Central Limit Theorem

$X_1, X_2, \ldots$ i i d random variables

with $EX_1 = \mu$, $Var(X_1) = \sigma^2$, then

$$\underbrace{\frac{S_n - n\mu}{\sigma\sqrt{n}}}_{= Z_n} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

$\Downarrow$ distribution

$Z_n \leftarrow F_n$ }.

Then we have $F_n(x) \longrightarrow \Phi(x)$

for every x.

This convergence is actually uniform

in x.

Polya's theorem: If Y_n is a sequence

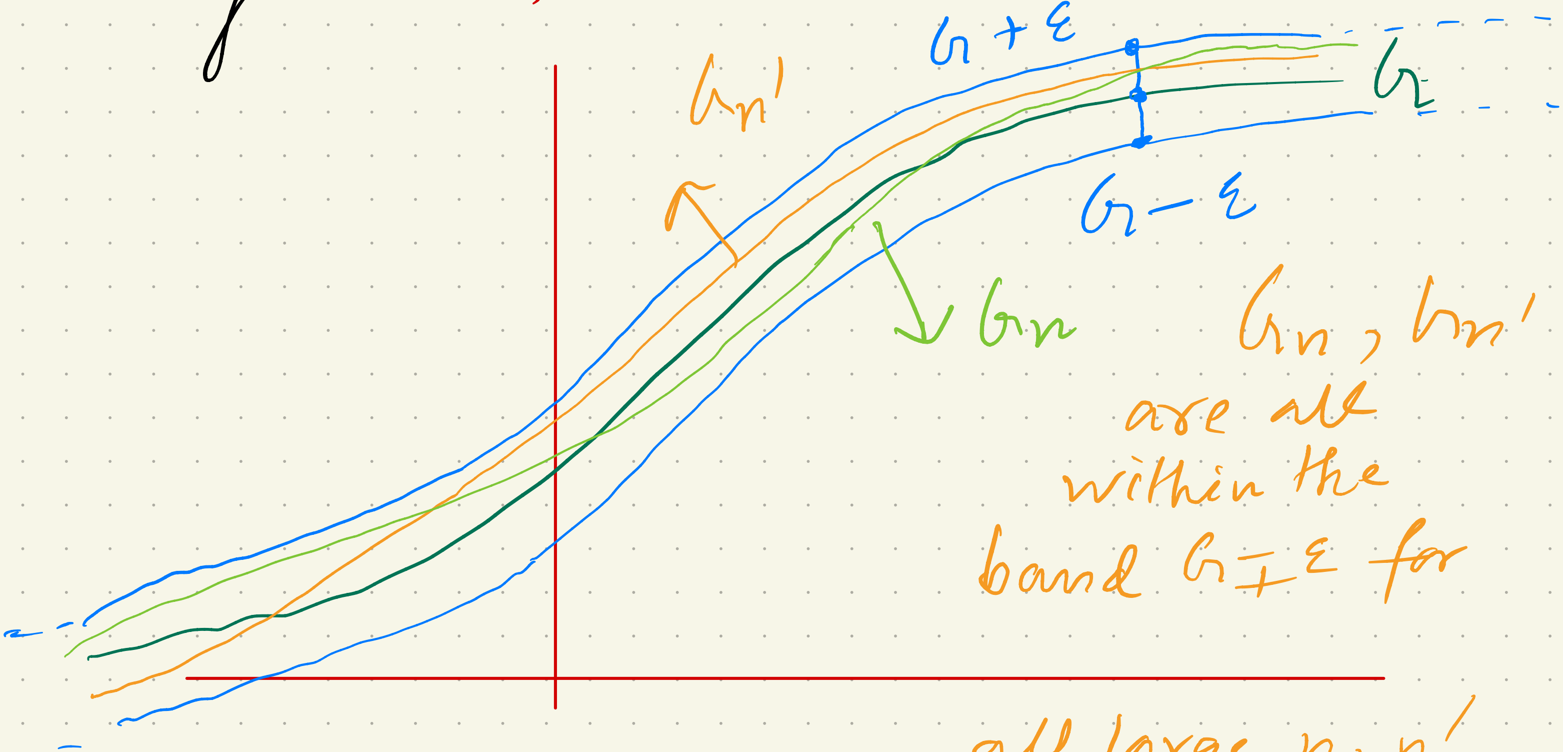
of random variables such that

$Y_n \xrightarrow{d} Y$ where Y is a continuous

random variable, G_n denotes the

cdf of Y_n and G denotes the cdf of Y, then:

