Chi-squared distribution. Student's t-distribution.

Gleb Karpov

October 3, 2022

Chi-squared distribution.

Let us have several random variables $Z_i \sim \mathcal{N}(0,1)$. Then, define the following function of these random variables:

$$\chi^2(k) = \sum_{i=1}^{k} Z_i^2. \tag{1}$$

Of course it is a random variable itself. Distribution of such random variable is called "Chi squared distribution", and allows us to perform sophisticated statistical procedures. Probability density function of chi-squared variable *heavily* depends on a number of summands in Eq. (1). We call their number **degrees of freedom**. So, the correct way to name the latter variable is "chi-squared random variable with k degrees of freedom", and we mention number of degrees of freedom in parentheses $\chi^2(k)$.

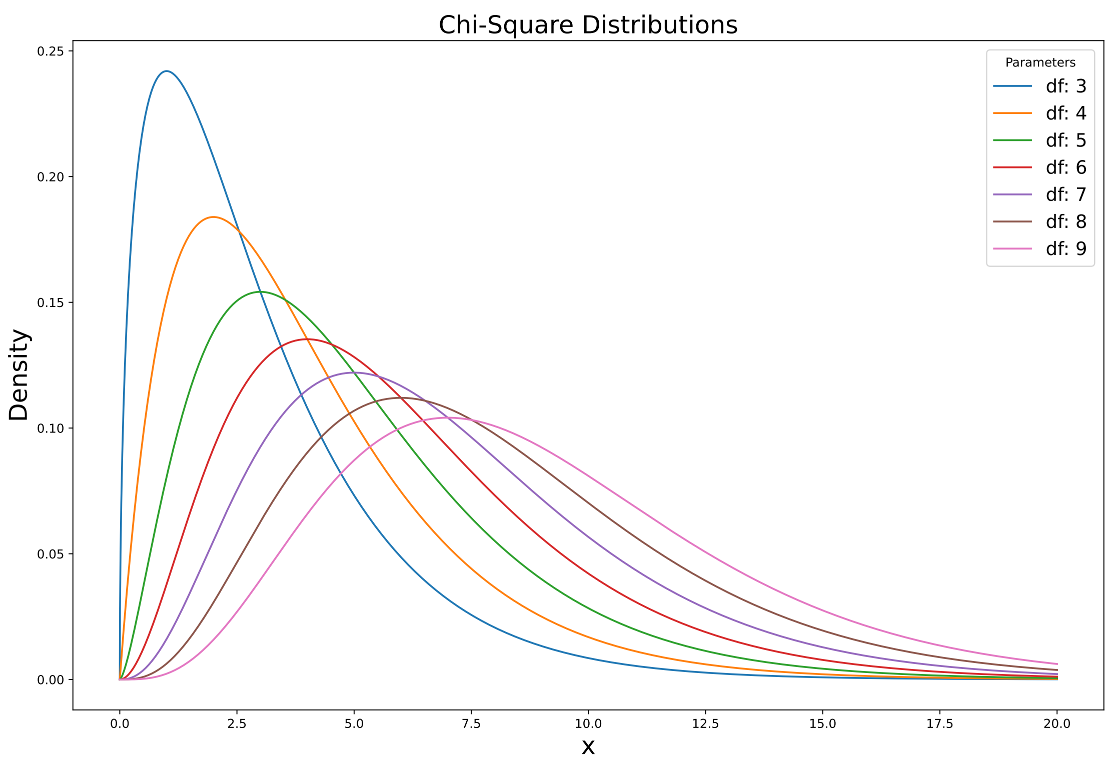

Figure 1: Change in the density function's shape with changing number of degrees of freedom.

Statement 1. *If two independent chi-squared variables are summed together, we obtain new random variable with number of degrees of freedom equal to the sum of these numbers from two variables.*

$$\chi^2(k+l) = \chi^2(k) + \chi^2(l) = \sum_{i=1}^{k} Z_i^2 + \sum_{j=1}^{l} Z_i^2 = \sum_{i=1}^{k+l} Z_i^2.$$

To bring all ties together, we need to uncover some properties of that random variable.

