

Lecture 12: Unbiased estimation and UMVUE

Lecturer: Yiqiao Zhong

Last update: 18:23:30, 2024-10-16

Disclaimer: materials in this pdf are used for STAT 709 “Mathematical Statistics” at the University of Wisconsin–Madison, but they are not guaranteed to be accurate. Consult textbooks whenever possible.

In previous lectures, we showed that bias-variance tradeoff is a general phenomenon in statistics. An optimal estimator may not exist or depends on a case-by-case analysis. However, if we only consider unbiased estimators, there is a systematic way of finding optimal estimators.

1 Complete statistics

Recall that an estimator $T(X)$ based on data $X = (X_1, \dots, X_n)$ drawn from $P \in \mathcal{P}$ is sufficient iff it retains relevant information of inferring the unknown distribution P , or in other words the conditional distribution $X|T(X)$ does not on the choice of P . However, there may be redundant information in $T(X)$; e.g., $T(X) = X$ is trivially sufficient but not practically useful.

We introduce completeness as a criterion for good summary statistics¹.

Definition 1 (Complete Statistics). $T(X)$ is complete for $P \in \mathcal{P}$ if and only if for all Borel functions f , $\mathbb{E}[f(T)] = 0, \forall P \in \mathcal{P}$ implies that $f(T) = 0$ a.s. P .

From this definition, if $T(X)$ is complete and ψ, ψ^{-1} are one-one measurable functions, then $\psi(T(X))$ is also complete. It is easy to find complete statistics in an exponential family.

Proposition 1. Let $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$ be an exponential family of full rank (contains an open set) with p.d.f. given by

$$f_\eta(x) = \exp\{\eta^T T(X) - \xi(\eta)\} h(x).$$

Then $T(X)$ is sufficient and complete for $\eta \in \Xi$.

Below are examples taken from [1], 2.15–2.17.

Example 1. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. The joint p.d.f. of $X_1, \dots, X_n$ is

$$(2\pi)^{-n/2} \exp\{\eta_1 T_1 + \eta_2 T_2 - n\xi(\eta)\},$$

where $T_1 = \sum_{i=1}^n X_i$, $T_2 = -\sum_{i=1}^n X_i^2$, and $\eta = (\eta_1, \eta_2) = (\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2})$. Thus, $T = (T_1, T_2)$ is complete and sufficient for η . Also, $(\bar{X}, S^2)$ is complete and sufficient for $\theta = (\mu, \sigma^2)$.

Example 2. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P_\theta = U(0, \theta)$, $\theta > 0$. Check that $X_{(n)}$ is complete and sufficient for $\theta \in [0, \infty)$.

Example 3. Let $\mathcal{P}$ be a family of distributions on $\mathbb{R}$ having Lebesgue p.d.f. Then $T(X) = (X_{(1)}, \dots, X_{(n)})$ is complete for $\mathcal{P}$.

¹A related notion is *minimal sufficient*. It is known that a complete and sufficient statistic is minimal sufficient, whereas a minimal sufficient statistic is not guaranteed to be complete and sufficient.

Proof. Let $\mathcal{P}_0$ be exponential family

$$f(x) = C(\theta_1, \dots, \theta_n) \exp(-x^{2n} + \theta_1 x + \theta_2 x^2 + \dots + \theta_n x^n),$$

where $C(\theta_1, \dots, \theta_n)$ is the normalizing constant. $\mathcal{P}_0 \subset \mathcal{P}$, and $(\sum X_i, \dots, \sum X_i^n)$ is complete for $\mathcal{P}_0$. Since a.s. $\mathcal{P}_0$ implies a.s. $\mathcal{P}$, $(\sum X_i, \dots, \sum X_i^n)$ is complete for $\mathcal{P}$. There exists one-one mapping $(\sum X_i, \dots, \sum X_i^n) \leftrightarrow (X_{(1)}, \dots, X_{(n)})$. Therefore, $(X_{(1)}, \dots, X_{(n)})$ is complete for $\mathcal{P}$. $\square$

2 UMVUE

Recall that we say $T(X)$ is unbiased for an unknown parameter ν if and only if

$$\mathbb{E}[T(X)] = \nu, \quad \forall P \in \mathcal{P}.$$

Definition 2 (UMVUE). *Suppose that $T(X)$ has finite second moment. We say $T(X)$ of ν is uniformly minimum variance unbiased estimator (UMVUE) if and only if*

$$\text{Var}(T(X)) \leq \text{Var}(U(X)), \quad \forall P \in \mathcal{P}, \quad \forall \text{ unbiased } U(X).$$

It is clear that UMVUE implies optimality among all unbiased estimators in terms of mean squared error (MSE). For general loss functions instead of the quadratic loss, “minimum variance” is replaced by “minimum risk”.