$$\sup_y |G_n(y) - G(y)| \to 0 \text{ as } n \to \infty$$



G_n, G_n' are all within the band $G \mp \varepsilon$ for all large n, n'.

Implication: $S_n = X_1 + \cdots + X_n$

Distribution of $S_n \underset{\text{appx}}{\sim} N[\underbrace{n\mu}_{\mathbb{E}S_n}, \underbrace{n\sigma^2}_{\operatorname{Var} S_n}]$

likewise

$$\overline{X}_n \underset{\text{appx}}{\sim} N\left[\underbrace{\mu}_{\mathbb{E}\overline{X}_n}, \underbrace{\frac{\sigma^2}{n}}_{\operatorname{var} \overline{X}_n}\right]$$

To see this consider:

$$\left| P(S_n \le t) - P(W_n \le t) \right|.$$

where $W_n \sim N(n\mu, n\sigma^2)$

Then $\dfrac{W_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$

$$= \left| P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{t - n\mu}{\sigma\sqrt{n}} \right) - P\left(\underbrace{\frac{W_n - n\mu}{\sigma\sqrt{n}}}_{= Z \sim N(0,1)} \le \frac{t - n\mu}{\sigma\sqrt{n}} \right) \right|$$

$$= \left| P\left(Z_n \le \frac{t - n\mu}{\sigma\sqrt{n}} \right) - \Phi\left(\frac{t - n\mu}{\sigma\sqrt{n}} \right) \right|$$

$$= \left| \underline{F_n}\left(\frac{t - n\mu}{\sigma\sqrt{n}} \right) - \Phi\left(\frac{t - n\mu}{\sigma\sqrt{n}} \right) \right|$$

$$\leq \sup_{x \in \mathbb{R}} \left| F_n(x) - \Phi(x) \right|$$

$\longrightarrow 0$ by Polya's theorem, as Φ is continuous.

What we've shown is that:

$$\sup_t \left| P(S_n \leq t) - P(N(n\mu, n\sigma^2) \leq t) \right| \longrightarrow 0 \text{ as } n \to \infty.$$

Immediate ~~that~~ $\text{Bin}(n, p)$ for p fixed and large n (and p not tremendously small) can be approximated by $N[np, np(1-p)]$

$\text{Bin}(n,p) =_d X_1 + \dots + X_n$

where X_i's are i.i.d $N(0,1)$.

But if p is tiny and np is moderate (rare events but lots of trials), the Poisson approximation works better.

Poisson $(n) \overset{d}{=} X_1 + \cdots + X_n$

where X_i's are i.i.d Poi(1)

and therefore CLT applies.

$\chi^2_n \overset{d}{=} X_1 + \cdots + X_n$ where

X_i's are i.i.d χ^2_1. So

$\chi^2_n \underset{appx}{\sim} N(n, 2n)$.

Connection between convergence in probability and convergence in distribution:

(i) If $X_n \to_p X$, then $X_n \xrightarrow{d} X$.

Continuous Mapping Theorem:

If $X_n \underset{p\ \ d}{\longrightarrow} X$ and h is a function such ~~that~~ $P(X \in D_h) = 0$, where $D_h = \{t : h \text{ is not continuous at } t\}$

Then $h(X_n) \underset{p\ \ d}{\longrightarrow} h(X)$

[For example: $h(x) = \begin{cases} 0 & \text{if } x \text{ rational} \\ 1 & \text{otherwise} \end{cases}$

then h is not continuous anywhere and so continuous mapping does not work at all]

The exact same statement for

continuous mapping holds replacing convergence in probability everywhere by convergence in distribution!

