ST5215 Advanced Statistical Theory, Lecture 12

HUANG Dongming

National University of Singapore

17 Sep 2020

Overview

Last time

- Minimax rules and Bayes rules
- UMVUE

Today

- Methods of finding the UMVUE

Recap: Minimax rules and Bayes rules

- To find a good decision rule, we can consider some characteristic R_T of $R_T(P)$ for a given decision rule T, and then minimize R_T over $T \in \mathfrak{J}$.
- T_* is minimax: if $\sup_{P\in\mathcal{P}} R_{T_*}(P) \le \sup_{P\in\mathcal{P}} R_T(P)$ for any rule T
- T_* is a Bayes rule w.r.t. Π : if $r_{T_*}(\Pi) \le r_T(\Pi)$ for any rule T
    - The Bayes risk w.r.t. Π is defined by $r_T(\Pi) = \int_{\mathcal{P}} R_T(P) d\Pi(P)$, where Π is a known probability measure on $(\mathcal{P}, \mathcal{F}_{\mathcal{P}})$ with an appropriate σ-field $\mathcal{F}_{\mathcal{P}}$

Recap: Unbiased Estimators

- For squared error loss, the risk of an unbiased estimator is equal to its variance
- An unbiased estimator $T(X)$ of θ is called the *uniformly minimum variance unbiased estimator (UMVUE)* if and only if $\mathrm{Var}(T(X)) \leq \mathrm{Var}(U(X))$ for any $P \in \mathcal{P}$ and any other unbiased estimator $U(X)$ of θ
- Lehmann-Scheffé Theorem: if there is a sufficient and complete statistic $T(X)$ and θ is estimable, then the UMVUE of θ is equivalent to the conditional expectation given $T(X)$ of an unbiased estimator

Symmetric Unbiased Estimator and Order Statistics

Let $X_1, ..., X_n$ be i.i.d. from an unknown population P in a nonparametric family $\mathcal{P}$. Consider the vector of order statistics $T = (X_{(1)}, ..., X_{(n)})$

- In many cases, T is sufficient and complete for $P \in \mathcal{P}$. (For example, $\mathcal{P}$ is the collection of all Lebesgue p.d.f.'s.)
- Property: An estimator $\varphi(X_1, ..., X_n)$ is a function of T iff the function φ is symmetric in its n arguments.
- If T is sufficient and complete for $\mathcal{P}$ and θ is an estimable parameter, then a symmetric unbiased estimator $\varphi(X)$ of θ is the UMVUE

Examples: (assuming T is sufficient and complete)

- $\bar{X}$ is the UMVUE of $\theta = EX_1$;
- S^2 is the UMVUE of $\mathrm{Var}(X_1)$;
- $n^{-1} \sum_{i=1}^{n} X_i^2 - S^2$ is the UMVUE of $(EX_1)^2$;
- $F_n(t) := \frac{1}{n} \sum_{i=1}^{n} I_{X_i \le t}$ is the UMVUE of $P(X_1 \le t)$ for any fixed t.

A Counterexample

- The previous conclusions are not true if the vector of order statistics T is **NOT** sufficient and complete for $P \in \mathcal{P}$.
- For example, if $n > 2$ and $\mathcal{P}$ contains all symmetric distributions having Lebesgue p.d.f.'s and finite means, then there is no UMVUE of $\mu = EX_1$
- Here we assume X_i's are i.i.d. from P, which has Lebesgue p.d.f. f such that $f(\mu + x) = f(\mu - x)$ for all $x \in \mathcal{R}$ and $\int |x| f(x) \, \mathrm{d}m < \infty$

High-level idea of proof:

- Construct two subfamilies $\mathcal{P}_1$ and $\mathcal{P}_2$
- Find the unique UMVUE T_j under $\mathcal{P}_j$ respectively
- If T is the UMVUE under $\mathcal{P}$, it follows from the definition that $R_T(P) = R_{T_j}(P)$ for any $P \in \mathcal{P}_j$. Then using the uniqueness of T_j, we have $T = T_j$, $\mathcal{P}_j$-a.s.
- If each p.d.f. in $\mathcal{P}_1$ is positive, we have $\mathcal{P}_1$-a.s. $\Leftrightarrow$ $\mathcal{P}$-a.s., and thus, $T = T_1$, $\mathcal{P}$-a.s.
- If "$T_1 = T_2$, $\mathcal{P}_2$-a.s." is false, then we arrive at a contradiction

A Counterexample (Cont.)

