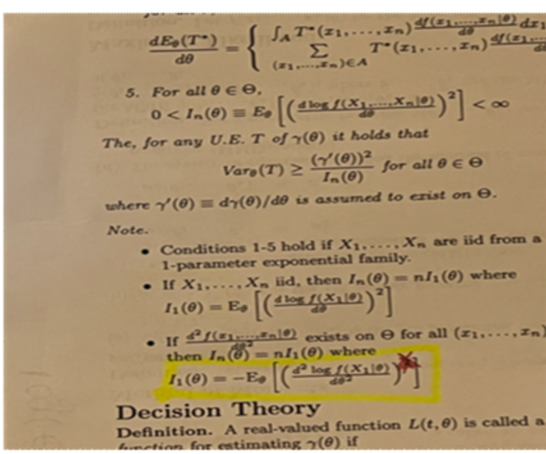
STAT 5430

Lecture 15, M, Feb 24

- Exam 1 is scheduled for W, Feb 26
  6:15-8:15 PM (Snedecor seminar room 3105)
  - No regular class on W, Feb 26
  - See Canvas for study guide, practice exams
  - can bring 1 page formula sheet (front/back) with anything on it
  - see Canvas for a "canned" sheet
  - I'll provide table with STAT 5420 distributions (see Canvas)



$$\frac{dE_\theta(T^*)}{d\theta} = \begin{cases} \int_A T^*(x_1,\ldots,x_n)\frac{df(x_1,\ldots,x_n|\theta)}{d\theta}dx_1 \\ \sum_{(x_1,\ldots,x_n)\in A} T^*(x_1,\ldots,x_n)\frac{df(x_1,\ldots}{d\theta} \end{cases}$$

5. *For all $\theta \in \Theta$,*
   $0 < I_n(\theta) \equiv E_\theta\left[\left(\frac{d\log f(X_1,\ldots,X_n|\theta)}{d\theta}\right)^2\right] < \infty$

The, for any U.E. T of $\gamma(\theta)$ it holds that

$$Var_\theta(T) \geq \frac{(\gamma'(\theta))^2}{I_n(\theta)} \text{ for all } \theta \in \Theta$$

where $\gamma'(\theta) \equiv d\gamma(\theta)/d\theta$ is assumed to exist on Θ.

Note.

- Conditions 1-5 hold if $X_1,\ldots,X_n$ are iid from a 1-parameter exponential family.
- If $X_1,\ldots,X_n$ iid, then $I_n(\theta) = nI_1(\theta)$ where $I_1(\theta) = E_\theta\left[\left(\frac{d\log f(X_1|\theta)}{d\theta}\right)^2\right]$
- If $\frac{d^2 f(x_1,\ldots,x_n|\theta)}{d\theta^2}$ exists on Θ for all $(x_1,\ldots,x_n)$ then $I_n(\theta) = nI_1(\theta)$ where $\boxed{I_1(\theta) = -E_\theta\left[\left(\frac{d^2\log f(X_1|\theta)}{d\theta^2}\right)^{\cancel{2}}\right]}$

Decision Theory

Definition. A real-valued function $L(t,\theta)$ is called a function for estimating $\gamma(\theta)$ if

fixed formula sheet (2nd version of Fisher Information $I_1(\theta)$ has no square)

Sufficiency and Point Estimation

Factorization Theorem, cont'd

$\underset{\sim}{1} = (1, \ldots, 1)'$

Example: Suppose $(X_1, \ldots, X_n) \sim MVN(\mu \cdot \underset{\sim}{1}, \sigma^2 \cdot A)$ where $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and A is a known $n \times n$ positive definite matrix. Find a sufficient statistic for (μ, σ^2).

