SDS7102: Linear Models and Extensions

Central Limit Theorems

Qiang Sun, Ph.D. <qiang.sun@mbzuai.ac.ae>

These slides are due to Eric Moulines.

August 19, 2025

MBZUAI

Convergence in probability

Definition

Let $\{X_n\}, X$ be random variables. Then $\{X_n\}$ converges in probability to X as $n \to \infty$ $(X_n \to_p X)$ if for each $\epsilon > 0$,

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

Convergence in distribution

Definition

Let $\{X_n\}, X$ be random variables. Then $\{X_n\}$ converges in distribution to X as $n \to \infty \, (X_n \to {}_d X)$ if

$$\lim_{n\to\infty} P\left(X_n \leq x\right) = P(X \leq x) = F(x)$$

for each continuity point of the distribution function $F(x)$.

Proving convergence in distribution

- Recall that a sequence of random variables $\{X_n\}$ converges in distribution to a random variable X if the corresponding sequence of distribution functions $\{F_n(x)\}$ converges to $F(x)$, the distribution function of X, at each continuity point of F.
- It is often difficult to verify this condition directly for a number of reasons. For example, it is often difficult to work with the distribution functions $\{F_n\}$.
- Also, in many cases, the distribution function F_n may not be specified exactly but may belong to a wider class; we may know, for example, the mean and variance corresponding to F_n but little else about F_n. (From a practical point of view, the cases where F_n is not known exactly are most interesting; if F_n is known exactly, there is really no reason to worry about a limiting distribution F unless F_n is difficult to work with computationally.)

Sheffe theorem

- Suppose that X_n has density function f_n (for $n \geq 1$) and X has density function f. Then $f_n(x) \to f(x)$ (for all but a countable number of x) implies that $X_n \to_d X$. Similarly, if X_n has frequency function f_n and X has frequency function f then $f_n(x) \to f(x)$ (for all x) implies that $X_n \to_d X$. (This result is known as Scheffé's Theorem.)
- The converse of this result is not true; in fact, a sequence of discrete random variables can converge in distribution to a continuous variable and a sequence of continuous random variables can converge in distribution to a discrete random variable.

Weak convergence of student distribution

- Suppose that $\{X_n\}$ is a sequence of random variables where X_n has Student's t distribution with n degrees of freedom. The density function of X_n is

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

- Stirling's approximation, which may be stated as

$$\lim_{y \to \infty} \frac{\sqrt{y}\Gamma(y)}{\sqrt{2\pi}\exp(-y)y^y} = 1$$

  allows us to approximate $\Gamma((n+1)/2)$ and $\Gamma(n/2)$ for large n.

- We then get

$$\lim_{n \to \infty} \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} = \frac{1}{\sqrt{2\pi}}$$

Weak convergence of student distribution

- Hence, we get

$$\lim_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2} = \exp\left(-\frac{x^2}{2}\right)$$

  and so

$$\lim_{n\to\infty} f_n(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

  where the limit is a standard Normal density function.

- Thus $X_n \to {}_d Z$ where Z has a standard Normal distribution.

Convergence for Continuous function

Theorem

Let $(X_n)_{n=1}^{\infty}$ *be a sequence of random variables and* X *a random variable.* $X_n \to_d X$ *if and only if for any bounded continuous function* f*,*

$$\lim_{n\to\infty} E(f(X_n)) = E(f(X)).$$

Rather than considering all bounded continuous functions, it suffices to establish that $\lim_{n\to\infty} E(f(X_n)) = E(f(X))$ for any differentiable function with a bounded derivative. More generally, this can be extended to indefinitely differentiable functions with all derivatives bounded.

