \quad d_n(x)$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$g_1(T)=E(d_1(x)|T) \quad g_2(T)=E(d_2(x)|T) \quad \cdots\cdots \quad g_n(T)=E(d_n(x)|T)$$

★ ⇒ Methods for finding UMVUE for θ

1) Find the complete T for θ

2) Find an estimator statistic $d_1(x)$ such that

$$E(d_1(x)|\theta)=\theta$$

3) Find $E(d_1(x)|T) \equiv g(T)$

$g(T)$ is the UMVUE

⇒ example: $X_1, \cdots X_n \overset{iid}{\sim} \text{poisson}(\lambda)$

Find the UMVUE

Review: $X \sim \text{poisson}(\lambda)$, $Y \sim \text{poisson}(\mu)$

$X \perp Y$

$X+Y \sim \text{poisson}(\lambda+\mu)$

For $k \geq x$

$$P(X=x \mid X+Y=k) = \frac{P(X=x,\ Y=k-x)}{P(X+Y=k)}$$

$$= \frac{P(X=x)\cdot P(Y=k-x)}{P(X+Y=k)}$$

$$= \frac{e^{-\lambda}\lambda^{x}}{x!}\cdot\frac{e^{-\mu}\mu^{k-x}}{(k-x)!} \Big/ \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^{k}}{k!}$$

$$= \binom{k}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^{x}\left(1-\frac{\lambda}{\lambda+\mu}\right)^{k-x}$$

$\Rightarrow \ X \mid X+Y = k \sim Bin\left(k, \frac{\lambda}{\lambda+\mu}\right)$

Solution:

(1) $T = \Sigma X_i$ is complete statistic, because poisson (λ) is in exp family

$E(X_1) = \lambda$, Let $d_1(x) = X_1$, the UMVUE for λ is $g(T) = E(X_1 \mid T)$

$X_1 \sim poisso(\lambda)$; $X_2 + \cdots + X_n \sim poisso((n-1)\lambda)$

$X_1 \perp X_2 + \cdots + X_n$

$X_1 \mid X_1 + \cdots + X_n = K \sim Bin\left(K \ \frac{\lambda}{n\lambda}\right)$

$X_1 \mid T \sim Bin($

$= Bin\left(K, \frac{1}{n}\right)$

in other words

$X_1 \mid T \sim Bin\left(T, \frac{1}{n}\right)$

$E(X_1 \mid T) = \frac{T}{n} = \bar{X}$

(2) Find an UMVUE for λ^2

$T = \Sigma X_i$ is complete for λ^2

$Var(X_1) = \lambda = E(X_1^2) - (E(X_1))^2$.

$\Rightarrow E(X_1^2 - X_1) = \lambda^2$

We will find $g(T) = E(X_1^2 - X_1 / T)$

$V(X_1 \mid T) + (E(X_1 \mid T))^2 = E(X_1^2 \mid T) - E(X_1 \mid T)$

$= \left[T \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) + \left(T \cdot \frac{1}{n}\right)^2\right] - \frac{T}{n}$

$$= \frac{1}{n^2}\left(T^2 - T\right)$$

$$= \bar{x}^2 + \frac{\bar{x}}{n}$$

Chapter 8 : « Likelihood theory »

« March 18, 2015 »   STAT 843.

Chapter8: « Likelihood theory »

$\Rightarrow$ $L(\theta; x) = f(x \mid \theta)$; L

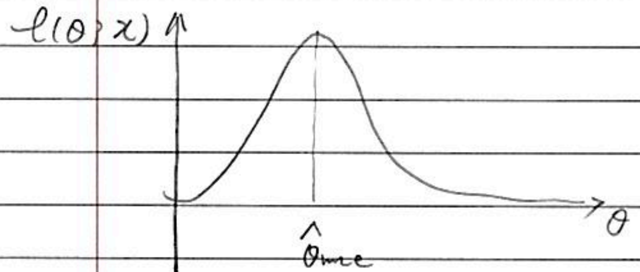
(1) score function :

log-likelihood :

$l(\theta, x) = \log L(\theta, x) = \log f(x \mid \theta)$

score function : $S(\theta; x) = \nabla_\theta l(\theta; x) = \frac{\partial}{\partial \theta} \log L(\theta, x)$

Maximum likelihood estimator (MLE)



$\hat{\theta}_{mle}(x) = \underset{\theta}{\arg\max}\, l(\theta; x)$

$= \arg\max L(\theta, x)$

$\Rightarrow$ An approach to finding $\hat{\theta}_{mle}$.

to solve $\nabla_\theta l(\theta, x) = 0$

$\Rightarrow$ example 1)

1) $x_1, \cdots x_n \mid \theta \sim unif(0, \theta)$

$L(\theta; x) = \prod_{i=1}^{n} f(x_i \mid \theta_i)$

$= \frac{1}{\theta^n} I(X_{(n)} < \theta)$ where $X_{(n)} = \max\{x_1, \cdots x_n\}$

$\hat{\theta}_{mle}(x) = X_{(n)}$

$E(X_{(n)})$

$= \int_0^{\theta} n \cdot x \left(\frac{x}{\theta}\right)^{n-1} dx$

$= n \cdot \frac{1}{\theta^{n-1}} \cdot \int_0^{\theta} x^n \, dx$

$= \frac{n}{\theta^{n-1}} \cdot \frac{x^{n+1}}{n+1} \Big|_0^{\theta} = \frac{n\theta}{n+1} < \theta$

$P((X_{(n)}) \leq x \mid \theta)$

$= \left(\frac{x}{\theta}\right)^n$

~~$f_{X_{(n)}}(x) = n \cdot \left(\frac{x}{\theta}\right)^{n-1}$ for $x < 0$~~

=> example 2:

$x_1, \cdots x_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

Likelihood: $L(\theta, x) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp - \frac{\Sigma(x_i - \mu)^2}{2\sigma^2}$

log-likelihood. $l(\theta, x) = -\frac{n}{2} \cdot \log 2\pi\sigma^2 - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$

Score-function:

$$\frac{\partial l}{\partial \mu} = - \frac{2\,\Sigma(x_i - \mu)}{2\sigma^2} = \frac{\Sigma(x_i-\mu)}{\sigma^2} = 0$$

$$\Rightarrow \mu = \Sigma x_i / n = \bar{x} \Rightarrow \hat{\mu}_{mle} = \mu$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{\partial (2l)}{\partial \sigma^2} = -\Sigma (x_i - \mu)^2 \cdot (\sigma^2)^{-2} (-1) - \frac{n}{\sigma^2} = 0$$

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n} = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$\sigma^2 = \Sigma(x_i - \bar{x})^2 / n \Rightarrow \hat{\sigma}^2_{mle} = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

Note: $E(\bar{x}) = \mu$

$Var(\bar{x}) = \frac{\sigma^2}{n}$

? $E(\hat{\sigma}^2_{mle}) = \frac{1}{n} \left[\frac{n-1}{n} \cdot \sigma^2 \right]$

$S^2 = \sum (x_i - \bar{x})^2 / n-1$

$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1} = \text{Gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right)$

$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1$

$E(S^2) = \sigma^2$; $Var\left(\frac{n-1}{\sigma^2} S^2\right) = 2(n-1)$; $V(S^2) = \left(\frac{\sigma^2}{n-1}\right)^2 \cdot 2(n-1)$

$E(\hat{\sigma}^2_{mle}) = E\left(\frac{n-1}{n} \cdot \frac{\sum(x_i - \bar{x})^2}{n-1}\right) = E\left(\frac{n-1}{n} \cdot S^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$

$\Rightarrow MSE(\hat{\sigma}^2_{mle})$

$= Var(\hat{\sigma}^2_{mle}) + Bias(\hat{\sigma}^2_{mle})^2$

$= \left(\frac{n-1}{n}\right)^2 \cdot \frac{2(\sigma^2)^2}{n-1} + \left(\frac{n-1}{n} \cdot \sigma^2 - \sigma^2\right)^2$

$= \left[\frac{n-1}{n^2} \times 2 + \frac{1}{n^2}\right](\sigma^2)^2$ $\quad 2n-1 / n^2$

$= \left(\frac{n-1}{n^2} \times 2 + \frac{1}{n^2}\right)(\sigma^2)^2 = \frac{3n-2}{n^2} \cdot (\sigma^2)^2$

? $MSE(S^2)$

$= Var(S^2)$ $\quad \frac{3n-2}{n^2} < \frac{2}{n-1}$

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

$MSE(\sigma^2_{mle}) \leq MSE(S^2)$

$\Rightarrow$ Some properties about $S(Q, x)$

$E\{S(Q; x) \mid \theta\} = 0$

$S_i(\theta; x) = \dfrac{\partial \ell(\theta; x)}{\partial \theta_i}$

$E\{S_i(\theta; \theta) \mid \theta\} = 0$

$\mathrm{cov}(S_i(\theta; x), S_j(\theta; x))$

$= E(S_i(\theta; x), S_j(\theta, x))$

$= -E\left(\dfrac{\partial^2 \ell(\theta; x)}{\partial \theta_i \partial \theta_j}\right)$

$I(\theta) = \mathrm{Cov}(S(\theta, x))$

$= \begin{pmatrix} \mathrm{Cov}(S_1, S_1) & \cdots & \mathrm{Cov}(S_1, S_d) \\ \vdots & & \\ \mathrm{Cov}(S_d, S_1) & \cdots & \mathrm{Cov}(S_d, S_d) \end{pmatrix}$

$I(\theta)$ is a function of θ alone fisher in function

$\dot{J}(\theta, x) = \left(-\dfrac{\partial^2}{\partial \theta_k \partial \theta_l} \ell(\theta; x)\right)$

$I(\theta) = E(\dot{J}(\theta; x))$

$J(\theta, x)$ - observed. Fisher information.

