Needs for asymptotic approach

- The distribution of given statistic $T_n(X)$ is needed for inference, but the exact distributions of $T_n(X)$ is not available or too complicated to deal with.
- The limiting distribution is used as an approximation to the distribution of $T_n(X)$ in the situation with a large but actually finite $n$.
- In addition to providing more theoretical results and/or simpler inference procedures, the asymptotic approach requires less stringent mathematical assumptions than does the exact approach.
- A major weakness of the asymptotic approach is that we cannot determine whether a particular $n$ is large enough to safely apply the asymptotic results. Asymptotic results should be used in combination with simulation studies.
Consistency

Definition 2.10 (Consistency of point estimators)
Let \( X = (X_1, \ldots, X_n) \) be a sample from \( P \in \mathcal{P} \) and \( T_n(X) \) be a point estimator of \( \vartheta \) for every \( n \).

(i) \( T_n(X) \) is called consistent for \( \vartheta \) iff \( T_n(X) \rightarrow_p \vartheta \) w.r.t. any \( P \in \mathcal{P} \).

(ii) Let \( \{a_n\} \) be a sequence of positive constants diverging to \( \infty \). \( T_n(X) \) is called \( a_n \)-consistent for \( \vartheta \) iff \( a_n[T_n(X) - \vartheta] = O_p(1) \) w.r.t. any \( P \in \mathcal{P} \), i.e.

\[
\sup_n \mathbb{P}\{|a_n(T_n(X) - \vartheta)| > C\epsilon\} < \epsilon.
\]

(iii) \( T_n(X) \) is called strongly consistent for \( \vartheta \) iff \( T_n(X) \rightarrow_{a.s.} \vartheta \) w.r.t. any \( P \in \mathcal{P} \).

(iv) \( T_n(X) \) is called \( L_r \)-consistent for \( \vartheta \) iff \( T_n(X) \rightarrow_{L_r} \vartheta \) w.r.t. any \( P \in \mathcal{P} \) for some fixed \( r > 0 \).
Remarks

- Consistency is actually a concept relating to a sequence of estimators, \( \{ T_n, n = n_0, n_0 + 1, \ldots \} \), but we usually just say “consistency of \( T_n \)” for simplicity.

- Each of the four types of consistency in Definition 2.10 describes the convergence of \( T_n(X) \) to \( \vartheta \) in some sense, as \( n \to \infty \).

- In statistics, consistency according to Definition 2.10(i), which is sometimes called weak consistency since it is implied by any of the other three types of consistency, is the most useful concept of convergence of \( T_n \) to \( \vartheta \).

- \( L_2 \)-consistency is also called consistency in mse, which is the most useful type of \( L_r \)-consistency.
Example 2.33

Let $X_1, \ldots, X_n$ be i.i.d. from $P \in \mathcal{P}$.
If $\vartheta = \mu$, which is the mean of $P$ and is assumed to be finite, then by the SLLN (Theorem 1.13), the sample mean $\bar{X}$ is strongly consistent for $\mu$ and, therefore, is also consistent for $\mu$.
If we further assume that the variance of $P$ is finite, then $\bar{X}$ is consistent in mse and is $\sqrt{n}$-consistent.

With the finite variance assumption, the sample variance $S^2$ is strongly consistent for the variance of $P$, according to the SLLN.

Consider estimators of $\mu$ of the form $T_n = \sum_{i=1}^{n} c_{ni} X_i$, where $\{c_{ni}\}$ is a double array of constants.
If $P$ has a finite variance, then $T_n$ is consistent in mse iff $\sum_{i=1}^{n} c_{ni} \to 1$ and $\sum_{i=1}^{n} c_{ni}^2 \to 0$.
If we only assume the existence of the mean of $P$, then $T_n$ with $c_{ni} = c_i/n$ satisfying $n^{-1} \sum_{i=1}^{n} c_i \to 1$ and $\sup_i |c_i| < \infty$ is strongly consistent (Theorem 1.13(ii)).
Methods of proving consistency

- One or a combination of the law of large numbers, the CLT, Slutsky’s theorem (Theorem 1.11), and the continuous mapping theorem (Theorems 1.10 and 1.12) are typically applied to establish consistency of point estimators.

