

## Lecture 15: The central limit Thm

The central limit theorem is most likely one of the most fundamental results in probability and statistics.

It is traditionally taught with Fourier methods (i.e. characteristic functions) but, in a way, this method doesn't really extend well for things like martingales, etc.

The method of proof we will give is based on Lindeberg's proof & has proven to be much more powerful and conceptual.

## Taylor with remainder for $C_c^\infty$ functions

For  $x \in \mathbb{R}$  and  $f \in C_c^\infty(\mathbb{R})$  Taylor's Theorem gives

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \frac{f''(x)}{2}\Delta x^2 + R_f(x, \Delta x)$$

$$\text{where } R_f(x, \Delta x) = \frac{f'''(x_*)}{3!} \Delta x^3$$

$$\text{since } |R_f(x, \Delta x)| = \underbrace{\left| \frac{f'''(x_*)}{3!} \Delta x \right|}_{\Delta x^2} / \Delta x^2$$

this term  $\rightarrow 0$  as  
 $\Delta x \rightarrow 0$  uniformly in  $x$   
 since  $\|f'''\|_\infty < \infty$

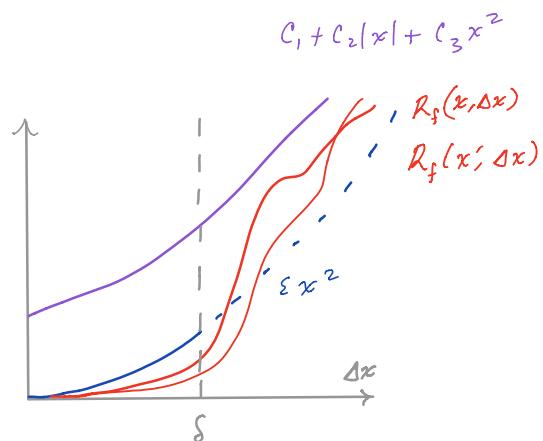
$$\therefore \forall \varepsilon > 0 \exists \delta > 0 \text{ s.t.}$$

$$|\Delta x| \leq \delta \Rightarrow |R_f(x, \Delta x)| \leq \varepsilon (\Delta x)^2$$

But we also have  $\max(\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty) < \infty$

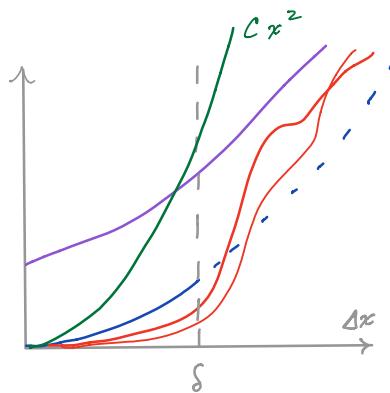
$$\therefore |\Delta x| \geq \delta \Rightarrow |R_f(x, \Delta x)| \leq c_1 + c_2 |\Delta x| + c_3 \Delta x^2$$

Here is the picture



Notice, however, we can find a  $C$  large enough so that

$$x \geq \delta \Rightarrow c_1 + c_2 x + c_3 x^2 \leq C x^2$$



$$\therefore \forall \varepsilon > 0 \exists \delta, C > 0 \text{ s.t.}$$

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \frac{f''(x)}{2}\Delta x^2 + R_f(x, \Delta x)$$

$$\sup_{x \in \mathbb{R}} |R_f(x, \Delta x)| \leq \varepsilon (\Delta x)^2 + C (\Delta x)^2 \int_{\{|x| \geq \delta\}} 1 dx$$

### Lindeberg's Method for CLT

As a warm up suppose  $X_1, X_2, \dots$  and  $Z_1, Z_2, \dots$  are all independent r.v.s  
s.t.  $E X_i = E Z_i$  &  
 $E X_i^2 = E Z_i^2 < \infty$

let  $g: \mathbb{R}^n \rightarrow \mathbb{R}$  be a smooth function and notice

$$\begin{aligned} & |Eg(X_1, \dots, X_n) - Eg(Z_1, \dots, Z_n)| \\ & \quad \pm Eg(Z_1, X_2, \dots, X_n) \\ & \quad \pm Eg(Z_1, Z_2, X_3, \dots, X_n) \\ & \quad \vdots \\ & \quad \pm Eg(Z_1, Z_2, \dots, Z_{n-1}, X_n) \\ & \leq \sum_{i=1}^n |Eg(\dots, Z_{i-1}, \textcolor{red}{X_i}, X_{i+1}, \dots) \\ & \quad - Eg(\dots, Z_{i-1}, \textcolor{red}{Z_i}, X_{i+1}, \dots)| \end{aligned}$$

$$= \sum_{i=1}^n |Eg_i(X_i) - Eg_i(Z_i)|$$

where  $g_i$  depends on  $\dots, Z_{i-1}$  &  $X_{i+1}, \dots$

Now let  $f \in C_c^\infty(\mathbb{R})$  and look what happens when we set

$$g(x_1, \dots, x_n) = f(x_1 + \dots + x_n)$$

(3)

$$\therefore g_i(x) = f(\dots + z_{i-1} + \textcolor{red}{x} + X_{i+1} + \dots)$$

(4)

$$\begin{aligned} & g_i(X_i) - g_i(0) \\ & = f(\dots + z_{i-1} + \textcolor{red}{X_i} + X_{i+1} + \dots) \\ & \quad - f(\dots + z_{i-1} + \textcolor{brown}{0} + X_{i+1} + \dots) \\ & = f'(zX_{(i)}) \textcolor{red}{X_i} + \underbrace{f''(zX_{(i)}) X_i^2}_{2} + R_f(zX_{(i)}, \textcolor{red}{X_i}) \end{aligned}$$

$$\text{where } zX_{(i)} := (\dots + z_{i-1} + X_{i+1} + \dots)$$

We also have

$$\begin{aligned} & g_i(z_i) - g_i(0) \\ & = f'(zX_{(i)}) \textcolor{red}{z_i} + \underbrace{f''(zX_{(i)}) z_i^2}_{2} + R_f(zX_{(i)}, \textcolor{red}{z_i}) \end{aligned}$$

