All proof

Lecture 21 : Strong laws of large numbers. (SLLN)

Theorem 2 (SLLN Kolmogorov; page 215).

Suppose $X, X_1, \ldots, X_n$ are i.i.d. Let $\mu = EX$.

Then :

① $E|X| < \infty \implies \bar{X}_n \xrightarrow{a.s.} \mu$.

② $E|X| = \infty \implies \overline{\lim}\, \bar{X}_n \overset{a.s.}{=} \infty$.

Also (a way to check the condition $E|X| < \infty$):

$$E|X| < \infty \iff \overline{\lim}\, \bar{X}_n < \infty \quad a.s.$$

$$\iff M_n \equiv \frac{1}{n} \max_{1 \le k \le n} |X_k| \xrightarrow{a.s.} 0.$$

(We only prove ①).

Proof. We start as in the proof of the WLLN.

- Let $Y_n \equiv X_n \cdot 1_{[|X| \leq n]}$.

Then:

- $E|X|$ is finite $\iff \sum_{n=1}^{\infty} P(|X| \geq n) < \infty$ (By ineq. 2.1 page 206)

$\iff \sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$ (By the def. of Y_n)

$\iff$ X_n and Y_n are K-equivalent.

- By Prop. 2.1 page 206:

$$\bar{X}_n \xrightarrow{a.s.} \mu \iff \bar{Y}_n \xrightarrow{a.s.} \mu.$$

Therefore the problem reduces, as before, to showing

$$① \quad \bar{Y}_n \xrightarrow{a.s.} \mu.$$

Proof of ①

- Let $\mu_i \equiv E Y_i$; $i = 1, 2, 3, \dots, n$.

- Let $\bar{\mu}_n = \frac{\mu_1 + \dots + \mu_n}{n}$

- Recall that, in the course of the proof of the WLLN, we proved that $\mu_n \to \mu$; where $\mu_n = E Y_n$.

- Result from real analysis: $\lim_{n \to \infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n}$, if $\{b_n\} \uparrow \infty$.

- We apply this result with $\begin{cases} a_n = \mu_1 + \dots + \mu_n \\ b_n = n \end{cases}$

Thus $\lim_{n\to\infty} \bar{\mu}_n = \lim_{n\to\infty} \frac{\mu_{n+1}}{1} = \mu$. ②

- Write now $\bar{Y}_n = \frac{\sum_{i=1}^{n} (Y_i - \mu_i)}{n} + \bar{\mu}_n$ ③

- From ② and ③, to show ① reduces to showing

$$\boxed{\frac{\sum_{i=1}^{n} (Y_i - \mu_i)}{n} \xrightarrow{a.s.} 0}$$ ④

Proof of 4

Notice now that we can further write.

$$\frac{1}{n}\cdot\sum_{i=1}^{n}(Y_i-\mu_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i-\mu_i)}{i}\cdot i$$

$$\equiv \frac{1}{b_n}\sum_{i=1}^{n} b_i\cdot Z_i,$$ where by

notation we have $b_n \equiv n$ and so $b_i \equiv i$

and

$$Z_i \equiv \frac{Y_i-\mu_i}{i}.$$

- To show 4 we therefore show that

$$\boxed{\frac{1}{b_n}\sum_{i=1}^{n} b_i\cdot Z_i \xrightarrow{a.s.} 0} \quad 5$$

- To show 5 we use Kronecker's Lemma

<u>If</u> $\sum_{i=1}^{n} Z_i \xrightarrow{a.s} \text{(some) } Z$ <u>then</u> $\frac{1}{b_n}\sum_{i=1}^{n} b_i Z_i \xrightarrow{a.s.} 0,$

for any $b_n \nearrow \infty$.

- To show (5) we therefore need to show that

$$Z_n \equiv \sum_{i=1}^{n} \frac{(Y_i - \mu_i)}{i} \xrightarrow{a.s.} (\text{some r.v.})\ Z. \qquad (6)$$

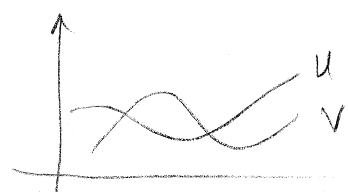
- To show (6) we use the Cauchy criterion of Proposition 2.3.3 page 30:

$$Z_n \xrightarrow{a.s.} (\text{some})\ Z \iff$$

$$P\Big(\max_{n \le m \le N} |Z_m - Z_n| \ge \varepsilon\Big) \le \varepsilon, \text{ for all } N \ge n_\varepsilon \text{ and all } \varepsilon > 0.$$

$$(7) \iff p^{\varepsilon}_{n,N} = \underline{P\Big(\max_{n < m \le N} \Big| \sum_{i=m+1}^{m} \frac{(Y_i - \mu_i)}{i} \Big| \ge \varepsilon\Big)} \xrightarrow[n \to \infty]{} 0.$$

The goal now becomes to show that (7) holds.



The problem is that the max of r.v may not be any one of those r.v. So you cannot use Markov ineq etc

Proof of 7

We use Kolmogorov's inequality (page ≥ 10, will be proved in the next Lecture).

$$P^{\varepsilon}_{n,N} \leq \frac{1}{\varepsilon^2} \sum_{i=n+1}^{N} \mathrm{Var}\left(\frac{Y_i - \mu_i}{i}\right)$$

$$= \frac{1}{\varepsilon^2} \sum_{i=n+1}^{N} \frac{\sigma_i^2}{i^2} \; ; \text{ for } \mathrm{var}(Y_i) \equiv \sigma_i^2$$

$$\leq \frac{1}{\varepsilon^2} \sum_{i=n+1}^{\infty} \frac{\sigma_i^2}{i^2} \xrightarrow[n\to\infty]{} 0, \text{ provided we}$$

show $\boxed{\displaystyle\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty \qquad 8}$.

Proof of 8.

$$LHS = \sum_{n=1}^{\infty} \frac{\mathrm{Var}\, Y_n}{n^2} \leq \sum_{n=1}^{\infty} \frac{E Y_n^2}{n^2} = \frac{1}{n^2} \sum_{n=1}^{\infty} \int_{|X| \leq n} x^2 \, dF$$

$$= \frac{1}{n^2} \sum_{n=1}^{\infty} \sum_{K=1}^{n} \int_{[K-1 \leq |X| \leq K]} x^2 \, dF$$

↓ continues

• $LHS \leq \sum_{K=1}^{\infty} \sum_{n=K}^{\infty} \frac{1}{n^2} \int_{[K-1 \leq |X| \leq K]} x^2 dF(x)$ (By changing the order of summation)

$$= \sum_{K=1}^{\infty} \left\{ \int_{[K-1 \leq |X| \leq K]} x^2 dF(x) \times \sum_{n=K}^{\infty} \frac{1}{n^2} \right\}$$

• Note now the following:

1) $\sum_{n=K}^{\infty} \frac{1}{n^2} = \sum_{n=K}^{\infty} \int_{n}^{n+1} \frac{dx}{n^2}$

2) $\frac{1}{n^2} \leq \frac{2}{x^2}$, for $n \leq x \leq n+1$, for any $n \geq 3$.

1) and 2) $\Rightarrow \sum_{n=K}^{\infty} \frac{1}{n^2} \leq \sum_{n=K}^{\infty} \int_{n}^{n+1} \frac{2dx}{x^2} = \int_{K}^{\infty} \frac{2dx}{x^2} = \frac{2}{K}$

• Therefore

$$LHS \leq \sum_{K=1}^{\infty} \frac{2}{K} \cdot \int_{[K-1 \leq |X| \leq K]} x^2 dF(x) \leq \sum_{K=1}^{\infty} \frac{2}{K} \cdot K \cdot \int_{[K-1 \leq |X| \leq K]} |x| dF(x)$$

$$= 2 \int |x| dF(x) = 2E|X| \underset{\substack{\uparrow \\ \text{By hypothesis}}}{<} \infty .$$ $\boxed{QED}$

Lecture 22

A maximal inequality and examples involving the SLLN and the Borel-Cantelli Lemmas

Proposition (Kolmogorov; Ineq. 3.1 page 210).

