

## 5.11: The F Distribution

In this section we will study a distribution that has special importance in statistics. In particular, this distribution arises from ratios of sums of squares when sampling from a normal distribution, and so is important in estimation and in the two-sample normal model and in hypothesis testing in the two-sample normal model.

### Basic Theory

#### Definition

Suppose that  $U$  has the chi-square distribution with  $n \in (0, \infty)$  degrees of freedom,  $V$  has the chi-square distribution with  $d \in (0, \infty)$  degrees of freedom, and that  $U$  and  $V$  are independent. The distribution of

$$X = \frac{U/n}{V/d} \quad (5.11.1)$$

is the  $F$  distribution with  $n$  degrees of freedom in the numerator and  $d$  degrees of freedom in the denominator.

The  $F$  distribution was first derived by George Snedecor, and is named in honor of Sir Ronald Fisher. In practice, the parameters  $n$  and  $d$  are usually positive integers, but this is not a mathematical requirement.

#### Distribution Functions

Suppose that  $X$  has the  $F$  distribution with  $n \in (0, \infty)$  degrees of freedom in the numerator and  $d \in (0, \infty)$  degrees of freedom in the denominator. Then  $X$  has a continuous distribution on  $(0, \infty)$  with probability density function  $f$  given by

$$f(x) = \frac{\Gamma(n/2 + d/2)}{\Gamma(n/2)\Gamma(d/2)} \frac{n}{d} \frac{[(n/d)x]^{n/2-1}}{[1 + (n/d)x]^{n/2+d/2}}, \quad x \in (0, \infty) \quad (5.11.2)$$

where  $\Gamma$  is the gamma function.

Proof

The trick, once again, is conditioning. The conditional distribution of  $X$  given  $V = v \in (0, \infty)$  is gamma with shape parameter  $n/2$  and scale parameter  $2d/nv$ . Hence the conditional PDF is

$$x \mapsto \frac{1}{\Gamma(n/2)(2d/nv)^{n/2}} x^{n/2-1} e^{-x(nv/2d)} \quad (5.11.3)$$

By definition,  $V$  has the chi-square distribution with  $d$  degrees of freedom, and so has PDF

$$v \mapsto \frac{1}{\Gamma(d/2)2^{d/2}} v^{d/2-1} e^{-v/2} \quad (5.11.4)$$

The joint PDF of  $(X, V)$  is the product of these functions:

$$g(x, v) = \frac{1}{\Gamma(n/2)\Gamma(d/2)2^{(n+d)/2}} \left(\frac{n}{d}\right)^{n/2} x^{n/2-1} v^{(n+d)/2-1} e^{-v(n/d+1)/2}; \quad x, v \in (0, \infty) \quad (5.11.5)$$

The PDF of  $X$  is therefore

$$f(x) = \int_0^\infty g(x, v) dv = \frac{1}{\Gamma(n/2)\Gamma(d/2)2^{(n+d)/2}} \left(\frac{n}{d}\right)^{n/2} x^{n/2-1} \int_0^\infty v^{(n+d)/2-1} e^{-v(n/d+1)/2} dv \quad (5.11.6)$$

Except for the normalizing constant, the integrand in the last integral is the gamma PDF with shape parameter  $(n+d)/2$  and scale parameter  $2d/(nx+d)$ . Hence the integral evaluates to

$$\Gamma\left(\frac{n+d}{2}\right) \left(\frac{2d}{nx+d}\right)^{(n+d)/2} \quad (5.11.7)$$

Simplifying gives the result.

Recall that the beta function  $B$  can be written in terms of the gamma function by

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a, b \in (0, \infty) \quad (5.11.8)$$

Hence the probability density function of the  $F$  distribution above can also be written as

$$f(x) = \frac{1}{B(n/2, d/2)} \frac{n}{d} \frac{[(n/d)x]^{n/2-1}}{[1 + (n/d)x]^{n/2+d/2}}, \quad x \in (0, \infty) \quad (5.11.9)$$

When  $n \geq 2$ , the probability density function is defined at  $x = 0$ , so the support interval is  $[0, \infty)$  in this case.

