Sampling distribution:-

(Idea:- For repeating a random experiment, we will have different values of the function of the random sample, i.e., we have the distn. of the fn. of the random sample. That is called sampling distn.).

Example:- Suppose $X_1, .., X_n \sim N(\mu, \sigma^2)$.

Define $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ & $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$.

Thm i) $\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$.

ii) $\overline{X_n} \perp\!\!\!\perp S_n^2$ (indep. of each other).

iii) $\frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X_n})^2}{\sigma^2} \sim \chi^2_{n-1}$.

Proof :- i) $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.

Want to know the distⁿ of $\bar{X}_n$.

$$M_{\bar{X}_n}(t) = E\left[e^{t\bar{X}_n}\right] = E\left[e^{t\frac{1}{n}\sum_{i=1}^{n} x_i}\right]$$

MGF of $\bar{X}_n$

$$= \prod_{i=1}^{n} E\left[e^{\frac{t}{n} x_i}\right]$$ since X_i's are indep.

$$= \prod_{i=1}^{n} \left(e^{\frac{t}{n}\mu + \frac{1}{2}\frac{t^2}{n^2}\sigma^2}\right)$$

$$= e^{t\mu + \frac{1}{2}t^2\left(\frac{\sigma^2}{n}\right)}.$$

MGF of $N(\mu, \frac{\sigma^2}{n})$.

Hence, $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$.

ii) We want to establish $\bar{X}_n \perp\!\!\!\perp S_n^2$. (Given that $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$)

The joint density of $X_1, \ldots, X_n$ is

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2}, \quad x_i \in \mathbb{R}, \ \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}^+$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left\{ (x_i - \bar{x}_n) + (\bar{x}_n - \mu) \right\}^2}$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \mu)^2 \right]}$$

[Product term $= 0$ since

$$2 \sum_{i=1}^{n} (x_i - \bar{x}_n)(\bar{x}_n - \mu)$$

$$= 2(\bar{x}_n - \mu) \sum_{i=1}^{n} (x_i - \bar{x}_n).$$

$$= 2(\bar{x}_n - \mu)\left[n\bar{x}_n - n\bar{x}_n\right] = 0]$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 - \frac{n}{2\sigma^2}(\bar{x}_n - \mu)^2\right]} \quad (*)$$

Tranformation:-

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \begin{aligned} Y_1 &= \frac{1}{n}(X_1 + \cdots + X_n) \\ Y_2 &= X_2 - \bar{X}_n \\ Y_3 &= X_3 - \bar{X}_n \\ &\vdots \\ Y_n &= X_n - \bar{X}_n \end{aligned} \longrightarrow \begin{aligned} X_1 &= Y_1 - Y_2 - \cdots - Y_n \\ X_2 &= Y_1 + Y_2 \\ X_3 &= Y_1 + Y_3 \\ &\vdots \\ X_n &= Y_1 + Y_n \end{aligned}$$

A few observations:- 1) $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = Y_1$

ii) $(n-1) S_n^2 = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

$$= \underbrace{(X_1 - \bar{X}_n)^2}_{\parallel} + \sum_{i=2}^{n} (X_i - \bar{X}_n)^2$$

$X_1 = Y_1 - Y_2 - \cdots - Y_n$

$\bar{X}_n = Y_1$

$$= (-Y_2 - \cdots - Y_n)^2 + \sum_{i=2}^{n} Y_i^2 \quad \text{since } Y_i = X_i - \bar{X}_n \;\; \forall\, i = 2, \ldots, n.$$

$$= \left(-\sum_{i=2}^{n} Y_i\right)^2 + \sum_{i=2}^{n} Y_i^2.$$

A crucial observation:-

$\bar{X}_n \perp\!\!\!\perp S_n^2 \iff Y_1 \perp\!\!\!\perp (Y_2, \ldots, Y_n)$.

Hence, we try to derive the joint density of $Y_1, \ldots, Y_n$.

Notice

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} 1 & -1 & - & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ & & & & \\ 1 & 0 & \cdots & \cdots & 1 \end{vmatrix}$$

$$\Rightarrow |J| = n.$$

From (*) $\Rightarrow$ $f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = c \times e^{\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 - \frac{n}{2\sigma^2}(\bar{x}_n - \mu)^2\right]}$

