Lecture 3. Limit Theorems

Useful Inequalities

Theorem 1 (Markov's inequality).

Let X be any nonnegative random variable such that $\mathbb{E}[X]$ exists. Then for any $t > 0$, we have

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}[X]}{t}$$

$\uparrow t$, small prob of getting there

$$EX = \int_0^\infty x f_X(x)\,dx = \int_0^t x f_X(x)\,dx + \int_t^\infty x f_X(x)\,dx$$

$$\geq \int_t^\infty x f_X(x)\,dx \geq t \int_t^\infty f_X(x)\,dx$$

$$= t\,P\{X \geq t\}$$

|| Prob of tail

Useful Inequalities

$EX = \mu$

Theorem 2 (Chebyshev's inequality).

For any random variable X with mean μ and finite variance and for any $t > 0$, we have

$$\mathbb{P}\{|X - \mu| \geq t\} \leq \frac{\mathbb{V}ar(X)}{t^2}$$

bounded by the spread

var small, prob of getting in tail v. small

$$P\{|X-\mu| \geq t\} = P\{(X-\mu)^2 \geq t^2\} \leq \frac{E[(X-\mu)^2]}{t^2}$$

① Convergence in probability and Law of Large Numbers

$x_i(w)$ $w \in \Omega$

equipped w prob. fn.

Definition 1.

Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables. We say that $\{X_n\}_{n=1}^{\infty}$ *converges* to X *in probability* (write $X_n \to^p X$) if for any $\varepsilon > 0$ as $n \to \infty$

For any ε, the seq. of prob. converges.

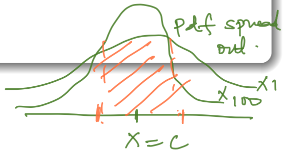
$$\underbrace{\mathbb{P}\{\,|X_n - X| > \varepsilon\}}_{a_n \to 0} \to 0$$

Theorem 3.

If $\mathbb{E}(X_n - X)^2 \to 0$, *then* $X_n \to^p X$.

$P\{|X_n - X| > \varepsilon\} = P\{(X_n - X)^2 > \varepsilon^2\}$

$\leq \dfrac{E(X_n - X)^2}{\varepsilon^2} \to 0$



more mass in ε area

The bigger the no, the more concentrated at center

Convergence in probability and Law of Large Numbers

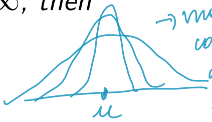
Theorem 4 (Law of Large Numbers).

If $\{X_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}ar(X_i) = \sigma^2 < \infty$, then

every x_i has same mean μ & σ^2

enough to assume just mean

$$\overline{X}_n := \sum_{i=1}^{n} X_i / n \to^p \mu$$



$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$ ← random.

$E\,\bar{X}_n = E\,\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} E X_i = \mu$

$E(\bar{X}_n - \mu)^2 = Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} Var\left(\sum x_i\right) = \frac{1}{n^2}\sum Var(x_i) = \frac{n\sigma^2}{n^2} = \sigma^2/n$

$P\{|\bar{X}_n - \mu| > \varepsilon\} \le \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$

$x_i = \begin{cases} 0 & 1/2 \\ 1 & 1/2 \end{cases}$

$(0, 0, 0, \dots 0) \to 0$ avg

$(0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0) \downarrow$ avg $3/10$

Prob. of happening is same

→ Extremely many outcomes get you to centre

→ LL nos: how likely avg realises.

HW: Try to prove.

☑ Check Yourself: you should be able to prove the following theorem

Theorem 5 (Law of Large Numbers).

Assume that $\{X_i\}_{i=1}^{\infty}$ are independent random variables with $\mathbb{V}ar(X_i) = \sigma_i^2$. Show that if $\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \to 0$ then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \to^p 0.$$

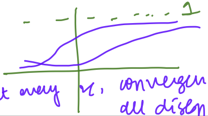
$X_n \longrightarrow X$ $X_n \Rightarrow X$

cdf : monotone function 0 & 1

Definition 2.

