Lecture 01: Convergence

Mathematical Statistics II, MATH 60062/70062

Thursday January 20, 2022

Reference: Casella & Berger, 5.5

Convergence concepts

- 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

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

- 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 $n \to \infty$).

<u>Proof:</u> 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 $n \to \infty$.

Convergence in distribution

A sequence of random variables $X_1, X_2, \ldots$, **converges in distribution** to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

Convergence in distribution

A sequence of random variables $X_1, X_2, \ldots$, **converges in distribution** to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

- We only need to consider the convergence at x that is a continuity point of F_X.
- It is really the CDFs that converge, not the random variables.

Continuity

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

Convergence in probability & convergence in distribution

If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X, the sequence also converges in distribution to X,

$$X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X.$$

The converse is not true in general. It is true when the limiting random variable is a constant.

Convergence in probability & convergence in distribution

If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X, the sequence also converges in distribution to X,

$$X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X.$$

The converse is not true in general. It is true when the limiting random variable is a constant.

Suppose that $X_n \sim \mathcal{N}(0,1)$ for all n and that $X \sim \mathcal{N}(0,1)$. Obviously, $F_{X_n}(x) \to F_X(x)$, for all $x \in \mathbb{R}$. However, this does not guarantee that X_n will be *close to X with high probability*.

E.g., if X_n and X are independent, then $Y = X_n - X$ is a $\mathcal{N}(0,2)$ random variable. For $\epsilon > 0$, $P(|X_n - X| \leq \epsilon) = P(|Y| \leq \epsilon)$ is a constant. This does *not* converge to 1.

Central Limit Theorem

Let $X_1, X_2, \ldots$, be a sequence of iid random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as $n \to \infty$.

Normal approximation to the sample proportion

Suppose $X_1, X_2, \ldots, X_n$ are iid $\mathrm{Bern}(p)$, where $0 < p < 1$. Recall that $E(X_1) = p$ and $\mathrm{Var}(X_1) = p(1-p)$.

For Bernoulli random variables, X_i's are zeros and ones, so $\bar{X}_n$ is a **sample proportion** (i.e., the proportion of ones in the sample).

The Central Limit Theorem says that

$$\sqrt{n}(\bar{X}_n - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)),$$

or

$$\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as $n \to \infty$. This is the foundation for the inference of categorical data.

Slutsky's Theorem

Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, where a is a constant. Then

1. $Y_n X_n \xrightarrow{d} aX$.
2. $X_n + Y_n \xrightarrow{d} X + a$.

Slutsky's Theorem

Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, where a is a constant. Then

1. $Y_n X_n \xrightarrow{d} aX$.
2. $X_n + Y_n \xrightarrow{d} X + a$.

Let $X_1, X_2, \ldots$, be a sequence of iid random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. The CLT says

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $n \to \infty$. In practice, we do not know σ and use the sample standard deviation S to replace σ for inference calculations.

By Slutsky's Theorem, we can show that

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Delta Method

Suppose X_n is a sequence of random variables that satisfy

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

as $n \to \infty$. For a given function g, suppose that $g'(\theta)$ exists and $g'(\theta) \neq 0$. Then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2\sigma^2),$$

as $n \to \infty$.

In other words, the distribution of $g(X_n)$ can be approximated by

$$\mathcal{N}\left(g(\theta), \frac{[g'(\theta)]^2\sigma^2}{n}\right)$$

for large n.

Variance of odds estimator

Suppose $X_1, X_2, \ldots, X_n$ are iid $\mathrm{Bern}(p)$ random variables, where $0 < p < 1$. Using $\frac{\hat{p}}{1-\hat{p}}$ as an estimate of the **odds** $\frac{p}{1-p}$, what is the variance of the estimate?

Variance of odds estimator

Suppose $X_1, X_2, \ldots, X_n$ are iid $\mathrm{Bern}(p)$ random variables, where $0 < p < 1$. Using $\frac{\hat{p}}{1-\hat{p}}$ as an estimate of the **odds** $\frac{p}{1-p}$, what is the variance of the estimate?

Let $\hat{p} = \bar{X}_n$. The CLT gives

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)), \quad \text{when } n \to \infty$$

Take $g(p) = \frac{p}{1-p}$, so $g'(p) = \frac{1}{(1-p)^2}$. The Delta Method says that

$$\begin{aligned} \mathrm{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right) &\approx [g'(p)]^2 \mathrm{Var}(\hat{p}) \\ &= \left[\frac{1}{(1-p)^2}\right]^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}. \end{aligned}$$

Multivariate extensions

All convergence concepts can be extended to handle sequences of random variables.

Central Limit Theorem: Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, is a sequence of iid random vectors (of dimension k) with $E(\boldsymbol{X}_1) = \boldsymbol{\mu}_{k\times 1}$ and $\mathrm{Cov}(\boldsymbol{X}_1) = \boldsymbol{\Sigma}_{k\times k}$. Let $\bar{\boldsymbol{X}}_n = (\bar{X}_{1+}, \bar{X}_{2+}, \ldots, \bar{X}_{k+})'$ denote the vector of sample means. Then $\sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{MVN}_k(\mathbf{0}, \boldsymbol{\Sigma})$.

Multivariate Delta Method: Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, is a sequence of iid random vectors (of dimension k) that satisfy $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{MVN}_k(\mathbf{0}, \boldsymbol{\Sigma})$. For a given function $g : \mathbb{R}^k \to \mathbb{R}$, suppose that g is differentiable at $\boldsymbol{\mu}$ and is not zero. Then

$$\sqrt{n}[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}} \boldsymbol{\Sigma} \frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}'}\right)$$

where

$$\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}} = \left(\frac{\partial g(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial g(\boldsymbol{x})}{\partial x_k}\right)\Big|_{\boldsymbol{x}=\boldsymbol{\mu}}.$$