Suppose $n > 2$ and X_i's are i.i.d. from $P \in \mathcal{P}$, which has Lebesgue p.d.f. f such that $f(\mu + x) = f(\mu - x)$ for all $x \in \mathcal{R}$ and $\int |x| f(x)\, \mathrm{d}m < \infty$. Then there is no UMVUE of $\mu = EX_1$.

Proof.

- Suppose that T is a UMVUE of μ
- Let $\mathcal{P}_1 = \{N(\mu, 1) : \mu \in \mathcal{R}\}$.
  The sample mean $\bar{X}$ is UMVUE when $\mathcal{P}_1$ is considered.
  The Lebesgue measure is dominated by any $P \in \mathcal{P}_1$.
  We conclude that $T = \bar{X}$ a.e.
- Let $\mathcal{P}_2$ be the family of uniform distributions on $(\mu - \delta, \mu + \delta)$, $\mu \in \mathcal{R}$, $\delta > 0$.
  $(X_{(1)} + X_{(n)})/2$ is the UMVUE when $\mathcal{P}_2$ is considered.
- Then $(X_{(1)} + X_{(n)})/2 = T = \bar{X}$, $\mathcal{P}_2$-a.s.
- This is impossible since $n > 2$. Hence, there is no UMVUE of μ.

$\square$

How to Find UMVUE? (1)

First method: solving equations for $h(T)$

- Find a sufficient and complete statistic T and its distribution
- Try some function h to see if $Eh(T)$ is related to θ
- Solve for h such that $Eh(T) = \theta$ for all P

Example Revisited: Uniform distributions

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on $(0, \theta)$, $\theta > 0$.

- The order statistic $X_{(n)}$ is sufficient and complete with Lebesgue p.d.f. $n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$
- Since

$$E_\theta X_{(n)} = n\theta^{-n}\int_0^\theta x^n dx = \frac{n}{n+1}\theta, \tag{1}$$

  we have $E_\theta\{(n+1)X_{(n)}/n\} = \theta$, for all $\theta > 0$
- By Lehmann-Scheffé Theorem, $\hat{\theta} = (n+1)X_{(n)}/n$ is the unique UMVUE of θ

Exercise

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on $(0, \theta)$, $\theta > 0$. Consider $\eta = g(\theta)$, where g is a differentiable function on $(0, \infty)$.

- The order statistic $X_{(n)}$ is sufficient and complete with Lebesgue p.d.f. $n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$
- An unbiased estimator $h(X_{(n)})$ of η must satisfy

$$\theta^n g(\theta) = n\int_0^\theta h(x)x^{n-1}dx \qquad \text{for all } \theta > 0. \tag{2}$$

- Differentiating both sides of the previous equation

$$n\theta^{n-1}g(\theta) + \theta^n g'(\theta) = nh(\theta)\theta^{n-1}, \text{ a.e.} \tag{3}$$

- Solving $h(t)$, we have $h(t) = g(t) + n^{-1}tg'(t)$, a.e.
- Hence, the UMVUE of η is given by

$$h(X_{(n)}) = g(X_{(n)}) + n^{-1}X_{(n)}g'(X_{(n)}) \tag{4}$$

Example: Poisson Distributions

Let $X_1, ..., X_n$ be i.i.d. from the Poisson distribution $P(\theta)$ with an unknown $\theta > 0$. Suppose that $\eta = g(\theta)$, where g is a smooth function such that $g(x) = \sum_{j=0}^{\infty} a_j x^j$, $x > 0$

