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^∞ functions

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

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

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

since $$|R_f(x,\Delta x)| = \underbrace{\left|\frac{f'''(x_*)}{3!}\Delta x\right|}_{\substack{\text{this term } \to 0 \text{ as} \\ \Delta x \to 0 \text{ uniformly in } x \\ \text{since } \|f'''\|_\infty < \infty}} \Delta x^2$$

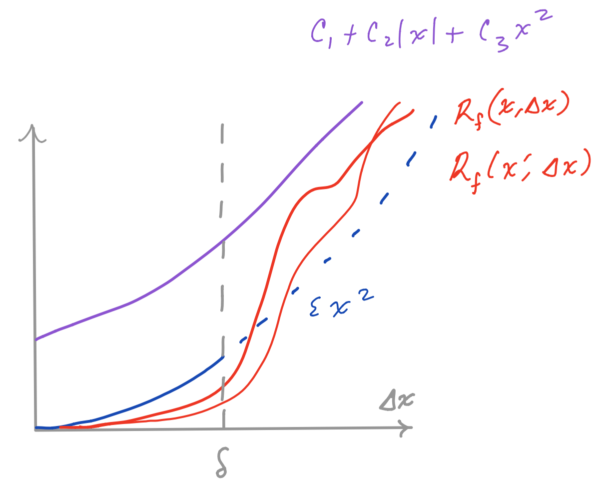
$\therefore$ $\forall \varepsilon > 0\ \exists \delta > 0$ s.t.

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

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

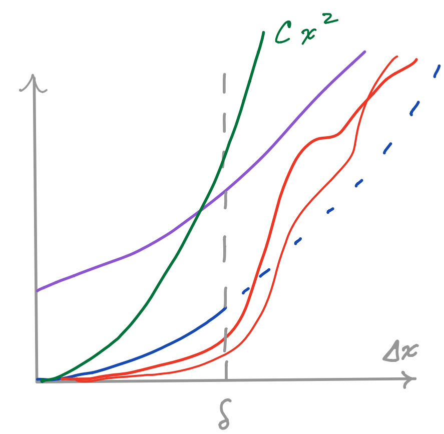
$\therefore$ $|\Delta x| \ge \delta \Rightarrow |R_f(x,\Delta x)| \le c_1 + c_2|\Delta x| + c_3\Delta x^2$

Here is the picture



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

$$x \ge \delta \Rightarrow c_1 + c_2 x + c_3 x^2 \le C x^2$$



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

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

$$\sup_{x\in\mathbb{R}} |R_f(x,\Delta x)| \le \varepsilon(\Delta x)^2 + C(\Delta x)^2 I_{\{|\Delta x| \ge \delta\}}$$

③

Lindeberg's Method for CLT

As a warmup suppose $X_1, X_2, \dots$ and $Z_1, Z_2, \dots$ are all independent r.v.s s.t. $EX_i = EZ_i$ &

$$EX_i^2 = EZ_i^2 < \infty$$

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

$$\begin{aligned}
&\left| Eg(X_1,\dots,X_n) - Eg(Z_1,\dots,Z_n) \right| \\
&\qquad \pm Eg(Z_1, X_2, \dots X_n) \\
&\qquad \pm Eg(Z_1, Z_2, X_3, \dots, X_n) \\
&\qquad \vdots \\
&\qquad \pm Eg(Z_1, Z_2, \dots, Z_{n-1}, X_n)
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^{n} \Big| Eg(\dots Z_{i-1}, X_i, X_{i+1}, \dots) \\
&\qquad - Eg(\dots Z_{i-1}, Z_i, X_{i+1}, \dots) \Big| \\
&= \sum_{i=1}^{n} \left| E g_i(X_i) - E g_i(Z_i) \right|
\end{aligned}$$