Proof I: Approximation of indicator function

- The key to the proof directly lies in approximating $P\left[X_n \leq x\right]$ by $E\left[f_\delta^+\left(X_n\right)\right]$ and $E\left[f_\delta^-\left(X_n\right)\right]$ where $f_\delta^+$and f_δ^- are two bounded, continuous functions.
- In particular, we define $f_\delta^+(y) = 1$ for $y \leq x$, $f_\delta^+(y) = 0$ for $y \geq x + \delta$ and $0 \leq f_\delta^+(y) \leq 1$ for $x < y < x + \delta$; we define $f_\delta^-(y) = f_\delta^+(y + \delta)$. If

$$g(y) = I(y \leq x)$$

it is easy to see that

$$f_\delta^-(y) \leq g(y) \leq f_\delta^+(y)$$

Proof II : Key inequalities

- Since $1_{\{y \leq x\}} \leq f_\delta^+(y)$, we get

$$\begin{aligned} P\left[X_n \leq x\right] &\leq E\left[f_\delta^+\left(X_n\right)\right] \\ &\leq E\left[f_\delta^+\left(X_n\right)\right] - E\left[f_\delta^+(X)\right] + E\left[f_\delta^+(X)\right] \\ &\leq \left|E\left[f_\delta^+\left(X_n\right)\right] - E\left[f_\delta^+(X)\right]\right| + P[X \leq x + \delta] \end{aligned}$$

- similarly, since $1_{\{y \leq x\}} \leq f_\delta^-(y)$, we get

$$P\left[X_n \leq x\right] \geq P(X \leq x - \delta) - \left|E\left[f_\delta^-\left(X_n\right)\right] - E\left[f_\delta^-(X)\right]\right|$$

Levy's continuity theorem

Theorem

Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables with corresponding characteristic functions $\varphi_n(t)$. Suppose that $(\varphi_n(t))_{t\geq 0}$ converges pointwise to some function $(\varphi(t))$ for all $t \in \mathbb{R}$. Then, the following statements are equivalent:

1. (X_n) *converges in distribution to some random variable* X.
2. $(\varphi(t))$ *is the characteristic function of some random variable* X.
3. $\varphi(t)$ *is continuous at* $t = 0$.

Central Limit theorems

Theorem (CLT for i.i.d. random variables)

Suppose that $X_1, X_2, \cdots$ are i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$ and define

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

Then $S_n \to_d Z \sim N(0,1)$ as $n \to \infty$.

Approximation of the binomial distribution

- Suppose that X is a Binomial random variable with parameters n and θ; X can be thought of as a sum of n i.i.d. Bernoulli random variables so the distribution of X can be approximated by a Normal distribution if n is sufficiently large.
- More specifically, the distribution of

$$\frac{X - n\theta}{\sqrt{n\theta(1-\theta)}}$$

is approximately standard Normal for large n.

Approximation of the binomial distribution

- We want to evaluate $P[a \leq X \leq b]$ for some integers a and b.
- A naive application of the CLT gives

$$
\begin{aligned}
& P[a \leq X \leq b] \\
& = P\left[\frac{a - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{b - n\theta}{\sqrt{n\theta(1-\theta)}}\right] \\
& \approx \Phi\left(\frac{b - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{a - n\theta}{\sqrt{n\theta(1-\theta)}}\right)
\end{aligned}
$$

Normal approximation of the binomial distribution

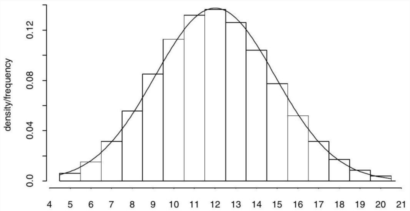

Figure 1: Binomial distribution ($n = 40, \theta = 0.3$) and approximating Normal density

Approximation of the binomial distribution

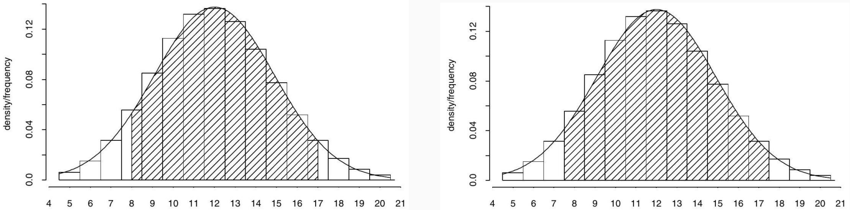

Figure 2: Left panel: Naive Normal approximation of $P(8 \leq X \leq 17)$; Right panel: Normal approximation of $P(8 \leq X \leq 17)$ with continuity correction

Continuity correction

- The distribution of X can be conveniently represented as a probability histogram with the area of each bar representing the probability that X takes a certain value.
- The naive Normal approximation given integrates the approximating Normal density from $a = 8$ to $b = 17$; It seems that the naive Normal approximation will underestimate the true probability.
- A better approximation may be obtained by integrating from $a - 0.5 = 7.5$ to $b + 0.5 = 17.5$. This corrected Normal approximation is

$$\begin{aligned} P[a \leq X \leq b] &= P[a - 0.5 \leq X \leq b + 0.5] \\ &\approx \Phi\left(\frac{b + 0.5 - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{a - 0.5 - n\theta}{\sqrt{n\theta(1-\theta)}}\right) \end{aligned}$$

- The correction used here is known as a continuity correction and can be applied generally to improve the accuracy of the Normal approximation for sums of discrete random variables.

