The Student's t-distribution

Contents

1 Hypothesis testing

To determine the statistical significance of regression estimates, hypothesis testing is used. We would like to see whether the true β_1 is equal to a constant β_1^0. However, this is not possible as β_1 is not observed; only its estimate $\hat{\beta}_1$ is. So, the hypothesis will be that $\hat{\beta}_1$ is equal to the constant $\hat{\beta}_1^0$. Then, the null hypothesis H_0 and alternative hypothesis H_1 respectively are

$$
\begin{aligned}
H_0 &: \hat{\beta}_1 = \beta_1^0 \\
H_1 &: \hat{\beta}_1 \neq \beta_1^0
\end{aligned}
$$

$\hat{\beta}_1$ is a random variable and has a sampling distribution. This means that there is a probability associated to the condition of the null hypothesis $H_0 : \hat{\beta}_1 = \beta_1^0$. To check if this equality statistically holds, a test statistic based on this random variable $\hat{\beta}_1$ needs to be constructed. This test statistic also is a random variable, and thus has a probability distribution.

The t test statistic is a function of $\boldsymbol{\varepsilon}$. So, the distribution of $\boldsymbol{\varepsilon}$ determines the distribution of the t statistic. Since $\boldsymbol{\varepsilon}$ is also not observed, assumptions about its distribution need to be made.

The simple linear model can be expressed in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

Here, $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are vectors consisting of n rows and one column. $\boldsymbol{\beta}$ is a 2×1 column vector containing β_1 and β_2 and $\boldsymbol{X}$ is a $n \times 2$ matrix: all x_{i1}'s in the first column and all x_{i2}'s in the second column.

When we assume that $\boldsymbol{\varepsilon}$ has a normal distribution, the t statistic has an exact distribution. This means that the probability distribution of t is the same for any finite n. The t test statistic has a t distribution if $\boldsymbol{\varepsilon}$ is normal. If $\boldsymbol{\varepsilon}$ is not normal distributed, the test statistic does not have an exact distribution. But if we require that n is large, then the test statistic has an asymptotic distribution that approximates an exact distribution.

2 Deriving the distribution of the t-statistic

To generalize the derivation of the t-statistic , we will use β_k, $\hat{\beta}_k$ and $\hat{\beta}_k^0$ from now on, where $k = 1, ..., K$ denotes the k-th coefficient. By the structure from OLS,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}.$$

Under the assumption that,

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left(0, \sigma^2 \boldsymbol{I}\right),$$

we get the following for $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

The above equation tells us that conditionally on $\boldsymbol{X}$, $\hat{\boldsymbol{\beta}}$ has a Normal distribution with mean $\boldsymbol{\beta}$ and variance $\sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. Subtracting $\boldsymbol{\beta}$ from both sides results in

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X} \sim N\left[\boldsymbol{0}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

The t-test is a hypothesis test which is performed on one specific estimator contained in the vector $\hat{\boldsymbol{\beta}}$. Letting $\hat{\beta}_k$ denote the k-*th* entry in $\hat{\boldsymbol{\beta}}$, we have that

$$\hat{\beta}_k - \beta_k \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right],$$

where

$$S^{kk} \equiv \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{k,k}$$

for ease of notation. Then, under the null hypothesis that $\beta_k = \beta_k^0$, we get:

$$\hat{\beta}_k - \beta_k^0 \mid \boldsymbol{X} \sim N\left[0, \sigma^2 S^{kk}\right].$$

The reason behind this rewriting is that under the null hypothesis we assume that the expected value of $\hat{\beta}_k$ conditional on $\boldsymbol{X}$ is β_k^0. So, under the null and conditional on $\boldsymbol{X}$, the expected value of the difference between $\hat{\beta}_k$ and β_k^0 is equal to zero. This allows us to test whether, with the given data, $\hat{\beta}_k - \beta_k^0$ is equal to zero. The next operation is dividing $\hat{\beta}_k - \beta_k^0$ by $\sqrt{\sigma^2 S^{kk}}$, which is done to normalize the distribution. As a result we get that:

$$z_k \mid \boldsymbol{X} \equiv \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}} \,\Bigg|\, \boldsymbol{X} \sim N[0, 1]. \tag{2}$$

Now the mean is zero and the variance is equal to 1. This means that the distribution is independent from $\hat{\beta}_k$, β_k^0, σ, and $\boldsymbol{X}$. Hence as a final form:

$$z_k \sim N[0, 1]$$

However, the true variance σ^2 is not observed. Therefore we have to substitute σ^2 with its unbiased estimator $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K}$.

