Kom 505E Week 10

G.Ü.

Studying:

Random vector : $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}_{N\times 1} \Rightarrow P_{\underline{X}}(\underline{x})$

$$P : \mathbb{R}^N \to \mathbb{R}$$

Conditional pdfs for Random vectors:

$$P_{X_N | X_1, \dots, X_{N-1}}(x_N | x_1, \dots, x_{N-1}) = \frac{P_{X_1 \dots X_N}(x_1, \dots, x_N)}{P_{X_1 \dots X_{N-1}}(x_1, \dots, x_{N-1})}$$

Note: many different conditional pdf's can be defined.

$$P_{X_1, X_2, \dots, X_N}(x_1, \dots, x_N) = P(x_N | x_1 \dots x_{N-1}) \dots P_{X_3|X_1 X_2}(x_3|x_1, x_2) \, P_{X_2|X_1}(x_2|x_1) \, P_{X_1}(x_1)$$

Applying the same rule again & again, we get the chain rule.

- Markov Property: When these r.v.s satisfy Markov property:

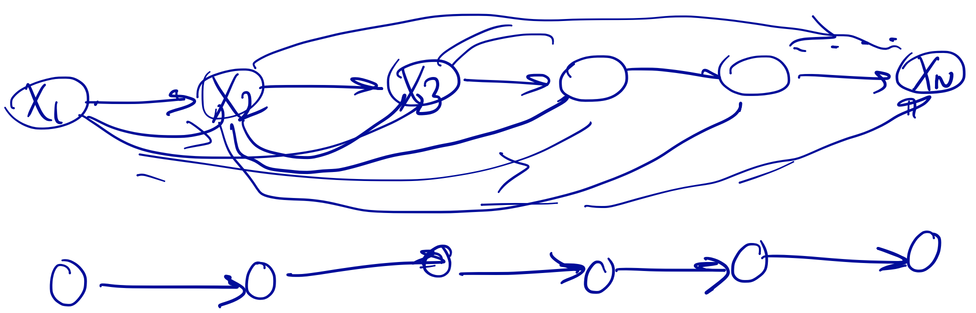
$$P_{X_N|X_{N-1}, \dots, X_1}(x_N | x_{N-1} \dots x_1) = P(x_N | x_{N-1}),$$

for $N = 2, 3, \dots$

$N = 1, 2, \dots$

↳ Chain rule becomes:

$$P_{X_1, X_2, \dots X_N}(\quad) = P_{X_1} P_{X_2|X_1} P_{X_3|X_2} \dots P_{X_N|X_{N-1}}.$$



w/o markov

w/ markov property

Expected Values:

$$E_{\underline{X}}[\underline{X}] = \begin{bmatrix} E_{X_1}[X_1] \\ \vdots \\ E_{X_N}[X_N] \end{bmatrix}_{N\times 1}$$

$g: \mathbb{R}^N \to \mathbb{R}$

$$E_{\underline{X}}[g(\underline{X})] = \int_{\underline{X}} g(\underline{X})\, p_{\underline{X}}(\underline{X})\, d\underline{X}$$

$$- E_{\underline{X}}\Big[\underbrace{\sum_{i=1}^{N} a_i X_i}\Big] = \sum_{i=1}^{N} a_i E_{X_i}[X_i] \quad , \quad a_i \in \mathbb{R}, \quad \underline{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}$$

$$E_{\underline{X}}[\underline{a}^T \underline{X}] = \underline{a}^T E_{\underline{X}}[\underline{X}]$$

$$- Var\Big(\sum_{i=1}^{N} X_i\Big) = E_{\underline{X}}\Big[\Big(\underbrace{\sum X_i - \sum \underbrace{E[X_i]}_{\mu_{x_i}}}\Big)^2\Big]$$

$$= \underbrace{E_{\underline{X}}\Big[\Big(\sum_i \underbrace{(X_i - \mu_{x_i})}_{\triangleq\, u_i}\Big)^2\Big]}$$

$$= E_{\underline{X}}\Big[\sum_i \sum_j u_i u_j\Big] \qquad = E_{\underline{X}}\Big[\Big(\sum_i u_i\Big)^2\Big]$$

$$= \sum_i \sum_j E_{\underline{X}}[u_i u_j]$$

$$E_{\underline{X}}[(X_i - \mu_{x_i})(X_j - \mu_{x_j})]$$

$$= Cov(X_i, X_j)$$

$$Var\Big(\sum X_i\Big) = \sum_i \sum_j Cov(X_i, X_j)$$

$$= \sum_{i=1}^{N} Var(X_i) + \sum_i \sum_{j\,(i\neq j)} Cov(X_i, X_j)$$

Note: 2nd term disappears when X_i are indep.

