29/3-21 HW 2 Endre Moen

Ex 1 Give def, explain utility

SS $\overset{\text{book}}{\hat{=}}$ Sufficient Statistics for θ:

A statistics $T(\underline{X})$ is SS for θ if the conditional distribution of the sample $\underline{X}$ given the value of $T(\underline{X})$ doesnt depend on θ

SS $\overset{\text{lecture}}{\hat{=}}$ $T(x)$ is SS for θ if

$$P_\theta(\cdot \mid T(\underline{X}) = x) \quad \forall x \in X$$

is independent of θ

Def independent: A, B independent if:

1) $P(A)P(B) = P(A)P(B)$

or equivelently 2) $P(A \mid B) = P(A)$

3) $P(B \mid A) = P(B)$

The definition is saying: "A statistics, T, is sufficient if you dont have to go back to the data, X, for more information"

Eg knowing the sample mean contains all the information. Going back to the data, $\underline{X}$, will not give more info about the sample mean."

Which is like the saying: $P(\theta \mid T; \text{data}) = P(\theta \mid T)$

while def in book say: $P(\text{data} \mid T; \theta) = P(\underset{\underset{\underline{X}}{\parallel}}{\text{data}} \mid T)$

Because of independence, given T, of: data, $\underline{X}$, and θ - As we have:

$$P(X \mid \theta) = P(X)$$

$$P(\theta \mid X) = P(\theta)$$

The Partition view of SS for θ

Looking at example of SS for θ

$X_1, X_2, X_3 \sim Ber(\theta)$, $T = X_1 + X_2 + X_3$

Partition #	X_1, X_2, X_3	t	$P_\theta(X=x \mid T=t)$
B_1	0 0 0	$t=0$	1
B_2	0 0 1	$t=1$	1/3
	0 1 0		1/3
	1 0 0		1/3
B_3	0 1 1	$t=2$	1/3
	1 0 1		1/3
	1 1 0		1/3
B_4	1 1 1	$t=3$	1

1. A partition is suff. if $f(x \mid X \in B) = P_\theta(X=x \mid T=t)$ is indep of θ
2. A statistics induces a partition $\{x \mid T(x) = t\}$
3. T is suff $\Leftrightarrow$ the partition is sufficient

$\tilde{T} = X_1$ is not SS for θ

B_1	0 0 0	$t=0$	$(1-\theta)^2$
	0 0 1		$\theta(1-\theta)$
	0 1 0		$\theta(1-\theta)$
	0 1 1		θ^2
B_2	1 0 0	$t=1$	$(1-\theta)^2$
	1 1 0		$\theta(1-\theta)$
	1 0 1		$\theta(1-\theta)$
	1 1 1		θ^2

t depends on θ

Knowing t is insufficient

calculated example : $X_i \sim Ber(\theta)$

let $(X_1, \ldots X_n)$ be the rs of $\mathcal{J}$

$$P(X_1 = x_1, \ldots X_n = x_n) = \prod \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i}$$

let $T = \Sigma X_i$, $T \sim Bin(n, p)$

$$P(T = t) = \binom{n}{t} \theta^t (1-\theta)^{n-t}$$

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

$\Rightarrow \Sigma x_i \neq y \Rightarrow P(X_1 = x_1 \ldots \mid Y = y) = 0$

else $\Rightarrow P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

$$\Rightarrow \frac{P(X_1 = x_1, \ldots \text{ And } T = t)}{P(T = t)}$$

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

And $\theta \notin \left\{ \frac{1}{\binom{n}{t}} \right\} \Rightarrow$ SS for θ

An easier method to work with ss for θ is :

<u>Factorisation thm</u> T is ss for θ iff

$$f(\underline{x}) = \phi(T(\underline{x}); \theta) \cdot h(\underline{x})$$

where ϕ dep only on $\underline{X}$ in T and h indep of θ

<u>ex</u> $$f(\underline{x}; \theta) = \underbrace{\theta^t (1-\theta)^{n-t}}_{\phi} \cdot \underbrace{1}_{h}$$

MSS - Minimal SS

MSS $\overset{\text{book}}{\Leftarrow}$ SS for θ is MSS if

$\forall T'$; $T'(\underline{x})$ ss for θ

$\Rightarrow$ $T(x)$ is a function of $T'(x)$

$\Leftrightarrow$ $T'(x) = T'(y) \Leftrightarrow T(x) = T(y)$

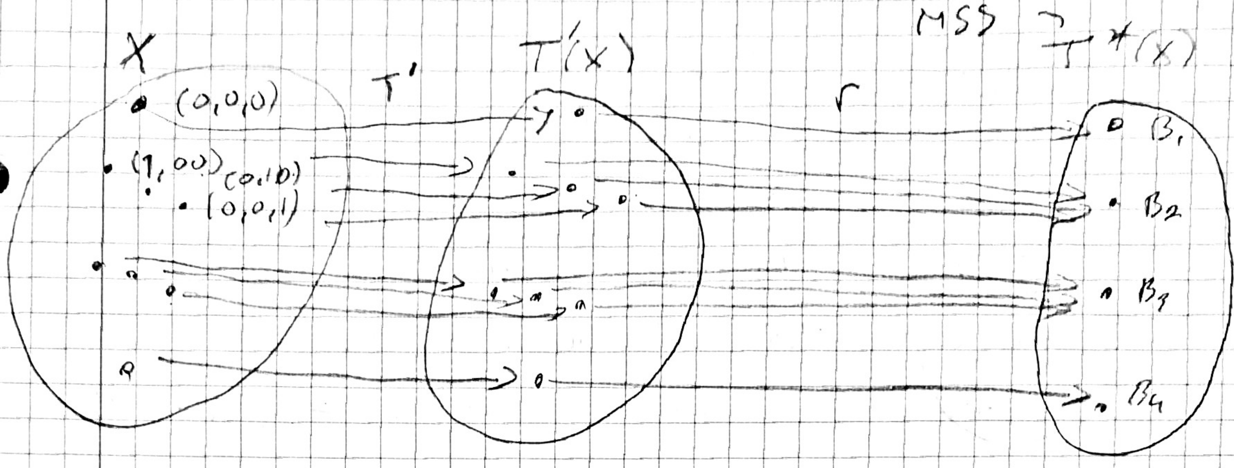
That is - $T(x)$ is the coarsest partition of $\mathcal{X}$ but still sufficient.

$\Rightarrow$ $T(\underline{x})$ is MSS $\Rightarrow$ it compresses the information in $\underline{X}$ the most

From book "... Also, any 1-1 function of a ss is ss. Suppose $T(x)$ is ss and define $T^*(\underline{x}) = r(T(x))$ $\forall x \in \underline{X}$, where r is 1-1 with inverse r^{-1}."

