Lecture 21: Asymptotic Evaluations of Point Estimators

Mathematical Statistics II, MATH 60062/70062

Tuesday April 19, 2022

Reference: Casella & Berger, 10.1

Large sample inference

Our discussions so far have been focused on **finite sample inference**. We will now investigate **large sample inference** and discuss three important topics in statistical inference:

- **Point estimation**
  - Consistency, efficiency
  - Asymptotic evaluations of maximum likelihood estimators
- **Hypothesis testing**
  - Wald test, score test, likelihood ratio test (LRT)
  - Asymptotic distributions of the test statistics
- **Confidence intervals**
  - Wald test, score test, LRT

Point estimation

We observe $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Suppose $X_1, \ldots, X_n$ is a random sample from a population $f_X(x \mid \theta)$, where the parameter θ is fixed and unknown.

Let

$$W_n = W_n(\boldsymbol{X}) = W_n(X_1, \ldots, X_n)$$

be a sequence of estimators (which depends on the sample size n).

For example,

$$\begin{aligned}
W_1 &= X_1 \\
W_2 &= \frac{X_1 + X_2}{2} \\
&\vdots \\
W_n &= \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.
\end{aligned}$$

Consistency

A sequence of estimators W_n is **consistent** for a parameter θ if

$$W_n \xrightarrow{p} \theta \text{ for all } \theta \in \Theta.$$

That is, for all $\epsilon > 0$ and for all $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| \geq \epsilon) = 0,$$

or equivalently,

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \epsilon) = 1.$$

Consistency

A sequence of estimators W_n is **consistent** for a parameter θ if

$$W_n \xrightarrow{p} \theta \text{ for all } \theta \in \Theta.$$

That is, for all $\epsilon > 0$ and for all $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| \geq \epsilon) = 0,$$

or equivalently,

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \epsilon) = 1.$$

Suppose W_n is a consistent estimator of θ. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then

$$g(W_n) \xrightarrow{p} g(\theta) \text{ for all } \theta \in \Theta,$$

by the **continuity** property. That is, $g(W_n)$ is consistent for $g(\theta)$.

A sufficient condition for consistency

By Markov's Inequality,

$$P_\theta(|W_n - \theta| \geq \epsilon) \leq \frac{E_\theta[(W_n - \theta)^2]}{\epsilon^2},$$

for all $\epsilon > 0$. Therefore, a sufficient condition for W_n to be consistent is that for all $\theta \in \Theta$,

$$\frac{E_\theta[(W_n - \theta)^2]}{\epsilon^2} \to 0,$$

as $n \to \infty$. Note that

$$E_\theta[(W_n - \theta)^2] = \mathrm{Var}_\theta(W_n) + [E_\theta(W_n) - \theta]^2 = \mathrm{Var}_\theta(W_n) + [\mathrm{Bias}_\theta(W_n)]^2.$$

Therefore, if W_n is a sequence of estimators satisfying

1. $\mathrm{Var}_\theta(W_n) \to 0$ as $n \to \infty$, for all $\theta \in \Theta$
2. $\mathrm{Bias}_\theta(W_n) \to 0$ as $n \to \infty$, for all $\theta \in \Theta$,

then W_n is a consistent estimator of θ.

Sample mean as consistent estimator

Suppose that $X_1, \dots, X_n$ are iid with mean $E_\theta(X_1) = \mu$ and variance $\mathrm{Var}_\theta(X_1) = \sigma^2 < \infty$. Let

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

be the sample mean, where $E_\theta(\bar{X}_n) = \mu$ and $\mathrm{Var}_\theta(\bar{X}_n) = \sigma^2/n$.

As an estimator of μ, we see that

1. $\mathrm{Var}_\theta(\bar{X}_n) \to 0$ as $n \to \infty$, for all $\theta \in \Theta$
2. $\mathrm{Bias}_\theta(\bar{X}_n) \to 0$ as $n \to \infty$, for all $\theta \in \Theta$.

Therefore, $\bar{X}_n$ is a consistent estimator of μ.

Consistency of MLEs

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta$. Let

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta \mid \boldsymbol{x})$$

be the maximum likelihood estimator (MLE) of θ. Under "certain regularity conditions," it follows that

$$\hat{\theta} \xrightarrow{p} \theta \text{ for all } \theta \in \Theta,$$

as $n \to \infty$. That is, *MLEs are consistent estimators*.