Therefore

$$\begin{aligned} & Eg_i(X_i) - Eg_i(z_i) \leftarrow \pm Eg_i(0) \\ & = E \left[ f'(zX_{(i)}) (\textcolor{red}{X_i} - \textcolor{red}{z_i}) \right] \\ & \quad + E \left[ \underbrace{f''(\frac{zX_{(i)}}{2})}_{2} (\textcolor{red}{X_i^2} - \textcolor{red}{z_i^2}) \right] \\ & \quad + ER_f(zX_{(i)}, \textcolor{red}{X_i}) - ER_f(zX_{(i)}, \textcolor{red}{z_i}) \end{aligned}$$

But  $(X_i, z_i)$  is indep of  $zX_{(i)}$  so

$$\underbrace{E \left[ f'(zX_{(i)}) (\textcolor{red}{X_i} - \textcolor{red}{z_i}) \right]}_{\rightarrow} = 0$$

$$\underbrace{E \left[ \underbrace{f''(\frac{zX_{(i)}}{2})}_{2} (\textcolor{red}{X_i^2} - \textcolor{red}{z_i^2}) \right]}_{\rightarrow} = 0$$

$$\text{since } E(X_i - z_i) = E(X_i^2 - z_i^2) = 0$$

$\therefore \forall \varepsilon > 0 \exists \delta, c > 0$  s.t.

(5)

$$\begin{aligned} & |Eg_i(X_i) - Eg_i(z_i)| \\ & \leq E|R_f(z_{X_i}, X_i)| + E|R_f(z_{X_i}, z_i)| \\ & \leq \varepsilon EX_i^2 + c E(X_i^2 I_{|X_i| \geq \delta}) \\ & \quad + \varepsilon EZ_i^2 + c E(Z_i^2 I_{|Z_i| \geq \delta}) \end{aligned}$$

Let's put these results together into a lemma.

Lemma 1:

If  $X_1, X_2, \dots, Z_1, Z_2, \dots$  are indep r.v.s s.t.  $EX_i = EZ_i$  &  $EX_i^2 = EZ_i^2$  and  $f \in C_c^\infty(\mathbb{R})$  then  $\forall \varepsilon > 0$

$\exists \delta, c > 0$  s.t.

$$\begin{aligned} & |Ef(X_1 + \dots + X_n) - Ef(z_1 + \dots + z_n)| \\ & \leq 2\delta \sum_{i=1}^n EX_i^2 + c \sum_{i=1}^n E X_i^2 I_{|X_i| \geq \delta} \\ & \quad + c \sum_{i=1}^n EZ_i^2 I_{|Z_i| \geq \delta} \end{aligned}$$

Now the key "reason" the limit in the CLT is Gaussian is the following lemma.

Lemma 2:

If  $Z, z_1, \dots, z_n \stackrel{iid}{\sim} N(0, \sigma^2)$  then

$$z \stackrel{D}{=} \frac{z_1}{\sqrt{n}} + \dots + \frac{z_n}{\sqrt{n}}.$$

Theorem: (The central limit theorem)

(6)

If  $X_1, X_2, \dots$  are iid r.v.s with  $EX_i = 0$  and  $EX_i^2 = \sigma^2 < \infty$ , then

$$\frac{S_n}{\sqrt{n}} \xrightarrow{D} Z \sim N(0, \sigma^2)$$

where  $S_n := X_1 + \dots + X_n$ .

Proof:

since  $\text{var}\left(\frac{S_n}{\sqrt{n}}\right) = \sigma^2$  Chebyshev's Thm implies

$$\left|\frac{S_n}{\sqrt{n}}\right| = O_p(1) \Rightarrow \left\{\frac{S_n}{\sqrt{n}}\right\}_{n \geq 1} \text{ is tight}$$

Now  $\mathbb{R}$  is locally compact and therefore Portmanteau II implies it will be sufficient to show that  $\forall f \in C_c^\infty(\mathbb{R})$

$$\begin{aligned} & \left|Ef\left(\frac{S_n}{\sqrt{n}}\right) - Ef(z)\right| \rightarrow 0 \\ & \text{if Lemma 2} \quad \downarrow \dots \downarrow \text{iid } N(0, \sigma^2) \\ & \left|Ef\left(\frac{X_1}{\sqrt{n}} + \dots + \frac{X_n}{\sqrt{n}}\right) - Ef\left(\frac{z_1}{\sqrt{n}} + \dots + \frac{z_n}{\sqrt{n}}\right)\right| \quad (\star) \end{aligned}$$

Now let  $c > 0$ .

By Lemma 1,  $\exists \delta, c > 0$  s.t.

$$\begin{aligned} (\star) & \leq 2\delta \sum_{i=1}^n \frac{EX_i^2}{n} + c \sum_{i=1}^n E \frac{X_i^2}{n} I_{|X_i| \geq \sqrt{n}\delta} \\ & \quad + c \sum_{i=1}^n E \frac{Z_i^2}{n} I_{|Z_i| \geq \sqrt{n}\delta} \\ & = 2c\sigma^2 + c EX_i^2 I_{|X_i| \geq \sqrt{n}\delta} \\ & \quad + c EZ_i^2 I_{|Z_i| \geq \sqrt{n}\delta} \end{aligned}$$

since  $X_i \sim Z_i$ ,  $Z_i \sim Z$  &  $EX_i^2 = EZ_i^2 = \sigma^2$

Now by the DCT we get

(7)

$$E X_i^2 I_{|X_i| \geq \sqrt{n}} \xrightarrow{n \rightarrow \infty} 0$$

$$E Z_i^2 I_{|Z_i| \geq \sqrt{n}} \xrightarrow{n \rightarrow \infty} 0.$$

Therefore

$$\limsup_{n \rightarrow \infty} \left| E f\left(\frac{S_n}{\sqrt{n}}\right) - E f(z) \right| \leq 2\epsilon\sigma^2$$

