Chapter 10: Asymptotic Evaluations
Lecture 22: Consistency

In this chapter, we consider a sample $(X_1, ..., X_n)$ not for fixed n, but as a member of a sequence corresponding to $n = n_0, n_0 + 1, ...$

We first consider limiting behaviors (as $n \to \infty$) of point estimators $T_n = T_n(X_1, ..., X_n)$.

Consistency

A reasonable point estimator is expected to perform better, at least on the average, if more information about the unknown population is available.

With a fixed model assumption and sampling plan, more data (larger sample size n) provide more information about the unknown population.

Thus, it is distasteful to use a point estimator T_n which, if sampling were to continue indefinitely, could possibly have a nonzero estimation error, although the estimation error of T_n for a fixed n may never be 0.

Definition 10.1.1.

A sequence of estimators T_n is a consistent sequence of estimators of $g(\theta)$ (or simply say T_n is a consistent estimator of $g(\theta)$) if, for every $\varepsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta(|T_n - g(\theta)| \geq \varepsilon) = 0.$$

Note that this consistency is related to convergence in probability, and is often called weak consistency.

It is implied by T_n converges almost surely to $g(\theta)$, which is often called strong consistency.

How to find or check consistency?

- Evaluate $P_\theta(|T_n - g(\theta)| \geq \varepsilon)$ directly.
- Use WLLN or SLLN; e.g., the sample moment is consistent for the population moment, as long as the population moment exists.
- Use continuity mapping: if T_n is consistent for $g(\theta)$, then $h(T_n)$ is consistent for $h(g(\theta))$ if h is continuous.

For example, if T_{n1} is consistent for $g_1(\theta)$ and T_{n2} is consistent for $g_2(\theta)$, then $T_{n1}+T_{n2}$ is consistent for $g_1(\theta)+g_2(\theta)$ and $T_{n1}T_{n2}$ is consistent for $g_1(\theta)g_2(\theta)$. (Theorem 10.1.5 is a special case of this.)

The next theorem provides a useful method.

Theorem 10.1.3.

An estimator T_n is consistent if $\lim_{n\to\infty}\text{Bias}_{T_n}(\theta)=\lim_{n\to\infty}E_\theta(T_n)-g(\theta)=0$ and $\lim_{n\to\infty}\text{Var}_\theta(T_n)=0$ for every $\theta\in\Theta$.

Proof.

By Chebychev's inequality, for every $\varepsilon>0$ and every $\theta\in\Theta$,

$$P_\theta(|T_n-g(\theta)|\geq\varepsilon)\leq\varepsilon^{-2}E_\theta[T_n-g(\theta)]^2=\varepsilon^{-2}\{\text{Var}_\theta(T_n)+[\text{Bias}_{T_n}(\theta)]^2\}$$

which converges to 0 as $n\to\infty$ under the given condition.

Example (consistency of the UMVUE).

Suppose that T_n is the UMVUE of $g(\theta)$ for any n based on a random sample $X_1,\ldots,X_n$.

Let $U_n=n^{-1}\sum_{i=1}^n T_1(X_i)$.

Then U_n is unbiased for $g(\theta)$ since $T_1(X_1)$ is unbiased for $g(\theta)$.

Since T_n is the UMVUE, $\mathrm{Var}_\theta(T_n) \leq \mathrm{Var}_\theta(U_n)$ and

$$\lim_{n\to\infty} \mathrm{Var}_\theta(T_n) \leq \lim_{n\to\infty} \mathrm{Var}_\theta(U_n) = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}_\theta(T_1(X_1)) = 0$$

Hence, T_n is consistent by Theorem 10.1.3.

What if T_n is defined only when $n \geq n_0$ for a positive integer n_0?

Example.

Let $X_1, ..., X_n$ be iid from a cdf F satisfying $F(\theta) = 1$ for some $\theta \in \mathscr{R}$ and $F(x) < 1$ for any $x < \theta$ (the *uniform*$(0, \theta)$ is a special case).

