Theorem 5.2 (Continuity Theorem, TBp. 181)

Let $F_n(x)$ be a sequence of cdfs with the corresponding mgfs $M_n(t)$. Let $F(x)$ be a cdf with the mgf $M(t)$. If $M_n(t) \to M(t)$ as $n \to \infty$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity point of F. → i.e., converge in distribution.

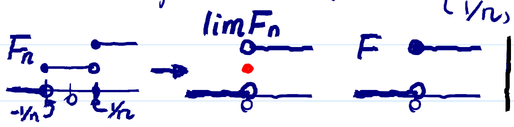
Notes.

1. The reverse of the continuity theorem also holds.
2. The continuity theorem still holds when the moment generating function is replaced by characteristics function (chf always exists).

The reason why $F_n \to F$ only at continuous points of F

counter example (discrete): $X_n = \begin{cases} -1/n, & p=1/2 \\ 1/n, & p=1/2 \end{cases}$ | p_n ½ ½ $-1/n$ $1/n$

Note. $\lim F_n$ is not a cdf



$X_n \xrightarrow{d} 0$ | 1

(for your information)

F_n, F : cdf; f_n, f : pdf; p_n, p : pmf;

But, $p_n(x) \to 0$, $\forall x \in \mathbb{R}$ ← not a pmf

Q: $F_n \xrightarrow{d} F$ implies $\lim_{n\to\infty} f_n(x) = f(x)$?

or $\lim_{n\to\infty} p_n(x) = p(x)$?

(continuous, Ec) $F_n = x - \frac{\sin(2n\pi x)}{2n\pi}$, $0<x<1$; $F_n \xrightarrow{d} U(0,1)$

But $f_n = 1 - \cos(2n\pi x)$ have no limit

Ans: In general, **NO.**

$f_n \to f$ ✗ → converge in dist. ← ✓ $M_n(t) \to M(t)$, $t \in (-a,a)$

$p_n \to p$ ✗ → converge in dist. ← ✓ $F_n(x) \to F(x)$ at cont. pts

Example 5.2 (Convergence of Poisson to Normal, TBp. 181-182)

Let $X_n \sim P(\lambda_n)$, $n = 1, 2, \ldots$ with $\lambda_n \to \infty$. We know that $E(X_n) = Var(X_n) = \lambda_n$ and $M_{X_n}(t) = e^{\lambda_n(e^t - 1)}$. Let

$$\left.\begin{aligned} E(Z_n)&=0 \\ Var(Z_n)&=1 \end{aligned}\right\} \leftarrow Z_n \overset{\text{Standardization}}{=} (X_n - \lambda_n)/\sqrt{\lambda_n}, = \frac{1}{\sqrt{\lambda_n}} X_n + (-\sqrt{\lambda_n})$$

Then $M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}} M_{X_n}\left(\frac{t}{\sqrt{\lambda_n}}\right) = e^{-t\sqrt{\lambda_n}} e^{\lambda_n(e^{t/\sqrt{\lambda_n}} - 1)}$. Because

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \lim_{n\to\infty} -t\sqrt{\lambda_n} + \lambda_n(e^{t/\sqrt{\lambda_n}} - 1) = \frac{t^2}{2},$$

Note: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

$$e^{t/\sqrt{\lambda_n}} = 1 + \frac{t}{\sqrt{\lambda_n}} + \frac{t^2}{2\lambda_n} + \frac{t^3}{6\lambda_n\sqrt{\lambda_n}} + \cdots$$

$M_{Z_n}(t) \to e^{t^2/2}$, which is the mgf of $N(0,1)$. By continuity theorem, $Z_n \xrightarrow{d} N(0,1)$, i.e., when λ is large, we can approximate the distribution of $P(\lambda)$ by $N(\lambda, \lambda)$.

i.e., $Z_n \xrightarrow{d} Z$, where $Z \sim N(0,1)$

compare the shape of their pmf & pdf (LNp.68 & 76)

- **LLN and CLT** ← limit theorems for sum ($\sum_{i=1}^{n} X_i$) or average ($\bar{X}_n$) of r.v.'s X_i's (data)

Theorem 5.3 (Weak Law of Large Numbers (WLLN), TBp. 178)

Data

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. ← not necessary to have identical distribution

cf.

Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then $\overline{X}_n \xrightarrow{P} \mu$. ← cf. "long-run average" in the explanation of mean (LNp.41)

Cauchy Note6 LNp.83

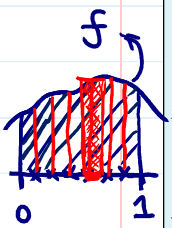
Proof: $E(\overline{X}_n) = \mu$, $Var(\overline{X}_n) = \sigma^2/n$

By Chebyshev's inequality, (LNp.43)

$$P(|\overline{X}_n - \mu| > \epsilon) \le \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0 \quad \text{as } n \to \infty$$

Notes. Under the same assumptions, a **strong law of large numbers (SLLN)**, which asserts that $\overline{X}_n \xrightarrow{a.s.} \mu$, can be proved.

Example 5.3 (Monte Carlo integration, TBp. 179)



To calculate $I(f) = \int_0^1 f(x)dx$, we can generate $X_1, X_2, \ldots, X_n$ i.i.d. $\sim U(0,1)$ and compute $\underbrace{\hat{I}(f)}_{\text{a r.v.} \leftarrow \overline{Y}_n} = \frac{1}{n}\sum_{i=1}^n \underbrace{f(X_i)}_{Y_i}$. By the LLN, ← $Y_1, \cdots, Y_n$ are i.i.d.

$\hat{I}(f)$ will be close to $E[f(X_i)] = \int_0^1 f(x) \times 1\, dx = I(f)$ as n is large.

$E(Y_i) = \mu = E[f(X_i)]$

Recall Thm4.1 in LNp.80

Example 5.4 (Repeated Measurements, TBp. 179-180)

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2, then

sample mean: an estimator of μ $\quad \overline{X}_n \xrightarrow{P} \mu$. (by WLLN) $\quad \left[E(Y_i) = E(X_i^2) = Var(X_i) + [E(X_i)]^2 = \sigma^2 + \mu^2\right]$

Let

sample variance: an estimator of σ^2 ($S_n^2 \xrightarrow{?} \sigma^2$)

$$S_n^2 \equiv \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \left(\frac{1}{n}\sum_{i=1}^n \underbrace{X_i^2}_{\equiv Y_i}\right) - \overline{X}_n^2. \leftarrow +\overline{X}_n^2 - 2\overline{X}_n^2$$

Because $g(x) = x^2$ is continuous, $\overline{X}_n^2 \xrightarrow{P} \mu^2$. Next, the r.v.'s $\underbrace{X_1^2}_{Y_1}, \ldots, \underbrace{X_n^2}_{Y_n}$ are i.i.d. with mean $\sigma^2 + \mu^2$. By WLLN

$$\frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \sigma^2 + \mu^2.$$

By Thm5.1, item 4, LNp.89

Therefore, $\boxed{\frac{n-1}{n} \cdot \frac{1}{n-1}}$

$$S_n^2 \xrightarrow{P} (\sigma^2 + \mu^2) - \mu^2 = \sigma^2.$$

(**Note.** $\frac{1}{n}$ in S_n^2 can be replaced by $\frac{1}{n-1}$.) ← cf. S_n^2 in Thm4.1, LNp.80

pdf graph in LNp.78

Example 5.5

If $X_n \sim t_n$, then $X_n \xrightarrow{d} N(0,1)$. ←

Let $Z \sim N(0,1)$, $U_1, \cdots, U_n \sim \chi_1^2$ and $Z, U_1, \cdots, U_n$ independent.

$Z \big/ \sqrt{(U_1 + \cdots + U_n)/n} \sim t_n \leftarrow N(0,1) \big/ \sqrt{\chi_n^2/n}$

Note. $\overline{U}_n \xrightarrow{P} 1$ (WLLN) & Slutsky's Thm (LNp.90)

(Ec) If $X_n \sim F_{m,n}$, then $mX_n \xrightarrow{d} \chi_m^2$ as $n \to \infty$. ← $(\chi_m^2/m)/(\chi_n^2/n)$

- Let $X_1, \ldots, X_n$ be i.i.d. $\sim$ Exponential(1), ← can be any distribution
  then $\mu_X = E(X_i) = 1$, $\sigma_X^2 = Var(X_i) = 1$, for $i = 1, \ldots, n$, and

$$n\bar{X}_n = \sum_{i=1}^n X_i \sim \text{Gamma}(n,1) \Rightarrow \bar{X}_n \sim \frac{1}{n}\text{Gamma}(n,1).$$

$\frac{1}{n}Z_n$, where $Z_n \sim \text{Gamma}(n,1)$

→ Gamma(n,n) by item 6, LNp.74

$$\Rightarrow \quad E(\bar{X}_n) = \mu_X = 1, \ Var(\bar{X}_n) = \sigma_X^2/n = 1/n$$