Since  $\epsilon > 0$  was arbitrary

$$\left| E f\left(\frac{S_n}{\sqrt{n}}\right) - E f(z) \right| \xrightarrow{n \rightarrow \infty} 0$$

as was to be shown. QED

### Corollary:

If  $Z$  is a r.v. that satisfies

$$(i) \text{ var}(Z) := \sigma^2 < \infty$$

$$(ii) Z = \frac{Z_1}{\sqrt{n}} + \dots + \frac{Z_n}{\sqrt{n}}$$

where  $Z_1, \dots, Z_n$  are indep. copies of  $Z$

then  $Z$  must be Gaussian:

$$Z \sim N(0, \sigma^2)$$

Proof:

Properties (i) & (ii) were the only properties of  $Z$  we use in the proof of the CLT.

By uniqueness of distributional limits it must be the only distribution with this property. QED

Theorem: (CLT for random vectors)

(8)

If  $X_1, X_2, \dots$  are iid d-dimensional r.v.s s.t.

$$E(X_i) = 0$$

$$E(X_i X_i^T) = \Sigma \in \mathbb{R}^{d \times d}$$

Then

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim N_d(0, \Sigma)$$

where  $S_n := X_1 + X_2 + \dots + X_n$ .

Proof:

One can use the same method of proof as in the CLT... but it requires d-dimensional Taylor approximations. Instead let's use the Cramér-Wold device & show

$$\forall k \in \mathbb{R}^d \quad \langle k, \frac{S_n}{\sqrt{n}} \rangle \xrightarrow{\mathcal{D}} \langle k, Z \rangle.$$

Now since

$$\langle k, \frac{S_n}{\sqrt{n}} \rangle = \frac{\langle k, X_1 \rangle}{\sqrt{n}} + \dots + \frac{\langle k, X_n \rangle}{\sqrt{n}}$$

where  $E \langle k, X_i \rangle = \langle k, E X_i \rangle = 0$

$$E \langle k, X_i \rangle^2 = k^T \Sigma k < \infty$$

$$\langle k, Z \rangle \sim N(0, k^T \Sigma k).$$

The univariate CLT applies.

$$\therefore \langle k, \frac{S_n}{\sqrt{n}} \rangle \xrightarrow{\mathcal{D}} \langle k, Z \rangle \text{ as } n \rightarrow \infty$$

as was to be shown. QED

### Theorem: (Lindeberg-Feller sufficiency)

Let  $X_{i,n}$  be a triangular array of independent r.v.s.

$$X_{11}$$

$$X_{12}, X_{22}$$

$$X_{13}, X_{23}, X_{33}$$

 $\vdots$ 
 $\ddots$ 

which satisfy

$$(a) E X_{i,n} = 0, \quad \forall n \quad \forall i \leq n$$

$$(b) \sum_{i=1}^n E X_{i,n}^2 = \sigma^2, \quad \forall n$$

$$(c) \sum_{i=1}^n E X_{i,n}^2 I_{|X_{i,n}| \geq \delta} \xrightarrow{n \rightarrow \infty} 0, \quad \forall \delta > 0$$

Then

$$X_{1n} + \dots + X_{nn} \xrightarrow{n \rightarrow \infty} N(0, \sigma^2).$$

Proof:

This proof is nearly identical to the proof we gave for the univariate CLT. I'll leave the details for a homework.

QED

In some sense the Lindeberg-Feller conditions are about as general as possible for independent r.v.s  $X_{i,n}$ . This can be seen by the following theorem

### Theorem: (Lindeberg-Feller necessity)

If  $\{X_{i,n}\}_{n=1, i=1}^{\infty, \infty}$  is a triangular array of independent r.v.s which satisfy (a) & (b) above and also  $\max_{1 \leq i \leq n} E X_{i,n}^2 \xrightarrow{n \rightarrow \infty} 0$  then

$$\sum_{i=1}^n X_{i,n} \xrightarrow{D} N(0, 1) \Rightarrow (\text{condition (c) above}).$$

(9)

Proof:

The standard proof uses characteristic functions which we don't want to get into. See Klunkel's book p. 319.

(10)

### Kolmogorov's 3 series theorem

We have been waiting to prove this theorem since we started studying the random series  $\sum_{k=1}^{\infty} \frac{R_k}{k}$  for Rademacher r.v.s  $R_k$ . Now that we have the Lindeberg CLT we can do it.

### Lemma: (Cauchy criterion for $\xrightarrow{a.e.}$ )

Let  $X_1, X_2, \dots$  be a sequence of real r.v.s. all defined on a probability space  $(\Omega, \mathcal{F}, P)$ .

Then the following are equivalent:

(i)  $\exists$  a real r.v.  $X$  on  $(\Omega, \mathcal{F}, P)$  s.t.

$$X_n \xrightarrow{a.e.} X$$

(ii)  $\forall \epsilon > 0,$

$$\lim_n \lim_m P(\max_{n \leq p \leq m} |X_n - X_p| > \epsilon) = 0$$

(iii)  $\forall \epsilon > 0,$

$$\lim_n \lim_m P(\max_{n \leq p \leq m} |X_n - X_p| \geq \epsilon) = 0.$$

Proof:

Clearly (i)  $\Leftrightarrow$  (ii).