Simple application:

$X_1, X_2, X_3, \ldots$ i.i.d random variables

$\mathbb{E}X_1 = \mu$, $\mathrm{Var}(X_1) = \sigma^2$.

look at the empirical estimate of σ^2 based on $X_1, \ldots, X_n$.

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

$$\xrightarrow{P} \sigma^2,$$

Would like to show that

Quick extend the idea of convergence in probability to multi-dimensional random variables:

Consider $\{\underset{\sim}{X}_n\}$: sequence of random vectors

$$\underset{\sim}{X}_n = \begin{pmatrix} X_{n,1} \\ \vdots \\ X_{n,d} \end{pmatrix}$$ where d is fixed.

Then $\underset{\sim}{X}_n \rightarrow_p \underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ if

$$P(\| \underset{\sim}{X}_n - \underset{\sim}{X} \| > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty$$

This is equivalent to $X_{n,j} \xrightarrow[p]{} X_j$ for $j = 1, 2, \ldots d$.

Continuous mapping continues to hold for random vectors!!

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

$$\left.\begin{aligned} Y_n &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\ Z_n &= \bar{x}_n \end{aligned}\right\}$$

$h(y, z) = y - z^2$ which is continuous everywhere.

Note $h(Y_n, Z_n) = \widehat{\sigma^2}$.

$$\begin{pmatrix} Y_n \\ Z_n \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \mathbb{E}X_1^2 \\ \mathbb{E}X_1 \end{pmatrix}$$

Note: $Y_n^2 = \frac{1}{n}\sum_{i=1}^{n} V_i$ where $V_i = X_i^2$, V_i's are iid

$\downarrow$ WLLN

$$\mathbb{E}V_1 = \mathbb{E}X_1^2$$

so $\begin{pmatrix} Y_n \\ Z_n \end{pmatrix} \xrightarrow{}_p \begin{pmatrix} \mathbb{E}X_1^2 \\ \mathbb{E}X_1 \end{pmatrix}$

so $h(Y_n, Z_n) = Y_n - Z_n^2$

$\xrightarrow{p} \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2$

$= \sigma^2.$

by the continuous mapping theorem.

Slutsky's theorem:

$\{X_n, Y_n, Z_n\}$: triplet of random variables.

$X_n \xrightarrow{}_p X$ (can be a proper random variable)

$Y_n \xrightarrow{p} c_1$. (constant)

$Z_n \xrightarrow{}_p c_2$ (constant)

Then $Y_n \cdot X_n + Z_n \xrightarrow[p]{} c_1 X + c_2$

Very useful result for asymptotic analysis (511).

An extremely useful inequality for probability and statistics is Jensen's Inequality:

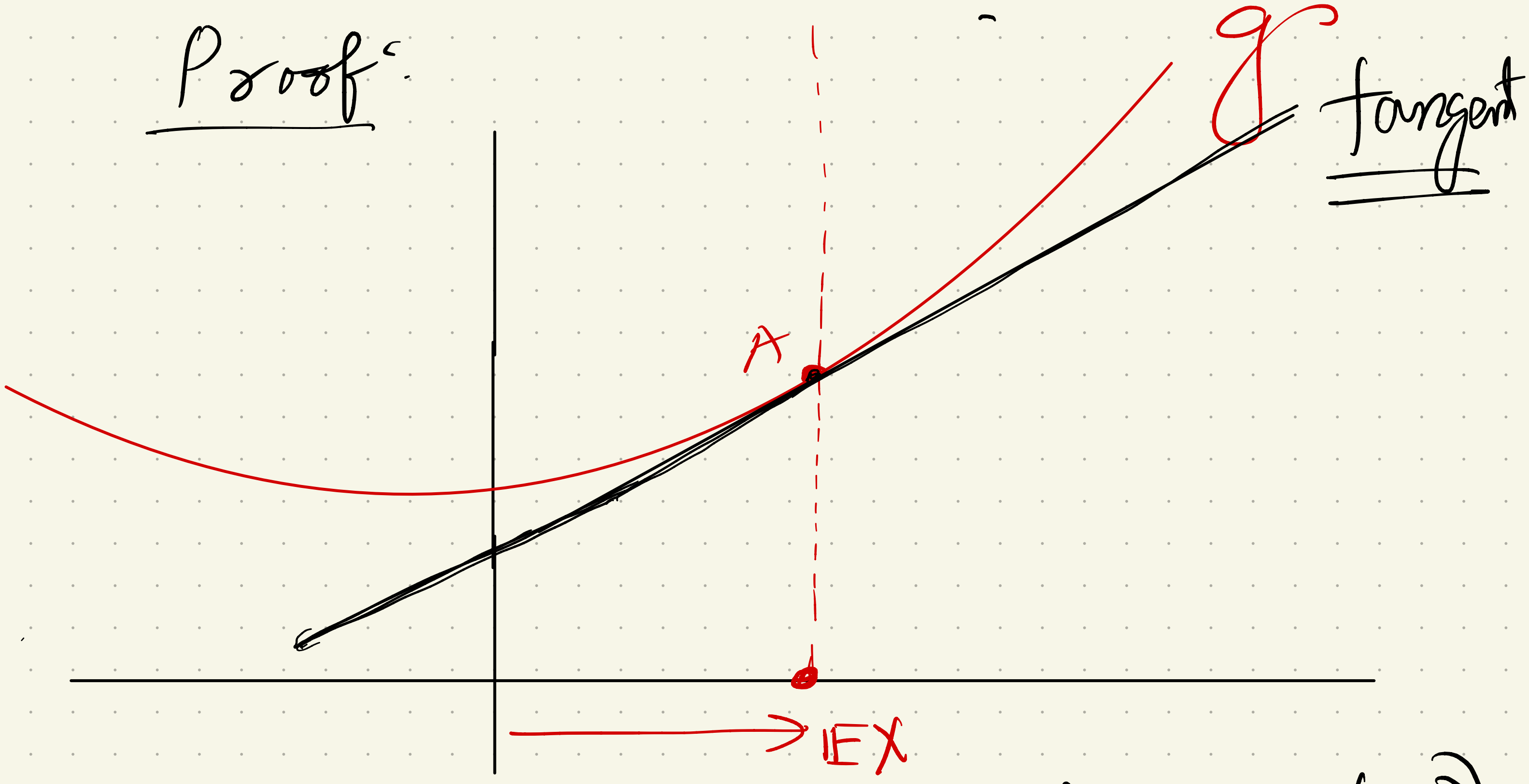
Consider a random X taking values in the interval (a, b) [could be $(-\infty, \infty)$] and suppose $\mathbb{E}X < \infty$. Let g be a convex function defined on (a, b) with $\mathbb{E}[g(X)] < \infty$.

Then $\mathbb{E}[g(X)] \geq g(\mathbb{E}X)$

Note: Inequality reverses for g concave, since then $-g$ is convex

Note: When g is linear i.e if is both concave and convex, equality holds i.e $\mathbb{E}[g(X)] = g(\mathbb{E}X)$

Proof:



co-ordinates of A are $(\mathbb{E}X, g(\mathbb{E}X))$

Tangent line to graph of g at point $(\mathbb{E}X, g(\mathbb{E}X))$ lies under the curve of g by convexity

i.e $g(t) \geq \alpha_0 + \beta_0 t$

where α_0, β_0 are the parameters of the line.

Notice: $\boxed{g(\mathbb{E}X) = \alpha_0 + \beta_0 \mathbb{E}X}$

so: $g(X) \geq \alpha_0 + \beta_0 X$ w.p. 1.

$$\Rightarrow \underline{\mathbb{E}[g(X)]} \geq \mathbb{E}[\alpha_0 + \beta_0 X]$$
$$= \alpha_0 + \beta_0 \mathbb{E}X$$
$$= g(\mathbb{E}X)$$

<u>Second part</u>: Equality i.e

$\mathbb{E}[g(X)] = g(\mathbb{E}X)$ ~~for a~~ proper convex function on (a,b) (i.e not linear) only if <u>X is</u> degenerate at <u>$\mathbb{E}X$</u> i.e X is constant w.p. 1.