In the special distribution simulator, select the  $F$  distribution. Vary the parameters with the scroll bars and note the shape of the probability density function. For selected values of the parameters, run the simulation 1000 times and compare the empirical density function to the probability density function.

Both parameters influence the shape of the  $F$  probability density function, but some of the basic *qualitative* features depend only on the numerator degrees of freedom. For the remainder of this discussion, let  $f$  denote the  $F$  probability density function with  $n \in (0, \infty)$  degrees of freedom in the numerator and  $d \in (0, \infty)$  degrees of freedom

in the denominator.

Probability density function  $f$  satisfies the following properties:

1. If  $0 < n < 2$ ,  $f$  is decreasing with  $f(x) \rightarrow \infty$  as  $x \downarrow 0$ .
2. If  $n = 2$ ,  $f$  is decreasing with mode at  $x = 0$ .
3. If  $n > 2$ ,  $f$  increases and then decreases, with mode at  $x = \frac{(n-2)d}{n(d+2)}$ .

Proof

These properties follow from standard calculus. The first derivative of  $f$  is

$$f'(x) = \frac{1}{B(n/2, d/2)} \left(\frac{n}{d}\right)^2 \frac{[(n/d)x]^{n/2-2}}{[1 + (n/2)x]^{n/2+d/2+1}} [(n/2-1) - (n/d)(d/2+1)x], \quad x \in (0, \infty) \quad (5.11.10)$$

Qualitatively, the second order properties of  $f$  also depend only on  $n$ , with transitions at  $n = 2$  and  $n = 4$ .

For  $n > 2$ , define

$$x_1 = \frac{d}{n} \frac{(n-2)(d+4) - \sqrt{2(n-2)(d+4)(n+d)}}{(d+2)(d+4)} \quad (5.11.11)$$

$$x_2 = \frac{d}{n} \frac{(n-2)(d+4) + \sqrt{2(n-2)(d+4)(n+d)}}{(d+1)(d+4)} \quad (5.11.12)$$

The probability density function  $f$  satisfies the following properties:

1. If  $0 < n \leq 2$ ,  $f$  is concave upward.
2. If  $2 < n \leq 4$ ,  $f$  is concave downward and then upward, with inflection point at  $x_2$ .
3. If  $n > 4$ ,  $f$  is concave upward, then downward, then upward again, with inflection points at  $x_1$  and  $x_2$ .

Proof

These results follow from standard calculus. The second derivative of  $f$  is

$$f''(x) = \frac{1}{B(n/2, d/2)} \left(\frac{n}{d}\right)^3 \frac{[(n/d)x]^{n/2-3}}{[1 + (n/d)x]^{n/2+d/2+2}} [(n/2-1)(n/2-2) - 2(n/2-1)(d/2+2)(n/d)x + (d/2+1)(d/2+2)(n/d)^2 x^2], \quad (5.11.13)$$

$$x \in (0, \infty)$$

The distribution function and the quantile function do not have simple, closed-form representations. Approximate values of these functions can be obtained from the special distribution calculator and from most mathematical and statistical software packages.

In the special distribution calculator, select the  $F$  distribution. Vary the parameters and note the shape of the probability density function and the distribution function. In each of the following cases, find the median, the first and third quartiles, and the interquartile range.

1.  $n = 5$ ,  $d = 5$
2.  $n = 5$ ,  $d = 10$
3.  $n = 10$ ,  $d = 5$
4.  $n = 10$ ,  $d = 10$

The general probability density function of the  $F$  distribution is a bit complicated, but it simplifies in a couple of special cases.

Special cases.