We obtain

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}}. \tag{3}$$

Due to this substitution, we need to figure out what the distribution of t_k is. This is done by rewriting the equation. We will multiply the equation by $\frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}}$ and afterwards multiply by $\frac{\sqrt{n-K}}{\sqrt{n-K}} = \frac{1}{\sqrt{n-K}/\sqrt{n-K}}$. The multiplications are valid since $\frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}} = \frac{\sqrt{n-K}}{\sqrt{n-K}} = 1$. The multiplications are colored to make the derivations easier to observe:

$$\begin{aligned}
t_k &= \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2 S^{kk}}} \\
&= \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{\sigma}^2}\sqrt{S^{kk}}} \frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}} \\
&= \frac{\hat{\beta}_k - \beta_k^0/\sqrt{\sigma^2 S^{kk}}}{\sqrt{\hat{\sigma}^2/\sigma^2}} \\
&= \frac{\left(\hat{\beta}_k - \beta_k^0\right)/\sqrt{\sigma^2 S^{kk}}}{\sqrt{\hat{\sigma}^2/\sigma^2}} \frac{1}{\sqrt{(n-K)}/\sqrt{(n-K)}} \\
&= \frac{\left(\hat{\beta}_k - \beta_k^0\right)/\sqrt{\sigma^2 S^{kk}}}{\sqrt{\left(\hat{\sigma}^2/\sigma^2\right)(n-K)/(n-K)}}.
\end{aligned}$$

From equation (2) we already know that the numerator of this fraction is distributed as

$$\left(\hat{\beta}_k - \beta_k^0\right) / \sqrt{\sigma^2 S^{11}} = z_k \sim N\left[0,1\right]$$

We also have that

$$\left(\hat{\sigma}^2/\sigma^2\right)(n-K) \sim \chi^2\left[n-K\right],$$

which is a chi-squared distribution with $n-K$ degrees of freedom, and is contained in the denominator of the fraction. Filling in the distribution gives:

$$t_k \sim \frac{N\left[0,1\right]}{\left\{\chi^2\left[n-K\right]/(n-K)\right\}^{1/2}} = t\left[n-K\right].$$

Hence, t_k follows a Student's t-distribution with $n-K$ degrees of freedom.

3 Generating values from the t-distribution in MATLAB

In this document we will make use of the Student's t-distribution. We first set the parameters that we will use later on. In this case the degrees of freedom is equal to five and `x` consists of five values: -5, -2.5, 0, 2.5 and 5. The related script file can be found here on GitHub.

```
% Define degrees of freedom
df = 5;
% Define x values range
x = linspace(-5, 5, 5);
```

To generate values from the Student's t-distribution, we need the corresponding probability density function. This function is defined as

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

$\Gamma(\cdot)$ denotes the gamma function and ν is the degrees of freedom. In our case, the values at which the gamma function will be evaluated are always positive valued. We can use the following Euler integral as an expression for the gamma function:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$$

In the code we will first calculate the two gamma functions and then determine the probability densities according to the theoretical method.

```
val1 = (df+1)/2;
fun1 = @(t) t.^(val1 - 1).*exp(-t);
gam1 = integral(fun1, 0, Inf);

val2 = df/2;
fun2 = @(t) t.^(val2 - 1).*exp(-t);
gam2 = integral(fun2, 0, Inf);

ymanual = gam1 / (gam2*sqrt(pi * df)) * (1 + x.^2 / df).^-((df + 1) ./ 2);
```

Instead of manually calculating the probability densities, another option is to make use of the built-in `pdf` function in MATLAB. The function requires three input arguments. The first one is the type of probability distribution. Since we want to use the Student's t-distribution, the input argument is `'T'`. The second input is the set of values at which the probability density needs to be evaluated, which is `x`. Thirdly, input for the parameter ν (the degrees of freedom) is required, which is denoted by `nu`.

```
ybuiltin = pdf('T', x, nu);
```

When comparing the calculated probability densities in the code below, MATLAB returns that the values are not the same.

```
isequal(ybuiltin,ymanual);
```

This difference is caused by rounding. The next code shows that the marginal error is less than 0.000001.

```
ismembertol(ybuiltin, ymanual, 1e-6);
abs(ybuiltin - ymanual);
```