$$\mathrm{Var}\left(\sum X_i\right) = \underbrace{[1\ 1\ \dots\ 1]_{1\times N}}_{\underline{\underline{1}}^T} \underbrace{\begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \dots & \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & & \\ & & \ddots & \\ \mathrm{Cov}(X_N,X_1) & - & \dots & \mathrm{Var}(X_N) \end{bmatrix}_{N\times N}}_{} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$\underline{\underline{C}}_X$: useful for understanding random vectors. $\Leftarrow$ $\underline{\underline{C}}_{\underline{X}}$: Covariance matrix of the r.v. $\underline{X}$.

(Note: we need the full (joint) pdf $\to p_{\underline{X}}(\underline{x})$ to know everything).

but knowing $\underline{\mu}_X$ & $\underline{\underline{C}}_X$ tells a lot.

$$\underline{\underline{C}}_X = E_{\underline{X}}\left[(\underline{X}-\underline{\mu}_X)(\underline{X}-\underline{\mu}_X)^T\right]$$

---

more generally: $\underline{Y} = \underline{\underline{A}}\,\underline{X}$; A: $M\times N$ matrix

$$\underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix}$$

$\underline{Q}$: Given $\underline{\mu}_X$ & $\underline{\underline{C}}_X$, $\underline{\mu}_Y$ & $\underline{\underline{C}}_Y$ = ?

$$Y_1 = a_1^T \underline{X} \qquad \underline{\underline{A}} = \begin{bmatrix} \underline{a_1}^T \\ \vdots \\ \underline{a_M}^T \end{bmatrix}$$

$$\vdots$$

$$Y_M = a_M^T \underline{X}$$

$$\underline{\mu}_Y = E_{\underline{Y}}[\underline{Y}] = E_{\underline{X}}[\underline{\underline{A}}\,\underline{X}] = \underline{\underline{A}}\,E_{\underline{X}}[\underline{X}] = \underline{\underline{A}}\,\underline{\mu}_X$$

$$\underline{\underline{C}}_Y = E\left[(\underline{Y}-\underline{\mu}_Y)(\underline{Y}-\underline{\mu}_Y)^T\right] = E_{\underline{X}}\left[(\underline{\underline{A}}(\underline{X}-\underline{\mu}_X))(\underline{\underline{A}}(\underline{X}-\underline{\mu}_X))^T\right]$$

$$= E_{\underline{X}}\left[\underline{\underline{A}}(\underline{X}-\underline{\mu}_X)(\underline{X}-\underline{\mu}_X)^T \underline{\underline{A}}^T\right]$$

$$\boxed{\underline{\underline{C}}_Y = \underline{\underline{A}}\,\underline{\underline{C}}_X\,\underline{\underline{A}}^T}$$

ex: $\underline{y} = \sum_{i=1}^{N} x_i = \underbrace{[1 \ldots 1]^T}_{\underline{A} = \underline{1}^T} \underline{x}$

$$\underline{\underline{C}}_y = \underline{1}^T \underline{\underline{C}}_x \underline{1}$$

Property of $\boxed{C_x}$: is positive semi-definite (p.s.d.) matrix. (also symmetric).

pf.: Define $y = \underline{a}^T \underline{x} \Rightarrow var(\underline{y}) = \underline{a}^T \underline{\underline{C}}_x \underline{a} \geq 0 \quad \forall \underline{a}$

Definition of p.s.d. matrix

$\Rightarrow \underline{\underline{C}}_x$ is a p.s.d. matrix.