Since r has an inverse, and is onto/surjection r is a bijection hense $r(T(\underline{x}))$ is isomorphic to $T^*(\underline{x})$ - so not so interesting. $r(T(x))$ can also be defined outside $\underline{X}$ that is also not so interesting since we look at the restriction $T|_{x \in X}$

Next: looking at the partition example:



where $T'(x) = \underline{X}$ which is ss for θ,
but not minimal, and $\{id, s_1, s_2, s_3\}$
$r : T(x) \longrightarrow T^*(x)$ $\Downarrow$
$(a,b,c) \longmapsto \sigma(a,b,c)$ where σ
and $a,b,c \in \{0,1\}$

We see from partition-example that the domain - the total number of orderings of 3 variables x_1, x_2, x_3 is of order 6, and that $r(\cdot)$ therefor is symmetric to the symmetric group on 3 letters, S_3. Therefor - no smaller partition is a group on 3 letters, but a larger ordering can be made into a smaller ordering, by Lagranges thm:

$$|G| = [G : H] \cdot |H|$$

when $B_1 \cong \langle id \rangle$
$B_2 \cong \langle s_1 \rangle = s_1, s^2, s^3,$
$B_3 \cong \langle s_2 \rangle$
$B_4 \cong \langle \cdot \rangle$

when $\langle \cdot \rangle$ is the elements generated by.

MSS = Minimal Sufficient Statistics : Data reduction

The data $\underline{X} = (X_1, \dots X_n)$ are always sufficient for θ

$\underline{X} \in \mathbb{R}^n$. Let T be a ss for θ. $T \in \mathbb{R}^k, k \le n$

If T has the smallest dimension of all ss for θ

Then T is MSS. Not always unique

Def T is ss for θ. T minimal if $T = f(V)$, V ss for θ

2 ways to prove

$$\forall T \qquad \text{fixed}$$

1) $\underline{X}, \underline{y} \in S$ and $T(x) = T(y) \Rightarrow T_2(x) = T_2(y)$

$\Rightarrow T_2$ minimal

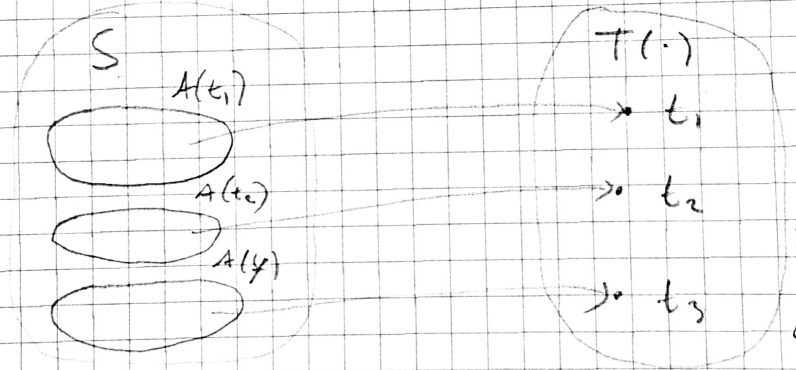
2) $X \sim f(x|\theta)$, $\theta \in \Omega$, $X \in S$, $T = T(X)$

T mss : $\dfrac{f(X|\theta)}{f(Y|\theta)}$ independent of $\theta \Leftrightarrow T(\underline{X}) = T(y)$ (*)

Pf Let T be statistics, assume (*) $\Rightarrow$ 1) show T ss for θ

2) show T MSS for θ

1) Let $A(t) = \{ \underline{X} \in \mathbb{R}^n \mid T(\underline{x}) = t \}$



$X_{t_1} \in A(t_1)$

$X_{t_2} \in A(t_2)$

$X_{T(S)} \in A(T(Y))$

if $T(\underline{y}) = T(X_{T(S)})$

$\Leftrightarrow \dfrac{f(\underline{Y}|\theta)}{f(\underline{X}|\theta)} = C(\underline{X}, \underline{y})$

$\Rightarrow f(Y|\theta) = f(X|\theta) \cdot C(x,y) \Rightarrow$ ss for θ

$= f(X_{T(y)}|\theta) = k(T(Y), \theta) h(y)$

2) Let V be ss for θ

$\Rightarrow f(\underline{X}|\theta) = g(V(\underline{x}), \theta) h(\underline{x}) = g(V(\underline{y}), \theta) h(\underline{y}) \cdot \dfrac{h(x)}{h(y)}$

$= f(Y|\theta)\, c(\underline{X}, \underline{y})$

by assumption $T(\underline{x}) = T(y)$ so if

$V(\underline{x}) = V(\underline{y}) \Rightarrow T(\underline{x}) = T(\underline{y})$

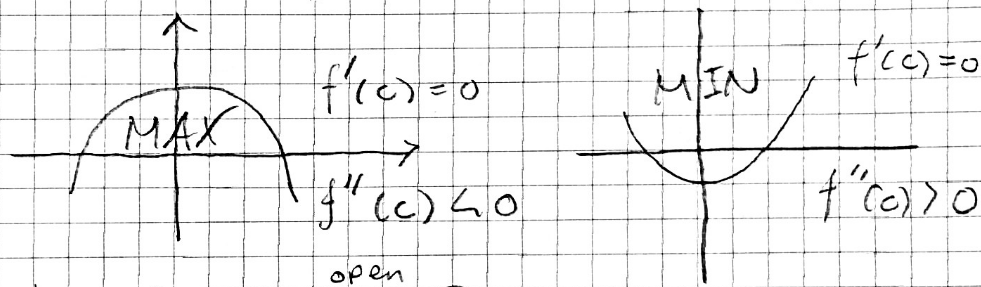
6) MLE $\overset{\text{loosely}}{=}$ Maximum Likelihood Estimator

For each sample point x, let $\hat{\theta}(x)$ be the parameter value at which $L(\theta|x)$ attains its maximum as a function of θ, with x fixed.
MLE of θ from sample $\underline{X}$ is $\hat{\theta}(\underline{X})$

MLE is found by:

1) $\frac{\partial}{\partial \theta_i} L(\theta|x) = 0 \quad \forall i \in \{1,..k\}$

2) check $\frac{\partial^2}{\partial \theta_i^2} L(\theta|x) < 0 \Rightarrow \frac{\partial}{\partial \theta_i}$ max



3) If $\underline{X} \overset{\text{open}}{\subset} \Omega$ - domain of $\underline{X}$

eg $\mathbb{R} \overset{\text{open}}{\subset} \mathbb{R}$ $\Longrightarrow$ Also check boundary conditions

4) Find global maximum

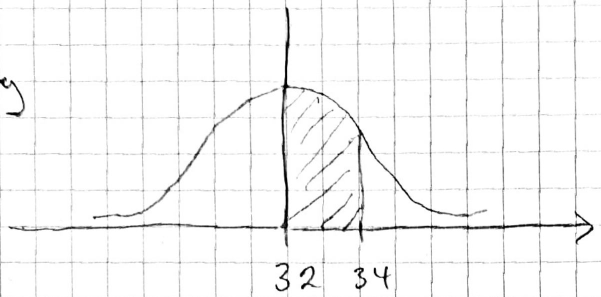
$$\sup_{c \in \underline{X}} \left\{ \frac{\partial}{\partial \theta_i} L(\theta|X) = 0 \right\}$$

Note its easier to work with $\text{Log}(L(\theta|x))$ And its analytically equiv as Log is monotone strictly increasing.

