Convergence
10/36-705 Intermediate Statistics
Lecture Notes 3

Xuezhe Ma

1 Introduction

The most important aspect of probability theory concerns the behavior of sequence of randoms. This part of probability is called **large sample theory**, or **limit theory**, or **asymptotic theory**. The basic question is: *what can we say about the limiting behavior of a sequence of random variables $X_1, X_2, X_3, \ldots$?* In this lecture, we will briefly discuss different types of convergence and introduce two fundamentally important theorems: the **law of large numbers** and the **central limit theorem**.

1.1 Random Samples

Let's first review some definitions of random samples.
Let $X_1, \ldots, X_n \sim P$. A **statistic** is any function $T_n = g(X_1, \ldots, X_n)$.
The **sample mean** is

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

and **sample variance** is

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

Let $\mu = \mathrm{E}[X_i]$ and $\sigma^2 = \mathrm{Var}[X_i]$. Recall that

$$\mathrm{E}[\overline{X}_n] = \mu, \quad \mathrm{Var}[\overline{X}_n] = \frac{\sigma^2}{n}, \quad \mathrm{E}[S_n^2] = \sigma^2.$$

Note that a statistic T_n is unnecessarily to be a single random variable, but to be a sequence of random variables. See the following example,

Example 1. Let $X_{(1)}, \dots, X_{(n)}$ denoted the ordered values:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

Then $T_n = (X_{(1)}, \dots, X_{(n)})$ is called the order statistic.

Lemma 1. *If* $X_1, \dots, X_n \sim N(\mu, \sigma^2)$*, then* $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$.

Proof. We know that the mgf of X_i is $M_{X_i}(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$. We have,

$$\begin{aligned} M_{\overline{X}_n}(t) &= \mathrm{E}[e^{t\overline{X}_n}] = \mathrm{E}[e^{\frac{t}{n}\sum_{i=1}^n X_i}] \\ &= (\mathrm{E}[e^{tX_i/n}])^n = (M_{X_i}(t/n))^n = e^{\mu t + \frac{\frac{\sigma^2}{n} t^2}{2}} \end{aligned}$$

which is the mgf of a random variable $X \sim N(\mu, \sigma^2/n)$. □

2 Convergence

The four main types of convergence are defined as follows.

Definition 1 (almost surely convergence).
X_n **converges almost surely to** X, written $X_n \overset{a.s.}{\to} X$, if for every $\epsilon > 0$,

$$\mathbb{P}\left(\lim_{n\to\infty} |X_n - X| < \epsilon\right) = 1 \tag{1}$$

X_n **converges almost surely to a constant** c, written $X_n \overset{a.s.}{\to} c$, if for every $\epsilon > 0$,

$$\mathbb{P}\left(\lim_{n\to\infty} |X_n - c| < \epsilon\right) = 1 \tag{2}$$

Note that the definition of *almost surely convergence* relies on the definition of *limit of a sequence of events* which is beyond the scope of this course. More information can be found in the course of **36-752 Advanced Probability Theory**. In this course, we are interested in the following three types of convergences.

Definition 2 (in probability convergence).
X_n **converges in probability to** X, written $X_n \overset{P}{\to} X$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0 \tag{3}$$

X_n **converges in probability to a constant** c, written $X_n \overset{P}{\to} c$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - c| > \epsilon\right) = 0. \tag{4}$$

Definition 3 (in quadratic convergence).
X_n **converges in quadratic to** X, written $X_n \overset{qm}{\to} X$, if

$$\lim_{n\to\infty} \mathrm{E}[(X_n - X)^2] = 0 \tag{5}$$

X_n **converges in quadratic to a constant** c, written $X_n \overset{qm}{\to} c$, if

$$\lim_{n\to\infty} \mathrm{E}[(X_n - C)^2] = 0. \tag{6}$$

Definition 4 (in distribution convergence).
X_n **converges in distribution to** X, written $X_n \rightsquigarrow X$, if at all $t \in \mathcal{R}$ where F is continuous,

$$\lim_{n\to\infty} F_n(t) = F(t) \tag{7}$$

where F_n and F are cdf of X_n and X, respectively.
X_n **converges in distribution to a constant** c, written $X_n \rightsquigarrow c$, if for all $t \neq c$,

$$\lim_{n\to\infty} F_n(t) = \delta_c(t) \tag{8}$$

where $\delta_c(t) = 0$ if $t < c$ and $\delta_c(t) = 1$ if $t \geq c$.

The relationships between the types of convergence can be summarized as follows: Make sure you can prove the above implications. In general, none of the reverse

$$\begin{array}{ccccc} & & \text{q.m.} & & \\ & & \downarrow & & \\ \text{a.s.} & \rightarrow & \text{prob} & \rightarrow & \text{distribution} \end{array}$$

implications hold except a special case introducing in the following theorem.