- In particular, Theorem 1.10 implies that if $T_n$ is (strongly) consistent for $\vartheta$ and $g$ is a continuous function of $\vartheta$, then $g(T_n)$ is (strongly) consistent for $g(\vartheta)$.

Example 2.33 (continued)

The point estimator $\bar{X}_2$ is strongly consistent for $\mu_2$.

To show that $\bar{X}_2$ is $\sqrt{n}$-consistent under the assumption that $\mathbb{P}$ has a finite variance $\sigma^2$, we can use the identity

$$\sqrt{n}(\bar{X}_2 - \mu_2) = \sqrt{n}(\bar{X} - \mu)(\bar{X} + \mu)$$

and the fact that $\bar{X}$ is $\sqrt{n}$-consistent for $\mu$ and $\bar{X} + \mu = O_p(1)$.

$\bar{X}_2$ may not be consistent in mse since we do not assume that $\mathbb{P}$ has a finite fourth moment.
Methods of proving consistency

- One or a combination of the law of large numbers, the CLT, Slutsky’s theorem (Theorem 1.11), and the continuous mapping theorem (Theorems 1.10 and 1.12) are typically applied to establish consistency of point estimators.

- In particular, Theorem 1.10 implies that if $T_n$ is (strongly) consistent for $\vartheta$ and $g$ is a continuous function of $\vartheta$, then $g(T_n)$ is (strongly) consistent for $g(\vartheta)$.

Example 2.33 (continued)

The point estimator $\bar{X}^2$ is strongly consistent for $\mu^2$. To show that $\bar{X}^2$ is $\sqrt{n}$-consistent under the assumption that $P$ has a finite variance $\sigma^2$, we can use the identity

$$\sqrt{n}(\bar{X}^2 - \mu^2) = \sqrt{n}(\bar{X} - \mu)(\bar{X} + \mu)$$

and the fact that $\bar{X}$ is $\sqrt{n}$-consistent for $\mu$ and $\bar{X} + \mu = O_p(1)$. $\bar{X}^2$ may not be consistent in mse since we do not assume that $P$ has a finite fourth moment.
Example 2.33 (continued)

Alternatively, we can use the fact that
\[ \sqrt{n}(\bar{X}^2 - \mu^2) \to_d N(0, 4\mu^2\sigma^2) \] (by the CLT and Theorem 1.12) to show the \( \sqrt{n} \)-consistency of \( \bar{X}^2 \).

The following example shows another way to establish consistency of some point estimators.
Example 2.33 (continued)

Alternatively, we can use the fact that
\[ \sqrt{n}(\bar{X}^2 - \mu^2) \rightarrow_d N(0, 4\mu^2\sigma^2) \] (by the CLT and Theorem 1.12) to show the \( \sqrt{n} \)-consistency of \( \bar{X}^2 \).

The following example shows another way to establish consistency of some point estimators.

Example 2.34

Let \( X_1, \ldots, X_n \) be i.i.d. from an unknown \( P \) with a continuous c.d.f. \( F \) satisfying \( F(\theta) = 1 \) for some \( \theta \in \mathbb{R} \) and \( F(x) < 1 \) for any \( x < \theta \).

Consider the largest order statistic \( X_{(n)} \) as an estimator of \( \theta \).

For any \( \epsilon > 0 \), \( F(\theta - \epsilon) < 1 \) and

\[ P(|X_{(n)} - \theta| \geq \epsilon) = P(X_{(n)} \leq \theta - \epsilon) = [F(\theta - \epsilon)]^n, \]

which imply (according to Theorem 1.8(v)) \( X_{(n)} \rightarrow_{a.s.} \theta \), i.e., \( X_{(n)} \) is strongly consistent for \( \theta \).
Example 2.34 (continued)

If we assume that \( F^{(i)}(\theta-) \), the \( i \)-th-order left-hand derivative of \( F \) at \( \theta \), exists and vanishes for any \( i \leq m \) and that \( F^{(m+1)}(\theta-) \) exists and is nonzero, where \( m \) is a nonnegative integer, then

\[
1 - F(X_{(n)}) = \frac{(-1)^m F^{(m+1)}(\theta-)}{(m+1)!} (\theta - X_{(n)})^{m+1} + o \left( |\theta - X_{(n)}|^{m+1} \right) \quad \text{a.s.}
\]