<< Mar 20, 2015 >> STAT 846.

$\Rightarrow$ $X_1, \cdots, X_n \mid \mu\ \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

$\hat{\mu}_{mle} = \bar{X}$ $\quad \hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$\hat{\sigma}^2_{mle} = \frac{n-1}{n} \cdot \frac{\Sigma(X_i - \bar{X})^2}{n-1} = \frac{n-1}{n} S^2$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ $\quad S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$

$E(S^2) = \frac{\sigma^2}{n-1} \cdot (n-1) = \sigma^2$

$Var(S^2) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2}{n-1}\sigma^4$

$MSE(S^2) = \frac{2}{n-1}\sigma^4$

$$
\begin{aligned}
MSE(\hat{\sigma}^2_{mle}) &= Var(\hat{\sigma}^2_{mle}) + [Bias(\hat{\sigma}^2_{mle})]^2 \\
&= \left(\frac{n-1}{n}\right)^2 V(S^2) + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\
&= \frac{(n-1)^2}{n^2}\frac{2}{n-1}\sigma^4 + \frac{1}{n^2}\sigma^4 \\
&= \left(\frac{2}{n} - \frac{1}{n^2}\right)\sigma^4 \\
&< \frac{2}{n-1}\sigma^4
\end{aligned}
$$

so, $\hat{\sigma}^2_{mle}$ ~~doro~~ dominates S^2

$\Rightarrow$ $S_i(\theta; x) = \frac{\partial}{\partial\theta} \ell(\theta; x)$

theorem: Given the support of X is free of θ

$E_X(S_i(\theta; x) \mid \theta) = 0$

$$
\begin{aligned}
COV(S_i(\theta; x)\ (S_j(\theta; x)) &= E\left(-\frac{\partial \ell(\theta, x)}{\partial\theta_i} \frac{\partial}{\partial\theta_j}\ell(\theta, x)\right) \\
&= I(\theta)
\end{aligned}
$$

proof:

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta, x) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(x;\theta))$$

$$= \frac{\partial}{\partial \theta_j}\left(\frac{\frac{\partial}{\partial \theta_j} f(x|\theta)}{f(x|\theta)} \right)$$

$$= \frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x|\theta)\cdot f(x|\theta) - \frac{\partial}{\partial \theta_i} f(x|\theta)\cdot \frac{\partial}{\partial \theta_j} f(x|\theta)}{f^2(x|\theta)}$$

For the R.H.S

$$= E\left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta, x)\right)$$

$$= -\left[\int \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x|\theta)\, dx - \int \frac{\partial}{\partial \theta_i} f \cdot \frac{\partial}{\partial \theta_j} f \cdot \frac{1}{f(x|\theta)}\, dx \right.$$

$$= \int \frac{\partial f}{\partial \theta_i} \cdot \frac{\partial f}{\partial \theta_j} \frac{1}{f(x|\theta)}\, dx$$

For the L. H. S.

$$= Cov(S_i, S_j)$$

$$= E(S_i, S_j) = \int \frac{\partial}{\partial \theta_i} l(\theta, x) \cdot \frac{\partial}{\partial \theta_j} l(\theta, x) \cdot f(x|\theta)\, dx$$

$$= \int \frac{\partial}{\partial \theta_i} f \cdot \frac{\partial f}{\partial \theta_j} \cdot \frac{1}{f(x|\theta)}\, dx.$$

$\Rightarrow$ Remark: if $X_1, \dots X_n | \theta \overset{iid}{\sim} f(X_i|\theta)$

Let X to be $X = (X_1, \dots X_n)$

$$f(x;\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$l(\theta; x) = \Sigma \log f(x_i|\theta) = \Sigma\, l(\theta; x_i)$$

$$S_K(\theta; x) = \frac{\partial}{\partial \theta_K} \ell(\theta; x)$$

$$= \frac{\partial}{\partial \theta_K} \Sigma\, \ell(\theta; x_i) = \Sigma \frac{\partial}{\partial \theta_K} \ell(\theta; x_i)$$

$I^{K,K}(\theta)$

$\Rightarrow Var(S_K(\theta; x)) = E\left(\Sigma \frac{\partial^2}{\partial \theta_K^2} \ell(\theta, x_i)\right)$

$= \sum_{i=1}^{n} E\left(-\frac{\partial^2}{\partial \theta_K^2} \ell(\theta; x_i)\right) = \sum_{i=1}^{n} I_1^{K,K}(\theta) = n\, I_1(\theta)$

where, $I_1^{K,K}(\theta) = E\left(-\frac{\partial^2}{\partial \theta_K^2} \ell(\theta, x_i)\right)$

⇒ **Cramer - Rao Lower bound (CRLB)**

suppose the support of X, Let θ be scalar.

Thw: ^ $W(X)$ is any estimator,

Let $m(\theta) = E(W(X))$, then

$$Var(W(X)) \geq \frac{m'(\theta)}{I(\theta)}$$

where $I(\theta) = E_X\left(-\frac{\partial^2}{\partial \theta^2} \ell(\theta; x)\right)$

⇒ particularly case Let $m(\theta) = \theta$

For any unbiased estimator of θ, $W(X)$

$Var(W(x)) \geq \frac{1}{I(\theta)}$.

Note that. $\frac{m'(\theta)}{I(\theta)}$ is called CRLB.

⇒ proof: Let $Z(X) = S(\theta; X) = \frac{\partial}{\partial \theta} \ell(\theta; X)$

We know that

$|cov(X, Y)|^2 \leq Var(X) \cdot Var(Y)$

$Cov(W(X), Z(X)) \leq Var(W(X)) \cdot Var(Z(X))$

$V(Z(x)) = V(S(\theta;x)) = E\left(\frac{\partial}{\partial\theta} \ell(\theta,x)\right) = I(\theta)$

$Cov(W(x), S(\theta;x))$

$= E(W(x) \cdot S(\theta;x))$ $\quad$ (b/c $E(S) = 0$)

$E(W(x)) \cdot E(S(\theta;x)) = 0$

$= \int W(x) \cdot S(\theta;x)\, dx$

$= \int W(x) \frac{\partial f(x;\theta)}{\partial\theta} \cdot \frac{1}{f(x;\theta)} f(x;\theta)\, dx$

$= \int W(x) \frac{\partial f(x;\theta)}{\partial\theta}\, dx$

$= \frac{\partial}{\partial\theta}\left\{\int W(x) \cdot f(x;\theta)\, dx\right\}$

$= \frac{\partial}{\partial\theta}(m(\theta)) = m'(\theta)$

therefore

$[m'(\theta)] \leq Var(W(x)) \cdot I(\theta)$

$\Rightarrow Var(W(x)) \geq \frac{m'(\theta)}{I(\theta)}$

$\Rightarrow$ example: $x_1 \cdots x_n \overset{iid}{\sim} exp(\theta)$

where θ is the scaler of x_i

Let $x = (x_1, \cdots x_n)$

$f(x;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}}$

$\ell(\theta;x) = -n \cdot \log\theta - \frac{1}{\theta} \Sigma x_i$

$S(\theta;x) = \frac{\Sigma x_i}{\theta^2} - \frac{n}{\theta} = 0 \Leftrightarrow \theta = \bar{x}$

$\hat{\theta}_{mle} = \bar{x}$

$I(\theta) = Var\left(\frac{\Sigma x_i}{\theta^2} - \frac{n}{\theta}\right) = \frac{1}{\theta^4} n \cdot Var(x_i) \cdot \theta^2 = \frac{n}{\theta^2}$

$$S'(\theta; x) = \frac{n}{\theta^2} - \frac{2\sum x_i}{\theta^3}$$

$$E(-S'(\theta; X)) = \frac{2}{n^3} \cdot n\theta - \frac{n}{\theta^2} = \frac{n}{\theta^2}$$

mLe $E(\bar{X}) = \theta$

$$Var(\hat{\theta}_{mle}) = Var(\bar{X}) = \frac{1}{n} Var(X_i) = \frac{1}{n}\theta^2$$

Here $Var(\hat{\theta}_{mle}) = \frac{1}{I(\theta)}$

$$CRLB = \frac{1}{I(\theta)}$$

<< Mar 23. 2015 >>   STAT 846.

