476 Statistics, Spring 2022.
Lecturer: Sergey Nadtochiy.

Lecture 4. Relative efficiency, sufficiency, completeness and MVUE. (Sections 9.1–9.5)

In this lecture, we learn how to choose a (point) estimator with the smallest MSE (i.e., the best estimator in terms of its finite-sample properties).

1 Relative efficiency

Recall that, for an unbiased estimator, its mean squared error (MSE) equals its variance.

Def 1. *For unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is*

$$\text{\textit{Eff}}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

If $\text{Eff}(\hat{\theta}_1, \hat{\theta}_2) > 1$, then $\hat{\theta}_1$ is better.

Ex 1. *Consider $Y_1, \dots, Y_n$ - i.i.d. $U(0, \theta)$ and the two unbiased estimators:*

$$\hat{\theta}_1 = 2\bar{Y}, \quad \hat{\theta}_2 = \frac{n+1}{n} Y_{(n)},$$

where $Y_{(k)}$ denotes the k-th smallest element of the sample (i.e., $Y_{(n)}$ is the largest one).

Q 1. *Find Eff$(\hat{\theta}_1, \hat{\theta}_2)$.*

We know that $\bar{Y}$ is unbiased. To verify that $\frac{n+1}{n} Y_{(n)}$ is unbiased We only need to recall the cdf of $Y_{(n)}$

$$\mathbb{P}(Y_{(n)} \leq y) = \mathbb{P}(Y_1 \leq y, \dots, Y_n \leq y) = (y/\theta)^n,$$

which gives us the pdf of $Y_{(n)}$:

$$\frac{n}{\theta}(y/\theta)^{n-1}, \quad y \in (0, \theta).$$

Thus,

$$\mathbb{E}\hat{\theta}_2 = \frac{n}{\theta^n} \int_0^\theta y^n dy = \frac{n}{n+1}\theta$$

Let's check the variances:

$$V(\hat{\theta}_1) = \frac{\theta^2}{3n},$$

$$\mathbb{E}\hat{\theta}_2^2 = \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{n}{n+2}\theta^2,$$

$$V(\hat{\theta}_2) = \frac{\theta^2}{n(n+2)},$$

$$\text{\textit{Eff}}(\hat{\theta}_1, \hat{\theta}_2) = \frac{3}{n+2}.$$

We conclude that for $n \geq 2$, the second estimator is better.

2 Sufficiency

Consider a sequence of indep. Bernoulli r.v.'s $\{X_n\}$ with unknown success probability p. Typically, we use $\bar{X}$ as an estimator for p.

Q 2. *Does there exist a different estimator with lower MSE?*

Note that

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n \,|\, \bar{X} = y/n) = \frac{p^y(1-p)^{n-y}}{C_n^y p^y (1-p)^{n-y}} = 1/C_n^y,$$

if $x_1 + \cdots x_n = y$ (and $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n \,|\, \bar{X} = y/n) = 0$ otherwise). As a side remark, in the above, we use C_n^y to denote the combinatorial expression "n choose y":

$$C_n^y = \frac{y!(n-y)!}{n!}.$$

The above means that, **given the value of $\bar{X}$, the conditional distribution of the sample is independent of** p! Intuitively, the latter means that, **given the value of $\bar{X}$, we cannot deduce any additional information about p from the sample values**.

Def 2. *A statistic U is* **sufficient** *for θ if the conditional distr. of $(Y_1, \dots, Y_n)$ given U does not depend on θ.*

The next theorem shows how one can improve any estimator by using a sufficient statistic.

Thm 1. *(Rao-Blackwell) Let $\hat{\theta}$ be an unbiased estimator for θ, with $V(\hat{\theta}) < \infty$, and let U be sufficient statistic. Then,*

$$\hat{\theta}^* := \mathbb{E}(\hat{\theta}|U)$$

is also an unbiased estimator, and $V(\hat{\theta}^*) \leq V(\hat{\theta})$.