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

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

④

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

& $g_i(X_i) - g_i(0)$

$$\begin{aligned}
&= f(\dots + Z_{i-1} + X_i + X_{i+1} + \dots) \\
&\quad - f(\dots + Z_{i-1} + 0 + X_{i+1} + \dots) \\
&= f'(ZX_{(i)})X_i + \frac{f''(ZX_{(i)})}{2}X_i^2 + R_f(ZX_{(i)}, X_i)
\end{aligned}$$

where $ZX_{(i)} := (\dots + Z_{i-1} + X_{i+1} + \dots)$

We also have

$$\begin{aligned}
&g_i(Z_i) - g_i(0) \\
&= f'(ZX_{(i)})Z_i + \frac{f''(ZX_{(i)})}{2}Z_i^2 + R_f(ZX_{(i)}, Z_i)
\end{aligned}$$

Therefore

$$\begin{aligned}
&E g_i(X_i) - E g_i(Z_i) \longleftarrow \pm E g_i(0) \\
&= E\left[f'(ZX_{(i)})(X_i - Z_i)\right] \\
&\quad + E\left[\frac{f''(ZX_{(i)})}{2}(X_i^2 - Z_i^2)\right] \\
&\quad + ER_f(ZX_{(i)}, X_i) - ER_f(ZX_{(i)}, Z_i)
\end{aligned}$$

But (X_i, Z_i) is indep of $ZX_{(i)}$ so

$$E\left[f'(ZX_{(i)})(X_i - Z_i)\right] = 0$$

$$E\left[\frac{f''(ZX_{(i)})}{2}(X_i^2 - Z_i^2)\right] = 0$$

since $E(X_i - Z_i) = E(X_i^2 - Z_i^2) = 0$

5

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

$$|E g_i(X_i) - E g_i(Z_i)|$$
$$\leq E|R_f(ZX_{(i)}, X_i)| + E|R_f(ZX_{(i)}, Z_i)|$$
$$\leq \varepsilon E X_i^2 + C E(X_i^2 I_{|X_i| \geq \delta})$$
$$+ \varepsilon E Z_i^2 + C E(Z_i^2 I_{|Z_i| \geq \delta})$$

Lets put these results together into a lemma.

Lemma 1:

If $X_1, X_2, \ldots$ $Z_1, Z_2, \ldots$ 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.

$$|E f(X_1 + \cdots + X_n) - E f(Z_1 + \cdots + Z_n)|$$
$$\leq 2\varepsilon \sum_{i=1}^n E X_i^2 + c \sum_{i=1}^n E X_i^2 I_{|X_i| \geq \delta} + c \sum_{i=1}^n E Z_i^2 I_{|Z| \geq \delta}$$

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

Lemma 2:

If $Z, Z_1, \ldots, Z_n \overset{iid}{\sim} N(0, \sigma^2)$ then

$$Z \overset{\mathcal{D}}{=} \frac{Z_1}{\sqrt{n}} + \cdots + \frac{Z_n}{\sqrt{n}}.$$

6

Theorem: (The central limit theorem)

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

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

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

Proof:

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

$$\left|\frac{S_n}{\sqrt{n}}\right| = \mathcal{O}_p(1) \Longrightarrow \left\{\frac{S_n}{\sqrt{n}}\right\}_{n \geq 1} \text{ is tight}$$

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

$$\left|E f\left(\frac{S_n}{\sqrt{n}}\right) - E f(Z)\right| \longrightarrow 0$$

|| Lemma 2   iid $N(0, \sigma^2)$

$$\left|E f\left(\frac{X_1}{\sqrt{n}} + \cdots + \frac{X_n}{\sqrt{n}}\right) - E f\left(\frac{Z_1}{\sqrt{n}} + \cdots + \frac{Z_n}{\sqrt{n}}\right)\right| \quad (*)$$

Now let $\varepsilon > 0$.