To prove the consistency of the largest order statistic $X_{(n)}$ as an estimator of θ, we can evaluate $P_\theta(|X_{(n)} - \theta| \geq \varepsilon)$ directly, for every $\varepsilon > 0$ and $\theta \in \mathscr{R}$,

$$P_\theta(|X_{(n)} - \theta| \geq \varepsilon) = P_\theta(X_{(n)} \leq \theta - \varepsilon) = [P_\theta(X_i \leq \theta - \varepsilon)]^n = [F(\theta - \varepsilon)]^n$$

which converges to 0 since $\theta - \varepsilon < \theta$ and hence $F(\theta - \varepsilon) < 1$.

In fact, $\sum_{n=1}^{\infty} [F(\theta - \varepsilon)]^n < \infty$ and hence $X_{(n)}$ is strongly consistent.

Example.

Let $X_1, ..., X_n$ be iid from a population with unknown mean $\mu \in \mathscr{R}$ and variance $\sigma^2 > 0$, and let $g(\mu) = 0$ if $\mu \neq 0$ and $g(0) = 1$.

Can we find a consistent estimator of $g(\mu)$?

By the SLLN, the sample mean $\bar{X}$ is consistent for μ, but g is not continuous at $\mu = 0$, so $g(\bar{X})$ is usually inconsistent, where

$$g(\bar{X}) = \begin{cases} 1 & \bar{X} = 0 \\ 0 & \bar{X} \neq 0 \end{cases}$$

When $\mu = 0$, $g(\mu) = g(0) = 1$ and, for any $0 < \varepsilon < 1$,

$$\begin{aligned} P(|g(\bar{X}) - 1| \geq \varepsilon) &= P(|g(\bar{X}) - 1| \geq \varepsilon, \bar{X} = 0) + P(|g(\bar{X}) - 1| \geq \varepsilon, \bar{X} \neq 0) \\ &= P(\bar{X} \neq 0) \end{aligned}$$

which usually does not converge to 0 (e.g., $P(\bar{X} \neq 0) = 1$ when X_i has a continuous cdf).

If $\mu \neq 0$, then $g(\mu) = 0$ and

$$P(|g(\bar{X})| \geq \varepsilon) = P(|g(\bar{X})| \geq \varepsilon, \bar{X} = 0) + P(|g(\bar{X})| \geq \varepsilon, \bar{X} \neq 0) = P(\bar{X} = 0)$$

which may converge to 0.

But $g(\bar{X})$ is still inconsistent because, for consistency, the probability has to converge to 0 for every parameter value.

Consider the estimator

$$T_n = \begin{cases} 1 & 0 \le |\bar{X}| < n^{-1/4} \\ 0 & \text{otherwise} \end{cases}$$

To show the consistency of T_n, we only need to show that

$$\lim_{n\to\infty} P(T_n = 1) = \begin{cases} 1 & \text{when } g(\mu) = 1 \text{ (i.e., } \mu = 0) \\ 0 & \text{when } g(\mu) = 0 \text{ (i.e., } \mu \neq 0) \end{cases}$$

If $\mu = 0$, by the CLT, $\sqrt{n}\bar{X}$ converges in distribution to $N(0, \sigma^2)$ and,

$$\lim_{n\to\infty} P(T_n = 1) = \lim_{n\to\infty} P(\sqrt{n}|\bar{X}| < n^{1/4}) = \lim_{n\to\infty} \Phi(n^{1/4}) = 1$$

where Φ is the cdf of $N(0,1)$.