=> Asympotic properties of MLE

$\hat{\theta}_n(x_1, \cdots, x_n) \to \theta$

$f_{\hat{\theta}}(x) \xrightarrow{dis.} f(x)$

☆ $\boxed{\hat{\theta}_n \to N(\theta, \frac{1}{I(\theta)})}$ where $I(\theta) = Var(S(\theta(x)))$

$= -E(\underbrace{\nabla \ell(\theta, x)}_{d \times d \text{ matrix}})$

=> Review :

1) $X_n \xrightarrow{prob.} X \iff \forall \varepsilon > 0 \; P(|X_n - X| > \varepsilon) \to 0$

2) $X_n \xrightarrow{distri.} X$; $F_{X_n}(x) = P(X_n \le x)$

$F_X(x) = P(X \le x)$

$X_n \xrightarrow{dis.} X \iff F_{X_n}(x) \to F_X(x)$, the continuity point of $F_{X_n}(x)$

3). Central Limit theorem and Large Number Law.

this is weak

LLN: $X_1, \cdots X_n | \theta \overset{iid}{\sim} f(x; \theta)$

$E(X_i) < +\infty$ then $\bar{X} = \sum X_i / n \xrightarrow{P/d} E(X_i)$

this is stronger

4) CLT: Given $X_1, \cdots, X_n | \theta \overset{iid}{\sim} f(x | \theta)$

$E(X_i) = \mu < \infty$

$Var(X_i) = \sigma^2 < \infty$

Then, $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

$\Rightarrow$ slutsky's theorem:

if $X_n \xrightarrow{d} X$

$Y_n \xrightarrow{d} a$

a is a non-Random Number without assuming any relationship btw X_n and Y_n; then

$g(X_n, Y_n) \xrightarrow{d} g(X, a)$

In particular:

$X_n + Y_n \rightarrow X + a$

$X_n \cdot Y_n \rightarrow X \cdot a$

$X_n / Y_n \rightarrow X/a \quad (a \neq 0)$

$X_n + \log(Y_n) \xrightarrow{d} X + \log(a)$

Example:

$X_1, \cdots X_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

$\hat{\mu} = \bar{X} \quad ; \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

proof: By LLN $\hat{\mu} = \bar{x} \xrightarrow{P} E(X_i) = \mu$

$\hat{\sigma}^2 = \frac{1}{n}(\Sigma X_i^2 - n\bar{x}^2) = \frac{\Sigma X_i^2}{n} - \bar{x}^2$

By LLN (Law Learge Number)

$\Sigma X_i^2 / n \xrightarrow{P} E(X_i^2) = \sigma^2 + \mu^2$

$\bar{X} \xrightarrow{P} \mu$

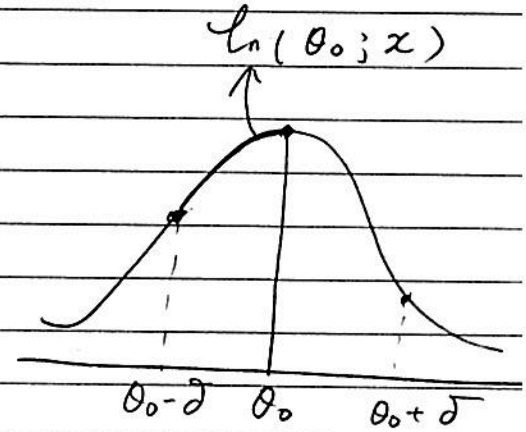
By slutsky's theorem.

$\hat{\sigma}^2 \rightarrow \sigma^2 + \mu^2 - \mu^2 = \sigma^2$

$\Rightarrow$ Thm: Given $x_1 \cdots x_n \mid \theta \overset{iid}{\sim} f(x \mid \theta)$

$\hat{\theta}_n$ is the MLE biased on $x_1, \cdots, x_n$

$\hat{\theta}_n \xrightarrow{P} \theta$



proof: Let $X = (x_1, \cdots x_n)$

$l_n(\theta; x) = \sum_{i=1}^{n} \log(f(x_i \mid \theta))$

For any fixed θ

$l_n(\theta; x)$ is a R.V

indeed is a sample mean of $Z_i = l(\theta; x_i)$
$= \log[f(x_i \mid \theta)]$

we will show that

$[l_n(\theta_0 - \delta) - l_n(\theta_0))] / n \to \mu_1 < 0$

$[l_n(\theta_0 + \delta) - l_n(\theta_0)] / n \to \mu_2 < 0$

proof: $\forall\ \theta \neq \theta_0$

$[l(\theta; x_n) - l_n(\theta_n; x_n)] / n$

$= [\Sigma[l(\theta; x_i) - l(\theta_0; x_i)]] / n$

$\xrightarrow{LLN} E[[l(\theta; x_i) - l(\theta_0; x_i)] \mid \theta]$

$E\left(\log \frac{f(x_i \mid \theta)}{f(x_i \mid \theta_0)} \Big| \theta_0\right) \leq \log E\left(\frac{f(x_i \mid \theta)}{f(x_i \mid \theta_0)} \Big| \theta\right)$

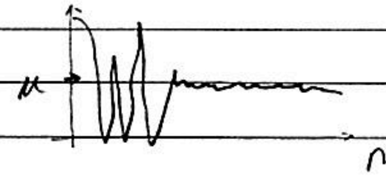
$= \log\left\{\int \frac{f(x \mid \theta)}{f(x \mid \theta_0)} \cdot f(x \mid \theta_0)\, dx\right) = \log \int f(x \mid \theta)\, dx)$
$= \log 1 = 0$

<< Mar. 25 2015 >> STAT 846

=> Review:

1) LLN (Law large Number)

$X_1 \cdots x_n \mid \theta \overset{iid}{\sim} f(x \mid \theta)$



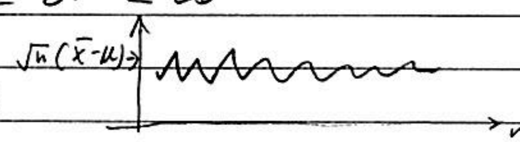
$E(x_i) = \mu < +\infty \qquad \bar{X}_n \xrightarrow{d/p} \mu$

2) CLT:

$X_1, \cdots x_n \mid \theta \overset{iid}{\sim} f(x \mid \theta)$

$E(x_i) = \mu < \infty \qquad V(x_i) = \sigma^2 < \infty$

$\sqrt{n}\,(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$



3). Slutsky's theorem:

Given $X_n \xrightarrow{d} X \qquad Y_n \xrightarrow{d} C$

then: $X_n + Y_n \xrightarrow{d} X + C$

$X_n \cdot Y_n \xrightarrow{d} X \cdot C$

$X_n / Y_n \xrightarrow{d} X / C$

=> example: $X_1 \cdots x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \Sigma (x_i - \bar{x})^2 / n \end{cases}$$

$\hat{\mu} \xrightarrow{d/p} \mu$ by (LLN)

$\hat{\sigma}^2 = \Sigma (x_i - \mu - (\bar{x} - \mu))^2 / n$

$= \Sigma (x_i - \mu)^2 / n - (\bar{x} - \mu)^2$

$\qquad\downarrow \qquad\qquad\qquad \downarrow$

$E((x_i - \mu)^2) = \sigma^2 \qquad (\mu - \mu)^2$

$\sqrt{n}\,(\bar{X}-\mu) \xrightarrow{d} N(0,\sigma^2)$ by CLT.

