STATS 300A: Theory of Statistics I **Autumn 2016/17**

Lecture 4: October 6

Lecturer: Joseph Romano *Scribes: Kenneth Tay*

4.1 Clarifications

- Completeness is a function of the family of distributions *induced* by the statistic T, not of the family of distributions for the data X.
- A statistic can be complete but not sufficient.

  For example, $X_1, \dots, X_n$ iid, $\sim \mathcal{N}(\theta, 1)$. Since this is a 1-parameter exponential family, $\bar{X}$ is sufficient and complete.

  Now, $X_1 \sim \mathcal{N}(\theta, 1)$, and if $\mathbb{E} f(X_1) = 0$, then $f = 0$. So X_1 is complete, but it is clearly not sufficient.

  In practice, there isn't a real reason to talk about complete statistics that are not sufficient.

4.2 Rao-Blackwell Theorem

Recall from last time:

Theorem 4.1 (Rao-Blackwell Theorem) *Assume that T is a sufficient statistic.*

Assume that we have a loss function $L(\theta, d)$ which is strictly convex in d, and that $\delta(X)$ is an estimator of $g(\theta)$ with finite risk $R(\theta, \delta)$.

Let $\eta(t) = \mathbb{E}\left[\delta(X) | T(X) = t\right]$. Then

$$R(\theta, \eta) < R(\theta, \delta)$$

unless $\delta = \eta$ with probability 1 (i.e. δ was a function of T to begin with).

Notes on Rao-Blackwell:

- Rao-Blackwell is not true without convexity of the loss function.
- Define a **randomized estimator** as $\delta = \delta(X, U)$, where U is an auxiliary random variable, of some known distribution, which is independent of X.

  Under Rao-Blackwell, we can dispense with randomized estimators. This is because $\mathbb{E}[\delta(X, U) \mid X]$ depends only on X.

4.3 Uniformly Minimum Variance Unbiased (UMVU) Estimators

Definition 4.2 *Let $X \sim P_\theta$, $\theta \in \Omega$, $g(\theta)$ real-valued. We say that $g(\theta)$ is* ***U-estimable*** *if there exists some $\delta(X)$ such that*

$$\mathbb{E}_\theta[\delta(X)] = g(\theta)$$

for all θ. (That is, $g(\theta)$ has an unbiased estimator.)

Note that there are some cases where $g(\theta)$ does not have an unbiased estimator!

Using Rao-Blackwell for squared error loss, we obtain the following corollary:

Corollary 4.3 *Suppose that $\delta(X)$ is an unbiased estimator for $g(\theta)$. Define*

$$\eta(T) = \mathbb{E}[\delta(X) \mid T].$$

Then η is also unbiased, and $Var_\theta \eta(T) \leq Var_\theta \delta(X)$ for all θ.

Definition 4.4 *We say that an estimator δ^* is **UMVU (uniformly minimum variance unbiased estimator)** or **UMRU (uniformly minimum risk unbiased estimator)** if for any other unbiased estimator δ,*

$$R(\theta, \delta^*) \leq R(\theta, \delta)$$

for all θ. It is uniquely UMVU if the above inequality is strict for some θ.

We can ask the following question: when does a U-estimable parameter $g(\theta)$ have an UMVUE, and how do we find it? The Lehmann-Scheffé Theorem answers this:

Theorem 4.5 (Lehmann-Scheffé Theorem, Take 1) *Suppose that there exists only one unbiased estimator of $g(\theta)$ based on a sufficient statistic T.*

Then, it must be UMVU.

Proof: Let $\eta^*(T)$ be unbiased, depending only on T. By assumption, it is the only such one.