We say that $\{X_n\}_{n=1}^{\infty}$ *converges* to X *in distribution* or *weakly* (write $X_n \Rightarrow X$) if for all $x \in \mathbb{R}$ where $F_X(x)$ is continuous

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



at every x, convergence of all distn

Theorem 6.

If $X_n \to^p X$, then $X_n \Rightarrow X$. || in lecture notes

- ☐ If c is some constant and $X_n \Rightarrow c$, then $X_n \to_p c$

Central Limit Theorem : may theorem

Theorem 7 (Central limit theorem).

Let $\{X_i\}$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then

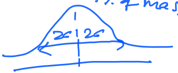
$$\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)} = \underbrace{\sqrt{n}(\overline{X}_n-\mu)} \Rightarrow N(0,\sigma^2)$$

$$\underbrace{\sqrt{n}\Big(\underset{\text{avg}}{\frac{1}{n}\sum_{i=1}^{n} x_i} - \mu\Big)}$$

: how close, how fast converge: | how far from the mean the avg is

multiply by big $\sqrt{n}$

$\Rightarrow \underbrace{N(0,\sigma^2)}_{\text{gaussian}}$



$\frac{1}{n}\sum x_i - \mu \simeq \frac{2\sigma}{\sqrt{n}}$ w prob 95%

$$\xi = \begin{cases} 0 & 1-p \to \text{no cancer} \\ 1 & p \to \text{cancer} \end{cases}$$

looking at each county

$\frac{1}{n}\sum_{i=1}^{n} \xi_i$ on avg how many cancer in county.

small population — have higher deviation from avg

$\frac{1}{n}\sum \xi_i - p \leftarrow$ dispersed as $\frac{\sigma}{\sqrt{n}}$

n is smaller for less populated $\Rightarrow$ more dispersed

Also CLT:

avg : $\bar{X} = \frac{1}{n} \sum_{i}^{n} x_i$

R.V

$X_i, \dots X_n$ iid

$E X_i = \mu$

$Var(X_i) = \sigma^2$

it has mean $E\bar{X} = \mu$

$Var(\bar{x}) = \sigma^2/n$

"kinda say" — $\bar{X}_n \overset{d}{\approx} N(\mu, \sigma^2/n)$ approx. dist.

avg is spread out in large samples.

but actually

To normalize $\frac{\bar{x}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} \overset{d}{\approx} N(0,1)$

stabilizes & becomes gaussian.

Central Limit Theorem

Theorem 8 (Linderberg- Feller's CLT).

Let $\{X_i\}_{i=1}^{\infty}$ be independent random variables with $\mathbb{E}X_i = \mu_i$ and $\mathbb{V}ar(X_i) = \sigma_i^2$. Denote $c_n^2 = \mathbb{V}ar(\sum_{i=1}^n X_i) = \sum_{i=1}^n \sigma_i^2$. If for any $\varepsilon > 0$

↑ sum of var

asy. negligibility condn *

↓ have xi overall kinda same

$$\lim_{n\to\infty} \frac{1}{c_n^2} \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu_i)^2 \mathbb{I}\{|X_i - \mu_i| > \varepsilon c_n\}\right] = 0$$

don't want to have outsized.

↙ centered & normalised

then

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{c_n} \Rightarrow N(0,1)$$

"asymptotic negligibility" implies $\max_{1\le i\le n} \frac{\sigma_i^2}{c_n^2} \to 0$

Lyapunov's condition (for $\delta > 0$):

$\lim_{n\to\infty} \frac{1}{c_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}|X_i - \mu_i|^{2+\delta} = 0$

How to use CLT.

1) Slutsky theorem and Continuous mapping theorem

Theorem 9.

