STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: Kaiser Fan and Chuchu Wang (Reviser)

Lecture 6

1 Unbiased Estimation

Definition 1 *An estimator is said to be unbiased if* $\mathbb{E}_\theta\{\delta(X)\} = g(\theta)$ *for all* θ.

Although it is very challenging to obtain uniformly best estimator, we can find an unbiased estimator with uniformly minimum risk, *i.e.* an unbiased estimator δ satisfying $R(\theta, \delta) \leq R(\theta, \delta')$ for all $\theta \in \Omega$ and any other unbiased estimator δ'.

Such an estimator is called a uniformly minimum risk unbiased estimator (UMRUE).

When $L(\theta, d) = (\theta - d)^2$, an UMRUE becomes a uniformly minimum variance unbiased estimator (UMVUE) because

$$\underbrace{\mathbb{E}_\theta\{(g(\theta) - \delta(X))^2\}}_{\text{Mean squared error (MSE)}} = \underbrace{\{\mathbb{E}_\theta\{\delta(X)\} - g(\theta)\}^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}_\theta\left\{(\delta(X) - \mathbb{E}_\theta\{\delta(X)\})^2\right\}}_{\text{Variance}}$$

The left-hand side of the equation is called the Mean Squared Error (MSE) of the estimator $\delta(X)$ for $g(\theta)$. Moreover, if $\delta(X)$ is unbiased, *i.e.* $\mathbb{E}_\theta\{\delta(X)\} - g(\theta) = 0$, then the MSE is reduced to

$$\mathbb{E}_\theta\{(g(\theta) - \delta(X))^2\} = \mathbb{E}_\theta\left\{(\delta(X) - \mathbb{E}_\theta\{\delta(X)\})^2\right\}$$

Definition 2 *If an unbiased estimator exists, then* $g(\cdot)$ *is called* U*-estimable.*

Example 1 *Suppose* $X \sim U(0, \theta)$. *Then* δ *is unbiased if*

$$\int_0^\theta \delta(x)\theta^{-1}\,dx = g(\theta), \qquad \forall\,\theta > 0$$

or if

$$\int_0^\theta \delta(x)\,dx = \theta g(\theta), \qquad \forall\,\theta > 0 \tag{1}$$

So, g *cannot be* U*-estimable unless* $\theta g(\theta) \to 0$ *as* $\theta \downarrow 0$. *If* g' *exists, then differentiating Equation 1, by fundamental theorem of calculus, we have*

$$\delta(x) = \frac{\partial}{\partial x}\left\{xg(x)\right\} = g(x) + xg'(x)$$

For, say $g(\theta) = \theta$, *then* $\delta(X) = X + X \cdot 1 = 2X$.

Example 2 *Suppose $X \sim \mathrm{Bin}(n, \theta)$. If $g(\theta) = \sin\theta$, then δ will be unbiased if*

$$\sum_{k=0}^{n} \delta(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sin\theta, \qquad \forall\, \theta \in (0,1) \tag{2}$$

The LHS of Equation 2 is a polynomial in θ with degree at most n. The sine function cannot be written as a polynomial of degree n, therefore, $\sin\theta$ is not U-estimable.

Definition 3 *An unbiased estimator δ is uniformly minimum variance unbiased (UMVU) if*

$$\mathrm{Var}_\theta(\delta) \leq \mathrm{Var}_\theta(\delta'), \qquad \forall\, \theta \in \Omega$$

for any other competing unbiased estimator δ'.

Theorem 4 ***(Lehmann-Scheffé Theorem)*** *If T is a complete and sufficient statistic, and $\mathbb{E}_\theta\{h(T(X))\} = g(\theta)$, i.e. $h(T(X))$ is unbiased for $g(\theta)$, then $h(T(X))$ is*

(a) the only function of $T(X)$ that is unbiased for $g(\theta)$;

(b) an UMRUE under any convex loss function;

(c) the unique UMRUE (hence UMVUE), up to a $\mathcal{P}$-null set, under any strictly convex loss function.

Proof

(a) Suppose $\mathbb{E}_\theta\{\tilde{h}(T(X))\} = g(\theta)$, then

$$\mathbb{E}_\theta\{\tilde{h}(T(X)) - h(T(X))\} = 0, \qquad \forall\, \theta \in \Omega$$

Thus, $\tilde{h}(T(X)) = h(T(X))$ almost surely for all $\theta \in \omega$ by completeness ($\mathbb{E}_\theta(f(T)) = 0 \Rightarrow f(T) = 0$).