- $T(X) = \sum_{i=1}^{n} X_i$ is sufficient and complete for $\theta > 0$
- Using properties of the m.g.f., we can show that T follows Poisson$(n\theta)$
- An unbiased estimator $h(T)$ of η must satisfy (for any $\theta > 0$):

$$\sum_{t=0}^{\infty} \frac{h(t)n^t}{t!}\theta^t = e^{n\theta} g(\theta)$$

$$= \sum_{k=0}^{\infty} \frac{n^k}{k!}\theta^k \sum_{j=0}^{\infty} a_j \theta^j \quad = \sum_{t=0}^{\infty} \left(\sum_{j,k:j+k=t} \frac{n^k a_j}{k!} \right) \theta^t.$$

- A comparison of coefficients in front of θ^t leads to

$$h(t) = \frac{t!}{n^t} \sum_{j,k:j+k=t} \frac{n^k a_j}{k!}. \tag{5}$$

How to Find UMVUE? (2)

Second method: Rao-Blackwellization

1. Find an unbiased estimator of θ, say $U(X)$
2. Find a sufficient and complete statistic $T(X)$
3. $E(U \mid T)$ is the UMVUE of θ by Lehmann-Scheffé Theorem

Remark.

- The distribution of T is not needed. We only need to work out the conditional expectation $E(U \mid T)$
- From the uniqueness of the UMVUE, it does not matter which $U(X)$ to use
- $U(X)$ should be chosen so as to make the calculation of $E(U \mid T)$ as easy as possible

Example: Exponential Distributions

Let $X_1, \ldots, X_n$ be i.i.d. sample from $\mathrm{Exp}(\theta)$. Let F_θ be the c.d.f. and $t > 0$. Find an UMVUE for the tail probability $p = 1 - F_\theta(t)$.

- $\overline{X} = n^{-1} \sum_{i=1}^n X_i$ is sufficient and complete, since $\mathrm{Exp}(\theta)$ is an exponential family of full rank
- $I_{(t,\infty)}(X_1)$ is unbiased: $EI_{(t,\infty)}(X_1) = P_\theta(X_1 > t) = 1 - F_\theta(t)$
- $E(I_{(t,\infty)}(X_1) \mid \overline{X})$ is a UMVUE
- Note that the distribution of $X_1/\overline{X}$ does not depend on θ – ancillary for θ
- By Basu's theorem, $X_1/\overline{X}$ and $\overline{X}$ are independent

$$\begin{aligned} P_\theta(X_1 > t \mid \overline{X} = \bar{x}) &= P_\theta(X_1/\overline{X} > t/\overline{X} \mid \overline{X} = \bar{x}) \\ &= P(X_1/\overline{X} > t/\bar{x}). \end{aligned}$$

- The UMVUE is $h(\bar{X})$, where $h(x) = P(X_1/\overline{X} > t/x)$

Example (Cont.)

- We need the unconditional distribution of $X_1/\overline{X}$:

$$P(X_1/\overline{X} > t/\bar{x}) = P\left(\frac{X_1}{\sum_{i=1}^n X_i} > \frac{t}{n\bar{x}}\right)$$

- $X_1 \sim \mathrm{gamma}(k = 1, \theta)$
- $\sum_{i=2}^n X_i \sim \mathrm{gamma}(k = n-1, \theta)$
- $\frac{X_1}{X_1+\sum_{i=2}^n X_i} \sim \mathrm{beta}(1, n-1)$, the p.d.f. is $(n-1)(1-x)^{n-2} I_{(0,1)}(x)$
  - See the supplementary reading material.
- Thus

$$P\left\{\frac{X_1}{\sum_{i=1}^n X_i} > \frac{t}{n\bar{x}}\right\} = (n-1)\int_{t/(n\bar{x})}^1 (1-x)^{n-2}\,\mathrm{d}x = \left(1 - \frac{t}{n\bar{x}}\right)^{n-1}$$

- So the UMVUE is

$$T(X) = \left(1 - \frac{t}{n\overline{X}}\right)^{n-1}$$

A Necessary and Sufficient Condition for UMVUE

Theorem (D)

Let $\mathcal{U}$ be the set of all unbiased estimators of 0 with finite variances and T be an unbiased estimator of η with $E(T^2) < \infty$.