Solution: joint pdf of $(X_1, \ldots, X_n)$ is

$$f(\underset{\sim}{x} \mid \mu, \sigma^2) = \frac{1}{(\sigma^2 2\pi)^{n/2}} \frac{1}{[\det(A)]^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{x} - \mu\underset{\sim}{1})' A^{-1} (\underset{\sim}{x} - \mu\underset{\sim}{1})\right]$$

$$= \underbrace{\frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\left[\underset{\sim}{x}' A^{-1} \underset{\sim}{x} + 2\mu \underset{\sim}{x}' A^{-1} \underset{\sim}{1} + \mu^2 \underset{\sim}{1}' A^{-1} \underset{\sim}{1}\right]\right.}_{g(\underset{\sim}{x}' A^{-1} \underset{\sim}{x},\ \underset{\sim}{x}' A^{-1} \underset{\sim}{1},\ \mu, \sigma^2)} \underbrace{\frac{1}{(2\pi)^{n/2}} \frac{1}{[\det(A)]^{1/2}} I(\underset{\sim}{x} \in \mathbb{R}^n)}_{h(\underset{\sim}{x})}$$

Hence, by Factorization Theorem,

$\underset{\sim}{S} = (\underset{\sim}{X}' A^{-1} \underset{\sim}{X},\ \underset{\sim}{X}' A^{-1} \underset{\sim}{1})$ are sufficient for (μ, σ^2)

Remarks:

1. The choice of $g(\underset{\sim}{S}, \underset{\sim}{\theta})$ and $h(\underset{\sim}{x})$ is not unique.
2. Any 1-to-1 function of a sufficient statistic is also sufficient.

   Example: In last example, suppose $A = I_{n \times n}$.

Then, $\underset{\sim}{S} = (\underset{\sim}{X}' A^{-1} \underset{\sim}{X},\ \underset{\sim}{X}' A^{-1} \underset{\sim}{1}) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$

Note $\underset{\sim}{T} = \left(\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2, \bar{X}_n\right)$ is a 1-to-1 function of $\underset{\sim}{S}$ $\Rightarrow$ $\underset{\sim}{T}$ is sufficient for (μ, σ^2)

Sufficiency and Point Estimation

Minimal Sufficiency

Question: Suppose $\underset{\sim}{S} \equiv (S_1, \ldots, S_k)$ is sufficient for $\underset{\sim}{\theta}$ and S_0 is another arbitrary statistic. Is $\underset{\sim}{S}^* \equiv (S_0, S_1, \ldots, S_k)$ is sufficient for $\underset{\sim}{\theta}$? Yes!

proof: Since $\underset{\sim}{S}$ is sufficient,

$f(\underset{\sim}{x}|\underset{\sim}{\theta}) = g(\underset{\sim}{S}, \underset{\sim}{\theta})h(\underset{\sim}{x})$ holds by Factorization Theorem

But $\underset{\sim}{S} = (S_1, .., S_k) = d(\underset{\sim}{S}^*)$ is a function of $\underset{\sim}{S}^*$

So, $f(\underset{\sim}{x}|\underset{\sim}{\theta}) = g(\underset{\sim}{S}, \underset{\sim}{\theta})h(\underset{\sim}{x})$

$= g(d(\underset{\sim}{S}^*), \underset{\sim}{\theta})h(\underset{\sim}{x}) = g_1(\underset{\sim}{S}^*, \underset{\sim}{\theta})h(\underset{\sim}{x})$

$\therefore$ $\underset{\sim}{S}^*$ is sufficient Fact. Theorem.

Definition: A vector of statistics $\underset{\sim}{S}$ is called **minimally sufficient** if

1. $\underset{\sim}{S}$ is sufficient for $\underset{\sim}{\theta}$, and

2. for any other vector $\underset{\sim}{T}$ of sufficient statistics for $\underset{\sim}{\theta}$, $\underset{\sim}{S}$ is a function of $\underset{\sim}{T}$.

(Later: we can check "minimally sufficient" using "completeness with sufficiency.")