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

$$(*) \leq 2\varepsilon \sum_{i=1}^n \frac{EX_i^2}{n} + c \sum_{i=1}^n \frac{E X_i^2 I_{|X_i| \geq \sqrt{n}\delta}}{n} + c \sum_{i=1}^n \frac{E Z_i^2 I_{|Z_i| \geq \sqrt{n}\delta}}{n}$$
$$= 2\varepsilon\sigma^2 + c E X_1^2 I_{|X_1| \geq \sqrt{n}\delta} + c E Z_1^2 I_{|Z_1| \geq \sqrt{n}\delta}$$

since $X_i \sim X_1$, $Z_i \sim Z_1$ & $EX_1^2 = EZ_1^2 = \sigma^2$

Now by the DCT we get

$$E X_1^2 I_{|X_1| \geq \sqrt{n}\delta} \xrightarrow{n\to\infty} 0$$

$$E Z_1^2 I_{|Z_1| \geq \sqrt{n}\delta} \xrightarrow{n\to\infty} 0.$$

Therefore

$$\limsup_{n\to\infty} \left| E f\left(\frac{S_n}{\sqrt{n}}\right) - E f(Z) \right| \leq 2\varepsilon\sigma^2$$

Since $\varepsilon > 0$ was arbitrary

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

as was to be shown.

QED.

Corollary:

If Z is a r.v. that satisfies

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

(ii) $Z \overset{d}{=} \frac{Z_1}{\sqrt{n}} + \cdots + \frac{Z_n}{\sqrt{n}}$

where $Z_1, \ldots, Z_n$ 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)

If $X_1, X_2, \ldots$ 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 $S_n := X_1 + X_2 + \cdots + X_n$.

Proof:

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

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

Now since

$$\left\langle k, \frac{S_n}{\sqrt{n}} \right\rangle = \underbrace{\frac{\langle k, X_1\rangle}{\sqrt{n}} + \cdots + \frac{\langle k, X_n\rangle}{\sqrt{n}}}_{\text{iid}}$$

where

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

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

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

the univariate CLT applies.

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

as was to be shown.

QED

Theorem: (Lindeberg-Feller sufficiency)

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

$$\begin{array}{l} X_{11} \\ X_{12}, X_{22} \\ X_{13}, X_{23}, X_{33} \\ \vdots \qquad\qquad \ddots \end{array}$$

which satisfy

(a) $EX_{in} = 0, \quad \forall n\ \forall i \le n$

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

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

Then

$$X_{1n} + \cdots + X_{nn} \xrightarrow[n\to\infty]{\mathcal{D}} N(0,\sigma^2).$$

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 sence the Lindgeberg-Feller conditions are about as general as possible for independent r.v.s X_{in}. This can be seen by the following Thm

Theorem: (Lindeberg-Feller necessity)

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

$$\sum_{i=1}^n X_{in} \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \left(\begin{array}{c}\text{condition} \\ (c) \text{ above}\end{array}\right).$$

Proof:

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

Kolmogorov's 3 series Theorem

We have been waiting to prove this theorem since we stared 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 $X_1, X_2 ..$ 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) $\forall \epsilon > 0$,

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

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

$$\lim_n \lim_m P\left(\max_{n\le p\le m} |X_n - X_p| \ge \epsilon\right) = 0.$$

Proof:

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

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

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

$$\mathbb{I}_n := \sup_{n\le p,q<\infty} |X_p - X_q|$$

and Note that $\mathbb{I}_n \xrightarrow{a.e.} 0 \iff$ (i)

Also notice that Fatu's lemma implies

$$P(I_{nm} > \varepsilon \text{ a.a.m}) \le \liminf_m P(I_{nm} > \varepsilon)$$
$$\le \limsup_m P(I_{nm} > \varepsilon)$$
$$\le P(I_{nm} > \varepsilon \text{ i.o.m})$$

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

$$\therefore \lim_m P(I_{nm} > \varepsilon) = P(I_{nm} > \varepsilon \text{ i.o.m})$$
$$= P(\lim_m I_{nm} > \varepsilon)$$

again by monotonicity of I_{nm} w.r.t m

$$= P(I_{n\infty} > \varepsilon)$$

$$\therefore \lim_n \lim_m P(I_{nm} > \varepsilon)$$
$$= \lim_n P(I_{n\infty} > \varepsilon)$$
$$= P(\lim \downarrow \{I_{n\infty} > \varepsilon\}), \text{ since } I_{n\infty} \downarrow \text{ as } n \uparrow$$