Let $X_1, \ldots, X_n, \ldots$ be independent, with $X_K \sim (0, \sigma_K^2)$.
Let $S_K \equiv X_1 + \ldots + X_K$. Then:

$$P\left(\max_{1 \le K \le m} |S_K| \ge \lambda\right) \le \frac{\operatorname{var} S_m}{\lambda^2} = \frac{\sum_{K=1}^{m} \sigma_K^2}{\lambda^2}, \text{ for all } \lambda > 0.$$

Do this proof

Proof Let

$$A_K \equiv \left[\max_{1 \le j < K} |S_j| < \lambda \le |S_K|\right]$$

- Notice that K is the first index for which $|S_K| \ge \lambda$. Such K is called "first passage time" in stochastic processes terminology.

Let $A \equiv \bigcup_{K=1}^{n} A_K = [\max_{1 \le K \le n} |S_K| \geq \lambda]$.

We thus need to show that

$$P(A) \le \frac{\operatorname{Var} S_n}{\lambda^2} \iff \operatorname{Var} S_n \geq \lambda^2 \cdot P(A) \quad (1).$$

We show now (1):

$$\operatorname{Var} S_n = \int S_n^2 \, dP \geq \int_A S_n^2 \, dP =$$

$$= \sum_{K=1}^{n} \int_{A_K} (S_n - S_K + S_K)^2 \, dP$$

$$= \sum_{K=1}^{n} \int \{ (S_n - S_K)^2 \cdot 1_{A_K} + 2 S_K \cdot (S_n - S_K) \cdot 1_{A_K} + S_K^2 1_{A_K} \} \, dP$$

$$\geq \sum_{K=1}^{n} \{ \int 0 \cdot dP + 2 E(S_K \cdot 1_{A_K}) \cdot E(S_n - S_K) + \int_{A_K} S_K^2 \, dP \}.$$

NOTE: $E(2 S_K (S_n - S_K) 1_{A_K}) \underset{\uparrow}{=} 2 E S_K 1_{A_K} \cdot E(S_n - S_K)$

Since $S_n - S_K$ is INDEPENDENT of S_K and A_K.
This is a commonly used trick in stochastic processes.

Now, since $E(S_n - S_k) = 0$ we further obtain

$$\operatorname{Var} S_n \geq \sum_{k=1}^{n} \int_{A_k} S_k^2 \, dP \underset{\uparrow}{\geq} \sum_{k=1}^{n} \lambda^2 \int_{A_k} dP \; \Rightarrow$$

Def. of A_k

$$\operatorname{Var} S_n \geq \lambda^2 \sum_{k=1}^{n} P(A_k) = \lambda^2 P(A)$$

q.e.d.

<u>Example 1</u>. If the identically distributed requirement is removed the SLLN (Kolmogorov) fails

Let $X_1, \dots, X_n, \dots$ be <u>independent</u> random variables.

Such that
$$P(X_n = -1) = 1 - \frac{1}{n^2}$$
$$P(X_n = n^2 - 1) = \frac{1}{n^2}.$$

Show that:

a) $E X_n = 0 \quad \forall n \geq 1$ (Immediate).

b) $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{a.s.} -1$

Does this result contradict the SLLN?

- To answer b) first note that

$$P(X_n = n^2 - 1) = P(X_n \neq -1) = \frac{1}{n^2} .$$

- Let $Y_n \equiv -1$, for all $n \geq 1$. $\Big\} \Rightarrow$

$$P(X_n \neq Y_n) = \frac{1}{n^2} \Rightarrow$$

$$\sum P(X_n \neq Y_n) < \infty \Rightarrow \{X_n\}_n \text{ and } \{Y_n\}_n$$

are K-equivalent. (1)

- Notice now that $\overline{Y}_n = \frac{-1 - 1 - \dots - 1}{n} = -1 \xrightarrow{a.s.} -1$ (2)

From (1), (2) and Proposition 1, page 6, Lecture 19 $\Rightarrow$

$$\overline{X}_n \xrightarrow{a.s.} -1 .$$

An application of both Borel-Cantelli Lemmas

(same as HW)

Exercise 2

Suppose that $X_1, \dots, X_n, \dots$ are i.i.d $Exp(1)$.

Show that : $\overline{\lim_n} \frac{X_n}{\log n} = 1$ a. s.

Solution: For any $\varepsilon > 0$

$$P(X_n > (1+\varepsilon)\log n) = e^{-(1+\varepsilon)\log n} = \frac{1}{n^{1+\varepsilon}}$$

Thus: $\sum_{n=1}^{\infty} P(X_n > (1+\varepsilon)\log n) = \sum_{n=1}^{\infty} \frac{1}{n^{1+\varepsilon}} < \infty \underset{\text{By B-C Lemma}}{\Longrightarrow}$

$$P\left(\left[\frac{X_n}{\log n} > 1+\varepsilon\right] \text{ i.o.}\right) = 0$$

$$\Downarrow$$

$$P\left(\overline{\lim} \frac{X_n}{\log n} \geq 1+\varepsilon\right) = 0$$

$$\Downarrow$$

$$\overline{\lim} \frac{X_n}{\log n} \leq 1+\varepsilon \text{ a.s} \underset{\varepsilon \to 0}{\Longrightarrow} \overline{\lim} \frac{X_n}{\log n} \leq 1 \text{ a.s} \quad (*)$$

Furthermore

$$P(X_n > (1-\varepsilon)\log n) = e^{-(1-\varepsilon)\log n} = \frac{1}{n^{1-\varepsilon}} \Rightarrow$$

$$\sum_{n=1}^{\infty} P\left(\frac{X_n}{\log n} > 1-\varepsilon\right) = \infty,$$ so by the second B-C lemma, for indep. random variables

$$\Downarrow$$

$$P\left(\frac{X_n}{\log n} > 1-\varepsilon \ \underline{i.o}\right) = 1 \Longleftrightarrow P\left(\overline{\lim}\left[\frac{X_n}{\log n} > 1-\varepsilon\right]\right) = 1$$

$$\Downarrow$$

$$\overline{\lim}\ \frac{X_n}{\log n} \geq 1-\varepsilon \quad a.s,$$ for all $\varepsilon > 0$, so taking $\varepsilon \to 0$ we have

$$\overline{\lim}\ \frac{X_n}{\log n} \geq 1, \ a.s. \ (**)$$

combining the "stars" completes the proof.

Q.ed.

General WLLN and SLLN: for independent, but not necessarily i.i.d r.v.s.

Let $X_1, \dots, X_n, \dots$ be $\boxed{\text{independent}}$ r.v.s.

a) If $\sum_{k=1}^{n} \frac{\sigma_k^2}{b_n^2} \to 0$, where $\sigma_k^2 = \operatorname{var} X_k$ and $\{b_n\}_n$ is any sequence of numbers s.t. $b_n \uparrow \infty$

then:

$$\frac{1}{b_n} \sum_{k=1}^{n} (X_k - \mu_k) \xrightarrow[n \to \infty]{P} 0 .$$

WLLN

b) If $\sum_{k=1}^{\infty} \frac{\sigma_k^2}{b_k^2} < \infty$, $\{b_k\}_k \uparrow \infty$ then.

SLLN

$$\frac{1}{b_n} \sum_{k=1}^{n} (X_k - \mu_k) \xrightarrow[n \to \infty]{a.s.} 0 ,$$

where $\mu_k = EX_k$.

Proof

1) Bonus problem (Hint: Use Chebyshev's inequality ; see page 222).

2) Recall that in the course of the proof of the SLLN we showed that:

$$(*) \quad \sum_{k=1}^{\infty} \frac{\sigma_k^2}{b_k^2} < \infty \Rightarrow \sum_{k=1}^{n} \frac{(X_k - \mu_k)}{b_k} \xrightarrow{a.s.} \text{(some) } U$$

- Use now Kronecker's Lemma, page 205 to conclude from $(*)$ that

$$\frac{1}{b_n} \sum_{k=1}^{n} (X_k - \mu_k) \xrightarrow{a.s.} 0 .$$

q.e.d.

NB 1 Notice the conditions on the variances σ_k^2. This is the price to pay if you want to drop the i.i.d. assumption.

NB 2 What sequence $\{b_n\}$ would you choose for Examples 1, page 3 ?

Lecture 23.

Applications of LLN's

Let $X_1, \ldots, X_n, \ldots$ be iid F.