Proof:

Note that $\hat{\theta}^*$ is a function of U and, hence, a function of the sample. This function, in general may depend on θ, as the conditional expectation depends on the joint distribution of $(\hat{\theta}, U)$, which in turn depends on θ. However, the defining property of sufficient statistic yields that the aforementioned function does not depend on θ. Thus, $\hat{\theta}^*$ is a statistic (and hence an estimator).

To see that $\hat{\theta}^*$ is unbiased, we use the tower property of conditional expectation:

$$\mathbb{E}\hat{\theta}^* := \mathbb{E}\left[\mathbb{E}(\hat{\theta}|U)\right] = \mathbb{E}\hat{\theta} = \theta.$$

Finally, let us check the variance. Note that

$$V(\hat{\theta}) = \mathbb{E}\hat{\theta}^2 - \theta^2, \quad V(\hat{\theta}^*) = \mathbb{E}(\hat{\theta}^*)^2 - \theta^2.$$

Hence, it suffices to show that $\mathbb{E}(\hat{\theta}^*)^2 \leq \mathbb{E}\hat{\theta}^2$, which is done as follows:

$$\mathbb{E}(\hat{\theta}^*)^2 = \mathbb{E}(\mathbb{E}(\hat{\theta}|U))^2 \leq \mathbb{E}(\mathbb{E}(\hat{\theta}^2|U)) = \mathbb{E}\hat{\theta}^2,$$

where we used Jensen's inequality and the tower property. ∎

Q 3. *How to find a sufficient statistic?*

Def 3. *For a sample of i.i.d. r.v.'s $Y_1, \dots, Y_n$, their* **likelihood function** $L(y_1, \dots, y_n; \theta)$ *is defined as*

$$L(y_1, \dots, y_n; \theta) := f(y_1; \theta) \cdots f(y_n; \theta),$$

where

- $f(\cdot; \theta)$ *is the density of Y_i, if the latter has absolutely continuous distribution (i.e., a distribution with a pdf),*
- *and $f(\cdot; \theta)$ is the probability function of Y_i, if the latter has a discrete distribution.*

Thm 2. *(Fisher-Neyman) U is a sufficient statistic if and only if there exist functions g and h such that*

$$L(y_1, \dots, y_n; \theta) = g(U(y_1, \dots, y_n), \theta) \cdot h(y_1, \dots, y_n).$$

(Note that **only g is allowed to depend on θ!***)*

Proof:

We will only show one implication: if U is sufficient, then the above decomposition holds. For concreteness, assume that Y_i has a discrete distribution. Consider any set of potential sample values $(y_1, \dots, y_n)$ and denote $z := U(y_1, \dots, y_n)$. Then,

$$\begin{aligned} &\mathbb{P}^\theta [Y_1 = y_1, \dots, Y_n = y_n] \\ &= \mathbb{P}^\theta [Y_1 = y_1, \dots, Y_n = y_n \,|\, U(Y_1, \dots, Y_n) = z] \cdot \mathbb{P}^\theta [U(Y_1, \dots, Y_n) = z] \\ &= h(y_1, \dots, y_n) \cdot g(U(y_1, \dots, y_n), \theta), \end{aligned}$$

where

$$h(y_1, \dots, y_n) := \mathbb{P}^\theta [Y_1 = y_1, \dots, Y_n = y_n \,|\, U(Y_1, \dots, Y_n) = z] \big|_{z=U(y_1,\dots,y_n)},$$

$$g(z, \theta) := \mathbb{P}^\theta [U(Y_1, \dots, Y_n) = z],$$

and we recall that h does not depend on θ by the definition of a sufficient statistic. ∎

Ex 2. *Let $\{Y_i\}_{i=1}^n$ be i.i.d. $Exp(1/\theta)$.*

Q 4. *Find a sufficient statistic for θ.*

Let us write the likelihood function:

$$L(y_1, \dots, y_n; \theta) = \theta^{-n} e^{-y_1/\theta} \cdots e^{-y_n/\theta} = \theta^{-n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^n y_i\right) = h(y_1, \dots, y_n) g(U(y_1, \dots, y_n), \theta),$$

where

$$U(y_1, \dots, y_n) := \bar{y}, \quad g(\bar{y}, \theta) := \theta^{-n} \exp\left(-\frac{n}{\theta}\bar{y}\right), \quad h(y_1, \dots, y_n) := 1.$$

Thus, $\bar{Y}$ is sufficient.