(*) $$= P(I_{n\infty} > \varepsilon \text{ i.o.n}), \text{ since:}$$

$$(\lim \downarrow \{I_{n\infty} > \varepsilon\})^c = \lim \uparrow \{I_{n\infty} \le \varepsilon\} = \bigcup_n \{I_{n\infty} \le \varepsilon\} \overset{\leftarrow I_{n\infty}\downarrow}{=} \{I_{n\infty} \le \varepsilon \text{ a.a.n}\}$$

To finish notice

(i) $\iff \mathbb{I}_n \xrightarrow{a.e.} 0$

$\iff I_{n\infty} \xrightarrow{a.e.} 0$ since $I_{n\infty} \le \mathbb{I}_n \le 2 I_{n\infty}$

$\iff \forall \varepsilon > 0 \; P(I_{n\infty} > \varepsilon \text{ i.o.n}) = 0$

$\iff \forall \varepsilon > 0 \; \lim_n \lim_m P(I_{nm} > \varepsilon)$

$\iff \forall \varepsilon > 0 \; \lim_n \lim_m P(I_{nm} \ge \varepsilon)$

QED

Theorem: (sum var $< \infty \Rightarrow$ sum $< \infty$)

Let $X_1, X_2, \dots$ be independent r.v.s s.t. $EX_n = 0 \; \forall n$. Then

$$\sum_{p=1}^{\infty} EX_p^2 < \infty \Rightarrow P(\underbrace{\sum_{p=1}^{\infty} X_p < \infty}) = 1$$

means the limit exist & is finite.

Proof:

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

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

$$P(\max_{n \le p \le m} |S_n - S_p| \ge \varepsilon)$$
$$= P(\max_{n < p \le m} |X_{n+1} + \dots + X_p| \ge \varepsilon)$$
$$\le \frac{1}{\varepsilon^2} \sum_{p=n+1}^{m} E(X_p^2)$$

requires mean ~~zero~~, indep r.v.s with finite variance.

Since $\lim_n \lim_m \text{RHS} \to 0 \; \forall \varepsilon > 0$, the Cauchy criterion lemma gives that S_n converges almost surely to a finite limite.

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

QED.

13

Theorem: (Kolmogorov's Three series)

Let $X_1, X_2, \ldots$ 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 $1 \iff$ each of the following hold

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

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

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

Proof:

In Lecture 9 we showed that Kolmogorov's 0-1 law established that

$$P\Big(\underbrace{\sum_{p=1}^{\infty} X_p < \infty}_{\text{Tail event.}}\Big) = 0 \text{ or } 1$$

($\Leftarrow$) Suppose (i) (ii) & (iii) hold.

By (iii) $\sum_{p=1}^{\infty} E\Big(\underbrace{X_p I_{|X_p| \le \delta} - E(X_p I_{|X_p| \le \delta})}_{\text{mean 0 indep r.v.s}}\Big)^2 < \infty$

$\therefore$ The previous theorem implies

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

By (ii) we have

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

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

$$\therefore \quad P\left(X_p I_{|X_p| \le \delta} = X_p \text{ a.a.}p\right) = 1$$

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

14

($\Longrightarrow$)

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

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

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

$$Y_{in} := \frac{Y_i - E(Y_i)}{\sum_{j=1}^{n} var(Y_j)}$$

bdd by a constant $\forall i$

$\longrightarrow \infty$ as $n \to \infty$

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

$$\sum_{i=1}^{n} E\left(Y_{in}^2 I_{|Y_{in}| \ge \varepsilon}\right) \longrightarrow 0 \quad \text{as } n \to \infty$$

since $I_{|Y_{in}| \ge \varepsilon} \overset{a.e}{=} 0 \ \forall i = 1, \ldots, n$ when n is suff large

$\therefore$ Lindeberg-Feller applies and gives

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

Also notice that

$\sum_{i=1}^{\infty} Y_i < \infty$ w.p. 1 since $\sum_{i=1}^{\infty} X_i < \infty$

$$\frac{\sum_{i=1}^{n} Y_i}{\sum_{j=1}^{n} var(Y_j)} \xrightarrow{a.e.} 0$$

goes to zero as $n \to \infty$

By Slutsky we have

$$\underbrace{\sum_{i=1}^n Y_{in} - \frac{\sum_{i=1}^n Y_i}{\sum_{j=1}^n var(Y_j)}}_{= - \frac{\sum_{i=1}^n E(Y_i)}{\sum_{j=1}^n var(Y_j)}} \xrightarrow{\mathcal{D}} N(0,1)$$

which is not random

$\therefore$ contradiction

$\therefore$ (iii) holds as well.