We define the empirical df of $X_1, \ldots, X_n$ by

$$F_n(x) \equiv F_n(x, \omega) \equiv \frac{1}{n} \sum_{K=1}^{n} 1_{(-\infty, x]}(X_K(\omega))$$

(1)

$$= \frac{1}{n} \sum_{K=1}^{n} 1_{[X_K \le x]}, \text{ for all } x \in \mathbb{R}.$$

Theorem (Glivenko - Cantelli) (A very important result!)

$$\| F_n - F \|_\infty \equiv \sup_{x \in \mathbb{R}} | F_n(x) - F(x) | \xrightarrow[n \to \infty]{a.s.} 0.$$

NB This is a uniform SLLN applied to the random function F_n. (That is: it holds uniformly over all x, but don't forget that F_n also depends on ω. So, the a.s. refers, as always, to ω.

Proof

Let $\xi_1, \ldots, \xi_n$ iid Uniform $(0,1)$.

Recall that $X_K = F^{-1}(\xi_K)$ $(\sim F)$ (So, every X_K can be represented in this way).

Let F_n, as above, be defined by

$$F_n(x) = \frac{1}{n}\sum_{K=1}^{n} 1_{[X_K \le x]}. \qquad (3)$$

Let G_n be the empirical distr. function of $\xi_1, \ldots, \xi_n$, that is:

$$G_n(t) = \frac{1}{n}\sum_{K=1}^{n} 1_{[\xi_K \le t]} \cdot \qquad (4)$$

Recall that we have proved (Theorem 3.1 page 111) that: (5) $1_{[X \le \cdot]} = 1_{[\xi \le F(\cdot)]}$, ~~for~~ on $\mathbb{R}$, for very ω. (Thus the "dot" above is in $\mathbb{R}$ and the Suppressed argument of X and ξ, respectively, is ω.)

• Therefore: $F_n(x) = \frac{1}{n} \sum_{K=1}^{n} 1_{[\xi_K \leq F(x)]}$

Let I be the identity function.

Then, by 3, 4 and 5 we have:

$$\boxed{F_n - F = G_n(F) - I(F) \quad \text{on } \mathbb{R}, \text{ for every } \omega.}$$

That is: $F_n(x) - F(x) = G_n(F(x)) - I(F(x))$.

Since F is a df, $0 \leq F(x) \leq 1$, for ~~all~~ any $x \in \mathbb{R}$.

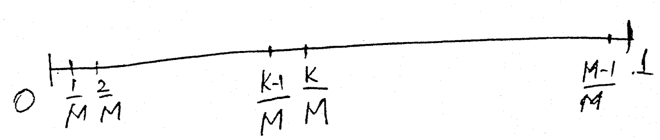
Thus:

$$\sup_{x \in \mathbb{R}} | F_n(x) - F(x) | \leq \sup_{t \in [0,1]} | G_n(t) - t |. \qquad (15)$$

In what follows we show that

$$\sup_{t \in [0,1]} | G_n(t) - t | \longrightarrow 0 \quad \text{a.s.}$$

Consider the following partitions of $[0,1]$, for M large, but fixed:



-11-

Then, for any $\frac{k-1}{M} \le t \le \frac{k}{M}$; $1 \le k \le M$ we have

$$\left.\begin{aligned} G_n(t) - t &\underset{G_n \uparrow}{\le} G_n\left(\frac{k}{M}\right) - \frac{k-1}{M} = G_n\left(\frac{k}{M}\right) - \frac{k}{M} + \frac{1}{M} \\ &\underline{\underline{\text{and}}} \\ G_n(t) - t &\ge G_n\left(\frac{k-1}{M}\right) - \frac{k}{M} = G_n\left(\frac{k-1}{M}\right) - \frac{k-1}{M} - \frac{1}{M}. \end{aligned}\right\} \quad (100)$$

Notice now that:

$$\begin{aligned} |G_n(t) - t| &\le \max\left(G_n(t) - t,\ t - G_n(t)\right) \\ &\underset{(100)}{\le} \max\left(G_n\left(\frac{k}{M}\right) - \frac{k}{M} + \frac{1}{M},\ -G_n\left(\frac{k-1}{M}\right) + \frac{k-1}{M} + \frac{1}{M}\right) \\ &\le \max\left(\left|G_n\left(\frac{k}{M}\right) - \frac{k}{M}\right| + \frac{1}{M},\ \left|G_n\left(\frac{k-1}{M}\right) - \frac{k-1}{M}\right| + \frac{1}{M}\right). \end{aligned}$$

<u>Thus</u>:

$$\begin{aligned} \sup_{t \in [0,1]} |G_n(t) - t| &\le \max\left\{ \max_{1 \le k \le M} \left|G_n\left(\frac{k}{M}\right) - \frac{k}{M}\right| + \frac{1}{M},\ \max_{1 \le k \le M} \left|G_n\left(\frac{k-1}{M}\right) - \frac{k-1}{M}\right| + \frac{1}{M} \right\} \\ &\le \max_{0 \le k \le M} \left|G_n\left(\frac{k}{M}\right) - \frac{k}{M}\right| + \frac{1}{M} \quad (11) \end{aligned}$$

- Assume now that M was chosen s.t. $\frac{1}{M} < \varepsilon$.
- We now show that
  $G_n\left(\frac{k}{M}\right) - \frac{k}{M} \xrightarrow{a.s.} 0$, for each $k = 0, \dots, M$.

il the definition of G_n:

$$G_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{[0 \le \xi_i \le t]} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} 1_{[0,t]}(\xi_i).$$

'w $G_n\left(\frac{K}{M}\right) = \frac{1}{n} \sum_{i=1}^{n} 1_{[0,\frac{K}{M}]}(\xi_i)$

If we let $Y_i \equiv 1_{[0,\frac{K}{M}]}(\xi_i) = \begin{cases} 1, & \text{if } \xi_i \in [0,\frac{K}{M}] \\ 0, & \text{otherwise.} \end{cases}$,

then $Y_i \sim$ Bernoulli with parameter $\frac{K}{M}$.

Also $EY_i = \frac{K}{M} < \infty$

Thus, by the SLLN:

$$G_n\left(\frac{K}{M}\right) - \frac{K}{M} \xrightarrow{a.s} 0 \quad (12)$$

Thus, by (11) and (12) and (15), since $\varepsilon > 0$ is arbitrary,

$$\|F_n - F\| \xrightarrow{a.s} 0$$

Example (Monte Carlo estimation).

Let $h: [0,1] \to [0,1]$ be a continuous function.
Let $(\xi_1, Z_1), \ldots, (\xi_n, Z_n)$ be iid pairs of $U(0,1)$ rv's (i.e. $\xi_i \sim \text{Unif}(0,1)$, $Z_i \sim \text{Unif}(0,1)$ for all i).

Show that $\bar{X}_n \xrightarrow{a.s} \int_0^1 h(t)\,dt$, where

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k$$

and

$$X_k = \mathbb{1}\{h(\xi_k) \geq Z_k\}.$$

Solution. Note that $X_1, \ldots, X_n$ are iid.
Thus, by the SLLN (since $E|X_i| < \infty$ for all i),

$$\bar{X}_n \xrightarrow{a.s} E(X_1).$$

$$EX_1 = E\,\mathbb{1}\{h(\xi_1) \geq Z_1\}$$
$$= \int_0^1 \int_0^1 \mathbb{1}\{h(t) \geq s\}\,ds\,dt = \int_0^1 \left(\int_0^{h(t)} ds \right) dt$$
$$= \int_0^1 h(t)\,dt.$$

$1-\frac{1}{3} = 2/3$

Lecture 24

$\log n^{2n}$

Theorem (Law of the iterated logarithm; LIL)

Let $X, X_1, \dots, X_n, \dots$ be iid rvs.

Let $S_n = X_1 + \dots + X_n$.

a) If $EX = 0$ and $\sigma^2 = \operatorname{var} X < \infty$ then

$$\overline{\lim} \frac{S_n}{\sqrt{2n \log\log n}} = \sigma \quad a.s.$$

$$\underline{\lim} \frac{S_n}{\sqrt{2n \log\log n}} = -\sigma \quad a.s.$$

Moreover:

$$\frac{S_n}{\sqrt{2n \log\log n}} \xrightarrow{a.s.} [-\sigma, \sigma]$$

(That is, a.s. ω, the limit set of the values of the LHS are in $[-\sigma, \sigma]$).