$\sqrt{n}\,(\hat{\sigma}^2-\sigma^2)$

$= \sqrt{n}\,\left(\Sigma (X_i-\mu)^2/n - (\bar{X}-\mu)^2 - \sigma^2 \right)$

$= \underbrace{\sqrt{n}\left(\frac{\Sigma E[(X_i-\mu)^2]}{n} - \sigma^2 \right)}_{\text{viewing } (X_i-\mu)^2 \text{ as } X_i \text{ in CLT}} - \sqrt{n}\,(\bar{X}-\mu)^2$

$= N(0, V[(X_i-\mu)^2]) - \frac{1}{\sqrt{n}}\,(\sqrt{n}\,(\bar{X}-\mu)]^2$

$\because [\sqrt{n}\,(\bar{X}-\mu)]^2 \xrightarrow{d} [N(0,\sigma^2)]^2$

$\frac{1}{\sqrt{n}} \xrightarrow{p/d} 0$

$\Rightarrow \sqrt{n}\,(\bar{X}-\mu)^2 \to 0$

By slusky theorem:

$$\boxed{\sqrt{n}\,(\hat{\sigma}^2-\sigma^2) \to N(0, 2\sigma^4)}$$

$Y_i = X_i - \mu \sim N(0,\sigma^2)$; $Z_i \sim N(0,1)$

$V(Y_i^2) = V(\sigma^2, Z_i^2)$

$= \sigma^4 \left(E(Z_i^4) - (E(Z_i^2))^2 \right) = \sigma^4 (3-1)$

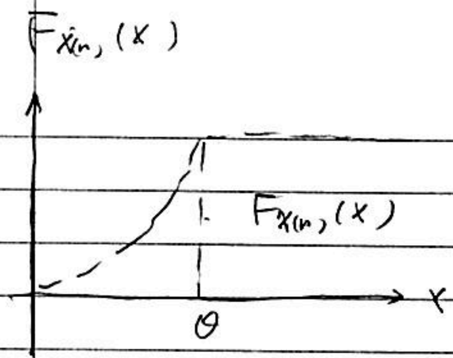
In formally we say

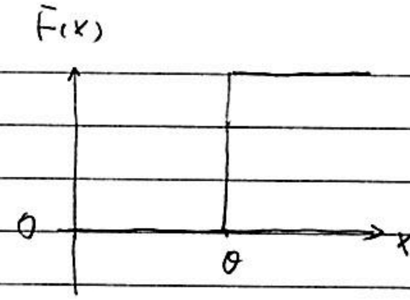
$\hat{\sigma}^2 \overset{\text{approximation}}{\sim} N(\sigma^2, \frac{2\sigma^4}{n})$

$\Rightarrow$ example: $X_1 \cdots X_n | \theta \sim \text{unif}(0, \theta)$

$\hat{\theta} = X_{(n)} = \max(X_1, \cdots, X_n)$

$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(\text{all } X_i \le x) = \left(\frac{x}{n}\right)^n$





~~For $0 < x \le 1$~~

~~F(x) is the CDF of~~

$F_{X_{(n)}}(x) \to 0$

X	θ
prob.	1

~~For $x \geq 1$~~

$F_{X_{(n)}}(x) \to 1$

$\hat{\theta} \xrightarrow{d} \theta$

$Y_n = \sqrt{n}(\hat{\theta} - \theta)$

$F_{Y_n}(y) = P(\sqrt{n}(\hat{\theta} - \theta) \le y) = P\left(\hat{\theta} \le \frac{y}{\sqrt{n}} + \theta\right)$

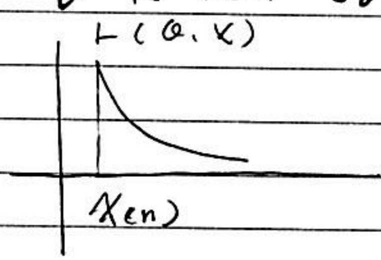
$= \left(\frac{y}{\sqrt{n}} + \theta\right)^n$ for $y/\sqrt{n} + \theta < \theta$

$\to \infty$

$\Rightarrow$ why $\sqrt{n}(\hat{\theta} - \theta)$ doesn't converge to $N(\ ;\ )$

1) the suppose of X_i depend on θ

2) $\hat{\theta}$ is not obtained at $\ell'(\theta) = 0$



$\Rightarrow$ Asymptotic Normality of MLE conditions:

1) the support of $f(x|\theta)$ doesn't depend on θ

2) $\ell'(\hat{\theta}_n) = 0$

~~3) ℓ is twice cont. differentiable at each θ~~

4) $|\ell'''(\theta; x)| \le g(x)$ for each θ, x.

$\exists$ a $g(x)$ with $E(g(x)) < +\infty$

then we have

$$\sqrt{n}(\hat{\theta}_{mle} - \theta) \xrightarrow{d} (0 \quad \frac{1}{I_1(\theta)})$$

where, $I_1(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ell(\theta, x_i)\right) = V\left(\frac{\partial}{\partial\theta}\ell(\theta, x_i)\right)$

That is $\hat{\theta}_{mle} \overset{approx}{\sim} N(\theta, \frac{1}{nI_1(\theta)})$

$\Rightarrow$ proof:

suppose θ_0 is the true parameter, $\hat{\theta}_n$ is the MLE

$x = (x_1, \cdots x_n)$ (Talayur extention)

$$\ell'(\theta, x) = \ell'(\hat{\theta}_n; x) + (\theta - \hat{\theta}_n)\ell''(\hat{\theta}_n^*)$$

where $\hat{\theta}_n^*(x)$ lies between θ_0 and $\hat{\theta}_n$

$$\ell'(\theta_0, x) = \ell'(\hat{\theta}_n; x) + (\theta_0 - \hat{\theta}_n)\ell''(\hat{\theta}_n^*)$$

$\ell'(\hat{\theta}_n) = 0$

$$\ell'(\theta_0, x) = (\theta_0 - \hat{\theta}_n)\ell''(\hat{\theta}_n^*) \Rightarrow \theta_0 - \hat{\theta}_n = \frac{\ell'(\theta_0, x)}{\ell''(\hat{\theta}_n^*)}$$

$$\sqrt{nI_1(\theta_0)}(\hat{\theta}_n - \theta_0) = -\frac{\ell'(\theta_0; x)}{\ell''(\hat{\theta}_n^*)} \cdot \sqrt{nI_1(\theta_0)}$$

$$= \left(\frac{\ell'(\theta_0; x)}{\sqrt{nI_1(\theta_0)}}\right) \cdot \left(\frac{\ell''(\theta_0)}{\ell''(\hat{\theta}_n^*)}\right)\left(-\frac{nI_1(\theta_0)}{\ell''(\theta_0)}\right)$$

We will show

$$C_1: \frac{\ell'(\theta_0, x)}{\sqrt{nI_1(\theta)}} \to N(0, 1)$$

$$C_2 : \quad \frac{-\ell''(\theta_0)}{-\ell''(\hat{\theta}_n^*)} \to 1$$

$$\sqrt{\frac{n}{\sigma^2}} \left(\bar{X} - \mu\right) \to N(0,1)$$

$$C_3 : \quad \frac{n I_1(\theta_0)}{\ell''(\theta_0)} \to 1$$

$\Rightarrow$ proof $C_1 \to N(0,1)$

$$\sqrt{n I_1^{-1}(\theta_0)} \left(\frac{\sum \ell'(\theta_0; x_i)}{n} \right) \xrightarrow{CLT} N(0,1)$$

because $E(\ell'(\theta_0; x_i)) = 0$; $Var(\ell'(\theta_0; x_i)) = I_1(\theta_0)$

$\Rightarrow$ proof C_3:

$$\frac{\frac{1}{n} \sum_{i=1}^{n} (\ell''(\theta_0; x_i))}{I_1(\theta)} \to \frac{E(-\ell''(\theta_0; x_i))}{I_1(\theta_0)}$$

$$\to 1$$

<< Mar 27. 2015 >> STAT 846

=> Asymptotic Normality under some regularity condition

$$\sqrt{n}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta))$$

$$\sqrt{n}\,(I(\theta_0))^{\frac{1}{2}}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} Nor\left[\underset{p\times p}{0}, \underset{p\times p}{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}}\right]$$

=> A generalization

$$\sqrt{n}\,(\hat{I}(\hat{\theta}_{mle}))^{\frac{1}{2}}\,(\hat{\theta}_{mle} - \theta_0) \to N(\underset{p\times p}{0}, I_p)$$

where, $\hat{I}(\hat{\theta}_{mle}) = -\sum_{i=1}^{n} \underbrace{\nabla_\theta^2 \, l(\hat{\theta}_{mle}; x_i)}_{p\times p} / n \xrightarrow{p/d} I(\theta_0)$ by LLN

note $I(\theta) = E(-\nabla_\theta^2 \, l(\theta; x))$

=> proof:

assume $p=1$.

