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.

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.

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 $h(X)$.

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.

When the limiting random variable is a constant, say c, use Markov's inequality; i.e., for $r \geq 1$,

$$P(|X_n - c| \geq \epsilon) \leq \frac{E(|X_n - c|^r)}{\epsilon^r}$$

and show the RHS converges to 0 as $n \to \infty$.

The most common case is $r = 2$, so that

$$\begin{aligned} E(X_n - c)^2 &= \mathrm{Var}(X_n) + (E(X_n) - c)^2 \\ &= \mathrm{Var}(X_n) + (\mathrm{Bias}(X_n))^2. \end{aligned}$$

Therefore, it suffices to show that both $\mathrm{Var}(X_n)$ and $\mathrm{Bias}(X_n)$ converge to 0.