ST5215 Advanced Statistical Theory, Lecture 13

HUANG Dongming

National University of Singapore

29 Sep 2020

Overview

Last time

- Methods of finding the UMVUE

Today

- More on finding UMVUEs
- Some remarks on earlier materials
- HW 2 solutions

Recap: Finding UMVUE (1)

- Lehmann-Scheffé Theorem: if $T(X)$ is a sufficient and complete statistic and $\hat{\theta}$ is an unbiased estimator of θ, then the UMVUE of θ is equivalent to $E(\hat{\theta} \mid T)$
- First method:
  1. Find a sufficient and complete statistic T and its distribution
  2. Find a function h such that $Eh(T) = \theta$ for all $P \in \mathcal{P}$
- Second method:
  1. Find an unbiased estimator of θ, say $\hat{\theta}$
  2. Find a sufficient and complete statistic T
  3. Compute $E(\hat{\theta} \mid T)$

Recap: Finding UMVUE (2)

In $L_2(\Omega, \mathcal{F}, P)$, the UMVUE is orthogonal to all "0-estimators"

- If S is an unbiased estimator of θ with finite variance, then
  S is a UMVUE of θ if and only if
  for any $P \in \mathcal{P}$ and any U that is unbiased for 0 with finite variance, it holds that
  $$E[SU] = 0$$

- If $S = h(T)$ is an unbiased estimator of θ with finite variance, where T is a sufficient statistic, then
  S is a UMVUE of θ if and only if
  for any $P \in \mathcal{P}$ and any $U(T)$ that is unbiased for 0 with finite variance, it holds that
  $$E[SU(T)] = 0$$

Example: Uniform Distributions

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on the interval $(0, \theta)$, where $\theta \in \Theta = [1, \infty)$. We know $X_{(n)}$ is sufficient for θ.

- To find a UMVUE of θ, we need to characterize the "0-estimators"
- Last lecture:
  - $U(X_{(n)})$ is an unbiased estimator of 0 $\Leftrightarrow$ $U(x) = 0$ a.e. on $[1, \infty)$ and

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

  - Since such a nontrivial $U(x)$ function exists, $X_{(n)}$ is not complete
- $T = h(X_{(n)})$ is a UMVUE $\Leftrightarrow$ $E[h(X_{(n)})U(X_{(n)})] = 0$ for any θ and any U that is "0-estimator"
- We need $h(x)U(x) = 0$ a.e. on $[1, \infty)$ and $\int_0^1 h(x)U(x)x^{n-1}dx = 0$
- We can choose

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

  so that $Eh(X_{(n)}) = \theta$

Example (Cont.)

We can show that $\max(X_{(n)}, 1)$ is complete and sufficient.
Sufficiency: The joint p.d.f. of $X_1, ..., X_n$ is

$$\frac{1}{\theta^n} I_{(0,\theta)}(X_{(n)}) = \frac{1}{\theta^n} I_{(0,\theta)}(\max(X_{(n)}, 1)),$$

because $\theta \geq 1$
Completeness: Suppose $E[f(\max(X_{(n)}, 1))] = 0$ for all $\theta > 1$, i.e.

$$0 = \int_0^1 f(1) x^{n-1} dx + \int_1^\theta f(x) x^{n-1} dx, \qquad \forall \theta > 1 \tag{2}$$

Letting $\theta \to 1$ we obtain that $f(1) = 0$. Then

$$0 = \int_1^\theta f(x) x^{n-1} dx, \qquad \forall \theta > 1, \tag{3}$$

which implies $f(x) = 0$ a.e. for $x > 1$.

Example: Uniform Distributions with Fixed Length

Let X be a sample (of size 1) from the uniform distribution Unif$(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathcal{R}$. Let g be a non-constant smooth function. Then there is no UMVUE of $\eta = g(\theta)$.