$$\sqrt{n \cdot \hat{I}(\hat{\theta}_{mle})}\,(\hat{\theta}_{mle} - \theta_0) = \sqrt{n I(\theta_0)}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N$$

$$\times \sqrt{\frac{\hat{I}(\hat{\theta}_{mle})}{I(\theta_0)}} \xrightarrow{p/d} 1$$

$\xrightarrow{d} N(0, 1)$ By Slutsky's theorem

=> example:

$X_1, \cdots x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

we have shown directly

$$\sqrt{n}\,(\hat{\sigma}^2_{mle} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$$

Now we apply our theorem

$$L(\theta; x_i) = (\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\ell(\theta; x_i) = -\frac{1}{2}\log\sigma^2 - \frac{(x_i-\mu)^2}{2\sigma^2}$$

Score function : $\frac{\partial}{\partial\sigma^2}\ell(\theta; x_i) = -\frac{1}{2}\cdot\frac{1}{\sigma^2} + \frac{(x_i-\mu)^2}{2(\sigma^2)^2}$

$$\frac{\partial}{\partial\mu}\ell(\theta; x_i) = \frac{2(x-\mu)}{2\sigma^2}$$

Hessian function : $\frac{\partial^2}{\partial\mu\partial\sigma^2}\ell = -\frac{(x_i-\mu)}{(\sigma^2)^2}$

$$\frac{\partial}{\partial(\sigma^2)^2}\ell = \frac{1}{2}\frac{1}{(\sigma^2)^2} - \frac{(x_i-\mu)^2}{(\sigma^2)^3}$$

$$\frac{\partial^2\ell}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

$$I(\theta) = -\begin{bmatrix} E\frac{\partial^2\ell}{\partial\mu^2} & E\frac{\partial^2\ell}{\partial\mu\partial\sigma^2} \\ E\frac{\partial^2\ell}{\partial\mu\partial\sigma^2} & E\frac{\partial^2\ell}{\partial(\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{bmatrix}$$

By Asymptotic Normality, we conclude :

$$\left[\begin{pmatrix} \hat{\mu}_{mle} \\ \hat{\sigma}^2_{mle} \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}\right]\sqrt{n} \xrightarrow{d} N\left(\underset{2\times 2}{0}, \underset{2\times 2}{\begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix}}\right)$$

⇒ $\boxed{\text{Three Asymptotic properties of MLE}}$

1) Consistency : $\hat{\theta}_{mle} \xrightarrow{p/d} \theta_0$

2) Normality : $\sqrt{n}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$

3) efficiency : $n\,Var(\hat{\theta}_{mle}) \to I^{-1}(\theta_0)$

⇒ likelihood Ratio test:

suppose that : $\theta = (\theta_1, \cdots \theta_p)$

Null hypothesis $H_0 : \theta_1 = \theta_1^0, \cdots, \theta_m = \theta_m^0$

alternative H_1 : no restriction

⇒ example1 (ANOVA)

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$ OR $H_0 : \theta_i = \mu_i - \mu_1 = 0$

$i = 2, \cdots p.$

⇒ example 2 : $X_1, \cdots, x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$\theta = (\mu, \sigma^2)$

$H_0 : \mu = \mu_0$ σ^2 can be anything

$H_1 : \mu \neq \mu_0$ σ^2 ———————.

point null hypothesis

⇒ Defination:

Likelihood Ratio Test

$$\Lambda = \max_{\theta \in H} f(x;\theta) \Big/ \max_{\theta \in H_0} f(x;\theta)$$

$$= f(x;\hat{\theta}_{mle1}) \Big/ f(x;\hat{\theta}_{mle0})$$

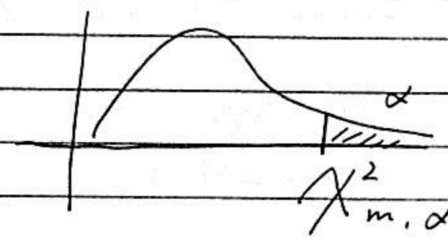
⇒ Wilk's theorem:

$$-2\log(\Lambda)\Big|\theta\in H_0 \xrightarrow{d} \chi^2_m$$

m = reduction of dimensions of θ from H_1 to H_0
$= \mathrm{Dim}(H_1) - \mathrm{Dim}(H_0)$

$\mathrm{Dim}(H_0) = p-m.$ $\mathrm{Dim}(H_1) = p$

Rejection Region of size α, $-2\log(\Lambda) > \chi^2_{m,\alpha}$.



⇒ example:

$x_1, \dots, x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

Given $\theta \in H_0$, $f(x;\theta) = \frac{1}{(\sqrt{2\pi})^n} \cdot \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$

$\mu = \mu_0$

$$\hat{\sigma}_0^2 = \frac{\Sigma(x_i-\mu_0)^2}{n}.$$

$$\Rightarrow f(x|\hat{\theta}_0) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{(\hat{\sigma}_0^2)^{\frac{n}{2}}} e^{-\frac{\Sigma(x_i-\mu_0)^2}{2\hat{\sigma}_0^2}} = e^{-\frac{n}{2}}$$

Given $\theta \in H_1$ $\hat{\mu}_1 = \bar{x}$

$$\hat{\sigma}_1^2 = \Sigma(x_i-\bar{x})^2/n$$

$$\Rightarrow f(x|\hat{\theta}_1) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{(\hat{\sigma}_1^2)^{\frac{n}{2}}} e^{-\frac{n}{2}}$$

so $\Lambda = f(x \mid \hat{\theta}_1) \big/ f(x \mid \hat{\theta}_0) = \dfrac{(\hat{\sigma}_0^2)^{\frac{n}{2}}}{(\hat{\sigma}_1^2)^{\frac{n}{2}}}$

$$= \left[\frac{\Sigma (x_i - \mu_0)^2}{\Sigma (x_i - \bar{x})^2} \right]^{\frac{n}{2}} = \left[\frac{\Sigma (x_i - \bar{x} + \bar{x} - \mu_0)^2}{\Sigma (x_i - \bar{x})^2} \right]^{n/2}$$

$$= \left[1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2} \right]^{n/2}$$

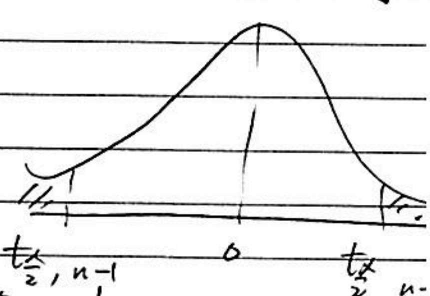
look at Λ directly

$$\Lambda > \lambda \iff \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2/n} > K/n$$

$$\boxed{\frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2/n} \text{ is T-test } \left(\because t = \frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}} \right)}$$

$$\iff \left| t = \frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}} \right| > t^*$$

To determine t^* we use.



$$\frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1} \Rightarrow t^* = t_{\frac{\alpha}{2}, n-1}$$

<< Mar 30, 2015 >> STAT 846

Schudule:

① No class 3rd Apr.

② Review on 6th Apr.

③ Test 2 on 8th Apr.

④ Final on 14th Apr.

⑤ Tuesday 9-12 AGRI 2D

=> Wilk Theorem:

$x_1, \cdots x_n \mid \theta \overset{iid}{\sim} f(x;\theta)$; $\theta = (\theta_1, \cdots, \theta_p)$

$H_0 : \theta_1 = \theta_1^0, \cdots, \theta_m = \theta_m^0$

H_1 : no restriction for θ

$$\Lambda = \max_{\theta \in H_1} f(x;\theta) \Big/ \max_{\theta \in H_0} f(x;\theta)$$

$$2\log\Lambda \mid \theta \in H_0 \xrightarrow{d} \chi^2_m$$

m = reduction of dimension of θ from H_1 to H_0

=> example. (ANOVA). $x_{ij} \mid \mu_i \sigma^2 \overset{iid}{\sim} N(\mu_i, \sigma^2)$, for $i = 1, 2, \cdots K$

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$

H_1 : No restriction.

