STAT 610: Discussion 1

1 Summary

- Minimal sufficient statistics
  - Definition: A statistic $T(\mathbf{X})$ is minimal sufficient if for any other sufficient statistics $T'(\mathbf{X})$, $T(\mathbf{x})$ is a function of $T'(\mathbf{x})$.
  - Theorem 6.2.13. $T(\mathbf{X})$ is minimal sufficient for $\boldsymbol{\theta}$ if T has the property that

$$T(\mathbf{x}) = T(\mathbf{y}) \Leftrightarrow \frac{f(\mathbf{x}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta})} \text{ does not depend on } \boldsymbol{\theta}.$$

  - Theorem 6.6.5. See lecture 3.

- In the case of exponential families

  Let $X_1, \ldots, X_n$ be i.i.d. from an exponential family with pdf

$$f_{\boldsymbol{\theta}}(x) = h(x)c(\boldsymbol{\theta}) \exp\{\boldsymbol{\eta}(\boldsymbol{\theta})^T \mathbf{T}(x)\},$$

  where $\boldsymbol{\eta}(\boldsymbol{\theta})^T = (w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta}))$, and $\mathbf{T}(x)^T = (t_1(x), \ldots, t_k(x))$. Then,
  - $T(\mathbf{X}) = (\sum_{i=1}^n t_1(X_i), \ldots, \sum_{i=1}^n t_k(X_i))$ is a sufficient statistic for $\boldsymbol{\theta}$.
  - $T(\mathbf{X}) = (\sum_{i=1}^n t_1(X_i), \ldots, \sum_{i=1}^n t_k(X_i))$ is also minimal sufficient if the parameter space contains an open set in $\mathbb{R}^k$.

- Ancillary statistics
  - Definition: $U(\mathbf{X})$ is ancillary of θ if the distribution of U does not depend on θ.

- Complete statistics
  - Definition: $T(\mathbf{X})$ is complete if T has the property that

$$\mathbb{E}_\theta g(T) = 0 \text{ for any measurable } g \text{ and } \theta \Rightarrow g(T) = 0 \text{ a.s.}$$

  - Basu's Theorem: Complete sufficient statistics and ancillary statistics for parameter θ are independent for all θ.

2 Questions

1. Let $(X_1, \ldots, X_n)$ be a random sample from density $\theta^{-1}e^{-(x-\theta)/\theta}I_{(\theta,\infty)}(x)$, where $\theta > 0$ is an unknown parameter.

   (a) Find a statistic that is minimal sufficient for θ.

   (b) Show whether the minimal sufficient statistic in (a) is complete.

2. Let X be a discrete random variable with

$$\mathbb{P}_\theta(X = x) = \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, \min\{\theta, n\}, n - x \leq N - \theta$$

   where n and N are known positive integers, $N \geq n$, and $\theta = 0, 1, \ldots, N$. Show that X is complete.

3. (a) If $\frac{X}{Y}$ and Y are independent random variables, show that

$$\mathbb{E}\left(\frac{X}{Y}\right)^k = \frac{\mathbb{E}(X^k)}{\mathbb{E}(Y^k)}.$$

   (b) Use Basu's theorem to show that if $X_1, ..., X_n$ are iid gamma(α, β), where α is known, then for $T = \sum_{j=1}^{i} X_j$

$$\mathbb{E}(X_{(i)}|T) = \mathbb{E}\left(\frac{X_{(i)}}{T}T|T\right) = T\frac{\mathbb{E}(X_{(i)})}{\mathbb{E}T}$$

4. Let $(X_1, \ldots, X_n)$, $n \geq 2$, be a random sample from a distribution having density $f_{\theta,j}$, where $\theta > 0$, $j = 1, 2$, $f_{\theta,1}$ is the density of $\mathcal{N}(0, \theta^2)$, and $f_{\theta,2}(x) = (2\theta)^{-1}e^{-|x|/\theta}$. Show that $T = (T_1, T_2)$ is minimal sufficient for (θ, j), where $T_1 = \sum_{i=1}^n X_i^2$ and $T_2 = \sum_{i=1}^n |X_i|$.

Q1

(a) $X_1, \dots, X_n \sim Exp(\theta, \theta)$

$$f(\underset{\sim}{x} \mid \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left\{-\frac{x_i-\theta}{\theta}\right\} \mathbb{1}\{x_i > \theta\}$$

$$= \theta^{-n} \exp\left\{-\frac{\sum x_i}{\theta} + n\right\} \prod \mathbb{1}\{x_i > \theta\} \quad \mathbb{1}\{x_{(1)} > \theta\}$$

$$= \theta^{-n} \exp\left\{-\frac{\sum x_i}{\theta} + n\right\} \mathbb{1}\{x_{(1)} > \theta\}$$

By F-T $\Rightarrow$ $\{\sum X_i, X_{(1)}\}$ is sufficient.

(Thm 6.2.13)

Given $\underset{\sim}{x} \neq \underset{\sim}{y}$.

$$\frac{f(\underset{\sim}{x} \mid \theta)}{f(\underset{\sim}{y} \mid \theta)} = \exp\left\{-\frac{(\sum X_i - \sum Y_i)}{\theta}\right\} \frac{\mathbb{1}\{x_{(1)} > \theta\}}{\mathbb{1}\{y_{(1)} > \theta\}}$$

- If $\sum X_i = \sum Y_i$ & $x_{(1)} = y_{(1)}$, $\frac{f(\underset{\sim}{x} \mid \theta)}{f(\underset{\sim}{y} \mid \theta)}$ is indep of θ.

- If $\frac{f(\underset{\sim}{x} \mid \theta)}{f(\underset{\sim}{y} \mid \theta)}$ is indep of θ,

we can see $\sum X_i = \sum Y_i$ & $x_{(1)} = y_{(1)}$

#

(b) Complete or not first order.

