Chi-Square Distribution

Egor Howell

Origins

The Chi-Square Distribution, ***χ2***, is the result of summing up ***v*** random independent variables from the ***Standard Normal Distribution***:

$$Y = \sum_{i=1}^{v} X_i^2 \sim \chi_v^2$$

Equation generated by author in LaTeX.

Origins

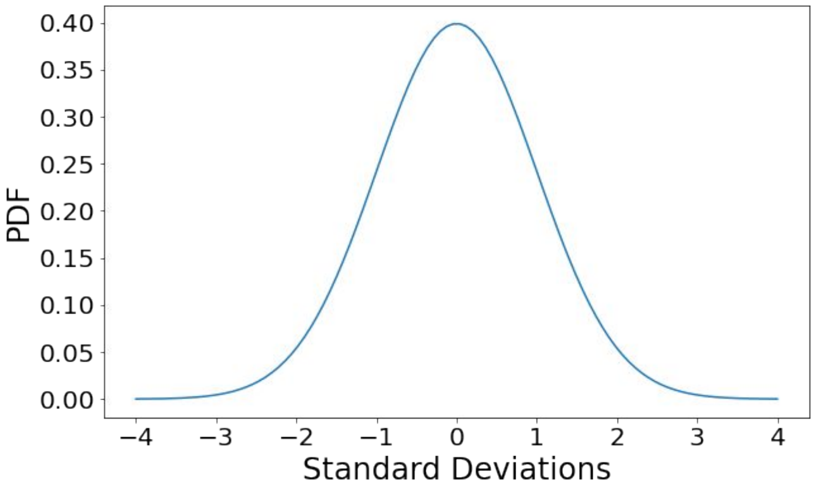
Degrees of freedom

Random variable from standard normal distribution

$$Y = \sum_{i=1}^{v} X_i^2 \sim \chi_v^2$$

Equation generated by author in LaTeX.

Standard Normal Distribution

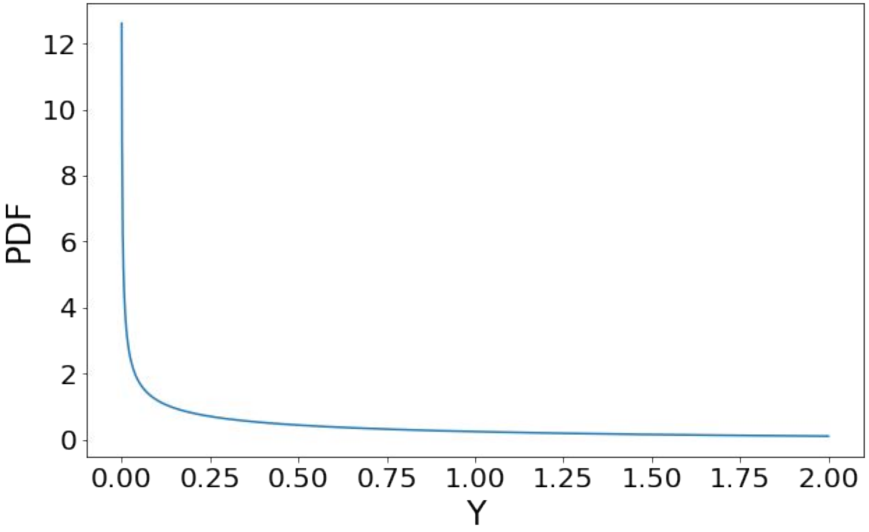


Example

For ***v* = *1***, the corresponding Ch-Square looks like:

$$Y = X^2 \sim \chi_1^2$$

Equation generated by author in LaTeX.



