Sampling distributions

Dr.E.S.Jeevanand
Associate Professor
U C College, Aluva

Statistics

Any function of the sample values is known as a statistics. Eg: Sample mean, Sample median, Sample variance etc. are all statistics.

Sampling Distribution

- A sampling distribution is a distribution of a statistic over all possible samples. That is sampling distribution is the probability distribution of the statistics.
- The commonly used sampling distributions are, Normal distribution, t distribution, χ^2 distribution and F distribution.

Students t distribution

- A continues random variable *t* is said to follow a student's t distribution with *n* degrees of freedom if its pdf is given by

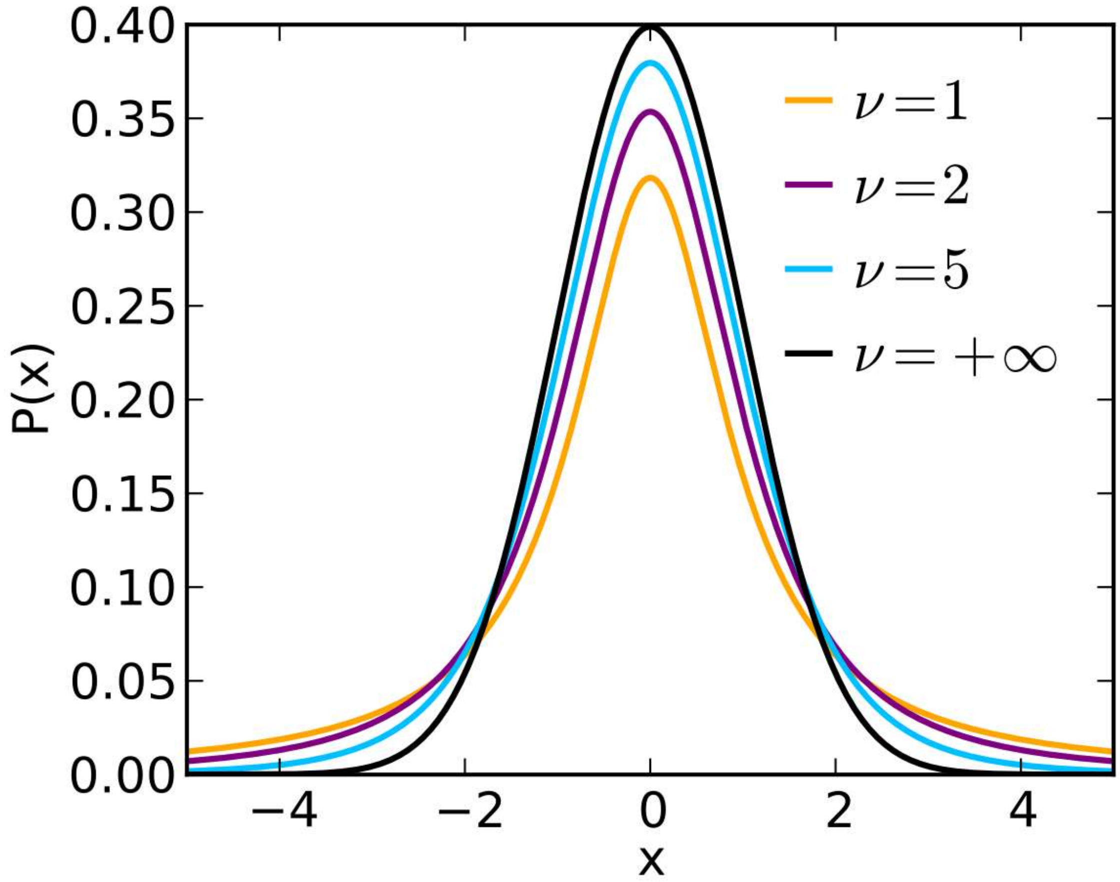
$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}, \quad -\infty < t < \infty$$

or

- $f(t) = \frac{1}{\sqrt{n}\mathrm{B}\left(\frac{1}{2},\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}, -\infty < t < \infty$

- the mean of the distribution is zero and variance is $\frac{n}{n-2}$ (n>2).

Image of T distribution



Properties

- A t-distribution is like a Z distribution, except has slightly fatter tails to reflect the uncertainty added by estimating σ.
- It is unimodel distribution
- It is a symmetric distribution (i.e. $\beta_1 = 0$)
- It is mesocourtic (i.e. $\beta_2 = 3$)
- The bigger the sample size (i.e., the bigger the sample size used to estimate σ), then the closer t becomes to Z.
- If n>100, t approaches Z.

