Probability and Statistics

November-1

Central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with mean μ and variance σ^2.

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

CLT $\Rightarrow$ $\lim\limits_{n\to\infty} \text{Prob}\left(\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right) = \Phi(x)$ ~~$\forall$~~

$\forall x \in \mathbb{R}$

where $\Phi(x)$ is the CDF of $N(0,1)$.

$$\boxed{S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)}$$

(approx)

$\sim$ (approx) CDF of S_n^* approaches to the CDF of $N(0,1)$ as $n\to\infty$ $\forall x \in \mathbb{R}$.

$S_n \underset{\text{(approx.)}}{\sim} N(n\mu, n\sigma^2)$

$$\boxed{X_1 + X_2 + \cdots + X_n \underset{\text{(approx)}}{\sim} N(n\mu, n\sigma^2)}$$

(As $n \to \infty$ CDF of $X_1 + \cdots + X_n = S_n$ approaches CDF of $N(n\mu, n\sigma^2)$)

Applications of CLT

i) $X_1, X_2, \cdots, \ldots$ be the sequence of iid Bernoulli (p).

$$X_i = \begin{cases} 0 & \text{with } 1-p \\ 1 & \text{with } p \end{cases}$$

$E(X_i) = p$, $\quad var(X_i) = p(1-p)$

$$S_n = X_1 + X_2 + \cdots + X_n \sim \underbrace{\text{Binomial } (n, p)}$$

$$\text{CLT} \Rightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} \underset{\text{approx}}{\sim} N(0,1)$$

$$\Rightarrow S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} \underset{\text{approx}}{\sim} N(0,1)$$

<u>H.W.</u> $S_1^*, S_2^*, S_3^*, \cdots, \underset{\downarrow}{\underbrace{S_{30}^*}}$ $\qquad S_1^* = \dfrac{S_1 - p}{\sqrt{p(1-p)}} = \dfrac{X_1 - p}{\sqrt{p(1-p)}}$

(Draw the pmf)

Ex: $Y \sim \text{Binomial}(\underbrace{20}, \tfrac{1}{2})$

$$\text{Prob}(8 \le Y \le 10) = \left(\tfrac{1}{2}\right)^{20} \underbrace{\left[\binom{20}{8} + \binom{20}{9} + \binom{20}{10}\right]} = \underbrace{0.457}$$

Notice: $Y = X_1 + X_2 + \cdots + X_{20} = S_{20}$

where $X_i \sim \text{Bernoulli}(\tfrac{1}{2})$ and

X_i's are independent.

$E(X_i) = p = \frac{1}{2} = \mu$; $\text{var}(X_i) = p(1-p) = \frac{1}{4} = \sigma^2$, $n = 20$

CLT $\Rightarrow$ $S_n^* = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} \underset{\text{approx}}{\sim} N(0,1)$

$$\text{Prob}(8 \le Y \le 10) = P\left(\frac{8-n\mu}{\sigma\sqrt{n}} \le \underbrace{\frac{Y-n\mu}{\sigma\sqrt{n}}} \le \frac{10-n\mu}{\sigma\sqrt{n}}\right)$$

$$= P\left(\frac{8-10}{\sqrt{5}} \le \frac{Y-n\mu}{\sigma\sqrt{n}} \le 0\right)$$

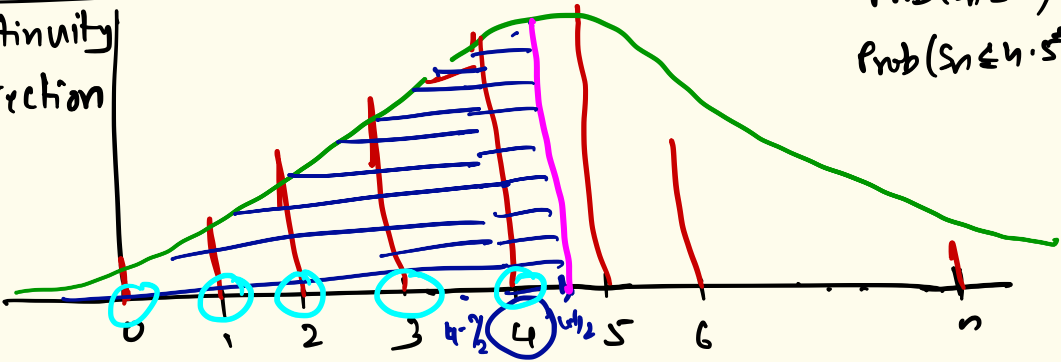
$$\underset{\text{CLT}}{\approx} \Phi(0) - \Phi\left(\frac{-2}{\sqrt{5}}\right)$$

$$= \underbrace{0.31}$$

Continuity correction

$\text{Prob}(S_n \le 4)$

$\text{Prob}(S_n \le 4.5)$



Comparison of $\text{Bin}(n, p) = S_n$ and $\underline{N(np, np(1-p))}$

$$P(8 \le Y \le 10) = P(7.5 \le Y \le 10.5)$$

$$= P\left(\frac{7.5 - n\mu}{\sigma\sqrt{n}} \le \frac{Y - n\mu}{\sigma\sqrt{n}} \le \frac{10.5 - n\mu}{\sigma\sqrt{n}}\right)$$

$$\overset{\text{CLT}}{\approx} \Phi\left(\frac{0.5}{\sqrt{5}}\right) - \Phi\left(\frac{-2.5}{\sqrt{5}}\right)$$

$$= \boxed{0.4567}$$

<u>Continuity correction</u> :

$$f_{S_n}(x) = \Phi\left(\frac{\boxed{x + 1/2} - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{\boxed{x - 1/2} - n\mu}{\sigma\sqrt{n}}\right)$$

$$\text{Prob}(S_n \le x) = \Phi\left(\frac{\boxed{x + \frac{1}{2}} - n\mu}{\sigma\sqrt{n}}\right)$$

Ex:

A bank teller serves customers standing in the queue one by one. The service time x_i for each customer i has mean $= 2$ and $var = 1$. Assuming that the serving times for different customers are indep, and letting Y be to be the total time bank teller spends serving 50 customers, find $(90 < Y < 110)$.

observe: $S_n = Y = X_1 + X_2 + X_3 + \cdots + X_{50}$; $n = 50$

$\underset{\mu}{E(X_i)} = 2$; $\underset{\sigma^2}{var(X_i)} = 1$ and X_i's are indep.

Required prob:

$$P(90 < Y < 110) = P\left(\frac{90-n\mu}{\sigma\sqrt{n}} \leq \frac{Y-n\mu}{\sigma\sqrt{n}} \leq \frac{110-n\mu}{\sigma\sqrt{n}}\right)$$

$$= P\left(\frac{90-100}{\sqrt{50}} \leq S_n^* \leq \frac{110-100}{\sqrt{50}}\right)$$

CLT

$$\approx \Phi\left(\frac{10}{\sqrt{50}}\right) - \Phi\left(\frac{-10}{\sqrt{50}}\right)$$

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

$$= 0.842$$