Lecture 20: Multivariate convergence and the Central Limit Theorem

Convergence in distribution for random vectors

Let $Z, Z_1, Z_2, ...$ be random vectors on $\mathscr{R}^k$.
If the cdf of Z is continuous, then we can define convergence in distribution of Z_n to Z by $\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$, for every $x \in \mathscr{R}^k$.
But this is not good enough if F_Z is not always continuous.
We can adopt the following definition.

Definition.

Let $Z, Z_1, Z_2, ...$ be random vectors on $\mathscr{R}^k$. If, for every $c \in \mathscr{R}^k$, $c'Z_n$ converges in distribution to $c'Z$, then we say that Z_n converges in distribution to Z.

- For any constant vector $c \in \mathscr{R}^k$, $c'Z_n$ is a linear combination of components of Z_n.
- Note that in this definition, the convergence of $c'Z_n$ has to be true for every c (not just some c).

- If Z_n converges in distribution to Z, then every component of Z_n converges in distribution to the corresponding component of Z.
- The converse is not true: if every component of Z_n converges in distribution to the corresponding component of Z, Z_n does not necessarily converge in distribution to Z, because each component corresponds to a particular c only. (A counter-example is given below.)

  This is different from convergence in probability and convergence almost surely.
- Theorems 5.5.12 and 5.5.13 can be extended to the case of random vectors.

A counter-example

Let $Z = (X, Y)$ have the joint pdf

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}[1 + g(x,y)], \qquad (x,y) \in \mathscr{R}^2$$

$$g(x,y) = \begin{cases} xy & -1 < x < 1,\ -1 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

In Chapter 4 we showed that $X \sim N(0,1)$ and $Y \sim N(0,1)$, but the joint distribution of Z is not normal.

Let $Z_n = (X_n, Y_n)$, where for each n, the random variables X_n and Y_n are independent, $X_n \sim N(0,1)$, and $Y_n \sim N(0,1)$.

Since X_n's have the same $N(0,1)$ distribution for all n, obviously that X_n "converges" in distribution to $X \sim N(0,1)$; similarly, Y_n converges in distribution to $Y \sim N(0,1)$.

But the joint distribution of $Z_n = (X_n, Y_n)$ for every n is

$$f_{Z_n}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \qquad (x,y) \in \mathscr{R}^2$$

since X_n and Y_n are independent.

Since Z_n's have the same distribution for every n, Z_n converges in distribution to Z_1, which has a different distribution from Z.

Thus, Z_n does not converge in distribution to Z although each component of Z_n converges in distribution to the corresponding component of Z.

The convergence in distribution is equivalent to the convergence in chf (Theorem C7 or M3(ii)) so that the convergence in chf is a tool to study convergence in distribution.

We can also use Theorem 2.3.12 or M3(i) to establish convergence in distribution by showing the convergence in mgf, but we have to know the existence of mgf's.

Example.

Let $X_1,...,X_n$ be iid random variables.
We want to show that there does not exist a sequence of real numbers $\{c_n\}$ such that $\lim_{n\to\infty}\sum_{i=1}^n(X_i - c_i)$ exists almost surely, unless $P(X_1 = c) = 1$ for a constant c.

Suppose that $Y = \lim_{n\to\infty}\sum_{i=1}^n(X_i - c_i)$ exists almost surely.
For any n, the chf of $Y_n = \sum_{i=1}^n(X_i - c_i)$ is

$$\phi_{Y_n}(t) = \prod_{i=1}^n \phi_{X_1}(t)e^{-itc_i} = [\phi_{X_1}(t)]^n e^{-it(c_1+\cdots+c_n)}$$

If Y exists almost surely, then Y_n converges in distribution to Y and

$$\lim_{n\to\infty}\left|[\phi_{X_1}(t)]^n e^{-it(c_1+\cdots+c_n)}\right| = \lim_{n\to\infty}|\phi_{X_1}(t)|^n = |\phi_Y(t)|.$$

However, $\lim_{n\to\infty}|\phi_{X_1}(t)|^n$ is either 0 or 1, depending on whether $|\phi_{X_1}(t)| < 1$ or $= 1$, which means that $|\phi_Y(t)|$ must be either 0 or 1.