Application or Uses of t-distribution

- To find out the confidence interval for population mean of a normal distribution when population standard deviation σ unknown
- To test sample mean $\bar{x}$ differs significantly from the hypothetical value of the population mean μ when population standard deviation σ unknown
- To the significance of the difference between two population means when population standard deviation σ unknown
- To test the significance of correlation coefficients and regression coefficients

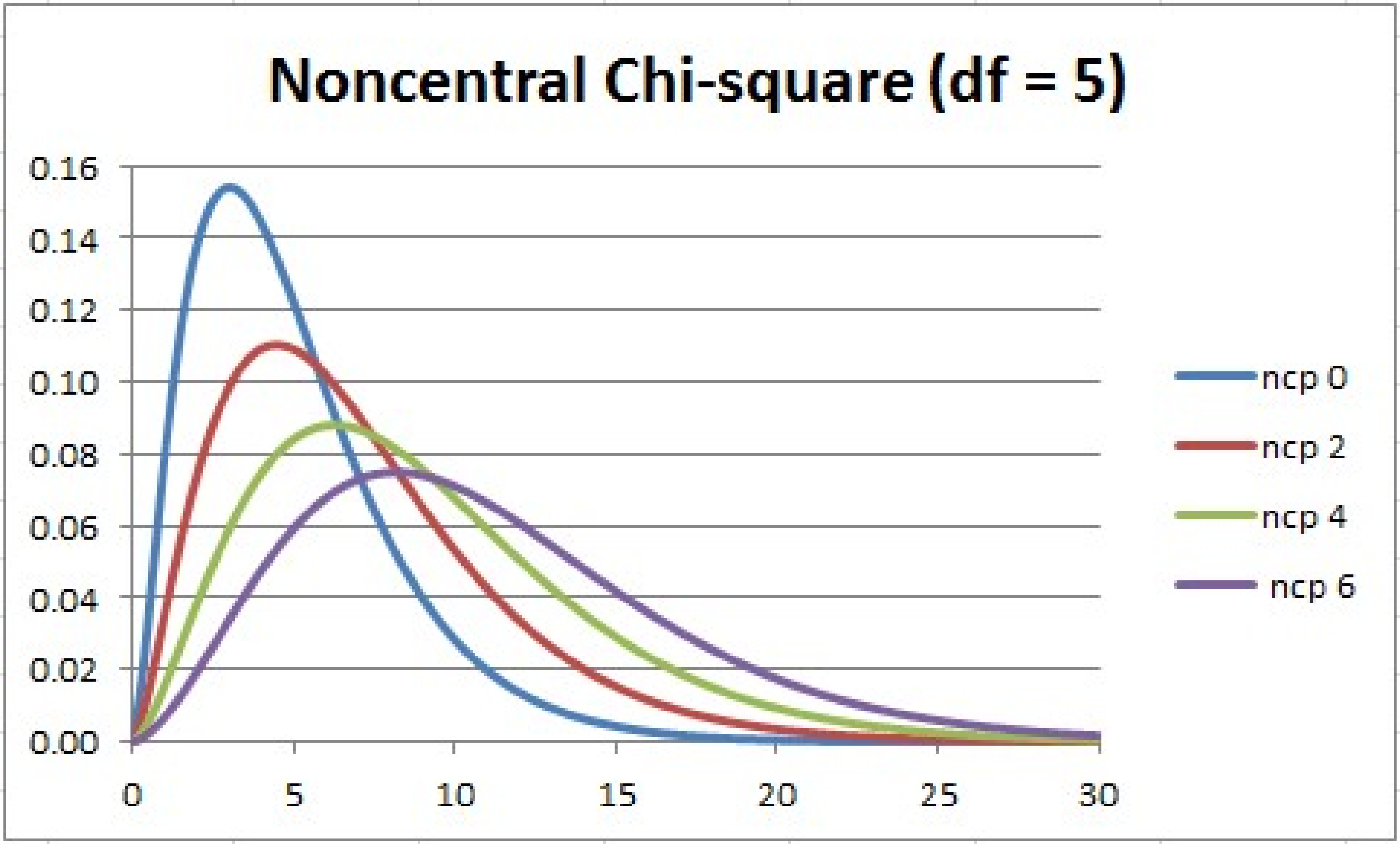
Chi-Square distribution

A continues random variable χ^2 is said to follow a chi-square distribution with *n* degrees of freedom if its pdf is given by

$$f(\chi^2) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\ e^{-\frac{\chi^2}{2}}\ (\chi^2)^{\left(\frac{n}{2}-1\right)},\ \ 0 \le \chi^2 < \infty.$$

the mean of the distribution is *n* and variance is 2*n*.