$$\text{Let } I_{nm} := \max_{n \leq p \leq m} |X_n - X_p|$$

$$I_{n\infty} := \sup_{n \leq p < \infty} |X_n - X_p|$$

$$I_{\infty\infty} := \sup_{n \leq p, q < \infty} |X_p - X_q|$$

and Note that  $I_{\infty\infty} \xrightarrow{a.e.} 0 \Leftrightarrow (i)$

Also notice that Fata's lemma implies

(11)

$$\begin{aligned} P(I_{nm} > \varepsilon \text{ a.s.m.}) &\leq \liminf_m P(I_{nm} > \varepsilon) \\ &\leq \limsup_m P(I_{nm} > \varepsilon) \\ &\leq P(I_{nm} > \varepsilon \text{ i.o.m.}) \end{aligned}$$

but these are equal  
since  $I_{nm} \uparrow$  as  $m \uparrow$

$$\begin{aligned} \therefore \lim_m P(I_{nm} > \varepsilon) &= P(I_{nm} > \varepsilon \text{ i.o.m.}) \\ &= P(\lim_m I_{nm} > \varepsilon) \\ &\quad \text{again by monotony} \\ &\quad \text{of } I_{nm} \text{ w.r.t. } m \\ &= P(I_{n\infty} > \varepsilon) \end{aligned}$$

$$\begin{aligned} \therefore \lim_n \lim_m P(I_{nm} > \varepsilon) &= \lim_n P(I_{n\infty} > \varepsilon) \\ &= P(\lim_n \{I_{n\infty} > \varepsilon\}), \quad \text{since } I_{n\infty} \downarrow \text{as } n \uparrow \\ (*) \quad &= P(I_{n\infty} > \varepsilon \text{ i.o.n.}), \quad \text{since:} \\ &\quad (\lim_n \{I_{n\infty} > \varepsilon\})^c \\ &\quad \uparrow \\ &\quad \lim \uparrow \{I_{n\infty} \leq \varepsilon\} \\ &\quad \uparrow \\ &\quad \bigcup_n \{I_{n\infty} \leq \varepsilon\} \end{aligned}$$

$\underbrace{\quad}_{\text{I}_{n\infty} \downarrow} \quad \{I_{n\infty} \leq \varepsilon \text{ a.a.n}\}$

To finish notice

$$\begin{aligned} (i) \iff I_n &\xrightarrow{\text{a.e.}} 0 & \{I_{n\infty} = \varepsilon \text{ a.a.n}\} \\ \iff I_{n\infty} &\xrightarrow{\text{a.e.}} 0 \quad \text{since } I_{n\infty} \leq I_n \leq 2I_{n\infty} \\ \iff \forall \varepsilon > 0 \quad &P(I_{n\infty} > \varepsilon \text{ i.o.n.}) = 0 \\ \iff \forall \varepsilon > 0 \quad &\lim_n \lim_m P(I_{nm} > \varepsilon) \\ \iff \forall \varepsilon > 0 \quad &\lim_n \lim_m P(I_{nm} > \varepsilon) \end{aligned}$$

QED

Theorem: (Sum van  $< \infty \Rightarrow$  sum  $< \infty$ )

(12)

Let  $X_1, X_2, \dots$  be independent r.v.s s.t.

$E X_n = 0 \quad \forall n$ . Then

$$\sum_{p=1}^{\infty} E X_p^2 < \infty \Rightarrow P\left(\sum_{p=1}^{\infty} X_p < \infty\right) = 1$$

means the limit exist & is finite.

Proof:

Suppose  $\sum_{p=1}^{\infty} E X_p^2 < \infty$  and let  $S_n := \sum_{p=1}^n X_p$

Notice Kolmogorov's maximal inequality from Lecture 11 gives that  $\forall \varepsilon > 0$

$$\begin{aligned} P\left(\max_{n \leq p \leq m} |S_n - S_p| \geq \varepsilon\right) \\ = P\left(\max_{n < p \leq m} |X_{n+1} + \dots + X_p| \geq \varepsilon\right) \end{aligned}$$

$$\leq \frac{1}{\varepsilon^2} \sum_{p=n+1}^m E(X_p^2) \quad \text{requires mean zero, indep r.v.s with finite variance.}$$

Since  $\lim_n \lim_m \text{RHS} \rightarrow 0 \quad \forall \varepsilon > 0$ , the

Cauchy criterion lemma gives that  $S_n$  converges almost surely to a finite limit.

$$\therefore P\left(\sum_{p=1}^{\infty} X_p < \infty\right) = 1$$

QED.

Theorem: (Kolmogorov's Three series)

(13)

Let  $X_1, X_2, \dots$  be independent r.v.s.  
and let  $\delta > 0$ .

Then  $P\left(\sum_{p=1}^{\infty} X_p < \infty\right) = 0$  or 1 and  
is 1  $\Leftrightarrow$  each of the following hold

$$(i) \quad \sum_{p=1}^{\infty} P(|X_p| > \delta) < \infty$$

$$(ii) \quad \sum_{p=1}^{\infty} E(X_p I_{|X_p| \leq \delta}) < \infty$$

$$(iii) \quad \sum_{p=1}^{\infty} \text{var}(X_p I_{|X_p| \leq \delta}) < \infty$$

Proof:

In Lecture 9 we showed that Kolmogorov's 0-1 law established that  $\sum_{p=1}^{\infty} X_p$  is a tail event.

$$P\left(\sum_{p=1}^{\infty} X_p < \infty\right) = 0 \text{ or } 1$$

$\Leftrightarrow$  Suppose (i), (ii) & (iii) hold.

$$\text{By (iii)} \quad \sum_{p=1}^{\infty} E(X_p I_{|X_p| \leq \delta} - E(X_p I_{|X_p| \leq \delta}))^2 < \infty$$

mean 0 indep r.v.s

$\therefore$  The previous theorem implies

$$\sum_{p=1}^{\infty} (X_p I_{|X_p| \leq \delta} - E(X_p I_{|X_p| \leq \delta})) < \infty \quad \text{w.p. 1}$$