Finally to show (ii) Notice that the previous Thm

$$(iii) \Longrightarrow \lim_n \sum_{p=1}^n (Y_p - E(Y_p)) \text{ exists \& is finite}$$

$$\Longrightarrow \lim_n \sum_{p=1}^n Y_p \text{ exists \& is finite}$$

since $\sum_{p=1}^\infty X_p < \infty \Rightarrow \sum_{p=1}^\infty Y_p < \infty$

QED

Application to series with coin flip signs

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

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

$\sum_{p=1}^\infty \frac{1}{p} = \infty$

- $\sum_{p=1}^\infty \frac{(-1)^p}{p} < \infty$ tuned cancelation
- $\sum_{p=1}^\infty \frac{R_p}{p} < \infty$ w.p. 1
  Random cancelation.
  We showed this in Lecture 7

$\sum_{p=1}^\infty \frac{1}{\sqrt{p}} = \infty$

- $\sum_{p=1}^\infty \frac{(-1)^p}{\sqrt{p}} < \infty$ tuned cancelation
- $\sum_{p=1}^\infty \frac{R_p}{\sqrt{p}}$? This was left unanswered.

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

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

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

$$\sum_{p=1}^\infty E\Big(\underbrace{\frac{R_p^2}{p}}_{\frac{1}{p}} \underbrace{I_{|R_p| \le \sqrt{p}\delta}}_{\text{always 1 when } \delta^2 \le p}\Big) < \infty$$

$\therefore$ condition (iii) does not hold

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

Poisson Law of Rare Events

In an exercise you showed

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

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

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

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

$$\left(\begin{array}{c} \sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2 \\ Z \sim N(0,\sigma^2),\ Z_i \overset{indep}{\sim} N(0,\sigma_i^2) \end{array}\right) \Longrightarrow Z \overset{\mathcal{D}}{=} Z_1 + \cdots + Z_n$$

$$\left(\begin{array}{c} \lambda = \lambda_1 + \cdots + \lambda_n \\ N \sim Poi(\lambda),\ N_i \overset{indep}{\sim} Poi(\lambda_i) \end{array}\right) \Longrightarrow N \overset{\mathcal{D}}{=} N_1 + \cdots + N_n$$

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

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

(ii) $\max_{1 \le i \le n} p_{in} \to 0$ as $n \to \infty$

Then

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

as $n \to \infty$.

This proof is pretty straight forward using characteristic functions. However one 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

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

$$\le \int_\Omega |p - q| \, d\mu \quad \text{when } dP = p\,d\mu, \; dQ = q\,d\mu$$

part of Scheffe's result

Theorem: (rates for Law of rare events).

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

$$\|\mathcal{L}(X_1 + \cdots + X_n) - \text{Poi}(p_1 + \cdots + p_n)\|_{TV} \le \sum_{i=1}^{n} p_i^2.$$

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 Ω-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

$$\|P - Q\|_{TV} = \frac{1}{2} \int_\Omega |p - q| \, d\mu$$

$$= 1 - \int_\Omega p \wedge q \, d\mu$$

$$= \inf_{(X,Y) \in \Pi(P,Q)} P^*(X \ne Y)$$

where X, Y is defined on $(\Omega^*, \mathcal{F}^*, P^*)$

with the infimum attained.