- To find a UMVUE of θ, we need to characterize the "0-estimators"
- An unbiased estimator $U(X)$ of 0 must satisfy

$$\int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} U(x)dx = 0 \qquad \text{for all } \theta \in \mathcal{R}.$$

- Differentiating both sides of the previous equation and applying the result of differentiation of an integral:

$$U(x) = U(x+1) \quad \text{a.e.}$$

- That is, U is unbiased estimator of 0 $\Leftrightarrow$ U has a period of 1 (a.e.) and $\int_0^1 U(x)\, \mathrm{d}x = 0$

Example (Cont.)

- If T is an UMVUE of $g(\theta)$, then for any $U(X)$ unbiased estimator of 0 with finite variance, $T(X)U(X)$ is unbiased for 0
- Using the result on the last slide, we have

$$T(x)U(x) = T(x+1)U(x+1) \quad \text{a.e.}$$

- Choosing $U(x) = sin(2\pi x + a)$ with any $a \in \mathcal{R}$, we conclude

$$T(x) = T(x+1) \quad \text{a.e.}$$

- Since T is unbiased for $g(\theta)$,

$$g(\theta) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} T(x)dx \qquad \text{for all } \theta \in \mathcal{R}$$

- Differentiating both sides and applying the result of differentiation of an integral

$$g'(\theta) = T\left(\theta + \tfrac{1}{2}\right) - T\left(\theta - \tfrac{1}{2}\right) \qquad \text{a.e.}$$

- But RHS=0, a.e., which forces g to be a constant function

Remarks on Early Materials

- Intuition for Factorization Theorem
- Sample variance
- Proving minimal sufficiency by Theorem A+B

Intuition for Factorization Theorem

Theorem

Suppose that X is a sample from $P \in \mathcal{P}$ and $\mathcal{P}$ is dominated by a σ-finite measure ν. Then $T(X)$ is sufficient for $P \in \mathcal{P}$ **if and only if** *there are nonnegative Borel functions $h(x)$ and $g_P(t)$ such that*

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x). \tag{4}$$

Consider the special case where ν is a counting measure on Ω, which is countable.

- "$\Rightarrow$": If T is sufficient, for any $x \in \Omega$, we have

  $$P(X = x) = P\left(T(X) = T(x)\right) \cdot P\left(X = x \mid T(X) = T(x)\right)$$

  Let $g_p(t) := P(T(X) = t)$ and $h(x) := P(X = x \mid T(X) = T(x))$
- "$\Leftarrow$": Suppose $P(X = x) = g_p(T(x))h(x)$ holds. $P\left(X = x \mid T(X) = T(x)\right) = \frac{P(X=x)}{\int I_{T(y)=T(x)} P(X=y)\ \mathrm{d}\nu} = \frac{h(x)}{\int I_{T(y)=T(x)} h(y)\ \mathrm{d}\nu}$, which does not depend on P. Hence, T is sufficient for $P \in \mathcal{P}$

Sample Variance (1)

The **sample variance** is defined as $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2$.

- If X_i's are i.i.d. with finite variance, then $ES^2 = \mathrm{Var}(X_1)$
- If $n = 1$, S^2 is not well-defined;it is unreasonable to estimate both the mean and the variance using only one sample
- If $n > 1$, $n - 1$ is usually referred as the *degree-of-freedom* (DOF)
    - The DOF of n samples is n. Estimating the mean by $\bar{X}$ takes up 1 DOF and there are $n - 1$ left
    - Originally, $x_1, \ldots, x_n$ can be n arbitrary numbers; once we computed $\bar{x}$, we had a constraint that $n\bar{x} = \sum_i x_i$
    - Under this constraint, $(x_1, \ldots, x_n)$ must lie on a $(n - 1)$-dimensional hyperplane

Sample Variance (2)

- Sometimes, we may want to divide the squared deviation by n rather than $(n-1)$
- To avoid confusion, we call $n^{-1}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2$ the **empirical variance** in this course
- The **empirical distribution** is defined as

$$P_n(A) = n^{-1}\#\{i : X_i \in A\}$$

- If $Y \sim P_n$ then $\mathrm{Var}(Y) = n^{-1}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2$.