Statement 2. *Consider random sample $X_1, \ldots, X_k$ from some population. Maybe it would seem suspicious, but still:*

$$\sum_{i=1}^{k}\left(X_i - \bar{X}\right) = 0. \tag{2}$$

Proof. Quite easy walk through.

$$\sum_{i=1}^{k}\left(X_i - \bar{X}\right) = \sum_{i=1}^{k} X_i - k\bar{X} = \sum_{i=1}^{k} X_i - k\frac{\sum\limits_{i=1}^{k} X_i}{k} = 0.$$

□

Statement 3. *Consider random sample $Z_1, \ldots, Z_k$ from the standard normal population, i.e. $Z_i \sim \mathcal{N}(0,1)$. Random variable $k\bar{Z}^2$ is a chi-squared variable, with one degree of freedom, i.e. $k\bar{Z}^2 \sim \chi^2(1\ df)$.*

Proof. We use the fact that $\text{Var}(aX) = a^2\text{Var}(X)$, and that the sum of normal random variables is normal random variable itself. Let us study properties of random variable $\bar{Z}$:

- $\text{Var}(\bar{Z}) = \sum\limits_{i=1}^{k} \text{Var}\left(\frac{Z_i}{k}\right) = \sum\limits_{i=1}^{k} \frac{1}{k^2}\text{Var}\left(Z_i\right) = \frac{1}{k^2}\sum\limits_{i=1}^{k} 1 = \frac{k}{k^2} = \frac{1}{k}.$
- $E(\bar{Z}) = \sum\limits_{i=1}^{k} E\left(\frac{Z_i}{k}\right) = \frac{1}{k}\sum\limits_{i=1}^{k} E\left(Z_i\right) = \frac{1}{k}\sum\limits_{i=1}^{k} 0 = 0.$
- Finally, we conclude that $\bar{Z} \sim \mathcal{N}\left(0, \frac{1}{k}\right)$.

Then consider random variable $\sqrt{k}\bar{Z}$. What are its properties?

- $\text{Var}(\sqrt{k}\bar{Z}) = k\text{Var}(\bar{Z}) = 1.$
- $E(\sqrt{k}\bar{Z}) = \sqrt{k}E(\bar{Z}) = 0.$

Which means that random variable $\sqrt{k}\bar{Z} \sim \mathcal{N}(0,1)$, and it is a standard normal variable. According to if some random variable $Z \sim \mathcal{N}(0,1)$ then $Z^2 \sim \chi^2(1\ df)$, we obtain:

$$\boxed{k\bar{Z}^2 \sim \chi^2(1\ df)}$$

□

Consider the following trick:

$$\sum_{i=1}^{k} Z_i^2 = \sum_{i=1}^{k}(Z_i - \bar{Z} + \bar{Z})^2 = \sum_{i=1}^{k}\left((Z_i - \bar{Z})^2 + 2(Z_i - \bar{Z})\bar{Z} + \bar{Z}^2\right) = \sum_{i=1}^{k}(Z_i - \bar{Z})^2 + k\bar{Z}^2 + 2\bar{Z}\underbrace{\sum_{i=1}^{k}(Z_i - \bar{Z})}_{0}$$

$$= \boxed{\sum_{i=1}^{k}(Z_i - \bar{Z})^2 + k\bar{Z}^2} \tag{3}$$

What does the latter tell us? As it was mentioned in Statement 1, when we sum up chi-squared variables, degrees of freedom also are being summed up. Then, if $\sum\limits_{i=1}^{k} Z_i^2 \sim \chi^2(k\ df)$, and $k\bar{Z}^2 \sim \chi^2(1\ df)$, it follows from Eq. (3), that:

$$\underbrace{\sum_{i=1}^{k} Z_i^2}_{k\ df} = \underbrace{\sum_{i=1}^{k}(Z_i - \bar{Z})^2}_{(k-1)\ df} + \underbrace{k\bar{Z}^2}_{1\ df}$$

$$\boxed{\sum_{i=1}^{k}(Z_i - \bar{Z})^2 \sim \chi^2(k-1\ df)} \tag{4}$$

Important! Pay attention, that in this situation, even if we sum from 1 to k, final random variable still has $(k-1)$ degrees of freedom! This is what I was saying during the seminars.

Student's t-distribution

This is artificial random variable, which also helps to handle with many sophisticated problems in Statistics. For now, we only need to know how it is constructed. Let us have standard normal variable $Z \sim \mathcal{N}(0,1)$, and chi-squared variable $\chi^2(k\ df)$ with k degrees of freedom. Consider the following variable:

$$\boxed{t(k\ df) = \frac{Z}{\sqrt{\frac{\chi^2(k\ df)}{k}}}} \tag{5}$$

The distribution of this random variable is called Student's t-distribution, and the variable itself is called a Student's t-variable. Important to note, that once it depends on chi-squared variable with arbitrary number of degrees of freedom, then t-variable also *heavily* depends on this number. The shape of density function is changing noticeably.

