F-Distribution

Egor Howell

Intuition

In this video we will cover the **F Distribution** (Fisher–Snedecor distribution) which is often used in the analysis of variance (ANOVA) tests and so is definitely worth knowing if you are a Data Scientist!

The F-Distribution is closely related to the Chi-Square Distribution. If you are unfamiliar with the Chi-Square Distribution, I would highly recommend checking out my previous video:

Origins

A random variable is part of the F-Distribution if it satisfies the following:

$$F \sim \frac{\chi_1^2/v_1}{\chi_2^2/v_2}$$

Equation produced by author in LaTeX.

$$F \sim F(v_1, v_2)$$

Equation produced by author in LaTeX.

Where ***χ2*** are two different Chi-Squared distributed random variables from two independent samples, which have degrees of freedom ***v_1*** and ***v_2*** respectively. In other words, it is the ratio of two Chi-Squared distributed random variables divided by their corresponding degrees of freedom.

Origins

A random variable is part of the F-Distribution if it satisfies the following:

$$F \sim \frac{\chi_1^2/v_1}{\chi_2^2/v_2}$$

Equation produced by author in LaTeX.

$$F \sim F(v_1, v_2)$$

Equation produced by author in LaTeX.

Remember the degrees of freedom for the Chi-Square distribution are the number of normally distributed random variables that we square and sum up.

The distribution over all possible F values gives rise to the F-Distribution!

Probability Density Function

The PDF for the F-distribution is this scary looking thing:

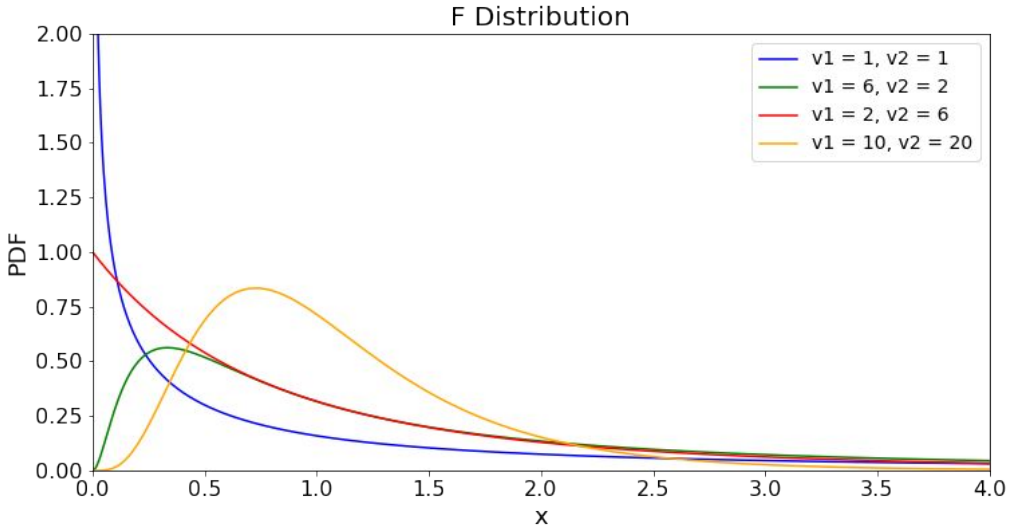
$$f(x) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right) v_1^{\frac{v_1}{2}} v_2^{\frac{v_2}{2}} x^{\frac{v_1}{2}-1}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)(v_2+v_1x)^{\frac{v_1+v_2}{2}}}$$

Equation generated in LaTeX by author.

Gamma Function

The interested reader can find a full derivation here, which includes a lot of fun things such as Jacobian matrices!

Plots



- As the degrees of freedom get larger, the distribution starts to converge towards a normal distribution.
- The distribution is often right skewed with heavy tails.
- It's always positive and is continuous.

```python
# import packages
import numpy as np
from scipy.stats import f
import matplotlib.pyplot as plt

# get x values
x = np.linspace(0, 4.5, 1000)

# get F-Distributions
f1 = f(1, 1, 0)
f2 = f(6, 2, 0)
f3 = f(2, 6, 0)
f4 = f(10, 20, 0)

# plot the distributions
plt.figure(figsize=(12, 6))
plt.plot(x, f1.pdf(x), label = 'v1 = 1, v2 = 1', color = 'blue')
plt.plot(x, f2.pdf(x), label = 'v1 = 6, v2 = 2', color = 'green')
plt.plot(x, f3.pdf(x), label = 'v1 = 2, v2 = 6', color = 'red')
plt.plot(x, f4.pdf(x), label = 'v1 = 10, v2 = 20', color = 'orange')
plt.xlim(0, 4)
plt.ylim(0.0, 2)
plt.xticks(fontsize=16)
plt.yticks(fontsize=16)
plt.xlabel('x', fontsize=18)
plt.ylabel('PDF', fontsize=18)
plt.title("F Distribution",fontsize=20)
plt.legend(fontsize=14)
plt.savefig('f_dist.png')
plt.show()
```

Thanks

Member-only story

F Distribution Simply Explained

A simple and concise description of the F-Distribution

Egor Howell
Published in Towards Data Science · 4 min read · May 20, 2022

230

Photo by Marin Tulard on Unsplash

@egorhowell

Newsletter

Dishing The Data