An Exercise on Page 20 in Lecture 7

Let $X_1, \ldots, X_n \sim P_\theta = U(\theta, \theta+1)$ for $\theta \in \mathcal{R}$, where $n > 1$. The joint p.d.f. is $f_\theta(x) = (x_i) = I_{(x_{(n)}-1, x_{(1)})}(\theta)$ and $T = (X_{(1)}, X_{(n)})$ is sufficient.

- Previously, we've used Theorem C to show that T is minimal sufficient. Here we use Theorem A+B
- Suppose $\mathbb{Q} = \{\theta_i : i = 1, 2, \ldots\}$. Define $\mathcal{P}_0 = \{f_{\theta_i}\}$. We need to show T is minimal sufficient for $\mathcal{P}_0$ and $\mathcal{P}_0$-a.s. implies $\mathcal{P}$-a.s.
- Let $f_\infty = \sum_i 2^{-i} f_i$. By Theorem B, $S = (f_1/f_\infty, f_2/f_\infty, \ldots,)$ is minimal sufficient for $\mathcal{P}_0$
- Note that

  $$S_i(x) > 0 \Leftrightarrow f_{\theta_i}(x) > 0 \Leftrightarrow x_{(1)} > \theta_i \text{ and } x_{(n)} - 1 < \theta_i,$$

  we have

  $$x_{(1)} = \sup\{\theta_i : S_i(x) > 0\}, \qquad x_{(n)} = 1 + \inf\{\theta_i : S_i(x) > 0\}$$

- In other words, $T(x) = \phi(S(x))$ for some measurable function ϕ (see Proposition 1.4 in JS); so T is minimal sufficient for $\mathcal{P}_0$

Tutorial

1. Exercise 2.6.35 in JS
2. Exercise 2.6.53 in JS
3. Let $(X_1, \ldots, X_n)$ be a random sample from $E(0, 100)$ (See Table 1.2). Use Basu's theorem to show that $X_n^4 / \sum_{j=1}^n X_j^4$ and $\sum_{j=1}^n X_j$ are independent, $i = 1, \ldots, n$
4. Let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim N\left(\alpha + \beta x_i, \sigma^2\right), i = 1, \ldots, n$ where $x_1, \ldots, x_n$ and σ^2 are known constants, and α and β are unknown parameters. We assume x_i's are not equal.
    1. Use the idea behind the method of moments to find an estimator of (α, β) (Hint: consider $\sum_i EY_i$ and $\sum_i E[Y_i x_i]$
    2. Find the maximum likelihood estimators $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ of $\theta = (\alpha, \beta)$
    3. Is the $\hat{\beta}$ you found in (b) unbiased? What is its MSE?
5. Exercise 2.6.63 in JS
6. Consider estimating success probability $\theta \in [0,1]$ from data $X \sim \text{Binomial}(n, \theta)$ under squared error loss. Define $\delta_{a,b}$ by

$$\delta_{a,b}(X) = a\frac{X}{n} + (1 - a)b.$$

Exercise 2.6.35 in JS

Let $X_1, \ldots, X_n$ be i.i.d. random variables having the Lebesgue p.d.f.

$$f_\theta(x) = (2\theta)^{-1}\left[I_{(0,\theta)}(x) + I_{(2\theta,3\theta)}(x)\right]$$

Find a minimal sufficient statistic for $\theta \in (0, \infty)$.

Solution:

- Let $\mathcal{P} = \{g_\theta : \theta > 0\}$, where $g_\theta(x) = \prod_{i=1}^n f_\theta(x_i)$
- Let $\Theta_0 = \{\theta_1, \theta_2, \ldots\}$ be the set of positive rational numbers and $\mathcal{P}_0 = \{g_\theta : \theta \in \Theta_0\}$
- $\mathcal{P}_0 \subset \mathcal{P}$ and $\mathcal{P}_0$-a.s. implies $\mathcal{P}$ -a.s.
- Let $g_\infty(x) = \sum_{i=1}^\infty 2^{-i} g_{\theta_i}(x)$
- Define $T = (T_1, T_2, \ldots)$ with $T_i(x) = g_{\theta_i}(x)/g_\infty(x)$. By Theorem B (Theorem 2.3(ii) in JS), T is minimal sufficient for $\mathcal{P}_0$
- By Theorem A (Theorem 2.3(i) in JS), it remains to show that T is sufficient for $\mathcal{P}$

Exercise 2.6.35 in JS (Cont.)