1. If  $n = 2$ ,

$$f(x) = \frac{1}{(1+2x/d)^{1+d/2}}, \quad x \in (0, \infty) \quad (5.11.14)$$

2. If  $n = d \in (0, \infty)$ ,

$$f(x) = \frac{\Gamma(n)}{\Gamma^2(n/2)} \frac{x^{n/2-1}}{(1+x)^n}, \quad x \in (0, \infty) \quad (5.11.15)$$

3. If  $n = d = 2$ ,

$$f(x) = \frac{1}{(1+x)^2}, \quad x \in (0, \infty) \quad (5.11.16)$$

4. If  $n = d = 1$ ,

$$f(x) = \frac{1}{\pi\sqrt{x(1+x)}}, \quad x \in (0, \infty) \quad (5.11.17)$$

## Moments

The random variable representation in the [definition](#), along with the moments of the chi-square distribution can be used to find the mean, variance, and other moments of the  $F$  distribution. For the remainder of this discussion, suppose that  $X$  has the  $F$  distribution with  $n \in (0, \infty)$  degrees of freedom in the numerator and  $d \in (0, \infty)$  degrees of freedom in the denominator.

### Mean

1.  $\mathbb{E}(X) = \infty$  if  $0 < d \leq 2$
2.  $\mathbb{E}(X) = \frac{d}{d-2}$  if  $d > 2$

### Proof

By independence,  $\mathbb{E}(X) = \frac{d}{n} \mathbb{E}(U) \mathbb{E}(V^{-1})$ . Recall that  $\mathbb{E}(U) = n$ . Similarly if  $d \leq 2$ ,  $\mathbb{E}(V^{-1}) = \infty$  while if  $d > 2$ ,

$$\mathbb{E}(V^{-1}) = \frac{\Gamma(d/2 - 1)}{2\Gamma(d/2)} = \frac{1}{d-2} \quad (5.11.18)$$

Thus, the mean depends only on the degrees of freedom in the denominator.

### Variance

1.  $\text{var}(X)$  is undefined if  $0 < d \leq 2$
2.  $\text{var}(X) = \infty$  if  $2 < d \leq 4$
3. If  $d > 4$  then

$$\text{var}(X) = 2 \left( \frac{d}{d-2} \right)^2 \frac{n+d-2}{n(d-4)} \quad (5.11.19)$$

### Proof

By independence,  $\mathbb{E}(X^2) = \frac{d^2}{n^2} \mathbb{E}(U^2) \mathbb{E}(V^{-2})$ . Recall that

$$\mathbb{E}(U^2) = 4 \frac{\Gamma(n/2 + 2)}{\Gamma(n/2)} = (n+2)n \quad (5.11.20)$$

Similarly if  $d \leq 4$ ,  $\mathbb{E}(V^{-2}) = \infty$  while if  $d > 4$ ,

$$\mathbb{E}(V^{-2}) = \frac{\Gamma(d/2 - 2)}{4\Gamma(d/2)} = \frac{1}{(d-2)(d-4)} \quad (5.11.21)$$

Hence  $\mathbb{E}(X^2) = \infty$  if  $d \leq 4$  while if  $d > 4$ ,

$$\mathbb{E}(X^2) = \frac{(n+2)d^2}{n(d-2)(d-4)} \quad (5.11.22)$$

The results now follow from the previous result on the [mean](#) and the computational formula  $\text{var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$ .

In the simulation of the special distribution simulator, select the  $F$  distribution. Vary the parameters with the scroll bar and note the size and location of the mean  $\pm$  standard deviation bar. For selected values of the parameters, run the simulation 1000 times and compare the empirical mean and standard deviation to the distribution mean and standard deviation..

### General moments. For $k > 0$ ,

1.  $\mathbb{E}(X^k) = \infty$  if  $0 < d \leq 2k$
2. If  $d > 2k$  then

$$\mathbb{E}(X^k) = \left( \frac{d}{n} \right)^k \frac{\Gamma(n/2 + k) \Gamma(d/2 - k)}{\Gamma(n/2) \Gamma(d/2)} \quad (5.11.23)$$

### Proof

By independence,  $\mathbb{E}(X^k) = \left( \frac{d}{n} \right)^k \mathbb{E}(U^k) \mathbb{E}(V^{-k})$ . Recall that

$$\mathbb{E}(U^k) = \frac{2^k \Gamma(n/2 + k)}{\Gamma(n/2)} \quad (5.11.24)$$

On the other hand,  $\mathbb{E}(V^{-k}) = \infty$  if  $d/2 \leq k$  while if  $d/2 > k$ ,

$$\mathbb{E}(V^{-k}) = \frac{2^{-k} \Gamma(d/2 - k)}{\Gamma(d/2)} \quad (5.11.25)$$

If  $k \in \mathbb{N}$ , then using the fundamental identity of the gamma distribution and some algebra,

$$\mathbb{E}(X^k) = \left( \frac{d}{n} \right)^k \frac{n(n+2) \cdots [n+2(k-1)]}{(d-2)(d-4) \cdots (d-2k)} \quad (5.11.26)$$

From the general moment formula, we can compute the skewness and kurtosis of the  $F$  distribution.