$m = \dim(H_1) - \dim(H_0)$

$= (K+1) - (1+1) = K-1$

Let $\theta_1 = \mu_1$

$\theta_2 = \mu_2 - \mu_1$

$\vdots$

$\theta_k = \mu_k - \mu_1$

$H_0: \ \theta_2 = 0, \ \theta_3 = 0, \ \cdots \ \theta_k = 0$

$H_1:$ no - restriction

$m = k - 1$

=> example:

$$x_1, \cdots x_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

$H_0: \quad \mu = \mu_0$

$H_1: \quad \mu \neq \mu_0$

$$\Lambda = \left[\frac{\sum_{i=1}^{n} (x_i - \mu_0)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]^{\frac{n}{2}}$$

$$\Lambda = \left[1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2} \right]^{\frac{n}{2}} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

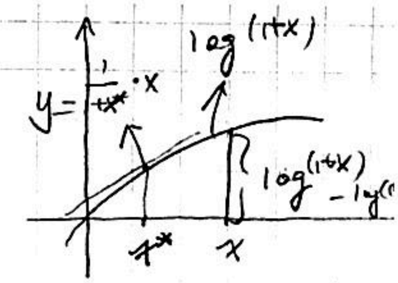
Rejection Region give size α.

$$\{ \Lambda > \lambda_\alpha \} \iff \left| \frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}} \right| > t_{\frac{\alpha}{2}, n-1}$$

$t_{\frac{\alpha}{2}, n-1}$ is determined by α and

the fact: $\dfrac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}} \Big| \mu = \mu_0 \sim t_{n-1}$

An illustration of Wilk's Theorem



- $$2\log(\Lambda) = n\log\left[1+\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2}\right]$$

$$\log(1+x) = \log(1+0) + (x^*-0)\left.\frac{\partial\log(1+x)}{\partial x}\right|_{x=x^*} = 0 + x\cdot\frac{1}{1+x^*}$$

$$= x\cdot\frac{1}{1+x^*} \qquad (0<x^*<x)$$

$$2\log\Lambda = n\cdot\log\left[1+\left(\frac{\bar{x}-\mu_0}{\hat{\sigma}}\right)^2\right]$$

$$= n\cdot\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2}\cdot\frac{1}{1+T^*} \quad \text{for some}$$

$$0 < T^* < \frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2}$$ (which is x is mean value theorem)

- $$2\log\Lambda = \frac{[\sqrt{n}(\bar{x}-\mu_0)]^2}{\sigma^2}\cdot\frac{\sigma^2}{\hat{\sigma}^2}\cdot\frac{1}{1+T^*}$$

$$\begin{array}{ccc} \downarrow \text{CLT} & \downarrow & \downarrow \\ [N(0,1)]^2 & 1 & 1 \end{array}$$

$$\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2} = \frac{(\bar{x}-\mu_0)^2}{\sigma^2}\cdot\frac{\sigma^2}{\hat{\sigma}^2}$$

$$\begin{array}{cc} \downarrow \text{LNN} & \downarrow \leftarrow \text{(converge)} \\ 0 & 1 \end{array}$$

$$B/C \quad 0 < T^* < \frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2} \Rightarrow T^* \to 0$$

$$\begin{array}{c} \downarrow \\ 0 \end{array}$$

- By Slutsky's Theorem $2\log\Lambda \xrightarrow{d} Z^2$ where $Z \sim N(0,1)$ that is, $2\log\Lambda \xrightarrow{d} \chi^2_1$

$\Rightarrow$ proof: Wilk's Theorem.

Let $d = m = 1$ for simplicity

θ is a scalar.

$H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$

$\Lambda_n = 2\log(f(x; \hat{\theta}_n) - f(x; \theta_0))$

$\hat{\theta}_n$ is M.L.E of θ under H_1

$\Lambda_n = 2[\ell(\hat{\theta}_n) - \ell(\theta_0)]$ where x is omitted.

$= 2 \; (\hat{\theta}_n - \theta_0)\ell'(\theta_n) + (\hat{\theta}_n - \theta_0)^2 \cdot f''(\theta_n^*) \quad \theta_0 < \theta_n^* < \theta_n$

$= n I_1(\theta_0) (\hat{\theta}_n - \theta_0)^2 \dfrac{\ell''(\theta_0)}{-n I_1(\theta_0)} \cdot \ell$

<< April 01. 2015 >>   STAT 846

Chapter : 7.2 Condition Inference

=> Example : δ indicater 2 meas tools

$\delta = 1$ -> the tool is more precise (e.g. using $n_1 = 90$ Sam

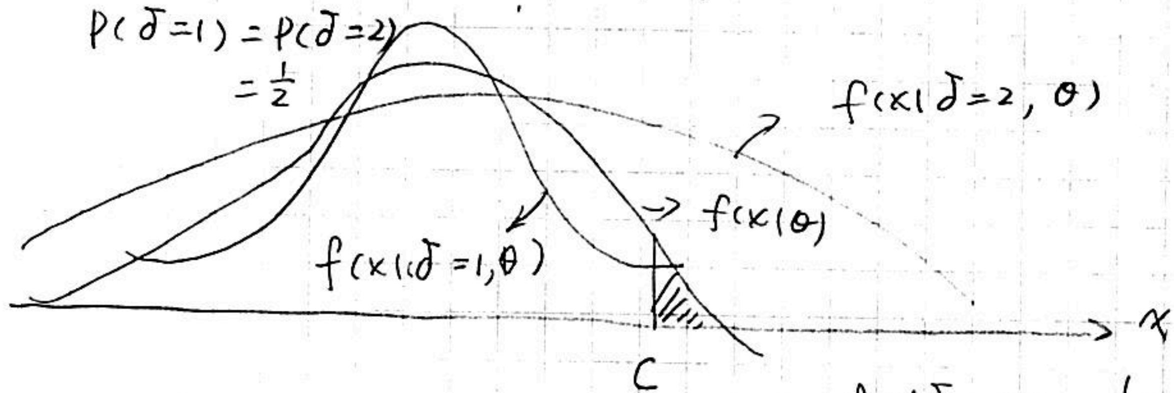
$\delta = 2$ -> the tool is less precise (e.g. using $n_2 = 10$ Sample

Model:

$X \mid \delta = 1 \sim N(\theta, \sigma_1^2)$, $\sigma_1 = 1$

$X \mid \delta = 2 \sim N(\theta, \sigma_2^2)$, $\sigma_2 = 3$

$P(\delta = 1) = P(\delta = 2) = \frac{1}{2}$



$H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$

$$f(x \mid \delta=1, \theta) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\theta)^2}{2\sigma_1^2}}$$

Test statistic: X

$$f(x \mid \delta=2, \theta) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\theta)^2}{2\sigma_2^2}}$$

Rejection Region: $\{ x > c \}$

$$f(x \mid \theta) = \frac{1}{2} f(x \mid \delta=1, \theta) + \frac{1}{2} f(x \mid \delta=2, \theta)$$

We will determine c by $Pr(x > c \mid \theta = 0) = 0.05$

=> Procedure 1 : using $f(x \mid \theta = 0)$
we will solve :

$$Pr(x > c \mid \theta = 0) = 0.05$$

$$\frac{1}{2} P(N(0, \sigma_1^2) > c) + \frac{1}{2} P(N(0, \sigma_2^2) > c) = 0.05$$

=> procedure 2 : observed ; x and δ

using $f(x \mid \delta, \theta)$ to determine C

If $\delta = 1$, $x \mid \theta = 0 \sim N(0, \sigma_1^2)$

$P(x > C_1 \mid \theta = 0) = 0.05$

$\Rightarrow C_1 = 0 + Z_\alpha \times 1 = 1.645$

which distribution we used.

If $\delta = 2$ $\quad x \mid \theta = 0 \sim N(0, \sigma_2^2)$

$P(x > C_2 \mid \theta = 0, \delta = 2) = 0.05$

$\Rightarrow C_2 = 0 + 3 \times Z_\alpha = 3 \times 1.645 = 4.92$

Rejection Region: $\quad x > 0 + Z_\alpha \cdot \sigma_\delta$

what's the problems for procedure 1:

1) ignore δ

2) The justification for C is based on $P(x > c) = \alpha$ without knowing δ.

The α is probability averaging all data sets.

3) C is determined before we observe any data value of x and δ

$\Rightarrow$ Definition: $\theta = (\varphi, \lambda)$

ψ is of our interest

λ is nuisance parameter

$T = (S, C)$ is MSS for (φ, λ)

If (a) the distribution of C doesn't depend on ψ but on λ

(b) the distribution of $S|C$ depends on ψ but not on λ.

Then, we say

C is an ancilliary statistic for ψ and S is conditionally sufficient for ψ given C,

=> example: θ ($\lambda = \phi$) the distribution of δ doesn't depend on θ.

The distribution $X|\delta$ depend on θ,

δ -> called ancilliary (not useless) statistic for θ

=> Conditional principle:

our inference should be conditional on all ancilliary statistic.

