χ^2- distribution

It was initially discovered by Helmert in 1875 and was again defined independently in 1900 by Karl Pearson, who gave this notation also.

If $X_1, X_2, \cdots, X_n$ are n independent normal variates with means $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, then the standard normal variates are

$$Z_i = \frac{X_i - E(X_i)}{\sqrt{Var(X_i)}} = \frac{X_i - \mu_i}{\sigma_i} \quad ; i = 1, 2, \cdots, n$$

The sum square of a standard normal variates is known as χ^2-variate with n degree of freedom, i.e.

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

and its probability density function is given by

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

Derivation of χ^2- distribution

If $X_1, X_2, \cdots, X_n$ are n independent normal variates with means $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, then the

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

Since X_i's $(i = 1, 2, \cdots, n)$ are independent, so Z_i's $(i = 1, 2, \cdots, n)$ are also independent, then the moment generating function of χ^2 about origin is

$$M_{\chi^2}(t) = E\left[e^{t\chi^2}\right]$$

$$= E\left[e^{t\sum_{i=1}^{n} Z_i^2}\right]$$

$$= E\left[e^{t(Z_1^2 + Z_2^2 + \cdots + Z_n^2)}\right]$$

$$= E\left[e^{tZ_1^2} \cdot e^{tZ_2^2} \cdots\cdots e^{tZ_n^2}\right]$$

$$= E(e^{tZ_1^2}) \cdot E(e^{tZ_2^2}) \cdots\cdots E(e^{tZ_n^2})$$

$$= M_{Z_1^2}(t) \cdot M_{Z_2^2}(t) \cdots\cdots M_{Z_n^2}(t)$$

$$= \prod_{i=1}^{n} M_{Z_i^2}(t)$$

$$= \left[M_{Z_i^2}(t)\right]^n$$

Now

$$M_{Z_i^2}(t) = E\left(e^{tZ_i^2}\right)$$

$$= \int_{-\infty}^{\infty} e^{tZ_i^2} f(Z_i)\, dZ_i$$

$$= \int_{-\infty}^{\infty} e^{tZ_i^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z_i^2}\, dZ_i$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tZ_i^2 - \frac{1}{2}Z_i^2}\, dZ_i$$

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

$$= \frac{\cancel{2}}{\sqrt{\cancel{2}\pi}} \int_{0}^{\infty} e^{-(1-2t)\frac{Z_i^2}{2}}\, dZ_i$$

$$= \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} e^{-(1-2t)\frac{Z_i^2}{2}}\, dZ_i$$

put $(1-2t)\frac{Z_i^2}{2} = T$

$$Z_i = \sqrt{\frac{2T}{1-2t}}$$

$$dZ_i = \frac{1}{\cancel{2}}\left(\frac{2T}{1-2t}\right)^{-\frac{1}{2}} \left(\frac{\cancel{2}dT}{1-2t}\right) = \frac{dT}{\sqrt{2T(1-2t)}}$$

If $Z_i \to 0$ then $T \to 0$ and $Z_i \to \infty$ then $T \to \infty$, so

$$M_{Z_i^2}(t) = \sqrt{\frac{\cancel{2}}{\pi}} \int_0^{\infty} e^{-T} \frac{dT}{\sqrt{\cancel{2}T(1-2t)}}$$

$$= \frac{1}{\sqrt{\pi}\sqrt{1-2t}} \int_0^{\infty} e^{-T}\, T^{-1/2}\, dT$$

$$= \frac{1}{\sqrt{\pi}\sqrt{1-2t}} \int_0^{\infty} e^{-T}\, T^{\frac{1}{2}-1}\, dT$$

$$= \frac{1}{\sqrt{\pi}(1-2t)^{\frac{1}{2}}} \Gamma\tfrac{1}{2}$$

$$= \frac{1}{\cancel{\sqrt{\pi}}(1-2t)^{1/2}} \cancel{\sqrt{\pi}}$$

$$= (1-2t)^{-1/2}$$

Therefore

$$M_{\chi^2}(t) = \left[M_{Z_i^2}(t)\right]^n$$

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

which is the m.g.f. of gamma variate with parameters $\frac{1}{2}$ and $\frac{n}{2}$.
Hence by uniqueness theorem of m.g.f. the