Thm 3. *If U is a sufficient statistic and G is a strictly monotone function, then $g(U)$ is also sufficient.*

Exercise 1. *Prove the above theorem using the Fisher-Neyman theorem.*

3 Completeness and MVUE

Recall that we aim to estimate an unknown parameter θ that takes values in a given set Θ. The most efficient estimator is known as MVUE.

Def 4. **Minimal variance unbiased estimator (MVUE)** *is an unbiased estimator for θ s.t. it has the smallest variance among all unbiased estimators of θ.*

Q 5. *How to find MVUE?*

For this, we need to introduce the notion of a complete statistic.

Def 5. *A statistic U is* **complete** *if, for any function g, the equality*

$$\mathbb{E}^{\theta} g(U) = 0 \text{ for all values of the unknown parameter } \theta \in \Theta$$

implies $g \equiv 0$.

Completeness means that the family of pdf's (or probability functions) $\{f(y;\theta)\}$ of U, over all θ, is sufficiently large, so that it is impossible to find a function that is orthogonal to $f(\cdot;\theta)$ for every θ.

Rem 1. *If U is complete, then $G(U)$ is also complete, for any strictly monotone function G.*

Once we have found a complete sufficient statistic, we can construct a MVUE from it.

Thm 4. *(Lehmann-Scheffe) If $\hat{\theta}$ is an unbiased estimator of θ and U is a sufficient and complete statistic. Then,*

$$\hat{\theta}^* := \mathbb{E}(\hat{\theta}|U)$$

is MVUE. In particular, if U **is complete and sufficient***, then, any unbiased estimator $G(U)$ is MVUE.*

Q 6. *How to find a complete sufficient statistic?*

Very often, the sufficient statistic obtained via Fischer theorem is also complete. This is because of the following result.

Thm 5. *Assume that the likelihood ratio $\frac{L(y_1,\ldots,y_n;\theta)}{L(x_1,\ldots,x_n;\theta)}$ is well defined for all possible $y_1,\ldots,y_n,x_1,\ldots,x_n$ and $\theta \in \Theta$, and that there exists a minimal sufficient statistic. Assume also that statistic U is such that*

$$\frac{L(y_1,\ldots,y_n;\theta)}{L(x_1,\ldots,x_n;\theta)} \text{ does not depend on } \theta \in \Theta \text{ if and only if } U(y_1,\ldots,y_n) = U(x_1,\ldots,x_n).$$

Then, U is sufficient and complete.

In this course, unless stated otherwise, we always assume that, if you are asked to compute MVUE, a minimal sufficient statistic does exist!

The sufficient statistics obtained from Fisher-Neyman thm are usually complete. Indeed, assume that L satisfies the representation in Fisher-Neyman thm:

$$L(y_1,\ldots,y_n;\theta) = g(U(y_1,\ldots,y_n),\theta) \cdot h(y_1,\ldots,y_n).$$

Then, the likelihood ratio is

$$\frac{L(y_1,\dots,y_n;\theta)}{L(x_1,\dots,x_n;\theta)} = \frac{g(U(y_1,\dots,y_n),\theta)}{g(U(x_1,\dots,x_n),\theta)}\frac{h(y_1,\dots,y_n)}{h(x_1,\dots,x_n)}.$$

Notice that, if $U(y_1,\dots,y_n) = U(x_1,\dots,x_n)$, then the above ratio does not depend on θ. However, it may happen that there exist $z \neq z'$ such that $g(z,\theta) = g(z',\theta)$ for all $\theta \in \Theta$. Then, there are two (or more) different values of U for which the above ratio still does not depend on θ. The above theorem says that if we can exclude such a (degenerate) situation, then U is not only sufficient but also complete.