Log transforms $\prod_{i=1}^{n} f(\theta_i|x) \overset{\text{Log}}{\Longrightarrow} \sum_{i=1}^{n} f(\theta_i|X)$

likelihood is **not** probability

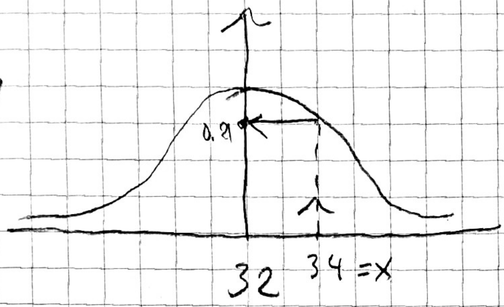
Probability



$P(X \in [32, 34] \mid \mu = 32, \sigma = 2.5)$

$X \sim N(32, 2.5^2)$

Likelihood



$L(\mu = 32 \wedge \sigma = 2.5 \mid x_1 = 34)$

$= 0.21$

$< L(\mu = 34 \wedge \sigma = 2.5 \mid x_1 = 34)$

but we also have the likelihood principle

$$L(\theta \mid x) = c(x, y)\, L(\theta \mid y)$$

Example $(X_1, \ldots, X_n)$ iid $\sim N(\theta, 1)$

1) $$L(\theta \mid X) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}(x_i - \theta)^2\right)$$

$$= \frac{1}{2\pi^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta)^2\right)$$

2) $$\frac{\partial}{\partial \theta} L(\theta \mid X) = \sum_{i=1}^{n} (x_i - \theta) = 0$$

$$\Leftrightarrow \sum X_i - n\theta = 0$$

$$\theta = \overline{X}$$

3) $$\frac{\partial^2}{\partial \theta^2} L(\theta \mid X) = -\theta < 0 \Rightarrow \theta = \overline{X} \text{ maximum}$$

4 $$\lim_{\theta \to \infty} L(\theta \mid X) = \lim_{\theta \to \infty} e^{-\theta} \xrightarrow{\theta \to \infty} 0$$

Hence $\hat{\theta}_{MLE} = \overline{X}$ ☒

MOM $\stackrel{\text{woah}}{\hat{=}}$ Method of Moment

let $\underline{X}=(X_1, .. X_n)$ have joint pdf $f(x|\theta)$

let $\theta=(\theta_1 .. \theta_k)$ Assume $E(X^i) < \infty$ $\forall i \in 1,..k$

Then

$$m_1 = n^{-1}\sum X_i = EX = \mu_1(\theta_1 ... \theta_k)$$

$$m_2 = n^{-1}\sum X_i^2 = EX^2 = \mu_2(\theta_1 .. \theta_k)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$m_k = n^{-1}\sum X_i^k = EX^2 = \mu_k(\theta_1 .. \theta_k)$$

is a system of k linear equations in k unknowns

Example $X_1 .. X_n$ iid $\sim N(\theta, \sigma^2)$.

$$m_1 = n^{-1}\sum x_i = \theta \quad \Rightarrow \hat{\theta} = \overline{X}$$

$$m_2 = n^{-1}\sum x_i^2 = \sigma^2 + \theta^2 \quad \Rightarrow \hat{\sigma}^2 = n^{-1}\sum x_i^2 - \theta^2$$

$$= n^{-1}\sum(x_i^2 - \overline{X})^2$$

MSE $\overset{\text{wade}}{\hat{=}}$ Mean squared error

of an estimator W of parameter θ

is the function θ defined by $E_\theta(W-\theta)^2$

$$\Rightarrow \quad MSE = Var_\theta W + (E_\theta W - \theta)^2$$

$$= \text{variance} + \text{bias } W^2$$

precision   accuracy

Bias $:\hat{=}$ $Bias_\theta W = E_\theta W - \theta$

when W - point estimator

θ - parameter of W

Example 7.3.3/4

$\underline{X}$ iid, $N(\mu, \sigma^2)$, $\bar{X}, S^2$ - unbiased

Let $\bar{X}$, X_1 estimate μ $E\bar{X} = \mu$, $ES^2 = \sigma^2$

$$MSE(\bar{X}) = E_\theta[(\bar{X} - \mu)^2] = Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$MSE(X_1) = E_\theta[(X_1 - \mu)^2] = Var(X_1) = \sigma^2$$

$$\frac{\sigma^2}{n} \leq \sigma^2 \quad \forall n \in \mathbb{Z}^+ \Rightarrow \bar{X} \text{ uniformly better}$$

$$\hat{\sigma}^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2$$

$$S^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$$

Using $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = E(\chi^2_{n-1}) = n-1$$

$$\frac{n-1}{\sigma^2} E(S^2) = n-1$$

$$E(S^2) = \frac{\sigma^2}{n-1} n-1 = \underline{\sigma^2}$$

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = Var(\chi^2_{n-1}) = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma^4} Var(S^2) = 2(n-1)$$

$$Var(S^2) = \frac{2(n-1)}{(n-1)^2} \cdot \sigma^4 = \underline{\frac{2\sigma^4}{n-1}}$$

$$\Rightarrow S^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 \Rightarrow \hat{\sigma}^2 = \frac{n-1}{n} S^2$$

$$\Rightarrow E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2 \Rightarrow Bias(\hat{\sigma}) = E(\hat{\sigma}^2) - \sigma^2$$

$$Var(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 Var(S) = \frac{(n-1)^2}{(n)^2} \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}$$

$$\xrightarrow{MSE} MSE(\hat{\sigma}) = E_\theta\left[(\hat{\sigma}^2 - \sigma^2)^2\right]$$

$$= \frac{2(n-1)\sigma^4}{n^2} - \left(E(\hat{\sigma}^2) - \sigma^2\right)^2$$

$$= \frac{2(n-1)\sigma^4}{n^2} - \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2$$

$$= \text{---}\,\|\,\text{---} - \left(-\frac{\sigma^2}{n}\right)^2$$

$$= \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}$$

$$\Rightarrow \lim_{n \to \infty} \frac{(2n-1)\sigma^4}{n^2} < \frac{2}{n-1}\sigma^4 = E(S^2 - \sigma^2)^2 \quad \forall n \geq 2$$

$\Rightarrow \hat{\sigma}^2$ is uniformly better than S^2

c) UMVUE = uniform minimum variance unbiased estimator

Def $\hat{\theta}$ UMVUE of $T(\theta)$ if

1) $E_\theta(\hat{\theta}) = T(\theta) \quad \forall \theta$ (unbiased)

2) $Var_\theta(\hat{\theta}) \le Var(\tilde{\theta}) \quad \forall \tilde{\theta}$ satisfying 1)

Can be difficult to find.