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

is a gamma variate with parameters $\frac{1}{2}$ and $\frac{n}{2}$. Then the probability differential function of χ^2 is

$$dP(\chi^2) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\frac{n}{2}} \cdot e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1} d\chi^2$$

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

proved

Properties of χ^2-distribution

(1) Moment generating function

The moment generating function about origin is

$$M_{\chi^2}(t) = E(e^{t\chi^2})$$

$$= \int_0^\infty e^{t\chi^2} f(\chi^2)\, d\chi^2$$

$$= \int_0^\infty e^{t\chi^2} \cdot \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1} d\chi^2$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \int_0^\infty e^{t\chi^2 - \chi^2/2} (\chi^2)^{\frac{n}{2}-1} d\chi^2$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \int_0^\infty e^{-(1-2t)\frac{\chi^2}{2}} (\chi^2)^{\frac{n}{2}-1} d\chi^2$$

put $(1-2t)\frac{\chi^2}{2} = T$

$$\chi^2 = \frac{2T}{(1-2t)}$$

$$d\chi^2 = \frac{2\,dT}{(1-2t)}$$

If $\chi^2 \to 0$ then $T \to 0$ and $\chi^2 \to \infty$ then $T \to \infty$, so

$$M_{\chi^2}(t) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \int_0^\infty e^{-T} \left(\frac{2T}{1-2t}\right)^{\frac{n}{2}-1} \left(\frac{2\,dT}{1-2t}\right)$$

$$= \frac{\cancel{2^{\frac{n}{2}}}}{\cancel{2^{\frac{n}{2}}}\Gamma\frac{n}{2}(1-2t)^{n/2}} \int_0^\infty e^{-T}\, T^{\frac{n}{2}-1} dT$$

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

The r^{th} moment about origin is

$$\mu'_r = \frac{\partial^r}{\partial t^r}\left[M_{\chi^2}(t)\right]_{t=0}$$

In particular,

$$\mu'_1 = \frac{\partial}{\partial t}\left[M_{\chi^2}(t)\right]_{t=0}$$

$$= \frac{\partial}{\partial t}\left[(1-2t)^{-n/2}\right]_{t=0}$$

$$= \left[-\frac{n}{\cancel{2}}(1-2t)^{-\frac{n}{2}-1}(-\cancel{2})\right]_{t=0}$$

$$= \left[n(1-2t)^{-\frac{n}{2}-1}\right]_{t=0}$$

$$= \left[n(1-0)^{-\frac{n}{2}-1}\right] = n = \text{Mean}$$

$$\mu'_2 = \frac{\partial^2}{\partial t^2}\left[M_{\chi^2}(t)\right]_{t=0}$$

$$= \frac{\partial}{\partial t}\left[\frac{\partial}{\partial t}\left\{M_{\chi^2}(t)\right\}\right]_{t=0}$$

$$= \frac{\partial}{\partial t}\left[n(1-2t)^{-\frac{n}{2}-1}\right]_{t=0}$$

$$= \left[n\left(-\frac{n}{2}-1\right)(1-2t)^{-\frac{n}{2}-2}(-2)\right]_{t=0}$$

$$= \left[2n\left(\frac{n}{2}+1\right)(1-2t)^{-\frac{n}{2}-2}\right]_{t=0}$$

$$= \left[\frac{\cancel{2}n(n+2)}{\cancel{2}}(1-2t)^{-\frac{n}{2}-2}\right]_{t=0}$$