Theorem 2. *If $X_n \rightsquigarrow c$, then $X_n \overset{P}{\to} c$, where c is a constant real number.*

Proof.

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n < c - \epsilon) + \mathbb{P}(X_n > c + \epsilon) \\ &= F_n(c - \epsilon) + 1 - F_n(c + \epsilon) \\ &\overset{n\to\infty}{\to} \delta_c(c - \epsilon) + 1 - \delta_c(c + \epsilon) \\ &= 0 + 1 - 1 = 0 \end{aligned}$$

□

Some convergence properties are preserved under transformations:

Theorem 3. *(a). If* $X_n \xrightarrow{P} X$ *and* $Y_n \xrightarrow{P} Y$*, then* $X_n + Y_n \xrightarrow{P} X + Y$ *.*
(b). If $X_n \xrightarrow{qm} X$ *and* $Y_n \xrightarrow{qm} Y$*, then* $X_n + Y_n \xrightarrow{qm} X + Y$ *.*
(c). If $X_n \xrightarrow{P} X$ *and* $Y_n \xrightarrow{P} Y$*, then* $X_n Y_n \xrightarrow{P} XY$ *.*

In general, $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ does not imply that $X_n + Y_n \rightsquigarrow X + Y$ or $X_n Y_n \rightsquigarrow XY$. But there are cases when it does:

Theorem 4 (Slutzky's Theorem). *If* $X_n \rightsquigarrow X$ *and* $Y_n \rightsquigarrow c$*, then*
(a). $X_n + Y_n \rightsquigarrow X + c$
(b). $X_n Y_n \rightsquigarrow cX$

Theorem 5 (Continuous Mapping Theorem). *Let* g *be a continuous function.*
(a). If $X_n \xrightarrow{P} X$*, then* $g(X_n) \xrightarrow{P} g(X)$*.*
(b). If $X_n \rightsquigarrow X$*, then* $g(X_n) \rightsquigarrow g(X)$*.*

Proof. (a). Since g is a comtinuous function, we have that $\forall \epsilon > 0$, $\exists \delta$ s.t. $\forall |x - y| < \delta$, $|g(x) - g(y)| < \epsilon$.
Then we have $\forall \epsilon > 0$,

$$\mathbb{P}(|g(X_n) - g(X)| < \epsilon) \geq \mathbb{P}(|X_n - X| < \delta).$$

As $X_n \xrightarrow{P} X$, we have

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

So we have $g(X_n) \xrightarrow{P} g(X)$.
(b). From Portmanteau theorem, $X_n \rightsquigarrow X$ is equivalent to for every closed set F,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$$

For closed set F, we denote $S = g^{-1}(F)$ as the pre-image of F under the mapping g. We first prove that S is a closed set when g is continuous.
Assume a sequence of points $x_1, \dots, x_n \in S$ and $x_n \to x$. We have $g(x_1), \dots, g(x_n) \in F$. Since g is continuous, we have $g(x_n) \to g(x)$. As F is closed, $g(x) \in F$. Then $x \in S$. So we have proved that S is closed.
Then,

$$\limsup_{n\to\infty} \mathbb{P}(g(X_n) \in F) = \limsup_{n\to\infty} \mathbb{P}(X_n \in S) \leq \mathbb{P}(X \in S) = \mathbb{P}(g(X) \in F)$$

So we get $g(X_n) \rightsquigarrow g(X)$. $\square$

3 The Law of Large Numbers

The law of large numbers (LLN) says that the mean of a large sample is close to the mean of the distribution, i.e. the distribution of $\overline{X}_n$ becomes more concentrated around its mean as n gets large.

Theorem 6 (**The Weak Law of Large Numbers (WLLN)**). *If $X_i, \ldots, X_n$ are iid, then $\overline{X}_n \xrightarrow{P} \mu$, where $\mu = \mathrm{E}[X_i]$.*

Proof. Here we only provide the proof for an easy case: the variance of X_i exists ($\mathrm{Var}[X_i] < \infty$).
Assume that $\sigma^2 = \mathrm{Var}[X_i] < \infty$. Using Chebyshev's inequality,

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\mathrm{Var}[\overline{X}_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

which tends to 0 as $n \to \infty$. □

We introduce the Strong Law of Large Numbers (SLLN), though it is beyound the scope of this course.

Theorem 7 (**The Strong Law of Large Numbers (WLLN)**). *If $X_i, \ldots, X_n$ are iid, then $\overline{X}_n \xrightarrow{as} \mu$, where $\mu = \mathrm{E}[X_i]$.*

4 The Central Limit Theorem