Central Limit theorem

Central Limit theorem gives the adequate reason for the everlasting presence of the normal distribution in probability and Statistic literature. In essence, it tells us that if we sum up iid random variables, the probability distribution of ~~of~~ the sum itself will eventually become a random variable with a normal distribution, no matter what was the probability distⁿ of the considered random variables.

Statement: Suppose that $x_1, x_2, \cdots, x_n$ are independent and identically (iid) distributed random variables with finite mean $(E(X_i) = \mu < \infty)$ and finite variance $(V(X_i) = \sigma^2 < \infty)$. Further let $\bar{X}_n$ is the sample mean of the sample $x_1, x_2, \cdots, x_n$ that is, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1) \text{ as } n \to \infty$$

Equivalently

$$P\left(\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \le z\right) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

as $n \to \infty$.

Remark: If we carefully relook at this celebrated CLT theorem then we find that whatever collection of iid random variables we start with, ultimately $\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}}$ will behave like a standard normal random variable. The only restrictions we need are that mean and variance of of that collection of rvs must be finite.

⊛ $X_1, X_2, \cdots, X_n$ iid rvs such that $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $i = 1, 2, \cdots, n$. then note that

$$E(\bar{X}_n) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n} E(X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{n} \cdot n\mu = \mu.$$

similarly $V(\bar{X}_n) = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$

$$= \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n)$$

$$V(\bar{X}_n) = \frac{1}{n^2}\left[V(X_1)+V(X_2)+\cdots+V(X_n)\right]$$

($\because X_i$ are independent $i=1,2,\cdots,n$.)

$$= \frac{1}{n^2}(\sigma^2+\sigma^2+\cdots+\sigma^2)$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

that's why we have

$$\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\left[\frac{\bar{X}_n-\mu}{\sigma}\right]$$

<u>Remark</u>: central limit theorem says that under given framework prob. distn of $\sqrt{n}\left[\frac{\bar{X}_n-\mu}{\sigma}\right]$ is given by standard normal $N(0,1)$.

from Here can you tell me distn of $\bar{X}_n$ itself

Prob distn of $\boxed{\bar{X}_n \sim N(\mu, \sigma^2/n)}$ — 2

from here can you provide ~~prob~~ prob. distn of S_n where

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

So prob. distⁿ of S_n is given by

$$\boxed{S_n \sim N(n\mu, \sigma^2 n)} \quad \text{--- (3)}$$

<u>Note</u>: Results stated in (2) & (3) can easily be obtained from Equation (1) by just using 'transformation of one variable method'.
So try to obtain results (2) & (3) from (1).

— — — x — — — x — — — x — —

Next we see some useful Applications of central limit theorem.

In particular we try to prove 'De-Moivre Laplace central limit theorem' for Bernoulli Random Variables.

Note that this particular case of CLT is a special case of the main statement.

De-Moivre Laplace CLT

Theorem: Suppose that $x_1, x_2, \cdots, x_n$ are iid Bernoulli $Ber(p)$ random variables. Define $S_n = \sum_{i=1}^{n} X_i = x_1 + x_2 + \cdots + x_n$. Then

$$P\left(\frac{S_n - np}{\sqrt{npq}} \leq x\right) = \Phi(x) \quad \text{as } n \to \infty$$

In other words, $\frac{S_n - np}{\sqrt{npq}} \sim N(0,1)$

Equivalently $S_n \simeq N(np, npq)$

Proof: we prove this result using the MGF method.

How to proceed ?? Try to compute the MGF of $\frac{S_n - np}{\sqrt{npq}}$. Thus we have

$$M_{\frac{S_n - np}{\sqrt{npq}}}(t) = E\left[e^{\left[\frac{S_n - np}{\sqrt{npq}}\right]t}\right]$$

$$= e^{-\frac{npt}{\sqrt{npq}}} E\left[e^{\frac{S_n}{\sqrt{npq}}t}\right]$$

$$= e^{-\frac{npt}{\sqrt{npq}}} E\left[e^{\frac{t}{\sqrt{npq}}[x_1 + x_2 + \cdots + x_n]}\right]$$

$$= e^{-\frac{npt}{\sqrt{npq}}} E\left[e^{\frac{tx_1}{\sqrt{npq}}} \cdot e^{\frac{tx_2}{\sqrt{npq}}} \cdots e^{\frac{tx_n}{\sqrt{npq}}}\right]$$

$$= e^{-\frac{npt}{\sqrt{npq}}} \left\{ E\left(e^{\frac{tx_1}{\sqrt{npq}}}\right) E\left(e^{\frac{tx_2}{\sqrt{npq}}}\right) \cdots E\left(e^{\frac{tx_n}{\sqrt{npq}}}\right) \right\}$$