$$= \left[n(n+2)(1-0)^{-\frac{n}{2}-2}\right]$$

$$= n(n+2)$$

Then

$$\text{Variance} = \mu_2 = \mu'_2 - \mu_1'^2$$

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

$$= \cancel{n^2} + 2n - \cancel{n^2} = 2n$$

hence

$$\text{Variance} = 2(\text{Mean}) \qquad (\text{Since mean} = n)$$

(2) Mode Mode is that value of the variate for which $f(\chi^2)$ is maximum, i.e, mode is the solution of

$$f'(\chi^2) = 0 \text{ and } f''(\chi^2) < 0$$

Therefore

$$f'(\chi^2) = \frac{d}{d\chi^2} f(\chi^2) = 0$$

$$\Rightarrow \frac{d}{d\chi^2}\left[\frac{1}{2^{n/2}\,\Gamma\frac{n}{2}}\, e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1}\right] = 0$$

$$\Rightarrow \frac{1}{2^{n/2}\,\Gamma\frac{n}{2}} \frac{d}{d\chi^2}\left[e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1}\right] = 0$$

$$\Rightarrow \left[e^{-\chi^2/2}\left(\frac{n}{2}-1\right)(\chi^2)^{\frac{n}{2}-2} + (\chi^2)^{\frac{n}{2}-1} e^{-\chi^2/2}\left(-\frac{1}{2}\right)\right] = 0$$

$$\Rightarrow \frac{(n-2)}{\cancel{2}} \cancel{e^{-\chi^2/2}} (\chi^2)^{\frac{n}{2}-2} = \frac{1}{\cancel{2}} \cancel{e^{-\chi^2/2}} (\chi^2)^{\frac{n}{2}-1}$$

$$\Rightarrow \frac{(\chi^2)^{\frac{n}{2}-2}}{(\chi^2)^{\frac{n}{2}-1}} = \frac{1}{(n-2)}$$

$$\Rightarrow (\chi^2)^{\cancel{\frac{n}{2}}-2-\cancel{\frac{n}{2}}+1} = \frac{1}{(n-2)}$$

$$\Rightarrow (\chi^2)^{-1} = \frac{1}{(n-2)}$$

$$\Rightarrow \chi^2 = (n-2) = \text{mode value}$$

(3) <u>Coefficient of Skewness</u>

Karl Pearson's Coefficient of Skewness is

$$S_k = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

$$= \frac{n-(n-2)}{\sqrt{2n}} = \frac{\cancel{n}-\cancel{n}+2}{\sqrt{2n}} = \frac{\cancel{2}}{\cancel{\sqrt{2n}}} = \sqrt{\frac{2}{n}}$$

(4) <u>Additive Property</u> The Sum of independent χ^2-variates is also a χ^2-variate. In other words, if $\chi_1^2, \chi_2^2, \dots, \chi_k^2$ are k independent χ^2-variates with $n_1, n_2, \dots, n_k$ degree of freedom respectively, then the $\chi_1^2 + \chi_2^2 + \dots + \chi_k^2$ is also a χ^2-variate with $n_1+n_2+\dots+n_k$ degree of freedom.

<u>Proof</u> If $\chi_1^2, \chi_2^2, \dots, \chi_k^2$ are k independent χ^2-variates with $n_1, n_2, \dots, n_k$ degree of freedom respectively, then their mgfs are

$$M_{\chi_1^2}(t) = (1-2t)^{-n_1/2}$$

$$M_{\chi_2^2}(t) = (1-2t)^{-n_2/2}$$

$$\dots\dots\dots$$

$$M_{\chi_k^2}(t) = (1-2t)^{-n_k/2}$$

then

$$M_{(\chi_1^2+\chi_2^2+\cdots+\chi_k^2)}(t) = E\left[e^{t(\chi_1^2+\chi_2^2+\cdots+\chi_k^2)}\right]$$
$$= E\left[e^{t\chi_1^2} \cdot e^{t\chi_2^2} \cdots\cdots e^{t\chi_k^2}\right]$$
$$= E\left(e^{t\chi_1^2}\right) \cdot E\left(e^{t\chi_2^2}\right) \cdots\cdots E\left(e^{t\chi_k^2}\right)$$