Remark 1. *UMVUE does not always exist a priori. For fixed $P \in \mathcal{P}$, we can solve*

$$\min_{\text{unbiased } T} \text{Var}(T(X)),$$

but not guaranteed to be uniform for every $P \in \mathcal{P}$.

The following Lehmann-Scheffé theorem [1][Thm. 3.1] states that given a sufficient and complete statistic, often we can find a UMVUE.

Theorem 1 (Lehmann-Scheffé). *Suppose that there exists a sufficient and complete statistic $T(X)$ for $\mathcal{P}$. If ν is estimable, then there exists a unique² UMVUE, which is of the form $h(T)$ where h is a Borel function.*

Proof. This is a consequence of Rao-Blackwell Theorem and completeness, which we reason below. If $U(X)$ is unbiased for ν , then

$$\text{Var}(\mathbb{E}[U|T]) \leq \text{Var}(U).$$

Note that $\mathbb{E}[U|T]$ is a function of T , i.e. $\mathbb{E}[U|T] = h(T)$. By completeness, h does not depend on the choice of U , so $h(T)$ has minimum variance among all unbiased estimators U , which implies that $h(T)$ is UMVUE.

If $\tilde{T}$ is another UMVUE, then the variance of $\mathbb{E}[\tilde{T}|T]$ must be no larger than that of $h(T)$, thus equal. Since MSE is strictly convex loss, $\tilde{T}$ must be a function of T (by Theorem 2.5). Therefore, $\tilde{T} = h(T)$ a.s. $\forall P \in \mathcal{P}$. $\square$

3 Three ways to find UMVUE

We will introduce three ways of finding UMVUEs. Examples are taken from [1][Section 3.1].

²Up to a.s., $\forall P \in \mathcal{P}$.

3.1 Method 1

Find sufficient and complete statistic $T(X)$, then find $h(T)$, such that

$$\mathbb{E}[h(T)] = \nu, \quad \forall P \in \mathcal{P}.$$

Example 4. (Example 3.1) Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Unif}([0, \theta])$, $\theta > 0$. Let $\nu = g(\theta) \in C^1$. We know that $X_{(n)} = \max_{1 \leq i \leq n} X_i$ is complete, with p.d.f. given by

$$f_\theta(x) = n \frac{1}{\theta^n} x^{n-1} I_{(0, \theta)}(x).$$

Our goal is to find h , such that

$$\theta^n g(\theta) = n \int_0^\theta h(x) x^{n-1} dx.$$

Taking derivative on both sides, we get

$$n\theta^{n-1}g(\theta) + \theta^n g'(\theta) = nh(\theta)\theta^{n-1}.$$

We can get a formula for h by re-arranging the equation above. In particular, if $\nu = \theta$, then

$$h(X_{(n)}) = (1 + \frac{1}{n})X_{(n)}.$$

Example 5. (Example 3.2) Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta)$, $\theta > 0$ unknown.

Example 6. (Example 3.6) Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P \in \mathcal{P}$, where $\mathcal{P}$ is nonparametric family with distributions having Lebesgue p.d.f. Let $T = (X_{(1)}, \dots, X_{(n)})$ be a complete and sufficient statistic. Note that $\varphi(X_1, \dots, X_n)$ is a function of T if and only if for any permutation π ,

$$\varphi(X_1, \dots, X_n) = \varphi(X_{\pi(1)}, \dots, X_{\pi(n)}).$$

We use T to construct UMVUE.

- (1) $\nu = \mathbb{E}[X_i]$. Then $\bar{X}$.
- (2) $\nu = \text{Var}(X_i)$. Then S^2 .
- (3) $\nu = (\mathbb{E}[X_i])^2$. Then $\frac{1}{n} \sum_{i=1}^n X_i^2 - S^2$. Note that generally $\mathbb{E}[\bar{X}^2] > (\mathbb{E}[\bar{X}])^2 = (\mathbb{E}[X_i])^2$.
- (4) $\nu = \mathbb{P}(X_i \leq t)$. Then $F_n(t) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(X_i)$.

3.2 Method 2

Find a sufficient and complete $T(X)$ and an unbiased $U(X)$, then apply conditioning (as in the proof at Sec 2). Note that any unbiased $U(X)$ (as long as its second moment exists) is okay. We have the freedom to choose one so that $\mathbb{E}[U|T]$ is easy to simplify.

Example 7 (Binomial). Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Bin}(k, \theta)$, where k is known, $\theta \in (0, 1)$ is unknown. Need to calculate

$$\mathbb{E}[I(X_i = 1)|T] = \mathbb{P}(X_i = 1|T),$$

where $T = \sum_{i=1}^n X_i$. You can then calculate the conditional probability (check the book).

