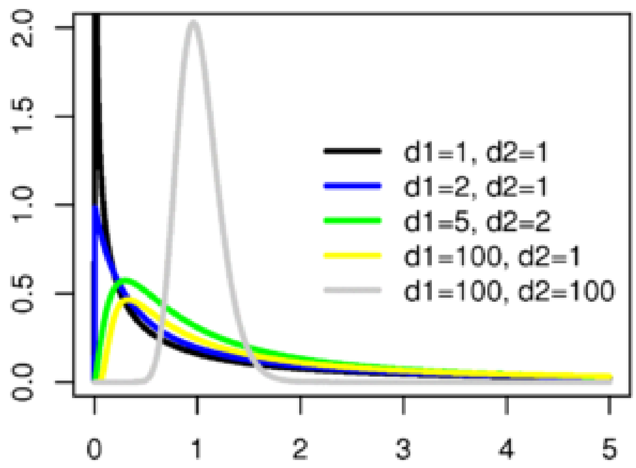
F-distribution

From Wikipedia, the free encyclopedia

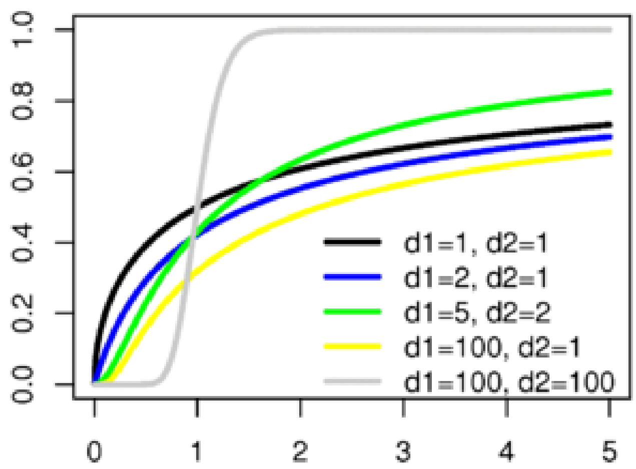
In probability theory and statistics, the ***F*-distribution** is a continuous probability distribution.[1][2][3][4] It is also known as **Snedecor's *F* distribution** or the **Fisher–Snedecor distribution** (after R. A. Fisher and George W. Snedecor). The *F*-distribution arises frequently as the null distribution of a test statistic, most notably in the analysis of variance; see *F*-test.

Fisher-Snedecor

Probability density function



Cumulative distribution function



Parameters	$d_1, d_2 > 0$ deg. of freedom
Support	$x \in [0, +\infty)$
pdf	$\frac{\sqrt{\frac{(d_1\,x)^{d_1}\,\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$
CDF	$I_{\frac{d_1 x}{d_1 x+d_2}}\left(\frac{d_1}{2},\frac{d_2}{2}\right)$
Mean	$\frac{d_2}{d_2-2}$ for $d_2 > 2$
Mode	$\frac{d_1-2}{d_1}\;\frac{d_2}{d_2+2}$ for $d_1 > 2$

Contents

Definition

If a random variable *X* has an *F*-distribution with parameters d_1 and d_2, we write $X \sim \mathrm{F}(d_1, d_2)$. Then the probability density function for *X* is given by

Variance	$\frac{2\,d_2^2\,(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ for $d_2 > 4$
Skewness	$\frac{(2d_1+d_2-2)\sqrt{8(d_2-4)}}{(d_2-6)\sqrt{d_1(d_1+d_2-2)}}$ for $d_2 > 6$
Ex. kurtosis	*see text*
MGF	*does not exist, raw moments defined in text and in* [1][2]
CF	*see text*

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1\,x)^{d_1}\,\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)} = \frac{1}{\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)}\left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1}\left(1+\frac{d_1}{d_2}\,x\right)^{-\frac{d_1+d_2}{2}}$$

for real $x \geq 0$. Here B is the beta function. In many applications, the parameters d_1 and d_2 are positive integers, but the distribution is well-defined for positive real values of these parameters.

