Lecture 16: Convergence

Mathematical Statistics I, MATH 60061/70061

Tuesday November 2, 2021

Reference: Casella & Berger, 5.5.1-5.5.2

- In statistical analysis, a key to the success of finding a good inferential procedure is being able to find some moments and/or distributions of various statistics.
- In many complicated problems, exact distributional results (i.e., "finite sample" results that are applicable for any fixed sample size n) of given statistics may not be available.
- When exact results are not available, we may be able to gain insight by examining the stochastic behavior as the sample size n becomes *infinitely large*. These are called **large sample** or **asymptotic** results.
- The asymptotic approach can also be used to obtain a procedure simpler (e.g., in terms of computation) than that produced by the exact approach.

Example: Exact and large sample results

Suppose $X_1, \ldots, X_n$ are iid $\text{Expo}(\lambda)$, where $\lambda > 0$, and let

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

the sample mean based on n observations. With MGF, we can show that (Lecture 6)

$$\bar{X}_n \sim \text{Gamma}(n, n\lambda).$$

This is an exact result (i.e., it is true for any finite sample size n). Later, we will use the **Central Limit Theorem** to show that

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

as $n \to \infty$. In other words, $\sqrt{n}(\bar{X}_n - \lambda)$ is *approximately Normal* for large n.

Convergence in probability

A sequence of random variables $X_1, X_2, \ldots$, **converges in probability** to a random variable X (written as $X_n \xrightarrow{p} X$) if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0,$$

that is, $P(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$. An equivalent definition is

$$\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1.$$

- For $\epsilon > 0$, quantities $P(|X_n - X| \geq \epsilon)$ and $P(|X_n - X| < \epsilon)$ are real numbers. Therefore, convergence in probability deals with the *non-stochastic* convergence of these sequences of real numbers.
- Informally, $X_n \xrightarrow{p} X$ means the probability of the event *"X_n stays away from X"* gets small as n gets large.
- In many cases, statisticians are concerned with situations where the limiting random variable X is a constant.

Weak Law of Large Numbers (WLLN)

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the sample mean

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

converges in probability to μ (i.e., $\bar{X}_n \xrightarrow{p} \mu$, as $n \to \infty$).

Weak Law of Large Numbers (WLLN)

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the sample mean

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

converges in probability to μ (i.e., $\bar{X}_n \xrightarrow{p} \mu$, as $n \to \infty$).

Proof: Suppose $\epsilon > 0$. By Markov's inequality,

$$\begin{aligned} P(|\bar{X}_n - \mu| \geq \epsilon) &= P((\bar{X}_n - \mu)^2 \geq \epsilon^2) \\ &\leq \frac{E(\bar{X}_n - \mu)^2}{\epsilon^2} \\ &= \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0, \end{aligned}$$

as $n \to \infty$.

Continuity

Suppose $X_n \xrightarrow{p} X$, as $n \to \infty$ and let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.

<u>Proof:</u> Suppose $\epsilon > 0$. By the definition of continuity, there exists $\delta(\epsilon) > 0$ such that

$$|x_n - x| < \delta(\epsilon) \Rightarrow |h(x_n) - h(x)| < \epsilon.$$

Note that $\{x : |x_n - x| < \delta(\epsilon)\} \subseteq \{x : |h(x_n) - h(x)| < \epsilon\}$. Therefore,

$$P(|x_n - x| < \delta(\epsilon)) \leq P(|h(x_n) - h(x)| < \epsilon).$$

Since $X_n \xrightarrow{p} X$, we have $P(|X_n - X| < \delta(\epsilon)) \to 1$, as $n \to \infty$. Based on the above inequality, $P(|h(x_n) - h(x)| < \epsilon) \to 1$ as well. Because $\epsilon > 0$ was arbitrary, we have $h(X_n) \xrightarrow{p} h(X)$.

Consistency of sample variance

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Let

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

denote the sample variance. Then $S_n^2 \xrightarrow{p} \sigma^2$, as $n \to \infty$.

Remark: When the limiting random variable is a constant, convergence in probability is sometimes referred to as **consistency**. We may say "X_n is a consistent estimator of μ" and "S_n^2 is a consistent estimator of σ^2."

Rewrite the sample variance as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n}{n-1}(\bar{X}_n - \mu)^2,$$

and denote

$$U_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2, \quad V_n = \bar{X}_n - \mu, \quad a_n = \frac{n}{n-1}.$$

By the WLLN, $U_n \xrightarrow{p} \sigma^2$, $V_n \xrightarrow{p} 0$ and $a_n \to 1$, as $n \to \infty$. Note that a_n is not a random variable, but we can view that a_n converges in probability to 1.

Consider the function $h(a_n, u_n, v_n) = a_n(u_n - v_n^2)$. By the continuity property, $S_n^2 \xrightarrow{p} h(a_n, U_n, V_n) = \sigma^2$, as $n \to \infty$.

Consistency of sample standard deviation

Consider the function $h(x) = \sqrt{x}$. If S_n^2 is a consistent estimator of σ^2, then the sample standard deviation $S_n = \sqrt{S_n^2} = h(S_n^2)$ is a **consistent estimator** of σ.

Note that S_n is a biased estimator of σ, but the *bias disappears asymptotically*.

Almost sure convergence

A sequence of random variables, $X_1, X_2, \ldots$, **converges almost surely** to a random variable X if, for any $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1.$$

- If a sample space S has elements denoted by s, then $X_n(s)$ and $X(s)$ are all functions defined on S.
- By "almost surely", it means that the functions $X_n(s)$ converge to $X(s)$ for all $s \in S$ except perhaps for $s \in N$, where $N \subset S$ and $P(N) = 0$

Continuity: Suppose X_n converges almost surely to X and let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then $h(X_n)$ converges almost surely to $h(X)$.

Strong Law of Large Numbers (SLLN)

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the sample mean

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

converges almost surely to μ.

- For both the WLLN and SLLN we had the assumption of a finite variance. It is in fact a stronger assumption than is needed.
- The only moment assumption that is needed is that $E|X_i| < \infty$.

Almost sure convergence vs. convergence in probability

Almost sure convergence is a very strong form of convergence (often stronger than is needed). It implies convergence in probability. The converse is not true in general.

Suppose $\hat{\theta}_n$ is a sequence of estimators for an unknown parameter θ. We can think of updating the value of $\hat{\theta}_n$ as data become available and wish that $\hat{\theta}_n$ has the following behavior:

- It becomes "close" to θ when n is sufficiently large.
- It never "stays away" from θ after further data collection.

Almost sure convergence guarantees this. Convergence in probability does not; it guarantees only that the probability that $\hat{\theta}_n$ "stays away" becomes small.

In practice, however, convergence in probability is all we need in most cases.