The Exact Sampling Distributions

Here, we study certain distributions that arise in sampling from a normal i.e., $N(\mu, \sigma^2)$ population. The entire topic is divided into 5 sub-divisions:

1. Objectives
2. Introduction
3. Types of Exact Sampling distributions
4. Properties of Sampling Distributions
5. Summary

1. Objectives :

i. To introduce the concepts of sampling distributions

ii. To know the use of sampling distributions to conduct tests of significance.

2. **Introduction :** Any probability distribution and therefore, any sampling distribution can be partially described by its mean and standard deviation. When the sample size (n)is small (say below 30) and standard deviation(σ) is unknown, the data no longer follow the properties of normal distribution and it may be distributed as some other distributions such as chi-square distribution, t- distribution, F-distribution and etc.

3. Types of Exact Sampling Distributions:

i. Chi-square distribution

ii. Student's t- distribution

iii. Snedecor's F-distribution

3.1. Chi-square distribution:

Definition : "The square of a standard normal variate is called a chi-square variate with 1 degree of freedom".

That is, if $X \sim N(\mu, \sigma^2)$, we define $Z = (X - \mu)/\sigma \sim N(0, 1)$, then $Z^2 \sim \chi^2$ with 1 d.f.

Thus in general, if $X_1, X_2, \ldots, X_n$ are n independent random variables, distributed as $N(\mu_i, \sigma_i^2)$, for $i = 1, 2, \ldots, n$, then the random variable χ^2 defined by

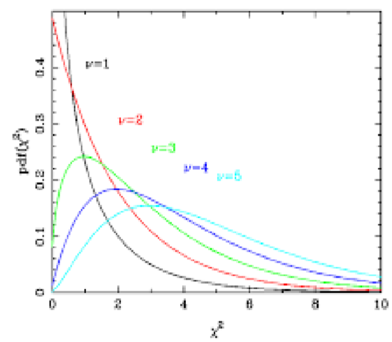
$$\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

is said to follow chi-square distribution with n d.f. and the pdf is

$$f(\chi^2)=\frac{1}{2^{n/2}\Gamma n/2}e^{-\chi^2/2}(\chi^2)^{n/2-1}, \text{for } 0\leq\chi^2<\infty;$$

Symbolically, $\chi^2 \sim \chi^2_{(n)}$ d.f.

3.1.a. Chi-square probability curve :

Chi square probability curve is drawn for various values of degrees of freedom(n). Here we observe that, for the degrees of freedom n = 1 and 2, the probability curve show decreasing in chi-square, that is chi-square distribution is severely skewed to the right, whereas, when n ≥ 3, the chi-square curves show increasing function of χ^2 and also the curves rapidly tend to be more symmetrical.	

3.1.b. Derivation of the p.d.f. of χ^2 :

Here, we shall derive the distribution of χ^2 using MGF method

Proof : We are given X_i's(i=1,2,…, n) are independent $N(\mu_i, \sigma_i^2)$ variates, we want the distribution of

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 = \sum_{i=1}^{n} Z_i^2,$$

where $Z_i = (X_i - \mu_i)/\sigma_i \sim N(0, 1)$.

Since X_i's are independent, and hence Z_i's are also independent. Therefore,

$$M_{\chi^2}(t) = M_{\sum Z_i^2}(t) = \prod_{i=1}^{n} M_{Z_i^2}(t) = [M_{Z_i^2}(t)]^n, \qquad (1)$$

where, $M_{Z_i^2}(t) = E[e^{tZ_i^2}] = \int_{-\infty}^{\infty} e^{tZ_i^2} f(x_i) dx_i$

$$= \int_{-\infty}^{\infty} e^{tZ_i^2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}[(x_i-\mu_i)/\sigma_i]^2} dx_i$$

Since $Z_i = (X_i - \mu_i)/\sigma_i$, we have

$$M_{Z_i^2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tZ_i^2} e^{-Z_i^2/2} dz_i,$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{1-2t}{2}\right)Z_i^2} dz_i$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{\left(\frac{1-2t}{2}\right)^{1/2}} = (1-2t)^{-1/2}, \qquad \text{because } \int_{-\infty}^{\infty} e^{-a^2x^2} dx = \frac{\sqrt{\pi}}{a}$$

Therefore, by equation (1), we have

$$M_{\chi^2}(t) = (1-2t)^{-n/2},$$

which is the mgf of a Gamma variate, with parameters 1/2 and n/2.