(i) *A necessary and sufficient condition for $T(X)$ to be a UMVUE of η is that $E[T(X)U(X)] = 0$ for any $U \in \mathcal{U}$ and any $P \in \mathcal{P}$.*

(ii) *Suppose that $T = h(\tilde{T})$, where $\tilde{T}$ is a sufficient statistic for $P \in \mathcal{P}$ and h is a Borel function. Let $\mathcal{U}_{\tilde{T}}$ be the subset of $\mathcal{U}$ consisting of Borel functions of $\tilde{T}$. Then a necessary and sufficient condition for T to be a UMVUE of η is that $E[T(X)U(X)] = 0$ for any $U \in \mathcal{U}_{\tilde{T}}$ and any $P \in \mathcal{P}$.*

- Use of this theorem:
  - find a UMVUE
  - check whether a particular estimator is a UMVUE
  - show the nonexistence of any UMVUE

Proof of (i)

$\Rightarrow$:

- Suppose that T is a UMVUE of η and $U \in \mathcal{U}$
- For any constant c, define $T_c = T + cU$, which is also unbiased for η
- Since T is a UMVUE, $\mathrm{Var}(T_c) \geq \mathrm{Var}(T)$, $\forall P \in \mathcal{P}$
- It reduced to

$$c^2\mathrm{Var}(U) + 2c\mathrm{Cov}(T, U) \geq 0, \qquad \forall P \in \mathcal{P} \tag{6}$$

- This holds for all $c \in \mathcal{R}$, so $0 = \mathrm{Cov}(T, U) = E(TU)$, $\forall P \in \mathcal{P}$

$\Leftarrow$:

- Suppose $E(TU) = 0$ for any $U \in \mathcal{U}$ and $P \in \mathcal{P}$
- Let T_0 be another unbiased estimator of η with $\mathrm{Var}(T_0) < \infty$
- Note that $T - T_0 \in \mathcal{U}$. So$E[T(T - T_0)] = 0, \quad P \in \mathcal{P}$
- Adding and subtracting $(ET)^2$, we have $\mathrm{Var}(T) = \mathrm{Cov}(T, T_0) \leq \sqrt{\mathrm{Var}(T)\mathrm{Var}(T_0)}$, by Cauchy-Schwartz inequality
- Hence $\mathrm{Var}(T) \leq \mathrm{Var}(T_0)$ for any $P \in \mathcal{P}$.

Proof of (ii)

- "$\Rightarrow$" is directly from Part (i). We only need to show "$\Leftarrow$"
- It suffices to show:
  $E(TU) = 0$ for any $U \in \mathcal{U}_{\tilde{T}}$ and $P \in \mathcal{P}$ implies that $E(TU) = 0$ for any $U \in \mathcal{U}$ and $P \in \mathcal{P}$
- Let $U \in \mathcal{U}$. Then $E(U|\tilde{T}) \in \mathcal{U}_{\tilde{T}}$
- Note that $T = h(\tilde{T})$. By tower property,

$$\begin{aligned} E(TU) &= E[E(TU|\tilde{T})] \\ &= E[E(h(\tilde{T})U|\tilde{T})] \\ &= E[h(\tilde{T})E(U|\tilde{T})] \\ &= E[TE(U|\tilde{T})] = 0 \end{aligned}$$

Corollary

(i) *Let T_j be a UMVUE of η_j, $j = 1, ..., k$, where k is a fixed positive integer. Then $T = \sum_{j=1}^k c_j T_j$ is a UMVUE of $\eta = \sum_{j=1}^k c_j \eta_j$ for any constants $c_1, ..., c_k$.*

(ii) *Let T_1 and T_2 be two UMVUE's of η with finite variances. Then $T_1 = T_2$ a.s. P for any $P \in \mathcal{P}$.*

Proof:

- For (i). Note that T is an unbiased estimator of η with finite variance. Then for any $U \in \mathcal{U}$ and any $P \in \mathcal{P}$, $E[TU] = \sum_j c_j E[T_j U] = 0$
- For (ii). Note that $T_1 - T_2 \in \mathcal{U}$.
  Thus, for any $P \in \mathcal{P}$, $E[T_1(T_1 - T_2)] = E[T_2(T_1 - T_2)] = 0$.
  This implies $E[(T_1 - T_2)(T_1 - T_2)] = 0$.
  So $T_1 - T_2 = 0$, a.s.