Image of χ^2 - distribution



Properties

- It is positively skewed distribution
- It is unimodel distribution
- If U and V are independent chi-square variables with n_1 and n_2 degrees of freedom then Z = U+V has chi-square variables with $n_1 + n_2$ degrees of freedom.
- If X is a standard normal variate (i.e. $X \sim N(0,1)$) then $Y = X^2$ is a χ^2 variate with 1 degree of freedom.

Application or Uses of χ^2 - distribution

1. To find out the confidence interval for population variance of a normal distribution
2. To test the hypothetical value of the population variance (i.e. to test whether $\sigma^2 = \sigma_0^2$)
3. To test the goodness of fit
4. To test the independence of attribute
5. To test the homogeneity of independent estimate of the population variance
6. To test the homogeneity of independent estimate of the population correlation coefficients.

Sendecor's F distribution

- A continues random variable *F* is said to follow a F distribution with (n_1 , n_2) degrees of freedom if its pdf is given by
- $f(F) = \dfrac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \dfrac{F^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}}, \ 0 \leq F < \infty.$

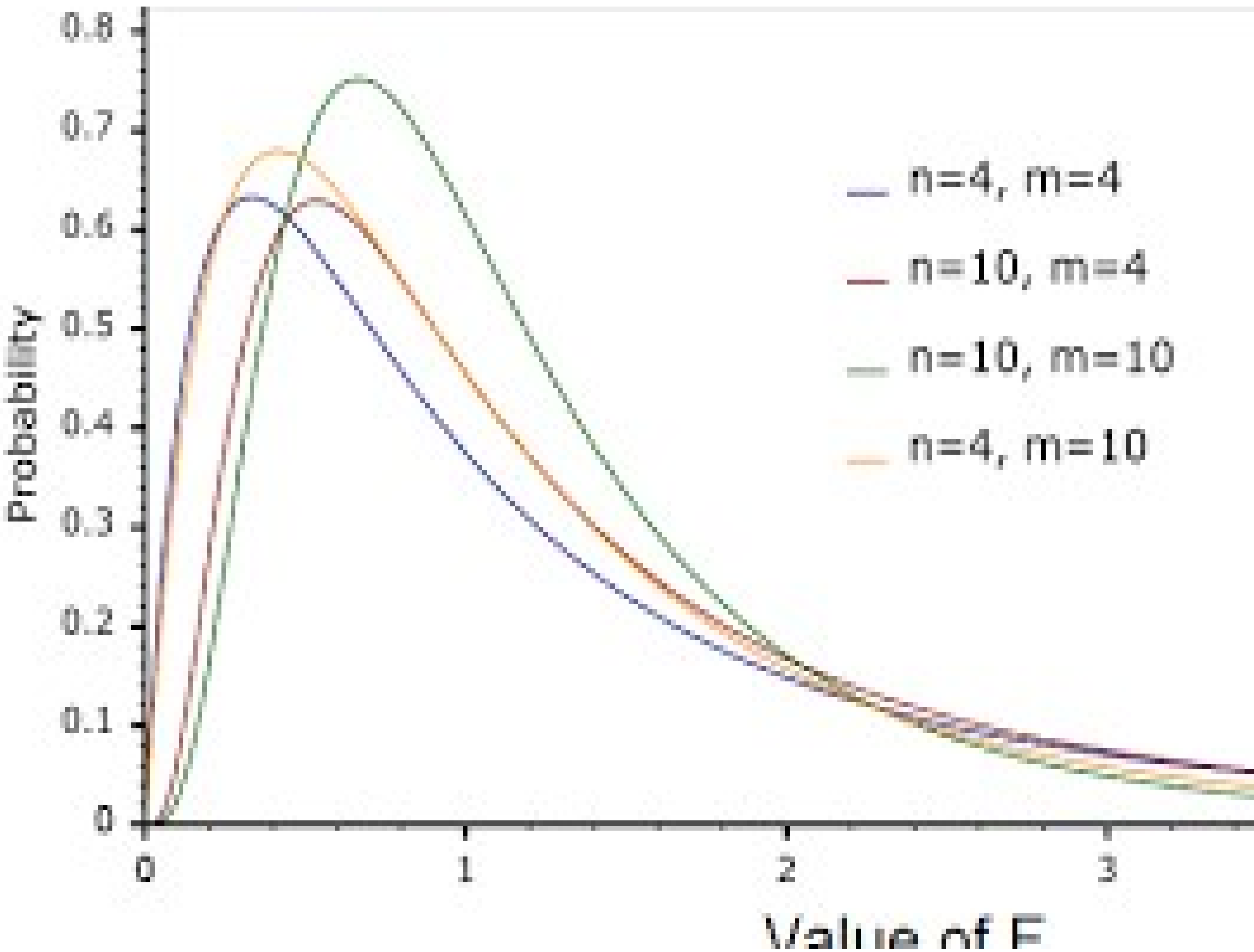
Mean and variance of F distribution

the mean of the distribution is $\frac{n_2}{n_2-2}$,

$n_2>2$ and

variance is $\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ $n_2>4$.

Image of F distribution



Properties

- It is positively skewed distribution
- It is unimodel distribution
- If U and V are independent chi-square variables with n_1 and n_2 degrees of freedom then F = $\frac{U/n_1}{V/n_2}$ has F distribution with (n_1,n_2) degrees of freedom.
- If t is a student's t variate with n degree of freedom then t^2 follows F distribution with $(1,n)$ degrees of freedom.

Application or Uses of F - distribution

1. To test the equality of two population variances (i.e. to test whether $\sigma_1^2 = \sigma_2^2$)
2. To test the significant of an observed multiple correlation
3. To test the significant of correlation ratio
4. To test the model fit or Linearity of Regression
5. To test the equality of several means

Relation between sampling distributions

- If X is a standard normal variate (i.e. $X \sim N(0,1)$) then $Y = X^2$ is a χ^2 variate with 1 degree of freedom.

- If X is a standard normal variate (i.e. $X \sim N(0,1)$) and Y is a χ^2 variate with n degree of freedom, then $t = \frac{X}{\sqrt{\frac{Y}{n}}}$ has t distribution with n degree of freedom.