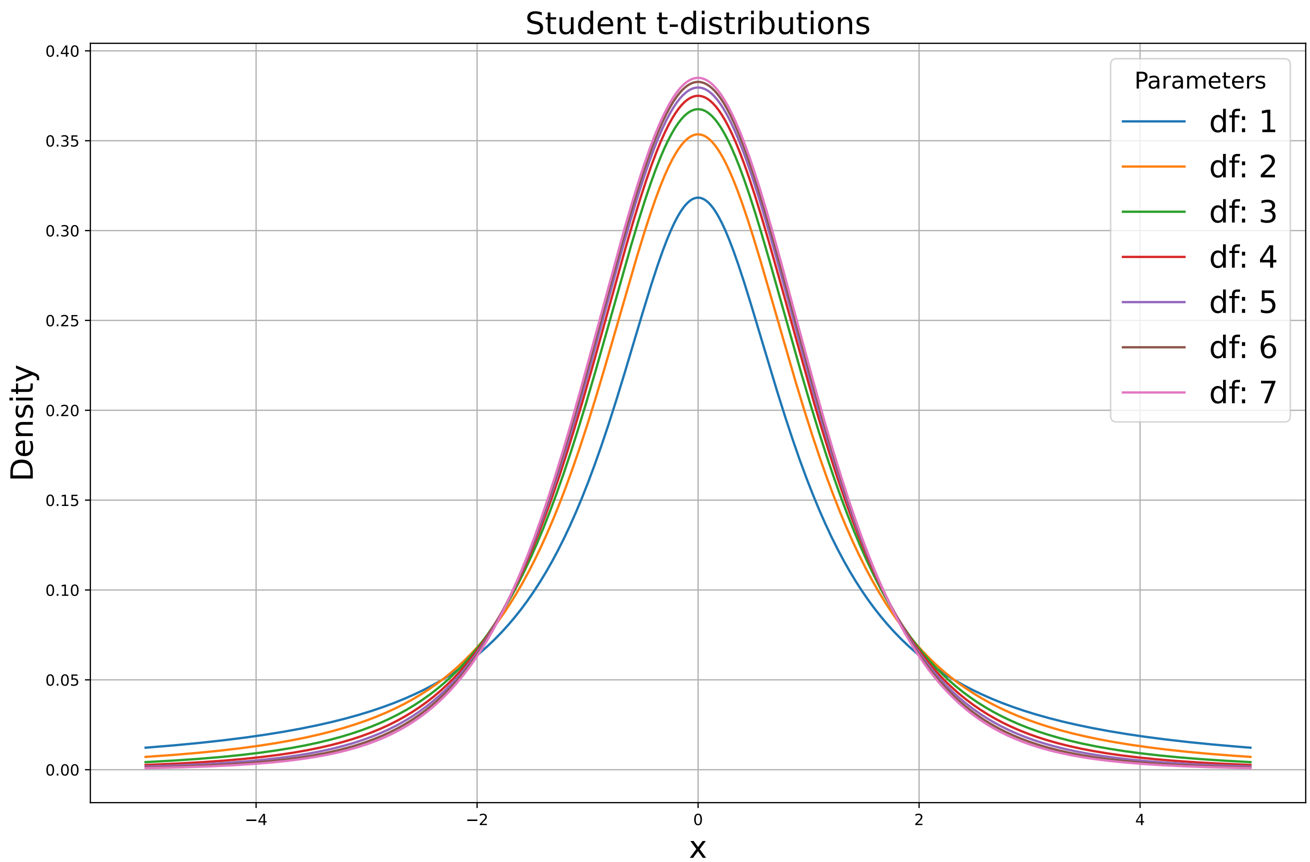

Figure 2: Change in the t-variable density function's shape with changing number of degrees of freedom.

Another property of t-distribution is that if number of degrees of freedom is largely increasing, then it behaves as standard normal distribution (you can see it on Figure 3):

$$t(n\ df) \xrightarrow{n\to\infty} Z \sim \mathcal{N}(0,1).$$

Application in Statistics

There is more theory, and properties about Chi-square and t distributions, but we are going to study, how we make a use of them in topics related to our course. Let us construct χ^2 variable from the new things we discovered in Statistics course!