C'3: Under the same hypotheses, the SLLN gives $\frac{S_n}{n} \xrightarrow{a.s} 0$, thus the factor by which we divide S_n is crucial.

Prf: Reading exercise, pages 235 – 237.

<u>Summary</u> (The importance of the normalizing factor).

Suppose $X_1, X_2, \dots, X_n, \dots$ are i.i.d. $(\mu, 1)$.
Then:

1) $$\frac{\sum_{K=1}^{n} (X_K - \mu)}{n} \xrightarrow{a.s.} 0, \quad \text{by the SLLN.}$$

2) $$\frac{\sum_{K=1}^{n} (X_K - \mu)}{n^{\frac{1}{r}}} \xrightarrow{a.s.} 0 \quad , \text{ for } 1 \leq r < 2,$$

by Marcinkiewicz - Zygmund Th. (Ex. 4.4 page 220)

(<u>NB</u>. $r = 1$ is the SLLN.
This result establishes the smallest power of n for which we still have a.s. convergence to the mean).

3) $\dfrac{\sum_{K=1}^{n}(X_K-\mu)}{\sqrt{2n\log\log n}} \xrightarrow{a.s.} [-1,1]$, by the LiL.

4) $\dfrac{\sum_{K=1}^{n}(X_K-\mu)}{\sqrt{n}} \xrightarrow{d} N(0,1)$, by the CLT.

<u>NB</u> By the LiL : $\underline{\lim} \dfrac{\sum_{K=1}^{n}(X_K-\mu)}{\sqrt{n}} = -\infty$

$\overline{\lim} \dfrac{\sum_{K=1}^{n}(X_K-\mu)}{\sqrt{n}} = \infty$ a.s.

There is no contradiction here, since the a.s. convergence refers to (a.s) pointwise conv of a sequence refers to <u>actual size of the terms</u> (so, we look at the convergence in the sequence), whereas ~~the~~ ~~CLT~~ convergence in distribution looks (roughly) at the frequency with which these terms occur.

Convergence of series of $\boxed{\text{independent}}$ r.v.s

Theorem 1 Let $X_1, \dots, X_n, \dots$ be independent.
Then, for some r.v. S we have

$$S_n \equiv \sum_{k=1}^{n} X_k \xrightarrow{a.s.} S \iff \sum_{k=1}^{n} X_k \xrightarrow{P} S.$$

Theorem 2 (The two series theorem) Use in Hw 4

1 Let $X_1, \dots, X_n, \dots$ be independent with $X_k \cong (\mu_k, \sigma_k^2)$.

I) If $\sum_{k=1}^{n} \mu_k \to \mu$ and $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$, then

$$S_n \equiv \sum_{k=1}^{n} X_k \xrightarrow{a.s.} (\text{some r.v.})\ S.$$

II) If, in addition to 1, $\exists\, c > 0$ s.t. $|X_k| \le c$ for all k:

a) $$\sum_{k=1}^{n} (X_k - \mu_k) \xrightarrow{a.s.} (\underset{\text{r.v.}}{\text{some}})\ S_0 \iff \sum_{k=1}^{\infty} \sigma_k^2 < \infty$$

b) $$\sum_{k=1}^{n} X_k \xrightarrow{a.s.} (\underset{\text{r.v.}}{\text{some}})\ S \iff \begin{cases} \sum_{k=1}^{n} \mu_k \to \mu \\ \sum_{k=1}^{\infty} \sigma_k^2 < \infty \end{cases}.$$

Important: In either I) or II) b) we have:

$$S = S_0 + \mu, \quad \text{where}$$
$$\mu \equiv \sum_{k=1}^{\infty} \mu_k \quad \text{and } S_0 \text{ is defined in II) a)}$$

and

1) $ES = \mu$

2) $\operatorname{var} S = \sum_{k=1}^{\infty} \sigma_k^2$.

Theorem 3 (The 3 series theorem).

Let $X_1, \dots, X_n, \dots$ be independent rvs.

Let
$$X_k^c = \begin{cases} X_k, & \text{if } |X_k| \le c \\ 0, & \text{if } |X_k| > c \end{cases}, \quad \text{for some } c>0 \text{ and all } k.$$

Then: (a) $S_n \equiv \sum_{k=1}^{n} X_k \xrightarrow{a.s.} (\text{some})\ S$

if and only if the following 3 conditions hold

$$(*) \begin{cases} I_c \equiv \sum_{k=1}^{\infty} P(|X_k|>c) \text{ converges.} \Rightarrow \{X_k\} \text{ \& } \{X_k^c\} \text{ are } k\text{-equivalent} \\ \text{and} \\ II_c \equiv \sum_{k=1}^{\infty} \operatorname{Var} X_k^c \text{ converges.} \\ \text{and} \\ III_c \equiv \sum_{k=1}^{\infty} E X_k^c \text{ converges, for some } c>0. \end{cases}$$

$$\left. II_c,\ III_c \right\} \Leftrightarrow \sum_{k=1}^{n} X_k^c \xrightarrow{a.s.} S$$

(b) Condition (*) holds for <u>one</u> $c > 0$ $\Longleftrightarrow$?
Condition (*) holds for <u>all</u> $c > 0$.

(c) If either of the series in (*) diverges, for any c, then $\sum_{k=1}^{n} X_k$ diverges a.s.

Lecture 25

Convergence in distribution
General CLTs

- Recall the multivariate central limit theorem (CLT):

Theorem 1 (page 201)

Let $X_1, \dots, X_n \dots$ be iid (μ, Σ). Then

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow[n \to \infty]{d} N(0, \Sigma).$$

- In what follows we relax the identically distributed assumption and present a more general result (for one dimensional variables, for the time being).

$\Sigma_{m \times m}$ matrix

$X_i \in \mathbb{R}^m$

$\mu \in \mathbb{R}^m$

Notation and introductory remarks.

- Let Φ denote the cdf a r.v. $Z \sim N(0,1)$.
- Let W_n be a random variable of interest.
- Showing that W_n is asymptotically normal means showing the following:

$$F_{W_n}(z) \xrightarrow[n \to \infty]{} \Phi(z), \text{ for any cont. point of } \Phi.$$

- Alternatively we can show a stronger result.

$$\| F_{W_n} - \Phi \|_\infty \equiv \sup_z | F_{W_n}(z) - \Phi(z) | \longrightarrow 0 .$$

- This is the type of result that we'll show below.

- A central limit theorem for independent variable.
- Note that we not only establish convergence in distribution, but we also give the rate (δ_n below) with which it is achieved.

<u>Theorem</u> (Berry- Esseen CLT).

Suppose $X_{nk} \cong (\mu_{nk}, \sigma^2_{nk})$, $k=1,\dots,n$, are independent.

Let
$$Y_k = \frac{X_{nk} - \mu_{nk}}{\left(\sum_{k=1}^{n} \sigma^2_{nk}\right)^{1/2}}.$$

Define $W_n \equiv \sum_{k=1}^{n} Y_k \cong (0,1)$.

Let
$$\delta_n \equiv \frac{\sum_{k=1}^{n} E|X_{nk}-\mu_{nk}|^3}{\left(\sum_{k=1}^{n}\sigma^2_{nk}\right)^{3/2}}. \quad (*)$$

If $\delta_n \to 0$, then $\| F_{W_n} - \Phi \|_\infty \to 0$, as $n \to \infty$.
(In fact $\| F_{W_n} - \Phi \|_\infty \le 9\,\delta_n$).

<u>Remark</u> Definition ⊛ above implicitly assumes the existence of a third moment. the following theorem relaxes that.

<u>Theorem</u> (Lindeberg — Feller CLT)

Let $X_{nk} \cong F_{nk}(\mu_{nk}, \sigma_{nk}^2)$ independent, for $1 \le k \le n$. Let $s_n^2 \equiv \sum_{k=1}^{n} \sigma_{nk}^2$.