### Skewness and kurtosis

1. If  $d > 6$ ,

$$\text{skew}(X) = \frac{(2n+d-2)\sqrt{8(d-4)}}{(d-6)\sqrt{n(n+d-2)}} \quad (5.11.27)$$

2. If  $d > 8$ ,

$$\text{kurt}(X) = 3 + 12 \frac{n(5d-22)(n+d-2) + (d-4)(d-2)^2}{n(d-6)(d-8)(n+d-2)} \quad (5.11.28)$$

Proof

These results follow from the formulas for  $\mathbb{E}(X^k)$  for  $k \in \{1, 2, 3, 4\}$  and the standard computational formulas for skewness and kurtosis.

Not surprisingly, the  $F$  distribution is positively skewed. Recall that the *excess kurtosis* is

$$\text{kurt}(X) - 3 = 12 \frac{n(5d-22)(n+d-2) + (d-4)(d-2)^2}{n(d-6)(d-8)(n+d-2)} \quad (5.11.29)$$

In the simulation of the special distribution simulator, select the  $F$  distribution. Vary the parameters with the scroll bar and note the shape of the probability density function in light of the previous results on skewness and kurtosis. For selected values of the parameters, run the simulation 1000 times and compare the empirical density function to the probability density function.

## Relations

The most important relationship is the one in the [definition](#), between the  $F$  distribution and the chi-square distribution. In addition, the  $F$  distribution is related to several other special distributions.

Suppose that  $X$  has the  $F$  distribution with  $n \in (0, \infty)$  degrees of freedom in the numerator and  $d \in (0, \infty)$  degrees of freedom in the denominator. Then  $1/X$  has the  $F$  distribution with  $d$  degrees of freedom in the numerator and  $n$  degrees of freedom in the denominator.

Proof

This follows easily from the random variable interpretation in the [definition](#). We can write

$$X = \frac{U/n}{V/d} \quad (5.11.30)$$

where  $U$  and  $V$  are independent and have chi-square distributions with  $n$  and  $d$  degrees of freedom, respectively. Hence

$$\frac{1}{X} = \frac{V/d}{U/n} \quad (5.11.31)$$

Suppose that  $T$  has the  $t$  distribution with  $n \in (0, \infty)$  degrees of freedom. Then  $X = T^2$  has the  $F$  distribution with 1 degree of freedom in the numerator and  $n$  degrees of freedom in the denominator.

Proof

This follows easily from the random variable representations of the  $t$  and  $F$  distributions. We can write

$$T = \frac{Z}{\sqrt{V/n}} \quad (5.11.32)$$

where  $Z$  has the standard normal distribution,  $V$  has the chi-square distribution with  $n$  degrees of freedom, and  $Z$  and  $V$  are independent. Hence

$$T^2 = \frac{Z^2}{V/n} \quad (5.11.33)$$

Recall that  $Z^2$  has the chi-square distribution with 1 degree of freedom.

Our next relationship is between the  $F$  distribution and the exponential distribution.

Suppose that  $X$  and  $Y$  are independent random variables, each with the exponential distribution with rate parameter  $r \in (0, \infty)$ . Then  $Z = X/Y$  has the  $F$  distribution with 2 degrees of freedom in both the numerator and denominator.