($\because$ $x_1, x_2, \cdots, x_n$ are independent).

$$= e^{-\frac{npt}{\sqrt{npq}}} \left\{ M_{x_1}\left(\frac{t}{\sqrt{npq}}\right) M_{x_2}\left(\frac{t}{\sqrt{npq}}\right) \cdots M_{x_n}\left(\frac{t}{\sqrt{npq}}\right) \right\}$$

$$= e^{-\frac{npt}{\sqrt{npq}}} \cdot \left(M_{x_1}\left(\frac{t}{\sqrt{npq}}\right)\right)^n$$

Also Note that $X \sim Ber(p)$

$M_X(t) = q + pe^t$

$$= e^{-\frac{npt}{\sqrt{npq}}} \left[q + pe^{\frac{t}{\sqrt{npq}}}\right]^n$$

$$= \left[q\, e^{-\frac{pt}{\sqrt{npq}}} + p\, e^{\frac{t(1-p)}{\sqrt{npq}}}\right]^n$$

$$= \left[q\, e^{-\frac{pt}{\sqrt{npq}}} + p\, e^{\frac{tq}{\sqrt{npq}}}\right]^n$$

$$= \left[q\left(1 - \frac{pt}{\sqrt{npq}} + \frac{p^2}{npq}\frac{t^2}{2} + \cdots\right) + p\left(1 + \frac{qt}{\sqrt{npq}} + \frac{q^2}{npq}\frac{t^2}{2} + \cdots\right)\right]^n$$

$$= \left[1 + \frac{qp(p+q)}{2npq} t^2 + O(1/n)\right]^n$$

$$\left[\text{Note: A function } f(x) \text{ is of } O(x) \text{ if } \frac{f(x)}{x} \to 0 \text{ as } x \to 0\right]$$

$$= \left[1 + \frac{t^2}{2n} + O(1/n)\right]^n$$

$$\to e^{t^2/2} \text{ as } n \to \infty. \qquad \left[\left(1+\frac{x}{n}\right)^n \to e^x \text{ as } n \to \infty\right]$$

Thus MGF of $\frac{S_n - np}{\sqrt{npq}}$ is given by

$$\boxed{M_{\frac{S_n - np}{\sqrt{npq}}}(t) = e^{t^2/2} \quad \text{as } n \to \infty}$$

But what about $e^{t^2/2}$ (recall yourself)??

[If $X \sim N(0,1)$ then $M_X(t) = e^{t^2/2}$]

Thus as $n \to \infty$ MGF of $\frac{S_n - np}{\sqrt{npq}}$ behaves like MGF of a Standard normal distⁿ.

Thus by uniqueness property of MGF we find that we have

$$\frac{S_n - np}{\sqrt{npq}} \sim N(0,1) \text{ as } n \to \infty. \quad \text{——} (I)$$

Thus result is proved.

Next we try to prove the Alternative Part.

Given the result of (I) let us show that

$$S_n \simeq N(np, npq).$$

Again use MGF technique. Let us now compute mgf of S_n. Before that define

$$Y = \frac{S_n - np}{\sqrt{npq}} \Rightarrow S_n = Y\sqrt{npq} + np$$

Now $M_{S_n}(t) = E\left[e^{S_n t}\right] = E\left[e^{(Y\sqrt{npq}+np)t}\right]$

$$= e^{npt} E\left(e^{(\sqrt{npq}\cdot t)Y}\right)$$

$$= e^{npt} M_Y(t\cdot\sqrt{npq}) = e^{npt}\cdot e^{\frac{t^2 npq}{2}}$$

$(\because Y \sim N(0,1))$

$$= e^{npt + \frac{1}{2}t^2 npq}$$

So, $$M_{S_n}(t) = e^{npt + \frac{1}{2}t^2 npq}$$

This is the MGF of a normal distn with mean np and variance npq. Thus we have

$$\boxed{S_n \sim N(np, npq)}$$

Ex: Let $X_1, X_2, \ldots, X_n$ be iid $\chi^2_{(1)}$ random variables. Define $S_n = X_1 + X_2 + \cdots + X_n$. Then using CLT show that $\frac{S_n - n}{\sqrt{2n}} \sim N(0,1)$ as $n \to \infty$.

Ex: Let $X_1, X_2, \ldots, X_n$ be iid Poisson $P(\lambda)$ random variables. Show that $\frac{S_n - n\lambda}{\sqrt{n\lambda}} \sim N(0,1)$ as $n \to \infty$ where $S_n = \sum_{i=1}^{n} X_i$.