Now we will take random draws from the Student's t-distribution. We will first only take ten samples and plot the corresponding frequency distribution. To be able to reproduce the output, we set a seed for the random number generator.

```
% Setting a seed to reproduce results
rng(14);
% Number of random samples
N = 10;
% Generate random samples
samples = random('T', nu, N, 1);
% Create histogram
figure;
histogram(samples,'Normalization','pdf','BinWidth', 0.5);
title("Empirical PDF of the t distribution (\nu = "+ nu +")");
xlabel('t');
ylabel('Probability density');
```

This gives us the following graph:

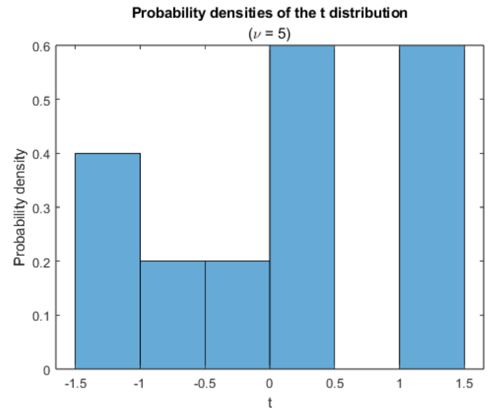

Figure 1: Frequency distribution of ten samples

We now increase the number of samples by setting N in the previous code to 100, resulting in Figure 2.

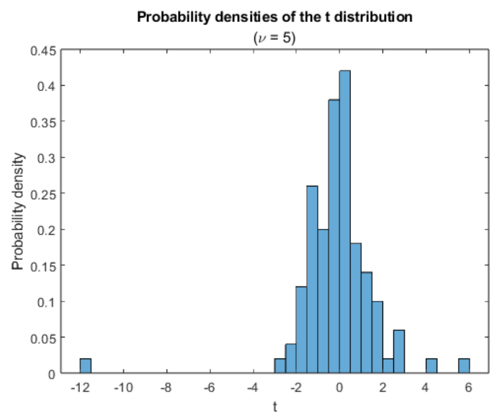

Figure 2: Frequency distribution of 100 samples

As a result of increasing the sample size, the frequency distribution looks smoother and has a distinguishable shape. However, it has a very noticeable outlier. When we keep on increasing the number of samples, we will start to approach the limiting distribution: the theoretical Student's t-distribution.

```
x = -5:0.1:5;
y = pdf('T', x, nu);

figure;
plot(x, y, 'LineWidth', 2);
title("Empirical PDF of the t distribution (\nu = "+ nu +")");
xlabel('t');
ylabel('Probability density');
```

The probability distribution looks like this:

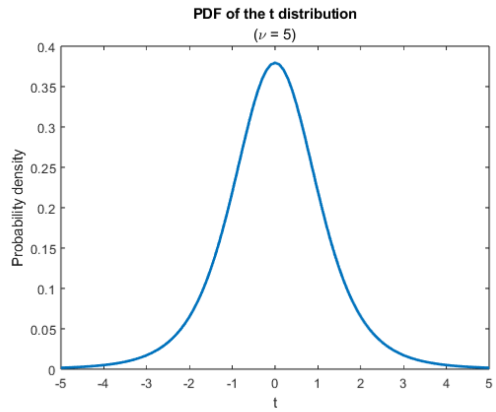

Figure 3: The limiting distribution

In this plot, the properties of the t-distribution can be observed: the distribution is centered around zero and is symmetric. It is also possible to increase the degrees of freedom, resulting in the following.

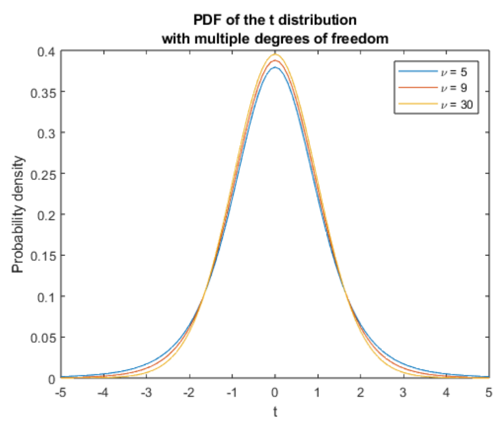

Figure 4: Limiting distributions with different degrees of freedom

As the degrees of freedom increase, the outer sides of the density functions, also known as the tails, contain less density: the tails are 'lighter' now and the density becomes more contained around zero.