LIMIT THEOREMS FOR RANDOM VECTORS

- For random vectors in X^d, SLLN holds w/o change:

$$X_n \text{ iid}, \quad \mathbb{E}X_n = \mu < \infty \quad \Rightarrow \quad \frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu \quad \text{as } n \to \infty$$

⌈Apply SLLN for each of the d coordinates⌋

WHAT ABOUT CLT ?

Def convergence in distribution for r. vectors in $\mathbb{R}^d$ is defined as

$$X_n \xrightarrow{d} X \quad \overset{def}{\Longleftrightarrow} \quad \mathbb{E}h(X_n) \to \mathbb{E}h(X) \quad \forall \text{ bdd, contin. } h: \mathbb{R} \to \mathbb{R}$$

Portmanteau Lemma generalizes to r. vectors (check!)

- Fourier analysis in $\mathbb{R}^d$ is similar: the F.T. of $f: \mathbb{R}^d \to \mathbb{R}$ is

$$\hat{f}(t) := \int_{\mathbb{R}^n} e^{-i 2\pi \langle t, x \rangle} \, dx, \qquad t \in \mathbb{R}^d$$

inner product.

- Fourier inversion formula extends to $\mathbb{R}^d$.

$\Rightarrow$ so does Levy continuity thm

$$X_n \xrightarrow{d} X \iff \mathbb{E}e^{i\langle t, X_n \rangle} \longrightarrow \underbrace{\mathbb{E}e^{i \langle t, X \rangle}}_{= \phi_X(t)} \quad \forall t \in \mathbb{R}^d \qquad (*)$$

$\phi_X(t)$ characteristic function

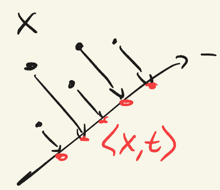
•

1D marginals determine the distribution:

THM (Cramer-Wold device)

(a) $X_n \xrightarrow{d} X \iff \langle X_n, t\rangle \to \langle X, t\rangle \quad \forall t \in \mathbb{R}^d$

(b) In particular, $X \overset{d}{=} Y \iff \langle X, t\rangle \overset{d}{=} \langle Y, t\rangle \quad \forall t \in \mathbb{R}^d$



$(\Rightarrow)$: $X_n \xrightarrow{d} X \Rightarrow \mathbb{E}h(\langle X_n, t\rangle) \to \mathbb{E}h(\langle X, t\rangle)$ $\forall$ bdd, continuous $h: \mathbb{R}^d \to \mathbb{R}$, $\forall t \in \mathbb{R}$ (since $x \mapsto h(\langle x, t\rangle)$ is bdd, continuous)

$\xRightarrow{\text{P.L}} \langle X_n, t\rangle \xrightarrow{d} \langle X, t\rangle$.

$(\Leftarrow)$: $\langle X_n, t\rangle \xrightarrow{d} \langle X, t\rangle \quad \forall t \in \mathbb{R} \xRightarrow{\text{P.L}} \mathbb{E}e^{i\langle X_n, t\rangle} \to \mathbb{E}e^{i\langle X, t\rangle} \; \forall t \xRightarrow{\text{Lévy cont. thm}} X_n \xrightarrow{d} X$. ∎

THM (Classical CLT in $\mathbb{R}^d$) Let $X_1, X_2, \dots$ be iid r.vectors in $\mathbb{R}^d$ with finite mean $\mu = \mathbb{E}X_n$ and covariance matrix $\Sigma = \mathbb{E}(X_n - \mu)(X_n - \mu)^T$.

Then
$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} Z \sim N(\mu, \Sigma).$$

WLOG $\mu = 0$. $\forall t \in \mathbb{R}^d$,

- $\langle X_n, t\rangle$ are iid r.variables with mean 0 and variance
$$\mathbb{E}\langle X_n, t\rangle^2 = \mathbb{E}\, t^T X_n X_n^T t = t^T \mathbb{E}[X_n X_n^T]\, t = t^T \Sigma t$$
and similarly
$$\langle Z, t\rangle \sim N(0, t^T \Sigma t).$$

- $\Rightarrow$ by 1D Classical CLT,
$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \langle X_k, t\rangle \xrightarrow{d} \langle Z, t\rangle$$
$$\|$$
$$\langle Z_n, t\rangle \quad \text{where} \quad Z_n = \frac{X_1 + \cdots + X_n}{t}.$$

- Cramer-Wold $\Rightarrow Z_n \xrightarrow{d} Z$. ∎

Made with Goodnotes

Cor (Jointly normal r.v's)

(a) R.v's $X_1, \dots, X_n$ are jointly normal $\iff$ $\forall$ linear combination $a_1X_1 + \dots + a_nX_n$ with fixed coeffs $a_i \in \mathbb{R}$ is normal.