=> example 2:

$x_1, \cdots, x_n | \theta \sim unif(\theta - \frac{1}{2}, \theta + \frac{1}{2})$

$f(x|\theta) = I(X_{(n)} - \frac{1}{2} < \theta < X_{(1)} + \frac{1}{2})$

So the $S = (X_{(1)}, X_{(n)})$ is MSS.

$T = \frac{X_{(1)} + X_{(n)}}{2}$ ← Not sufficient

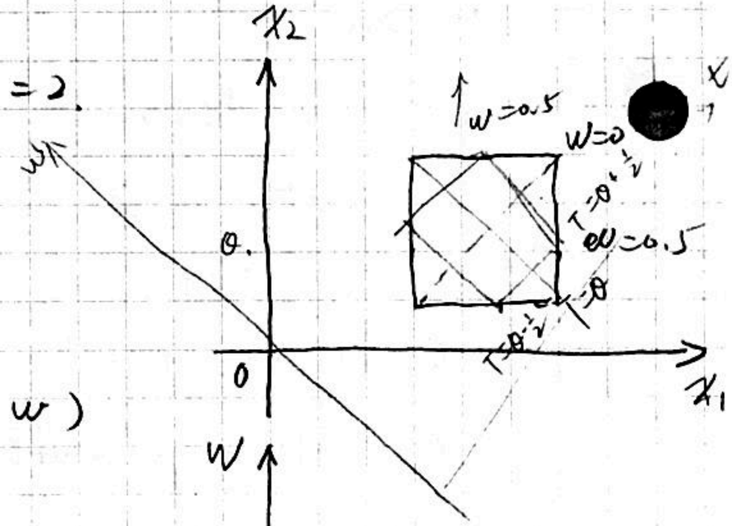
$W = X_{(n)} - X_{(1)}$ ← ancilliary

$f_T(t)$ depend on θ; $f_W(w)$ does not depend on θ

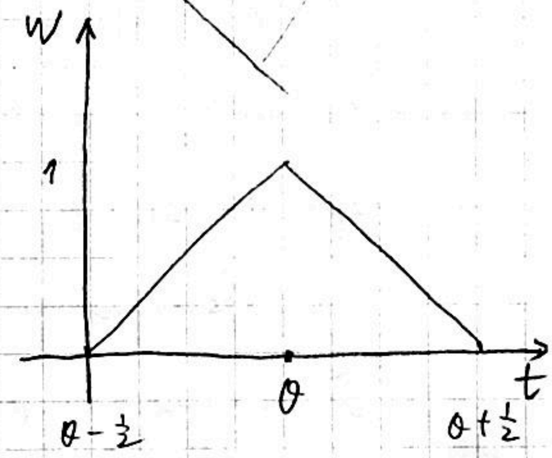
(T, W) is MSS.

$\Rightarrow$ look at a special case $n = 2$.

$$\begin{cases} T = \dfrac{x_1 + x_2}{2} \\ W = |x_1 - x_2| \end{cases}$$



the joint distribution of (T, W)

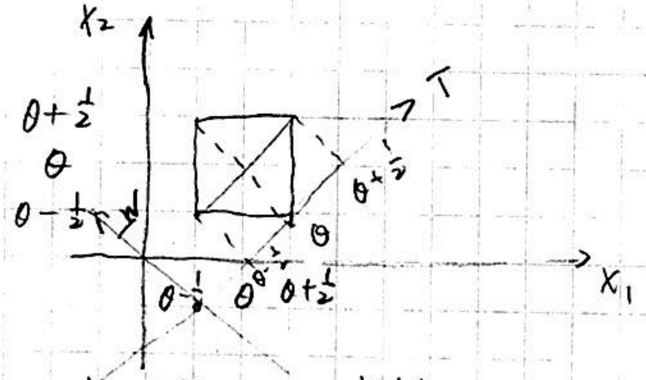


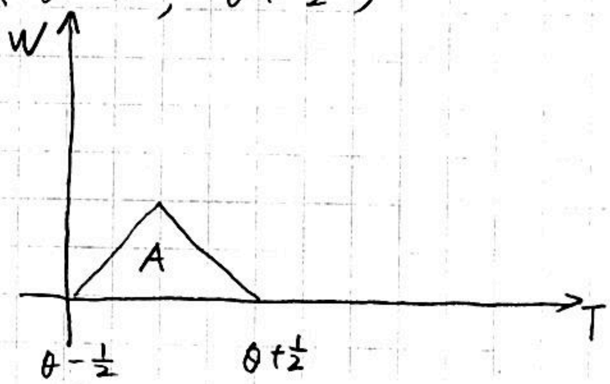
$(T, W) \sim \text{Unif}(A)$

A is the triangle.

<< Mar 6, 2015 >> STAT846:

$\Rightarrow$ example: $X_1, \cdots X_2 \overset{iid}{\sim} unif(\theta - \frac{1}{2}, \theta + \frac{1}{2})$





$(T, W) \sim unif(A)$

$$T = \frac{X_{(1)} + X_{(2)}}{2} = \frac{X_1 + X_2}{2}$$

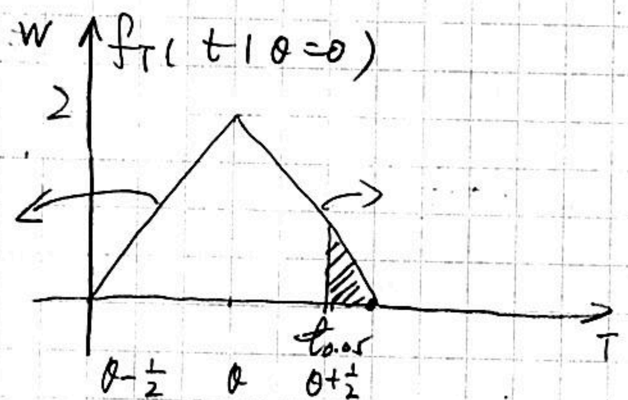
$$W = X_{(2)} - X_{(1)} = |X_{(1)} - X_{(2)}|$$

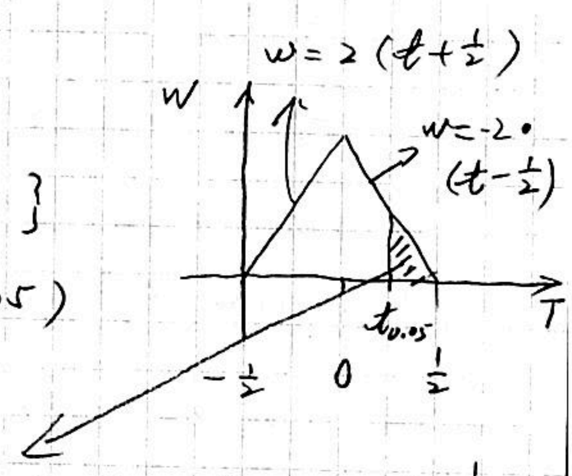
$\Rightarrow$ hypothesis ~~test~~:

$H_0: \theta \le 0$ against $H_1: \theta > 0$

procedure 1: (unconditional)

using marginal distribution of T





Rejection Region: $\{ T > t \}$

$P(T > t_{0.05}) = 0.05 \quad (\alpha = 0.05)$

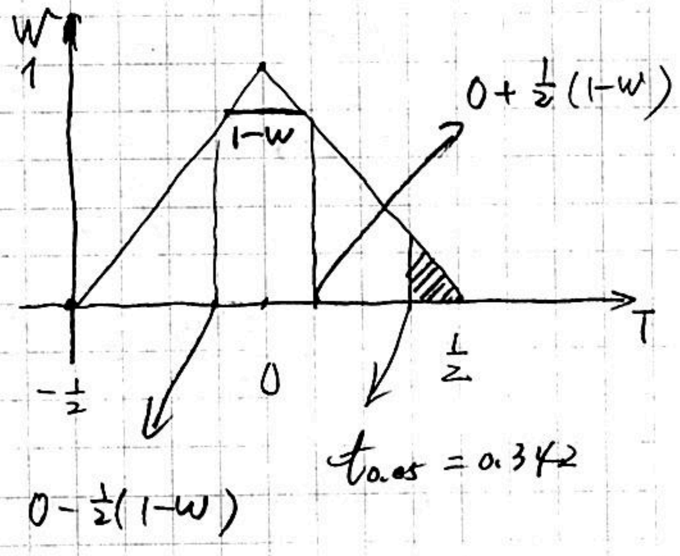
where, $t_{0.05} = 0.342$

$$(\frac{1}{2} - t_{0.05}) \times [-2(t_{0.05} - \frac{1}{2})] \times \frac{1}{2}$$

$$= 0.025$$

procedure 2 : (Conditional inference)

