STAT 543 :)

Lec 18, W, Mar 5

sufficiency & completeness → Homework 4 posted, due M, Mar 10

- Exam 1 solutions, grading key, summary posted

So, T^* is UMVUE of $\tau(\theta)$ by L-S theorem.

Sufficiency and Point Estimation

Exponential Families (for Checking Sufficiency/Completeness)

Definition: A family of pdf/pmf $\{f(\underset{\sim}{x}|\underset{\sim}{\theta}) : \underset{\sim}{\theta} \in \Theta\}$, $\Theta \subset \mathbb{R}^p$, is called an **exponential family** if it can be written in the form

pmf/pdf of data

$$f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \begin{cases} c(\underset{\sim}{\theta})h(\underset{\sim}{x})\exp\left[\underbrace{\sum_{i=1}^{k} q_i(\underset{\sim}{\theta})t_i(\underset{\sim}{x})}_{t_i(\underset{\sim}{x}),\ q_i(\underset{\sim}{\theta}),\ i=1,\dots,k}\right] & \underset{\sim}{x} \in A \leftarrow \text{support} \\ 0 & \text{otherwise} \end{cases}$$

where

$\leftarrow$ support

$A \equiv \{\underset{\sim}{x} : f(\underset{\sim}{x}|\underset{\sim}{\theta}) > 0\}$ does NOT depend on $\underset{\sim}{\theta}$,

$c(\underset{\sim}{\theta}) > 0$ and $h(\underset{\sim}{x}) > 0$ are positive-valued functions,

and $q_i(\underset{\sim}{\theta})$, $t_i(\underset{\sim}{x})$ are real-valued functions for $i = 1, \dots, k$.

tool to determine/find complete & sufficient statistics

Theorem: Let $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ be a (possibly vector-valued) random iid sample from $f(\underset{\sim}{x}|\underset{\sim}{\theta})$, where $\{f(\underset{\sim}{x}|\underset{\sim}{\theta}) : \underset{\sim}{\theta} \in \Theta\}$ is an exponential family admitting a representation as above. If

set of all k-tuples over $\underset{\sim}{\theta} \in \Theta$ must an "open set" in $\mathbb{R}^k$

$$\left\{\overbrace{\left[q_1(\underset{\sim}{\theta}), \dots, q_k(\underset{\sim}{\theta})\right]}^{k\text{-tuple}} : \underset{\sim}{\theta} \in \Theta\right\} \supset \underbrace{(a_1, b_1) \times \cdots \times (a_k, b_k)}_{\text{open set/rectangle in } \mathbb{R}^k}$$

for some $a_i < b_i$, $i = 1, \dots, k$, then

$$\underset{\sim}{S} = \left(\sum_{j=1}^{n} t_1(\underset{\sim}{X}_j), \dots, \sum_{j=1}^{n} t_k(\underset{\sim}{X}_j)\right) \leftarrow k \text{ statistics}$$

is complete and sufficient.

Sufficiency and Point Estimation

Exponential Families: Illustration

Example. Let $X_1, \ldots, X_n$ be iid Gamma(α, β), $\alpha, \beta > 0$. Show that $\underset{\sim}{T} = (\sum_{i=1}^n X_i, \prod_{i=1}^n X_i)$ is complete and sufficient.

Solution: $X_1, \ldots, X_n$ iid, so consider the pdf of X_1

$$f(x|\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & , x>0 \\ 0 & \text{o.w.} \end{cases}$$

$e^{\log x^\alpha}$

$$= \begin{cases} c(\underset{\sim}{\theta}) h(x) \exp\left[-x/\beta + \alpha \log x\right] & , x>0 \\ 0 & \text{o.w} \end{cases}$$

$\underset{\sim}{\theta} = (\alpha, \beta)$, $c(\underset{\sim}{\theta}) = \frac{1}{\Gamma(\alpha)\beta^\alpha}$, $h(x) = x^{-1}$, $A \overset{\text{support}}{=} (0, \infty)$

$t_1(x) = x$, $q_1(\underset{\sim}{\theta}) = -1/\beta$, $t_2(x) = \log x$, $q_2(\underset{\sim}{\theta}) = \alpha$

check

$\{ [q_1(\underset{\sim}{\theta}), q_2(\underset{\sim}{\theta})] \in \mathbb{R}^2 : \underset{\sim}{\theta} = (\alpha, \beta) \in (0,\infty) \times (0, \infty) \}$

$= \{ [-1/\beta, \alpha] \in \mathbb{R}^2 : \alpha, \beta > 0 \} = (-\infty, 0) \times (0, \infty)$

$\supset (-1, 0) \times (0, 1)$

$(-10, -\pi) \times (1, 100)$

contains some open interval

By Theorem,

$\underset{\sim}{S} = \left(\sum_{j=1}^n t_1(X_j), \sum_{j=1}^n t_2(X_j) \right) = \left(\sum_{j=1}^n X_j, \sum_{j=1}^n \log X_j \right)$

is sufficient & complete

So, $\underset{\sim}{T} = \left(\sum_{j=1}^{n} X_j , \prod_{j=1}^{n} X_i\right)$ is one-to-one function with

$$\underset{\sim}{S} = \left(\sum_{j=1}^{n} X_j , \sum_{j=1}^{n} \log X_j\right)$$

& so $\underset{\sim}{T}$ is complete & sufficient

Note: looked at the problem as "n" iid X_i's (real-valued) & worked with $f(x|\theta)$
$\uparrow$ 1 obs.

or could have "1" vector $\underset{\sim}{x} = (x_1, \ldots, x_n)$ & worked with $f(\underset{\sim}{x}|\theta)$
$\uparrow$ all data

e.g. Suppose $X_1, \ldots, X_n$ are independent & $X_i \sim \text{Poisson}(i\theta)$, $\theta > 0$, $i = 1, \ldots, n$

$\underset{\sim}{X} = (X_1, \ldots, X_n) \leftarrow$ 1 vector

$$f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} \frac{e^{-i\theta}(i\theta)^{x_i}}{x_i!} = \underbrace{e^{-\sum_{i=1}^{n} i\theta}}_{c(\theta)} \underbrace{\left(\prod_{i=1}^{n} \frac{(i)^{x_i}}{x_i!}\right)}_{h(\underset{\sim}{x})} \exp\Big(\underbrace{\log\theta}_{q_1(\theta)} \cdot \underbrace{\sum_{i=1}^{n} x_i}_{t_1(\underset{\sim}{x})}\Big)$$

$e^{\log \theta^{\sum x_i}}$

check $\{q_1(\theta) \equiv \log\theta : \theta > 0\} = (-\infty, \infty)$
$\supset (0,1)$

$\Rightarrow t_1(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is complete & sufficient by theorem (applied to 1 $\underset{\sim}{x}$)