Define $$Z_n \equiv \frac{\sum_{k=1}^{n} (X_{nk} - \mu_{nk})}{s_n}.$$

The following are equivalent. (thus 1) and 2) are equivalent):

1) $\| F_{Z_n} - \Phi \| \to 0$ <u>and</u>

u.a.n.) $\max_{1 \le k \le n} P(|X_{nk} - \mu_{nk}|/s_n \ge \varepsilon) \to 0$

2) $$LF_n^{\varepsilon} \equiv \sum_{k=1}^{n} \int_{[|x-\mu_{nk}|/s_n \ge \varepsilon]} \left(\frac{x-\mu_{nk}}{s_n}\right)^2 dF_{nk}(x) \to 0,$$

for all $\varepsilon > 0$. (the Lindeberg condition)

Proof : 2) $\Rightarrow$ 1) (We'll prove 1) $\Rightarrow$ 2) later)

Note first that, by Chebyshev's inequality we have

$$P(|X_{nk} - \mu_{nk}|/s_n \geq \varepsilon) =$$

$$= P(|X_{nk} - \mu_{nk}| \geq \varepsilon \cdot s_n) \leq \frac{\sigma_{nk}^2}{\varepsilon^2 s_n^2}, \text{ for each } k = 1, \dots, n$$

Thus, to prove the (u. a. n) conditions it is enough to show that

$$\max_{1 \leq k \leq n} \frac{\sigma_{nk}^2}{s_n^2} \xrightarrow[n \to \infty]{} 0. \quad (2)$$

- Let now $Y_k \equiv \dfrac{X_{nk} - \mu_{nk}}{s_n}$.
- Let F_k denote the df of Y_k.
- Notice that, with this notation, we can re-write the L-F condition as

• (3') $\displaystyle\sum_{k=1}^{n} \int_{[|Y|\geq\varepsilon]} y^2\, dF_k(y) \longrightarrow 0$, for all $\varepsilon > 0$, and

So $\displaystyle\max_{1\leq k\leq n} \int_{[|Y|\geq\varepsilon]} y^2\, dF_k(y) \longrightarrow 0$. (3)

(Since $\max\limits_{1\leq k\leq n}$ {positive numbers} $\leq \sum\limits_{k=1}^{n}$ positive #

Notice now that

$$\operatorname{var}(Y_k) = \frac{\sigma_{nk}^2}{s_n^2} \leq E\, Y_k^2 .$$

Thus:

$$\max_k \frac{\sigma_{nk}^2}{s_n^2} \leq \max_k \int y^2\, dF_k(y) = \max_k \int_{[|Y|\leq\varepsilon]} y^2\, dF_k(y) + \max_k \int_{[|Y|\geq\varepsilon]} y^2\, dF_k(y)$$

$$\leq \varepsilon^2 + \varepsilon \leq \varepsilon ,$$

By (3)

which, by (2), proves that the u.a.n. condition holds.

We prove now that 2) ($L F_n^{\varepsilon} \to 0$) implies

① $\| F_{Z_n} - \Phi \| \longrightarrow 0$.

- Recall that $W \equiv Z_n \equiv \sum_{k=1}^{n} Y_k \cong (0,1)$. ;

$E\, Y_k = 0$, for all k

$\underline{\underline{\text{and}}}$

$$\sum_{k=1}^{n} E\, Y_k^2 = \frac{1}{S_n^2} \sum_{k=1}^{n} E\,(X_{n,k} - \mu_{n,k})^2 = \frac{S_n^2}{S_n^2} = 1.$$

- In what follows we use Slutsky's Lemma:

$$\left.\begin{array}{l} a)\ Z_n' \xrightarrow{d} Z \\ b)\ Z_n - Z_n' \xrightarrow{P} 0 \end{array}\right\} \Longrightarrow Z_n \xrightarrow{d} Z \ ; \text{ for } Z \sim N(0,1)$$

Thus, to show ① we'll construct a sequence Z_n' such that a) and b) hold.

- Define $Y_k' \equiv Y_k \mathbb{1}[|Y_k| \le \varepsilon_n]$. Let $Z_n' \equiv \sum_{k=1}^{n} Y_k'$.

<u>We first show b)</u>:

$$P\Big(\Big|\sum_{k=1}^{n} Y_k' - \sum_{k=1}^{n} Y_k\Big| \ge \varepsilon_n\Big) \le P\Big(\sum_{k=1}^{n} Y_k' \ne \sum_{k=1}^{n} Y_k\Big)$$
$$\le P(\exists\, k \in \{1,..,n\} \text{ such that } Y_k \ne Y_k')$$
$$\le \sum_{k=1}^{n} P(|Y_k| > \varepsilon_n), \text{ for some } \varepsilon_n \text{ to be chosen later}$$

- Notice that

$$E|Y_k|^2 \mathbb{1}[|Y_k| > \varepsilon_n] = E|Y_k|^2 \mathbb{1}[|Y_k|^2 > \varepsilon_n^2]$$
$$\ge \varepsilon_n^2 P(|Y_k| > \varepsilon_n)$$

• **Thus**:

$$P\left(\left|\sum_{k=1}^{m} Y'_k - \sum_{k=1}^{n} Y_k\right| \geq \varepsilon_n\right) \leq \frac{1}{\varepsilon_n^2} \sum_{k=1}^{n} E|Y_k|^2 1[|Y_k| > \varepsilon_n].$$

• Recall now that condition (3') page 3 says that.

$$LF_n^{\varepsilon} \equiv \sum_{k=1}^{n} E|Y_k|^2 1[|Y_k| > \varepsilon] \xrightarrow[n \to \infty]{} 0, \text{ for } \underline{\text{all}} \ \varepsilon > 0.$$

• Choose now ε_n s.t. $\frac{1}{\varepsilon_n^2} \cdot LF_n^{\varepsilon_n} \to 0$. Then

$$Z_n - Z'_n \xrightarrow{P} 0.$$

<u>We show now a)</u>:

We will use Berry-Esseen's Theorem to show that

$$Z'_n \xrightarrow{d} Z (\sim N(0,1)).$$

$$Z'_n = \sum_{k=1}^{n} \underbrace{Y_k 1[|Y_k| \leq \varepsilon_n]}$$

• Let $\mu_k^* \equiv E\, Y_k 1[|Y_k| \leq \varepsilon_n]$,

$\sigma_k^{2,*} = \text{var}\, Y_k 1[|Y_k| \leq \varepsilon_n]$.

• Let $U_k \equiv \dfrac{Y_k 1[|Y_k| \leq \varepsilon_n] - \mu_k^*}{\left(\sum_{k=1}^{n} \sigma_k^{2*}\right)^{1/2}}$; Notice that $\sum_{k=1}^{n} U_k \cong (0,1)$.

•

Thus $\left(\sum_{k=1}^{n} \sigma_k^{2,*}\right)^{\frac{1}{2}} U_k + \mu_k^* = Y_k \, 1[|Y_k| \le \varepsilon_n] \Rightarrow$

$$\left(\sum_{k=1}^{n} \sigma_k^{2,*}\right)^{1/2} \sum_{k=1}^{n} U_k + \sum_{k=1}^{n} \mu_k^* = Z_n' .$$

Thus : If we show
$$\left.\begin{array}{l} \sum_{k=1}^{n} U_k \xrightarrow{d} N(0,1) \\ (100) \quad \sum_{k=1}^{n} \mu_k^* \to 0 \\ (101) \quad \sum_{k=1}^{n} \sigma_k^{2*} \to 1 \end{array}\right\} \xrightarrow{\text{again by Slutsky}}$$

we will have that $Z_n' \to Z$.