(b) The joint normal distribution is uniquely determined by the means and covariances of X_i

promised on p. 65

$\ulcorner$ $(\Rightarrow)$ $\therefore$ $X = (X_1, \dots, X_n)$ is normal $\overset{\text{def}}{\iff}$ $X =$ affine transf. of $N(0, I_n)$

$\Rightarrow a_1X_1 + \dots + a_nX_n = \langle X, a\rangle =$ affine transf. of $N(0, I_n)$. $\square$

$(\Leftarrow)$ Assume: $\forall a \in \mathbb{R}^n$: $\langle X, a\rangle$ is a normal r. variable.

Its mean and variance are:

$$\begin{cases} \mathbb{E}\langle X, a\rangle = \langle \mu, a\rangle, & \text{where } \mu := \mathbb{E}X \\ \operatorname{Var}(\langle X, a\rangle) = \mathbb{E}\langle X-\mu, a\rangle^2 = \mathbb{E}\, a^T(X-\mu)(X-\mu)^T a = a^T\Sigma a & \text{where } \Sigma := \operatorname{cov}(X) \end{cases}$$

Define $Y \sim N(\mu, \Sigma) \Rightarrow \langle Y, a\rangle$ is also a normal r.v. with

$$\begin{cases} \mathbb{E}\langle Y, a\rangle = \langle \mu, a\rangle \\ \operatorname{Var}(\langle Y, a\rangle) = a^T\Sigma a \end{cases} \quad \text{(by the same argument)}$$

$\Rightarrow \langle X, a\rangle$ and $\langle Y, a\rangle$ are both normal r.var's with same mean, variance.

$\Rightarrow \langle X, a\rangle \overset{d}{=} \langle Y, a\rangle$ $\forall a$. Cramer-Wold $\Rightarrow$ $X \overset{d}{=} Y \sim N(\mu, \Sigma)$ $\lrcorner$

SKIP

Lévy's continuity thm $\Rightarrow$ distr. of X is determined by ch.f. General formula:

<u>THM</u> (<u>Inversion formula</u>) let μ be a prob. measure on $\mathbb{R}$,

$$\varphi(t) := \int_{\mathbb{R}} e^{itx} \mu(dx) \qquad (= \varphi_X(t) \text{ where } X \text{ has law } \mu)$$

Then $\forall\, a<b$:

$$\mu(a,b) + \tfrac{1}{2}\mu(\{a,b\}) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\,\varphi(t)\,dt$$

Γ RHS = limit of

$$\frac{1}{2\pi}\int_{-T}^{T}\int_{\mathbb{R}} \underbrace{\frac{e^{-ita}-e^{-itb}}{it}}_{\substack{= \int_a^b e^{-itx}dx \;\Rightarrow\; |\cdot| \le b-a \\ \Rightarrow \text{integrand is bounded}}} \cdot e^{itx}\mu(dx)\,dt \overset{\text{Fubini}}{=} \frac{1}{2\pi}\int_{\mathbb{R}}\int_{-T}^{T} \underbrace{\frac{e^{-it(x-a)}-e^{-it(x-b)}}{it}}_{\parallel}\,dt\,\mu(dx) =$$

$$\underbrace{\frac{\cos(t(x-a))}{it}}_{\text{odd} \Rightarrow \int_{-T}^{T}=0} + \underbrace{\frac{\sin(t(x-a))}{t}}_{\text{even}} - \underbrace{\frac{\cos(t(x-b))}{it}}_{\text{odd} \Rightarrow \int_{-T}^{T}=0} - \underbrace{\frac{\sin(t(x-b))}{t}}_{\text{even}}$$

... **COMPLETE THE PROOF**

Moreover, if $\int_{\mathbb{R}} |\varphi(t)|\,dt < \infty$ then X has bounded, continuous density

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi(t)\,dt. \quad \ldots$$

$\Rightarrow$ Cor on random harmonic series (Example (b) p. 111)

STEIN'S METHOD

- Recall "Stein's identity" [HW 1]:

$$\mathbb{E} f'(Z) = \mathbb{E} Z f(Z) \qquad \text{if } Z \sim N(0,1)$$

- Stein's identity characterizes $N(0,1)$ distribution.

  Moreover, a stability result holds:

  "if $\mathbb{E} f'(W) \approx \mathbb{E} W f(W)$ $\forall$ bounded, smooth function $f: \mathbb{R} \to \mathbb{R}$

  then $W \approx Z \sim N(0,1)$ in distribution"

  measured in Wasserstein metric

$$W_1(W, Z) = \sup_{\|h\|_{Lip} \le 1} |\mathbb{E} h(W) - \mathbb{E} h(Z)|$$

by approximation, equivalent to $\|h'\|_\infty \le 1$

Precisely:

"Stein's Lemma" let $Z \sim N(0,1)$ and let W be $\forall$ r.v. Then

$$W_1(W, Z) \le \sup_{f \in \mathcal{F}} |\mathbb{E} f'(W) - \mathbb{E} W f(W)|$$

where $\mathcal{F} = \{ f: \mathbb{R} \to \mathbb{R}: \|f\|_\infty \le C, \|f'\|_\infty \le C, \|f''\|_\infty \le C \}$

abs. constant

Proof Fix $\forall$ h with $\|h'\|_\infty \le 1$.