Ex: Sample mean of i.i.d. r.v.s (independent & identically distributed)

$X_1, X_2, \ldots, X_N$ (i.i.d) $\left\{ \begin{array}{l} E_{X_i}[X_i] = \mu \quad \forall i \\ Var(X_i) = \sigma^2 \end{array} \right.$

Transformation: $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$, $E[\bar{X}] = ?$, $Var(\bar{X}) = ?$

$$E[\bar{X}] = E\left[\frac{1}{N} \sum_i X_i\right] = \frac{1}{N} \underbrace{\sum_i E[X_i]}_{N.\mu} = \mu$$

$$Var(\bar{X}) = Var\left(\frac{1}{N} \sum_i X_i\right) = \frac{1}{N^2} \underbrace{\sum_i \underbrace{Var(X_i)}_{\sigma^2}}_{N.\sigma^2} = \frac{\sigma^2}{N}$$

As $N \to \infty$, $Var(\bar{X}) \to 0$, $E[\bar{X}] \to \mu$

Decorrelation of a random vector $\underline{X}$:

- $\underline{X}$ is an uncorrelated random vector if $\underline{\underline{C}}_x$ is diagonal $\equiv \mathrm{Cov}(X_i, X_j) = 0$, $i \neq j$.

- If X is not uncorrelated;

$\underline{Y} = \underline{\underline{W}}\,\underline{X}$; can be made uncorrelated

if we choose $\boxed{\underline{\underline{W}} = \underline{\underline{V}}^T}$; $\boxed{\underline{V}:\ \text{eigenvector matrix}}$

$$\underline{\underline{C}}_x = \underline{\underline{V}}\,\underbrace{\underline{\underline{\Lambda}}}_{\text{diagonal matrix}}\,\underline{\underline{V}}^T \Rightarrow \underline{\underline{C}}_x \underline{\underline{V}} = V \Lambda \underbrace{V^T V}_{I}$$

$$\underline{\underline{V}}^T \underline{\underline{C}}_x \underline{\underline{V}} = \underbrace{V^T V}_{I}\Lambda$$

$$\underline{\underline{C}}_y = \underline{\underline{W}}\,\underline{\underline{C}}_x\,\underline{\underline{W}}^T = \underbrace{\underline{\underline{V}}^T \underline{\underline{C}}_x \underline{\underline{V}}}_{\text{diagonal}} = \Lambda$$

Recall: Λ : diagonal matrix of eigenvalues.

Summary: Transform $\underline{X} : (\underline{\mu}_x, \underline{\underline{C}}_x)$ w/ $\underline{\underline{V}}^T$

to obtain $\underline{Y} = \underline{\underline{V}}^T \underline{X}$: $(\underline{\mu}_y = \underline{\underline{V}}^T \mu_x,\ \underline{\underline{C}}_y = \Lambda)$

$\underline{\underline{C}}_y = \underline{\underline{V}}^T \underline{\underline{C}}_x \underline{\underline{V}}$

Go further: Pick $\underline{\underline{W}} = \Lambda^{-1/2}\underline{\underline{V}}^T$

$$\underline{\underline{C}}_{y} = \Lambda^{-1/2}\underbrace{V^T C_x V}_{\Lambda}\Lambda^{-1/2} = I \quad : \text{whitening.}$$

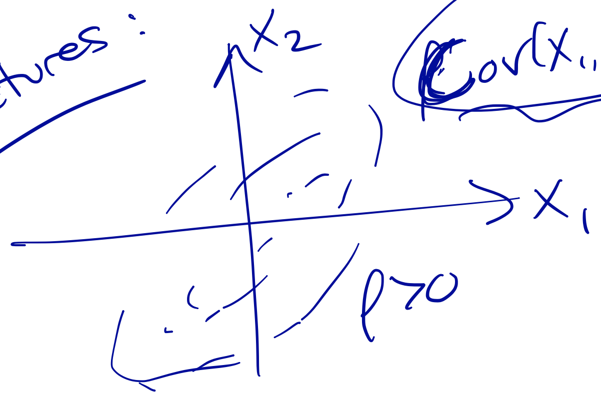
Note: possible when $\underline{\underline{C}}_x$ is p.d. (positive-definite) ie $\forall \lambda_i > 0$. (all of its eigenvalues are positive)