Since $|\phi_Y(t)|$ is continuous and $\phi_Y(0) = 1$, $|\phi_Y(t)| = 1$ for all t and hence $|\phi_{X_1}(t)| = 1$ for all t.

This proves that $P(X_1 = c) = 1$ for a constant c.

The Central Limit Theorem (CLT) is one of the most important theorems in probability and statistics.

It derives the limiting distribution of a sequence of normalized random variables/vectors.

Theorem 5.5.15 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be iid random variables with $E(X_1) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$. Then, for any $x \in \mathscr{R}$,

$$\lim_{n\to\infty} P(\sqrt{n}(\bar{X}-\mu)/\sigma \le x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

That is, $\sqrt{n}(\bar{X}-\mu)/\sigma$ converges in distribution to $Z \sim N(0,1)$.

- Normality comes from sums of iid random variables without distributional assumption except the finiteness of the variance.
- The assumption of finite variance is essentially necessary for convergence of a sume of iid random variables to normality.

Proof.

We apply Theorem C7, i.e., we try to show that the chf of

$$Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i-\mu}{\sigma}$$

converges to the chf of $N(0,1)$, which is $e^{-t^2/2}$.

Let ϕ be the chf of $(X_i - \mu)/\sigma$ (not depending on i since X_i's are iid).

From the properties of chf, the chf of Z_n is

$$\phi_{Z_n}(t) = \left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n .$$

It remains to show that $\lim_{n\to\infty} \phi_{Z_n}(t) = e^{-t^2/2}$ for $t \in \mathscr{R}$.

Since $E((X_i-\mu)/\sigma) = 0$ and $\mathrm{Var}((X_i-\mu)/\sigma) = 1$, by Theorem C1 and Taylor's expansion,

$$\phi(s) = 1 - \frac{s^2}{2} + R(s) \qquad \text{as } |s| \to 0$$

where

$$\lim_{|s|\to 0} R(s)/s^2 = 0$$

Then, for any $t \in \mathscr{R}$,

$$\lim_{n\to\infty} nR\left(\frac{t}{\sqrt{n}}\right) = 0$$

and hence

$$\begin{aligned}
\lim_{n\to\infty} \phi_{Z_n}(t) &= \lim_{n\to\infty} \left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \lim_{n\to\infty} \left[1 - \frac{t^2}{2n} + R\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \lim_{n\to\infty} \left[1 - \frac{1}{n}\left\{\frac{t^2}{2} + nR\left(\frac{t}{\sqrt{n}}\right)\right\}\right]^n \\
&= \lim_{n\to\infty} \left[1 - \frac{1}{n}\left(\frac{t^2}{2}\right)\right]^n \\
&= e^{-t^2/2}
\end{aligned}$$

The multivariate CLT

Let $X_1, X_2, \ldots$ be iid random vectors on $\mathscr{R}^k$ with $E(X_1) = \mu$ and finite covariance matrix Σ. Then $\sqrt{n}(\bar{X} - \mu)$ converges in distribution to a random vector $X \sim N(0, \Sigma)$, the k-dimensional normal distribution with mean 0 and covariance matrix Σ.

Proof.

By the definition of convergence in distribution for random vectors, we need to show that for any constant vector $c \in \mathscr{R}^k$, $\sqrt{n}c'(\bar{X} - \mu)$ converges in distribution to $c'X$.

From the result discussed earlier, we know that $c'X \sim N(0, c'\Sigma c)$.

By the CLT,

$$\frac{\sqrt{n}(c'\bar{X} - E(c'\bar{X}))}{\sqrt{\mathrm{Var}(c'X)}} = \frac{\sqrt{n}(c'\bar{X} - c'\mu)}{\sqrt{c'\Sigma c}}$$

converges in distribution to $Z \sim N(0, 1)$.

Note that $\sqrt{c'\Sigma c}Z \sim N(0, c'\Sigma c)$.

The proof is then completed because we can show from the definition that, if Y_n converges in distribution to Y, then aY_n converges in distribution to aY.

Example 5.5.16.

Let $X_1, ..., X_n$ be a random sample from a *negative-binomial*(r, p) distribution.