Consider Stein's differential equation

$$f'(w) - w f(w) = h(w) - \mathbb{E} h(Z), \qquad w \in \mathbb{R}$$

If $\exists$ solution $f \in \mathcal{F}$ (*)

then we can set $w := W$ and take expectation on both sides. $\Rightarrow$

$$\mathbb{E} f'(W) - \mathbb{E} W f(W) = \mathbb{E} h(W) - \mathbb{E} h(Z)$$

Take $|\cdot|$ on both sides, we complete the proof, modulo (*). $\square$

Proof of $(*)$:

① SOLVING STEIN'S D.E. by the method of integrating factors:

$e^{\int(-w)dw} = e^{-w^2/2}$ in our case

- Multiply both sides by $e^{-w^2/2}$ $\Rightarrow$

$$\left(e^{-w^2/2}f(w)\right)' = e^{-w^2/2}\left(h(w)-\mathbb{E}h(Z)\right)$$

$$\Rightarrow \quad e^{-w^2/2}f(w) = \int_{-\infty}^{w} e^{-x^2/2}\left(h(x)-\mathbb{E}h(Z)\right)dx + \underset{0}{\underset{\shortparallel}{C}}$$

$$\Rightarrow \quad f(w) = e^{w^2/2}\int_{-\infty}^{w} e^{-x^2/2}\left(h(x)-\mathbb{E}h(Z)\right)dx \qquad (*)$$

- Note that the total integral $\int_{-\infty}^{\infty} e^{-x^2/2}\left(h(x)-\mathbb{E}h(Z)\right) = \mathbb{E}h(\overbrace{X}^{N(0,1)})-\mathbb{E}h(Z)=0$. Thus,

$$\boxed{f(w) = -e^{w^2/2}\int_{w}^{\infty} e^{-x^2/2}\left(h(x)-\mathbb{E}h(Z)\right)dx =: \mathrm{Stein}(h)(w)} \qquad (**)$$

$(*)$ is useful for $w\le 0$, $(**)$ is useful for $w\ge 0$.

Unique up to $ce^{w^2/2}$
$\Rightarrow$ unique if we require $f(w) = o(e^{w^2/2})$ as $w\to\pm\infty$

PROPERTIES OF THE SOLUTION:

hides an absolute constant factor

② $\boxed{\|\mathrm{Stein}(h)\|_\infty \lesssim \|h'\|_\infty}$

WLOG $h(0)=0$, $\|h'\|_\infty = 1$ $\Rightarrow$ $|h(x)|\le|x|$, $|\mathbb{E}h(Z)|\le\mathbb{E}|Z|\le 1$

$$\Rightarrow \forall w>0: |f(w)| \overset{(**)}{\le} e^{w^2/2}\underbrace{\int_{w}^{\infty} e^{-x^2/2}(x+1)\,dx}_{\underbrace{\int_w^\infty e^{-x^2/2}x\,dx}_{=e^{-w^2/2}} + \underbrace{\int_w^\infty e^{-x^2/2}dx}_{\lesssim e^{-w^2/2}}} \lesssim 1.$$

$\lesssim e^{-w^2/2}$ by Gaussian tail bound ("Mills ratio" p. 28)

For $w<0$, proceed similarly but use $(*)$.

$$\int_w^\infty e^{-x^2/2}dx \le \frac{1}{w}e^{-w^2/2} \quad \forall w>0 \quad (\circ)$$

$\lesssim e^{-w^2/2}$ since LHS ≤ 1 $\forall w$

③ $\boxed{\|\mathrm{Stein}(h)'\|_\infty \lesssim \|h\|_\infty}$

WLOG $\|h\|_\infty = 1$ $\Rightarrow$ $|\mathbb{E}h(Z)|\le 1$

$$\Rightarrow \forall w>0: f'(w) \overset{(**)}{=\!=} -we^{w^2/2}\cdot\int_w^\infty e^{-x^2/2}\underbrace{\left(h(x)-\mathbb{E}h(Z)\right)}_{|\cdot|\le 2}dx + e^{w^2/2}\cdot e^{-w^2/2}\underbrace{\left(h(w)-\mathbb{E}h(Z)\right)}_{|\cdot|\le 2}$$

$$\Rightarrow |f'(w)| \le 2we^{w^2/2}\int_w^\infty e^{-x^2/2}dx + 2 \lesssim 1 \text{ by Gaussian tail bd } (\circ).$$

④ $\|\mathrm{Stein}(h)'\|_\infty \lesssim \underset{\substack{\| \\ 1\ \text{WLOG}}}{\|h'\|_\infty}$