Standardization: Given $\underline{X}$ w/ mean $\underline{\mu}_x$, $\underline{\underline{C}}_x$ (p.d)

(★) $$\underline{U} = \underline{\underline{\Lambda}}^{-1/2}\, \underline{\underline{V}}^T (\underline{X} - \underline{\mu}_x) \Rightarrow$$ has mean $= 0$
$\underline{\mu}_u = 0$

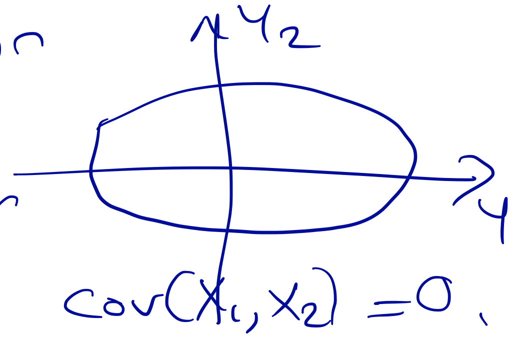
$\underline{\underline{C}}_u = \underline{\underline{I}}$.

In pictures:



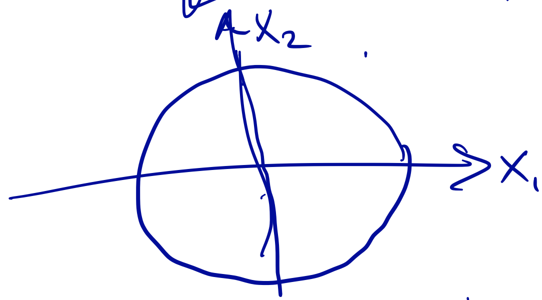
$Cov(x_1, x_2) = \rho \sqrt{VarX\, VarY} > 0$

Decorrelation $\longrightarrow$



$cov(x_1, x_2) = 0$.

whitening



Def: A standardized r.v. is defined to be one for which the mean is zero and the covariance is $\underline{\underline{I}}$ → random vectors
(for a scalar r.v. (1-D) $\equiv$ Var $= 1$)

★ Any r.v. can be standardized by subtracting the mean and for 1-D r.v. dividing the result by $\sqrt{Var(X)}$:

$$\boxed{X_s = \frac{X - \mu_x}{\sqrt{Var\,X}}}$$ for 1-D r.v.s.

For N-Dim $\underline{X}$; use (★).

9.9 Real-World Example (Image Coding)

Reading Exercise:

Image $\rightarrow$ N-dim. random vector: $\underline{X}$:

(Karhunen-Loeve Transform):

1. Xform random vector $\underline{X}$ into an uncorrelated:

$$\underline{Y} = \underline{\underline{A}}\,\underline{X} = \underline{\underline{V}}^T \underline{X}$$

2. Discard r.v.'s whose variance is small w.r.t. others ($\underline{\underline{C}}$ matrix, diagonals), by setting corresp. elements of $\underline{Y}$ to zero.

$\hat{Y} \Rightarrow$ will be stored or xmitted.

3. Xform back to $\hat{X} = \underline{\underline{A}}^{-1} \hat{Y} = \underline{\underline{V}}\,\hat{Y}$

to recover an approx to the original r.v. (at the receiver).

Chapter 15 Limit Theorems:

① LOLN : Law of Large Numbers :
② CLT : Central Limit Theorem

For r.v.s that are sum of a LARGE # i.i.d. r.v.'s

Ex: $X_i \sim$ Bernoulli, $p = \frac{1}{2}$, $p_X(k) = \begin{cases} \frac{1}{2}, & k=0 \\ \frac{1}{2}, & k=1 \end{cases}$

X_i, i.i.d.