Recall that

$$\mu = E(X_1) = \frac{r(1-p)}{p}, \qquad \sigma^2 = \mathrm{Var}(X_1) = \frac{r(1-p)}{p^2}$$

Then, the CLT tells us that the sequence of normalized random variables

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\sqrt{n}(\bar{X} - r(1-p)/p)}{\sqrt{r(1-p)/p^2}}, \qquad n = 1, 2, ...$$

converges in distribution to $Z \sim N(0, 1)$.

This provides a way to approximately calculate probabilities that are very difficult to compute exactly.

For example, if $r = 10$, $p = 1/2$ and $n = 30$, using the fact that $\sum_{i=1}^{30} X_i \sim$ *negative-binomial*(nr, p), we can exactly compute

$$P(\bar{X} \le 11) = P\left(\sum_{i=1}^{30} X_i \le 330\right) = \sum_{x=0}^{330} \binom{300+x-1}{x} (0.5)^{300+x} = 0.8916$$

The CLT gives us the approximation

$$\begin{aligned} P(\bar{X} \le 11) &= P\left(\frac{\sqrt{30}(\bar{X}-10)}{\sqrt{20}} \le \frac{\sqrt{30}(11-20)}{\sqrt{20}}\right) \\ &\approx \Phi\left(\frac{\sqrt{30}(11-20)}{\sqrt{20}}\right) = \Phi(1.2247) = 0.8888 \end{aligned}$$

where $\Phi(x)$ is the cdf of $N(0,1)$.

Normal approximation to binomial

Let $Z_n \sim binomial(n,p)$, $n = 1,2,...$

To apply the CLT, note that each Z_n is a sum of n independent Bernoulli random variables $X_1,...,X_n$, i.e.,

$$Z_n = X_1 + \cdots + X_n, \qquad n = 1,2,...$$

$$X_1, X_2, ... \quad \text{are iid with} \quad E(X_i) = p, \ \ \mathrm{Var}(X_i) = p(1-p)$$

Then, $\bar{X} = Z_n/n$ and

$$\frac{Z_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \text{ converges in distribution to } Z \sim N(0,1)$$

Consequently, for any m,

$$P(Z_n \le m) = P\left(\frac{Z_n - np}{\sqrt{np(1-p)}} \le \frac{m - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{m - np}{\sqrt{np(1-p)}}\right)$$

From the previous two examples, to approximate a probability of the form $P(Z_n \le x)$, we always first normalize Z_n to $[Z_n - E(Z_n)]/\sqrt{\mathrm{Var}(Z_n)}$ and then approximate the probability by $\Phi([x - E(Z_n)]/\sqrt{\mathrm{Var}(Z_n)})$.

In fact, a little bit more can be done using the following result.

Pólya's theorem

If a sequence of random vectors Z_n converges in distribution to Z and Z has a continuous cdf F_Z on $\mathscr{R}^k$, then

$$\lim_{n\to\infty} \sup_{x\in\mathscr{R}^k} |F_{Z_n}(x) - F_Z(x)| = 0.$$

- If the limiting cdf F_Z is continuous, then the convergence of F_{Z_n} to F_Z is not only for every x, but also uniformly for all $x \in \mathscr{R}^k$.
- This result implies the following useful result:
  If Z_n converges in distribution to Z with a continuous F_Z and $c_n \in \mathscr{R}^k$ with $c_n \to c$, then $F_{Z_n}(c_n) \to F_Z(c)$.

Asymptotic normality

If Y_n, $n = 1, 2, ...$, is a sequence of random variables and μ_n and $\sigma_n > 0$, $n = 1, 2, ...$, are constants (typically $\mu_n = E(Y_n)$ and $\sigma_n^2 = \mathrm{Var}(Y_n)$ if they are finite) such that $(Y_n - \mu_n)/\sigma_n$ converges in distribution to $N(0,1)$, then for any sequence $\{c_n\}$ of constants,

$$\lim_{n\to\infty} \left| P(Y_n \le c_n) - \Phi\left(\frac{c_n - \mu_n}{\sigma_n}\right) \right| = 0,$$

i.e., $P(Y_n \le c_n)$ can be approximated by $\Phi(\frac{c_n - \mu_n}{\sigma_n})$, regardless of whether c_n, μ_n, or σ_n has a limit.