Variance Stabilizing transform for Bernoulli random variables

- Suppose that $X_1, \cdots, X_n$ are i.i.d. Bernoulli random variables with parameter θ. Then

$$\sqrt{n}\left(\bar{X}_n - \theta\right) \to_d Z \sim N(0, \theta(1-\theta))$$

- Find g such that $\sqrt{n}\left(g\left(\bar{X}_n\right) - g(\theta)\right) \to_d N(0,1)$.
- We solve the differential equation

$$g'(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}}$$

- The general form of the solutions to this differential equation is

$$g(\theta) = \sin^{-1}(2\theta - 1) + c$$

where c is an arbitrary constant that could be taken to be 0. (The solutions to the differential equation can also be written $g(\theta) = 2\sin^{-1}(\sqrt{\theta}) + c.\Big)$.

CLT for weighted sums

Theorem

Suppose that $X_1, X_2, \cdots$ *are i.i.d. random variables with* $E(X_i) = 0$ *and* $\operatorname{Var}(X_i) = 1$ *and let* $\{c_i\}$ *be a sequence of constants. Define*

$$S_n = \frac{1}{s_n} \sum_{i=1}^{n} c_i X_i \quad \text{where} \quad s_n^2 = \sum_{i=1}^{n} c_i^2$$

Then $S_n \to_d Z$*, a standard Normal random variable, provided that*

$$\max_{1 \le i \le n} \frac{c_i^2}{s_n^2} \to 0$$

as $n \to \infty$*.*

Lyapunov CLT

Theorem

Suppose that $X_1, X_2, \cdots$ *are independent random variables with* $E(X_i) = 0, E\left(X_i^2\right) = \sigma_i^2$ *and* $E\left(|X_i|^3\right) = \gamma_i$ *and define*

$$S_n = \frac{1}{s_n} \sum_{i=1}^{n} X_i$$

where $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$*. If*

$$\lim_{n \to \infty} \frac{1}{s_n^{3/2}} \sum_{i=1}^{n} \gamma_i = 0$$

then $S_n \to_d Z$*, a standard Normal random variable.*

Cramér-Wold device

Theorem (Cramér-Wold device)

Suppose that $\{\boldsymbol{X}_n\}$ and $\boldsymbol{X}$ are random vectors. Then $\boldsymbol{X}_n \to_d \boldsymbol{X}$ if, and only if,

$$\boldsymbol{t}^T \boldsymbol{X}_n \to {}_d \boldsymbol{t}^T \boldsymbol{X}$$

for all vectors $\boldsymbol{t}$.

Multivariate CLT

Theorem

Suppose that $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \cdots$ *are i.i.d. random vectors with mean vector* $\boldsymbol{\mu}$ *and variancecovariance matrix* C *and define*

$$\boldsymbol{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu}) = \sqrt{n} \left(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu} \right).$$

Then $\boldsymbol{S}_n \to_d \boldsymbol{Z}$ *where* $\boldsymbol{Z}$ *has a multivariate Normal distribution with mean* $\mathbf{0}$ *and variance-covariance matrix* C*.*

Convergence in probability of random vectors

Definition

We will say that $\boldsymbol{X}_n \to_p \boldsymbol{X}$ if each coordinate of $\boldsymbol{X}_n$ converges in probability to the corresponding coordinate of $\boldsymbol{X}$. Equivalently, we can say that $\boldsymbol{X}_n \to_p \boldsymbol{X}$ if

$$\lim_{n\to\infty} P\left[\|\boldsymbol{X}_n - \boldsymbol{X}\| > \epsilon\right] = 0$$

where $\|\cdot\|$ is the Euclidean norm of a vector.