WLLN

→ Proof of this we assumed $\sigma^2 < \infty$. General proof can be done.

→ X_n : i.i.d with mean $\mu < \infty$. Then

No bounds on variance:

$$\lim_{n\to\infty} P\left(\left|\underbrace{\frac{\sum_{i=1}^{n} X_i}{n}}_{M_n} - \mu\right| > \varepsilon\right) = 0.$$

Sample mean

The difference b/w M_n & μ being greater than +ve ε is 0. So M_n lies in $\mu \pm \varepsilon$ with probability 1.

→ $P(|M_n - \mu| > \varepsilon) \leq \dfrac{E[|M_n-\mu|^2]}{\varepsilon^2} = \dfrac{\sigma^2}{n\varepsilon^2}$ (By Chebyshev's ineq.)

→ $var(X_i) = \sigma^2 \; \forall i \Rightarrow var(M_n) = \dfrac{\sigma^2}{n}$

$var(aX_1 + bX_2)$, $X_1 \perp\!\!\!\perp X_2$

$= a^2 var X_1 + b^2 var X_2$.

WLLN.

$X_i \sim$ iid mean

$$\frac{\sum_{i=1}^{n} X_i}{n} \to \mu \Rightarrow P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) = 0$$

Paretto distri. (Finite mean, infinite var.)

• Application of WLLN

X, P_X, $x \in \{a, b, c\}$.

$X_1, X_2, \ldots, X_n$ i.i.d from P_X. Estimate P_X (i.e., $P_X(a), P_X(b), P_X(c)$)

Given a seq., we assume the empirical estimate to be the best (Pretty much it is).

Empirical distri.

I_A → Indicator RV of event A

$$\frac{\sum_{i=1}^{n} \overbrace{I\{X_i = a\}}^{Y_i}}{n} \xrightarrow{n\to\infty} P_X(a) \quad \text{(Assuming WLLN)}.$$

→ Freq. of a's in $X_1, \ldots, X_n$.

$\{X_i = a\} = \{\omega : X_i(\omega) = a\}$.

$$E[Y_i] = E[I\{X_i = a\}]$$
$$= 1 \cdot P_X(a) + 0$$
$$= P_X(a).$$

- Convergence in distribution

$X_n \to X$ in distribution (CDF) if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$

(Seq. of RV)

for all points x at which $F_X(x)$ is continuous.

We say RV converge when the CDFs converge.

Limit of a fun^c exists when fun^c is contin.

- Central limit theorem

Let $X_1, X_2, \ldots,$ be a seq. of i.i.d RV with mean μ & var. σ^2

Then $$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{\text{in dist.}} N(0,1)$$

Standard Gaussian.

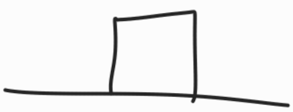
i.e., $$\lim_{n\to\infty} P(Z_n \le x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

Interpretation

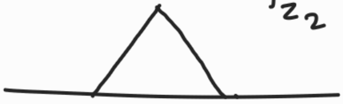
$X_i \sim$ i.i.d Uniform $[0,1]$.

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \frac{n}{2}}{\sqrt{n/12}}, \qquad f_{z_1}, \quad z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$$

$$z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}} \qquad f_{z_1}$$



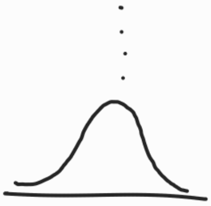
$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$$



f_{Z_2}

Something related to convolution (See what actually is convolution plz)

$$Z_3 = X_1 + X_2 + X_3 - \frac{3}{2}$$



Gaussian.

→ Consider $\mu = 0$, $\sigma = 1$

So CLT:

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \xrightarrow{\text{in dist.}} N(0,1).$$

Importance of this $\sqrt{n}$

Can we say $\sum_{i=1}^{n} X_i$ is Gaussian? (Transpose $\sqrt{n}$ to other side).

Consider $X_i \in \{-1, 1\}$.

Not all real numbers.

↳ This will takes only integer values.

$\sqrt{n}$ brings fractions into pictures.

→ $a_1, a_2, \ldots, a_n$ $n \in \mathbb{N}$.