By (ii) we have

$$\sum_{p=1}^{\infty} X_p I_{|X_p| \leq \delta} < \infty \quad \text{w.p. 1}$$

Also (i) implies  $P(|X_p| > \delta \text{ i.o.p.}) = 0$  by  
the first Borel-Cantelli lemma.

$$\therefore P(X_p I_{|X_p| \leq \delta} = X_p \text{ a.a.p.}) = 1$$

$$P\left(\sum_{p=1}^{\infty} X_p < \infty\right) = P\left(\sum_{p=1}^{\infty} X_p I_{|X_p| \leq \delta} < \infty\right) = 1$$

( $\Rightarrow$ )

(14)

$$\text{Suppose } P\left(\sum_{p=1}^{\infty} X_p < \infty\right) = 1.$$

If (i) were not true then

$$\sum_{p=1}^{\infty} P(|X_p| > \delta) = \infty \Rightarrow P(|X_p| > \delta \text{ i.o.p.}) = 1$$

by the second Borel-Cantelli lemma

$$\Rightarrow |S_n - S_{n-1}| = |X_n| > \delta \text{ i.o.w.}$$

w.p. 1 where  $S_n = \sum_{p=1}^n X_p$

$\Rightarrow S_n$  doesn't have a limit  
 $\Rightarrow$  contradiction

So we have that (i) is true.

Now suppose (iii) is not true. For each  $n \in \mathbb{N}$  set

$$Y_i := X_i I_{|X_i| \leq \delta}$$

$$Y_{in} := \frac{Y_i - E(Y_i)}{\sqrt{\text{var}(Y_i)}} \quad \begin{matrix} \xrightarrow{i \text{ d.d.}} \\ \xrightarrow{n \rightarrow \infty} \end{matrix} \text{ by a constant}$$

so that  $Y_{in}$ 's are indep,  $EY_{in} = 0$  &  
 $\sum_{i=1}^n E(Y_{in})^2 = 1$ . Moreover  $\forall \varepsilon > 0$

$$\sum_{i=1}^n E(Y_{in}^2 I_{|Y_{in}| \geq \varepsilon}) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

since  $I_{|Y_{in}| \geq \varepsilon} = 0 \quad \forall i = 1, \dots, n$  when  $n$  is sufficiently large

$\therefore$  Lindeberg-Feller applies and gives

$$\sum_{i=1}^n Y_{in} \xrightarrow{D} N(0, 1)$$

Also notice that

$$\begin{aligned} \sum_{i=1}^n Y_i &\xrightarrow{\text{w.p. 1}} \sum_{i=1}^n X_i & \xrightarrow{\text{a.e.}} 0 \\ \text{since } \sum_{i=1}^n X_i &< \infty & \xrightarrow{\sum_{j=1}^n \text{var}(Y_j)} 0 \\ \xrightarrow{\text{goes to zero as } n \rightarrow \infty} & & \xrightarrow{\sum_{j=1}^n \text{var}(Y_j)} 0 \end{aligned}$$

By Slutsky we have

$$\begin{aligned} \sum_{i=1}^n Y_{in} - \frac{\sum_{i=1}^n Y_i}{\sum_{j=1}^n \text{var}(Y_j)} &\xrightarrow{\mathcal{D}} N(0,1) \\ = - \frac{\sum_{i=1}^n E(Y_i)}{\sum_{j=1}^n \text{var}(Y_j)} &\text{ which is not random} \end{aligned}$$

$\therefore$  contradiction

$\therefore$  (iii) holds as well.

Finally to show (ii) Notice that the previous Thm

$$\begin{aligned} (\text{iii}) &\Rightarrow \lim_n \sum_{p=1}^n (Y_p - E(Y_p)) \text{ exists \& is finite} \\ &\Rightarrow \lim_n \sum_{p=1}^n Y_p \text{ exists \& is finite} \\ &\text{since } \sum_{p=1}^{\infty} X_p < \infty \Rightarrow \sum_{p=1}^{\infty} Y_p < \infty \quad \text{QED} \end{aligned}$$

(15)

(16)

To analyze  $\sum_{p=1}^{\infty} \frac{R_p}{\sqrt{p}}$  we can use Kolmogorov Three series Theorem. Condition (iii) was

$$\sum_{p=1}^{\infty} \text{var}(X_p I_{|X_p| \leq \delta}) < \infty$$

$\Updownarrow$  setting  $X_p = \frac{R_p}{\sqrt{p}}$

$$\sum_{p=1}^{\infty} E\left(\frac{R_p^2}{p} I_{|R_p| \leq \sqrt{p}\delta}\right) < \infty$$

$\frac{1}{p}$  always 1 when  $\delta^2 \leq p$

$\therefore$  condition (iii) does not hold

$$\therefore P\left(\sum_{p=1}^{\infty} \frac{R_p}{\sqrt{p}} \text{ converges}\right) = 0 \text{ by 3 serie}$$

### Application to series with coin flip signs

Let  $R_1, R_2, \dots$  be the Rademacher r.v.s developed in Lecture 1 which are

independent &  $R_p = \begin{cases} -1 & \text{w.p. } \frac{1}{2} \\ 1 & \text{w.p. } \frac{1}{2} \end{cases}$

$$\begin{aligned} \sum_{p=1}^{\infty} \frac{1}{p} &= \infty \quad \xrightarrow{\sum_{p=1}^{\infty} \frac{(-1)^p}{p} < \infty} \text{tuned cancellation} \\ &\quad \xrightarrow{\sum_{p=1}^{\infty} \frac{R_p}{p} < \infty} \text{w.p. 1} \\ &\quad \text{Random cancellation.} \end{aligned}$$

We showed this in Lecture 7

$$\begin{aligned} \sum_{p=1}^{\infty} \frac{1}{\sqrt{p}} &= \infty \quad \xrightarrow{\sum_{p=1}^{\infty} \frac{(-1)^p}{\sqrt{p}} < \infty} \text{tuned cancellation} \\ &\quad \xrightarrow{\sum_{p=1}^{\infty} \frac{R_p}{\sqrt{p}}} \text{This was left unanswered.} \quad \text{?} \end{aligned}$$