Hence, by uniqueness theorem of mgfs

$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$, is a Gamma variate with parameters1/2 and n/2.

Thus, we have the probability differential as

$$dF(\chi^2) = \frac{(1/2)^{n/2}}{\Gamma n/2} e^{-\chi^2/2} (\chi^2)^{n/2-1} d\chi^2$$

$$= \frac{1}{2^{n/2}\Gamma n/2} e^{-\chi^2/2} (\chi^2)^{n/2-1} d\chi^2, \text{for } 0 \le \chi^2 < \infty;$$

which is the required p.d. of chi-square distribution with n d.f.

4. Properties of chi-square distribution :

If , the random variable $X \sim \chi^2_{(n)}$ d.f., then

1. Mean = n and variance =2n.
2. MGF = $(1\text{-}2t)^{-n/2}$, provided $|2t| < 1$.
3. Mode = $n - 2$, for $n > 2$.
4. Sum of k independent chi-square variates with n_i d.f. is a chi-square variate with

   $n = \sum_{i=1}^{k} n_i$ d.f.
5. If $X \sim \chi^2$ with n d.f. and $Y \sim \chi^2$ with m d.f.,
   then $(X/Y) \sim \beta_2(n/2, m/2)$ and $X/(X+Y) \sim \beta_1(n/2, m/2)$.
6. Let $X \sim U(0, 1)$, then $Y = -2\log X$ has χ^2 with 2 d.f.

4.i. Moment generating Function(MGF) of χ^2 distribution:

By definition of MGF of a random variable X

$$M_X(t) = E[e^{tX}] = \int_{\infty}^{\infty} e^{tx} f(x)dx$$

Since X~ χ^2 distribution with n d.f., we have

$$M_X(t) = \int_0^{\infty} e^{tx} \frac{1}{2^{n/2}\Gamma n/2} e^{-\frac{x}{2}} x^{n/2\ -1} dx;\ \ 0 \leq x < \infty;$$

$$= \frac{1}{2^{n/2}\Gamma n/2} \int_0^{\infty} e^{-\frac{1}{2}(1-2t)x} x^{n/2\ -1} dx;\ ;$$

$$= \frac{1}{2^{n/2}\Gamma n/2} \times \frac{\Gamma n/2}{[(1-2t)/2]^{n/2}}$$

$$\therefore M_X(t) = (1-2t)]^{-n/2},\ \text{iff}\ |2t| < 1. \qquad (1)$$

Which is the required mgf of a chi-square distribution

Note : Using binomial expansion for negative index, equation (1) can be written as

$$M_X(t) = 1 + \frac{n}{2}(2t) + \frac{\frac{n}{2}(\frac{n}{2}+1)}{2!}(2t)^2 + \ldots + \frac{\frac{n}{2}(\frac{n}{2}+1)(\frac{n}{2}+2)\ldots(\frac{n}{2}+r-1)}{r!}(2t)^r.$$

Therefore,

$\mu_r' =$ coefficient of $\frac{t^r}{r!}$ in the expansion of $M_X(t)$

$$=2^r \tfrac{n}{2}(\tfrac{n}{2}+1)(\tfrac{n}{2}+2)\ldots(\tfrac{n}{2}+r-1)$$

$=n(n+2)(n+4)\ldots.(n+2r-2)$

When r = 1 then

$\mu_1' =$ coefficient of 't' in the expansion of $M_X(t)$

$= n =$ mean

When r = 2

$\mu_2' =$ coefficient of '$t^2/2$' in the expansion of $M_X(t)$

$= n(n+2)$

Therefore, Variance= $\mu_2' - (\mu_1')^2$

$= n(n+2) - n^2 = 2n.$

Similarly, the remaining higher order moments can be obtained.