- Let $\phi(t) = \overline{\lim_{k\to\infty}} t_k$ for $t = (t_1, t_2, \ldots)$
- For any $\theta > 0$, there is a nondecreasing sub-sequence $\{\theta_{i_k}\} \subset \Theta_0$ such that $\lim_k \theta_{i_k} = \theta$
- For this θ, define a function $\psi_\theta(t) = (t_{i_1}, t_{i_2}, \ldots, \theta_{i_k}, \ldots)$ for $t = (t_1, t_2, \ldots)$
- Then for any $x \in \mathcal{R}_+^n$ such that $x_j \neq \theta$ for all j,

$$g_\theta(x) = \lim_k g_{\theta_{i_k}}(x) = g_\infty(x) \lim_k T_{i_k}(x) = g_\infty(x)\phi(\psi_\theta(T(x)))$$

- By the factorization theorem, T is sufficient for $\mathcal{P}$

Exercise 2.6.53 in JS

Let X be a discrete random variable with p.d.f.

$$f_\theta(x) = \begin{cases} \theta & x = 0 \\ (1-\theta)^2\theta^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $\theta \in (0,1)$. Show that X is boundedly complete, but not complete.

Proof:

- Suppose $h(x)$ such that $E[h(X)] = 0$ for all $\theta \in (0,1)$
- Then

$$h(0)\theta + \sum_{x=1}^{\infty} h(x)(1-\theta)^2\theta^{x-1} = 0, \qquad \forall\theta \in (0,1)$$

- Rewriting the LHSin the ascending order of the powers of θ, we obtain that

$$h(1) + \sum_{x=1}^{\infty}[h(x-1) - 2h(x) + h(x+1)]\theta^x = 0, \qquad \forall\theta \in (0,1)$$

Exercise 2.6.53 in JS (Cont.)

$$h(1) + \sum_{x=1}^{\infty} [h(x-1) - 2h(x) + h(x+1)]\theta^x = 0, \qquad \forall \theta \in (0,1)$$

- Comparing the coefficients of both sides, we obtain that $h(1) = 0$ and $h(x-1) - h(x) = h(x) - h(x+1)$ for all $x = 1, 2, \ldots,$
- Therefore $h(x) = (1-x)h(0)$ for all $x = 1, 2, \ldots,$
- If $h(x)$ is assumed to be bounded, then $h(0) = 0$, and thus $h(x) = 0$, for all $x \in \mathcal{N}$. This means that X is boundedly complete.
- If $h(x) = 1 - x$, we have $E[h(X)] = 0$ for any θ but $h(X) \neq 0$. Therefore, X is not complete

Problem 3

Let $(X_1, \ldots, X_n)$ be a random sample from $E(0, 100)$ (See Table 1.2). Use Basu's theorem to show that $X_n^4 / \sum_{j=1}^n X_j^4$ and $\sum_{j=1}^n X_j$ are independent, $i = 1, \ldots, n$

Proof:

- Let $\mathcal{P} = \{\mathrm{E}(0, \theta) : \theta > 0\}$. We can postulate the model $P \in \mathcal{P}$ because $P = \mathrm{E}(0, 100)$
- Since $\mathcal{P}$ is a natural exponential family of full rank, $\sum_{j=1}^n X_j$ is sufficient and complete
- Represent $X_i = \theta Y_i$ where $Y_i \sim \mathrm{E}(0, 1)$
- Since $X_n^4 / \sum_{j=1}^n X_j^4 = \theta^4 Y_n^4 / (\theta^4 \sum_{j=1}^n Y_j^4) = Y_n^4 / \sum_{j=1}^n Y_j^4$ does not depend on θ, it is ancillary
- By Basu's theorem, $X_n^4 / \sum_{j=1}^n X_j^4$ and $\sum_{j=1}^n X_j$ are independent

Problem 4

Let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim N\left(\alpha + \beta x_i, \sigma^2\right)$, $i = 1, \ldots, n$ where $x_1, \ldots, x_n$ and σ^2 are known constants, and α and β are unknown parameters. We assume x_i's are not equal.