Ex $\underline{X}$ iid $\sim U(0,\theta)$

$\hat{\theta}_{MLE} = \max X_i$

$Y = \frac{n+1}{n} \hat{\theta}_{MLE}$ unbiased

Thm 7.3.23 $T = T(\underline{x})$ ss for θ
Assume that T's family of pdf/pmf with param θ is complete
$\Rightarrow$ If $E_\theta(\phi(T)) = \theta$ And $\phi(T)$ is UMVUE for θ,

* Def complete $\hat{=}$ $f(t|\theta)$ family of pdfs of $T(\underline{x})$. The family of prob dist is complete if $E_\theta\, g(T) = 0 \quad \forall \theta$
$\Rightarrow P_\theta(g(T) = 0) = 1$
equiv $T(\underline{X})$ is called complet statistics

(*) $\Rightarrow$ max X_i complete

And Ex 6.23 $\Rightarrow$ Y is UMVUE

Chap 8.2 book

d) LRT $\hat{=}$ Likelihood Ratio Test

LRT statistics for testing

$H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_0^c$

where $\Theta = \Theta_0 \cup \Theta_0^c$ and is all possible parameters.

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta|X)}{\sup_{\Theta_0^c} L(\theta|X)}$$

with rejection region $\{x : \lambda(x) \leq c\}$, $c \in [0,1]$

LRT based on test-statistics

$$\Delta_n = 2(\ell(\hat{\theta} - \ell(\hat{\theta})) \xrightarrow{d} \chi^2 \text{ if } H_0 \text{ true}$$

$$\Rightarrow \lambda(x) \simeq 1 \text{ if } H_0 \text{ true}$$
$$\simeq 0 \text{ if } H_0 \text{ false}$$

Rejection region

$$R = \{ \underline{X} \mid 2\Delta_n > \mu_{\frac{\alpha}{2}} \}$$

$$\Leftrightarrow R = \{ \underline{X} \mid \sqrt{2\Delta_n} > z_{\frac{\alpha}{2}} \}$$

ex $\quad \underline{X}$ iid, $f(x|\theta) = e^{-(x-\theta)} \mathbb{1}_{\{x \geq \theta\}} I_{\{\theta > 0\}}$

$\Rightarrow L(\theta|\underline{x}) = e^{-\sum x_i + n\theta} \mathbb{1}(\theta \leq x_{(1)})$

$H_0: \theta < \theta_0 \qquad H_1: \theta > \theta_0$

θ_0 - fixed.

L - is increasing on $\theta \in (-\infty, x_{(1)})$

$x_{(1)} < \theta_0 \Rightarrow \quad \dfrac{L(x_{(1)}|\underline{x})}{L(x_{(1)}|\underline{x})} = \dfrac{e^{-\sum x_i + n x_{(1)}}}{e^{-\sum x_i + n x_{(1)}}} = 1$

$x_{(1)} > \theta_0 \Rightarrow \quad \dfrac{L(\theta_0|\underline{x})}{L(x_{(1)}|\underline{x})} \overset{?}{=} \dfrac{e^{-\sum \theta_0 + n x_{(1)}}}{e^{-\sum x_i + n x_{(1)}}}$

$$\Rightarrow \lambda(\underline{x}) = \begin{cases} 1 & \theta_0 \geq x_{(1)} \\ \exp(-n(x_{(1)} - \theta_0)) & \theta_0 < x_{(1)} \end{cases}$$

8.3.11

UMP $\hat{=}$ Uniformly Most Powerful

Let G be a class of tests for

$H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$

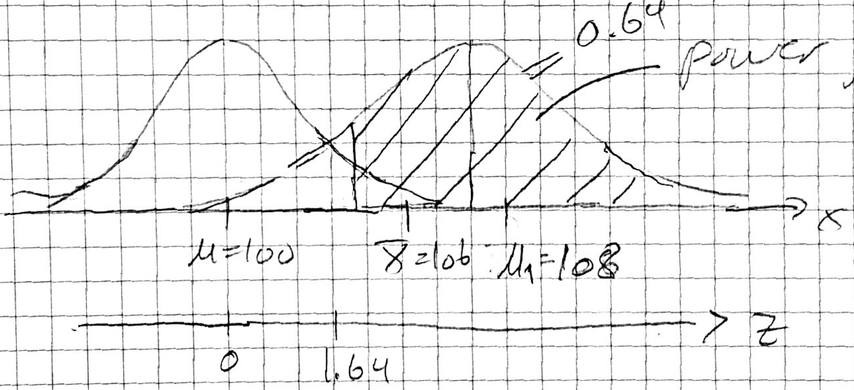
A test in G, with power function $\beta(\theta)$ is UMP class G test if $\beta(\theta) \geq \beta'(\theta)$ $\forall\ \theta \in \Theta_0^c$ And $\forall\ \beta'(\theta)$ a power function for G

8.3.1

Def Power function $\hat{=}$ of a hypotesis test with rejection region R is $\beta(\theta) = P_\theta(X \in R)$

$\beta(\theta) \approx 1$ good $\forall\ \theta \in \Theta_0^c$

$\beta(\theta) \approx 0$ good $\forall\ \theta \in \Theta_0$



Power – probability of rejecting H_0 when $\mu = 108$ is eg 0.64

Power – probability of rejecting H_0 when H_0 = false

	H_0 is True	H_0 is False
Reject H_0	Type I	V — Power, $\beta(\theta)$
Accept H_0	V	Type II

Continues

Power – probability of making correct decision

A deviation from H_0

UMP is the most likely to reject H_0 when H_1 is true.