- We now verify that the hypotheses of the B-E theorem hold for $\sum_{k=1}^{n} U_k$. We need to check that

$$\delta_n \equiv \frac{\sum_{k=1}^{n} E|Y_k \, 1[|Y_k| \le \varepsilon_n] - \mu_k^*|^3}{\left(\sum_{k=1}^{n} \sigma_k^{2,*}\right)^{3/2}} \to 0$$

By the C_r inequality we have:

$$E|Y_k \, 1[|Y_k| \le \varepsilon_n] - \mu_k^*|^3 \le 4\left(E|Y_k \, 1[|Y_k| \le \varepsilon_n]|^3 + (|\mu_k^*|)^3\right)$$

- Now: $\sum_{k=1}^{n}(|\mu_k^*|)^3 = \sum_{k=1}^{n} |E\, Y_k 1[|Y_k| \le \varepsilon_n]|^3$

$$= \sum_{k=1}^{n} |E\, Y_k 1[|Y_k| \le \varepsilon_n]| \cdot |E\, Y_k 1[|Y_k| \le \varepsilon_n]|^2$$

$$\underset{\substack{\uparrow \\ C-S}}{\le} \varepsilon_n \underbrace{\sum_{k=1}^{n} (EY_k^2)^{1/2 \cdot 2}}_{=1} \underbrace{\left[E(1[|Y_k| \le \varepsilon_n])^2\right]^{1/2 \cdot 2}}_{\le 1} = \varepsilon_n \to 0.$$

Thus, it remains to check that

$$(100) \quad \sum_{k=1}^{n} \mu_k^{*} \longrightarrow 0$$

$$(101) \quad \sum_{k=1}^{n} \sigma_k^{2,*} \longrightarrow 1$$

and

$$(102) \quad \sum_{k=1}^{n} E\left(\left|Y_k \mathbb{1}_{[|Y_k| \leq \varepsilon_n]}\right|^3\right) \longrightarrow 0 .$$

Proof of (100) - (102) $\longrightarrow$ See. Hwk. 5.

Then, indeed $\sum_{k=1}^{n} U_k \xrightarrow{d} N(0,1)$.

q.e.d.

Lecture 26

1. A discussion of the Lindeberg – Feller theorem.
2. The δ- method.
3. Some consequences of convergence in distribution

Discussion

1) What if Lindeberg's condition fails?

We construct a sequence of random variables $X_1, \dots, X_n$, independent ~~of~~ and with zero means such that, for

$$S_n \equiv X_1 + \dots + X_n$$

and

$$\sigma_n^2 = \mathrm{var}(S_n) \text{ we have:}$$

a) Lindeberg's condition fails

BUT

b) $\dfrac{S_n}{\sigma_n} \xrightarrow{d} N(0, a^2)$, with $a^2 < 1$ and,

moreover, $\max\limits_{1 \le k \le n} \dfrac{\sigma_k^2}{\sigma_n^2} \longrightarrow 0$.

Solution: • Let $U_1, \dots, U_n$ be independent $(0,1)$ random variables, where

$$U_n \begin{pmatrix} -cn & 0 & cn \\ \frac{1}{2n^2} & 1-\frac{1}{n^2} & \frac{1}{2n^2} \end{pmatrix}, \quad c>0.$$

• Notice that

$$\sum_{n=1}^{\infty} P(\underset{U_n \neq 0}{\underset{\Updownarrow}{|U_n| \geq \varepsilon}}) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \xrightarrow{\text{Borel-Cantelli}}$$

$$P(|U_n| \geq \varepsilon \ i.o.) = 0, \quad \text{for any } \varepsilon > 0.$$

Thus: $U_n \to 0$ a.s., and so

$$\sqrt{n} \cdot \frac{U_1 + \dots + U_n}{n} = \sqrt{n}\, \overline{U}_n \xrightarrow{P} 0. \qquad (2)$$

• Let $Y_1, \dots, Y_n$ be iid $(0,1)$, independent of the U_n's. By the CLT we have:

$$\sqrt{n}\, \overline{Y}_n \xrightarrow{d} N(0,1) \qquad (3)$$

- Construct now the desired sequence

$X_1, \ldots, X_n, \ldots$ independent

$$X_n \equiv U_n + Y_n .$$

- Let $S_n = \sum_{k=1}^{n} X_k$.

- $$\sigma_n^2 = \mathrm{var}(S_n) \underset{\text{indep}}{=} \sum_{k=1}^{n} \mathrm{var}(U_k + Y_k) =$$

$$\underset{\text{indep}}{=} \sum_{k=1}^{n} \left\{ \mathrm{var}(U_k) + \mathrm{var}(Y_k) \right\}$$

$$= n(c^2+1) \Longrightarrow \boxed{\sigma_n = \sqrt{n} \cdot \sqrt{c^2+1}}$$

- Then:

$$\frac{S_n}{\sigma_n} = (n \cdot \bar{U}_n + n \cdot \bar{Y}_n) \cdot \frac{1}{\sqrt{n}\sqrt{c^2+1}} =$$

$$= \frac{1}{\sqrt{c^2+1}} \cdot \sqrt{n}\,\bar{U}_n + \frac{1}{\sqrt{c^2+1}} \cdot \sqrt{n}\,\bar{Y}_n \xrightarrow[\text{(2) + (3)}]{\text{Slutsky's Lemma +}}$$

(4) $\frac{S_n}{\sqrt{n}} \xrightarrow{d} N\left(0, \frac{1}{1+c^2}\right) \equiv N(0, a^2)$, with $a^2 < 1$.

However, Lindeberg's condition fails:

$$LF_n^\varepsilon \equiv \sum_{K=1}^{n} \int_{[|x| \geq \varepsilon \cdot \frac{\sqrt{n}}{a}]} \left(\frac{x}{\sqrt{n}} \cdot a\right)^2 d F_{X_K}(x)$$

$\mu_n = 0$

$\sigma_n = \sqrt{n} \cdot \sqrt{c^2+1}$; $a^2 \equiv (c^2+1)^{-1} \iff \frac{1}{a} = \sqrt{c^2+1} \iff a = \frac{1}{\sqrt{c^2+1}}$

$$= \frac{a^2}{n} \sum_{K=1}^{n} \int_{[|x| \geq \varepsilon \frac{\sqrt{n}}{a}]} x^2 d F_{X_K}(x) \approx \rightarrow$$ see next page

Now, recall that $X = U + Y$

$U \downarrow$ discrete, $Y \downarrow$ continuous, $N(0,1)$

$$\underset{(5)}{\approx} \frac{a^2}{n} \sum_{K=1}^{n} \frac{c^2 K^2}{K^2} + \theta(1)$$

The "U" contribution The "Y" contribution

$$= \frac{a^2}{n} \cdot n \cdot c^2 + \theta(1) = \frac{c^2}{c^2+1} + \theta(1).$$

That is: $\angle F_n^{\varepsilon} \to \frac{c^2}{c^2+1} > 0$.

__hwk exercise__: Give a rigurous explanation of (5).

This is a __bonus__ exercise

__Morals__: We have shown that there exist sequences

$$X_n \xrightarrow{d} N(0, a^2) \quad \text{with } a^2 < 1$$

for which $\angle F_n^{\varepsilon} \not\to 0$.

__NB__: $a^2 = 1$ would correspond to $c^2 = 0$, that is to $\angle F_n^{\varepsilon} \to 0$.

Moral 2

This example also shows that, $\star$ in general:

$$V_n \xrightarrow[n\to\infty]{d} V \not\Longrightarrow var(V_n) \xrightarrow[n\to\infty]{} var(V).$$

- Take $V_n = \frac{S_n}{\sigma_n}$, for S_n and σ_n defined above.
- We showed that $V_n \xrightarrow{d} V$, for $V \sim N(0, \frac{1}{1+c^2})$ (*)

<u>Hence</u> $var(V) = \frac{1}{1+c^2}$.

- On the other hand, since $E V_n = 0$, we have:

$$var(V_n) = E(V_n^2) = \frac{E(S_n^2)}{\sigma_n^2} = \frac{\sigma_n^2}{\sigma_n^2} = 1.$$

Definition of σ_n^2

- Therefore, (*) holds, BUT:

$var(V_n) = 1$ <u>does NOT</u> converge (or equals) $var(V) = \frac{1}{1+c^2} < 1$

• For the "moreover part" notice that:

$$\sigma_K^2 = 1+c^2, \text{ for } K=\overline{1,n} \text{ and so}$$

$$\max_{1\le K\le n} \frac{\sigma_K^2}{\sigma_n^2} = \frac{1}{n} \to 0,$$ although L's condition fails.

2). An example in which both L's condition and "the max" condition fail, but we do have asympt. normality.

• Consider the following construction:

Independent
$$\begin{cases} X_{n1} \cong N(0, pn), \text{ for some } 0<p<1. \\ X_{n2} \equiv 0 \\ \vdots \\ X_{n[pn]} \equiv 0 \\ X_{n,[pn]+1} \cong N(0,1) \\ \vdots \\ X_{nn} \cong N(0,1) \end{cases}$$

Check that: 1) $\frac{S_n}{\sigma_n} \xrightarrow{d} N(0,1)$

2) L's cond. fails and

$$\max_{K=\overline{1,n}} \frac{\sigma_{nK}^2}{\sigma_n^2} = \frac{pn}{n} = p \neq 0$$

Some important results related to convergence in distribution (Pages 288 – 290).

