

5. The Student t Distribution

In this section we will study a distribution that has special importance in statistics. In particular, this distribution will arise in the study of a standardized version of the [sample mean](#) when the underlying distribution is normal.

The Probability Density Function

Suppose that Z has the [standard normal distribution](#), V has the [chi-squared distribution](#) with n degrees of freedom, and that Z and V are [independent](#). Let

$$T = \frac{Z}{\sqrt{V/n}}$$

In the following exercise, you will show that T has [probability density function](#) given by

$$f(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad t \in \mathbb{R}$$

1. Show that T has the given probability density function by using the following steps.
- Show first that the conditional distribution of T given $V = v$ is normal with mean 0 and variance $\frac{n}{v}$.
 - Use (a) to find the joint probability density function of (T, V) .
 - Integrate the joint probability density function in (b) with respect to v to find the probability density function of T .

The distribution of T is known as the **Student t distribution** with n degree of freedom. The distribution is well defined for any $n > 0$, but in practice, only positive integer values of n are of interest. This distribution was first studied by [William Gosset](#), who published under the pseudonym *Student*. In addition to supplying the proof, [Exercise 1](#) provides a good way of thinking of the t distribution: the t distribution arises when the variance of a mean 0 normal distribution is randomized in a certain way.

2. In the [random variable experiment](#), select the student t distribution. Vary n and note the shape of the density function. For selected values of n , run the simulation 1000 times with an update frequency of 10. Note the apparent convergence of the empirical density function to the true density function.

3. Sketch the graph of the Student [probability density function](#). In particular, let $a_n = \sqrt{\frac{n}{n+2}}$ and show that
- f is symmetric about $t = 0$.
 - f is increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$.
 - The maximum occurs at $t = 0$.
 - f is concave upward on $(-\infty, -a_n)$ and on (a_n, ∞) ; f is concave downward on $(-a_n, a_n)$.
 - The inflection points occur at $t = \pm a_n$.

- f. $f(t) \rightarrow 0$ as $t \rightarrow \infty$ and as $t \rightarrow -\infty$
- g. Note that $a_n \rightarrow 1$ as $n \rightarrow \infty$.

In particular, it follows that the distribution is unimodal with mode and median at $t = 0$

4. The t distribution with 1 degree of freedom is known as the **Cauchy distribution**, named after **Augustin Cauchy**.

Show that the probability density function is

$$f(t) = \frac{1}{\pi(1+t^2)}, \quad t \in \mathbb{R}$$

The probability density function of the Cauchy distribution can be obtained by normalizing the function

$$g(t) = \frac{1}{1+t^2}, \quad t \in \mathbb{R}$$

We recognize g , of course, as the derivative of the arctangent function. However, the graph of g is also known as the **witch of Agnesi**, named for the Italian mathematician **Maria Agnesi**.

The **distribution function** and the **quantile function** of the general t distribution do not have simple, closed-form representations. Approximate values of these functions can be obtained from the **table of the t distribution**, from the **quantile applet**, and from most mathematical and statistical software packages. However, we can find simple formulas in the special case of the Cauchy distribution.

5. Let F denote the distribution function of the Cauchy distribution. Show that

- a. $F(t) = \frac{1}{\pi} \arctan(t) + \frac{1}{2}$, $t \in \mathbb{R}$
- b. $F^{-1}(p) = \tan\left(\pi\left(p - \frac{1}{2}\right)\right)$, $p \in (0, 1)$
- c. The first quartile is $t = -1$ and the third quartile is $t = 1$

6. In the **quantile applet**, select the student distribution. Vary the parameter and note the shape of the density function and the distribution function. In each of the following cases, find the median, the first and third quartiles, and the interquartile range.

- a. $n = 2$
- b. $n = 5$
- c. $n = 10$
- d. $n = 20$



Moments

Suppose that T has the t -distribution with n degrees of freedom. The basic **random variable representation** can be used to find the **mean** and **variance** and other moments of T .

7. Show that

- $\mathbb{E}(T)$ is undefined if $n \in (0, 1]$
- $\mathbb{E}(T) = 0$ if $n \in (1, \infty)$

In particular, the Cauchy distribution does not have a mean.

8. Show that

- $\text{var}(T)$ is undefined if $n \in (0, 1]$
- $\text{var}(T) = \infty$ if $n \in (1, 2]$
- $\text{var}(T) = \frac{n}{n-2}$ if $n \in (2, \infty)$

Note that $\text{var}(T) \rightarrow 1$ as $n \rightarrow \infty$.

9. In the simulation of the **random variable experiment**, select the student t distribution. Vary n and note the location and shape of the mean-standard deviation bar. For the following values of n , run the simulation 1000 times with an update frequency of 10. Compare the behavior of the empirical moments with the theoretical results in [Exercise 7](#) and [Exercise 8](#).

- $n = 3$
- $n = 2$
- $n = 1$

10. Show that

- $\mathbb{E}(T^k)$ is undefined if k is odd and $n \in (0, k]$
- $\mathbb{E}(T^k) = \infty$ if k is even and $n \in (0, k]$
- $\mathbb{E}(T^k) = 0$ if k is odd and $n \in (k, \infty)$
- Finally, if k is even and $n \in (k, \infty)$ then

$$\mathbb{E}(T^k) = \frac{\Gamma((k+1)/2) \Gamma((n-k)/2)^{k/2}}{\sqrt{\pi} \Gamma(n/2)}$$

Normal Approximation

You probably noticed that, qualitatively at least, the t density function is very similar to the standard normal density function. The similarity is quantitative as well:

11. Use a basic limit theorem from calculus to show that for fixed t ,

$$f(t) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \text{ as } n \rightarrow \infty$$

Note that the function on the right is the probability density function of the standard normal distribution.

12. In the basic [random variable representation](#), use the [strong law of large numbers](#) to show that, with probability 1,
- $\frac{V}{n} \rightarrow 1$ as $n \rightarrow \infty$
 - $T \rightarrow Z$ as $n \rightarrow \infty$

The t distribution has more probability in the tails, and consequently less probability near 0, compared to the standard normal distribution.

13. Suppose that T has the t distribution with $n = 10$ degrees of freedom. For each of the following, compute the true value using the [quantile applet](#) and then compute the normal approximation. Compare the results.
- $\mathbb{P}(-0.8 < T < 1.2)$
 - The 90th percentile of T .