Find an not trivial ancillary statistics.

$$E \sum_{i=1}^{n} (X_i - \theta) = n\theta \qquad (\because \ X_i - \theta \sim Exp(\theta)).$$

$$\Rightarrow E(\sum X_i) = 2n\theta.$$

$$X_{(1)} - \theta \sim Exp(0, \frac{\theta}{n}).$$

So $E(X_{(1)} - \theta) = \frac{\theta}{n}$. $\Rightarrow E(X_{(1)}) = (1+\frac{1}{n})\theta$.

$$E\left[\sum_{i=1}^{n} X_i - \frac{2n}{(1+\frac{1}{n})} X_{(1)}\right] = 0 \quad \text{for all } \theta.$$

$$\underbrace{\sum_{i=1}^{n} X_i - \frac{2n}{(1+\frac{1}{n})} X_{(1)}}_{\parallel}$$

$$g(\sum X_i, X_{(1)})$$

But g is not zero.

So, $(\sum X_n, X_{(1)})$ is not complete statistics for θ.

$X_1 \dots X_n \sim Exp(\theta)$

$$P(X_{(1)} > t) = P(X_i > t)^n$$

#

Q_2:

Let $$h(\theta) = E_\theta g(X) = \sum_{x=0}^{\min(\theta,n)} g(x) \frac{\binom{\theta}{x}\binom{N-x}{n-x}}{\binom{n}{x}}$$

Show that. $h(\theta)=0 \quad \forall\, \theta$

$$\Rightarrow \quad g(x)=0 \quad \forall\, x=0,1,\cdots,n\}$$

(which can imply X is complete)

Pf) Plug in different θ.

Let $\theta=0$

$$h(0) = g(0)\frac{\binom{0}{0}\binom{N}{n}}{\binom{N}{n}} = 0 \quad \Rightarrow \quad g(0)=0$$

Let $\theta=1$

$$h(1) = \cancelto{0}{g(0)\frac{\binom{1}{0}\binom{N}{n}}{\binom{N}{n}}} + g(1)\frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = 0$$

$$\Rightarrow \quad g(1)=0$$

Continuous ~~the~~ procedure

We can get $\quad g(k)=0 \quad \forall\, k=0,1,\cdots,n$ #

Q3:

a)

$$\frac{X}{Y} \perp\!\!\!\perp Y$$

$$EX^n = E\left(\frac{X}{Y}Y\right)^n \stackrel{\downarrow}{=} E\left(\frac{X}{Y}\right)^n EY^n$$

b)

$$E[X_{(i)} \mid T] = T\, E\left[\frac{X_{(i)}}{T} \,\Big|\, T\right]$$

If $\frac{X_{(i)}}{T} \perp\!\!\!\perp T$

$$\Rightarrow = T\, E\left[\frac{X_{(i)}}{T}\right] \stackrel{(a)}{=} T \frac{E X_{(i)}}{E T}$$

Claim: $\frac{X_{(i)}}{T} \perp\!\!\!\perp T$

Pf) It is sufficient to claim

- T is complete for β
- $\frac{X_{(i)}}{T}$ is ancillary statistics for β.

$\Rightarrow$ Apply Basu thm

☆ T is complete

$$T = \sum_{j=1}^{n} X_j \sim \text{Gamma}(n\alpha, \beta)$$

$$f_T(t) = \frac{1}{\Gamma(n\alpha)\, \beta^{n\alpha}} t^{n\alpha - 1} \exp\{-t/\beta\}$$

which is a full rank exponential family $\xRightarrow{\text{Thm}}$ T is complete.

☆ $$\frac{X_{(i)}/\beta}{T/\beta} = \frac{Y_{(i)}}{\sum_{j=1}^{n} Y_j}$$, where $Y_1, \dots Y_n \sim \text{Gamma}(\alpha, 1)$

so $\frac{X_{(i)}}{T}$ is an ancillary statistics #

Q_4:

$$f(\underset{\sim}{x} \mid j, \theta) = \left(\frac{1}{\sqrt{2\pi\theta^2}}\right)^n \exp\left\{-\frac{\sum x_i^2}{2\theta^2}\right\} 1\{j=1\}$$

$$+ \left(\frac{1}{2\theta}\right)^n \exp\left\{-\frac{\sum |x_i|}{\theta}\right\} 1\{j=2\}$$

which is a function of $(\sum x_i^2, \sum |x_i|)$

$\xrightarrow{FT}$ $(\sum x_i^2, \sum |x_i|)$ is sufficient for θ and j

- Use thm 6.2.13

$$\frac{f(\underset{\sim}{x} \mid j, \theta)}{f(\underset{\sim}{y} \mid j, \theta)} = \frac{\left(\frac{1}{\sqrt{2\pi\theta^2}}\right)^n \exp\left\{-\frac{\sum x_i^2}{2\theta^2}\right\} 1\{j=1\} + \left(\frac{1}{2\theta}\right)^n \exp\left\{-\frac{\sum |x_i|}{\theta}\right\} 1\{j=2\}}{\left(\frac{1}{\sqrt{2\pi\theta^2}}\right)^n \exp\left\{-\frac{\sum y_i^2}{2\theta^2}\right\} 1\{j=1\} + \left(\frac{1}{2\theta}\right)^n \exp\left\{-\frac{\sum |y_i|}{\theta}\right\} 1\{j=2\}}$$

which is indep of (θ, j)

$$\iff (\sum x_i^2, \sum |x_i|) = (\sum y_i^2, \sum |y_i|)$$

So $(\sum x_i^2, \sum |x_i|)$ is a minimal statistic #