Let δ be any other unbiased estimator. By Rao-Blackwell, we can improve on it:

$$\eta(T) = \mathbb{E}[\delta \mid T].$$

However, η is now a function of T which is unbiased, hence it must be equal to η^*! Therefore

$$R(\theta, \eta^*) = R(\theta, \eta) \leq R(\theta, \delta)$$

for all θ. ■

The condition in Theorem 4.5 may seem strange, but the following proposition shows that completeness implies it:

Proposition 4.6 *If T is a complete statistic, then there is at most one unbiased estimator of $g(\theta)$ based on T.*

Proof: Suppose that $\delta_1(T)$ and $\delta_2(T)$ are both unbiased estimators of $g(\theta)$ based on T. Then for all θ,

$$\mathbb{E}_\theta \left[\delta_1(T) - \delta_2(T)\right] = 0.$$

By completeness, this means that $\delta_1 = \delta_2$ with probability 1. ■

We can now rewrite Theorem 4.5 in a slightly more useful form:

Theorem 4.7 (Lehmann-Scheffé Theorem, Take 2) *If we have an unbiased estimator of $g(\theta)$ based on a sufficient and complete statistic T, then it must be UMVU.*

We'll now go through a series of examples to demonstrate how this theorem can be used to find UMVUs.

4.3.1 Example: $\mathcal{N}(\mu, \sigma^2)$

Let $X_1, \dots, X_n$ iid, $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

Case 1: μ unknown, σ known. Want to estimate μ.

We know that $\bar{X}$ is sufficient and complete, and is an unbiased estimator of μ. Hence, it is UMVU.

Now, let $\delta(X_1, \dots, X_n) = \text{Median}(X_1, \dots, X_n)$. This is an unbiased estimator of μ. The Rao-Blackwell Theorem says that we can improve on this estimator (or at least do no worse) by using $\mathbb{E}[\delta \mid \bar{X}]$ instead.

However, we know that $\mathbb{E}[\delta \mid \bar{X}]$ is an unbiased estimator of μ, and is a function of $\bar{X}$ which is sufficient and complete. Hence, we must have $\mathbb{E}[\text{Median}(X_1, \dots, X_n) \mid \bar{X}] = \bar{X}$.

Case 2: μ and σ unknown. Want to estimate $g(\mu, \sigma) = \mu$.

We know that $T = \left(\sum X_i, \sum X_i^2\right)$ is complete and sufficient. Since $\bar{X}$ is a function of T, it is again UMVU.

Case 3: μ and σ unknown. Want to estimate $g(\mu, \sigma) = \sigma^2$. (Assume $n > 1$.)

The sample variance

$$\frac{\sum (X_i - \bar{X})^2}{n-1}$$

is an unbiased estimator of σ^2, and is a function of T from Case 2. Hence, it is UMVU.

4.3.2 Example: Poisson(λ)

Let $X_1, \dots, X_n$ iid, $X_i \sim \text{Poisson}(\lambda)$. This is a 1-parameter exponential family with $\sum X_i$ being complete and sufficient.

To estimate $g(\lambda) = \lambda$, note that $\dfrac{\sum X_i}{n}$ and $\dfrac{\sum (X_i - \bar{X})^2}{n-1}$ are both unbiased estimators of λ, but only $\dfrac{\sum X_i}{n}$ is UMVU.

Let's say we want to estimate $g(\lambda) = e^{-\lambda}$. Let $T = \sum X_i$. Then $T \sim \text{Poisson}(n\lambda)$.

First consider the case where $n > 1$. In order for $\mathbb{E}_\lambda \eta(T) = e^{-\lambda}$ for all λ, we must have, for all λ,

$$\begin{aligned}
\sum_{t=0}^{\infty} \eta(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!} &= e^{-\lambda}, \\
\sum_{t=0}^{\infty} \frac{\eta(t)(n\lambda)^t}{t!} &= e^{(n-1)\lambda} \\
&= \sum_{t=0}^{\infty} \frac{[(n-1)\lambda]^t}{t!},
\end{aligned}$$

i.e. $\eta(t) = \left(\dfrac{n-1}{n}\right)^t$. Hence, when we set $\eta(T) = \left(1 - \dfrac{1}{n}\right)^T$, we get that $\eta(T)$ is UMVU.