(Since $\chi_1^2, \chi_2^2, \cdots, \chi_k^2$ are independent,

$$= M_{\chi_1^2}(t) \cdot M_{\chi_2^2}(t) \cdots\cdots M_{\chi_k^2}(t)$$
$$= (1-2t)^{-n_1/2} (1-2t)^{-n_2/2} \cdots\cdots (1-2t)^{-n_k/2}$$
$$= (1-2t)^{-(n_1+n_2+\cdots+n_k)/2}$$

which is the m.g.f. of χ^2-variate with $(n_1+n_2+\cdots+n_k)$ degree of freedom. Hence by uniqueness theorem of m.g.f., $\chi_1^2+\chi_2^2+\cdots+\chi_k^2$ is also a χ^2-variate with $n_1+n_2+\cdots+n_k$ degree of freedom.

proved

(5) Limiting property

χ^2-distribution tends to a normal distribution as $n \to \infty$.

Proof. We know that the mean and variance of χ^2-distribution with n degree of freedom is given by

$$E(\chi^2) = n \quad \text{and} \quad Var(\chi^2) = 2n$$

then we define a standard χ^2-variate

$$Z = \frac{\chi^2 - E(\chi^2)}{\sqrt{Var(\chi^2)}} = \frac{\chi^2 - n}{\sqrt{2n}}$$

Therefore

$$M_Z(t) = E\left(e^{tZ}\right)$$
$$= E\left[e^{t\left(\frac{\chi^2-n}{\sqrt{2n}}\right)}\right]$$
$$= E\left[e^{\frac{t\chi^2}{\sqrt{2n}} - \frac{nt}{\sqrt{2n}}}\right]$$
$$= E\left[e^{\frac{t\chi^2}{\sqrt{2n}} - t\sqrt{\frac{n}{2}}}\right]$$
$$= E\left[e^{\frac{t\chi^2}{\sqrt{2n}}} \, e^{-t\sqrt{\frac{n}{2}}}\right]$$
$$= e^{-t\sqrt{\frac{n}{2}}} \, E\left[e^{\frac{t\chi^2}{\sqrt{2n}}}\right]$$

$$= e^{-t\sqrt{\frac{n}{2}}} M_{\chi^2}\left(\frac{t}{\sqrt{2n}}\right)$$

$$= e^{-t\sqrt{\frac{n}{2}}} \left(1-2\cdot\frac{t}{\sqrt{2n}}\right)^{-n/2}$$

$$= e^{-t\sqrt{\frac{n}{2}}} \left(1-t\sqrt{\frac{2}{n}}\right)^{-n/2}$$

taking log both the sides

$$\log M_Z(t) = -t\sqrt{\frac{n}{2}} - \frac{n}{2}\log\left(1-t\sqrt{\frac{2}{n}}\right)$$

$$= -t\sqrt{\frac{n}{2}} - \frac{n}{2}\left[-t\sqrt{\frac{2}{n}} - \frac{t^2}{2}\left(\frac{2}{n}\right) - \frac{t^3}{3}\left(\frac{2}{n}\sqrt{\frac{2}{n}}\right) - \cdots\right]$$

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

$$= \cancel{-t\sqrt{\frac{n}{2}}} + \cancel{t\sqrt{\frac{n}{2}}} + \frac{t^2}{2} + \frac{t^3}{3}\left(\sqrt{\frac{2}{n}}\right) + \cdots$$

$$= \frac{t^2}{2} + \frac{t^3}{3}\left(\sqrt{\frac{2}{n}}\right) + \cdots$$

as $n \to \infty$, then

$$\lim_{n\to\infty} \log M_Z(t) = \lim_{n\to\infty}\left[\frac{t^2}{2} + \frac{t^3}{3}\left(\sqrt{\frac{2}{n}}\right) + \cdots\right]$$

$$= \frac{t^2}{2} + 0 + \cdots$$

$$= \frac{t^2}{2}$$

Therefore

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

which is the m.g.f. of standard normal variate. Hence by uniqueness theorem of m.g.f., standard χ^2-variate tends to standard normal variate as $n \to \infty$. In otherwords, χ^2-distribution tends to normal distribution as $n \to \infty$.