Let $X, X_1, ..., X_n, ...$ and $Y, Y_1, ..., Y_n, ...$ be some random variables. Let g be some continuous function. Let c be some constant. Then

- ☐ *If $X_n \to^p X$ and $Y_n \to^p Y$, then $X_n + Y_n \to^p X + Y$ and $X_n Y_n \to^p XY$.* (can − & ÷ too)
- ☐ *If $X_n \Rightarrow X$ and $Y \to^p c$, then $X_n + Y_n \Rightarrow X + c$ and $X_n Y_n \Rightarrow cX$.* have weak convergence
- ☐ *If $X_n \to^p X$, then $g(X_n) \to^p g(X)$.*
- ☐ *If $X_n \Rightarrow X$, then $g(X_n) \Rightarrow g(X)$*

slutsky

*cannot discuss joint by just discussing marginals

$X_n \Rightarrow X$ (weakly converge) $Y_n \Rightarrow Y$ (marginal con.) then In general $X_n + Y_n \not\Rightarrow X + Y$

if $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix}$ then $X_n + Y_n \Rightarrow X + Y$

Joint distbn

Joint convergence

Empty slide

$\square$

Delta method

second way to use CLT

★ will always use this

Tay Exp → CLT

Theorem 10.

Assume that for a sequence of random variables X_n and constants μ and σ we have $\sqrt{n}(X_n - \mu) \Rightarrow N(0, \sigma^2)$. If $g(\cdot)$ is twice continuously differentiable and $g'(\mu) \neq 0$, then

$$\sqrt{n}(g(X_n) - g(\mu)) \Rightarrow N(0, \sigma^2 (g'(\mu))^2)$$

ass. $\begin{cases} X_n \xrightarrow{P} \mu \\ \sqrt{n}(X_n - \mu) \Rightarrow N(0, \sigma^2) \end{cases}$

g - smooth

then $g(X_n) \xrightarrow{P} g(\mu)$

$\underbrace{\sqrt{n}\,[g(X_n) - g(\mu)]}$ // Taylor

Taylor Expansion use always

$u_n \xrightarrow{P} \mu$

multiply

$= \sqrt{n}\,[g'(u_n^*)(X_n - \mu)]$

derivative at point u_n^* is betw x_n & μ

$= g'(u_n^*) \cdot \sqrt{n}(X_n - \mu)$

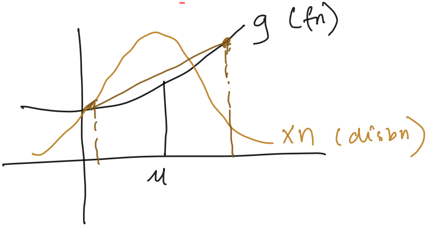
$\downarrow P$

$= g'(\mu)$

Empty slide

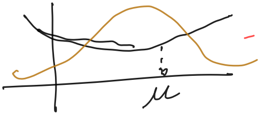
Slutsky $\Rightarrow g'(u) \cdot N(0, 6^2)$ ——— ⊛

when delta method works or not?



- Taylor approx:
  fn in this are mostly linear
  apply linear transf.
  & is still Gaussian.

- Delta not works when
  g not well approx.
  by linear fn.
  when non. linear fn eg



- derivative is 0.

Delta method

* today

Theorem 11.

Let $X_1, \ldots, X_n, \ldots$ be a sequence of iid $k \times 1$ random vectors such that $\sqrt{n}(X_n - \mu) \Rightarrow N(0, \Sigma)$. Let $g : \mathbb{R}^k \to \mathbb{R}$ be a twice continuously differentiable function. Let $\tau^2 = (\partial g(\mu)/\partial\mu)^T \Sigma (\partial g(\mu)/\partial\mu)$. Here $\partial g(\mu)/\partial\mu$ is a $k \times 1$ vector with i-th component equals $\partial g(\mu)/\partial\mu_i$. Then

gradient of partial derivatives

$$\sqrt{n}(g(X_n) - g(\mu)) \Rightarrow N(0, \tau^2)$$

con. T, $1 \times k$, $k \times k$, $k \times 1$

X_n $k \cdot 1$ $\quad \mu$ $k \cdot 1$ $= \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$

$\sqrt{n}(X_n - \mu) \Rightarrow N(0, \Sigma)$ $k \times k$

Empty slide

$\square$