Differentiate Stein's DE (here f = solution (**)):

$$f''(w) - f(w) - wf'(w) = h'(w)$$

$$\Rightarrow f''(w) - wf'(w) = f(w) + h'(w) =: H(w)$$

Use again the method of integrating factors. Multiply by $e^{-w^2/2} \Rightarrow$

$$\left(e^{-w^2/2} f'(w)\right)' = e^{-w^2/2} H(w)$$

$$\Rightarrow e^{-w^2/2} f'(w) = -\int_w^\infty e^{-x^2/2} H(x)\,dx + C$$

$$\Rightarrow f'(w) = -e^{w^2/2} \int_w^\infty e^{-x^2/2} \underbrace{H(x)}_{|H(x)| \le |f(w)| + |h'(w)| \lesssim 1 \text{ by ② \& assumption}}\,dx + Ce^{w^2/2}$$

$$\Rightarrow |f'(w)| \le \underbrace{e^{w^2/2} \int_w^\infty e^{-x^2/2}\,dx}_{\lesssim 1 \text{ by Mills ratio (p.2)}} + \underbrace{Ce^{w^2/2}}_{=0}$$

(If $C \ne 0$, $f'(w) \approx Ce^{w^2/2}$ as $w \to +\infty$ $\Rightarrow f(w)$ is unbounded, contradicting ②)

⑤ $\|\mathrm{Stein}(h)''\|_\infty \le \|h'\|_\infty$

As in ④, $f''(w) - wf'(w) = f(w) + h'(w) =: H(w)$. Substitute $w = Z$, take $\mathbb{E} \Rightarrow$

$\Rightarrow \mathbb{E}H(Z) = \mathbb{E}f''(Z) - \mathbb{E}Zf'(Z) = 0$ by Stein's identity for f'.

$$\Rightarrow f''(w) - wf'(w) = H(w) - \mathbb{E}H(Z)$$

i.e. $f' = \mathrm{Stein}(H)$

$$\Rightarrow \|f''\|_\infty = \|\mathrm{Stein}(H)'\|_\infty \overset{③}{\lesssim} \|H\|_\infty \overset{\Delta}{\le} \|\underset{\mathrm{Stein}(h)}{f}\|_\infty + \|h'\|_\infty \overset{②}{\lesssim} \|h'\|_\infty.$$

Stein's lemma is now completely proved.

• As an application, let's strengthen quantitative CLT (lem p. 118 for 3 times diff'ble h):

Wasserstein CLT Let $X_1, \dots, X_n$ be independent mean zero random variables with $\mathbb{E}|X_i|^3 < \infty$. Then $W = X_1 + \dots + X_n$ satisfies

$$W_1(W, Z) \le \underbrace{C}_{\text{absolute const.}} \sum_{i=1}^n \mathbb{E}|X_i|^3 \quad \text{where } Z \sim N(0, Var(W))$$

Proof WLOG $Var(W) = 1$.

By Stein's lemma (p. 1), it is enough to prove that

$$\left| \mathbb{E} W f(W) - \mathbb{E} f'(W) \right| \le C \sum_{i=1}^n \mathbb{E}|X_i|^3 \quad \text{whenever } \|f''\|_\infty \le 1.$$

$$W f(W) = \sum_i X_i f(W).$$

$$W_i \xrightarrow{X_i} W$$

Taylor approximation around $W_i := \sum_{j: j \ne i} X_j = W - X_i$:

$$f(W) = f(W_i) + (W - W_i) f'(W_i) + (W - W_i)^2 \frac{A_i^2}{2} \quad \text{where } |A_i| \le \|f''\|_\infty \le 1$$

Multiply both sides by X_i, sum over i, take $\mathbb{E}$

$$\mathbb{E} W f(W) = \sum_i \underbrace{\mathbb{E} X_i f(W_i)}_{\overset{\text{independence}}{=} 0} + \sum_i \underbrace{\mathbb{E} X_i (\overbrace{W - W_i}^{= X_i}) f'(W_i)}_{\overset{\text{independ.}}{=} \mathbb{E}[X_i^2]\, \mathbb{E} f'(W_i)} + \sum_i \mathbb{E} X_i (\overbrace{W - W_i}^{= X_i})^2 \frac{A_i^2}{2}$$

$$|f'(W_i) - f'(W)| \le |W_i - W| \cdot \|f''\|_\infty \le |X_i| \Rightarrow \mathbb{E} f'(W_i) = \mathbb{E} f'(W) + B_i \text{ where } |B_i| \le \mathbb{E}|X_i|$$

$$\Rightarrow \mathbb{E} W f(W) = \underbrace{\sum_i \mathbb{E}[X_i^2]}_{= Var(W) = 1} \mathbb{E} f'(W) + \sum_i \mathbb{E}[X_i^2] B_i + \sum_i \mathbb{E} X_i^3 \frac{A_i^2}{2}.$$

$$\overset{\Delta}{\Rightarrow} \left| \mathbb{E} W f(W) - \mathbb{E} f'(W) \right| \le \sum_i \underbrace{\mathbb{E}[X_i^2]\, \mathbb{E}|X_i|}_{\le (\mathbb{E}|X_i|^3)^{2/3} (\mathbb{E}|X_i|^3)^{1/3} = \mathbb{E}|X_i|^3} + \frac{1}{2} \sum_i \mathbb{E}|X_i|^3 \le \frac{3}{2} \sum_i \mathbb{E}|X_i|^3$$