$\bar{X}_n \xrightarrow{P} 1$

$\Updownarrow$

$\bar{X}_n \xrightarrow{d} 1$

cdf

1

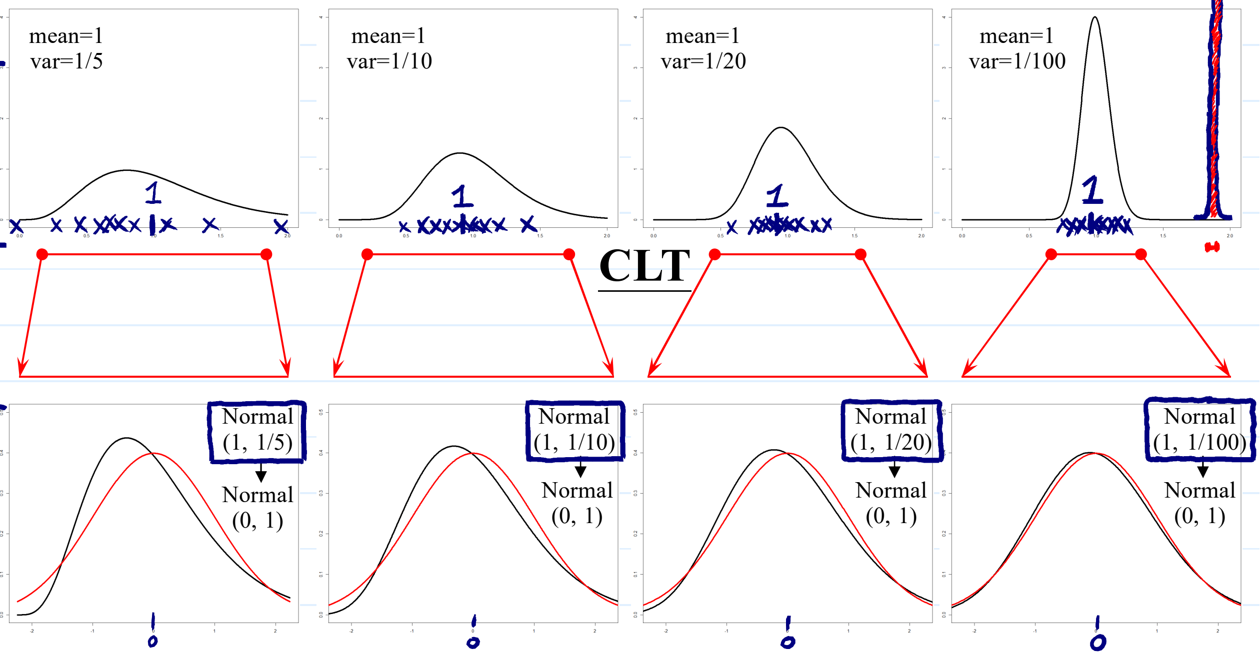
LLN

n=5 n=10 n=20 n=100

pdf of $\bar{X}_n$

$\frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$ (standardization)

pdf of $\frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$



Theorem 5.4 (Central Limit Theorem, TBp. 169)

Data → Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Let ← cf. Thms 3 (LNp 93)

$E(\bar{X}_n) = \mu$, $Var(\bar{X}_n) = \sigma^2/n$ — $\bar{X}_n \left(= \frac{1}{n}\sum_{i=1}^n X_i\right)$ and $T_n \left(= n\bar{X}_n = \sum_{i=1}^n X_i\right)$ — $E(T_n) = n\mu$, $Var(T_n) = n\sigma^2$

be the average and the sum of data, respectively. Then,

cdf $\quad \frac{(T_n - n\mu)/n}{\sigma\sqrt{n}/n} \quad$ cdf

$$\lim_{n\to\infty} P\left(\frac{T_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \lim_{n\to\infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le x\right) = \Phi(x),$$

standardization

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

$\boxed{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{T_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)}$

Proof. Let $W_i = \frac{X_i - \mu}{\sigma}$, then $E(W_i) = 0$ and $Var(W_i) = 1$. Let

standardization $\qquad Z_n \equiv \frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i$. $\quad E(W_i^2)$ → $W_1, W_2, \cdots$ are i.i.d.