4.ii. Limiting form of χ^2 Distribution for large n(n-- >∞):

Let X~ χ^2 distribution with n d.f. then the mgf

$$M_X(t) = (1-2t)]^{-n/2}, \text{ iff } |2t|<1.$$

The MGF of standard χ^2 variate say $Z=(x-n)/\sqrt{2n}$ is

$$M_Z(t) = e^{-nt/\sqrt{2n}}[(1 - t\sqrt{\tfrac{2}{n}})]^{-n/2}, \text{ iff } |2t|<1.$$

Taking logarithm on both sides of $M_Z(t)$, we get

$$Log\,M_Z(t) = -t\sqrt{\tfrac{n}{2}} - \frac{n}{2}\log(1-t\sqrt{\tfrac{2}{n}}),$$

Since, $\log(1-x)= - x - x^2/2 - x^3/3 \ \ldots$; so, we have

$$= -t\sqrt{\tfrac{n}{2}} + \frac{n}{2}\left[t\sqrt{\tfrac{2}{n}} + \frac{t^2}{2}\frac{2}{n} + \frac{t^3}{3}\left(\frac{2}{n}\right)^{3/2} + ... \right]$$

$$= -t\sqrt{\tfrac{n}{2}} + t\sqrt{\tfrac{n}{2}} + \frac{t^2}{2} + O(n^{-1/2})$$

$$= \frac{t^2}{2} + O(n^{-1/2})$$

Where $O(n^{-1/2})$ be the order of n containing $n^{1/2}$ and higher powers of n in the denominator.

$$\therefore \lim_{n\to\infty} \log M_Z(t) = \frac{t^2}{2}$$

$$\Rightarrow \lim_{n\to\infty} M_Z(t) = e^{\frac{t^2}{2}},$$

which is the MGF of a standard normal variate. Thus by uniqueness theorem of mgf, χ^2 distribution tends to normal distribution asymptotically. That is, as n-- >∞, χ^2 distribution tends to normal distribution asymptotically.

4.iii. Additive or reproductive property of χ^2 distribution:

The sum of independent χ^2 variates is also a chi-square variate. That is, if X_i (i=1,2,…,k) are k - independent χ^2 variates with n_i d.f. respectively then, the sum $\sum_{i=1}^{k} X_i$ is also a chi-square variate with $n = \sum_{i=1}^{k} n_i$ d.f.

Proof:

Given X~ χ^2 distribution with n d.f. then by definition of mgf,

$$M_{X_i}(t) = (1-2t)]^{-n_i/2}, \text{ iff } |2t|<1 \text{ and } \forall\, i=1,2,..k$$

Then by uniqueness theorem of mgf, when X_i's are independent, we have

$$M_{\sum_{i=1}^{k} X_i}(t) = \prod_{i=1}^{k} M_{X_i}(t) = [M_{X_i}(t)]^k$$

$$= (1-2t)^{-n_1/2} \times (1-2t)^{-n_2/2} \times ... \times (1-2t)^{-n_k/2}$$

$$=(1-2t)^{-(n_1+n_2+..+n_k)/2}=(1-2t)^{-\frac{1}{2}\sum_{i=1}^{k}n_i}$$

Which is the mgf of a chi-square variate with $n=\sum_{i=1}^{k}n_i$ d.f.

Note : Converse of the above result is also true. That is if X_i 's(i=1,2,…,k) are χ^2 variates with n_i d.f. respectively and if the sum $\sum_{i=1}^{k}X_i$ is a chi square variate with $n=\sum_{i=1}^{k}n_i$ d.f. then X_i's are independent.