8.3.16

MLR = Monotone likelihood ratio

A family of pdf/pmf $\{g(t|\theta)\}$ for a univariate rv T has MLR if

$$\forall \theta_2 > \theta_1 \quad f(t) = \frac{g(t|\theta_2)}{g(t|\theta_1)} \text{ is monotone}$$

Ex in exponential family

$$g(t|\theta) = h(t)\, c(\theta) \exp(w(\theta) t)$$

with $w(\theta)$ nondecreasing

Transparent
open
honest
vounerable

Ex Work hard / slack off

Suppose you work on project — you can work-hard $\nearrow H$ or slack-off $\downarrow L$

let e - be effort

q - quality of project deliverable

$\Rightarrow$ let $e \in \{H, L\}$

hence q drawn from $f(q|e)$

bayes law $\Rightarrow$ $$P(e=H|q) = \frac{f(q|H)}{f(q|H) + f(q|L)}$$

then probability worker worked hard?

$$\frac{1}{1 + \frac{f(q|L)}{f(q|H)}}$$ which is monoton in increasing q.

Monotone likelihood functions are used to construct UMP tests:

<u>Karlin - Rubin Thm</u> 8.3.17

$H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$

let T ss for θ and its family of dist. form MLR

Then the test $T > t_0$ is UMP level α test where $\alpha = P_{\theta_0}(T > t_0)$

exercise 2

a) Q: 1) Recall def exponential family

2) Show $(\chi^2_k)_{k \in \mathbb{Z}^+}$ form exponential family

A: Def exponential family:

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

Where $h(x) > 0$, $t_i(x) \in \mathbb{R}^*$

$c(\theta) > 0$, $w_i(\theta) \in \mathbb{R}^*$

Chi-squared pdf with k degrees freedom

$$f(x|k) = \frac{1}{(2^{k/2})\ \Gamma\left(\frac{k}{2}\right)} x^{\left(\frac{k}{2}-1\right)} \exp\left(\frac{-x^2}{2}\right) \qquad x \geq 0, \quad k \in \mathbb{Z}^+$$

$$= \left[2^{k/2}\, \Gamma\left(\frac{k}{2}\right)\right]^{-1} e^{\left(\frac{k}{2}-1\right)\ln x}\, e^{-\frac{x^2}{2}}$$

$$= \left[2^{k/2}\, \Gamma\left(\frac{k}{2}\right)\right]^{-1} e^{-\frac{x^2}{2}}\, e^{\left(\frac{k}{2}-1\right)\ln x}$$

$$h(x) = e^{-\frac{x^2}{2}}$$

$$c(k) = \frac{1}{2^{k/2}\, \Gamma\left(\frac{k}{2}\right)}$$

$$w_i(k) = \frac{k}{2} - 1$$

$$t_i(x) = \ln x$$

2b) Q: let $x_1 .. x_n$ be iid $\sim \chi_k^2$
find MSS T for k

A: $f(x|\theta) = h(x) \exp(\eta(\theta) T(x) - A(\theta))$

$$f(x|k) = [2^{k/2} \Gamma(\tfrac{k}{2})]^{-1} x^{\frac{k}{2}-1} e^{-x^2/2}$$

$$= \underbrace{e^{-x^2/2}}_{h(x)} e^{\underbrace{(\frac{k}{2}-1)}_{\eta(k)} \underbrace{\ln x}_{T(x)} - \underbrace{[\frac{k}{2}\ln 2 + \ln\Gamma(\frac{k}{2})]}_{A(k)}}$$

$$\frac{f(x|k)}{f(y|y)} = \frac{\prod_{i=1}^n e^{-\frac{x_i^2}{2}} e^{(\frac{k}{2}-1)\ln x_i} \cdot \cancel{e^{-\frac{k}{2}\ln 2 + \ln\Gamma(\frac{k}{2})}}}{\prod_{i=1}^n e^{-y_i^2/2} e^{(\frac{k}{2}-1)\ln y_i} \cancel{e^{-k/2 \ln 2 + \ln\Gamma(\frac{k}{2})}}}$$

$$= \frac{e^{-\Sigma(x_i^2/2)} e^{\Sigma(\frac{k}{2}-1)\ln x_i}}{e^{-\Sigma(y_i^2/2)} e^{\Sigma(k/2-1)\ln y_i}}$$

$$= \exp\left(-\frac{1}{2}\sum_{i=1}^n (x_i^2 - y_i^2)\right) \exp\left(n \cdot (\tfrac{k}{2}-1) \sum_{i=1}^n \ln x_i - \ln y_i\right)$$

Thm 6.2.12 extended from notes
let $\underline{x}$ be r.s. from full-rank expo family with
param $\theta = (\theta_1 .. \theta_k)$ $k \le n$
$\Rightarrow T(x_1 .. x_n) = (\sum_{j=1}^n t_1(x_j), \dots, \sum_{j=1}^n t_k(x_j))$ is MSS for θ

in a) we found $t_1 = \ln x_i$

Thm 6.2.12 $\Rightarrow$ $T(\underline{x}) = (\sum_{j=1}^n \ln x_1, \dots \sum_{j=1}^n \ln x_n))$ is MSS for θ

$\Rightarrow T(\underline{x}) = \sum_{i=1}^n \ln X_i$

d) Q: How can you apply b) to find

A: $\underbrace{Y = E(\bar{X} \mid T)}_{\text{See this is better estimator}}$ where $T = \overset{\text{MSS}}{\sum \ln(X_i)}$

MoM

$\bar{X}_{MoM}$ $\hat{\theta}_{MOM} = \bar{X}$ Thm 7.3.13

Rao-Blackwell 7.3.17

$Y_{\wedge}$ better uniformly $= \begin{cases} T \text{ SS for } \theta \\ Y = E(\hat{\theta} \mid T) \\ \hat{\theta} \text{ unbiassed of } \theta \\ \hat{\theta} \text{ unbiassed of } T(\theta) \text{ works} \end{cases}$

(T is complete statistics)

c) Q: 1) Find an estimator for k by use of the method of Moments.

2) Calculate its MSE

A: $m_1 = \frac{1}{n}\sum X_i \quad , M_1 = E x$

$$f(x \mid k) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} \cdot x^{\left(\frac{k}{2}-1\right)} e^{-x^2/2}$$

$$M_1 = \frac{1}{k}\sum_{i=1}^{k} x_i = \bar{x} \Rightarrow \hat{k} = \frac{1}{\bar{x}}\sum_{i=1}^{k} x_i$$