If $\mu \neq 0$, then by the WLLN and continuity map theorem, $|\bar{X}|$ converges in probability to $|\mu| > 0$ and by Slutsky's theorem, $n^{-1/4}/|\bar{X}|$ converges in probability to 0, and

$$\lim_{n\to\infty} P(T_n = 1) = \lim_{n\to\infty} P(1 < n^{-1/4}/|\bar{X}|) = 0.$$

Theorem 10.1.6 (consistency of MLEs)

Let $X_1, ..., X_n$ be iid with pdf or pmf $f_\theta(x)$ and $\widehat{\theta}_n$ be the MLE of θ. Under some conditions (see §10.6.2), $g(\widehat{\theta}_n)$ is consistent for $g(\theta)$ for any continuous function g.

Proof.

We only need to prove that $\widehat{\theta}_n$ converges in probability to θ.

We prove the case where Θ is a compact (bounded close) set in $\mathscr{R}^k$.

The proof for general case is complicated and omitted.

Since Θ is compact, every sub-sequence of $\widehat{\theta}_n, n = 1, 2...$ has a limit in Θ, say ϑ.

The proof is completed if we can show that $\vartheta = \theta$ for any arbitrary sub-sequence of $\widehat{\theta}_n, n = 1, 2....$

The log likelihood function is

$$\log L(\theta|X) = \log\left(\prod_{i=1}^{n} f_\theta(X_i)\right) = \sum_{i=1}^{n} \log f_\theta(X_i)$$

By the SLLN, $n^{-1}\log L(\theta|X)$ converges almost surely to $E_\theta[\log f_\theta(X_i)]$.

Also, $n^{-1}\log L(\vartheta|X)$ converges almost surely to $E_\theta[\log f_\vartheta(X_i)]$ for any $\vartheta \neq \theta$.

One of the conditions in §10.6.2 is that $f_\theta(x)$ is a continuous function of θ for every x.

Hence, $n^{-1}\log L(\theta|X)$ is a continuous function of θ.

Then, for a sub-sequence n_j, $j = 1, 2, \ldots$ with $\widehat{\theta}_{n_j} \to \vartheta$,

$$n^{-1}\log L(\widehat{\theta}_{n_j}|X) - n^{-1}\log L(\vartheta|X) \to 0$$

Since $\widehat{\theta}_{n_j}$ is the MLE,

$$n^{-1}\log L(\widehat{\theta}_{n_j}|X) \geq n^{-1}\log L(\theta|X)$$

By the earlier results, we obtain that

$$E_\theta[\log f_\vartheta(X_i)] \geq E_\theta[\log f_\theta(X_i)]$$

which is the same as

$$\int_{\mathscr{X}} [\log f_\vartheta(x)] f_\theta(x) dx \geq \int_{\mathscr{X}} [\log f_\theta(x)] f_\theta(x) dx$$

or the same as

$$\int_{\mathscr{X}} \left[\log \frac{f_\vartheta(x)}{f_\theta(x)}\right] f_\theta(x)dx \geq 0$$

By Jensen's inequality and the example given in Chapter 3 (lecture 16 of stat 609),

$$\int_{\mathscr{X}} \left[\log \frac{f_\vartheta(x)}{f_\theta(x)}\right] f_\theta(x)dx \leq 0$$

Combing the two inequalities, we obtain that

$$\int_{\mathscr{X}} \left[\log \frac{f_\vartheta(x)}{f_\theta(x)}\right] f_\theta(x)dx = 0$$

i.e., the equality in Jensen's inequality holds.

In Jensen's inequality, since the function $-\log t$ is strictly convex, the equality holds iff $f_\theta(x)/f_\vartheta(x) = c$ is a constant for all $x \in \mathscr{X}$.

Since $f_\theta(x)$ is a pdf, we must have $c = 1$, i.e., $f_\theta(x) = f_\vartheta(x)$, $x \in \mathscr{X}$.

Hence, $\theta = \vartheta$ and the proof is completed.

The technique used in this proof can also be used to prove the consistency of $\widehat{\theta}_n$ obtained under the GMM approach.

Example (consistency of MLE's)

In many cases the consistency of MLE's may be directly checked, especially the required conditions for Theorem 10.1.6 are not satisfied.

We consider the following example.