□

Use Thm for $X_i = \frac{Y_i - \mu}{\sigma}$ ⇓

Cor If $Y_1, Y_2, \dots$ are iid r.v.'s with mean μ, variance σ, and $\mathbb{E}|Y_1|^3 =: \rho^3 < \infty$, then $S_n = Y_1 + \dots + Y_n$ satisfies

$$W_1\left(\frac{S_n - \mu n}{\sigma \sqrt{n}}, \underset{N(0,1)}{Z} \right) \le \frac{C (\rho/\sigma)^3}{\sqrt{n}} \quad \forall n \in \mathbb{N}.$$

Remark (Berry–Esseen CLT) A modification of this argument yields the same bound in Kolmogorov metric $d_K(W, Z) = \sup_{x \in \mathbb{R}} |P\{W \le x\} - P\{Z \le x\}|$ (use $h :=$ smoothing of $1_{(-\infty, x]}$)

[E. Bolthausen, An estimate of the remainder in a combinatorial CLT '1984]

CLT FOR DEPENDENT R.V'S

CLT Let $X_1, \dots, X_n$ be mean zero random variables such that each X_i may depend on at most d r.v's in $\{X_1, \dots, X_n\}$. Let $W = X_1 + \dots + X_n$ have $\mathrm{Var}(W) = 1$. Then

$$W_1(W, Z) \le C d^2 \sum_{i=1}^{n} \mathbb{E}|X_i|^3$$

where $Z \sim N(0,1)$.

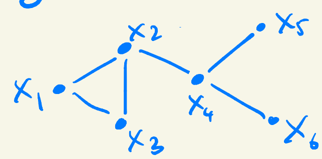
Formally: consider a graph $G = (V, E)$ with $V = \{1, \dots, n\}$. ("Dependency graph")

$\forall i \in V$, let $N_i := \{ j \in V : (i,j) \in E \}$ ("dependency neighborhood")

Assume that:

(a) $\forall i \in V$, $X_i \perp\!\!\!\perp \{X_j : j \notin N_i\}$

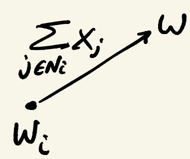
(b) Max. degree of G is $\le d$.



Proof By Stein's lemma (p.133), it is enough to bound

$$\left| \mathbb{E} W f(W) - \mathbb{E} f'(W) \right| \quad \text{for any } f \text{ satisfying } \|f'\|_\infty \le 1, \ \|f''\|_\infty \le 1.$$

- $\mathbb{E} W f(W) = \sum_i X_i f(W)$.



$\boxed{W_i := \sum_{j \notin N_i} X_j}$ Taylor approximation about W_i:

$$f(W) = f(W_i) + (W - W_i) f'(W_i) + (W - W_i) \frac{A_i^2}{2!} \quad \text{where} \quad |A_i| \le \|f''\|_\infty \le 1$$

Multiply both sides by X_i, sum over i, take $\mathbb{E}$ $\Rightarrow$

$$\mathbb{E} W f(W) = \underbrace{\sum_i \mathbb{E} X_i f(W_i)}_{\text{independence} \ = \ 0} + \underbrace{\sum_i \mathbb{E} \underbrace{X_i \overbrace{(W - W_i)}^{\sum_{j \in N_i} X_j}}_{\leftarrow \text{ independent}} \underbrace{f'(W_i)}_{\to}}_{= \ \mathrm{I}} + \underbrace{\sum_i \mathbb{E} X_i \overbrace{(W - W_i)}^{\sum_{j \in N_i} X_j}{}^2 \frac{A_i^2}{2}}_{= \ \mathrm{II}} \qquad (*)$$

• $\text{I} = \sum_i \sum_{j\in N_i} \mathbb{E}[X_i X_j] \cdot \mathbb{E} f'(W_i)$

$|f'(W_i) - f'(W)| \le |W_i - W| \cdot \|f''\|_\infty \le |X_i| \Rightarrow \mathbb{E} f'(W_i) = \mathbb{E} f'(W) + B_i$ where $|B_i| \le \mathbb{E}|X_i|$

$$\Rightarrow \text{I} = \underbrace{\sum_i \sum_{j\in N_i} \mathbb{E}[X_i X_j]}_{\sum_{i,j=1}^n \mathbb{E}[X_i X_j] = \mathbb{E}\left(\sum_{i=1}^n X_i\right)^2 = Var(W) = 1} \cdot \mathbb{E} f'(W) + \sum_i \sum_{j\in N_i} \mathbb{E}[X_i X_j] \cdot B_i$$