Proved

Features of χ^2- Curve

The Curve

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1} \quad ; \quad 0 \le \chi^2 < \infty$$

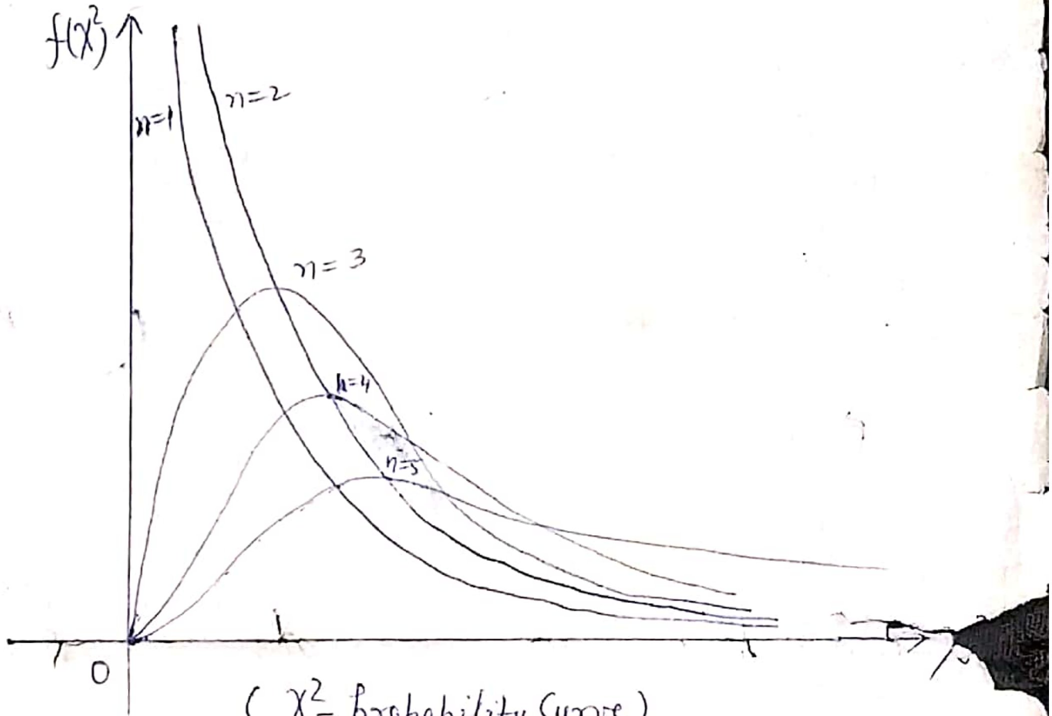
has the following main features:

(1) The distribution has no parameter and its shape depends upon the degree of freedom.

(2) When the degree of freedom is less than 2, i.e. $n \le 2$, the density decreases very fast. $(n-2)$ is the mode value of the distribution. So for $n>2$, it is always positively skewed. For $n=2$ it is exponential distribution.

(3) Skewness decreases or the distribution becomes more and more Symmetrical as the number of degree of freedom (n) increases.

(4) The curve starts tangentially from the origin ($\chi^2 = 0$) rises to the maximum value (where χ^2 takes mode value) and then falls slowly to become again tangential at ∞. The curve has no mode value for $n=1$



(χ^2- probability Curve)