χ^2 distribution

This is the distribution of sum of squares of independent random variables *which are themselves distributed* by a Normal distribution with mean 0 and standard deviation 1.

$z_i \sim \mathrm{N}(0,1)$

$$x = \sum_{i=1}^{n} z_i^2$$

$x \sim \chi^2(k)$; $k \le n$

$$p(x) = \frac{x^{\frac{1}{2}k-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}k}\Gamma\left(\frac{1}{2}k\right)}$$

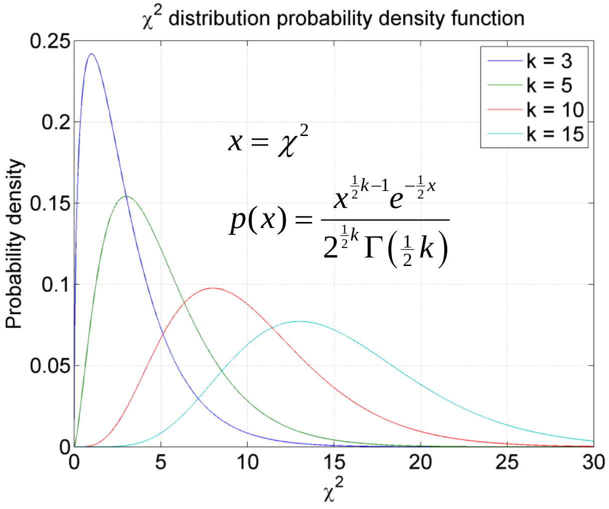
$F(x,k) = \gamma\left(\frac{1}{2}x, \frac{1}{2}k\right)$

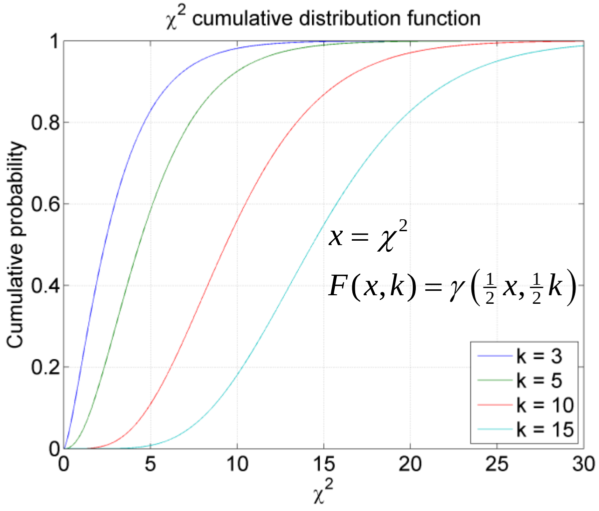
$F^{-1}(x,k) = 2\gamma^{-1}\left\{x, \frac{1}{2}k\right\}$

$M_x(t) = (1-2t)^{-\frac{1}{2}k}$; $t < \frac{1}{2}$

$\mu = k$ mean

$\sigma^2 = 2k$ variance

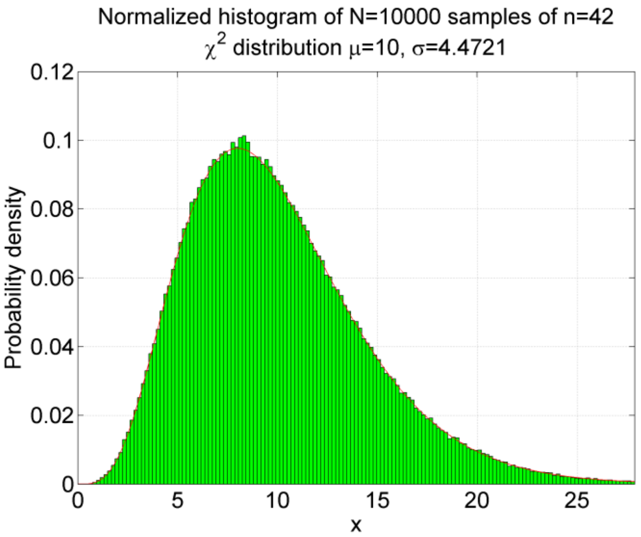




Cumulative distribution function

$F(x,k) = P(u \le x)$

$$F(x,k) = \int_0^{\infty} p(u)du$$



This distribution is useful in **comparing measurements** Y_i to **theoretical predictions** y_i. If the measurements are independent, the random variable z_i , given measurement *i* of *n*, might be expected to be distributed by N(0,1).

$$\boxed{z_i = \frac{(Y_i - y_i)^2}{y_i}}$$

Hence we expect the sum of these values to be distributed by the **χ^2 distribution.** $x = \sum_{i=1}^{n} z_i^2 \sim \chi^2(k)$

i.e. have probability density $$\boxed{p(x) = \frac{x^{\frac{1}{2}k-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}k}\Gamma\left(\frac{1}{2}k\right)}}$$

k is the number of 'degrees of freedom'. If the data has been used to create the theoretical prediction, then *k* may be *less* than the number of measurements (and corresponding predictions) *n*.

For example, if a data sample represents the expected frequencies of goals per game in a UK Premiership football team, a *Poisson* distribution might be an appropriate model. This distribution requires a single parameter, which is the mean number of goals per game in this case. Therefore $k = n$ -1 in this example, as the data will be used to compute the mean. If a Normal distribution is used, $k = n$ - 2 since the data would be used to find *both* mean and variance of the distribution.