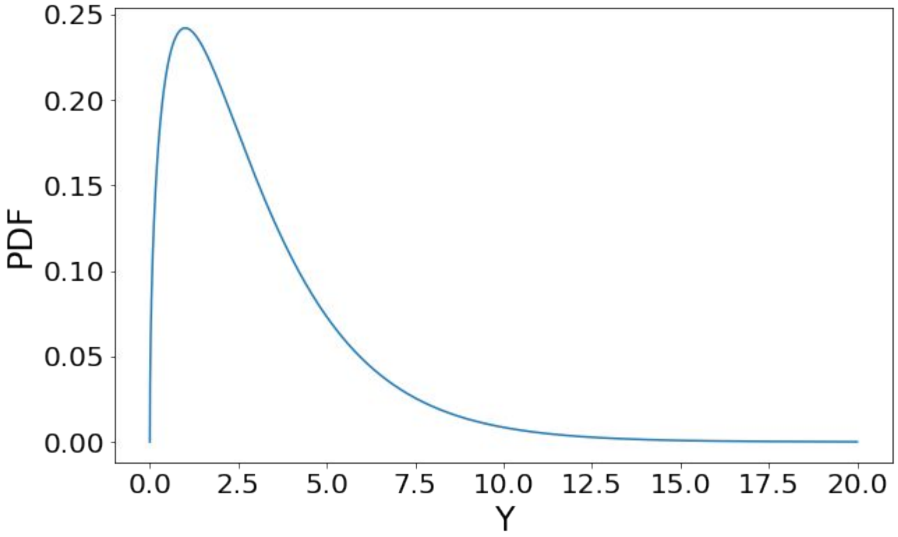
As we are squaring the values they are **going to significantly decrease or increase and become positive.**

Example

For ***v* = *3***, the corresponding Ch-Square looks like:

$$Y = X_1^2 + X_2^2 + X_3^2 \sim \chi_3^2$$

Equation generated by author in LaTeX.



The values are on average a lot greater and more left skewed. This makes sense as we are more likely to sample a high value number when we have more degrees of freedom. The summation also naturally leads to higher values.

Derivation

Now we are going to derive the Probability Density Function (PDF) for a Chi-Square Distribution with one degree of freedom, ***v = 1***.

PDF for standard normal distribution:

$$X \sim \mathcal{N}(0,1)$$

Equation generated by author in LaTeX.

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Equation generated by author in LaTeX.

Derivation

To calculate the PDF we need to find the Cumulative Density Function (CDF) and then differentiate it. The definition of the CDF is:

$$F(x) = P(X \leq x)$$

Equation generated by author in LaTeX.

Applying it to the Chi-Square formula we want to find the PDF of X^2:

$$f(x) = \frac{d}{dx} P(X^2 \leq x) = \frac{d}{dx} P(X \leq |x|) = \frac{d}{dx} P(-\sqrt{x} \leq X \leq \sqrt{x})$$

Equation generated by author in LaTeX.

Derivation

Now this derivation requires many integrals and numerous substitutions, therefore I have omitted the full proof as I don't want this to be a boring maths lecture! However, there is a great post here and webpage here that goes through the whole proof thoroughly.

Nonetheless, by carrying out the differentiation the final PDF equals:

$$f(x) = \frac{e^{-\frac{x}{2}}}{\sqrt{2\pi x}} = \frac{x^{\frac{1}{2}-1}\, e^{-\frac{x}{2}}}{\sqrt{2\pi}}$$

Equation generated by author in LaTeX.

Derivation

The PDF for ***v*** degrees of freedom is:

$$f(x) = \frac{x^{\frac{v}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})}$$

Equation generated by author in LaTeX.

Where **Γ** is the gamma function. For half positive integer values, the function is of the form:

$$\Gamma\left(v+\frac{1}{2}\right) = \binom{v-\frac{1}{2}}{v} v! \sqrt{\pi}$$

Equation generated by author in LaTeX.

Link To Gamma Distribution

We know that the Gamma Distribution is parametrised by two values:

- λ, The rate of events
- n, The number of the events you are waiting for

$$f(x) = \frac{\lambda^n \, x^{n-1} \, e^{-\lambda x}}{\Gamma(n)}$$

Image generated by author in LaTeX.

Link To Gamma Distribution

By setting ***n = v/2*** and ***λ = 1/2***:

$$f(x) = \frac{\frac{1}{2}^{\frac{v}{2}} \, x^{\frac{v}{2}-1} \, e^{-\frac{1}{2}x}}{\Gamma(\frac{v}{2})} = \frac{x^{\frac{v}{2}-1} \, e^{-\frac{1}{2}x}}{2^{\frac{v}{2}} \, \Gamma(\frac{v}{2})}$$

Image generated by author in LaTeX.

We show that the Chi-Square Distribution is just a special case of the Gamma Distribution!

Thanks

Chi-Square Distribution Simply Explained

A simple explanation of the Chi-Square Distribution and its origins

Egor Howell
Published in Towards Data Science · 5 min read · May 9, 2022

Photo by Michał Parzuchowski on Unsplash

@egorhowell

Newsletter

Dishing The Data