Example: Uniform Distributions

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on the interval $(0, \theta)$.

- In previous lecture, we have shown that $(1 + n^{-1})X_{(n)}$ is the UMVUE for θ when the parameter space is $\Theta_0 = (0, \infty)$.

Suppose now that $\Theta = [1, \infty)$.

- $X_{(n)}$ is not complete, although it is still sufficient for θ
- Lehmann-Scheffé Theorem does not apply to $X_{(n)}$
- $(1 + n^{-1})X_{(n)}$ does not make use of the information about $\theta \geq 1$

We now use Theorem D (ii) to find a UMVUE of $\theta \in \Theta$.

Consider $T = h(X_{(n)})$ such that $E(TU) = 0$ for $U \in \mathcal{U}_{X_{(n)}}$.

- Let $U(X_{(n)})$ be an unbiased estimator of 0.
- Since $X_{(n)}$ has the Lebesgue p.d.f. $n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$,

$$0 = \int_0^1 U(x)x^{n-1}dx + \int_1^\theta U(x)x^{n-1}dx \quad \text{for all } \theta \geq 1. \tag{7}$$

- This implies that $U(x) = 0$ a.e. Lebesgue measure on $[1, \infty)$ and

$$\int_0^1 U(x)x^{n-1}dx = 0. \tag{8}$$

For T to be the UMVUE, we must have

$$\int_0^1 h(x)U(x)x^{n-1}dx = 0. \tag{9}$$

Consider the following function, where c and b are some constants

$$h(x) = \begin{cases} c & 0 \leq x \leq 1 \\ bx & x > 1, \end{cases} \tag{10}$$

From the previous discussion, the following holds

$$E[h(X_{(n)})U(X_{(n)})] = 0, \qquad \theta \geq 1. \tag{11}$$

For $h(X_{(n)}$ to be unbiased, we need

$$\begin{aligned}\theta &= E[h(X_{(n)})] \\ &= cP(X_{(n)} \leq 1) + bE[X_{(n)} I_{(1,\infty)}(X_{(n)})] \\ &= c\theta^{-n} + [bn/(n+1)](\theta - \theta^{-n}), \quad \forall \theta \geq 1.\end{aligned}$$

Thus, we only need to choose $c = 1$ and $b = (n+1)/n$.
The UMVUE of θ is then

$$h(X_{(n)}) = \begin{cases} 1 & 0 \leq X_{(n)} \leq 1 \\ (1 + n^{-1})X_{(n)} & X_{(n)} > 1. \end{cases} \tag{12}$$

This estimator is better than $(1 + n^{-1})X_{(n)}$, which is the UMVUE when the parameter space is $\Theta_1 = (0, \infty)$.
In fact, $h(X_{(n)})$ is complete and sufficient for $\theta \in [1, \infty)$.

Tutorial

1. Let $X_1, \ldots, X_n$ be i.i.d. random variables having the exponential distribution $E(0, \theta), \theta \in (0, \infty)$. Consider estimating θ under the squared error loss. Calculate the risks of the sample mean $\bar{X}$ and $cX_{(1)}$, where c is a positive constant.
   Is $\bar{X}$ better than $cX_{(1)}$ for some c?
2. Consider the estimation of an unknown parameter $\theta \geq 0$ under the squared error loss. Show that if T and U are two estimators such that $T \leq U$ and $R_T(P) < R_U(P)$, then $R_{T_+}(P) < R_{U_+}(P)$, where f_+ denotes the positive part of f.
3. Let X be a random variable having the binomial distribution $\text{Bi}(p, n)$ with an unknown $p \in (0, 1)$ and a known n. Consider the problem of estimating $\theta = p^{-1}$. Show that there is no unbiased estimator of θ.