$E[X_i] = p = \frac{1}{2}$

$Var(X_i) = p(1-p) = \frac{1}{4}$

Consider $\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i$

$$E[\bar{X}_N] = \frac{1}{N} \sum_{i=1}^{N} E[X_i] = \frac{1}{N} \cdot N \cdot \frac{1}{2} = \frac{1}{2} = p.$$

$$Var(\underbrace{\bar{X}_N}_{\frac{1}{N}\sum X_i}) = \frac{1}{N^2}\left(N \cdot Var(X_i)\right) = \frac{1}{N^2} N \cdot \frac{1}{4} = \frac{1}{4N}$$

$Var(\bar{X}_N) \to 0$ as $N \to \infty$.

$E[\bar{X}_N] \xrightarrow[N\to\infty]{} \frac{1}{2}$ (in some sense)

See Fig 15.2 for different N ;

Recall: Sum of N iid Bernoulli r.v.'s is a Binomial r.v.

$$S_N = \sum_{i=1}^{N} X_i \sim \text{Binomial}(N, p)$$

$$\bar{X}_N = S_N / N$$

gets more concentrated around the mean

Theorem (Law of Large Numbers):

more generally $X_i \to$ any pdf.

If $X_1, X_2, \ldots X_N$ are I.I.D (i.i.d) r.v.s w/ mean $E_{X_i}[X_i] = \mu$, $Var(X_i) = \sigma^2 < \infty$,

then $\lim_{N\to\infty} \bar{X}_N = E_X[X]$ (convergence in probability)

Convergence in Probability (Def)

A seq. $X_1, X_2, \ldots$ of r.v.s. converges to a number a in probability as $N \to \infty$,

$$P(|X_N - a| > \varepsilon) \xrightarrow[N\to\infty]{} 0$$

$$\equiv P(|X_N - a| \leq \varepsilon) \xrightarrow[N\to\infty]{} 1$$

pf: Given an $\varepsilon > 0$; $P(|\bar{X}_N - \mu_x| > \varepsilon) \leq \frac{Var X}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2}$

Use Chebyshev Ineq.

$$\lim_{N\to\infty} P(|\bar{X}_N - \mu_x| > \varepsilon) \leq \lim_{N\to\infty} \frac{\sigma^2}{N\varepsilon^2} = 0$$

Since a prob must be ≥ 0.

$$\lim_{N\to\infty} P[|\bar{X}_N - \mu_x| > \varepsilon] = 0$$

$$\equiv \bar{X}_N \xrightarrow[N\to\infty]{} \mu_x \text{ in probability}$$

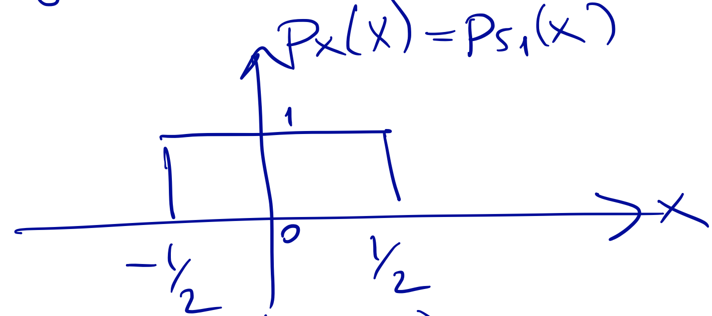
$\boxed{\text{LOLN : sample mean converges (in prob) to the true mean.}}$

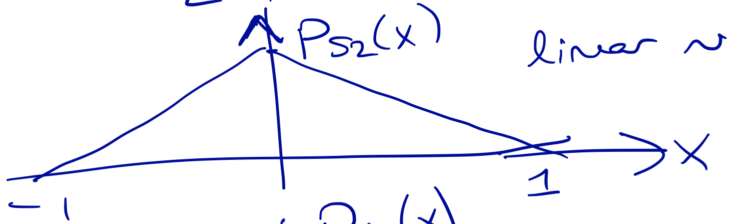
CLT: By LOLN, pdf of the sample mean r.v. width decreased. (variance $\downarrow$)

CLT $\rightarrow$ what happens to the pdf?