Let $X_1, ..., X_n$ be a random sample from *uniform*$(\theta, \theta+1)$, $\theta \in \mathscr{R}$.

Note that the cdf F satisfies $F(\theta+1) = 1$ and $F(x) < 1$ if $x < \theta+1$.

From the previous example we conclude that $X_{(n)}$ is strongly consistent for $\theta+1$, or $X_{(n)} - 1$ is strongly consistent for θ.

A similar argument shows that $X_{(1)}$ is strongly consistent for θ.

We have shown previously that any $T(X_1, ..., X_n)$ satisfies

$$X_{(n)} - 1 \leq T(X_1, ..., X_n) \leq X_{(1)}$$

is an MLE of θ.

Then

$$1 = P\left(\lim_{n\to\infty} X_{(n)} - 1 = \theta\right) = P\left(\lim_{n\to\infty} T(X_1, ..., X_n) = \theta\right) = P\left(\lim_{n\to\infty} X_{(1)} = \theta\right)$$

Hence, any MLE is strongly consistent for θ, although MLE's are not unique.

Consistency of the LSE in a linear model

Large sample results in a general linear model, $Y = X\beta + \mathscr{E}$, are useful in cases where $\mathscr{E}$ is not normal or the matrix $V = \mathrm{Var}(\mathscr{E})$ is complex.

We consider $n \to \infty$ and a fixed p (the dimension of β).

The consistency of the LSE $\widehat{\beta}$ can be easily established under very weak conditions: since $\widehat{\beta}$ is unbiased for β, we only need to show that $\mathrm{Var}(l'\widehat{\beta}) \to 0$ as $n \to \infty$ for any fixed $l \in \mathscr{R}^p$, $l \neq 0$.

Assuming that $V = \mathrm{Var}(\mathscr{E})$ exists, for any $l \in \mathscr{R}^p$, we have

$$\mathrm{Var}(l'\widehat{\beta}) = l'(X'X)^{-1}X'VX(X'X)^{-1}l \leq \lambda_+(V)l'(X'X)^{-1}l$$

where $\lambda_+(V)$ is the largest eigenvalue of V.

Hence, for a particular l, $\mathrm{Var}(l'\widehat{\beta}) \to 0$, i.e., $l'\widehat{\beta}$ is consistent for $l'\beta$, if $l'(X'X)^{-1}l \to 0$.

For the consistency of $l'\widehat{\beta}$ for all $l \in \mathscr{R}^p$, which is equivalent to the convergence of $\widehat{\beta}$ in probability to β, a necessary and sufficient condition is the largest eigenvalue of $(X'X)^{-1}$, which is the same as the smallest eigenvalue of $(X'X)^{-1}$, tends to 0 as $n \to \infty$.

Consistency of the error variance estimator in a linear model

Consider the linear model $Y = X\beta + \mathscr{E}$ with $\mathrm{Var}(\mathscr{E}) = \sigma^2 I_n$.

We now show that $\widehat{\sigma}^2 = \mathrm{SSR}/(n-p)$ is a consistent estimator of σ^2. (without normality).

We know that $\widehat{\sigma}^2$ is unbiased; if we show the consistency by proving the variance of $\widehat{\sigma}^2$ tends to 0, we need the finiteness of the 4th moment of ε_i.

Instead, we consider

$$\frac{\mathrm{SSR}}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - x_i'\widehat{\beta})^2 = \frac{1}{n-p}\sum_{i=1}^{n}[\varepsilon_i - x_i'(\widehat{\beta}-\beta)]^2$$

$$= \frac{1}{n-p}\sum_{i=1}^{n}\varepsilon_i^2 + \frac{1}{n-p}\sum_{i=1}^{n}[x_i'(\widehat{\beta}-\beta)]^2 - \frac{2}{n-p}\sum_{i=1}^{n}\varepsilon_i x_i'(\widehat{\beta}-\beta)$$

By the SLLN, the first term converges almost surely to σ^2.