$X_1, \ldots, X_n \xrightarrow{\text{data reduction}} \underset{\sim}{T}$ is sufficient for $\underset{\sim}{\theta}$ (more)

$\Big\downarrow$ data reduction $g(\underset{\sim}{T}) = \underset{\sim}{S}$
($\underset{\sim}{S}$ is a function of $\underset{\sim}{T}$)

minimally sufficient $\underset{\sim}{S}$ (less)

e.g. In last MVN example, $A = I_{n \times n}$, it turns out that $\underset{\sim}{S} = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$ is minimally sufficient for (μ, σ^2).
Later: We can show "minimal sufficiency" using "completeness"

Sufficiency and Point Estimation

Remarks on Sufficiency

1. If $X_1, \ldots, X_n$ is a random sample (iid) from pdf/pmf $f(x|\underset{\sim}{\theta})$, $\underset{\sim}{\theta} \in \Theta$, then the order statistics $X_{(1)}, \ldots, X_{(n)}$ are sufficient for $\underset{\sim}{\theta}$.

   proof: By the factorization theorem, $X_{(1)}, \ldots, X_{(n)}$ are sufficient for $\underset{\sim}{\theta}$ because we can write

$$\begin{aligned}\text{the joint pdf/pmf } f(\underset{\sim}{x}|\underset{\sim}{\theta}) &= \prod_{i=1}^{n} f(x_i|\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(x_{(i)}|\underset{\sim}{\theta}) \\ &= \underbrace{g(x_{(1)}, \ldots, x_{(n)}, \underset{\sim}{\theta})}_{\prod_{i=1}^{n} f(x_{(i)}|\underset{\sim}{\theta})} \underbrace{h(\underset{\sim}{x})}_{1}, \quad \text{for all } \underset{\sim}{x}, \underset{\sim}{\theta}\end{aligned}$$

2. If $\underset{\sim}{S} = (S_1, S_2, \ldots, S_k)$ is sufficient for real-valued $\theta \in \Theta \subset \mathbb{R}$, then any Bayes estimator is a function of $\underset{\sim}{S}$.

$$\begin{aligned} f_{\Theta|\underset{\sim}{x}}(\theta) &\propto f(\underset{\sim}{x}|\theta)\pi(\theta) \\ &\propto g(\underset{\sim}{s}, \theta)\pi(\theta) \end{aligned}$$

   Example: From homework, consider $X_1, \ldots, X_n$ iid Bernoulli(θ), $0 < \theta < 1$; loss $L(t, \theta) = \frac{(t-\theta)^2}{\theta(1-\theta)}$; and uniform(0,1) prior $\pi(\theta)$.

   Then the Bayes estimator is $T_0 = \bar{X}_n$, which is sufficient for θ (by factorization theorem).

3. If $\underset{\sim}{S} = (S_1, S_2, \ldots, S_k)$ is sufficient for $\underset{\sim}{\theta} \in \Theta \subset \mathbb{R}^p$ and $\hat{\theta}$ is the unique MLE of $\underset{\sim}{\theta}$, then $\hat{\theta}$ is a function of $\underset{\sim}{S}$.

$f(\underset{\sim}{x}|\theta) = g(\underset{\sim}{s}, \theta)h(\underset{\sim}{x})$ by Fact theorem

Sufficiency and Point Estimation

Rao-Blackwell Theorem & Sufficiency

Rao-Blackwell Theorem. Let $f(\underset{\sim}{x}|\underset{\sim}{\theta}) = f(x_1, \ldots, x_n|\underset{\sim}{\theta})$ be the joint pdf/pmf of $(X_1, \ldots, X_n)$ and $\underset{\sim}{S} = (S_1, S_2, \ldots, S_k)$ be sufficient for $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p$. Also let T be any UE of a real-valued $\gamma(\underset{\sim}{\theta})$ and $T^* = \mathrm{E}(T|\underset{\sim}{S})$ *(this conditional expectation does not depend on $\underset{\sim}{\theta}$, since $\underset{\sim}{S}$ is sufficient, and so is a statistic).*

new estimator (just depend on $\underset{\sim}{S}$ through data & NOT on $\underset{\sim}{\theta}$)

Then,

1. T^* is a function of $\underset{\sim}{S}$ and an UE of $\gamma(\underset{\sim}{\theta})$.

2. $\mathrm{Var}_{\underset{\sim}{\theta}}(T^*) \leq \mathrm{Var}_{\underset{\sim}{\theta}}(T)$, for all $\underset{\sim}{\theta} \in \Theta$.