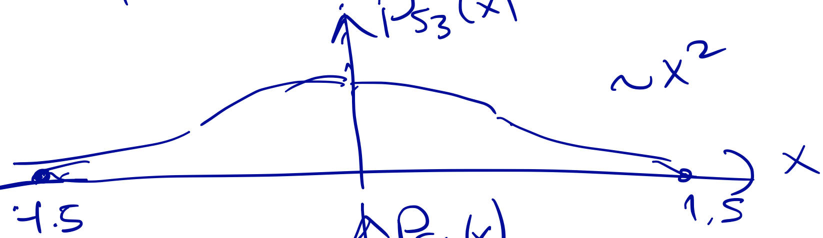
Now consider: $S_N = \sum_{i=1}^{N} X_i$ sum of N IID r.v.s

eg. let $X_i \sim U[-\frac{1}{2}, \frac{1}{2}]$ i.i.d.



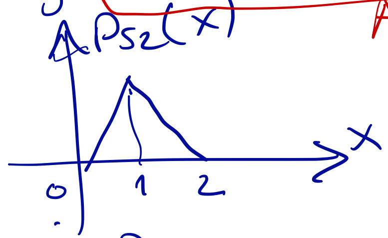


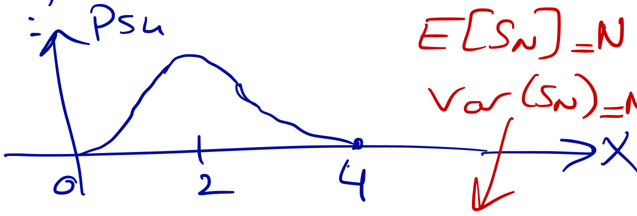
linear $\sim x$

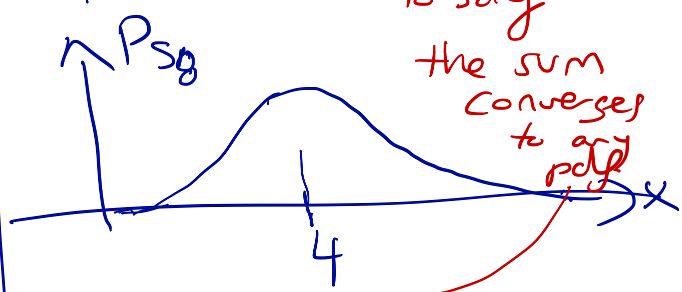


$\sim x^2$

eg. $\boxed{X_i \sim U[0,1]}$







As N $\nearrow$ the pdf shifts to the right

$E[S_N] = N \cdot \frac{1}{2}$

$Var(S_N) = N/12$

$\downarrow$

not possible to say the sum converges to any pdf

$S_1 = X_1$;

$S_2 = X_1 + X_2$; $p_{S_2} = p_{X_1} * p_{X_2}$

$S_3 = X_1 + X_2 + X_3$; $p_{S_3} = p_{X_1} * p_{X_2} * p_{X_3}$

$\vdots$

$$\boxed{E[S_N] = N \underbrace{E[X]}_{\mu}}$$

$$Var(S_N) = N \cdot \sigma^2$$

$\Longrightarrow$ We need a standardized sum

$\Rightarrow$ To overcome this problem, we define:

Standardized Sum: $\tilde{S}_N = \dfrac{S_N - E[S_N]}{\sqrt{Var(S_N)}}$

$$\tilde{S}_N = \frac{S_N - N\mu_x}{\sqrt{N\sigma_x^2}}$$

$$\Rightarrow E[\tilde{S}_N] = 0$$
$$Var(\tilde{S}_N) = 1$$

CLT says that $\tilde{S}_N \xrightarrow[N\to\infty]{}$ Gaussian r.v. w/ $\mu=0$, $\sigma^2=1$.

Theorem (Central Limit Theorem):

If $X_1, X_2, \dots, X_N$ are cont. IID r.v.s (general not uniform) each w/ mean $E_X[X]=\mu_x$, & variance $Var(X)=\sigma_x^2$, $S_N = \sum_{i=1}^{N} X_i$ then as $N\to\infty$:

$$\frac{S_N - E[S_N]}{\sqrt{Var(S_N)}} = \frac{S_N - N\mu_x}{\sqrt{N\sigma_x^2}} \xrightarrow[N\to\infty]{} \overbrace{\mathcal{N}(0,1)}^{\text{std normal}}$$

convergence in distrib.