↳ Seq. of rational nos. converges to an irrational no. (Involves RA).

Proof of CLT.

$X_i : \mu = 0, \sigma = 1$

WLOG we prove it for $\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \xrightarrow{\text{in dist.}} N(0,1)$

$$\frac{\sum X_i - n\mu}{\sqrt{n}\,\sigma} \xrightarrow[\text{dist}]{\text{in}} N(0,1)$$

$$Y_i = \frac{X_i - \mu}{\sigma}$$

$$\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}} \xrightarrow{\text{in dist}} N(0,1)$$

Let $Z_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$

$$M_{Z_n}(s) = E[e^{sZ_n}] = \left(E[e^{sX_1/\sqrt{n}}]\right)^n$$ ($\because$ for every i, distri. is same coz. i.i.d).

$$= \left(M_X\left(\frac{s}{\sqrt{n}}\right)\right)^n$$

If we show convergence of MGF to $\mathcal{N}(0,1)$, we're done.

$$M_{Z_n} \longrightarrow M_N \quad N \sim N(0,1)$$
$$\Downarrow$$
$$F_{Z_n} \rightarrow F_N \quad N \sim N(0,1)$$

($\because$ MGF $\Leftrightarrow$ CDF)

$$M_Z(s) = e^{s^2/2}$$

$$L(s) = \ln M_X(s)$$

$$\lim_{n\to\infty} \ln M_{Z_n}(s) = s^2/2 \longrightarrow \text{To show.}$$

$L(0) = 0$, $L'(0) = 0$, $L''(0) = 1$ (Variance). } Exercise.

$\left(M_X(0) = 1,\ M_X'(0) = E[X],\ M_X''(0) = E[X^2]\right)$

$$\lim_{n\to\infty} n \ln M_X\left(\frac{s}{\sqrt{n}}\right) = \lim_{n\to\infty} n \cdot L\left(\frac{s}{\sqrt{n}}\right).$$

$$= \lim_{n\to\infty} \frac{L\left(\frac{s}{\sqrt{n}}\right)}{\frac{1}{n}}$$ (Applying L'Hospital rule).

$$= \lim_{n\to\infty} \frac{L'\left(\frac{s}{\sqrt{n}}\right) \cdot s\left(\frac{-1}{2}\right)n^{-3/2}}{\left(\frac{-1}{n^2}\right)} = \lim_{n\to\infty} \frac{L'\left(\frac{s}{\sqrt{n}}\right)s}{2n^{-1/2}}$$

$L(0)=0$
$L'(0)=0$
$L''(0)=1$

(Applying L'Hospital's rule again).

$$= \lim_{n\to\infty} \frac{s^2 L''\left(\frac{s}{\sqrt{n}}\right)n^{-3/2}}{2n^{-3/2}} = \lim_{n\to\infty} \frac{s^2 L''\left(\frac{s}{\sqrt{n}}\right)}{2} = \frac{s^2 L''(0)}{2} = \frac{s^2}{2}.$$

MGF may not exist for all RV.
Characteristic func.
$\Downarrow$
Exists for every RV.

More general proof using characteristic func.

- Normal approximation

$S_n = \sum_{i=1}^{n} X_i$, $X_i \sim$ i.i.d (μ, σ) ; $P(S_n \le c) = ?$

(Approx. also fine).

$$P(S_n \le c) = P\left(\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \le \frac{c - n\mu}{\sqrt{n}\,\sigma}\right) \approx \Phi\left(\frac{c-n\mu}{\sqrt{n}\,\sigma}\right)$$

$$\Phi(t) = \int_{-\infty}^{t} \frac{e^{-x^2/2}\,dx}{\sqrt{2\pi}}$$

Exercise

$S_{50} = \sum_{i=1}^{50} X_i$ $\qquad P(90 < S_{50} < 100) = ?$

$\hookrightarrow = -\Phi(-\sqrt{2}) + \Phi(\sqrt{2})$

$\mu = 2$, var $= 1$

- Convergence in mean square

$X_n \to X$ in mean square.

$\lim_{n\to\infty} E[(X_n - X)^2] = 0$. $\quad X_n \to X$ in prob. if $P(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$.

Thm: P.T convergence in mean square implies convergence in probability.