The cumulative distribution function is

$$F(x; d_1, d_2) = I_{\frac{d_1 x}{d_1 x + d_2}}\left(\tfrac{d_1}{2}, \tfrac{d_2}{2}\right),$$

where I is the regularized incomplete beta function.

The expectation, variance, and other details about the F(d_1, d_2) are given in the sidebox; for $d_2 > 8$, the excess kurtosis is

$$\gamma_2 = 12\frac{d_1(5d_2-22)(d_1+d_2-2)+(d_2-4)(d_2-2)^2}{d_1(d_2-6)(d_2-8)(d_1+d_2-2)}.$$

The k-th moment of an F(d_1, d_2) distribution exists and is finite only when $2k < d_2$ and it is equal to [5]

$$\mu_X(k) = \left(\frac{d_2}{d_1}\right)^k \frac{\Gamma\left(\tfrac{d_1}{2}+k\right)}{\Gamma\left(\tfrac{d_1}{2}\right)} \frac{\Gamma\left(\tfrac{d_2}{2}-k\right)}{\Gamma\left(\tfrac{d_2}{2}\right)}$$

The F-distribution is a particular parametrization of the beta prime distribution, which is also called the beta distribution of the second kind.

The characteristic function is listed incorrectly in many standard references (e.g., [2]). The correct expression [6] is

$$\varphi^F_{d_1,d_2}(s) = \frac{\Gamma(\frac{d_1+d_2}{2})}{\Gamma(\frac{d_2}{2})} U\left(\frac{d_1}{2}, 1-\frac{d_2}{2}, -\frac{d_2}{d_1}\imath s\right)$$

where $U(a, b, z)$ is the confluent hypergeometric function of the second kind.

Characterization

A random variate of the F-distribution with parameters d_1 and d_2 arises as the ratio of two appropriately scaled chi-squared variates:[7]

$$X = \frac{U_1/d_1}{U_2/d_2}$$

where

- U_1 and U_2 have chi-squared distributions with d_1 and d_2 degrees of freedom respectively, and
- U_1 and U_2 are independent.

In instances where the F-distribution is used, for example in the analysis of variance, independence of U_1 and U_2 might be demonstrated by applying Cochran's theorem.

Equivalently, the random variable of the F-distribution may also be written

$$X = \frac{s_1^2}{\sigma_1^2} \;/\; \frac{s_2^2}{\sigma_2^2}$$

where s_1^2 and s_2^2 are the sums of squares S_1^2 and S_2^2 from two normal processes with variances σ_1^2 and σ_2^2 divided by the corresponding number of χ^2 degrees of freedom, d_1 and d_2 respectively.[*citation needed*]

In a frequentist context, a scaled F-distribution therefore gives the probability $p(s_1^2/s_2^2 \mid \sigma_1^2, \sigma_2^2)$, with the F-distribution itself, without any scaling, applying where σ_1^2 is being taken equal to σ_2^2. This is the context in which the F-distribution most generally appears in F-tests: where the null hypothesis is that two independent normal variances are equal, and the observed sums of some appropriately selected squares are then examined to see whether their ratio is significantly incompatible with this null hypothesis.

The quantity X has the same distribution in Bayesian statistics, if an uninformative rescaling-invariant Jeffreys prior is taken for the prior probabilities of σ_1^2 and σ_2^2.[8] In this context, a scaled F-distribution thus gives the posterior probability $p(\sigma_2^2/\sigma_1^2 | s_1^2, s_2^2)$, where now the observed sums s_1^2 and s_2^2 are what are taken as known.

Generalization

A generalization of the (central) F-distribution is the noncentral F-distribution.