Proof:

First let

$$\bar{P} = \int_{\cdot} (p - p\wedge q)\, d\mu, \qquad \bar{Q} = \int_{\cdot} (q - p\wedge q)\, d\mu \quad \Bigg\} \text{ finite measures}$$

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

$$P(A) - Q(A) = \left(P(A) - \int_A p\wedge q\, d\mu\right) - \left(Q(A) - \int_A p\wedge q\, d\mu\right)$$

$$= \int_A (p - p\wedge q)\, d\mu - \int_A (q - p\wedge q)\, d\mu$$

$$= \bar{P}(A) - \bar{Q}(A).$$

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

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

Notice

$$\underbrace{0 - \bar{Q}(\Omega)}_{= -\left(1 - \int_\Omega p\wedge q\, d\mu\right)} \le \bar{P}(A) - \bar{Q}(A) \le \underbrace{\bar{P}(\Omega) - 0}_{= 1 - \int_\Omega p\wedge q\, d\mu}$$

$$\therefore \quad \sup_{A\in\mathcal{F}} |\bar{P}(A) - \bar{Q}(A)| \le 1 - \int_\Omega p\wedge q\, d\mu$$

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

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

$$\therefore \quad |\bar{P}(A) - \bar{Q}(A)| = |0 - \bar{Q}(\Omega)|$$

$$= 1 - \int_\Omega p\wedge q\, d\mu$$

In summary

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

$$= 1 - \int_\Omega p\wedge q\, d\mu$$

$$= \int_\Omega \frac{p+q}{2} - p\wedge q\, d\mu$$

$$= \int_\Omega \tfrac{1}{2}|p-q|\, d\mu$$

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

To finish we need to show

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

with the infimum attained.

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

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

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

$$\therefore \quad Q(A) = P^*(Y\in A) \quad \text{by def of } \Pi(P,Q)$$

$$\le P^*(X\in A) + P^*(X\ne Y)$$

$$= P(A) + P^*(X\ne Y)$$

$$\therefore \quad Q(A) - P(A) \le P^*(X\ne Y)$$

By symmetry we also have

$$P(A) - Q(A) \le \mathbb{P}^*(X \ne Y)$$

$$\therefore \quad \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \le \inf_{(X,Y) \text{ in } \Pi(P,Q)} \mathbb{P}^*(X \ne Y) \qquad (*)$$

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

sample $\theta \sim Ber\left(\int_\Omega p \wedge q \, d\mu \right)$

If $\theta = 1$: Return X, Y by

$$X \sim \frac{\int_\cdot p \wedge q \, d\mu}{\int_\Omega p \wedge q \, d\mu}$$

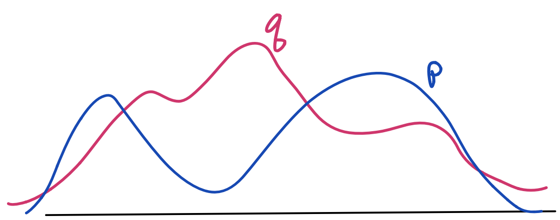
$$Y = X$$

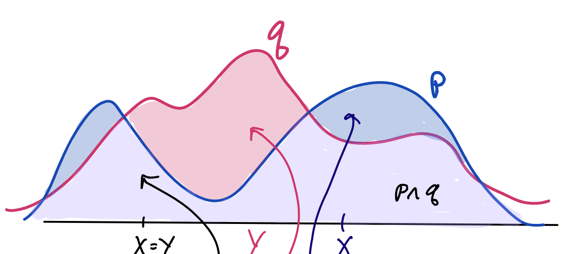
If $\theta = 0$: Return X, Y by

$$X \sim \frac{\int_\cdot (p - p \wedge q) \, d\mu}{\int_\Omega (p - p \wedge q) \, d\mu}$$

indep.

$$Y \sim \frac{\int_\cdot (q - p \wedge q) \, d\mu}{\int_\Omega (q - p \wedge q) \, d\mu}$$





If $\theta = 1$ pick random & let $X = Y$ be the x-coord

If $\theta = 0$ pick indep points X, Y from the blue, red areas separately.

In the case $0 < \int_\Omega p \wedge q \, d\mu < 1$ we don't need to worry about conditioning on Prob zero events (the other boundary cases are easy).