Statement 4. *Consider random sample $X_1, \ldots, X_n$ of length n from the normal population with mean μ and variance σ^2. Random variable $\frac{S^2(n-1)}{\sigma^2}$ is a chi-squared variable, with $(n-1)$ degree of freedom, where S^2 is a sample variance, $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$.*

Proof. Let us make some transformations with sample variance:

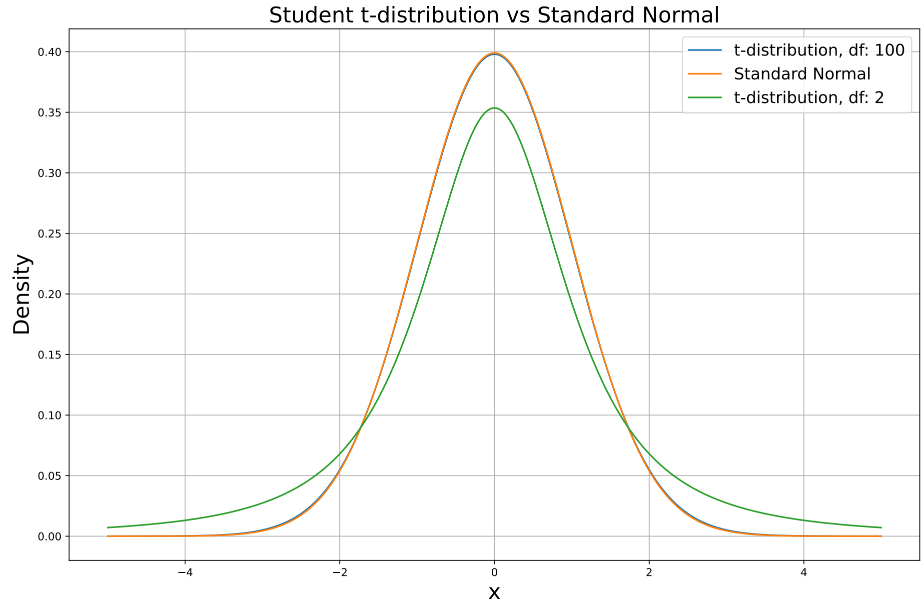

Figure 3: Converging of the t-distribution density to the standard normal with increasing of the degrees of freedom.

$$\frac{S^2(n-1)}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}-\frac{\bar{X}-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(Z_i-\frac{\bar{X}-\mu}{\sigma}\right)^2 =$$

$$= \sum_{i=1}^{n}\left(Z_i-\bar{Z}\right)^2 \rightarrow \text{ according to consequence of Eq. (3)} \rightarrow \sum_{i=1}^{n}\left(Z_i-\bar{Z}\right)^2 \sim \chi^2(n-1\ df).$$

Finally,

$$\boxed{\frac{S^2(n-1)}{\sigma^2} \sim \chi^2(n-1\ df)} \tag{6}$$

□

Important! Notice, that here as well, even if have n elements in the sample and so sum from 1 to n in the sample variance formula, the final chi-squared variable still has $(n-1)$ degrees of freedom!

Statement 5. *Consider random sample $X_1,\ldots,X_n$ of length n from the normal population with mean μ and variance σ^2. Random variable*

$$t = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$$

follows Student's t-distribution with $(n-1)$ degrees of freedom.

Proof. Let us start with definition of the t-variable:

$$t(k\ df) = \frac{Z}{\sqrt{\frac{\chi^2(k\ df)}{k}}}.$$

Now we will deal with each colored part separately.

1. Z, we create it from idea that sample mean $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$. We can perform transformation to the standard normal variable as we used to do in Probability Theory:

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

2. $\chi^2(k\ df)$, we can use result of Statement 4 and create such variable from the sample variance. Because number of degrees of freedom can be arbitrary, it also can be $(n-1)$:

$$\chi^2(n-1\ df) = \frac{S^2(n-1)}{\sigma^2}$$

3. k, basically it means we need to divide on the number of degrees of freedom that χ^2 variable has. In our case:

$$k = n - 1.$$

Let's put it all together.

$$t(n-1\ df) = \frac{Z}{\sqrt{\frac{\chi^2(n-1\ df)}{n-1}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2(n-1)}{\sigma^2(n-1)}}\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}.$$

$\square$