Def: If $X_1, X_2, \dots$ and X r.v.s w/ distributions $(F_1, F_2, \dots)$ & F; distribution of X X_N converges to the distribution of X as $N\to\infty$ if $F_N(x) \xrightarrow[N\to\infty]{} F(x)$

Convergence in distribution

Equivalently CLT:
$$\underbrace{F_{\tilde{S}_N}(x)}_{P(\tilde{S}_N \le x)} \longrightarrow \Phi(x)$$
$$P(\tilde{S}_N \le x) \xrightarrow[N\to\infty]{} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$$

Q: Where can we use CLT?

Ex: Let $X_i \sim N(0,1)$ i.i.d.

Define: $Y_N = \sum_{i=1}^{N} X_i^2$

Approximate Y_N pdf by a Gaussian (using CLT)

$$\tilde{Y}_N = \frac{Y_N - N\,E_x[X^2]}{\sqrt{N\,Var(X^2)}}, \quad \text{CLT} \quad \tilde{Y}_N \to N(0,1)$$

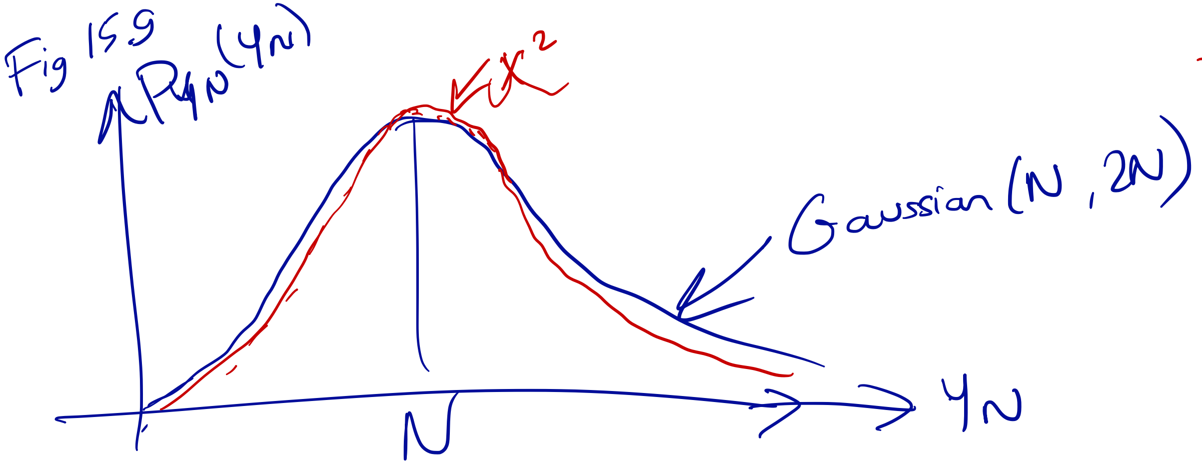
* X_i's are i.i.d so are the X_i^2's : Why?
b/c Functions of indep r.v.'s are indep.

$$E[X^2] = 1$$
$$Var(X^2) = \underbrace{E[(X^2)^2]}_{\underbrace{E[X^4]}_{=3 \text{ (show as HW)}}} - \underbrace{(E[X^2])^2}_{1} = 2$$

$$\Rightarrow \tilde{Y}_N = \frac{Y_N - N}{\sqrt{2N}} \longrightarrow N(0,1) \text{ by CLT.}$$

$$\Rightarrow Y_N = \sqrt{2N}\,\tilde{Y}_N + N \sim N(N, 2N)$$

Note: Exact $Y_N \sim \chi^2_N$ (Chi-Squared distrib)



Approx will become better as $N \uparrow$.

CLT: expressed into distributions: (CLT for discrete r.v.s)

$$P\left[\frac{S_N - E[S_N]}{\sqrt{Var(S_N)}} \le x\right] \xrightarrow[N \to \infty]{} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Reading Assignment

Section 15.6 : Opinion Polls :

Catch how CLT is used.