Inference 2, 2023, lecture 9

Rolf Larsson

November 27, 2023

Today

Chap. 4. Estimation (continued).

- Asymptotic properties
  - Consistency
  - Asymptotic normality

Consistency

Definition (4.12)

We say that a sequence $\{T_n\}$ of estimators for a parameter $\gamma = g(\theta)$ is **weakly consistent**, if T_n converges in probability to γ, that is: If for any $\varepsilon > 0$ and for all $\theta \in \Theta$

$$\lim_{n\to\infty} \mathrm{P}_\theta(|T_n - \gamma| > \varepsilon) = 0.$$

If $\{T_n\}$ converges with probability one (almost surely, a.s.) to γ, that is, for all $\theta \in \Theta$

$$\mathrm{P}_\theta\left(\lim_{n\to\infty} T_n = \gamma\right) = 1,$$

then it is **strongly consistent**.

Strong consistency implies weak consistency.

Write $T_n \xrightarrow{\mathrm{p}} \gamma$ for convergence in probability and $T_n \xrightarrow{\mathrm{a.s.}} \gamma$ for convergence almost surely.

Consistency

Example 1: Weak consistency *does not* imply strong consistency.

- Let $X_1, X_2, \ldots$ be a sequence of independent Bernoulli variables, where $\mathrm{P}(X_k = 1) = 1/k$ for all k.
- X_n converges in probability to 0 as $n \to \infty$, because
  $\mathrm{P}(X_n = 0) = 1 - 1/n \to 1$,
  implying $\lim_{n\to\infty} \mathrm{P}(|X_n - 0| > \varepsilon) = 0$ for any positive $\varepsilon < 1$.
- X_n **does not** converge almost surely to 0, because
  - $X_n \xrightarrow{\text{a.s.}} 0$ is equivalent to $\mathrm{P}(|X_n| > \varepsilon \text{ i.o.}) = 0$ for all $\varepsilon > 0$, where i.o. means infinitely often.
  - But $\sum_n \mathrm{P}(X_n = 1) = \sum_n \frac{1}{n} = \infty$ implies $\mathrm{P}(X_n = 1 \text{ i.o.}) = 1$.
  - Hence, $\mathrm{P}(|X_n| > \varepsilon \text{ i.o.}) = 1$ for any positive $\varepsilon < 1$, and we can not have $X_n \xrightarrow{\text{a.s.}} 0$.

Consistency

Theorem (4.8)

(The Continuous Mapping Theorem.) *Let $\{S_n\}$ be a sequence of random variables, S_0 a random variable and h a continuous function. Then,*

$$S_n \xrightarrow{\text{p}} S_0 \Rightarrow h(S_n) \xrightarrow{\text{p}} h(S_0),$$
$$S_n \xrightarrow{\text{a.s.}} S_0 \Rightarrow h(S_n) \xrightarrow{\text{a.s.}} h(S_0).$$

- Let $X_1, \ldots, X_n$ be independent random variables, distributed as X.
- Let $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.
- Let $\bar{X}_n$ be the mean and S_n^2 be the sample variance of $(X_1, \ldots, X_n)$.

Then,

1. $\bar{X}_n \xrightarrow{\text{p}} \mu$
2. $S_n^2 \xrightarrow{\text{p}} \sigma^2$
3. $S_n \xrightarrow{\text{p}} \sigma$

Asymptotic normality

Convergence in distribution:

$X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$ if $F_n(x) = \mathrm{P}(X_n \leq x) \to \mathrm{P}(X \leq x) = F(x)$
for all points x at which $F_n(x)$ is continuous.

Definition (4.13)

A sequence of estimators $\{T_n\}$ for a m-dimensional parameter $\gamma = g(\theta)$ is **asymptotically normal** if for all $\theta \in \Theta$,

$$\sqrt{n}(T_n - \gamma) \xrightarrow{\mathcal{D}} \mathrm{N}_m\{0, \Sigma(\theta)\},$$

where $\Sigma(\theta)$ is a positive definite $m \times m$ (covariance) matrix.

Asymptotic normality

- An estimator is said to be **asymptotically efficient** if it is asymptotically normal with

$$\Sigma(\theta) = (D_\theta g)(\theta)\{I_{\mathbf{X}}(\theta)\}^{-1}\{(D_\theta g)(\theta)\}^{\mathrm{T}}$$

  for all $\theta \in \Theta$ (cf the Cramér-Rao lower bound).
- Under regularity conditions, the MLE is asymptotically efficient.

Asymptotic normality

Theorem (4.10)

The Delta method (scalar case)

Suppose T_n is an estimator of the form $T_n = h(S_n)$ where the sequence $\{S_n\}$ is asymptotically normal, i.e.

$$\sqrt{n}(S_n - \mu) \xrightarrow{\mathcal{D}} N(0, \tau^2)$$

for some constants μ and $\tau^2 > 0$.

If h has a continuous nonzero derivative h' at μ, then

$$\sqrt{n}\{T_n - h(\mu)\} \xrightarrow{\mathcal{D}} N\left\{0, h'(\mu)^2\tau^2\right\}.$$

Asymptotic normality

Let $X_1, ..., X_n$ be independent random variables, distributed as X which is exponential with intensity β.

1. What is the asymptotic distribution of $\bar{X}$?
2. Derive the asymptotic distribution of $\hat{\beta}_{\mathrm{MLE}}$ via 1. and the delta method.
3. Verify that $\hat{\beta}_{\mathrm{MLE}}$ is asymptotically efficient.

News of today

- Consistency, weak and strong
- The continuous mapping theorem
- Asymptotically normal estimator
- Asymptotically efficient estimator
- Under regularity conditions, the MLE is asymptotically efficient.
- The delta method