In the case where $n = 1$, note that X is complete and sufficient. Note also that

$$P(X = j) = \frac{e^{-\lambda}\lambda^j}{j!}, \quad \Rightarrow \quad P(X = 0) = e^{-\lambda}.$$

Hence, the indicator function

$$1_{\{X=0\}} = \begin{cases} 1 & \text{if } X = 0, \\ 0 & \text{otherwise} \end{cases}$$

is UMVU.

4.3.3 Example: Uniform Distribution

Let $X_1, \ldots, X_n$ iid, $X_i \sim U(0, \theta)$. We want to estimate $g(\theta) = \frac{\theta}{2}$.

Previously we showed that $T = \max(X_1, \ldots, X_n)$ is a complete and sufficient statistic for this model.

Note that X_1 is an unbiased estimator of $\frac{\theta}{2}$. Thus,

$$\begin{aligned} \mathbb{E}[X_1 \mid T] &= \frac{1}{n}T + \frac{n-1}{n}\left(\frac{T}{2}\right) \\ &= \frac{n+1}{2n}T \end{aligned}$$

is UMVU for $\frac{\theta}{2}$. Equivalently, $\frac{n+1}{n}T$ is UMVU for θ.

4.3.4 Example: $\mathcal{N}(\mu, 1)$

Let $X_1, \ldots, X_n$ iid, $X_i \sim \mathcal{N}(\mu, 1)$. We want to estimate

$$g(\mu) = P_\mu \{X_i \leq u\} = \Phi(u - \mu),$$

where u is some critical fixed value and Φ is the standard normal CDF.

Note that $P\left\{X_1 \leq u \mid \bar{X}\right\}$ is UMVU. We use a trick to calculate this.

$$P\left\{X_1 \leq u \mid \bar{X} = \bar{x}\right\} = P\left\{X_1 - \bar{X} \leq u - \bar{x} \mid \bar{X} = \bar{x}\right\}.$$

Note that the distribution of $X_1 - \bar{X}$ does not depend on μ, and so it is ancillary. By Basu's Theorem, $X_1 - \bar{X}$ must be independent of the complete sufficient statistic $\bar{X}$, and so

$$P\left\{X_1 \leq u \mid \bar{X} = \bar{x}\right\} = P\left\{X_1 - \bar{X} \leq u - \bar{x}\right\}.$$

Now, let's find the distribution of $X_1 - \bar{X}$. It is normally distributed with mean 0. The variance can be computed:

$$\begin{aligned} \text{Var}(X_1 - \bar{X}) &= \text{Var}X_1 + \text{Var}\bar{X} - 2\text{Cov}(X_1, \bar{X}) \\ &= 1 + \frac{1}{n} - 2\text{Cov}\left(X_1, \frac{X_1}{n}\right) \\ &= 1 + \frac{1}{n} - \frac{2}{n} \\ &= \frac{n-1}{n}. \end{aligned}$$

Thus, we have

$$P\left\{X_1 - \bar{X} \le u - \bar{x}\right\} = \Phi\left(\frac{u - \bar{x}}{\sqrt{\frac{n-1}{n}}}\right).$$

4.3.5 Example: Bernoulli Trials

Let $X_1, \ldots, X_n$ iid, $X_i \sim \text{Bernoulli}(p)$.

$T = \sum X_i$ is complete and sufficient, and $T \sim \text{Binom(n, p)}$.

$\dfrac{T}{n}$ is UMVU for p.

To get a UMVU for the variance $g(p) = p(1-p)$: We know that $\dfrac{1}{n-1}\sum(X_i - \bar{X})^2$ is always an unbiased estimator for the variance. But since we always have $X_i^2 = X_i$ for Bernoulli random variables,

$$\begin{aligned}
\frac{1}{n-1}\sum(X_i - \bar{X})^2 &= \frac{1}{n-1}\left(\sum X_i^2 - n\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum X_i - n\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(T - \frac{T^2}{n}\right) \\
&= \frac{T(n-T)}{n(n-1)}
\end{aligned}$$

is a function of T, and hence is UMVU.