### Poisson Law of Rare Events

In an exercise you showed

$$\text{Bin}(n, p_n) \xrightarrow{\mathcal{D}} \text{Poi}(\lambda)$$

when  $n p_n \rightarrow \lambda > 0$ . Notice that this can be rephrased to say the following

$$\underbrace{X_{1n} + X_{2n} + \dots + X_{nn}}_{\text{each one of these is an indep Ber}(p_n)} \xrightarrow{\mathcal{D}} \text{Poi}(\lambda)$$

Another thing to notice is that  $\text{Poi}(\lambda)$  has a similar invariance to  $N(0, \sigma^2)$

$$\left( Z \sim N(0, \sigma^2), z_i \stackrel{\text{indp}}{\sim} N(0, \sigma_i^2) \right) \Rightarrow Z = z_1 + \dots + z_n$$

$$\left( \lambda = \lambda_1 + \dots + \lambda_n, N_i \stackrel{\text{indp}}{\sim} \text{Poi}(\lambda_i) \right) \Rightarrow N = N_1 + \dots + N_n$$

Therefore it seems natural that one would get a similar extension of  $\text{Bin}(n, p_n) \xrightarrow{\mathcal{D}} \text{Poi}(\lambda)$  to a Lindeberg-Feller type result:

Theorem: (Law of rare events)

If  $X_{in} \stackrel{iid}{\sim} \text{Ber}(p_{in})$  satisfying

$$(i) \quad \sum_{i=1}^n p_{in} \rightarrow \lambda > 0 \text{ as } n \rightarrow \infty$$

$$(ii) \quad \max_{1 \leq i \leq n} p_{in} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Then

$$X_{1n} + X_{2n} + \dots + X_{nn} \xrightarrow{\mathcal{D}} \text{Poi}(\lambda)$$

as  $n \rightarrow \infty$ .

This proof is pretty straight forward using characteristic functions.

However we can get a much better result, with a much more interesting proof, that gives rates of convergence.

Recall that in Scheffé's Thm we defined the "Total-variation" distance on probability measures over a common measure space

$$\|P - Q\|_{TV} := \sup_{A \in \mathcal{F}} |P(A) - Q(A)|$$

$$\leq \int_{\Omega} |P - Q| d\mu \quad \text{when } dP = p d\mu \quad dQ = q d\mu$$

part of Scheffé's result

(17)

Theorem: (rates for Law of rare events).

If  $X_i \stackrel{\text{indep}}{\sim} \text{Ber}(p_i)$  then

$$\begin{aligned} \|\mathbb{E}(X_1 + \dots + X_n) - \text{Poi}(p_1 + \dots + p_n)\|_{TV} \\ \leq \sqrt{\frac{n}{\sum_{i=1}^n p_i^2}}. \end{aligned}$$

The proof of this theorem uses the notion of coupling.

Definition: Suppose  $P$  &  $Q$  are two probability measures on a measurable space  $(\Omega, \mathcal{F})$ . A coupling of  $P$  &  $Q$  is a pair of  $\mathbb{R}$ -valued r.v.s  $X, Y$  defined on a common probability space s.t.  $X \sim P$  &  $Y \sim Q$ .

The set of all couplings is denoted  $\Pi(P, Q)$ .

The Total variation distance has a nice expression in terms of coupling.

Lemma: (TV & coupling)

If  $P$  and  $Q$  are two probability measures defined on a measurable space  $(\Omega, \mathcal{F})$  with  $dP = p d\mu$  &  $dQ = q d\mu$  then

$$\begin{aligned} \|P - Q\|_{TV} &= \frac{1}{2} \int_{\Omega} |P - Q| d\mu \\ &= 1 - \int_{\Omega} P(X \neq Y) d\mu \\ &= \inf_{(X, Y) \in \Pi(P, Q)} P^*(X \neq Y) \quad \text{where } X, Y \text{ is defined on } (\Omega^*, \mathcal{F}^*, P^*) \\ &\text{with the infimum attained.} \end{aligned}$$

Proof:

First let

$$\begin{aligned}\bar{P} &= \int_{\Omega} (P - P \wedge Q) d\mu && \left. \right\} \text{finite measures} \\ \bar{Q} &= \int_{\Omega} (Q - P \wedge Q) d\mu\end{aligned}$$

Notice that  $\forall A \in \mathcal{F}$

$$\begin{aligned}P(A) - Q(A) &= \left( P(A) - \int_A P \wedge Q d\mu \right) \\ &\quad - \left( Q(A) - \int_A P \wedge Q d\mu \right) \\ &= \int_A (P - P \wedge Q) d\mu \\ &\quad - \int_A (Q - P \wedge Q) d\mu \\ &= \bar{P}(A) - \bar{Q}(A).\end{aligned}$$

$$\therefore \|P - Q\|_{TV} = \sup_{A \in \mathcal{F}} |\bar{P}(A) - \bar{Q}(A)|$$

$$\text{Now we show } \sup_{A \in \mathcal{F}} |\bar{P}(A) - \bar{Q}(A)| = 1 - \int_{\Omega} P \wedge Q d\mu.$$

Notice

$$\underbrace{0 - \bar{Q}(\Omega)}_{= - \left( 1 - \int_{\Omega} P \wedge Q d\mu \right)} \leq \bar{P}(\Omega) - \bar{Q}(\Omega) \leq \underbrace{\bar{P}(\Omega)}_{= 1 - \int_{\Omega} P \wedge Q d\mu} - 0$$

$$\therefore \sup_{A \in \mathcal{F}} |\bar{P}(A) - \bar{Q}(A)| = 1 - \int_{\Omega} P \wedge Q d\mu$$