- If U and V are independent chi-square variables with n_1 and n_2 degrees of freedom then $F = \frac{U/n_1}{V/n_2}$ has F distribution with (n_1, n_2) degrees of freedom.
- If t is a student's t variate with n degree of freedom then t^2 follows F distribution with $(1, n)$ degrees of freedom.

Standard Error

The standard deviation of the sampling distribution of a statistic is called standard error (SE)

Uses of Standard Error

1. It is used for finding confidence intervals
2. It is used for testing a given hypothesis
3. It gives an idea about the unreliability of the sample. Reciprocal of SE is taken as a measure of reliability.

Sampling distribution of the variance of sample chosen from a Normal population N(μ,σ)

Let x_1, x_2, …, x_n be a sample chosen from Normal population N(μ,σ) and let $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$. Define $u = \frac{ns^2}{\sigma^2}$ then *u* has chi-square variables with *n* -1 degrees of freedom.

Sampling distribution of the mean of sample chosen from a Normal population N(μ,σ)

- Let $x_1, x_2, \ldots, x_n$ be a sample chosen from Normal population N(μ,σ) we requires the distribution of
- $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

Case-I when σ known

In this case $\bar{x}$ has normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (i.e $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$).

Proof:

Given $x_i \sim N(\mu, \sigma)$ then $M_{x_i}(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}$ and

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

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

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

which is the mgf of $N(\mu, \frac{\sigma}{\sqrt{n}})$ so $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

Case-II when σ unknown

In this case $t = \frac{(\bar{x}-\mu)\sqrt{n-1}}{S}$ has Students t distribution with '*n*-1' degrees of freedom.

Proof

We have already proved that $\bar{x} \sim$ N(μ,$\frac{\sigma}{\sqrt{n}}$)

So $u = \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma}$ has N(0,1) distribution. Also $Y = \frac{ns^2}{\sigma^2}$ has chi-square distribution with *n* -1 degrees of freedom. Hence $t = \frac{U}{\sqrt{\frac{Y}{n-1}}}$ has t distribution with *n*-1 degree of freedom

ie. $t = \frac{\frac{(\bar{x}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{\frac{ns^2}{\sigma^2}}{n-1}}} = \frac{(\bar{x}-\mu)\sqrt{n-1}}{S} \sim t_{n-1}$.

The End

- **Contact details** **esjmaths@gmail.com**
- **9847915101, 8089684892**