By the Cauchy-Schwartz inequality, the squared last term is bounded by $4\times$ the product of the first and second terms.

Hence, it remains to show that the second term tends to 0.

The result follows from

$$
\begin{aligned}
E\left(\sum_{i=1}^{n}[x_i'(\widehat{\beta}-\beta)]^2\right) &= \sum_{i=1}^{n} E[x_i'(X'X)^{-1}X'\mathscr{E}]^2 \\
&= \sum_{i=1}^{n} E[\mathscr{E}'X(X'X)^{-1}x_i x_i'(X'X)^{-1}X'\mathscr{E}] \\
&= E[\mathscr{E}'X(X'X)^{-1}X'\mathscr{E}] \\
&= \mathrm{trace}[X(X'X)^{-1}X'E(\mathscr{E}\mathscr{E}')] \\
&= \sigma^2 p
\end{aligned}
$$

Note that $\mathrm{Var}(\widehat{\beta}) = \sigma^2(X'X)^{-1}$.

Hence, a consistent estimator of the variance matrix of the LSE is $\widehat{\sigma}^2(X'X)^{-1}$, which will be useful in large sample inference under a linear model without normality assumption.

When estimating a $p \times p$ full rank matrix V that depends on n by $\widehat{V}$, consistency of $\widehat{V}$ means

$$\|\widehat{V}V^{-1} - I_p\|_{\max} \text{ converges in probability to } 0,$$

where $\|A\|_{\max}$ is the maximum of the absolute values of elements in A.

We still consider the linear model, but assume that ε_i's are independent with mean 0 and for each i, $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$.

In this case,

$$\mathrm{Var}(\widehat{\beta}) = \sum_{i=1}^{n} \sigma_i^2 (X'X)^{-1} x_i x_i' (X'X)^{-1}$$

How do we obtain a consistent estimator of this variance matrix?

The unknown quantities are σ_i's.

If we estimate each σ_i^2 by $(Y_i - x_i'\widehat{\beta})^2$, then

$$\widehat{V} = \sum_{i=1}^{n} (Y_i - x_i'\widehat{\beta})^2 (X'X)^{-1} x_i x_i' (X'X)^{-1}$$

is a consistent estimator of $\mathrm{Var}(\widehat{\beta})$ under some weak conditions.

The consistency can be proved similarly, because

$$\begin{aligned} \widehat{V} &= \sum_{i=1}^{n} \varepsilon_i^2 (X'X)^{-1} x_i x_i' (X'X)^{-1} + \sum_{i=1}^{n} [x_i'(\beta - \widehat{\beta})]^2 (X'X)^{-1} x_i x_i' (X'X)^{-1} \\ &\quad + 2 \sum_{i=1}^{n} \varepsilon_i x_i'(\beta - \widehat{\beta}) (X'X)^{-1} x_i x_i' (X'X)^{-1} \end{aligned}$$

Estimation of the distribution function

Let $X_1, ..., X_n$ be iid from a distribution F.

If we have a parametric model $F = F_\theta$, then a consistent estimator of F can be obtained using a consistent estimator of θ.

We now consider the nonparametric case and want to estimate F.

The empirical distribution defined previously is

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$$

By the SLLN, for each fixed x, $\widehat{F}_n(x)$ is a consistent estimator of $F(x)$.

What can we say about $\widehat{F}_n$ as an estimator of the function F?

We use the following metric:

$$\|\widehat{F}_n - F\|_\infty = \sup_x |\widehat{F}_n(x) - F(x)|$$

From Dvoretzky, Kiefer and Wolfowitz (1956) inequality,

$$P\left(\|\widehat{F}_n - F\|_\infty > z\right) \le Ce^{-2nz^2}, \quad z > 0,\ n = 1, 2, ...$$

we have

$$P\left(\lim_{n\to\infty} \|\widehat{F}_n - F\|_\infty = 0\right) = 1.$$