Therefore.

$$\mathbb{P}^*(X \in A)$$

$$= \mathbb{P}^*(X \in A \mid \theta = 1) \mathbb{P}^*(\theta = 1) + \mathbb{P}^*(X \in A \mid \theta = 0) \mathbb{P}^*(\theta = 0)$$

$$= \frac{\int_A p \wedge q \, d\mu}{\cancel{\int_\Omega p \wedge q \, d\mu}} \cancel{\int_\Omega p \wedge q \, d\mu} + \frac{\int_A (p - p \wedge q) \, d\mu}{\cancel{\int_\Omega (p - p \wedge q) \, d\mu}} \cancel{\left(1 - \int_\Omega p \wedge q \, d\mu\right)}$$

$$= \int_A p \, d\mu$$

$$= P(A)$$

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

Now all that needs to be shown is that $\mathbb{P}^*(X \ne Y) = 1 - \int_\Omega p \wedge q \, d\mu$.

We have

$$\mathbb{P}^*(X = Y) \ge \int_\Omega p \wedge q \, d\mu$$

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

Taking complements gives

$$1 - \int_\Omega p \wedge q \, d\mu \ge \mathbb{P}^*(X \ne Y)$$

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

$$= 1 - \int_\Omega p \wedge q \, d\mu$$

QED.

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

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

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

If μ denotes the counting measure on $\Omega = \{0,1,2,\ldots\}$ then

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

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

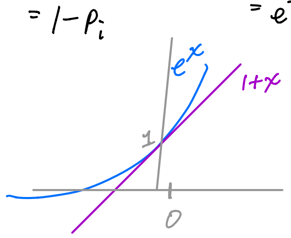
$$\frac{d\mathcal{L}(X_i)}{d\mu} \wedge \frac{d\mathcal{L}(Y_i)}{d\mu} = \min(1-p_i, e^{-p_i}) I_{\{0\}} + \min(p_i, e^{-p_i} p_i) I_{\{1\}}$$

By the TV Lemma we have

$$\|\mathcal{L}(X_i) - \mathcal{L}(Y_i)\|_{TV}$$

$$= 1 - \int_\Omega \frac{d\mathcal{L}(X_i)}{d\mu} \wedge \frac{d\mathcal{L}(Y_i)}{d\mu} \, d\mu$$

$$= 1 - \underbrace{\min(1-p_i, e^{-p_i})}_{=1-p_i} - \underbrace{\min(p_i, e^{-p_i} p_i)}_{=e^{-p_i} p_i}$$



$$= p_i (1 - e^{-p_i}) \leq p_i^2$$

Also By TV lemma $\exists X_i^* \sim X_i$ & $Y_i^* \sim Y_i$ s.t.

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

and letting

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

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

indep

indep.

we have

$$\|\mathcal{L}(X_1 + \cdots + X_n) - Poi(p_1 + \cdots + p_n)\|_{TV}$$

$$= \|\mathcal{L}(X_1^* + \cdots + X_n^*) - \mathcal{L}(Y_1^* + \cdots + Y_n^*)\|_{TV}$$

$$= \|\mathcal{L}(X^*) - \mathcal{L}(Y^*)\|_{TV}$$

$$\leq P(\underbrace{X^* \neq Y^*}_{\Longrightarrow \exists \text{ some } i \text{ s.t. } X_i^* \neq Y_i^*})$$

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

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

QED.

Notice this also shows that

$$\left.\begin{array}{l} Y_n \sim Poi(np_n) \\ np_n \to \lambda \end{array}\right\} \Longrightarrow \frac{Y_n - np_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \lambda)$$

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

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

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

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

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

$$\parallel$$

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

where $|\varepsilon_n| \le \|\mathcal{L}(X_n) - \mathcal{L}(Y_n)\|_{TV}$

$$\le np_n^2 = \underbrace{(np_n)}_{\to\lambda}\underbrace{p_n}_{\to 0}$$

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

Stein's Method

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

⋮

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

This is an old problem (first proved by Erdös & Kac I think) which we will revist later once we prove the existance 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.

⋮

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

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

⋮