Theorem 1 (the Helly - Bray Theorem).

$X_n \xrightarrow{d} X$ if and only if $Eg(X_n) \underset{n\to\infty}{\longrightarrow} Eg(X)$,

for all bounded and continuous functions g.

Theorem 2 Let $X \sim F$. Suppose that $X_n \xrightarrow{d} X$. If g is continuous a.s. F then

$g(X_n) \xrightarrow{d} g(X)$.

(This Theorem is called either the Mann-Wald theorem or the continuous mapping theorem).

Theorem 3: Theorem 11.7.4 page 190 (reading exercise).

Portmanteau

The delta-method

Let $\{W_n\}_n$, V be random variables.
Let $a \in \mathbb{R}$ be a fixed number.
Let $c_n \xrightarrow[n\to\infty]{} \infty$ be a sequence of real numbers.
Let g be a function on $\mathbb{R}$, differentiable at a.

- Assume that

$$c_n(W_n - a) \xrightarrow{d} V. \quad (4)$$

- Then:

$$c_n(g(W_n) - g(a)) \xrightarrow{d} g'(a)\, V. \quad (3)$$

Proof
- Recall that "g diff at a" means

$$(1) \quad g(x) - g(a) = g'(a)(x-a) + R(x-a),$$

where $\frac{R(x-a)}{|x-a|} \xrightarrow[x\to a]{} 0$, or, equivalently, $\underbrace{R(x-a) = o(|x-a|)}_{(5)}$.

- Replacing now x by W_n in (1) we obtain

$$g(W_n) - g(a) = g'(a)(W_n - a) + R(W_n - a) \Rightarrow$$

$$\therefore \; c_n(g(W_n) - g(a)) = c_n g'(a)(W_n - a) + c_n R(W_n - a).$$

- If we now show that

$$(2) \quad c_n R(X_n - a) \xrightarrow{P} 0 ,$$

then by Slutsky's lemma and (4) we have (3).

- To show (2), we first show that (5) still holds, with x replaced by X_n, and "o" replaced by "o_p".

(*) For this we shall show (a result worth remembering!) that:

$$(11) \quad c_n U_n \xrightarrow{d} (\text{some}) U ; \quad c_n \to \infty \longrightarrow U_n \xrightarrow{P} 0 .$$

- First recall that for ANY random vector U we have
$\forall \varepsilon > 0, \quad \exists M \text{ s.t. } P(|U| \geq M) < \varepsilon .$
- By Theorem 11.7.4 (page 290): $\forall \varepsilon > 0$
$P(c_n|U_n| \geq M) - P(|U| \geq M) < \varepsilon$, for n sufficiently large.
$\Downarrow$
$P(c_n|U_n| \geq M) \leq 2\varepsilon$, for $n \geq N$. Thus, increasing (if needed) M:

$$(10) \quad \forall \varepsilon > 0, \exists M > 0 \text{ st. } P(c_n|U_n| \geq M) < 2\varepsilon .$$

$\boxed{NB}$ This $\underline{M}$ does $\underline{\text{NOT}}$ depend on $\underline{\underline{n}}$.

Recall now that $c_n \to \infty \Rightarrow$

$\forall \varepsilon > 0, \quad \frac{1}{c_n} < \varepsilon, \quad$ for all $n \geq N$.

$\Rightarrow \forall c > 0$ and all $n \geq N \Rightarrow$

$$\frac{M}{c_n} < Mc \Rightarrow$$

$P(|U_n| \geq Mc) \leq P(c_n |U_n| \geq M) < 2\varepsilon$, for all $n \geq N$ and <u>all</u> $c > 0$. $\Rightarrow (U_n) = O_p(1)$

<u>Thus</u> $U_n \xrightarrow{P} 0$.

(**) Another VERY useful result is:
If $R(h) \equiv o(|h|)$, as $h \to 0$, then $R(U_n) = o_p(|U_n|)$, if $U_n \xrightarrow{P} 0$. (12)

<u>Proof</u> Let $t(h) = \begin{cases} \frac{R(h)}{|h|}, & \text{if } h \neq 0 \\ 0, & \text{if } h = 0 \end{cases}$. This t is continuous (~~we~~ we define $R(0) = 0$) and by the continuous mapping theorem we have the result.

Applying now (11) and (12) to $U_n \equiv W_n - a$, we thus have that $R(W_n - a) = o_p(|W_n - a|)$.

We can now re-write ① as:

$$c_n(g(W_n)-g(a)) = c_n g'(a)(W_n-a) + c_n\, o_p(|W_n-a|)$$
$$= g'(a)[c_n(W_n-a)] + o_p(c_n|W_n-a|)$$
$$\underset{(10)}{=} g'(a)[c_n(W_n-a)] + o_p(O_p(1))$$
$$= g'(a)[c_n(W_n-a)] + \underline{o_p(1)}.$$

q.e.d.

<u>Remark</u> <u>"The vector version" of the delta-method</u>

- Let $\Phi : D \subset \mathbb{R}^k \to \mathbb{R}^m$ be differentiable at $a \in D$.
- Let W_n be random vectors taking values in $D \subset \mathbb{R}^k$.
- Assume $c_n[W_n - a] \xrightarrow{d} V$, where $a \in \mathbb{R}^k$, $c_n \to \infty$, V takes values in $\mathbb{R}^k$.
- <u>Then</u>:

$$c_n[\Phi(W_n) - \Phi(a)] \xrightarrow{d} \underbrace{\underset{m\times k}{\Phi'(a)}\cdot \underset{k\times 1}{V}}_{m\times 1}, \quad \text{where}$$

$$\Phi'(a) = \begin{pmatrix} \frac{\partial \Phi_1(a)}{\partial x_1} & \cdots & \frac{\partial \Phi_1(a)}{\partial x_k} \\ \vdots & & \\ \frac{\partial \Phi_m(a)}{\partial x_1} & \cdots & \frac{\partial \Phi_m(a)}{\partial x_k} \end{pmatrix}.$$

READ for HW

Lecture 27: Some statistical applications of convergence in distribution

1) The bootstrap

Let $X_1, \ldots, X_n$ iid F.

Let $\vec{X}_n \equiv (X_1, \ldots, X_n)$. Let F_n be the empirical df based on $\vec{X}_n$.

Note that $\bar{X}_n = \int x \, dF_n(x) \left(= \frac{1}{n} \int x \, d \sum_{k=1}^{n} 1_{[X_k \le x]} = \frac{X_1 + \ldots + X_n}{n} \right)$.

Denote the variance w.r.t to F_n by S_n^2.

(So $S_n^2 = E_{F_n}(X - E_{F_n}(X))^2$; with $E_{F_n}(X) = \bar{X}_n$.

Def A *bootstrap sample* is an iid sample from F_n.

Let $\vec{X}_n^* \equiv (X_{n1}^*, \ldots, X_{nn}^*)$ be a bootstrap sample.

(*Note* Creating a bootstrap sample from the original sample is equivalent with sampling with replacement from $(X_1, \ldots, X_n)$.)

Once we get the bootstrap sample, we can again compute the "bootstrap" empirical distribution ; $F_n^*(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{[X_k^* \le x]}$ and the associated $\bar{X}_n^*$ and S_n^*.

Let $D_n \equiv \max_{1 \le k \le n} \frac{X_k - \bar{X}_n}{\sqrt{n}\, S_n}$.

Define: $$\bar{Z}_n^* = \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{S_n}$$

$$T_n^* = \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{S_n^*}.$$

Theorem

1) $\bar{Z}_n^* \xrightarrow{d} N(0,1)$ iff $D_n \xrightarrow{P} 0$.

2) If $D_n \xrightarrow{P} 0$ then $T_n^* \xrightarrow{d} N(0,1)$.

Remark: The above convergences are wrt the joint probability on $\Omega \times \Omega_n^*$. This type of requirement is called "weak bootstrap". For "strong bootstrap" and a proof see pages 274-275.

2) Asymptotic normality of the sample variance

Let $X_1, \ldots, X_n$ be iid (μ, σ^2).

Assume $\mu_4 \equiv EX^4 < \infty$ and $\sigma^2 > 0$.