$a^2 b \le a^3 + b^3$ (Young's inequality)

$$\Rightarrow |\text{I} - \mathbb{E} f'(W)| \le \sum_i \sum_{j\in N_i} \underbrace{\mathbb{E}|X_i X_j| \cdot \mathbb{E}|X_i|}_{\|X_i\|_2 \cdot \|X_j\|_2 \cdot \|X_i\|_1}$$

CS; $L_p - L_q$

$\|X_i\|_2 \cdot \|X_j\|_2 \cdot \|X_i\|_1 \le \|X_i\|_3^2 \cdot \|X_j\|_3 \le \|X_i\|_3^3 + \|X_j\|_3^3$

$$\le 2d \cdot \sum_{i=1}^n \|X_i\|^3 = 2d \sum_{i=1}^n \mathbb{E}|X_i|^3 .$$

• $|\text{II}| \underset{\Delta \,\&\, |A_i| \le 1}{\le} \sum_i \mathbb{E}|X_i (W - W_i)^2| = \sum_i \mathbb{E}\Big|X_i \underbrace{\Big(\sum_{j\in N_i} X_j\Big)^2}_{\text{expand}}\Big| \underset{\Delta}{\le} \sum_i \sum_{j,k\in N_i} \underbrace{\mathbb{E}|X_i X_j X_k|}_{\frac{\mathbb{E}|X_i|^3 + \mathbb{E}|X_j|^3 + \mathbb{E}|X_k|^3}{3}}$

am-gm

$$\le d^2 \sum_i \mathbb{E}|X_i|^3 .$$

• Put into $(*)$ $\Rightarrow |\mathbb{E} W f(W) - \mathbb{E} f'(W)| \le |\text{I} - \mathbb{E} f'(W)| + |\text{II}| \le 2d^2 \sum_i \mathbb{E}|X_i|^3$. QED

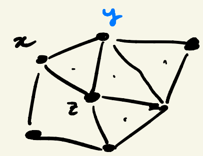
• <u>**Example**</u>: $W = X_1 X_2 + X_2 X_3 + \cdots + X_{n-1} X_n$ where all X_i are mean 0 indep.

$d = 3 \Rightarrow W$ satisfies CLT. More generally, time series.

HW: $W = \sum_{i,j=1}^n X_i X_j$ is NOT always approx. normal.

Application: Triangles in Random Graphs

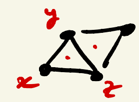
- Consider an Erdös–Rényi random graph $G \sim G(n,p)$ with $p = \text{const}$, $n \to \infty$.



$\Delta = 4$ here

$\Delta := \#(\text{triangles in } G)$ satisfies CLT?

$= \sum_{xyz} \mathbb{1}_{\Delta_{xyz}}$ where $\Delta_{xyz} = \{\text{triple } xyz \text{ is a triangle}\}$



sum over all $\binom{n}{3}$ triples

- $\boxed{d \lesssim n}$ (for a fixed triple xyz, there are $\lesssim n^2$ other triples that share a pair with xyz)

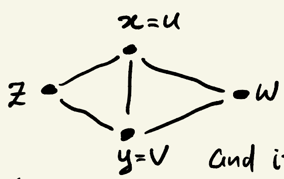
- $\mathbb{P}(\Delta_{xyz}) = p^3 \Rightarrow \boxed{\mathbb{E}\Delta = \binom{n}{3} p^3 \asymp n^3}$

- $\mathrm{Var}(\Delta) = \sum_{xyz,\, uvw} \underbrace{\mathrm{Cov}(\mathbb{1}_{\Delta_{xyz}}, \mathbb{1}_{\Delta_{uvw}})}_{\overset{?}{=}}$

(a) All covariances are ≥ 0 [ijk being a Δ may only increase the prob. of lmn being a Δ]

(b) There are $\gtrsim n^4$ covariances that are $\gtrsim 1$

[For such two triples



$\mathrm{cov}(\mathbb{1}_{\Delta_{xyz}}, \mathbb{1}_{\Delta_{uvw}}) > 0$, and its value only depends on $p = \text{const} \Rightarrow \gtrsim 1$

There are $\binom{n}{4} \asymp n^4$ such pairs of triples]

(a) & (b) $\Rightarrow \boxed{\mathrm{Var}(\Delta) \gtrsim n^4}$
$\overset{\shortparallel}{\sigma^2}$