(b) Consider any unbiased estimator $\delta(X)$ and let $\tilde{h}(T(X)) = \mathbb{E}_\theta\{\delta(X) \mid T(X)\}$. Then

$$\mathbb{E}_\theta\{\tilde{h}(T(X))\} = \mathbb{E}_\theta\{\mathbb{E}_\theta\{\delta(X) \mid T(X)\}\} = \mathbb{E}_\theta\{\delta(X)\} = g(\theta)$$

by tower property of conditional expectation ("smoothing"). By (a), $\tilde{h}(T(X)) = h(T(X))$, then $R(\theta, \tilde{h}(T(X))) = R(\theta, h(T(X)))$. By Rao-Blackwell Theorem, we have $R(\theta, \tilde{h}(T(X))) \leq R(\theta, \delta)$ for all $\theta \in \Omega$ if the loss function is convex. It follows that

$$R(\theta, h(T(X))) \leq R(\theta, \delta),\ \forall \theta \in \Omega$$

Therefore, $h(T(X))$ is an UMRUE under any convex loss function.

(c) If the loss function is strictly convex, $R(\theta, h(T(X))) < R(\theta, \delta)$ unless $\delta(X) \overset{a.s.}{=} h(T(X))$. Thus, $h(T(X))$ is the unique UMRUE (resp. UMVUE if the loss function adopted is the squared loss function).

Remark 1 *For $\mathcal{P}$-null set, it is with measure 0 s.t. we need not to consider that set, because it would not affect the expectation nor the risk. This theorem provides us with some useful strategies ("educated guesses") for finding UMRUEs under convex loss functions:*

1. *Rao-Blackwellisation.*
2. *Solve for the (unique) δ satisfying $\mathbb{E}_\theta\{\delta(T(X))\} = g(\theta)$ for all $\theta \in \Omega$.*
3. *Guess (the right form of unbiased function of $T(X)$)*

Example 3 ***(Rao-Blackwellisation)*** *Suppose $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \text{Bernoulli}(\theta)$. We know that $T(X) = \sum_{i=1}^n X_i$ is a complete sufficient statistic, we also know that $n^{-1}T(X)$ is an unbiased estimator for θ, i.e.*

$$\mathbb{E}_\theta\left(\frac{T(X)}{n}\right) = \frac{1}{n}\mathbb{E}_\theta\left(\sum_{i=1}^n X_i\right) = \frac{n\theta}{n} = \theta.$$

Therefore, $n^{-1}T(X)$ is an UMRUE for θ under any convex loss function. Suppose that, instead, we want to estimate θ^2. Let's examine: $\delta(X) = \mathbb{1}\{X_1 = X_2 = 1\} = X_1 \cdot X_2$ (GUESS). Consider $\mathbb{E}_\theta\{\delta(X)\}$, we have

$$\mathbb{E}_\theta\{\delta(X)\} = \mathbb{E}_\theta\{\mathbb{1}\{X_1 = X_2 = 1\}\} = \mathbb{E}_\theta(X_1 \cdot X_2) \overset{i.i.d.}{=} \{\mathbb{E}_\theta(X_1)\}^2 = \theta^2.$$

By conditioning on $T(X)$, we can find the UMRUE via

$$\begin{aligned}
\mathbb{E}_\theta\{\delta(X) \mid T(X) = t\} &= \mathbb{P}_\theta(X_1 = X_2 = 1 \mid T(X) = t) \\
&= \frac{\mathbb{P}_\theta\left(X_1 = X_2 = 1, \sum_{i=1}^n X_i = t-2\right)}{\mathbb{P}_\theta(T(X) = t)} \\
&= \frac{\theta^2\binom{n-2}{t-2}\theta^{t-2}(1-\theta)^{n-t}\mathbb{1}\{t \geq 2\}}{\binom{n}{t}\theta^2(1-\theta)^{n-t}} \\
&= \frac{t(t-1)\mathbb{1}\{t \geq 2\}}{n(n-1)}.
\end{aligned}$$

Note that in this case $\mathbb{1}\{t \geq 2\}$ is redundant as for $t = 0$ or 1, the term $t(t-1) = 0$. Hence, we conclude that the UMRUE for θ^2 is

$$\frac{T(X)\{T(X)-1\}}{n(n-1)}.$$

Example 4 ***(Rao-Blackwellisation)*** *Suppose $X_1, \cdots, X_n \overset{i.i.d.}{\sim} U(0, \theta)$. In this case, $T(X) = X_{(n)}$ is a complete and sufficient statistic, and $\delta(X) = 2X$ is an unbiased estimator for θ, i.e.*

$$\mathbb{E}_\theta(2X) = 2 \times \frac{\theta}{2} = \theta.$$

Given $X_{(n)}$, X_1 is equal to $X_{(n)}$ with probability $1/n$ and follows uniform $(0, X_{(n)})$ with probability $1-1/n$. Hence,

$$\mathbb{P}_\theta(X_1 = x_1 \mid T(X)) = \frac{1}{n} \times \mathbb{1}\{T(X) = x_1\} + \left(1 - \frac{1}{n}\right) \times \frac{1}{T(X)}\mathbb{1}\{0 < x_1 < T(X)\}$$