Since $\Phi(\frac{t - \mu_n}{\sigma_n})$ is the cdf of $N(\mu_n, \sigma_n^2)$, Y_n is said to be **asymptotically distributed** as $N(\mu_n, \sigma_n^2)$ or simply **asymptotically normal**.

Other forms of CLT

There are other forms of CLT for non-iid sequences of random variables/vectors.
We introduce without proof the following three.

Lindeberg's CLT

Let X_n, $n = 1, 2, ...$, be independent random variables with

$$0 < \sigma_n^2 = \mathrm{Var}\left(\sum_{j=1}^n X_j\right) < \infty, \quad n = 1, 2, ...$$

If

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{j=1}^n E\left[(X_j - E(X_j))^2 I_{\{|X_j - E(X_j)| > \varepsilon\sigma_n\}}\right] = 0 \quad \text{for any } \varepsilon > 0,$$

then

$$\frac{1}{\sigma_n} \sum_{j=1}^n (X_j - E(X_j)) \text{ converges in distribution to } N(0, 1).$$

Liapounov's CLT

Let X_n, $n = 1, 2, \ldots$, be independent random variables with

$$\lim_{n\to\infty} \frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{n} E|X_j - EX_j|^{2+\delta} = 0 \quad \text{for some } \delta > 0.$$

Then the same conclusion as in Lindeberg's CLT holds.
The condition in Liapounov's CLT is stronger than the condition in Lindeberg's CLT, Liapounov's condition is easier to verify.

The CLT for a linear combination of iid random variables

Let X_n, $n = 1, 2, \ldots$, be iid random variables with $E(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$. Let $c_n \in \mathscr{R}$, $n = 1, 2, \ldots$, be constants such that

$$\lim_{n\to\infty} \left(\max_{1\le i\le n} c_i^2 \Big/ \sum_{i=1}^{n} c_i^2 \right) = 0.$$

Then

$$\sum_{i=1}^{n} \frac{c_i(X_i - \mu)}{\sigma} \Big/ \sqrt{\sum_{i=1}^{n} c_i^2} \text{ converges in distribution to } N(0,1).$$

Example.

We apply Liapounov's CLT to independent random variables $X_1, X_2, \ldots$ satisfying $P(X_j = \pm j^a) = P(X_j = 0) = 1/3$, where $a > 0$, $j = 1, 2, \ldots$.

Note that $E(X_j) = 0$ and for all j

$$\sigma_n^2 = \mathrm{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \mathrm{Var}(X_j) = \frac{2}{3}\sum_{j=1}^n j^{2a}.$$

We need to show that, for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^n E|X_j - EX_j|^{2+\delta} = 0$$

For any $\delta > 0$,

$$\sum_{j=1}^n E|X_j - E(X_j)|^{2+\delta} = \frac{2}{3}\sum_{j=1}^n j^{(2+\delta)a}.$$

Since

$$\sum_{j=1}^{n-1} j^t \le \sum_{j=1}^{n-1} \int_j^{j+1} x^t dx \le \sum_{j=2}^n j^t$$

and

$$\sum_{j=1}^{n-1} \int_j^{j+1} x^t dx = \int_1^n x^t dx = \frac{n^{t+1} - 1}{t+1},$$

we conclude that

$$\lim_{n\to\infty} \frac{1}{n^{t+1}} \sum_{j=1}^{n} j^t = \frac{1}{t+1}$$

for any $t > 0$.

Then

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{n} E|X_j - EX_j|^{2+\delta} &= \lim_{n\to\infty} \frac{\frac{2}{3}\sum_{j=1}^{n} j^{(2+\delta)a}}{\left(\frac{2}{3}\sum_{j=1}^{n} j^{2a}\right)^{1+\delta/2}} \\
&= \lim_{n\to\infty} \left(\frac{3}{2}\right)^{\delta/2} \frac{(2a+1)^{1+\delta/2}}{(2+\delta)a+1} \frac{n^{(2+\delta)a+1}}{n^{(2a+1)(1+\delta/2)}} \\
&= \left(\frac{3}{2}\right)^{\delta/2} \frac{(2a+1)^{1+\delta/2}}{(2+\delta)a+1} \lim_{n\to\infty} \frac{1}{n^{\delta/2}} \\
&= 0.
\end{aligned}$$