Let $M(t)$ be the mgf of W_i's and $M_{Z_n}(t)$ be the mgf of Z_n, then

$$M_{Z_n}(t) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[\underbrace{M(0)}_{1} + \underbrace{M'(0)}_{0}\frac{t}{\sqrt{n}} + \frac{1}{2}\underbrace{M''(0)}_{1}\left(\frac{t}{\sqrt{n}}\right)^2 + \underbrace{o\left(\frac{1}{n}\right)}_{\text{little o}}\right]^n$$

$\left(1 + \frac{a_n}{n}\right)^n$, $a_n = \frac{t^2}{2} + b_n$, $a_n \to t^2/2$

$$= \left[1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \longrightarrow e^{t^2/2} \quad \left[\text{because if } a_n \to a, \left(1 + \frac{a_n}{n}\right)^n \to e^a\right]$$

$o\left(\frac{1}{n}\right) = \frac{b_n}{n}$, where $b_n \to 0$ $\qquad$ mgf of $N(0,1)$ → by Thm 5.2 (LNp.91)

Notes.

1. When mgf's do not exist, we can use chf's to prove it instead.
2. This is one of the simplest versions of CLT.

Example 5.6 (Normal approximation to Binomial distribution, TBp.187)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim B(1,p)$, then $T_n \sim B(n,p)$. Note that $E(X_i) = p, Var(X_i) = p(1-p)$ and $E(T_n) = np, Var(T_n) = np(1-p)$. By CLT,

standardization of a $B(n,p)$ random variable

$$\frac{\overset{\sim B(n,p)}{T_n} - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1),$$

i.e., when n is large enough, we can approximate the distribution of $B(n,p)$ by $N(np, np(1-p))$.

any p

cf.

p small

Note: 1. how about those distributions that can be generated from a sum of some i.i.d. random variables? (example?)

2. **(cf.)** Poisson in Def. 4.7 (LNp.66) and Example 5.2 (LNp.92)

Example 5.7 (measurement error (or called sampling error), TBp. 186)

- Suppose that you want to know the average income of families living in Taipei. = population
- If you can ask every families their incomes, you will get the exact value of the average, denoted by μ. ← unknown

census

sampling

- However, what if you only take a random sample of, say, 1000 families? → assume $X_1, \cdots, X_{1000}$ are i.i.d. (∵ 1000 ≪ population size) with mean μ. intuition # of all families
- The average income of the 1000 families, denoted by $\overline{X}_{1000}$, is a random variable. It has an error $\overline{X}_{1000} - \mu$, which is called measurement error or sampling error. unknown
- By CLT, the error will be distributed normally, and we can approximate $P(|\overline{X}_{1000} - \underset{E(\overline{X}_{1000})}{\mu}| < c)$ using normal distribution no matter what the distribution of incomes is. What normal? see Ex.5.9 (LNp.99)

Example 5.8 (experimental error)

- It is usually true that an experimental error ϵ is a function of a number of component errors $\epsilon_1, \ldots, \epsilon_n$. → $\epsilon = f(\epsilon_1, \cdots, \epsilon_n)$
- for example, errors in the settings of experimental conditions, errors due to variation in raw materials, and so on.
- If each individual component error is fairly small, it is possible to approximate the overall error ϵ as a linear function of independently distributed component errors

$$\epsilon \approx a_1\epsilon_1 + \ldots + a_n\epsilon_n.$$

approximation

advanced version of CLT

- By CLT, the distribution of ϵ will tend to normal as the number of component errors becomes large.
- This argument also offers a good justification for why in many statistical methods, such as in ANOVA or linear regression, the error part is assumed to be distributed normally.

Can it be applied to other quantities with normal distribution? e.g., height, weight, ...

Example 5.9 (cont. Ex.5.4 in LNp.94, Repeated Measurements)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2. Then, by LLN, CLT, and Slutsky's theorem, (LNp90)

cf. standardization

$$\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}\cdot\frac{\sigma}{S_n} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1),$$

cf. Thm 4.1, item 5 LNp.80 & Example 5.5 LNp.94

cf. Why is it useful?

because $\dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$ and $S_n^2 \xrightarrow{P} \sigma^2$ $\left(\Rightarrow \dfrac{\sigma^2}{S_n^2} \xrightarrow{P} 1\right)$.

Ex.5-4 (LNp.94)

$\boxed{\text{Ex5.7, LNp.98}}$ $\boxed{\overline{X}_n - \mu \overset{d}{\approx} N(0, \frac{S_n^2}{n})}$

❖ **Reading**: textbook, chapter 5

❖ **Further reading**: Roussas, chapter 8