1. Use the idea behind the method of moments to find an estimator of (α, β) (Hint: consider $\sum_i EY_i$ and $\sum_i E[Y_i x_i]$
2. Find the maximum likelihood estimators $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ of $\theta = (\alpha, \beta)$
3. Is the $\hat{\beta}$ you found in (b) unbiased? What is its MSE?

Solution: Part (1)

- Note that $EY_i = \alpha + \beta x_i$ and $E(x_i Y_i) = x_i \alpha + \beta x_i^2$
- Equate

$$\sum_i Y_i = n\hat{\alpha} + \hat{\beta} \sum_i x_i,$$
$$\sum_i Y_i x_i = \hat{\alpha} \sum_i x_i + \hat{\beta} \sum_i x_i^2,$$

- Solve: let $\overline{Yx} = n^{-1} \sum_i Y_i x_i$ and $\overline{x^2} = n^{-1} \sum_i x_i^2$

$$\hat{\beta} = \frac{\overline{Yx} - \bar{Y}\bar{x}}{\overline{x^2} - \bar{x}^2}, \qquad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$$

Problem 4 (MLE)

Part (2).

- The log likelihood is

$$\ell(\alpha, \beta) = -\frac{1}{2\sigma^2}\sum_i (Y_i - \alpha - \beta x_i)^2 - \frac{n}{2}\log\left(2\pi\sigma^2\right)$$

- The maximum of $\ell(\alpha, \beta)$ is the minimum of the sum of squared residuals

$$r(\alpha, \beta) = \sum_i (Y_i - \alpha - \beta x_i)^2$$

- Set $\partial r/\partial\alpha$ and $\partial r/\partial\beta$ to 0:

$$\sum_i -(Y_i - \alpha - \beta x_i) = 0,$$

$$\sum_i -(Y_i - \alpha - \beta x_i)x_i = 0$$

- The solution is the same as the MOM estimator and is the unique MLE because $r(\alpha, \beta)$ is convex

Problem 4 ($\hat{\beta}$ is unbiased)

- Note that

$$E\left(\overline{Yx}\right) = n^{-1}\sum_i E(Y_i x_i)$$
$$= n^{-1}\sum_i (\alpha x_i + \beta x_i^2)$$
$$= \alpha\bar{x} + \beta\overline{x^2},$$

and $E\overline{Y} = n^{-1}\sum_i E(Y_i) = n^{-1}\sum_i(\alpha + \beta x_i) = \alpha + \beta\bar{x}$

-

$$E\hat{\beta} = E\left(\frac{\overline{Yx} - \bar{Y}\bar{x}}{\overline{x^2} - \bar{x}^2}\right)$$
$$= \frac{1}{\overline{x^2} - \bar{x}^2}\left[E\left(\overline{Yx}\right) - E\left(\overline{Y}\right)\bar{x}\right]$$
$$= \frac{1}{\overline{x^2} - \bar{x}^2}\left(\alpha\bar{x} + \beta\overline{x^2} - \alpha\bar{x} - \beta\bar{x}\bar{x}\right)$$
$$= \beta$$

Problem 4 (Compute MSE)

- Since $\hat{\beta}$ is unbiased, its MSE equals to its variance
- Note that $\hat{\beta}$ is a linear combination of Y_i's, and the coefficient of each Y_i is

$$\frac{n^{-1}}{\overline{x^2} - \bar{x}^2}(x_i - \bar{x}),$$

so the variance of $\hat{\beta}$ is

$$\sum_i \left[\frac{n^{-1}}{\overline{x^2} - \bar{x}^2}(x_i - \bar{x})\right]^2 \sigma^2.$$