3. If $\mathrm{Var}_{\underset{\sim}{\theta}_0}(T^*) = \mathrm{Var}_{\underset{\sim}{\theta}_0}(T)$ holds for some $\underset{\sim}{\theta}_0 \in \Theta$, then $P_{\underset{\sim}{\theta}_0}(T = T^*) = 1$.

Idea: T is UE of $\gamma(\theta)$ $\xrightarrow{\text{condition on sufficient } \underset{\sim}{S}}$ new $T^* = E(T|\underset{\sim}{S})$

"Rao-Blackwellization" or we say we "Rao-Blackwellize" T using $\underset{\sim}{S}$

Remarks

- Given an UE T of $\gamma(\underset{\sim}{\theta})$, the theorem shows how to obtain an UE T^* that is at least as good as T in terms of variance (in fact, better than T unless $T = T^*$ with probability 1 for all $\underset{\sim}{\theta}$). That is, you can "Rao-Blackwellize" an UE T by conditioning on a sufficient statistic $\underset{\sim}{S}$.

- For finding an UMVUE of $\gamma(\underset{\sim}{\theta})$ we may restrict attention to the class of estimators that are functions of a sufficient statistic.

Sufficiency and Point Estimation

Rao-Blackwell Theorem: Illustration

$n=2$

Example: Suppose X_1, X_2 are iid Exponential(θ). Note $T = X_1$ is an UE of θ and $\text{Var}_\theta(T) = \text{Var}_\theta(X_1) = \theta^2$.

$E_\theta T = E_\theta X_1 = \theta$

Also note that $S = X_1 + X_2$ is sufficient for θ by factorization theorem & S is GAMMA(2,θ)-distributed.

Verify that

1. $T^* = \text{E}_\theta(T|S) = \text{E}_\theta(X_1|S)$ is a function of S;
2. T^* doesn't depend on θ;
3. T^* is unbiased for θ;
4. and compare $\text{Var}_\theta(T)$ and $\text{Var}_\theta(T^*)$

Solution: Given $S = s > 0$, first find the conditional pdf of $X_1|S = s$ as

$$f(x_1|S=s) = \frac{f_{X_1,S}(x_1, s|\theta)}{f_S(s|\theta)} = \frac{f_{X_1,X_2}(x_1, x_2 = s - x_1|\theta)}{f_S(s|\theta)}$$

joint pdf of (X_1, S)

$$= \begin{cases} \dfrac{\theta^{-2}e^{-x_1/\theta}e^{-(s-x_1)/\theta}}{\theta^{-2}se^{-s/\theta}} = s^{-1} & \text{if } 0 < x_1 < s \\ 0 & \text{otherwise} \end{cases}$$

Gamma pdf

So, given $S = s > 0$, the conditional distribution of X_1 is UNIF(0, S)

Hence, the conditional expectation is $\text{E}_\theta(X_1|S = s) = \frac{0+S}{2} = S/2$

Now, treating S as a random variable, we have $T^* = \text{E}_\theta(X_1|S) = \frac{S}{2} = \frac{X_1+X_2}{2} = \bar{X}_2$

(1) T^* is function of S

(2) T^* doesn't depend on θ

(3) T^* is UE of θ $(E_\theta(\bar{X}_2) = \theta)$

(4) $\text{Var}_\theta(\bar{X}_2) = \theta^2/2 < \text{Var}_\theta(X_1) = \theta^2$