This result and the fact that \( P \left( n[1 - F(X_{(n)})] \geq s \right) = (1 - s/n)^n \) imply that \( (\theta - X_{(n)})^{m+1} = O_p(n^{-1}) \), i.e., \( X_{(n)} \) is \( n^{(m+1)^{-1}} \)-consistent.
Example 2.34 (continued)

If we assume that $F^{(i)}(\theta-)$, the $i$th-order left-hand derivative of $F$ at $\theta$, exists and vanishes for any $i \leq m$ and that $F^{(m+1)}(\theta-)$ exists and is nonzero, where $m$ is a nonnegative integer, then

$$1 - F(X_{(n)}) = \frac{(-1)^m F^{(m+1)}(\theta-)}{(m+1)!} (\theta - X_{(n)})^{m+1} + o \left( |\theta - X_{(n)}|^{m+1} \right) \ a.s.$$  

This result and the fact that $P \left( n \left[ 1 - F(X_{(n)}) \right] \geq s \right) = (1 - s/n)^n$ imply that $(\theta - X_{(n)})^{m+1} = O_p(n^{-1})$, i.e., $X_{(n)}$ is $n^{(m+1)^{-1}}$-consistent.

If $m = 0$, then $X_{(n)}$ is $n$-consistent, which is the most common situation. If $m = 1$, then $X_{(n)}$ is $\sqrt{n}$-consistent. The limiting distribution of $n^{(m+1)^{-1}}(X_{(n)} - \theta)$ can be derived as follows. Let

$$h_n(\theta) = \left[ \frac{(-1)^m (m + 1)!}{n F^{(m+1)}(\theta-)} \right]^{(m+1)^{-1}}.$$  

For $t \leq 0$, by Slutsky’s theorem,
Example 2.34 (continued)

\[ \lim_{n \to \infty} P \left( \frac{X(n) - \theta}{h_n(\theta)} \leq t \right) = \lim_{n \to \infty} P \left( \left[ \frac{\theta - X(n)}{h_n(\theta)} \right]^{m+1} \geq (-t)^{m+1} \right) \]

\[ = \lim_{n \to \infty} P \left( n \left[1 - F(X(n))\right] \geq (-t)^{m+1} \right) \]

\[ = \lim_{n \to \infty} \left[ 1 - \frac{(-t)^{m+1}}{n} \right]^n \]

\[ = e^{-(-t)^{m+1}}. \]

Remarks

▶ It can be seen from the previous examples that there are many consistent estimators.

▶ Like the admissibility in statistical decision theory, consistency is a very essential requirement in the sense that any inconsistent estimators should not be used, but a consistent estimator is not necessarily good.

▶ Thus, consistency should be used together with one or a few more criteria.
Importance of consistent estimators

We discuss a situation in which finding a consistent estimator is crucial. Suppose that an estimator $T_n$ of $\vartheta$ satisfies

$$c_n[T_n(X) - \vartheta] \rightarrow_d \sigma Y,$$  \hfill (1)

where $Y$ is a random variable with a known distribution, $\sigma > 0$ is an unknown parameter, and $\{c_n\}$ is a sequence of constants.

For example, in Example 2.33, $\sqrt{n}(\bar{X} - \mu) \rightarrow_d N(0, \sigma^2)$; in Example 2.34, (1) holds with $c_n = n^{(m+1)^{-1}}$ and $\sigma = \frac{(-1)^m(m + 1)!/F^{(m+1)}(\theta -)}{(m+1)^{-1}}$.

If a consistent estimator $\hat{\sigma}_n$ of $\sigma$ can be found, then, by Slutsky’s theorem,

$$c_n[T_n(X) - \vartheta] / \hat{\sigma}_n \rightarrow_d Y$$

and, thus, we may approximate the distribution of $c_n[T_n(X) - \vartheta] / \hat{\sigma}_n$ by the known distribution of $Y$. 