Asymptotic normality of MLEs

Suppose $X_1, \dots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta$. Let $\hat{\theta}$ be the MLE of θ. Under "certain regularity conditions," it follows that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)),$$

as $n \to \infty$, where the asymptotic variance

$$v(\theta) = \frac{1}{I_1(\theta)}$$

depends on $I_1(\theta)$, the **Fisher Information** based on one observation, which is given by

$$I_1(\theta) = E_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log f_X(X \mid \theta) \right]^2 \right\} = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_X(X \mid \theta) \right]$$

Regularity conditions

- The regularity conditions presented in the next slide relate to the differentiability of the density function and the ability to interchange integration and differentiation.
- The first four are sufficient for the consistency of MLEs.
- In addition to the four conditions for consistency, the last two are required to prove asymptotic normality of MLEs.
- These conditions are often used in proofs of theorem. They generally hold for Exponential families that are of full rank.
- If the support $\mathcal{X}$ depends on the parameter, then some of the conditions will not hold.

Regularity conditions

1. $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$.
2. The parameter θ is **identifiable**. That is, for $\theta_1, \theta_2 \in \Theta$,
$$f_X(x \mid \theta_1) = f_X(x \mid \theta_2) \implies \theta_1 = \theta_2.$$
3. The family of PDF $\{f_X(x \mid \theta) : \theta \in \Theta\}$ has common support $\mathcal{X}$. In addition, the PDF $f_X(x \mid \theta)$ is differentiable with respect to θ.
4. The parameter space Θ contains an open set where the true value of θ, say θ_0, resides as an interior point.
5. The PDF/PMF is three times differentiable with respect to θ, the third derivative is continuous in θ, and $\int_{\mathbb{R}} f_X(x \mid \theta) dx$ can be differentiated three times under the integral sign.
6. For any $\theta_0 \in \Theta$, there exists a positive number c a function $M(x)$, both of which may depend on θ_0 such that
$$\left| \frac{\partial^3}{\partial \theta^3} \log f_X(x \mid \theta) \right| \leq M(x)$$
for all $x \in \mathcal{X}$ and for all $\theta_0 - c < \theta < \theta_0 + c$ with $E_{\theta_0}[M(X)] < \infty$.

Asymptotic normality of functions of MLEs

Under certain regularity conditions, an MLE $\hat{\theta}$ satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)),$$

where the asymptotic variance

$$v(\theta) = \frac{1}{I_1(\theta)}.$$

By the **Delta Method**, if $g : \mathbb{R} \to \mathbb{R}$ is differentiable at θ and $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} \mathcal{N}\left(0, [g'(\theta)]^2 v(\theta)\right).$$

Functions of MLEs are asymptotically Normal.

A common large sample technique

Suppose a sequence of estimators $\hat{\theta}$ (not necessarily an MLE) satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)).$$

It follows that

$$Z_n = \frac{\hat{\theta} - \theta}{\sqrt{\frac{v(\theta)}{n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Suppose that $v(\hat{\theta})$ is a consistent estimator of $v(\theta)$, that is,

$$v(\hat{\theta}) \xrightarrow{p} v(\theta),$$

for all $\theta \in \Theta$ as $n \to \infty$. Because of the continuity property,

$$\sqrt{\frac{v(\theta)}{v(\hat{\theta})}} \xrightarrow{p} 1.$$

By Slutsky's Theorem,

$$Z_n^* = \frac{\hat{\theta} - \theta}{\sqrt{\frac{v(\hat{\theta})}{n}}} = \frac{\hat{\theta} - \theta}{\sqrt{\frac{v(\theta)}{n}}} \sqrt{\frac{v(\theta)}{v(\hat{\theta})}} \xrightarrow{d} \mathcal{N}(0, 1).$$

MLE of Normal variance

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(0, \theta)$, where $\theta > 0$. The MLE of θ is

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

We know that the MLE is consistent, $\hat{\theta} \xrightarrow{p} \theta$, as $n \to \infty$. We also know that the centered and scaled asymptotic distribution of $\hat{\theta}$ is

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)),$$

where

$$v(\theta) = \frac{1}{I_1(\theta)}.$$

MLE of Normal variance

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(0, \theta)$, where $\theta > 0$. The MLE of θ is

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

We know that the MLE is consistent, $\hat{\theta} \xrightarrow{p} \theta$, as $n \to \infty$. We also know that the centered and scaled asymptotic distribution of $\hat{\theta}$ is