- Note that $n^{-1}\sum_i (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$, the variance equals to

$$\frac{\sigma^2}{n(\overline{x^2} - \bar{x}^2)}$$

Exercise 2.6.63 in JS

Let $X_1, \ldots, X_n$ be i.i.d. from the $N\left(\mu, \sigma^2\right)$ distribution, where $\mu \in \mathcal{R}$ and $\sigma > 0$. Consider the estimation of σ^2 with the squared error loss. Show that $\frac{n-1}{n} S^2$ is better than S^2, the sample variance. Can you find an estimator of the form cS^2 with a nonrandom c such that it is better than $\frac{n-1}{n} S^2$?

Proof:

- In Lecture 8 (Page 16), we know that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$
- We can represent $(n-1)S^2 = \sigma^2 \sum_{i=1}^{n-1} Z_i^2$ where Z_i's are i.i.d. standard normal r.v.s.
- Since $ES^2 = \sigma^2$, the bias of cS^2 is $(c-1)\sigma^2$
- Since the variance of Z_i^2 is 2, the variance of cS^2 is

$$\frac{c^2\sigma^4}{(n-1)^2} \times (n-1) \times 2$$

- The MSE of cS^2 is $\sigma^4\left[(c-1)^2 + 2c^2/(n-1)\right]$
- The quadratic function $(c-1)^2 + 2c^2/(n-1)$ is minimized at $c_* = (n-1)/(n+1)$ and is increasing on $(c_*, 1]$.
- $c_* < (n-1)/n < 1$: $c_* S^2$ is $\mathfrak{J}$-optimal and $\frac{n-1}{n} S^2$ is better than S^2

Problem 6

Consider estimating success probability $\theta \in [0,1]$ from data $X \sim \text{Binomial}(n, \theta)$ under squared error loss. Define $\delta_{a,b}$ by

$$\delta_{a,b}(X) = a\frac{X}{n} + (1-a)b.$$

which might be called a linear estimator, because it is a linear function of X

1. Find the variance and bias of $\delta_{a,b}$.
2. If $a > 1$, show that $\delta_{a,b}$ is inadmissible by finding a competing linear estimator with better risk. Hint: Find an unbiased estimator with smaller variance.
3. If $b > 1$ or $b < 0$, and $a \in [0, 1)$, show that $\delta_{a,b}$ is inadmissible by finding a competing linear estimator with better risk. Hint: Find an estimator with the same variance but better bias.
4. If $a < 0$, find a linear estimator with better risk than $\delta_{a,b}$

Solution: Part (1).

- $E\delta_{a,b} = a\theta + (1-a)b$. The bias is $(1-a)(b-\theta)$
- The variance is $a^2/n^2 \times n\theta(1-\theta) = a^2\theta(1-\theta)/n$

Problem 6 (Cont.)

The MSE of $\delta_{a,b}$ is $(1-a)^2(b-\theta)^2 + a^2\theta(1-\theta)/n$.

Part (2).

- If $a > 1$, consider $\delta_{1,b}$
- Its squared bias is 0 and its variance is $\theta(1-\theta)/n$, each of which is smaller than that of $\delta_{a,b}$
- So $\delta_{1,b}$ is better than $\delta_{a,b}$

Part (3).

- If $b > 1$ and $a \in [0, 1)$, consider $\delta_{a,1}$
- Its variance equals to that of $\delta_{a,b}$
- Its squared bias is $(1-a)^2(1-\theta)^2 < (1-a)^2(b-\theta)^2$ because $\theta \leq 1$
- So $\delta_{a,1}$ is better than $\delta_{a,b}$ if $b > 1$
- Similarly, $\delta_{a,0}$ is better than $\delta_{a,b}$ if $b < 0$

Part (4).

- If $a < 0$, consider $\delta_{-a,b}$
- Its variance equals to that of $\delta_{a,b}$
- Its squared bias is $(1+a)^2(b-\theta)^2 < (1-a)^2(b-\theta)^2$, $\forall\theta \neq b$
- So $\delta_{-a,b}$ is better than $\delta_{a,b}$