Proof

We first find the distribution function  $F$  of  $Z$  by conditioning on  $X$ :

$$F(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(Y \geq X/z) = \mathbb{E}[\mathbb{P}(Y \geq X/z | X)] \quad (5.11.34)$$

But  $\mathbb{P}(Y \geq y) = e^{-ry}$  for  $y \geq 0$  so  $F(z) = \mathbb{E}(e^{-rX/z})$ . Also,  $X$  has PDF  $g(x) = re^{-rx}$  for  $x \geq 0$  so

$$F(z) = \int_0^\infty e^{-rx/z} re^{-rx} dx = \int_0^\infty re^{-rx(1+1/z)} dx = \frac{1}{1+1/z} = \frac{z}{1+z}, \quad z \in (0, \infty) \quad (5.11.35)$$

Differentiating gives the PDF of  $Z$

$$f(z) = \frac{1}{(1+z)^2}, \quad z \in (0, \infty) \quad (5.11.36)$$

which we recognize as the PDF of the  $F$  distribution with 2 degrees of freedom in the numerator and the denominator.

A simple transformation can change a variable with the  $F$  distribution into a variable with the beta distribution, and conversely.

Connections between the  $F$  distribution and the beta distribution.

1. If  $X$  has the  $F$  distribution with  $n \in (0, \infty)$  degrees of freedom in the numerator and  $d \in (0, \infty)$  degrees of freedom in the denominator, then

$$Y = \frac{(n/d)X}{1 + (n/d)X} \quad (5.11.37)$$

has the beta distribution with left parameter  $n/2$  and right parameter  $d/2$ .

2. If  $Y$  has the beta distribution with left parameter  $a \in (0, \infty)$  and right parameter  $b \in (0, \infty)$  then

$$X = \frac{bY}{a(1-Y)} \quad (5.11.38)$$

has the  $F$  distribution with  $2a$  degrees of freedom in the numerator and  $2b$  degrees of freedom in the denominator.

Proof

The two statements are equivalent and follow from the standard change of variables formula. The function

$$y = \frac{(n/d)x}{1 + (n/d)x} \quad (5.11.39)$$

maps  $(0, \infty)$  one-to-one onto  $(0, 1)$ , with inverse

$$x = \frac{d}{n} \frac{y}{1-y} \quad (5.11.40)$$

Let  $f$  denote the PDF of the  $F$  distribution with  $n$  degrees of freedom in the numerator and  $d$  degrees of freedom in the denominator, and let  $g$  denote the PDF of the beta distribution with left parameter  $n/2$  and right parameter  $d/2$ . Then  $f$  and  $g$  are related by

1.  $g(y) = f(x) \frac{dx}{dy}$
2.  $f(x) = g(y) \frac{dy}{dx}$

The  $F$  distribution is closely related to the beta prime distribution by a simple scale transformation.

Connections with the beta prime distributions.

1. If  $X$  has the  $F$  distribution with  $n \in (0, \infty)$  degrees of freedom in the numerator and  $d \in (0, \infty)$  degrees of freedom in the denominator, then  $Y = \frac{a}{d}X$  has the beta prime distribution with parameters  $n/2$  and  $d/2$ .
2. If  $Y$  has the beta prime distribution with parameters  $a \in (0, \infty)$  and  $b \in (0, \infty)$  then  $X = \frac{b}{a}Y$  has the  $F$  distribution with  $2a$  degrees of the freedom in the numerator and  $2b$  degrees of freedom in the denominator.

Proof

Let  $f$  denote the PDF of  $X$  and  $g$  the PDF of  $Y$ .

1. By the change of variables formula,

$$g(y) = \frac{d}{n} f\left(\frac{d}{n}y\right), \quad y \in (0, \infty) \quad (5.11.41)$$

Substituting into the beta  $F$  PDF shows that  $Y$  has the appropriate beta prime distribution.

2. Again using the change of variables formula,

$$f(x) = \frac{a}{b} g\left(\frac{a}{b}x\right), \quad x \in (0, \infty) \quad (5.11.42)$$

Substituting into the beta prime PDF shows that  $X$  has the appropriate  $F$  PDF.

## The Non-Central $F$ Distribution

The  $F$  distribution can be generalized in a natural way by replacing the ordinary chi-square variable in the numerator in the [definition above](#) with a variable having a non-central chi-square distribution. This generalization is important in analysis of variance.