So, the joint density of $Y_1, \ldots, Y_n$ is

$x_1 = y_1 - y_2 \cdots - y_n$

$\bar{x}_n = y_1$

$$f_{Y_1 \ldots Y_n}(y_1, \ldots, y_n) = c \times e^{\left[-\frac{1}{2\sigma^2}(x_1 - \bar{x}_n)^2 - \frac{1}{2\sigma^2}\sum_{i=2}^{n}(x_i - \bar{x}_n)^2 - \frac{n}{2\sigma^2}(\bar{x}_n - \mu)^2\right]}$$

$$= c \times e^{\left[-\frac{1}{2\sigma^2}(-y_2 \cdots - y_n)^2 - \frac{1}{2\sigma^2}\sum_{i=2}^{n} y_i^2 - \frac{n}{2\sigma^2}(y_1 - \mu)^2\right]}$$

$$= \underset{\substack{\text{some } K}}{K} \times \underbrace{e^{-\frac{n}{2\sigma^2}(y_1-\mu)^2}}_{\text{p.d.f. of } Y_1} \times \underbrace{\underset{\text{some } K^*}{K^*}\, e^{-\frac{1}{2\sigma^2}(-y_2-\cdots-y_n)^2 - \frac{1}{2\sigma^2}\sum_{i=2}^{n} y_i^2}}_{\text{Joint p.d.f. of } (Y_2, \ldots, Y_n).}$$

K & K^* are such that $C = K \times K^*$

Hence $Y_1 \perp\!\!\!\perp (Y_2, \ldots, Y_n)$.

$\quad\;\; \parallel$

Further, Hence, $\bar{X}_n \perp\!\!\!\perp S_n^2$

The proof is complete.

iii) Want to show that $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$

$$\iff \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)}{\sigma^2} \sim \chi^2_{n-1}.$$

Proof :-

1^{st} observation :-

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

$\Rightarrow \dfrac{X_i - \mu}{\sigma} \sim N(0,1) \quad \forall \; i = 1, \ldots, n.$

$\Rightarrow \dfrac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$

Therefore $M_{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}}(t) \overset{\text{check}}{=} (1-2t)^{-\frac{n}{2}}$

This implies that

$$(1-2t)^{-\frac{n}{2}} = E\left[e^{t \times \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2}\right]$$

$$= E\left[e^{\frac{t}{\sigma^2}\left[\sum_{i=1}^{n}(x_i-\bar{x}_n)^2+n(\bar{x}_n-\mu)^2\right]}\right]$$ (product term $=0$).

$$= E\left[e^{t\left\{\frac{(n-1)S_n^2}{\sigma^2}+\frac{n(\bar{x}_n-\mu)^2}{\sigma^2}\right\}}\right]$$

$$= E\left[e^{\frac{t(n-1)S_n^2}{\sigma^2}}\right]\times E\left[e^{\frac{t\,n(\bar{x}_n-\mu)^2}{\sigma^2}}\right].$$ since $\bar{X}_n \perp\!\!\!\perp S_n^2$

$$(1-2t)^{-\frac{n}{2}} = E\left[e^{\frac{t(n-1)S_n^2}{\sigma^2}}\right]\times(1-2t)^{-\frac{1}{2}}$$

since $\frac{n(\bar{x}_n-\mu)^2}{\sigma^2} \sim \chi^2_1$ → Reason.

$\bar{X}_n \sim N(\mu,\sigma^2)$

$\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$\Rightarrow \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \sim N(0,1)$

$\Rightarrow \frac{n(\bar{X}_n-\mu)^2}{\sigma^2} \sim \chi^2_1$.

$$\Leftrightarrow E\left[e^{\frac{t(n-1)S_n^2}{\sigma^2}}\right] = (1-2t)^{-(n-1)/2}$$ → MGF of χ^2_{n-1}.

Hence $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

Significance of this result:-

If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.
Want to test $H_0: \sigma = \sigma_0$ ag. $H_1: \sigma \neq \sigma_0$, where σ_0 is specified to us. To test H_0 ag. H_1, we can use this result.

Another fact:-

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

$$\Rightarrow \bar{X_n} \sim N(\mu, \sigma^2/n).$$

$$E(\bar{X_n}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$$

$$Var(\bar{X_n}) = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) + \frac{1}{n^2}\sum_{i \neq j} cov(X_i, X_j)$$

$$= \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n} \text{ since } X_i \perp\!\!\!\perp X_j \ \forall i \neq j.$$

A few well-known dist^{n}s :- (χ^2, t & F) :-

χ^2 dist^n with n degrees of freedom :-

Let $X_1, \ldots, X_n$ i.i.d. $N(0,1)$, then

$$U = X_1^2 + \cdots + X_n^2 \sim \chi^2_n .$$

The p.d.f. of U is $f_U(u) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-\frac{u}{2}} u^{\frac{n}{2}-1}$, $u \geq 0$.

Remark :- The explicit form of F_U (CDF of U) is not possible to derive, and hence, in practice you cannot generate random number by inverse transformation method.