Tutorial for 11-10 and 11-13

1. Completeness.

- Definition

①

Definition 5 (One-parameter cases) Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s. from pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$, where $\theta \in \Theta \subset R$. A statistic $T = T(\boldsymbol{X})$ is said to be ***complete*** if and only if

$$E[\alpha(T)] = 0 \text{ for all } \theta \in \Theta \text{ implies that } P(\alpha(T) = 0) = 1 \text{ for all } \theta \in \Theta,$$

where $\alpha(T)$ is any statistic.

Definition 6 (Multi-parameter cases) Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s. from pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$, where $\theta \in \Theta \subset R^k$. A vector of statistics $T_1 = T_1(\boldsymbol{X}), \dots, T_r = T_r(\boldsymbol{X})$, where $r \geq k$, is said to be ***jointly complete*** if and only if

$$E[\alpha(T_1, \dots, T_r)] = 0 \text{ for all } \theta \in \Theta \text{ implies that } P(\alpha(T_1, \dots, T_r) = 0) = 1 \text{ for all } \theta \in \Theta,$$

where $\alpha(T_1, \dots, T_r)$ is any statistic.

②

From an reference book.

DEFINITION **2.4** *Let $f_\theta(t)$ be a family of pdf's for a statistic T.*

1. T is *complete* for θ if

$$E_\theta g(T) = 0 \quad \Longrightarrow \quad g(T) = 0 \ \ a.s. \quad \forall \theta.$$

2. T is *boundedly complete* if the previous statement holds for all bounded g. ∎

Completeness is not intuitive !!!

2. exponential family

- Definition:

Definition 7 (One-parameter cases) Suppose that a random variable X has a pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$, where $\theta \in \Theta \subset R$. Denote *supp(X)* by $\{x: f(x|\theta) > 0 \;\; or \; p(x|\theta) > 0\}$, which is also known as ***the support of*** X. **If (i)** supp(X) does not depend on θ, **and (ii)** the pdf or pmf of X can be written in form of

$$exp[a(\theta) + b(x) + c(\theta)d(x)],$$

where $a(\cdot)$, $b(\cdot)$, $c(\cdot)$, and $d(\cdot)$ are real-valued functions, then **the distribution of X is said to be an member of the (one-parameter) exponential family.**

Definition 8 (Multi-parameter cases) Suppose that a random variable X has a pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$, where $\theta = (\theta_1, \dots, \theta_k)' \in \Theta \subset R^k$ and k is a finite integer greater than 1.
If (i) supp(X) does not depend on θ, **and (ii)** the pdf or pmf of X can be written in form of

$$exp\left[a(\theta) + b(x) + \sum_{j=1}^{k} c_j(\theta)d_j(x)\right],$$

where $a(\cdot)$, $b(\cdot)$, $c_j(\cdot)$, and $d_j(\cdot)$, for $j = 1, \dots, k$, are real-valued functions, then **the distribution of X is said to be an member of the (k-parameter) exponential family.**

- properties

☆ A good property is that:

It is easy to find a complete and minimal statistic.

Theorem 6 (without proof): Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s from a distribution in an (full-rank/ one-parameter) exponential family with pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$ that can be written in form of $exp[a(\theta) + b(x) + c(\theta)d(x)]$, where $\theta \in \Theta \subset R$. Then, $\sum_{i=1}^{n} d(X_i)$ is a **complete and minimal sufficient statistic.**

Theorem 7 (without proof): Let $\boldsymbol{X} = \{X_i: i = 1, \dots, n\}$ be a r.s from a distribution in an (full-rank) exponential family with pdf $f(\cdot\,|\theta)$ or pmf $p(\cdot\,|\theta)$ that can be written in form of $exp\left[a(\theta) + b(x) + \sum_{j=1}^{k} c_j(\theta)d_j(x)\right]$, where $\theta \in \Theta \subset R^k$. Then, $\{\sum_{i=1}^{n} d_1(X_i), \dots, \sum_{i=1}^{n} d_k(X_i)\}$ is **a set of jointly complete and minimal sufficient statistics.**

- Examples

belong to: Normal, gamma, Poisson, binomial

not belong to: student-t, Uniform$(0, \theta)$, Cauchy, Double exponential

- Examples

Let $\mathbf{X} = \{X_i, i=1,2,\dots,n\}$ be a r.s from $N(\mu, \sigma^2)$

According Them 7 to find a set of jointly complete and minimal sufficient statistics

Method 1:

$$f(x|\theta) = f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\cdot x - \frac{\mu^2}{2\sigma^2}\right)$$

$C(\theta_2)\, d(x) + C(\theta_1)\cdot C(\theta_2)\cdot d(x) - a(\theta)$

$d_1(x) = x^2$

$d_2(x) = x$

$$\Rightarrow T(\mathbf{X}) = \left\{\sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} x_i\right\}$$

Method 2.

$$\prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot \exp\left[\frac{1}{\theta}\sum_{i=1}^{n} x_i - \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2\right]$$

$$T(\mathbf{X}) = \left\{\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} x_i^2\right\}$$

3. Lehman-Scheffé Theorem.

★ For consistency, I will rewrite this theorm. (Only change the notations)

- Definition:

① **Theorem 8 (Lehmann-Scheffé Theorem):** Let CS be a complete and (minimal) sufficient statistic. If there exists a function $h(CS)$ which is unbiased for $g(\theta)$, then it is the unique UMVUE of $g(\theta)$.

② **Lehmann-Scheffe Theorem**

THEOREM **3.3** *If T is C-S, and $E_\theta[\eta(T)] = g(\theta)$, then $\eta(T)$ is the UMVUE for $g(\theta)$.*

From another lecture note.

Let's recall Rao-Blackwell Theorem.

$\eta(T) = E(\underset{\text{Randomly selected}}{S(X)} \mid \underset{\text{unbiased}}{T})$

Check

$E_\theta(\eta(T)) = g(\theta)$?

Y → UMVUE

N → not UMVUE.

① find a sufficient statistic by Rao Blackwell

⇒ ② prove the statistics are complete

③ get a UMVUE by Lehman-Scheffé

Summary of Chapter 2.

- Ways to find UMVUE.

1. Guess the correct form of the function of CS. (CRLB)
2. Solve for $h(CS)$ in the equation $E[h(CS)] = g(\theta)$ directly.
3. Use Rao-Blackwell theorem to construct $h(CS)$ by
   a. first guessing or finding any unbiased estimator T for $g(\theta)$, and then
   b. evaluating $h(CS) = E[T|CS]$.

1. Simple guess and then <u>prove</u> it.
   - C-S : exponential family
   - unbiased:

2. Lehman - Scheffé Them.

   To solve $\eta(T)$ by $E\eta(T) = g(\theta)$ if we can know the distribution of T.

   $\boxed{\text{see example later.}}$

3. Rao-Blackwell Them.

   $\begin{cases} ① \text{ unbiased } T. \\ ② \text{ C-S} \end{cases}$

   $\Rightarrow \quad h(CS) = E(T/CS)$ is a UMVUE