Bias: $E[k] = E[\bar{x}] = \bar{x} \Rightarrow$ unbiased

variance $Var(\bar{x}) = \frac{2k}{n}$

$$MSE = \frac{2k}{n} + 0 = \frac{2k}{n}$$

$$MSE = E_\theta(W-\theta)^2$$

$$E_k\left[\hat{k}-k\right] = E\left[\frac{1}{\bar{x}}\sum_{i=1}^{k} x_i - k\right]^2$$

$$= E\left[\left(\frac{1}{\frac{1}{k}\sum_{i=1}^{k} x_i}\right)\cdot\sum_{i=1}^{k} x_i - k\right]^2 =$$

$$= E\left[\frac{1}{\frac{1}{k}} - k\right]^2 = E\left[k-k\right]^2 = 0$$

exc 3 let $x_1 .. x_n$ be random sample from a distribution with density

$$f(x|\theta) = C\theta x^{-2} \quad , \quad 0 < \theta \leq x < \infty$$

a) Q: find $C > 0$ st $f(x|\theta)$ is pdf

A: $$\int_\theta^\infty C\theta x^{-2}\, dx = 1$$

$$\theta \int_\theta^\infty x^{-2}\, dx = \frac{1}{C}$$

$$C = \frac{1}{\theta \int_\theta^\infty x^{-2}\, dx} = \frac{1}{\theta\left[-x^{-1}\Big|_\theta^\infty\right.}$$

$$= \frac{1}{\theta\left[-\infty^{-1} - (-\theta)\right]}$$

$$= \frac{1}{\theta[0+\theta]} = \theta^{-2}$$

$$C = \frac{1}{\theta^2}$$

3b) Q: Show that the Method of Moments estimator is not well-def:

A:
$$E(x) = \int_{\theta}^{\infty} x \, f(x|\theta) \, dx$$

$$= c\theta \int_{\theta}^{\infty} x \frac{1}{x^2} \, dx$$

$$= c\theta \int_{\theta}^{\infty} \frac{1}{x} \, dx = c\theta \left[\ln x \right|_{\theta}^{\infty}$$

$$= c\theta \left[\ln \infty - \ln \theta \right)$$

$$= \infty$$

$\Rightarrow$ So the MoM estimator doesn't exist.

3 c) Q: Compute the Maximum Likelihood Estator for θ and show it is biased:

A: $L(\theta|x) = \prod_{i=1}^{n} f(x_i | \theta_1 .. \theta_k)$

from a) $c = \frac{1}{\theta^2} \Rightarrow f(x|\theta) = \frac{1}{\theta^2} \cdot \theta x^{-2} \mathbb{1}(\theta < x)$
$= \theta^{-1} x^{-2} \mathbb{1}(\quad)$

$L(\theta|\underline{x}) = \prod_{i=1}^{n} \theta^{-1} x_i^{-2}$

$= \theta^{-n} \prod_{i=1}^{n} x_i^{-2}$

$\ln L = -n \ln \theta - 2 \sum_{i=1}^{n} \ln x_i$

$\frac{\partial}{\partial \theta} \ln L = -\frac{n}{\theta}$

$\frac{\partial}{\partial \theta} \ln L = 0 \Rightarrow -\frac{n}{\theta} = 0 \Rightarrow \theta \rightarrow \infty$

but we have $\theta < X_i \quad \forall i$

$\Rightarrow \hat{\theta} = \min_i x_i$

To show $\hat{\theta}$ unbiased - $E(\hat{\theta}) = E(\min X_i)$

$\hat{\theta} = X_{(1)}$ is necessarily biased because

$P(X_{(1)} > \theta) > 0$ but $P(X_{(1)} < \theta) = 0$

so $P(X_{(1)} = \theta) = 0$

but $E(X_{(1)}) = P(X_{(1)} > x) = P(X_1 > x) \cap P(X_2 > x) \ldots$
$\cap P(X_n > x) = 0$

$= \inf_x P(X_i > x) = 0$

$\Rightarrow \hat{\theta}$ unbiased.

Ex 4 let $X_1 .. X_n$ each take values in $1..k$

a) Q: Prove that for this case there always exists MSS for

A: $Bernoulli(p) = Bin(1, p)$

X_1 takes one value in $1..k$ with probability 1

let $X_1 = i$, $i \in 1..k$ then $P(X_1 = j) = 0$, $j \neq i$

so $f(x|p) = \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$

Suppose $x \in \{0, 1\}$, $P(x=1) = \theta$, $P(x=0) = 1-\theta$

$x \sim p(x|\theta) = \theta^x (1-\theta)^{1-x}$

observe n trials : $x = x_1 .. x_n$

with $p(x_1 .. x_n | \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$

$= \theta^k (1-\theta)^{n-k}$

$k = \sum x_i$, k is rv $k \in 1..n$

$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \sim bin(n, \theta)$

Joint pmf $X_1 .. X_n$, and k

$$p(x_1 .. x_n, k | \theta) = \begin{cases} p(x_1 .. x_n | \theta) & \text{if } k = \sum x_i \\ 0 & \text{else} \end{cases}$$

$\Rightarrow p(x_1 .. x_n | k, \theta) = \dfrac{p(x, k|\theta)}{p(k|\theta)}$

$= \dfrac{\theta^k (1-\theta)^{n-k}}{\binom{n}{k} \theta^k (1-\theta)^{n-k}} = \dfrac{1}{\binom{n}{k}} \not\ni \theta$

$\Rightarrow$ conditional prob $X_1 .. X_n$ given $k = \sum x_i$ is uniformly distributed over the $\binom{n}{k}$ sequence that have k - ones.

So $X_1 .. X_n$ given k is indep of θ

$k = \sum x_i$ compresses $\{0,1\}^n$ (n bits) into $\log_2 n$ - bits on average.

Alternativly

4a) $X_1 .. X_n$ takes values in $\{1 .. k\}$

Q: Prove there always exists MSS

SS for θ 1) the data in it-self $T = (X_1 .. X_n)$ is SS for θ
Its the finest partition.

Complete
Def 6.2.21
Ex 6.2.22
2) Since T takes on values $\{1 .. k\}$ with probability 1, this yields $P_\theta(g(T) = 0) = 1 \quad \forall \theta$
$\Rightarrow T$ complete

3) Lecture 14 page 5
- Does there always exist MSS?
- No but if X is RV which takes discret values $\Rightarrow$ yes.