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)),$$

where

$$v(\theta) = \frac{1}{I_1(\theta)}.$$

The PDF of X is

$$f_X(x \mid \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta},$$

and

$$\log f_X(x \mid \theta) = -\frac{1}{2} \log(2\pi\theta) - \frac{x^2}{2\theta}.$$

The derivatives of $\log f_X(x \mid \theta)$ are

$$\frac{\partial}{\partial \theta} \log f_X(x \mid \theta) = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}$$
$$\frac{\partial^2}{\partial \theta^2} \log f_X(x \mid \theta) = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3}.$$

Therefore,

$$I_1(\theta) = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_X(X \mid \theta) \right] = E_\theta \left(\frac{X^2}{\theta^3} - \frac{1}{2\theta^2} \right) = \frac{\theta}{\theta^3} - \frac{1}{2\theta^2} = \frac{1}{2\theta^2}$$

and

$$v(\theta) = \frac{1}{I_1(\theta)} = 2\theta^2.$$

We have

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 2\theta^2) \iff Z_n = \frac{\hat{\theta} - \theta}{\sqrt{\frac{2\theta^2}{n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

By continuity, a consistent estimator of $v(\theta) = 2\theta^2$ is $v(\hat{\theta}) = 2\hat{\theta}^2$. Therefore, by Slutsky's Theorem,

$$Z_n^* = \frac{\hat{\theta} - \theta}{\sqrt{\frac{2\hat{\theta}^2}{n}}} = \frac{\hat{\theta} - \theta}{\sqrt{\frac{2\theta^2}{n}}} \sqrt{\frac{2\theta^2}{2\hat{\theta}^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Asymptotic relative efficiency

Suppose we have two competing sequences of estimators (neither of which is necessarily an MLE sequence) denoted by W_n and V_n that satisfy

$$\sqrt{n}(W_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_W^2)$$
$$\sqrt{n}(V_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_V^2).$$

Both estimators are consistent estimators of θ. The **asymptotic relative efficiency (ARE)** is defined as

$$\mathrm{ARE}(W_n \text{ to } V_n) = \frac{\sigma_W^2}{\sigma_V^2}.$$

- If $\mathrm{ARE} < 1$, then W_n is more efficient than V_n.
- If $\mathrm{ARE} = 1$, then W_n is as efficient as V_n.
- If $\mathrm{ARE} > 1$, then W_n is less efficient than V_n.

Asymptotic evaluation of Beta estimators

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Beta}(\theta, 1)$, where $\theta > 0$.

- The Method Of Moments (MOM) estimator of θ is

$\bar{X} = \frac{\theta}{\theta + 1}$

$$\hat{\theta}_{\mathrm{MOM}} = \frac{\bar{X}}{1 - \bar{X}}$$

  and (by CLT and Delta Method) $\hat{\theta}_{\mathrm{MOM}}$ satisfies

$$\sqrt{n}(\hat{\theta}_{\mathrm{MOM}} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\theta(\theta+1)^2}{\theta+2}\right).$$

- The MLE of θ is

$$\hat{\theta}_{\mathrm{MLE}} = -\frac{n}{\sum_{i=1}^{n} \log X_i}$$

  and $\hat{\theta}_{\mathrm{MLE}}$ satisfies

$$\sqrt{n}(\hat{\theta}_{\mathrm{MLE}} - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

Asymptotic evaluation of Beta estimators

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Beta}(\theta, 1)$, where $\theta > 0$.

- The Method Of Moments (MOM) estimator of θ is

$$\hat{\theta}_{\mathrm{MOM}} = \frac{\bar{X}}{1 - \bar{X}}$$

  and (by CLT and Delta Method) $\hat{\theta}_{\mathrm{MOM}}$ satisfies

$$\sqrt{n}(\hat{\theta}_{\mathrm{MOM}} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\theta(\theta+1)^2}{\theta+2}\right).$$

- The MLE of θ is

$$\hat{\theta}_{\mathrm{MLE}} = -\frac{n}{\sum_{i=1}^{n} \log X_i}$$

  and $\hat{\theta}_{\mathrm{MLE}}$ satisfies

$$\sqrt{n}(\hat{\theta}_{\mathrm{MLE}} - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

What is the $\mathrm{ARE}(\hat{\theta}_{\mathrm{MOM}}$ to $\hat{\theta}_{\mathrm{MLE}})$?