Example 8. (Example 3.3) Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} E(\theta)$. We know that

$$F_\theta(X) = (1 - e^{-x/\theta})I_{(0,\infty)}(x).$$

Let $\nu = 1 - F_\theta(t)$, then

$$T(X) = \mathbb{P}(X_1 > t | \bar{X}),$$

is UMVUE. Note that $\bar{X}$ is complete, and moreover, $\frac{X_1}{\bar{X}}$ is ancillary since $E(\theta)$ only has a scale parameter. By Basu's Theorem (see Theorem 3), $\bar{X}$ is independent of $\frac{X_1}{\bar{X}}$. So,

$$\mathbb{P}(X_1 > t | \bar{X} = \bar{x}) = \mathbb{P}\left(\frac{X_1}{\bar{X}} > \frac{t}{\bar{x}} | \bar{X} = \bar{x}\right) = \mathbb{P}\left(\frac{X_1}{\bar{X}} > \frac{t}{\bar{x}}\right).$$

It remains to derive the p.d.f. of

$$\frac{X_1}{X_1 + \sum_{i>1} X_i}.$$

Note that X_1 is independent of $\sum_{i>1} X_i$, and $\sum_{i>1} X_i$ has a Gamma distribution. We can use the change of variables formula (Proposition 1 in Lecture 3) to calculate the explicit p.d.f of this transformation ([exercise](#)).

Example 9. (Nonexistence of UMVUE) See slide 16, page 12.

Remark 2. Let $\mathcal{U}$ be the set of all unbiased estimators,

$$\text{Var}(T(X)) \leq \text{Var}(U(X)), \quad \forall P \in \mathcal{P}, \quad \forall U \in \mathcal{U}.$$

If $\mathcal{P}$ is small, $\mathcal{U}$ may be large; and conversely if $\mathcal{P}$ is large, then $\mathcal{U}$ may be small. A larger parameter space does not imply it is harder to find a UVMUE.

3.3 Method 3

Find UMVUE without knowing complete statistics.

Theorem 2 (Theorem 3.2 of [1]). Let

$$\mathcal{U} = \{U : \mathbb{E}[U(X)] = 0, \text{Var}(U(X)) < \infty, \forall P \in \mathcal{P}\}.$$

T is unbiased for ν with $\mathbb{E}[T(X)^2] < \infty$.

(1) $T(X)$ is UMVUE if and only if

$$\mathbb{E}[T(X)U(X)] = 0, \quad \forall U \in \mathcal{U}, \quad \forall P \in \mathcal{P}.$$

(2) If $\tilde{T}$ is sufficient for $\mathcal{P}$, let

$$\tilde{\mathcal{U}} = \mathcal{U} \cap \{g(\tilde{T}) : g \text{ Borel}\}.$$

Then, $T = h(\tilde{T})$ is UMVUE if and only if

$$\mathbb{E}[T(X)U(X)] = 0, \quad \forall U \in \tilde{\mathcal{U}}, \quad \forall P \in \mathcal{P}.$$

proof idea. Calculus of variations. In optimization,

$$\min_{x \in \Omega} f(x) \Rightarrow \langle \nabla f(x^*), x - x^* \rangle = 0.$$

Strategy is true if Ω is in Hilbert space, where

$$f(x) \approx f(x^*) + \langle \nabla f(x^*), x - x^* \rangle.$$

□

As a consequence, UMVUE satisfies linearity; see [1][Corollary 3.1]. We can also use this theorem to show UMVUE does not exist in some cases.

Example 10. Let $X \sim \text{Unif}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathbb{R}$. If $\nu = g(\theta)$, where g is not a constant, no UMVUE exists. See proofs in [1][Example 3.8].

4 Further reading

Ancillary statistics

Definition 3. $V(X)$ is called ancillary if $V(X)$ does not depend on $\mathcal{P}$, i.e. its distribution remains the same for every $P \in \mathcal{P}$.

Given a family $\mathcal{P}$, an ancillary statistic provides no information for $\mathcal{P}$.

Example 11. If $X_1, \dots, X_n \sim \mathcal{N}(\mu, 1)$, then $X_i - X_j$ is ancillary.

Example 12. If $X_1, \dots, X_n \sim \mathcal{N}(0, \sigma^2)$, then $\frac{X_i}{X_j}$ and $\frac{X_{(i)}}{X_{(j)}}$ are both ancillary.

Theorem 3 (Basu's Theorem). If V is ancillary and T is bounded complete and sufficient for P . Then $V \perp T$ w.r.t. $\forall P \in \mathcal{P}$.

Example 13. Let $X_1, \dots, X_n \sim \mathcal{N}(\mu, 1)$. $\bar{X}$ is complete and sufficient, and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is ancillary. This implies $\bar{X} \perp S^2$.

References

[1] Jun Shao. *Mathematical Statistics*. Springer Science & Business Media, 2nd edition, 2013.