Suppose that  $U$  has the non-central chi-square distribution with  $n \in (0, \infty)$  degrees of freedom and non-centrality parameter  $\lambda \in [0, \infty)$ ,  $V$  has the chi-square distribution with  $d \in (0, \infty)$  degrees of freedom, and that  $U$  and  $V$  are independent. The distribution of

$$X = \frac{U/n}{V/d} \quad (5.11.43)$$

is the *non-central  $F$  distribution* with  $n$  degrees of freedom in the numerator,  $d$  degrees of freedom in the denominator, and *non-centrality parameter*  $\lambda$ .

One of the most interesting and important results for the non-central chi-square distribution is that it is a Poisson mixture of ordinary chi-square distributions. This leads to a similar result for the non-central  $F$  distribution.

Suppose that  $N$  has the Poisson distribution with parameter  $\lambda/2$ , and that the conditional distribution of  $X$  given  $N$  is the  $F$  distribution with  $N + 2n$  degrees of freedom in the numerator and  $d$  degrees of freedom in the denominator, where  $\lambda \in [0, \infty)$  and  $n, d \in (0, \infty)$ . Then  $X$  has the non-central  $F$  distribution with  $n$  degrees of freedom in the numerator,  $d$  degrees of freedom in the denominator, and non-centrality parameter  $\lambda$ .

Proof

As in the theorem, let  $N$  have the Poisson distribution with parameter  $\lambda/2$ , and suppose also that the conditional distribution of  $U$  given  $N$  is chi-square with  $n+2N$  degrees of freedom, and that  $V$  has the chi-square distribution with  $d$  degrees of freedom and is independent of  $(N, U)$ . Let  $X = (U/n)/(V/d)$ . Since  $V$  is independent of  $(N, U)$ , the variable  $X$  satisfies the condition in the theorem; that is, the conditional distribution of  $X$  given  $N$  is the  $F$  distribution with  $n+2N$  degrees of freedom in the numerator and  $d$  degrees of freedom in the denominator. But then also, (unconditionally)  $U$  has the non-central chi-square distribution with  $n$  degrees of freedom in the numerator and non-centrality parameter  $\lambda$ ,  $V$  has the chi-square distribution with  $d$  degrees of freedom, and  $U$  and  $V$  are independent. So by definition  $X$  has the  $F$  distribution with  $n$  degrees of freedom in the numerator,  $d$  degrees of freedom in the denominator, and non-centrality parameter  $\lambda$ .

From the last result, we can express the probability density function and distribution function of the non-central  $F$  distribution as a series in terms of ordinary  $F$  density and distribution functions. To set up the notation, for  $j, k \in (0, \infty)$  let  $f_{jk}$  be the probability density function and  $F_{jk}$  the distribution function of the  $F$  distribution with  $j$  degrees of freedom in the numerator and  $k$  degrees of freedom in the denominator. For the rest of this discussion,  $\lambda \in [0, \infty)$  and  $n, d \in (0, \infty)$  as usual.

The probability density function  $g$  of the non-central  $F$  distribution with  $n$  degrees of freedom in the numerator,  $d$  degrees of freedom in the denominator, and non-centrality parameter  $\lambda$  is given by

$$g(x) = \sum_{k=0}^{\infty} e^{-\lambda/2} \frac{(\lambda/2)^k}{k!} f_{n+2k,d}(x), \quad x \in (0, \infty) \quad (5.11.44)$$

The distribution function  $G$  of the non-central  $F$  distribution with  $n$  degrees of freedom in the numerator,  $d$  degrees of freedom in the denominator, and non-centrality parameter  $\lambda$  is given by

$$G(x) = \sum_{k=0}^{\infty} e^{-\lambda/2} \frac{(\lambda/2)^k}{k!} F_{n+2k,d}(x), \quad x \in (0, \infty) \quad (5.11.45)$$

This page titled 5.11: The F Distribution is shared under a CC BY 2.0 license and was authored, remixed, and/or curated by Kyle Siegrist (Random Services) via source content that was edited to the style and standards of the LibreTexts platform.