Feynman Liang
STAT 210A
Lecture #5
9/8/2016

1 Last class

Theorem 1.1. *T sufficient + T complete $\implies$ T minimal*

Not very useful because proving completeness is difficult: need to show for *any* function f.

2 More properties of statistics

Definition 2.1. $V(X)$ is *ancillary* if its distribution does not depend on θ

Example 2.2. For $X_1, \cdots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, 1)$, the sample variance is ancillary and does not tell us anything about θ

Theorem 2.3 (Basu)**.** *T complete and sufficient, V ancillary, then T, V are independent $\forall \theta \in \Theta$.*

Proof. NTS $P_\theta(T \in B, V \in A) = P_\theta(T \in B) P_\theta(V \in A)$, $\forall A, B$ measurable, $\forall \theta$.

$$\begin{aligned} P_\theta(T \in B, V \in A) &= \mathbb{E}_\theta 1_B(T) 1_A(V) && (2.1)\\ &= \mathbb{E}_\theta \mathbb{E}_\theta \left[1_B(T) 1_A(V) | T\right] && (2.2)\\ &= \mathbb{E}_\theta \left\{1_B(T) \mathbb{E}_\theta \left[1_A(V) | T\right]\right\} && (2.3) \end{aligned}$$

By completeness of T

$$\mathbb{E}_\theta \left[1_A(V) | T\right] = q_A(T) \tag{2.4}$$

Hence

$$\mathbb{E}_\theta 1_A(V) = \mathbb{E}_\theta q_A(T) \tag{2.5}$$

But since V is ancillary, the LHS is a constant wrt θ so we can define a constant $p_A := \mathbb{E}_\theta 1_A(V)$ and write

$$\mathbb{E}_\theta q_A(T) = p_A \tag{2.6}$$

But since T is complete, we have $q_A(T) = p_A$ is constant as well. Hence, eq. (2.3) becomes

$$\begin{aligned} P_\theta(T \in B, V \in A) &= \mathbb{E}_\theta 1_B(T) p_A && (2.7)\\ &= P_\theta(T \in B) P_\theta(V \in A) && (2.8) \end{aligned}$$

so T and V are independent. $\square$

Example 2.4. Consider $\mathcal{P}_\mu = \{\mathcal{N}(\mu, \sigma^2)^n : \mu \in \mathbb{R}\}$. The joint distribution is

$$\frac{1}{(2\pi\sigma^2}^{n/2} \exp\left[\underbrace{\frac{n\mu}{\sigma^2}}_{\eta'T}\bar{x} - \underbrace{\frac{n\mu^2}{2\sigma^2}}_{A(\eta)} - \underbrace{\frac{1}{2\sigma^2}\sum_i^n x_i^2}_{h(x)}\right] \tag{2.9}$$

$\bar{x}$ is sufficient, also complete (can check if full-rank).

We will show $S^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ is ancillary i.e. should not depend on μ.

Define $Y_i = X_i - \mu$, $\bar{Y} = \frac{1}{n}\sum_i Y_i$. $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. Notice $X_i - \bar{X} = Y_i - \bar{Y}$ so $S^2 = \frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2$. As $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, the distribution S^2 does not depend on μ (i.e. is ancillary) and hence (by Basu's Theorem) $\bar{X} - S^2$ is independent for all μ i.e. over the family $\mathcal{P} = \mathcal{N}(\mu, \sigma^2)$.

NOTE: This result is only true for μ but for a *fixed* σ^2. The general result *does* hold for $\mathcal{P}_{\mu,\sigma^2}$ but needs to be shown. Could show the result by taking union of above result over all $\sigma^2 > 0$.

This is often useful for the t test, where $\bar{X}\,\mathcal{N}$, $S^2 \sim \chi^2$, and the statistic used is $t = \frac{\bar{X}}{\sqrt{S^2}}$.

3 The case for convex loss

Definition 3.1. f is *convex* if $\forall x, y \in \operatorname{dom} f, \gamma \in [0, 1]$

$$f(\gamma x + (1-\gamma)y] \le \gamma f(x) + (1-\gamma)f(y) \tag{3.1}$$

Theorem 3.2 (Jensen). *f convex*

$$f(\mathbb{E}X) \le \mathbb{E}f(x) \tag{3.2}$$

Definition 3.3. A *loss function* $L(\theta, \delta)$ takes a parameter θ and estimator $\delta(X)$ for $g(\theta)$ which evaluates how "good" the estimator is.

Convex loss means that L is convex in the second argument (i.e. $L(\theta, \cdot)$ is convex).

Theorem 3.4 (Rao-Blackwell). *T sufficient, loss L convex, $\delta(X)$ estimator for $g(\theta)$. Let*

$$\eta(T) = \mathbb{E}\left[\delta(X)|T\right] \tag{3.3}$$

Then

$$R(\theta, \eta) \le R(\theta, \delta) \tag{3.4}$$

(Strict "$<$" if L strictly convex, unless $\delta(X) = \eta(T)$)

Proof.

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}_\theta L(\theta, \delta) && \text{(3.5)} \\ &= \mathbb{E}_\theta \mathbb{E}_\theta\left[L(\theta, \delta)|T\right] && \text{Tower prop.} \quad \text{(3.6)} \\ &\ge \mathbb{E}_\theta L(\theta, \mathbb{E}_\theta\left[\delta|T\right]) && \text{Jensen} \quad \text{(3.7)} \\ &= R(\theta, \eta) && \text{(3.8)} \end{aligned}$$

□

Definition 3.5 (See Keener 4.1)**.** δ is *unbiased* for $g(\theta)$ if

$$\mathbb{E}_\theta \delta(X) = g(\theta) \quad \forall \theta \in \Theta \tag{3.9}$$

g is *U-estimable* or *estimable* if an unbiased estimator exists.

Example 3.6 (Non-estimable statistic)**.** $X \sim \text{Binom}(n, \theta)$, n fixed. Consider $g(\theta) = \sqrt{\theta}$. Suppose we have an estimator $\delta(X)$.

$$\mathbb{E}_\theta \delta(X) = \sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} \delta(X) \tag{3.10}$$

This is a polynomial wrt θ, so it will not be equal to $\sqrt{\theta}$ $\forall \theta \in \Theta$. Hence, $g(\theta)$ is not estimable.