Lecture 21 – Asymptotic properties of estimators

Last Modified: November 10, 2026

Outline:

21.1 Asymptotic Properties of Estimators 1
21.2 Asymptotic Properties of MLE 4

21.1 Asymptotic Properties of Estimators

Consider a statistical model $\{f_\theta(x) \mid \theta \in \Theta\}$ and an estimator T_n based on a sample of size n. We study the properties of T_n as n becomes large.

Definition 1. Let T_n be an estimator for $g(\theta)$. Then T_n is **asymptotically unbiased** if

$$\lim_{n\to\infty} \mathbb{E}[T_n] = g(\theta).$$

Remark 1. Of course if T_n is unbiased, then it is also asymptotically unbiased.

Definition 2. Let T_n be an estimator $g(\theta)$. Then T_n is **consistent** if

$$T_n \xrightarrow{p} g(\theta).$$

Example 1. By the LLN, the sample mean of iid random variables $\{X_i\}$ is consistent for $\theta = \mathbb{E}[X_1]$.

Remark 2.

- Consistent estimators are not unique: If T_n is consistent for $g(\theta)$, then $T_n^* = T_n + \frac{1}{n}$ is also consistent for $g(\theta)$.
- If T_n is consistent for $g(\theta)$ and h is a continuous function, then the continuous mapping theorem implies that
$$h(T_n) \xrightarrow{p} h(g(\theta))$$
and thus $h(T_n)$ is consistent for $h(g(\theta))$.
- Consistency and asymptotic unbiasedness are different properties, and neither implies the other.
- Consistency does not say anything about the rate of convergence, i.e., we do not know how large n needs to be in order for the estimator to be reasonably "close" to the parameter.

Example 2. Let $X_i \overset{\text{iid}}{\sim} \mathsf{Unif}(0, \theta)$. We show that the MLE for θ, given by

$$T_n = X_{(n)}$$

is both asymptotically unbiased and consistent.

For $0 \le t \le \theta$, the distribution function of T_n is

$$F_{T_n}(t) = \mathbb{P}[T_n \le t] = \prod_{i=1}^{n} \mathbb{P}[X_i \le t] = \left(\frac{t}{\theta}\right)^n$$

By differentiation, the pdf is

$$f_{T_n}(t) = \frac{d}{dt} F_{T_n}(t) = \frac{n}{\theta^n} t^{n-1} 1_{\{t \in [0,\theta]\}}$$

Thus

$$\mathbb{E}[T_n] = \int_0^\theta \frac{n}{\theta^n} t^n dt = \frac{n}{\theta^n} \int_0^\theta t^n dt = \frac{n}{\theta^n} \frac{t^{n+1}}{n+1}\Big|_0^\theta = \frac{n}{n+1}\theta \xrightarrow{n\to\infty} \theta$$

and so T_n is asymptotically unbiased.

Consistency follows from the fact that $F_{T_n}(t)$ converges to the step function $F(t)$ which equals zero for $t < \theta$ and one when $t \geq \theta$. This is precisely the cdf of the nonrandom θ. Hence $X_{(n)}$ converges in distribution to θ, and hence also in probability to θ.

Definition 3. Let T_n be an estimator of $g(\theta)$. The estimator T_n is said to be **asymptotically normal** if

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

for some $v > 0$. We call v the **asymptotic variance**.

Let's show asymptotic normality implies consistency, and is therefore stronger than it. Indeed,

$$T_n = g(\theta) + \frac{1}{\sqrt{n}}\big\{\sqrt{n}(T_n - g(\theta))\big\} \xrightarrow{d} g(\theta)$$

by Slutsky's theorem applied to $X_n Y_n$: here, $X_n = 1/\sqrt{n}$ converges to zero in probability and $Y_n = \sqrt{n}(T_n - g(\theta))$ converges to $\mathcal{N}(0, v)$ in distribution.

Remark 3. Asymptotic normality can be described informally by saying that the distribution of T_n is approximately $\mathcal{N}(g(\theta), \frac{v(\theta)}{n})$ for large n. This gives the correct intuition but is not mathematically precise. Also, in contrast to consistency, asymptotic normality *does* give information about a rate of convergence. Continuing with the informal reasoning, it tells us that

$$T_n \approx g(\theta) + \sqrt{\frac{v(\theta)}{n}} Z$$

for $Z \sim \mathcal{N}(0,1)$. Therefore, T_n converges to $g(\theta)$ *at rate* $1/\sqrt{n}$.