To see why the infimum is attained  
set  $A = \{P < Q\}$  and notice

$$\begin{aligned}\bar{P}(A) &= \int_{P < Q} (P - P \wedge Q) d\mu = 0 && \left. \right\} \Rightarrow \bar{P}(A^c) = \bar{P}(\Omega) \\ \bar{Q}(A^c) &= \int_{P \geq Q} (Q - P \wedge Q) d\mu = 0 && \left. \right\} \Rightarrow \bar{Q}(A) = \bar{Q}(\Omega)\end{aligned}$$

(19)

$$\therefore |\bar{P}(A) - \bar{Q}(A)| = |0 - \bar{Q}(\Omega)| \\ = 1 - \int_{\Omega} P \wedge Q d\mu$$

In summary

$$\|P - Q\|_{TV} = \sup_{A \in \mathcal{F}} |\bar{P}(A) - \bar{Q}(A)|$$

$$= 1 - \int_{\Omega} P \wedge Q d\mu$$

$$= \int_{\Omega} \frac{P+Q}{2} - P \wedge Q d\mu$$

$$= \int_{\Omega} \frac{1}{2}|P-Q| d\mu$$

where the last line follows since for any  $t, s \geq 0$ ,  $\frac{1}{2}(t+s - |s-t|) = s \wedge t$ .

To finish we need to show

$$\|P - Q\|_{TV} = \inf_{(X,Y) \in \Pi(P,Q)} P^*(X \neq Y)$$

with the infimum attained.

Let  $(X,Y) \in \Pi(P,Q)$  and let  $(\mathbb{D}^*, \mathcal{F}^*, P^*)$  denote the base measure for  $(X,Y)$ .

Now for any  $A \in \mathcal{F}$  we have

$$\{Y \in A\} \subset \{X \in A\} \cup \{X \neq Y\}$$

$$\begin{aligned}\therefore Q(A) &= P^*(Y \in A) \quad \text{by def of } \Pi(P,Q) \\ &\leq P^*(X \in A) + P^*(X \neq Y) \\ &= P(A) + P(X \neq Y)\end{aligned}$$

$$\therefore Q(A) - P(A) \leq P(X \neq Y)$$

(20)

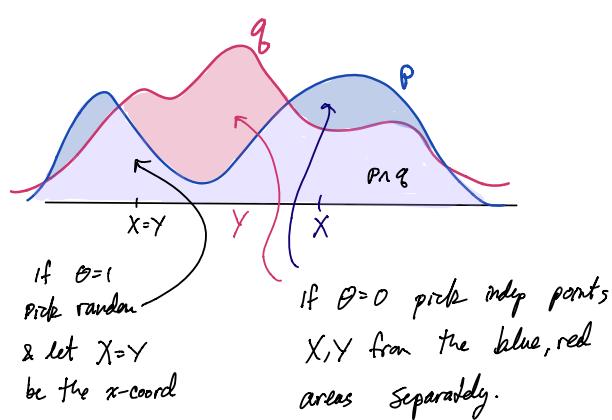
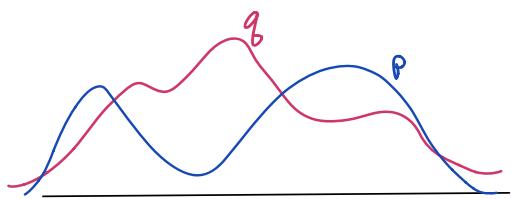
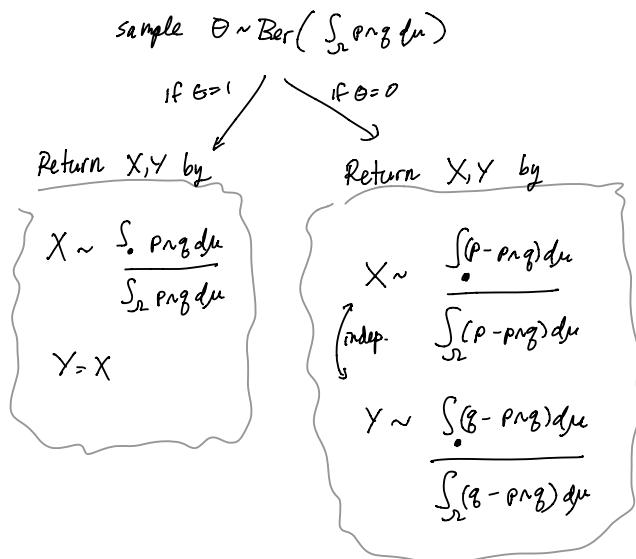
By symmetry we also have

(21)

$$P(A) - Q(A) \leq P^*(X \neq Y)$$

$$\therefore \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \leq \inf_{\substack{(X,Y) \in \\ \Pi(P,Q)}} P^*(X \neq Y) \quad (\star)$$

To show the minimum is attained consider the following r.v. coupling described by a randomized algorithm:



In the case  $0 < \int_{\Omega} p_Θ q_Θ dμ < 1$  we don't need to worry about conditioning on Prob zero events (The other boundary cases are easy). (22)

Therefore.

$$P^*(X \in A)$$

$$\begin{aligned} &= P^*(X \in A \mid \Theta=1) P^*(\Theta=1) + P^*(X \in A \mid \Theta=0) P^*(\Theta=0) \\ &= \frac{\int_A p_Θ q_Θ dμ}{\int_{\Omega} p_Θ q_Θ dμ} \cancel{\int_{\Omega} q_Θ dμ} + \frac{\int_A (p - p_Θ q_Θ) dμ}{\int_{\Omega} (p - p_Θ) dμ} \cancel{\left(1 - \int_{\Omega} p_Θ q_Θ dμ\right)} \\ &= \int_A p dμ \\ &= P(A) \end{aligned}$$

In a similar way  $P^*(Y \in A) = Q(A)$   
So that  $(X, Y) \in \Pi(P, Q)$ .

