Lecture 17: Convergence in Distribution

Mathematical Statistics I, MATH 60061/70061

Tuesday November 16, 2021

Reference: Casella & Berger, 5.5.3

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.

Example: Maximum of Uniforms

Let $X_1, X_2, \ldots$ are iid $\mathrm{Unif}(0,1)$ and $X_{(n)} = \max_{i=1,\ldots,n} X_i$. For every $\epsilon > 0$,

$$\begin{aligned}
P(|X_{(n)} - 1| \geq \epsilon) &= P(X_{(n)} \geq 1 + \epsilon) + P(X_{(n)} \leq 1 - \epsilon) \\
&= 0 + P(X_{(n)} \leq 1 - \epsilon) \\
&= P(X_i \leq 1 - \epsilon, i = 1, \ldots, n) \\
&= (1 - \epsilon)^n,
\end{aligned}$$

which goes to 0 as $n \to \infty$. So, $X_{(n)}$ *converges in probability* to 1.

Example: Maximum of Uniforms

Let $X_1, X_2, \ldots$ are iid $\mathrm{Unif}(0,1)$ and $X_{(n)} = \max_{i=1,\ldots,n} X_i$. For every $\epsilon > 0$,

$$
\begin{aligned}
P(|X_{(n)} - 1| \geq \epsilon) &= P(X_{(n)} \geq 1 + \epsilon) + P(X_{(n)} \leq 1 - \epsilon) \\
&= 0 + P(X_{(n)} \leq 1 - \epsilon) \\
&= P(X_i \leq 1 - \epsilon, i = 1, \ldots, n) \\
&= (1 - \epsilon)^n,
\end{aligned}
$$

which goes to 0 as $n \to \infty$. So, $X_{(n)}$ *converges in probability* to 1.

However, if we take $\epsilon = t/n$, we then have

$$P(X_{(n)} \leq 1 - t/n) = (1 - t/n)^n \to e^{-t}$$

$$\Rightarrow P(n(1 - X_{(n)}) \leq t) \to 1 - e^{-t}.$$

This means, the random variable $n(1 - X_{(n)})$ *converges in distribution* to $\mathrm{Expo}(1)$.

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,

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

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,

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

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.

Convergence in probability & convergence in distribution

If X_n converges in distribution to a constant μ, then

$$\lim_{n\to\infty} P(X_n \leq x) = \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x > \mu. \end{cases}$$

Convergence in probability & convergence in distribution

If X_n converges in distribution to a constant μ, then

$$\lim_{n\to\infty} P(X_n \le x) = \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x > \mu. \end{cases}$$

For every $\epsilon > 0$,

$$\begin{aligned} P(|X_n - \mu| \ge \epsilon) &= P(X_n - \mu \ge \epsilon) + P(X_n - \mu \le -\epsilon) \\ &= P(X_n \ge \mu + \epsilon) + P(X_n \le \mu - \epsilon) \\ &= 1 - P(X_n < \mu + \epsilon) + P(X_n \le \mu - \epsilon) \\ &\le 1 - P(X_n < \mu + \epsilon/2) + P(X_n \le \mu - \epsilon) \\ &\to 1 - 1 + 0 = 0. \end{aligned}$$

Thus, X_n converges in probability to μ.

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 $n \to \infty$.

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 $n \to \infty$.

To prove the central limit theorem, we will assume that the MGF of X_i exists. This assumption is not necessary, but it does make the proof easier. A more general proof would involve characteristic functions that we do not cover.

We will show that the MGF of Z_n converges to $M_Z(t) = e^{t^2/2}$, the MGF of $Z \sim \mathcal{N}(0,1)$.

Define

$$Y_i = \frac{X_i - \mu}{\sigma},$$

for $i = 1, 2, \ldots, n$, and let $M_Y(t)$ denote the common MGF of Y (since X_i's are iid, Y_i's are iid). Note that the expected value and variance of Y_i are $E(Y_i) = 0$ and $\mathrm{Var}(Y_i) = 1$.

Express Z_n in terms of Y_i's:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

The MGF of Z_n is given by

$$\begin{aligned} M_{Z_n}(t) = E(e^{tZ_n}) &= E\left(e^{\frac{t}{\sqrt{n}} \sum_{i=1}^{n} Y_i}\right) \\ &= E\left(e^{\frac{t}{\sqrt{n}} Y_1} e^{\frac{t}{\sqrt{n}} Y_2} \ldots e^{\frac{t}{\sqrt{n}} Y_n}\right) \\ &= \left(E\left(e^{\frac{t}{\sqrt{n}} Y_1}\right)\right)^n \\ &= (M_Y(t/\sqrt{n}))^n. \end{aligned}$$

Expanding $M_Y(t/\sqrt{n})$ in a Taylor series around 0, we have

$$M_Y(t/\sqrt{n}) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{\left(\frac{t}{\sqrt{n}} - 0\right)^k}{k!},$$

where

$$M_Y^{(k)}(0) = \frac{d^k}{dt^k} M_Y(t)\Big|_{t=0}.$$

We have

$$\begin{aligned} M_Y^{(0)}(0) &= M_Y(0) = 1 \\ M_Y^{(1)}(0) &= E(Y) = 0 \\ M_Y^{(2)}(0) &= E(Y^2) = 1. \end{aligned}$$

Therefore, the expansion becomes

$$M_Y(t/\sqrt{n}) = 1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y(t/\sqrt{n}),$$

where the remainder term

$$R_Y(t/\sqrt{n}) = \sum_{k=3}^{\infty} M_Y^{(k)}(0) \frac{\left(\frac{t}{\sqrt{n}} - 0\right)^k}{k!}.$$

An application of Taylor's Theorem (see Theorem 5.5.21, CB) shows that, for fixed $t \neq 0$, we have

$$\lim_{n\to\infty} \frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0.$$

Since t is fixed, we also have

$$\lim_{n\to\infty} \frac{R_Y(t/\sqrt{n})}{(1/\sqrt{n})^2} = \lim_{n\to\infty} nR_Y\left(\frac{t}{\sqrt{n}}\right) = 0,$$

The above is also true at $t = 0$ since $R_Y(0/\sqrt{n}) = 0$

Thus, for any fixed t, we can write

$$\begin{aligned}
\lim_{n\to\infty}\left(M_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n &= \lim_{n\to\infty}\left[1+\frac{(t/\sqrt{n})^2}{2!}+R_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \lim_{n\to\infty}\left[1+\frac{1}{n}\left(\frac{t^2}{2}+nR_Y\left(\frac{t}{\sqrt{n}}\right)\right)\right]^n \\
&= e^{t^2/2}.
\end{aligned}$$

Since $e^{t^2/2}$ is the the MGF of the $\mathcal{N}(0,1)$ distribution, the theorem is proved.

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

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