T - Distribution

Egor Howell

What is the t-distribution?

The t-distribution, is a continuous probability distribution that is very similar to the normal distribution, however has the following key differences:

- **Heavier tails**: *More of its probability mass is located at the extremes (higher kurtosis). This means that it is more likely to produce values far from its mean.*
- **One parameter**: *The t-distribution has only one parameter, the degrees of freedom, as it's used when we are unaware of the population's variance.*

Origin

The origin behind the t-distribution comes from the idea of modelling normally distributed data without knowing the population's variance of that data.

For example, say we sample ***n*** data points from a normal distribution, the following will be the mean and variance of this sample respectively:

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

Equation by author in LaTeX.

Sample Mean

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Equation by author in LaTeX.

Sample Standard Deviation

Origin

Combining the two equations, we can construct the following random variable:

Sample Mean

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

Equation by author in LaTeX.

Sample Standard Deviation

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Equation by author in LaTeX.

t-statistic

Population Mean

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Equation by author in LaTeX.

Probability Density Function

The t-distribution is parameterised by only one value, the degrees of freedom, *v*, and its probability density function looks like this:

$$f(t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\cdot\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

Equation by author in LaTeX.

Where:

- ***t*** *is the random variable (the t-statistic).*
- ***v*** *is the degrees of freedom, which is equal to* ***n−1****, where* ***n*** *is the sample size.*
- ***Γ(z)*** *is the gamma function, which is:*

Characteristics

The mean is defined as follows for ***v > 1***:

$$E(T) = 0$$

Equation by author in LaTeX.

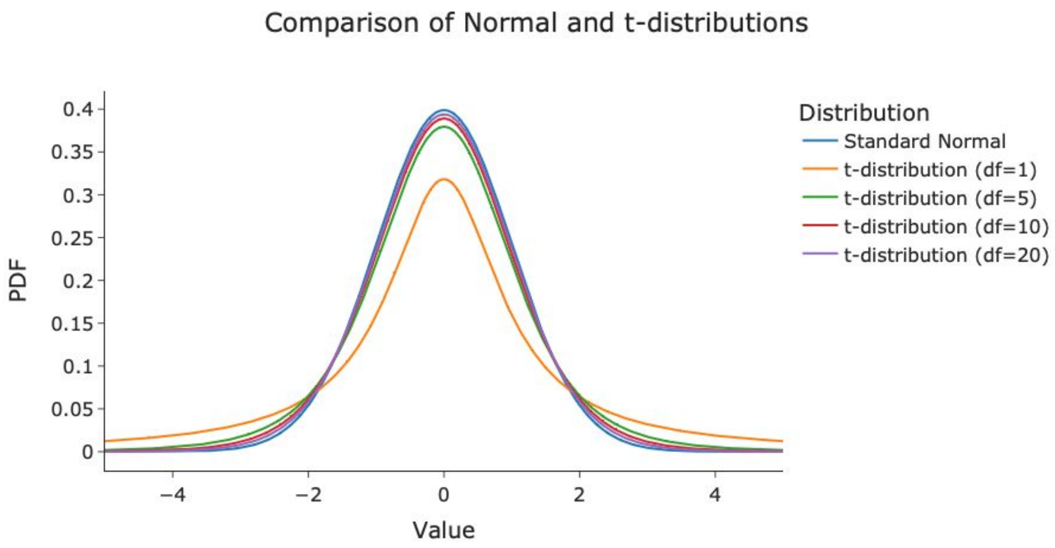
And the variance is defined as follows for ***v > 2***:

$$Var(T) = \frac{\nu}{\nu - 2}$$

Equation by author in LaTeX.

Plots

Below is an example plot of the t-distribution as a function of various degrees of freedom and also compared to the standard normal distribution:




```python
# Import packages
import numpy as np
from scipy.stats import t, norm
import plotly.graph_objects as go

# Generate data
x = np.linspace(-5, 5, 1000)
normal_pdf = norm.pdf(x, 0, 1)

# Create plot
fig = go.Figure()

# Add standard normal distribution to plot
fig.add_trace(go.Scatter(x=x, y=normal_pdf, mode='lines', name='Standard Normal

# Add t-distributions to plot for various degrees of freedom
for df in [1, 5, 10, 20]:
    t_pdf = t.pdf(x, df)
    fig.add_trace(go.Scatter(x=x, y=t_pdf, mode='lines', name=f't-distribution

fig.update_layout(title='Comparison of Normal and t-distributions',
                  xaxis_title='Value',
                  yaxis_title='PDF',
                  legend_title='Distribution',
                  font=dict(size=16),
                  title_x=0.5,
                  width=900,
                  height=500,
                  template="simple_white")
fig.show()
```

Applications

- **T-test**: *The most famous application of the t-distribution is [hypothesis testing](#) through use of the t-test, which measures the statistical difference between two sample means.*
- **Confidence intervals**: *For small sample sizes (typically less than 30), it is used to compute the [confidence interval](#) for that certain statistic with increased uncertainty.*
- **Regression**: *The t-distribution is used to determine if we should add certain covariates to our regression model and calculate hypothesis tests around the significance of their coefficients.*
- **Bayesian Statistics**: *The t-distribution is sometimes used as a prior distribution in [bayesian inference](#), which can be applied in all areas of data science, particularly reinforcement learning.*

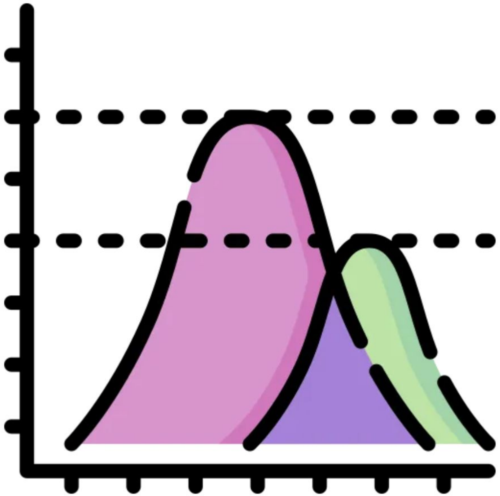
Thanks

What is the t-distribution

Discover the origins, theory and uses behind the famous t-distribution *

Egor Howell
Published in Towards Data Science · 5 min read · Sep 2

493 2



Probability icons created by Freepik — Flaticon. https://www.flaticon.com/free-icons/probability.

@egorhowell

Newsletter

Dishing The Data