- Use CLT for $W := \frac{\Delta - \mathbb{E}\Delta}{\sqrt{\mathrm{Var}\Delta}} = \sum_{xyz} \frac{\overbrace{\mathbb{1}_{\Delta_{xyz}}}^{\mathrm{Ber}(p^3)} - p^3}{\sigma} \Rightarrow$

$$W_1(W, Z) \lesssim \frac{d^2}{\sigma^3} \underbrace{\sum_{xyz}}_{\lesssim n^3 \text{ terms}} \underbrace{\mathbb{E}\left|\mathbb{1}_{\Delta_{xyz}} - p^3\right|^3}_{\leq 1} \lesssim \frac{n^2}{n^6} \cdot n^3 \lesssim \frac{1}{n}.$$

$\Rightarrow$ **THM** The # of triangles Δ in Erdös–Rényi graph $G(n,p)$ is approximately normal:

$$d\left(\frac{\Delta - \mathbb{E}\Delta}{\sqrt{\mathrm{Var}\Delta}}, Z\right) \leq \frac{C(p)}{n} \quad \text{where } Z \sim N(0,1)$$

Remark A finer analysis $\Rightarrow$ CLT holds iff $np^3 \to \infty$.

APPLICATION: A Probabilistic Proof of Stirling's Approximation

Thm (Stirling's Approximation) $\quad n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1+o(1))$ as $n\to\infty$

Proof following [Nils Lid Hjort, Emil Aas Stoltenberg, Probability Proofs of Stirling '2024]

$Y_n \sim \text{Poisson}(n)$ can be expressed as a sum of n iid Poisson(1)

$\Rightarrow \quad W := \frac{Y_n - n}{\sqrt{n}}$ is a sum of n iid rv's with mean 0, var 1, third moment $= O(1)$ (check!)

$\Rightarrow$ By Wasserstein CLT,

$$W_1(W, Z) \to 0 \quad \text{where } Z \sim N(0,1)$$

The function $h(x) = \max(x, 0)$ is 1-Lipschitz $\Rightarrow$

$$\Big| \mathbb{E}h(W) - \underbrace{\mathbb{E}h(Z)}_{=\frac{1}{\sqrt{2\pi}}\int_0^\infty x e^{-x^2/2}\,dx \overset{y=x^2/2}{=\!=} \frac{1}{\sqrt{2\pi}}} \Big| \to 0 \quad \text{as } n\to\infty$$

$$\mathbb{E}h(W) = \sum_{k=n}^\infty \frac{k-n}{\sqrt{n}} \overbrace{p(k)}^{=e^{-n}n^k/k!}$$

$$= \frac{1}{\sqrt{n}}\Big(np(n) - \underbrace{np(n) + (n+1)p(n+1)}_{=} - \underbrace{np(n+1) + (n+2)p(n+2)}_{=} - \underbrace{np(n+2) + \cdots}_{=} \Big)$$

check! $\qquad$ check!

$$= \frac{np(n)}{\sqrt{n}} = \frac{\sqrt{n}\, e^{-n} n^n}{n!} \longrightarrow \frac{1}{\sqrt{2\pi}} \quad \Rightarrow \text{Q.E.D.}$$

CONDITIONAL EXPECTATION

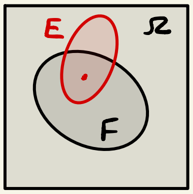
- Fix a prob. space $(\Omega, \Sigma, \mathbb{P})$ throughout.

1. Conditioning on an event

Def Let E, F be events. The conditional probability of E given F is

$$\mathbb{P}(E|F) := \frac{\mathbb{P}(E\cap F)}{\mathbb{P}(F)} \qquad \text{as long as } \mathbb{P}(F) \neq 0$$

- Interpetation:



$(F, \underbrace{\Sigma \cap F}_{= \{A\cap F \,:\, A \in \Sigma\}})$ is a new prob. space.

$E \mapsto \mathbb{P}(E|F)$ is a new prob. measure on $(F, \Sigma\cap F)$ (check!)

- Ex:

	Cancer	No
Smoker	8	32
No	16	304

$$\mathbb{P}(\text{Cancer} \mid \text{Smoker}) = \frac{8}{8+32} = 0.2$$

$$\mathbb{P}(\text{Cancer} \mid \text{No smoker}) = \frac{16}{16+304} = 0.05$$

- Relation with independence:

$$E \perp\!\!\!\perp F \iff \mathbb{P}(E|F) = \mathbb{P}(E) \qquad \text{if } \mathbb{P}(F) \neq 0.$$

- **Ex** Suppose you know that your friend has 2 children.
You saw one of them, and it was a girl.
What is the probability that the other child is also a girl?

$\Omega = \{GG, GB, BG, BB\}$ (older first), $\mathbb{P}$ = uniform.

F = "at least one child is a girl" $= \{GG, GB, BG\}$, E = "both children are girls" $= \{GG\}$

$$\Rightarrow \mathbb{P}(E|F) = \frac{1/4}{3/4} = \frac{1}{3}.$$

?! Why not ½?