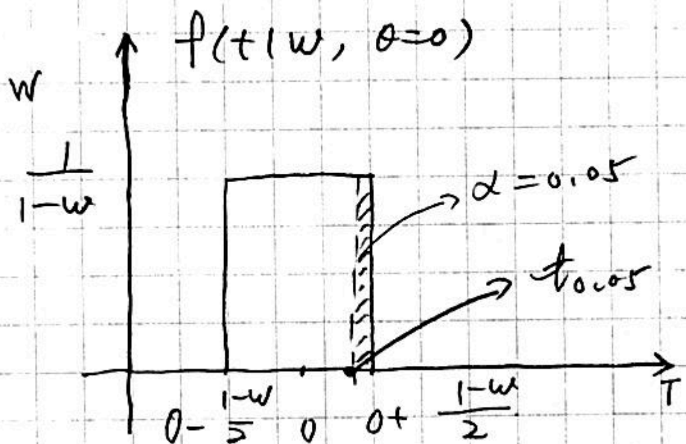
using the distribution $f(t \mid w, \theta)$



$T \mid w, \theta=0 \sim \text{unif}\left(\frac{w}{2}-\frac{1}{2}, \frac{1}{2}-\frac{w}{2}\right)$

$w = 2\left(t+\frac{1}{2}\right) \Rightarrow t = \frac{w}{2} - \frac{1}{2}$.

$w = -2\left(t-\frac{1}{2}\right) \Rightarrow t = \frac{1}{2} - \frac{w}{2}$



$$\left(\frac{1-w}{2} - t^*_{0.05}\right) = 0.05 \times (1-w)$$

$$\frac{1-w}{2} - 0.05(1-w) = t^*_{0.05}$$

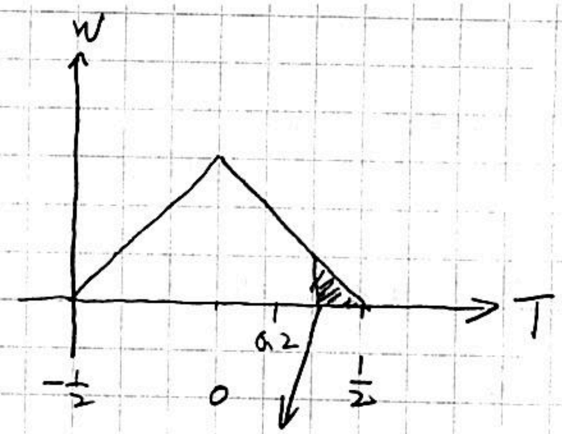
$$t^*_{0.05} = 0.45(1-w)$$

we should reject when:

Rejection Region is : $t > 0.45(1-w) = t^*_{0.05}$

Looking at specific data set : $\begin{cases} t = 0.2 \\ w = 0.7 \end{cases}$

$$\begin{cases} t = 0.2 \\ w = 0.7 \end{cases}$$

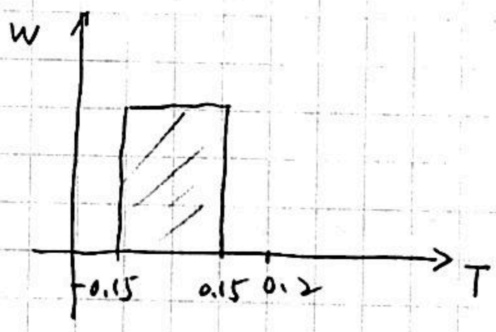


- By the procedure 1
  No reject

$\Rightarrow$ procedure 2:

$T \mid W = 0.7 \;\; \theta = 0 \sim \text{Unif}\left(0 - \frac{0.3}{2},\ 0 + \frac{0.3}{2}\right)$ $\quad t_{0.05} = 0.342$

$= \text{Unif}(-0.15,\ 0.15)$



By procedure 2. We will reject H_0

$t^*_{0.05} = 0.45 \times 0.3 = 0.135$

the procedure 2. is better than procedure 1.

- $\Rightarrow$ the procedure 3 (Bayesian)

$Pr(H_0 \mid X) < C$, we will Reject H_0 (in our case $C = \frac{1}{2}$)

We will compute $Pr(\theta \leq 0 \mid X_1, X_2) = ?$

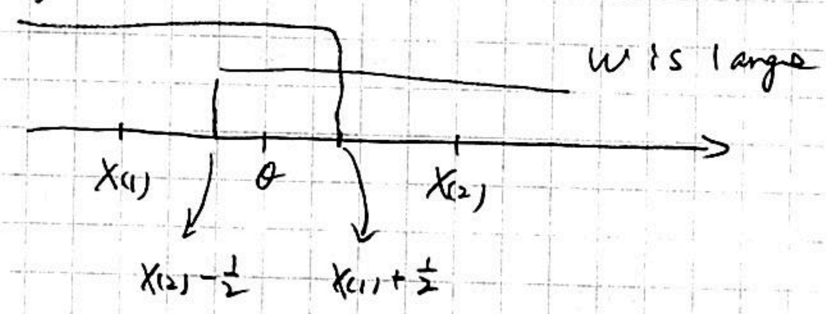
$f(X_1, X_2 \mid \theta) = I\left(X_{(2)} - \frac{1}{2} < \theta < X_{(1)} + \frac{1}{2}\right)$

$\theta \sim I(-M < \theta < M)$

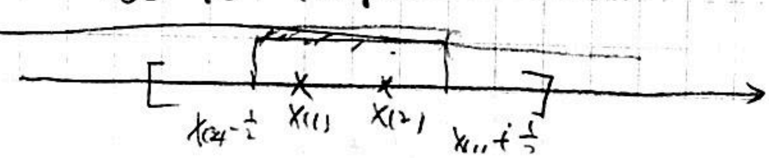
$\theta \mid X_1, X_2 \sim \text{Unif}\left(X_{(2)} - \frac{1}{2},\ X_{(1)} + \frac{1}{2}\right)$

range $(\theta \mid X_1, X_2) = 1 - W$

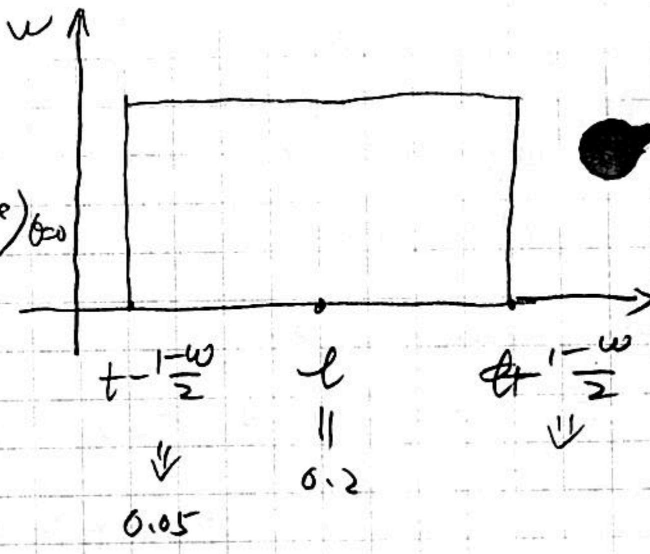
another way.



- $\theta \mid X_1, X_2 \sim \text{unif}\left(t - \frac{1-W}{2},\ t + \frac{1-W}{2}\right)$



Given $\begin{cases} \ell = 0.2 \\ w = 0.7 \end{cases}$



$\Pr(\theta < 0 \mid x_1, x_2) = 0$ (it is impossible)

We reject H_0.

Review:

For test 2:

Concept	Skill
Chapter 4: ✓✓size, ✓power ✓	Find UMP in MLR family
✓UMP	
✓N-P Lemma	
✓MLR.	
Chapter 5: ✓exp. family.	verify exp family
✓natural statistic.	
✓natural parameter	
Chapter 6. ✓MSS.	Find UMVUE with R-B theorem
✓Complete sufficient S.	
✓R-B theorem	$E(d_1(x) \mid T)$
Chapter 8: likelihood function	① Fisher information $I(\theta)$
	② CRLB
	③ Asymptotic distribution of $\hat{\theta}_{MLE}$
including threshold by Wilk's theorem ←	④ likelihood Ratio test

Chapter 2 : Concept . Skill

Loss function.

Risk function

Baye's risk

admissible.

minimax rule

identify all decision rules

Find risk set

Find adm. rule

minimax rule

bayes Rule

Chapter3 :

prior.

posterior

marginal distribution

predictive.

find posterior.

find Baye's Rule

Find Baye's risk.

Find predictive distribution