Now all that needs to be shown is that  $P^*(X \neq Y) = 1 - \int_{\Omega} p_Θ q_Θ dμ$ .

We have

$$P^*(X \neq Y) \geq \underbrace{\int_{\Omega} p_Θ q_Θ dμ}$$

since this is the chance of getting  $\Theta=1$  which means  $X \neq Y$

Taking complements gives

$$1 - \int_{\Omega} p_Θ q_Θ dμ \geq P^*(X \neq Y)$$

$$\geq \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \quad \text{by } (\star)$$

$$= 1 - \int_{\Omega} p_Θ q_Θ dμ$$

QED.

Proof of the rate bounds for the law of rare events:

Suppose  $X_i \stackrel{\text{indep}}{\sim} \text{Ber}(p_i)$  and let

$Y_i \stackrel{\text{indep}}{\sim} \text{Poi}(p_i)$ .

If  $\mu$  denotes the counting measure on  $\mathcal{S} = \{0, 1, 2, \dots\}$  then

$$\frac{d\mathbb{Z}(X_i)}{d\mu} = (1-p_i)I_{\{0\}} + p_i I_{\{1\}}$$

$$\frac{d\mathbb{Z}(Y_i)}{d\mu} = \sum_{k=0}^{\infty} e^{-p_i} \frac{p_i^k}{k!} I_{\{k\}}$$

$$\begin{aligned} \frac{d\mathbb{Z}(X_i)}{d\mu} \wedge \frac{d\mathbb{Z}(Y_i)}{d\mu} &= \min(1-p_i, e^{-p_i}) I_{\{0\}} \\ &\quad + \min(p_i, e^{-p_i} p_i) I_{\{1\}} \end{aligned}$$

By the TV Lemma we have

$$\begin{aligned} &\|\mathbb{Z}(X_i) - \mathbb{Z}(Y_i)\|_{TV} \\ &= 1 - \int \underbrace{\frac{d\mathbb{Z}(X_i)}{d\mu}}_{= 1-p_i} \wedge \underbrace{\frac{d\mathbb{Z}(Y_i)}{d\mu}}_{= e^{-p_i} p_i} d\mu \\ &= 1 - \underbrace{\min(1-p_i, e^{-p_i})}_{e^x} - \underbrace{\min(p_i, e^{-p_i} p_i)}_{1+x} \\ &\quad \text{Graph: A blue curve } e^x \text{ and a purple curve } 1+x \text{ intersect at } (0,1). \\ &= p_i (1 - e^{-p_i}) \leq p_i^2 \end{aligned}$$

(23)

Also By TV lemma  $\exists X_i^* \sim X_i$  &

$Y_i^* \sim Y_i$  s.t.

$$P^*(X_i^* \neq Y_i^*) = \|\mathbb{Z}(X_i) - \mathbb{Z}(Y_i)\|_{TV} \leq p_i^2$$

and letting

$$X^* = X_1^* + \dots + X_n^*$$

$$Y^* = Y_1^* + \dots + Y_n^*$$

we have

$$\|\mathbb{Z}(X_1 + \dots + X_n) - \text{Poi}(p_1 + \dots + p_n)\|_{TV}$$

$$= \|\mathbb{Z}(X_1^* + \dots + X_n^*) - \mathbb{Z}(Y_1^* + \dots + Y_n^*)\|_{TV}$$

$$= \|\mathbb{Z}(X^*) - \mathbb{Z}(Y^*)\|_{TV}$$

$$\leq P(X^* \neq Y^*)$$

$\Rightarrow \exists \text{ some } i \text{ s.t. } X_i^* \neq Y_i^*$

$$\leq \sum_{i=1}^n P(X_i^* \neq Y_i^*)$$

$$\leq \sum_{i=1}^n p_i^2$$

QED-

Notice this also shows that

$$\left. \begin{aligned} Y_n &\sim \text{Poi}(np_n) \\ np_n &\rightarrow \lambda \end{aligned} \right\} \Rightarrow \frac{Y_n - np_n}{\sqrt{n}} \xrightarrow{D} N(0, \lambda)$$

This follows since if  $X_n \sim \text{Bin}(n, p_n)$  then

$$\frac{X_n - np_n}{\sqrt{n}} \xrightarrow{D} N(0, \lambda)$$

$\therefore \forall x \in \mathbb{R}$

(25)

$$P\left(\frac{X_n - np_n}{\sqrt{n}} \leq x\right) \xrightarrow[n \rightarrow \infty]{} P(Z \leq x)$$

where  $Z \sim N(0, \lambda)$

||

$$P\left(\frac{Y_n - np_n}{\sqrt{n}} \leq x\right) + \varepsilon_n$$

$$\text{where } |\varepsilon_n| \leq \|J(X_n) - J(Y_n)\|_{TV}$$

$$\leq np_n^2 = \underbrace{(np_n)}_{\rightarrow \lambda} \underbrace{p_n}_{\rightarrow 0}$$

$$\therefore \frac{Y_n - np_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \lambda).$$

$$\max_{i \leq n} \frac{s_i}{\sqrt{n}} \xrightarrow{\mathcal{D}} |Z| \text{ where } Z \sim N(0, 1)$$

This is an old problem (first proved by Erdős & Kac I think) which we will revisit later once we prove the existence of the Wiener process. In this section I'll sketch a different proof that illustrates how flexible Lindeberg's method is for establishing distributional limits.

•  
•  
•

$$e^{-\|h-g\|_E^2}$$

$e^{-\|h-g\|_E^2}$  is positive definite on Hilbert spaces

### Stein's Method

Totally different method that can be used to get rates of distributional convergence of the CLT

•  
•  
•

uses our old friend Sheffé's theorem to give a proof that is far easier than the one devised by Von-Neumann & Schauder in the 1930's. It will also help us understand some of the Hilbert space isomorphisms associated with random fields in a later lecture.

•  
•  
•