Vast array of applications of this result...

A.M $\geq$ G.M $\geq$ H.M inequality

Let $x_1, x_2, \ldots, x_n$ be n numbers (assume +ve) $\longrightarrow$ distinct points.

Then:

$$A.M = \underbrace{\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \left(x_1 \cdot x_2 \cdots x_n\right)^{1/n}}_{1}$$

$$\underset{2}{\geq} \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \cdots + \frac{1}{x_n}\right)}$$

look up Example 4.7.8 for complete proof

A.M $\geq$ H.M

Jensen with $g(t) = \frac{1}{t}$

Inequality (1):

suffices to show:

$$\boxed{\log\left(\frac{x_1 + \cdots + x_n}{n}\right)}$$

$\log(\mathbb{E}X)$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \log(x_i) = \mathbb{E}[\log X]$$

Let X be a random variable that satisfies $P(X = x_i) = \frac{1}{n}$.

The result is immediate from the concavity of the log function

Another inequality.

Let X be a random variable and let g and h both be non-decreasing functions defined on the range of X.

Then $\text{Cov}(\underbrace{g(X)}_{u}, \underbrace{h(X)}_{v}) \geq 0$

Neat little trick:

Take X_1 and X_2 to be i.i.d. copies of X.

look at $g(X_1) - g(X_2)$ and $h(X_1) - h(X_2)$ and consider

$$\mathbb{E}\left[(g(X_1) - g(X_2))(h(X_1) - h(X_2))\right]$$

$$\geq 0$$

Now expand to get:

$$\mathbb{E}[g(X_1)h(X_1)] + \mathbb{E}[g(X_2)h(X_2)]$$

$$- \mathbb{E}[g(X_2)h(X_1)] - \mathbb{E}[g(X_1)h(X_2)] \geq 0$$

$$\mathbb{E}[g(X_2)h(X_1)] = \mathbb{E}[g(X_2)]\,\mathbb{E}[h(X_1)] = \mathbb{E}[g(X_1)]\,\mathbb{E}[h(X_1)]$$

$$\mathbb{E}[g(X_1)h(X_2)] = \mathbb{E}[g(X_1)]\,\mathbb{E}[h(X_2)] = \mathbb{E}[g(X_1)] \times \mathbb{E}[h(X_1)]$$

also $\mathbb{E}[g(X_1)h(X_1)]$

$$= \mathbb{E}[g(X_2)h(X_2)]$$

From these observations, we get:

$$2\,\mathbb{E}[g(X_1)h(X_1)] - 2\,\mathbb{E}[g(X_1)]\,\mathbb{E}[h(X_1)] \geq 0$$

i.e. $2 \cdot \mathrm{Cov}(g(X_1), h(X_1)) \geq 0$.

Corollary: If g is $\uparrow$ and h is $\downarrow$, then $\mathrm{Cov}(g(X_1), h(X_1)) \leq 0$.

an interesting exercise:

Let $(X_1, X_2, \ldots, X_n)$ be a random vector with

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N\left[\underset{\sim}{0}_{n\times 1}, \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & & & & \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} \right]$$

equicorrelation matrix

(i) Show that ~~the~~

$$(X_1, \ldots, X_n) \overset{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(n)})$$

where π is a permutation of $\{1, 2, \ldots, n\}$.

This is called exchangeability i.e. the joint distribution is invariant to permutations.

(ii) Show that $\rho \geq -\frac{1}{n-1}$ }

Hint: What is the variance of $\bar{X}_n$?

(iii) Find the joint distribution of $(X_1, \ldots, X_{n-1}, \bar{X}_n)$.

(use properties of the multivariate normal distribution)]

(iv) Explicitly determine the conditional distribution of $(X_1, \ldots, X_{n-1} \mid \bar{X}_n)$.]

(v) What happens to the conditional distribution as $n \to \infty$?