Exercise 1

Let $X_1, \ldots, X_n$ be i.i.d. random variables having the exponential distribution $E(0, \theta), \theta \in (0, \infty)$. Consider estimating θ under the squared error loss. Calculate the risks of the sample mean $\bar{X}$ and $cX_{(1)}$, where c is a positive constant. Is $\bar{X}$ better than $cX_{(1)}$ for some c?

Solution:

- The c.d.f. of $X_{(1)}$ is
  $1 - P(X_{(1)} \geq t) = 1 - \prod_i P(X_i \geq t) = 1 - \exp(-nt/\theta), \forall t > 0$
- So $X_{(1)} \sim E(0, \theta/n)$
- Using the decomposition of MSE=bias^2+Variance, we have

$$R_{\bar{X}}(\theta) = \theta^2/n,$$
$$R_{cX_{(1)}}(\theta) = (\theta - c\theta/n)^2 + c^2\theta^2/n^2 = (1 - 2c/n + 2c^2/n^2)\theta^2$$

- $R_{\bar{X}}(\theta) < R_{cX_{(1)}}(\theta) \Leftrightarrow \theta^2/n < (1 - 2c/n + 2c^2/n^2)\theta^2$
  $\Leftrightarrow 0 < 2c^2 - 2cn + n^2 - n.$
- If $n > 2$, this always holds. If $n = 2$, this holds if $c \neq 1$. If $n = 1$, this holds iff $c > 1$.

Exercise 2

Consider the estimation of an unknown parameter $\theta \geq 0$ under the squared error loss. Show that if T and U are two estimators such that $T \leq U$ and $R_T(P) < R_U(P)$, then $R_{T_+}(P) < R_{U_+}(P)$, where f_+ denotes the positive part of f.

Proof:

- Note that $T = T_+ - T_-$, we have

$$\begin{aligned} R_T(P) &= E(T-\theta)^2 \\ &= E\left(T_+ - T_- - \theta\right)^2 \\ &= E\left(T_+ - \theta\right)^2 + E\left(T_-^2\right) + 2\theta E\left(T_-\right) - 2E\left(T_+ T_-\right) \\ &= R_{T_+}(P) + E\left(T_-^2\right) + 2\theta E\left(T_-\right) \end{aligned}$$

- Similarly, $R_U(P) = R_{U_+}(P) + E\left(U_-^2\right) + 2\theta E\left(U_-\right)$
- $T \leq U \Rightarrow T_- \geq U_-$. Also note that $\theta \geq 0$. Hence,

$$E\left(T_-^2\right) + 2\theta E\left(T_-\right) \geq E\left(U_-^2\right) + 2\theta E\left(U_-\right)$$

- Since $R_T(P) < R_U(P)$, we have $R_{T+}(P) < R_{U_+}(P)$

Exercise 3

Let X be a random variable having the binomial distribution $\mathrm{Bi}(p, n)$ with an unknown $p \in (0, 1)$ and a known n. Consider the problem of estimating $\theta = p^{-1}$. Show that there is no unbiased estimator of θ.

Proof:

- Suppose that $T(X)$ is an unbiased estimator of p^{-1}, i.e.,

$$\frac{1}{p} = E[T(X)] = \sum_{k=0}^{n} \binom{n}{k} T(k) p^k (1-p)^{n-k}, \quad \forall p \in (0,1) \tag{13}$$

- Note that for all p, we have an upper-bound on the RHS of Eq (13).

$$\sum_{k=0}^{n} \binom{n}{k} T(k) p^k (1-p)^{n-k} \leq \sum_{k=0}^{n} \binom{n}{k} T(k) < \infty$$

- But letting $p \to 0$, the LHS of Eq (13) diverges to ∞. So Eq (13) is impossible and there is no unbiased estimator of p^{-1}