Lec - 43

Theo : (Strong Law of Large Numbers).

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean μ. Then

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1$$

where $S_n = X_1 + X_2 + \cdots + X_n$.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean μ and non-zero variance σ^2.

Define $S_n := X_1 + X_2 + \cdots + X_n$, $\quad Z_n := \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$

Then $\displaystyle\lim_{n\to\infty} P(Z_n \le x) = N(x) \quad \forall x \in \mathbb{R}$

where $N(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$.

Example ①: Let X_i's be independent Bernoulli (p) random variables. Then $E[X_i] = p$ and $Var(X_i) = p(1-p)$. Also

$$S_n = X_1 + X_2 + \cdots + X_n$$

has Binomial (n, p) distribution. Thus

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$$

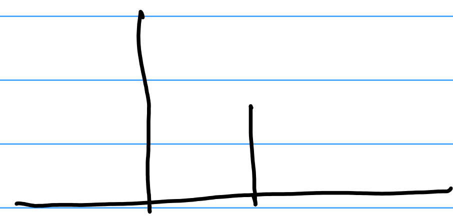
Let $p = \frac{1}{3}$.

We plot the pmf of Z_n for different values of n.

$n=1$:

$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

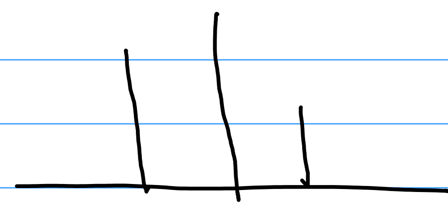
PMF of Z_1:



$n=2$:

$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

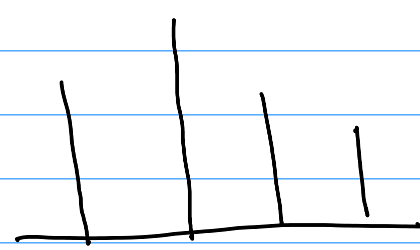
PMF of Z_2:



pmf of Z_3:

$n=3$:

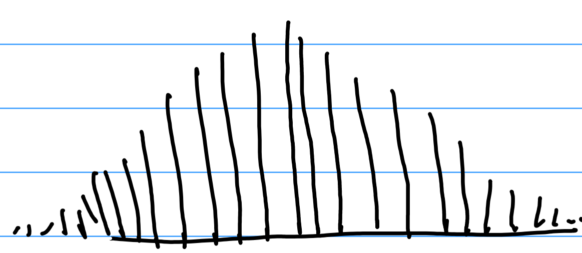
$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$



pmf of Z_{30}:

$n=30$:

$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$



Hence, the cdf of Z_n will converge to the std. normal cdf.

$$F_{Z_n}(x) = \sum_{z \in R_{Z_n} : z \leq x} f_{Z_n}(x) \longrightarrow N(x)$$

This is preciesly what the CLT states.

<u>Example ②</u>: Let X_i's be independent uniform $(0,1)$ random variables.

Then $E(X_i) = \frac{1}{2}$, $Var(X_i) = \frac{1}{12}$.

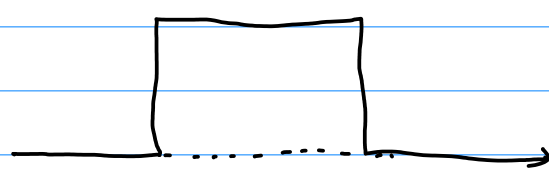
Let $S_n = X_1 + X_2 + \cdots + X_n$. In this case

$$Z_n = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

We plot the pdf of Z_n for different values of n.

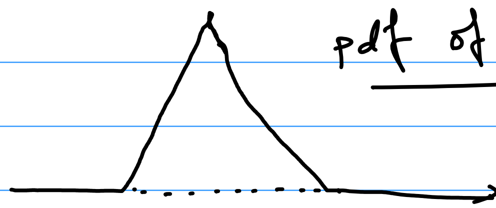
<u>$n=1$</u>: $Z_1 = \dfrac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$

<u>pdf of Z_1</u>:



<u>$n=2$</u>: $Z_2 = \dfrac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$