Related distributions and properties

- If $X \sim \chi^2_{d_1}$ and $Y \sim \chi^2_{d_2}$ are independent, then $\frac{X/d_1}{Y/d_2} \sim \mathrm{F}(d_1, d_2)$
- If $X \sim \mathrm{Beta}(d_1/2, d_2/2)$ (Beta distribution) then $\frac{d_2 X}{d_1(1-X)} \sim \mathrm{F}(d_1, d_2)$
- Equivalently, if $X \sim \mathrm{F}(d_1, d_2)$, then $\frac{d_1 X/d_2}{1 + d_1 X/d_2} \sim \mathrm{Beta}(d_1/2, d_2/2)$.
- If $X \sim \mathrm{F}(d_1, d_2)$ then $Y = \lim_{d_2 \to \infty} d_1 X$ has the chi-squared distribution $\chi^2_{d_1}$
- $\mathrm{F}(d_1, d_2)$ is equivalent to the scaled Hotelling's T-squared distribution $\frac{d_2}{d_1(d_1 + d_2 - 1)} \mathrm{T}^2(d_1, d_1 + d_2 - 1)$.
- If $X \sim \mathrm{F}(d_1, d_2)$ then $X^{-1} \sim \mathrm{F}(d_2, d_1)$.
- If $X \sim \mathrm{t}(n)$ then

$$X^2 \sim \mathrm{F}(1, n)$$
$$X^{-2} \sim \mathrm{F}(n, 1)$$

- *F*-distribution is a special case of type 6 Pearson distribution
- If X and Y are independent, with $X, Y \sim$ Laplace(μ, b) then

$$\tfrac{|X-\mu|}{|Y-\mu|} \sim \mathrm{F}(2, 2)$$

- If $X \sim \mathrm{F}(n, m)$ then $\tfrac{\log X}{2} \sim \mathrm{FisherZ}(n, m)$ (Fisher's z-distribution)
- The noncentral *F*-distribution simplifies to the *F*-distribution if $\lambda = 0$.
- The doubly noncentral *F*-distribution simplifies to the *F*-distribution if $\lambda_1 = \lambda_2 = 0$
- If $\mathrm{Q}_X(p)$ is the quantile p for $X \sim \mathrm{F}(d_1, d_2)$ and $\mathrm{Q}_Y(1-p)$ is the quantile $1-p$ for $Y \sim \mathrm{F}(d_2, d_1)$, then

$$\mathrm{Q}_X(p) = \frac{1}{\mathrm{Q}_Y(1-p)}.$$

See also

- Chi-squared distribution
- Chow test
- Gamma distribution
- Hotelling's T-squared distribution
- Student's t-distribution
- Wilks' lambda distribution
- Wishart distribution

References

1. ^ ***a*** ***b*** Johnson, Norman Lloyd; Samuel Kotz, N. Balakrishnan (1995). *Continuous Univariate Distributions,*

Volume 2 (Second Edition, Section 27). Wiley. ISBN 0-471-58494-0.

2. ^ ***a b c*** Abramowitz, Milton; Stegun, Irene A., eds. (1965), "Chapter 26" (http://www.math.sfu.ca/~cbm/aands/page_946.htm), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, p. 946, ISBN 978-0486612720, MR 0167642 (//www.ams.org/mathscinet-getitem?mr=0167642).
3. ^ NIST (2006). Engineering Statistics Handbook – F Distribution (http://www.itl.nist.gov/div898/handbook/eda/section3/eda3665.htm)
4. ^ Mood, Alexander; Franklin A. Graybill, Duane C. Boes (1974). *Introduction to the Theory of Statistics (Third Edition, pp. 246–249)*. McGraw-Hill. ISBN 0-07-042864-6.
5. ^ Taboga, Marco. "The F distribution" (http://www.statlect.com/F_distribution.htm).
6. ^ Phillips, P. C. B. (1982) "The true characteristic function of the F distribution," *Biometrika*, 69: 261–264 JSTOR 2335882 (http://www.jstor.org/stable/2335882)
7. ^ M.H. DeGroot (1986), *Probability and Statistics* (2nd Ed), Addison-Wesley. ISBN 0-201-11366-X, p. 500
8. ^ G.E.P. Box and G.C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley. p.110

External links

- Table of critical values of the *F*-distribution (http://www.itl.nist.gov/div898/handbook/eda/section3/eda3673.htm)
- Earliest Uses of Some of the Words of Mathematics: entry on *F*-distribution contains a brief history (http://jeff560.tripod.com/f.html)
- Free calculator for *F*-testing (http://www.waterlog.info/f-test.htm)