Ex 3. *Consider a sample $Y_1,\dots,Y_n$ of i.i.d. r.v.'s from $Ber(p)$ distr..*

Q 7. *Find MVUE for p.*

Let us compute the likelihood ratio:

$$L(y_1,\dots,y_n;p) = \mathbb{P}^p(Y_1 = y_1,\dots,Y_n = y_n) = p^{n\bar{y}}(1-p)^{n-n\bar{y}},$$

$$\frac{L(y_1,\dots,y_n;p)}{L(x_1,\dots,x_n;p)} = \frac{p^{n\bar{y}}(1-p)^{n-n\bar{y}}}{p^{n\bar{x}}(1-p)^{n-n\bar{x}}} = (p/(1-p))^{n(\bar{y}-\bar{x})}$$

The above does not depend on p if and only if $\bar{y} = \bar{x}$. Therefore, by Theorem 5, $\bar{Y}$ is a complete sufficient statistic. As $\bar{Y}$ is also an unbiased estimator of p, it is a MVUE.

Ex 4. *Consider a sample $Y_1,\dots,Y_n$ of i.i.d. r.v.'s with the Weibull distribution, that has the pdf*

$$f(y;\theta) = e^{-y^2/\theta}(2y/\theta) \quad y > 0,$$

and zero for $y \leq 0$.

Q 8. *Find MVUE for θ.*

$$L(y_1,\dots,y_n;p) = 2^n\theta^{-n}e^{-\frac{1}{\theta}\sum_{i=1}^n y_i^2}\prod_{i=1}^n y_i, \quad y_1,\dots,y_n > 0,$$

$$\frac{L(y_1,\dots,y_n;p)}{L(x_1,\dots,x_n;p)} = \frac{e^{-\frac{1}{\theta}\sum_{i=1}^n y_i^2}\prod_{i=1}^n y_i}{e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^2}\prod_{i=1}^n x_i} = e^{-\frac{1}{\theta}(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2)}\frac{\prod_{i=1}^n y_i}{\prod_{i=1}^n x_i}$$

The above does not depend on θ if and only if $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n x_i^2$. Therefore, by Theorem 5, $U = \sum_{i=1}^n Y_i^2$ is a complete sufficient statistic.

We need to construct an unbiased estimator of θ in a form if a function of U. It is easy to check that

$$\mathbb{E}Y_i^2 = \theta.$$

Then, we the statistic

$$\frac{1}{n}\sum_{i=1}^n Y_i^2$$

is unbiased and s a function of U. Thus, it is a MVUE for θ.

Exercise 2. *Consider a sample $Y_1, \ldots, Y_n$ of i.i.d. r.v.'s with the uniform distribution $U(0, \theta)$. Show that $\frac{n+1}{n} Y_{(n)}$ is a MVUE for θ.*

The notions of sufficient and complete statistics extend in a straightforward way to vector-valued functions, which is relevant when θ is a vector.

Ex 5. *Consider a sample $Y_1, \ldots, Y_n$ of i.i.d. r.v.'s with distr. $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown.*

Q 9. *Find MVUE for μ and σ^2.*

$$L(y_1, \ldots, y_n; p) = (2\pi)^{-n/2} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2},$$

$$\frac{L(y_1, \ldots, y_n; p)}{L(x_1, \ldots, x_n; p)} = \frac{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2}}{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}} = \frac{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2\mu y_i)}}{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i^2 - 2\mu x_i)}} = e^{-\frac{1}{2\sigma^2} (\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2) + \frac{\mu}{\sigma^2} (\sum_{i=1}^n y_i - \sum_{i=1}^n x_i)}$$

The above does not depend on (μ, σ^2) if and only if $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n x_i^2$ and $\sum_{i=1}^n y_i = \sum_{i=1}^n x_i$. Thus, Thm 5 yields that the pair of statistics $U = (\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i^2)$ is sufficient and complete for (μ, σ^2).

Notice that $\bar{Y}$ is an unbiased estimator of μ and S^2 is an unbiased estimator of σ^2, and both are functions of U. Thus, $(\bar{Y}, S^2)$ is a MVUE of (μ, σ^2).