Example 3. Let $\{X_i\}$ be iid random variables with $\mu = \mathbb{E}[X_1]$ and $v = \mathrm{Var}(X_1)$. From the CLT, we get

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, v)$$

Thus, the sample mean is an asymptotically normal estimator for μ.

Example 4. We showed above that $X_{(n)} = \max(X_1, \ldots, X_n)$, where $X_i \overset{\text{iid}}{\sim} \mathsf{Unif}(0, \theta)$, is asymptotically unbiased and consistent for θ. But it cannot be asymptotically normal! To see why, note that

$$\mathbb{P}_\theta(\sqrt{n}(X_{(n)} - \theta) \geq 0) = \mathbb{P}_\theta(X_{(n)} \geq \theta) = 0,$$

but asymptotic normality would require that this probability converge to $\mathbb{P}(\mathcal{N}(0,1) \geq 0) = 1/2$.

Remark 4. Using the delta method, we can construct an asymptotically normal estimator for $g(\theta)$ from one for θ:

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} \mathcal{N}(0, v) \quad \Longrightarrow \quad \sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, v[g'(\theta)]^2)$$

Example 5. Suppose $X_i \overset{\text{iid}}{\sim} \mathsf{Bernoulli}(\theta)$. Find an asymptotically normal estimator for

$$g(\theta) = \log \frac{\theta}{1-\theta}$$

The mean and variance are

$$\mathbb{E}[X_1] = \theta, \quad \mathbb{E}[X_1^2] = \theta, \quad \mathsf{Var}(X_1) = \mathbb{E}\big[X_1^2\big] - \mathbb{E}[X_1]^2 = \theta - \theta^2 = \theta(1-\theta)$$

The MLE for θ is given by the sample mean

$$T_n = \bar{X}_n$$

By the Central Limit Theorem

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta(1-\theta))$$

Next, we observe that $g(\theta)$ is continuous with

$$g'(\theta) = \frac{d}{d\theta}[\log\theta - \log(1-\theta)] = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

Using the delta method, we conclude that

$$\sqrt{n}\left(\log\frac{\bar{X}_n}{1-\bar{X}_n} - \log\frac{\theta}{1-\theta}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\theta(1-\theta)}\right)$$

and so $\log\frac{\bar{X}_n}{1-\bar{X}_n}$ is asymptotically normal for $g(\theta)$.

Definition 4. Let S_n and T_n be asymptotically normal estimators for $g(\theta)$, with asymptotic variances $v_S(\theta)$ and $v_T(\theta)$, respectively. We say that T_n is **asymptotically more efficient** than S_n if

$$v_T(\theta) \le v_S(\theta), \quad \text{for all } \theta.$$

What is the "most" asymptotically efficient an estimator can be? Recall the Cramér–Rao lower bound: if $T_n = T(X_1, \ldots, X_n)$ is an *unbiased* estimator of $g(\theta)$, then

$$\mathsf{Var}(T_n) \;\ge\; \frac{g'(\theta)^2}{nI_X(\theta)},$$

where $I_X(\theta)$ is the Fisher information for one observation. We called T_n efficient if its variance was equal to this CR lower bound. There is a corresponding notion of *asymptotic* efficiency.

Definition 5. Let T_n be an estimator of $g(\theta)$. The estimator T_n is *asymptotically efficient* if

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \frac{(g'(\theta))^2}{I_X(\theta)}\right)$$

Remark 5. Thus an asymptotically efficient estimator T_n achieves the Cramér–Rao lower bound in an asymptotic sense. Indeed, using the informal interpretation of above limiting statement, we have $T_n \approx \mathcal{N}(g(\theta), \frac{(g'(\theta))^2}{nI_X(\theta)})$, which informally means $\mathsf{Var}(T_n) \approx \frac{(g'(\theta))^2}{nI_X(\theta)}$.

Remark 6. The Cramér–Rao lower bound pairs naturally with the delta method. If T_n is asymptotically efficient for θ, then

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_X(\theta)}\right).$$

For a differentiable function g with $g'(\theta) \neq 0$, the delta method gives

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

This is exactly what we need to show $g(T_n)$ is asymptotically efficient for $g(\theta)$! In other words, the delta method gives the exact right asymptotic variance.