4.iv. The theorem:

Theorem 1: Independence of a sample mean and a variance in random sampling from a normal population.

Let $x_1, x_2, \ldots, x_n$ be a random sample from normal population with mean μ and variance σ^2, then

i) $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ and

ii) $\frac{ns^2}{\sigma^2} = \sum_{i=1}^{n}(x_i - \bar{x})^2$ = (n-1)S^2 is a chi-square variate with (n-1)d.f. and these two are independently distributed.

Proof: The joint probability differential of $x_1, x_2, \ldots, x_n$ is given by

$$dP(x_1.x_2...x_n)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_i)^2}dx_1dx_2...dx_n \quad -\infty< x_i <\infty;\ \forall\ i=1,2,...,n.$$

Consider the transformation to the variables Y_i(i = 1,2,…,n)by means of a linear orthogonal transformation Y = AX, where

$$A = \begin{bmatrix} a_{11} & a_{12} & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & a_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1} & a_{n2} & . & . & a_{nn} \end{bmatrix}$$

In particular, $a_{11} = a_{12} = \ldots = a_{1n} = 1/\sqrt{n}$, which

=> $y_1 = (1/\sqrt{n})(x_1 + x_2 + \ldots, +x_n) = \sqrt{n}\,\bar{x}$ (1)

Then, $dy_1 = \sqrt{n}\, d\bar{x}$

It can be easily shown that, the above choice of $a_{11}, a_{12}, \ldots, a_{1n}$ satisfies the condition of orthogonality, so that, $\sum_{j=1}^{n} a_{ij}^2 = 1$. Since, the transformation is orthogonal, we have

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n\bar{x}^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + y_1^2, \quad \text{from (1)}$$

$$=> \sum_{i=2}^{n} y_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2. \qquad \textbf{(2)}$$

and

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2 = . \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 = \sum_{i=2}^{n} y_i^{\ 2} + n(\bar{x} - \mu)^2, \ \text{from(2)}$$

Now $|A'A| = |I_n| = 1$, and therefore the jacobian transformation J = ± 1. Thus the joint probability differential of $(y_1, y_2, \ldots, y_n)$ is given by

$$dG(y_1.y_2 \ldots y_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . e^{-\frac{1}{2\sigma^2}\left(\sum_{i=2}^{n} y_i^{\ 2} + n(\bar{x} - \mu)^2\right)} |J| dy_1.dy_2 \ldots . dy_n \quad .$$

$$= \frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}} e^{-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}} d\bar{x}. \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} e^{-\frac{1}{2\sigma^2}\sum_{i=2}^{n} y_i^2} dy_2.dy_3 \ldots dy_n$$

Therefore, we have on simplification,

$g(y_1.y_2.y_3....y_n) = g(y_1).g(y_2.y_3....y_n)$ => y_1 and $(y_2,y_3,....,y_n)$ are independent

where $g(y_1)$ is the p.d.f of $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ and

$$\sum_{i=2}^{n} y_i^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 = ns^2 = (n-1)S^2$$

are independently distributed. Moreover, $\sum_{i=2}^{n} y_i^2 / \sigma^2 = ns^2 / \sigma^2 \sim \chi^2_{(n-1)}\ d.f.$

4.v. Applications of chi-square distribution :

chi-square distribution has large number of applications, some of them are listed below:

1. To test the significance of population variance $\sigma^2 = \sigma_0^2$.
2. To test the goodness of fit.
3. To test the independence of attributes.
4. To test the homogeneity of independent estimates of the population variance.
5. To test the homogeneity of independent estimates of the population correlation coefficient.

5. Summary :

We have discussed thoroughly the definition, probability curve, derivation of the pdf, MGF and proved, some of the properties of Chi-square distribution. Also, we have listed out, some of the applications of Chi-square distribution.