Prop (Law of Total Probability) If $\Omega = F_1 \cup \cdots \cup F_n$ then

n could be ∞

$$P(E) = \sum_i P(E|F_i)\cdot P(F_i) \qquad \forall E \in \Sigma$$

$$E = \bigsqcup_i (E\cap F_i) \Rightarrow P(E) = \sum P(E\cap F_i)$$

Computing probabilities by conditioning

Ex Two players take turns flipping a coin. The first player to obtain a head wins. What is the prob. that $\underbrace{\text{the player who starts wins}}_{E}$?

Condition on the first flip

player 1 wins — player 1 flips H — or T

$$P(E) = \underbrace{P(E|H)}_{1}\underbrace{P(H)}_{\frac{1}{2}} + \underbrace{P(E|T)}_{}\underbrace{P(T)}_{\frac{1}{2}}$$

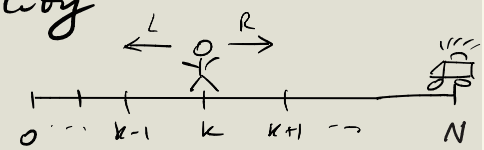
|| game resets, player 2 starts

$P(\text{the player who starts loses}) = 1 - P(E)$

$\Rightarrow P(E) = \frac{1}{2} + (1-P(E))\cdot\frac{1}{2}$. Solving gives $P(E) = 2/3$

Ex (Gambler's ruin) Consider a simple random walk starting at $k \in [0, N]$. What is the probability of reaching N before reaching 0? bankruptcy

$=: E_k$



Condition on 1st step, L or R:

$$P(E_k) = P(E_k \mid L)\, P(L) + P(E_k \mid R)\, P(R)$$

$$= P(E_{k-1}) \cdot \frac{1}{2} + P(E_{k+1}) \cdot \frac{1}{2}$$

walk "resets" at $k-1$

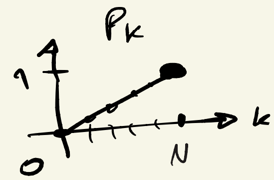
walk "resets" at $k+1$

Denoting $p_k = P(E_k)$, we obtain

$$\begin{cases} p_k = \frac{1}{2}(p_{k-1} + p_{k+1}), & k = 1, \dots, N-1 \\ p_0 = 0; \quad p_N = 1 \end{cases}$$

$N+1$ linear equations in $N+1$ unknowns. Solve →

$$\boxed{p_k = \frac{k}{N}}$$



Ex Secretary problem, a.k.a. Best prize problem

- We are presented with n prizes, in sequence.
- Upon seeing a prize, we must accept it (and end the game) or reject it (and move to the next prize). No going back.
- The only info we have at $\forall$ time is how the current prize compares to the prizes already seen.
- We want to $\underbrace{\boxed{\text{pick the best prize.}}}_{E}$ What shall we do?

- Strategy: | reject the first k prizes;
  | accept the first one that is better than all those k.

Let's compute $P(E)$ and optimize k.

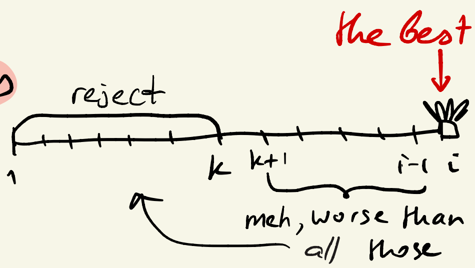
- Condition on the position of best prize:

$$\boxed{B_i = \text{"i-th prize is the best"}}.$$

$$\text{L.T.P.} \Rightarrow P(E) = \sum_{i=1}^{n} \underbrace{P(E|B_i)}_{=?} \underbrace{P(B_i)}_{=1/n}$$

- $\forall\ i \le k$ $\quad P(E|B_i) = 0$ (we reject the first k prizes)
- $\forall\ i > k$: assume B_i occurs, i.e. i-th prize is the best.
  We pick it iff each prize $k+1, \dots, i-1$ is worse than some of the first k prizes (Otherwise we lose it)

$\Rightarrow$ E occurs $\Leftrightarrow$ the best prize among the first $i-1$ prizes is among the first k prizes.



This happens with prob. $\dfrac{k}{i-1}$

$$\Rightarrow \quad P(E|B_i) = \frac{k}{i-1}$$

$$\Rightarrow P(E) = \sum_{i=k+1}^{n} \frac{k}{i-1} \cdot \frac{1}{n} = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i} \simeq \frac{k}{n} \int_k^n \frac{dx}{x} = \frac{k}{n} \ln \frac{n}{k} = -\lambda \ln \lambda \quad \text{where } \lambda = \frac{k}{n}$$

Maximize $\Rightarrow \lambda = 1/e$, $P(E) \simeq 1/e$ $\Rightarrow$

optional stopping

Ans | Strategy = reject the first $\frac{n}{e}$ prizes, accept the first better than all rejected.
| Probability to pick the best prize $= 1/e \simeq 0.37$. Regardless of n!