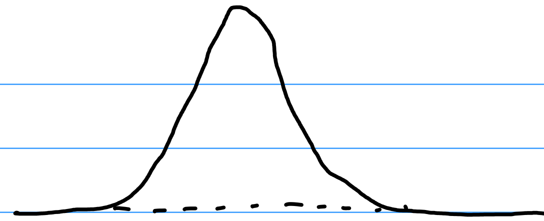
<u>pdf of Z_2</u>:



$n=3$:

$$Z_3 = \frac{X_1 + X_2 + X_3 - \frac{2}{3}}{\sqrt{\frac{3}{12}}}$$

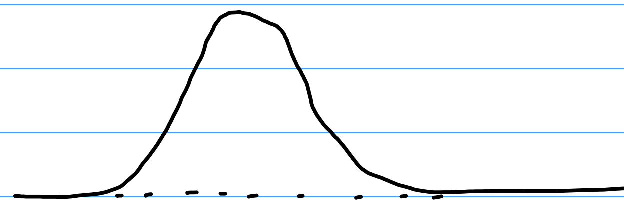
pdf of Z_3:



$n=30$:

$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{\frac{30}{12}}}$$

pdf of Z_{30}:



Hence the cdf of Z_n will converge to the standard normal cdf.

$$F_{Z_n}(x) = \int_{-\infty}^{x} f_{Z_n}(z)\,dz \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz$$

This is what the CLT says.

(Qn) Suppose (X_n) is a sequence of iid discrete (continuous) random variables with finite mean and non-zero variance. Then does CLT say that pmf (or pdf) of Z_n converges to pdf of standard normal random variable?

Ans: NO. CLT does not say that pmf (pdf) of Z_n converges to pdf of standard normal random variable. CLT is about convergence of distributions functions.

Suppose (X_i) are continuous with pdf f.

Then S_n has pdf $(*f)^n$. CLT says that $\forall\, x \in \mathbb{R}$

$$\lim_{n\to\infty} \int_{-\infty}^{\frac{x-n\mu}{\sigma\sqrt{n}}} (*f)^n (t)\,dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

Normal Approximation based on the Central Limit Theorem:

CLT says $P(Z_n \le x) \approx N(x)$ for large values of n

Note that $S_n = \sigma\sqrt{n}\, Z_n + n\mu$.

Since normality is preserved under linear transformations & since $Z_n \sim N(0,1)$, this is equivalent to treating S_n as a normal random variable with mean $n\mu$ and variance $n\sigma^2$.

X_i's iid.
mean μ & var. σ^2

$P(S_n \le c) :=$

Step ① Calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n.

Step ② Use the approximation

$$P(S_n \le c) \approx N\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right).$$

Example ③: We load on a plane 100 packages whose weights are independent random variables that are uniformly distributed betn 5 and 50 kg. What is the probability that the total weight will exceed 3000 kg?

Soln: Let X_i denote the weights of the i^{th} package.

$X_1, X_2, \dots, X_{100}$ are iid uniform random variables with density

$$f(x) = \begin{cases} \frac{1}{45}, & \text{if } 5 \le x \le 50 \\ 0 & \text{o.w.} \end{cases}$$

Let $S = X_1 + X_2 + \dots + X_{100}$ denote the total weight. Then the question is to calculate

$$P(S > 3000).$$

Let f, g and h be functions on the reals and suppose the convolutions $(f*g)*h$ and $f*(g*h)$

exist. Then we have

$$(f * g) * h = f * (g * h)$$

Convolution:

$$(f * g)(u) = \sum_{x+y=u} f(x) g(y)$$

Using this result and S is sum of independent random variables, we can see that pdf of S is 100-fold convolution of f. (Difficult!)

So instead of finding out the actual probability which is very difficult in this case we find an approximate answer with the help of CLT.

Treat S as normal random variable.

$\therefore$ Mean of $S = 100\mu$ & variance $= 100\sigma^2$.

$$E(X_i) = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{45}\int_{5}^{50} x\,dx = \frac{5+50}{2}$$

$$= 27.5$$

$$E(X_i^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \frac{1}{45}\int_{5}^{50} x^2\,dx = 925$$

$$Var(X_i) = 925 - (27.5)^2 = 168.75$$

$\therefore$ Now $P(S > 3000) = 1 - P(S \leq 3000)$

$$= 1 - N\left(\frac{3000 - 2750}{10\sqrt{168.75}}\right)$$

$$= 1 - N(1.92) \text{ Ans.}$$