4 b) Prove that for this case MSS for θ partition is unique

A: T is the coarsest partition of $\underline{X}$

<u>Hint</u>: Def Coars: $A \subset B \iff A$ coarser than B

So A is Minimal Sufficient Partition (MSP)

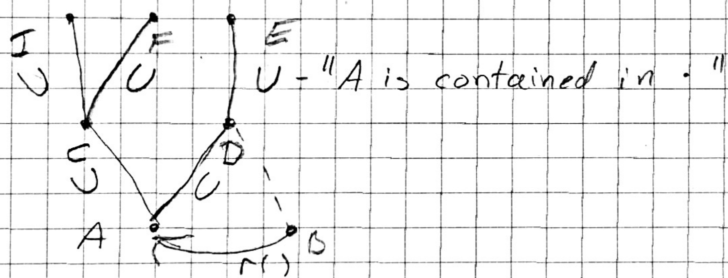
$\iff \forall G \in \Omega$ then $A \subset G$

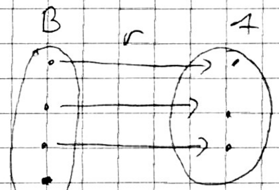
space of partitions on $\{1..k\}$

st each element is a sufficient statistics

That is, A is the greatest lower bound (GLB)

on the family of sufficient statistics Ω





Assume B is also MSP

then there exist a 1-1 mapping

$r: B \to A$. If r is <u>not</u> onto

then A is coars<u>er</u> than B $\nearrow\!\!\!\!\!\!\searrow$

Hence r is 1-1 and onto

$\implies$ r is a bijection

Since $1..k$ is finite

$\implies$ r is the identity mapping $r = id$

Hence $A = B$ ∎

Ex 5 let $X_1 .. X_n$ be iid $\sim U(0, \theta)$

a) Q: Explain why $U[a,b]$ not in expo. fam.

A: $$f(x|\theta) = \frac{1}{\theta} \mathbb{1}(0 < x < \theta)$$

Def expo fam:

$$p(x|\theta) = h(x) g(\theta) \exp\left(\sum w_i(\theta) t_i(x)\right)$$

but $\mathbb{1}(0 < x < \theta)$ is not of the form

$\exp\left(\sum w_i(\theta) t_i(x)\right)$

We have shown in lecture $\hat{\theta}_{MLE} = \max_i X_i$

$\theta_{MoM} = 2\overline{X}$

b) Q: Compare these two estimators:

1) Which of the two point estimators is better?

2) Are there alternatives?

A: $f(x|\theta) = \frac{1}{\theta} \mathbb{1}\{0 < x < \theta\}$

let $M = \max x_i \Rightarrow P(M < \theta) = 1 \Rightarrow M$ biased

let $\hat{\theta} = M$ pdf of $X_{(n)}$

$$F(x) = P(\hat{\theta} \leq x) = P(X_1 \leq x, X_2 \leq x, \dots X_n \leq x) = \left(\frac{x}{\theta}\right)^n$$

$$F'(x) = f(x) = n\left(\frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \begin{cases} \frac{n x^{n-1}}{\theta^n} & 0 < x < \theta \\ 0 & \text{else} \end{cases}$$

$$E[X_{max}] = \int_0^\theta x \cdot \frac{n x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n}\int_0^\theta x^n dx$$

$$= \frac{n}{\theta^n} \cdot \left[\frac{x^{n+1}}{n+1}\right|_0^\theta = \frac{n \cdot \theta^{n+1}}{\theta^n \, n+1} = \frac{n}{n+1}\theta \neq \theta$$

$\Rightarrow \hat{\theta}_{MLE}$ biased

$$Var(\hat{\theta}) = \int_0^\theta x^2 \frac{n x^{n-1}}{\theta^n} dx - E[\hat{\theta}]^2$$

$$= \frac{n}{\theta^n} \frac{x^{n+2}}{n+2}\Big|_0^\theta - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} - \left(\frac{n}{n+1}\theta\right)^2$$

$$= \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+2}\theta\right)^2$$

$$MSE = Var(\hat{\theta}) + bias\,\hat{\theta} = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+2}\theta\right)^2 + \left(\frac{n}{n+1}\theta - \theta\right)$$

$\hat{\theta}_{MOM} = 2\bar{x}$

$$E[2\bar{x}] = 2E[\bar{x}]$$
$$= \frac{2}{n}\sum E[x_i]$$
$$= \frac{2}{n}\cdot n \cdot \frac{\theta}{2} = \theta \Rightarrow \text{not biased}$$

$$Var(\hat{\theta}_{MOM}) = Var\left(2\cdot\frac{1}{n}\sum x_i\right)$$
$$= \frac{4}{n^2}\sum Var(x_i)$$
$$= \frac{4}{n^2}\, n\, V_\theta(x)$$

$$V_\theta(x) = E(x^2) - E(x)^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}$$

$$Var(\hat{\theta}_{MOM}) = \frac{4}{n^2}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n^2} = MSE \quad (\text{bias} = 0)$$

MSE of $\hat{\theta}_{MOM}$ < MSE $\hat{\theta}_{MLE}$

$\Rightarrow$ $\hat{\theta}_{MOM}$ better

2) Alternatives: $E(\max X) = \frac{n}{n+1}\theta$ so $\frac{n+1}{n}\hat{\theta}_{MLE}$ unbiased

5 c) Q: 1) Find the likelihood ratio statistics for the testing problem

$H_0: \theta \leq 1$ vs $H_1: \theta > 1$ $\qquad \theta_0 = 1$

2) find test statistics for $\hat{\theta}_{MOM}$ which is asymptotically normal distributed

A: The LRT statistics is

$$\lambda((x_1 .. x_n)) = \frac{\sup_{\theta \leq 1} L(\theta \mid x_1 .. x_n)}{\sup_{\theta > 1} L(\theta \mid x_1 .. x_n)}$$

Note $\hat{\theta}_{MLE} = X_{(n)}$, $\hat{\theta}_{MOM} = 2\bar{X}$

$$f(x \mid \theta) = \frac{1}{\theta} \mathbb{1}\{0 < x < \theta)$$

$$\lambda(\underline{x}) = \frac{\prod \frac{1}{\theta_0} \mathbb{1}\{0 < x_i \leq \theta_0}{\prod \frac{1}{\theta} \mathbb{1}\{0 < x_i \leq \theta\}} =$$

$$= \mathbb{1}\{X_{(n)} < \theta_0\} = \frac{\mathbb{1}\{X_{(n)} \leq 1\}}{\left(\frac{1}{\theta}\right)^n}$$

Assume $X_{(n)} \leq 1$

$$\Rightarrow \frac{1}{\left(\frac{1}{\theta}\right)^n} = \left(\frac{1}{\theta}\right)^{-n} = \theta^n = X_{(n)}^n = \hat{\theta}_{MLE}^n$$

Test of size α by def

$$R = \{X \mid \lambda(x) < e^{-\frac{1}{2} u_{\frac{\alpha}{2}}}\}$$

$$\underline{\underline{R = \{X \mid \lambda(x) > 1\}}} = \{X \mid X_{(n)} > 1\}$$

5 c)

2) construct test statistics based on $\hat{\theta}_{MOM}$ which is asymptotically normal dist

$\hat{\Theta}_{MOM} = 2\bar{x}$

$\sqrt{n}(\hat{\Theta}_{MOM} - \Theta) \longrightarrow ?$

$Var(X_i) = \frac{1}{12} \cdot \Theta^2$

$\sqrt{n}(\bar{X} - \frac{\Theta}{2}) \xrightarrow{CLT} N(0, \frac{\Theta^2}{12})$

5d) Q: Show that the likelihood ratio test in c) is UMP test?

