Opinionated Lessons in Statistics

by Bill Press

#19 The Chi-Square Statistic

Let's talk about **chi-square**. It means two different (but related) things. Recall that a t-value is (by definition) a deviate from $\mathrm{N}(0,1)$

χ^2 is a "statistic" defined as the sum of the squares of n independent t-values.

$$\chi^2 = \sum_i \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2, \qquad x_i \sim \mathrm{N}(\mu_i, \sigma_i)$$

$\mathrm{Chisquare}(\nu)$ is a distribution (special case of Gamma), defined as

$$\chi^2 \sim \mathrm{Chisquare}(\nu), \qquad \nu > 0$$

$$p(\chi^2)d\chi^2 = \frac{1}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)} (\chi^2)^{\frac{1}{2}\nu - 1} \exp\left(-\tfrac{1}{2}\chi^2\right) d\chi^2, \qquad \chi^2 > 0$$

The important theorem is that $\chi 2$ is in fact distributed as Chisquare.

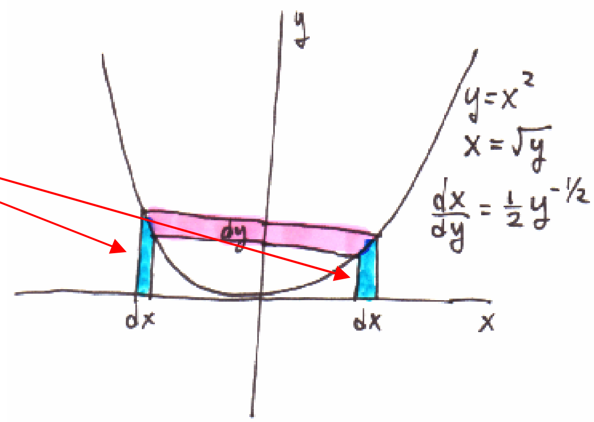
Let's prove it.

Prove first the case of ν=1:

Suppose $$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \Rightarrow \quad x \sim \mathrm{N}(0,1)$$

and $y = x^2$

$$p_Y(y)\,dy = 2p_X(x)\,dx$$



$$\text{So, } p_Y(y) = y^{-1/2} p_X(y^{1/2}) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}$$
$$\sim \mathrm{Chisquare}(1)$$

To prove the general case for integer ν, compute the characteristic function

$$\chi^2 \sim \text{Chisquare}(\nu), \qquad \nu > 0$$

$$p(\chi^2)d\chi^2 = \frac{1}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)}(\chi^2)^{\frac{1}{2}\nu-1}\exp\left(-\tfrac{1}{2}\chi^2\right)d\chi^2, \qquad \chi^2 > 0$$

```
In[9]:= pchi2 = (1 / (2^(nu / 2) Gamma[nu / 2])) y^(nu / 2 - 1) Exp[-y / 2]
```

Out[9]= $\frac{2^{-nu/2}\, e^{-y/2}\, y^{-1+\frac{nu}{2}}}{\text{Gamma}\left[\frac{nu}{2}\right]}$

```
In[10]:= Integrate[pchi2, {y, 0, Infinity}, GenerateConditions → False]
```

Out[10]= 1

```
In[11]:= Integrate[pchi2 Exp[I t y], {y, 0, Infinity},
          GenerateConditions → False]
```

Out[11]= $(1 - 2\, i\, t)^{-nu/2}$

Since we already proved that ν=1 is the distribution of a single t^2-value, this proves that the general ν case is the sum of ν t^2-values.

Question: What is the generalization of

$$\chi^2 = \sum_i \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2, \qquad x_i \sim \mathrm{N}(\mu_i, \sigma_i)$$

to the case where the x_i's are normal, **but not independent**?
I.e., **x** comes from a multivariate Normal distribution?

Answer:

$$\chi^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}), \qquad \mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Proof is one of those Cholesky things,

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T, \quad \mathbf{L}\mathbf{y} = \mathbf{x} - \boldsymbol{\mu},$$

show that $\mathbf{y}$ is product of independent N(0,1)'s, as we did before, and that

$$\chi^2 = \mathbf{y}^T \mathbf{y} = \sum y_i^2$$