$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ Gamma function $\gamma(x,a) = \frac{1}{\Gamma(a)}\int_0^{x} t^{a-1}e^{-t}dt$ Lower incomplete gamma function

χ^2 test example

The frequency of goals scored per match by Liverpool football team in the UK Premiership 1995-2014, is predicted to be *Poisson* distributed. A frequency table for the number of goals scored is given below.

Goals per game y	Frequency f	Expected frequency f_E
0	56	51.0701
1	97	99.8767
2	90	97.6634
3	66	63.6661
4	34	31.1276
5	14	12.1751
6	3	3.9684
7	1	1.1087

The mean number of goals per game is given by:

$$\bar{y} = \frac{0\times56+1\times97+2\times90+3\times66+4\times34+5\times14+6\times3+7\times1}{56+97+90+66+34+14+3+1}$$

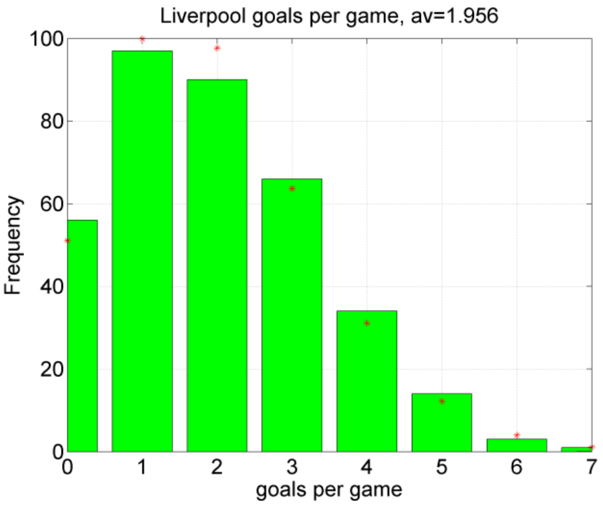
$$\bar{y} = 1.956$$

If the numbers of goals per game is Poisson distributed then the predicted frequencies are given by:

$$\boxed{f_E(y) = 361\times\frac{1.956^y}{y!}e^{-1.956}}$$

Define: $$\boxed{\chi^2 = \sum_{x=0}^{7}\frac{\left(f(y)-f_E(y)\right)^2}{f_E(y)} = 2.3727}$$

$$\boxed{P\left(\chi^2 \le 2.3727\right) = 0.0419}$$



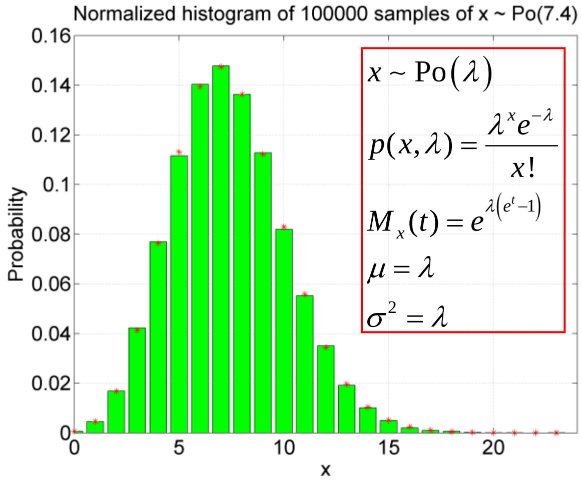
Since the data has been used to compute the mean value, the number degrees of freedom is seven, rather than eight, which is the number of possible values of goals per game.

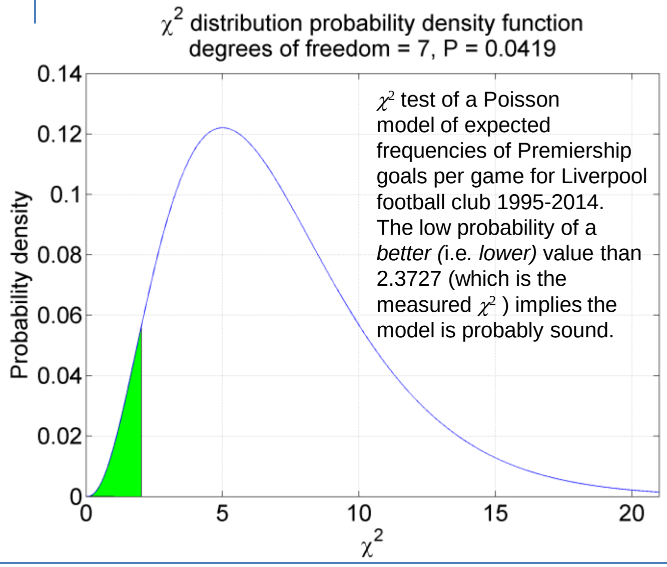
The χ^2 distribution can tell us the probability of the measured value being less than or equal to it, 'by chance alone.' In this example, the probability is only 0.0419. From this we can infer the Poisson distribution is probably a 'good' model.

'Good' theoretical models imply low P values*, although a P value 'too low' might imply spurious i.e. potentially fabricated data!

Poisson distribution

The random variable x is the number occurrences (e.g. goals, telephone calls ....) in a set interval of time, given a mean rate of occurrence λ.





χ^2 test of a Poisson model of expected frequencies of Premiership goals per game for Liverpool football club 1995-2014. The low probability of a *better* (i.e. *lower)* value than 2.3727 (which is the measured χ^2) implies the model is probably sound.

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \qquad \gamma(x,a) = \frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt$$

*Or high if we use 1-*P*