A: given θ, $F(X_{(n)}) = P(X_{(n)} \leq x) = \left(\frac{x}{\theta}\right)^n$

if $0 \leq x \leq \theta$

$$f(X_{(n)}) = \frac{n x^{n-1}}{\theta^n} \mathbb{1}_{(0 \leq x \leq \theta)}$$

$$L(\theta \mid X_{(n)}) = \frac{1}{\theta^n} \mathbb{1}_{\{\theta \geq X_{(n)}\}}$$

$L(\theta) = 0$ if $\theta < X_{(n)}$

and $\frac{1}{\theta^n}$ is decreasing in n if $\theta \geq X_{(n)}$

Since any region $\geq X_{(n)}$ has likelihood than $\theta = X_{(n)}$ and $\frac{1}{\theta^n}$ monotonically decreasing

by Karlin-Ruben $\Longrightarrow$ the LR-test in c) is UMP

e) see r script

$\hat{\theta}_{MLE}$ is asymptotically better

$\hat{\theta}_{MoM}$ is better for smal $n \approx 5$

ex 6) let $X_1 \dots X_n$ be iid with pmf given by $p(X=a_i) = p_i$ for $i = 1..k$

a) Q: let $Y_i = \sum_{j=1}^{n} \mathbb{1}_{\{X_j = a_i\}}$

Argue that $Y = (Y_1 .. Y_n)$ is multinomial dist.
In the case $k=3$ and $p_1 = p_2 = p_3$ and $n=9$,
calculate $E(Y)$ and $\operatorname{cov}(Y)$

A:

$$P(X_1 = a_1, X_2 = a_2, \dots, X_k = a_k)$$
$$= \frac{n!}{a_1! \, a_2! \dots a_k!} p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$$

$n=9$, $k=3$, $p_1 = p_2 = p_3 = 1/3$ and $p_1 + p_2 + p_3 = 1$

$$\Rightarrow \frac{9}{x_1! \, x_2! \, x_3!} \left(\frac{1}{3}\right)^{a_1} \left(\frac{1}{3}\right)^{a_2} \left(\frac{1}{3}\right)^{a_3}$$

From def 4.6.2: Multinomial dist is model of:

1) n independent trials $(n=9)$
2) each trial is of k distinct outcomes
   $K = 3$
3) The probability is p_i, $p_i = 1/3$
4) Y_i is the count of the number of times the ith outcome occurs in the n trials

$$Y_i = \sum_{j=1}^{n} \mathbb{1}_{\{X_j = a_i\}}$$

$Y_i \in \{0..9\}$ $\qquad i \in \{1,2,3\}$

$E(Y_i) = n \cdot p_i = 9 \cdot 1/3 = 3$

$cov(Y_i, Y_j) = -n p_i p_j (1 - p_i)$

$= -9 \cdot 1/3 \cdot 1/3 (1 - 1/3) = -9 \cdot \frac{1}{9} \cdot \frac{2}{3}$

if $i \neq j$ $\qquad = \underline{-\frac{2}{3}}$

else

$cov(Y_i, Y_i) = Var(Y_i) = n p_i (1 - p_i)$

$= 9 \frac{1}{3} (1 - 1/3) = 9 \cdot \frac{1}{3} \cdot \frac{2}{3}$

$= 9 \cdot \frac{2}{9} = \underline{2}$

6 b) In general, the likelihood function of Y is

$$L(n_1, n_2, \dots n_{k-1} | p_1 \dots p_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

and $n_k = n - \sum_{j=1}^{k-1} n_j$, $n \in \mathbb{Z}^+$ - fixed/known

Q: 1) Is Y SS for p?

2) Is Y MSS for p?

A: 1)

Because of $\sum p_i = 1$ there is only $k-1$ free parameters $n_1, \dots n_{k-1}$ by dependens $n_k = n - \sum n_j$

$$L(\underline{n} | \underline{p}) = c \prod_{i=1}^{k} p_i^{n_i}$$

$$\frac{L(n|p)}{L(m|p)} = \frac{c \prod^{n} p_i^{n_i}}{c' \prod^{m} p_i^{m_i}} = c(n, m)$$

only if $n_i = m_i$ $i \in \{1 \dots k-1\}$

So $(n_1, \dots n_{k-1})$ is SS for n_i

and $L(n_1 \dots n_{k-1})$ is ——"——

2) by thm 6.2.13 $\ell(n|p)$ is MSS for n as $n_i = m_i$ $i \in 1 \dots k-1$

only if $\ell(n|p) = \ell(m|p)$

6 c) Q: Show that MLE of $(p_1 \dots p_k)$ is given by

$$\hat{\theta}_{MLE} = \left(\frac{Y_1}{n}, \dots \frac{Y_k}{n}\right)$$

A $L(n_1 \dots n_{k-1} \mid p) = \frac{n!}{n_1! \cdot \dots \cdot n_k!} \cdot p_1^{n_1} \cdot \dots \cdot p_k^{n_k}$

$$\ell = \ln L = \ln \frac{n!}{n_1! \dots n_k!} + n_1 \ln p_1 + \dots + n_k \ln p_k$$

$$= \ln n! - \ln n_1! - \dots - \ln n_k! + n_1 \ln p_1 + \dots + n_k \ln p_k$$

$$\frac{\partial}{\partial n_i} \ell = -\frac{1}{n_i!}\left(\frac{\partial}{\partial n_i} n_i!\right) + \ln p_i \quad ?$$

$$\Rightarrow \ell = \operatorname{Log} L(p_1 \dots p_k \mid n_1 \dots n_{k-1}, n) = \log\left(\frac{n!}{n_1! \dots n_k!}\right) + \sum_{i=1}^{n} \log p_i^{n_i}$$

$$\sum p_i = 1 \quad \xrightarrow[\text{Multiplier ?}]{\text{Lagrange}} \quad \log c + \sum \log p_i^{n_i} + \lambda(1 - \sum p_i)$$

$$\frac{\partial}{\partial p_i} \ell = \frac{n_i}{p_i} - \lambda \Rightarrow \frac{\partial}{\partial p_i} \ell = 0 \Rightarrow n_i = \lambda p_i$$

$$\xRightarrow{\text{Summin}} \sum_{i=1}^{n} \lambda p_i = \sum n_i$$

$$\lambda = n$$

$$\Rightarrow \frac{n_i}{p_i} - n = 0 \qquad n_i = n p_i \iff p_i = \frac{n_i}{n}$$

$$\xRightarrow{n_i = Y_i} \frac{Y_i}{n} = \hat{\theta}_{MLE}$$