Let $S_n^2 \equiv \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2 =$ the sample variance

Let $Z_k \equiv \frac{X_k - \mu}{\sigma} \cong (0, 1)$

$$Y_k \equiv Z_k^2 = \left[\frac{X_k - \mu}{\sigma}\right]^2 \cong \left(1, \frac{\mu_4}{\sigma^4} - 1\right)$$

$$= \left(1, 2\left(1 + \frac{\gamma_2}{2}\right)\right);$$

where $\gamma_2 \equiv \frac{(\mu_4 - 3\sigma^4)}{\sigma^4} \equiv$ the kurtosis, measure the tail heaviness of X.

- We write: $U_n =_a V_n$ if $U_n - V_n \xrightarrow{P} 0$.

Theorem

1) $\sqrt{n}\,[S_n^2-\sigma^2]\cdot\frac{1}{\sqrt{2}\,\sigma^2} =_a \frac{1}{2}\sqrt{n}\,[\bar{Y}_n-1] \xrightarrow{d} N(0,1+$

2) $\sqrt{n}\,[S_n-\sigma]\cdot\frac{2}{\sigma} =_a \sqrt{n}\,[\bar{Y}_n-1] \xrightarrow{d} \sqrt{2}\,N(0,1+\frac{\delta_2}{2}$

Proof: We just show 1). The proof of 2) is similar

First note that, by rearranging the sum, we obtain

$$\frac{S_n^2}{\sigma^2} = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{k=1}^{n}(Z_k-\bar{Z}_n)^2 \Rightarrow$$

$$\frac{\sqrt{n}(S_n^2-\sigma^2)}{\sqrt{2}\,\sigma^2} = \frac{n}{n-1}\cdot\frac{1}{\sqrt{2}}\cdot\sqrt{n}(\bar{Y}_n-1)$$

$$-\frac{n}{n-1}\cdot\frac{1}{\sqrt{2}}\cdot\sqrt{n}\cdot\bar{Z}_n^2 - \frac{\sqrt{n}}{\sqrt{2}(n-1)}$$

$$\underset{\uparrow}{=_a} \frac{1}{\sqrt{2}}\sqrt{n}(\bar{Y}_n-1) \underset{\uparrow}{\xrightarrow{d}} N\left(0,\frac{var(Y)}{2}\right) = N\left(0,1+\frac{\delta_2}{2}\right)$$

Bonus exercise! CLT

qed

3) A Berry-Esseen CLT for the LS estimates of the regression coefficients

Consider the regression model

$$Y = X\beta + \varepsilon \text{ ; where}$$

- $Y = (Y_1, \dots, Y_n)'$
- X is a $n \times p$ design matrix with deterministic entries $\{x_{ij}\}_{\substack{1 \le i \le n \\ 1 \le j \le p}}$.
- $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ where ε_i iid, $i = 1, \dots, n$ and

We assume: $E\,\varepsilon_i = 0$; $E\,\varepsilon_i^2 = \sigma^2$; $\tau_3 \equiv E\left|\frac{\varepsilon_1}{\sigma}\right|^3 < \infty$.

- Let $M \equiv (X'X)^{-1} \equiv (m_{ij})_{1 \le i, j \le p}$.
- $\beta \in \mathbb{R}^p$ is the unknown regression coefficient.
- Let $\hat{\beta}$ be the LS estimator of β; i.e.

$$\hat{\beta} = (X'X)^{-1} X' Y.$$

- Let $H \equiv X(X'X)^{-1}X \equiv (h_{ij})_{1 \le i,j \le n}$.

- <u>Result</u> Let G_{ni} be the distribution function of $\dfrac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{m_{ii}}}$. Then

$$\| G_{ni} - \Phi \|_\infty \le 9 \tau_3 \max_{1 \le K \le n} (h_{KK})^{1/2}.$$

- Outline of the proof (See the next pages for full details).

1) We first notice that, for every $1 \le i \le p$ we can write:

- $(*)$ $$\frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{m_{ii}}} \underset{\uparrow}{=} \sum_{K=1}^{n} \frac{a_{Ki}\,\varepsilon_K}{\sigma} \underset{\uparrow}{=} \sum_{K=1}^{n} \frac{a_K\,\varepsilon_K}{\sigma}$$

See pages o and oo below. Suppressing the index i for clarity

- Recall the Berry-Esseen Th., Lecture 25, page 3.

...

- Notice that :
    - $E\left(\frac{a_K \varepsilon_K}{\sigma}\right) = 0$ (here $\{a_K\}_K$ are fixed numbers).
    - $Var\left(\frac{a_K \varepsilon_K}{\sigma}\right) = a_K^2$

- It turns out (page o below) that

$$\sum_{K=1}^{n} a_k^2 = 1.$$

2) We write now the LHS in ⊛ in the form required by the B-E theorem.:

$$\frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{m_{ii}}} = \sum_{K=1}^{n} \underbrace{\frac{a_K \varepsilon_K}{\sigma}}_{\parallel \atop Y_K} = \sum_{K=1}^{n} Y_K \equiv W_n$$

- Note that in this case

$$\frac{X_{nK} - \mu_{nK}}{\left(\sum_{K=1}^{n} \sigma_{nK}^2\right)^{1/2}} = X_{nK} \equiv Y_K .$$

3) We apply now the theorem to obtain

$$\|G_{ni} - \Phi\|_\infty = \| F_{W_n} - \Phi \|_\infty \leq 9\, \gamma_n .$$

4) We evaluate

$$\gamma_n \equiv \frac{\sum_{k=1}^{n} E\,|X_{nk} - \mu_{nk}|^3}{\left(\sum_{k=1}^{n} \sigma^2_{nk}\right)^{3/2}}$$

$$= \sum_{k=1}^{n} E \left| \frac{a_k\, \varepsilon_{k.}}{\sigma} \right|^3$$

In this case.

$$\leq 9\, \tau_3 \max_{1 \leq k \leq n} (h_{kk})^{1/2} .$$

Page (00)

Qed

Proposition 6.1. *Assume that $\tau_3 \stackrel{\text{def}}{=} \mathbb{E}|\varepsilon_1/\sigma|^3 < \infty$. Let G_{ni} be the distribution function of $(\widehat{\beta}_i - \beta_i)/(\sigma\sqrt{m_{ii}})$. Then*

$$\|G_{ni} - \Phi\|_\infty \le 9\tau_3 \max_{1\le k\le n}\left(h_{kk}\right)^{1/2}.$$

Proof. Write U for the matrix with the eigen-vectors of M as its column vectors, and let Λ be the diagonal matrix with the eigen-values of M as its diagonal elements. Then

$$M = U\Lambda U^T$$

is the eigen-value decomposition of $M^{-1} = (X^T X)^{-1}$ in $\mathbb{R}^p$, and write $B = B_n = U\Lambda^{1/2}U^T$, so that $M^{-1} = B^2$. Let e_i be the ith unit vector and observe that

$$\frac{\widehat{\beta}_i - \beta_i}{\sigma\sqrt{m_{ii}}} = e_i^T \sigma^{-1} B(\widehat{\beta} - \beta).$$

Furthermore, write

$$B(\widehat{\beta} - \beta) = \underbrace{U\Lambda^{-1/2}U^T}_{p\times p}\underbrace{X^T}_{p\times m}\underbrace{\varepsilon}_{m\times 1} \stackrel{\text{def}}{=} \underbrace{F}_{p\times m}\underbrace{\varepsilon}_{m\times 1},$$

and

$$e_i^T F\varepsilon = e_i^T B(\widehat{\beta} - \beta) \stackrel{\text{def}}{=} \sum_{k=1}^n a_k \varepsilon_k,$$

with $a_k = a_{nk} = \langle e_i, f_k \rangle$ and $f_k \in \mathbb{R}^p$ is the kth column vector of F, $k = 1, \ldots, pn$. It is easily verified that $F^T F = H$ and $\mathrm{Cov}(F\varepsilon) = \sigma^2 I$, whence by Cauchy-Schwarz,

$$\max_k |a_k| \le \max_k \|f_k\|\|e_i\| = \max_k (h_{kk})^{1/2}$$

and

$$\sum_{k=1}^n a_k^2 = \|F^T e_i\|^2 = \|e_i\|^2 = 1.$$