Sometimes, we can re-use a complete sufficient statistic for θ to find a MVUE for a function of θ (provided this function is strictly monotone).

Ex 6. *Consider a sample $Y_1, \ldots, Y_n$ of i.i.d. r.v.'s with $Exp(1/\theta)$ distribution.*

Q 10. *Find MVUE for θ^2.*

First we find a complete sufficient statistic for θ:

$$L(y_1, \ldots, y_n; p) = \theta^{-n} e^{-\frac{1}{\theta} \sum_{i=1}^n y_i}, \quad y_1, \ldots, y_n > 0,$$

$$\frac{L(y_1, \ldots, y_n; p)}{L(x_1, \ldots, x_n; p)} = e^{-\frac{n}{\theta}(\bar{y} - \bar{x})},$$

The above does not depend on θ if and only if $\bar{y} = \bar{x}$. Thus, $\bar{Y}$ is a complete sufficient statistic for θ.

Since on $\theta \geq 0$, the function $\theta \mapsto \theta^2$ is strictly monotone. Therefore, $\bar{Y}$ is a complete sufficient statistic for θ^2.

Thus, we only need to find an unbiased function of $\bar{Y}$. Since $V(Y_i) = \theta^2$, we compute $\mathbb{E}\bar{Y}^2$ and reverse-engineer a MVUE

$$\frac{n}{n+1} \bar{Y}^2.$$

Indeed, the above is a function of a complete sufficient statistic $\bar{Y}$ and

$$\mathbb{E} \frac{n}{n+1} \bar{Y}^2 = \frac{n}{n+1} \mathbb{E}\bar{Y}^2 = \frac{n}{n+1} (V(\bar{Y}) + (\mathbb{E}\bar{Y})^2) = \frac{n}{n+1} (\theta^2/n + \theta^2) = \theta^2.$$

Sufficient (especially, if also complete) statistic can be used to construct a pivot (if its distr. is known) and, in turn, a confidence interval. Such confidence intervals are typically the shortest at a given confidence level, for a given sample size.

Ex 7. *The lifetime of an electronic component in a navigation system of a missile (measued in 100s of hrs) has the Weibull pdf*

$$f(y;\theta) = e^{-y^2/\theta}(2y/\theta), \quad y > 0.$$

A sample of 10 i.i.d. realizations is taken:

$$.637,\ 1.531,\ .733,\ 2.256,\ 2.364,\ 1.601,\ .152,\ 1.826,\ 1.8681.126$$

Q 11. *Construct a* 0.95*-confidence interval for* θ*.*

Recall that $\sum_{i=1}^n Y_i^2$ *is a complete sufficient statistic for* θ*. Note also that*

$$\frac{2Y_i^2}{\theta} \sim Exp(1/2) = Gamma(1,2) = \chi^2(2).$$

To verify the above, we compute the cdf of $\frac{2Y_i^2}{\theta}$*:*

$$\mathbb{P}\left(\frac{2Y_i^2}{\theta} \leq y\right) = \mathbb{P}(Y_i \leq \sqrt{y\theta/2}) = F_{Y_i}(\sqrt{y\theta/2}),$$

where F_{Y_i} *is the cdf of* Y_i*. Then, the pdf of* $\frac{2Y_i^2}{\theta}$ *is*

$$\frac{d}{dy}F_{Y_i}(\sqrt{y\theta/2}) = \frac{\sqrt{\theta/2}}{2}y^{-1/2}e^{-(y\theta/2)/\theta}2\sqrt{y\theta/2}/\theta = \frac{1}{2}e^{-y/2}, \quad y > 0,$$

which is the pdf of $Exp(1/2)$*.*

Therefore,

$$\sum_{i=1}^{10} \frac{2Y_i^2}{\theta} \sim \chi^2(20)$$

Using the above as a pivot, we obtain

$$0.95 = \mathbb{P}\left(\frac{2\sum_{i=1}^{10} Y_i^2}{34.17} \leq \theta \leq \frac{2\sum_{i=1}^{10} Y_i^2}{9.591}\right),$$

which gives the realized confidence interval

$$[1.442, 5.139]$$