To find the UMVUE, we calculate

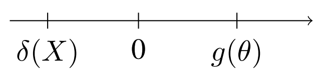
$$
\begin{aligned}
\mathbb{E}_\theta\{\delta(X) \mid T(X)\} = 2\mathbb{E}_\theta\{X_n \mid T(X)\} &= 2\left\{\frac{1}{n}T(X) + \left(1-\frac{1}{n}\right)\int_0^{T(X)} \frac{x_1}{T(X)}\,dx_1\right\} \\
&= 2\left\{\frac{T(X)}{n} + \left(1-\frac{1}{n}\right)\frac{T(X)}{2}\right\} \\
&= \left(\frac{n+1}{n}\right)T(X),
\end{aligned}
$$

which gives the desired result.

Example 5 ***(Solve for unique*** $\boldsymbol{\delta}$***)*** *Let* $X \sim \mathrm{Poisson}(\theta)$*.* X *is a complete and sufficient statistic,* X *is also unbiased and therefore UMVU for* θ*. Suppose we are interested in estimating* $g(\theta) = e^{-a\theta}$ *for* $a \in \mathbb{R}$*, known instead. We need to find an estimator* δ *such that* $\mathbb{E}_\theta\{\delta(X)\} = g(\theta)$ *for all* $\theta > 0$*. Under this model, we write*

$$
\begin{aligned}
& & \mathbb{E}_\theta\{\delta(X)\} = \sum_{x=0}^{\infty}\delta(x)\frac{e^{-\theta}\theta^x}{x!} = e^{-a\theta}, \qquad & \forall\,\theta > 0 \\
&\Leftrightarrow & \sum_{x=0}^{\infty}\frac{\delta(x)\theta^x}{x!} = e^{(1-a)\theta} = \sum_{x=0}^{\infty}\frac{(1-a)^x\theta^x}{x!}, \qquad & \forall\,\theta > 0 \\
&\Rightarrow & \delta(X) = (1-a)^X \text{ is the UMVUE for } g(\theta) &
\end{aligned}
$$

Note that the estimator is not ideal in the sense that if $a = 2$*, the estimator* $\delta(X) = (-1)^X$ *will change sign according to* X*'s "evenness" even though the estimand* $e^{-a\theta}$ *is non-negative. The estimator is hence inadmissible when* $a > 1$ *and dominated by* $\max\{\delta(X), 0\}$*.*



If $\delta(X)$ *is negative, then it will be far from the unbiased estimator* $g(\theta)$*.*

Example 6 ***(Guess)*** *Let* $X_1, \cdots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$*. Consider the case where* $\theta = (\mu, \sigma^2)$ *is unknown.*

(a) The UMVUE for $\sigma^2 is \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = S^2$

(b) How about the UMVUE for σ*?*

(c) What is the UMVUE for μ^2*?*

(b) Observe that

$$
X_i - \bar{X}_n \sim N\left(0, \frac{n-1}{n}\sigma^2\right) \qquad \Rightarrow \qquad \mathbb{E}\left\{\mid X_i - \bar{X}_n|\right\} = \sigma\sqrt{\frac{2}{\pi}} \times \sqrt{\frac{n-1}{n}}
$$

This implies

$$
\delta' = \frac{\sqrt{\pi n}}{\sqrt{2(n-1)}} \mid X_i - \bar{X}_n|
$$

is unbiased for σ*. At this point, we can Rao-Blackwellise this term. But the calculation is very tedious. Instead, we can observe another fact that*

$$
S_*^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = (n-1)S^2.
$$

We know that $S_*^2 \sim \sigma^2\chi_{n-1}^2$. Hence $\mathbb{E}(S_*) = \sigma\mathbb{E}(\chi_{n-1})$, which in turns implies that

$$\frac{\mathbb{E}(S_*)}{\mathbb{E}(\chi_{n-1})} = \sigma$$

meaning that $\frac{S_*}{\mathbb{E}(\chi_{n-1})}$ is unbiased for σ and hence UMVU.

(c) Taking the expecatation of the UMVUE for μ and squaring it, we obtain

$$\mathbb{E}(\bar{X}_n^2) = \mu^2 + \frac{\sigma^2}{n}$$

So

$$\delta_n(X) = \bar{X}_n^2 - \frac{S_*^2}{n(n-1)}$$

is the UMVUE. However, $\delta_n(X)$ can be negative even though the estimand is a non-negative quantity. The estimator is in fact inadmissible and dominated by the biased estimator $\max(0, \delta_n(X))$ since

$$\mathbb{E}\{\max(0, \delta_n(X))\} \neq \max\{\mathbb{E}(\delta(X)), 0\}.$$

Remark 2 *From the above example, with MLE,*

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \left(\frac{n-1}{n}\right) S_*^2$$