



# ORF 245 Fundamentals of Statistics

## $\chi^2$ , $t$ , and $F$ Distributions

Robert Vanderbei

Fall 2015

Slides last edited on October 17, 2015

# $\chi^2$ Distribution

Let  $X_1, X_2, \dots, X_n$  be a sequence of iid standard normal random variables and let

$$Z = \sum_{i=1}^n X_i^2$$

The random variable  $Z$  is said to have a  $\chi_n^2$  distribution.

For  $n = 1$ , the distribution is easy to compute explicitly. We begin with the cdf:

$$\begin{aligned} P(Z \leq z) &= P(X^2 \leq z) = P(-\sqrt{z} \leq X \leq \sqrt{z}) \\ &= \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 2 \int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \end{aligned}$$

Differentiating, we get the density function:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} z^{-1/2} e^{-z/2}$$

We omit the derivation, but for general values of  $n$  the density has the form:

$$f_Z(z) \approx z^{n/2-1} e^{-z/2}, \quad z \geq 0$$

Note: This is a Gamma distribution with parameters “ $n$ ” =  $n/2$  and  $\lambda = 1/2$ .

# Student's $t$ -Distribution

Let  $Z \sim N(0, 1)$  and let  $U \sim \chi_n^2$ .

Suppose that  $Z$  and  $U$  are independent.

The distribution of

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

is called the  $t$  *distribution* with  $n$  *degrees of freedom*.

We omit the derivation, but the density function has an “explicit” formula

$$f_T(t) = c \left( 1 + \frac{t^2}{n} \right)^{-(n+1)/2}$$

where  $c$  is a normalization constant to make the area under the density equal to one.

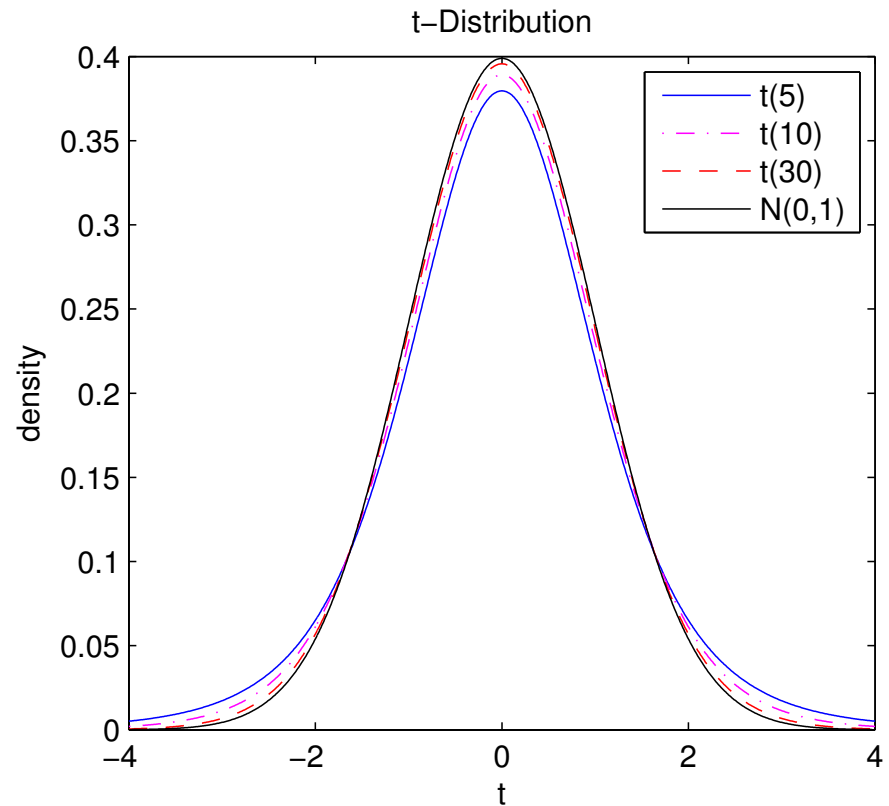
Note: When  $n = 1$ , this is the Cauchy distribution.

# The $t$ Distribution vs. $N(0,1)$

```
n = 10;  
t = -4:0.01:4 ;  
c = gamma((n+1)/2);  
c = c / (sqrt(n*pi) * gamma(n/2) );  
f = c*(1+t.^2/n).^(-(n+1)/2);  
plot(t,f,'g--');
```

Alternatively...

```
n = 10;  
t = -4:0.01:4 ;  
f = pdf('t',t,n);  
plot(t,f,'g--');
```



For  $n \geq 30$  or so, the  $t$ -distribution is very well approximated by the standard normal distribution.

# The $F$ Distribution

Let  $U \sim \chi_m^2$  and let  $V \sim \chi_n^2$ .

The distribution of

$$W = \frac{U/m}{V/n}$$

is called the *F distribution* with  $m$  and  $n$  degrees of freedom and is denoted by  $F_{m,n}$ .

We omit the derivation, but the density function has an “explicit” formula

$$f_W(w) = c w^{m/2-1} \left(1 + \frac{m}{n}w\right)^{-(m+n)/2}, \quad w \geq 0$$

where  $c$  is a normalization constant to make the area under the density equal to one.

# Sample Mean and Sample Variance

Let  $X_1, X_2, \dots, X_n$  be independent  $N(\mu, \sigma^2)$  random variables.

Such a collection is often called a *sample* from the distribution.

---

The *sample mean* is defined as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Recall that

$$\mathbb{E}(\bar{X}) = \mu \quad \text{and} \quad \text{Var}(\bar{X}) = \sigma^2/n$$

---

The *sample variance* is defined as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Why  $n - 1$  in the denominator?

# Biased vs. Unbiased Estimator

We compute:

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) &= \sum_{i=1}^n \mathbb{E}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= n\mathbb{E}(X^2) - 2\sum_{i,j} \mathbb{E}\left(\frac{X_i X_j}{n}\right) + n\sum_{j,k} \mathbb{E}\left(\frac{X_j X_k}{n^2}\right) \\ &= n\mathbb{E}(X^2) - \frac{1}{n}\sum_{i,j} \mathbb{E}(X_i X_j) \\ &= n\mathbb{E}(X^2) - \frac{1}{n}\sum_{i \neq j} \mathbb{E}(X_i X_j) - \frac{1}{n}\sum_{i=j} \mathbb{E}(X_i X_j) \\ &= n\mathbb{E}(X^2) - n(n-1)\frac{\mu^2}{n} - \mathbb{E}(X^2) \\ &= (n-1)\left(\mathbb{E}(X^2) - \mu^2\right) \\ &= (n-1)\sigma^2\end{aligned}$$

Hence,  $\mathbb{E}(S^2) = \sigma^2$ .

# Three Results for Later

## Corollary A

The random variables  $\bar{X}$  and  $S^2$  are *independent*.

## Theorem B

The random variable  $(n - 1)S^2/\sigma^2$  is *chi-square* with  $n - 1$  degrees of freedom.

## Corollary B

The random variable

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

has a *t-distribution* with  $n - 1$  degrees of freedom.