Note on F-distribution and F-test:

Mathematical Framework:

F distribution:

$$f(F;m,n) = \frac{m^{m/2}\, n^{n/2}\, \Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\,\Gamma(n/2)} \cdot \frac{F^{\frac{m}{2}-1}}{(mF+n)^{\frac{m+n}{2}}}$$

$$\begin{cases} m,n \in \mathbb{N} \\ F \in \mathbb{R}^+ \end{cases}$$

Since the Beta Function is:

$$B(x,y) = \int_0^1 t^{y-1}(1-t)^{x-1}\,dt$$

$$= \boxed{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{(x-1)!\,(y-1)!}{(x+y-1)!}$$

$$\longrightarrow B(x,y) = B(y,x)$$

Then the other form would be:

$$f(F;m,n) = \frac{m^{m/2}\, n^{n/2}}{B(m/2, n/2)} \cdot \frac{F^{\frac{m}{2}-1}}{(mF+n)^{\frac{m+n}{2}}}$$

Or in the explicit form:

$$f(F;m,n) = \frac{m^{m/2}\, n^{n/2}\left(\frac{m+n}{2}-1\right)!}{\left(\frac{m}{2}-1\right)!\left(\frac{n}{2}-1\right)!} \cdot \frac{F^{\frac{m}{2}-1}}{(mF+n)^{m+n/2}}$$

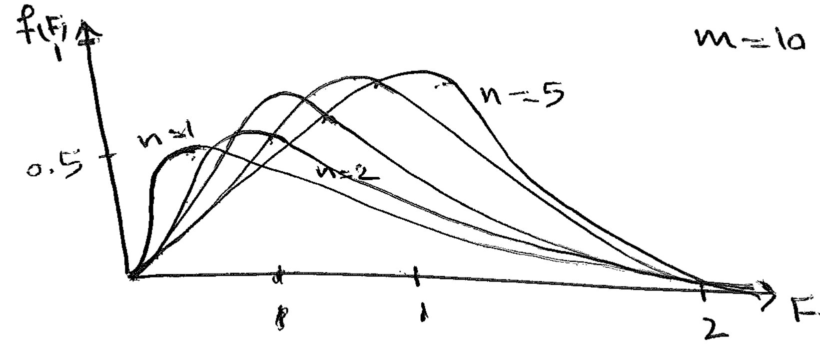
* This distribution is often called the Fisher F-distribution, after the famous British statistician Sir Ronald Aylmer Fisher (1890-1962), sometimes the Snedecor F-distribution and sometimes Fisher-Snedecor F-distribution.

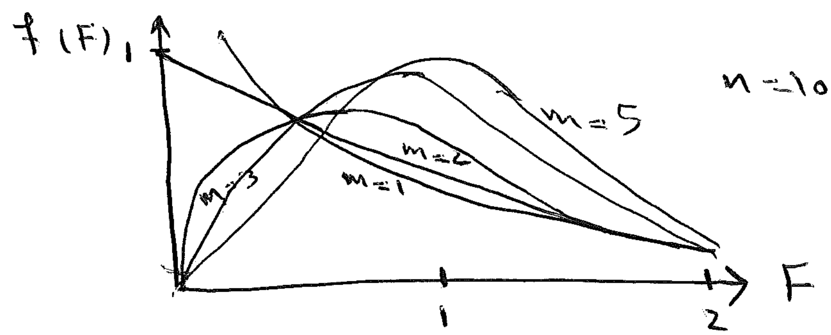
* m and n are called "degrees of freedom"

@ In statistics, the number of degrees of freedom is the number of independent pieces of data being used to make a calculation.

* It is customary to plot the f distribution (any of the formulae given for $f(F; m, n)$ here) against F, for different values of 'n' and 'm'.

EX.





*** Mathematical Background:

Definitions:

(terms "density" here is interchangable with "distribution")

* Probability Mas Function (PMF):

Probability that "discrete" random variable will exactly equal a discrete value.

* Probability Density Function (PDF):

Probability that "Continuous" random variable will fall within an interval.

* Cumulative Density Function (CDF):

Probability that either a <u>discrete</u> or <u>continuous</u> random variable will take a value less than or equal to a certain value.

* Characteristic Function:

Fourier transform of the probability density function.

* Chi-distribution:

A direct relation exists between Chi and Gaussian distribution of random variables. If 'x' is a Gaussian random variable and $y = x^2$, then y has a Chi-distribution with <u>one</u> degree of freedom.

** Notation:

* Probability Density Function:

$$P(r) = \sum_{z=-\infty}^{r} P(z) \quad \& \quad P(r) = \int_{-\infty}^{r} f(t)\,dt$$

discrete continuous

These functions are assumed to be properly normalized:

$$\sum_{r} P(r) = 1 \quad \& \quad \int_{-\infty}^{\infty} f(x)\,dx$$

* Characteristic Function:

It is denoted either by Φ or φ:

$$\begin{cases} \Phi(0) = 1 \\ |\Phi(t)| \leq 1 \end{cases} \qquad \Phi(t) = \int_{-\infty}^{+\infty} e^{ixt} f(x)\, dx$$

The Inverse: $f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi(t)\, e^{-ixt}\, dt$

+++ In some texts, characteristic Function is denoted as a function of 'ix'. Some of the notations include it as a function of frequency (ω).

$$\varphi(j\omega) = \int_{-\infty}^{+\infty} f(x)\, e^{j\omega x}\, dx$$

where $\underline{j}$ is used for $\sqrt{-1}$ instead of $\underline{i}$ to avoid confusion (with i index).

* Derivation of the Chi-Square distribution:

We work our way out, with the Cumulative Density Function (CDF). CDF is denoted by $F(x)$:

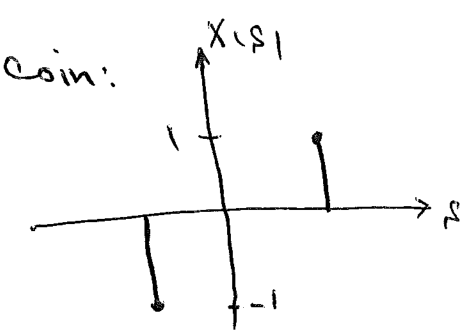
$$F(x) = P(X \leq x) \qquad \text{obviously: } 0 \leq F(x) \leq 1$$

$$F(-\infty) = 0 \quad , \quad F(\infty) = 1$$

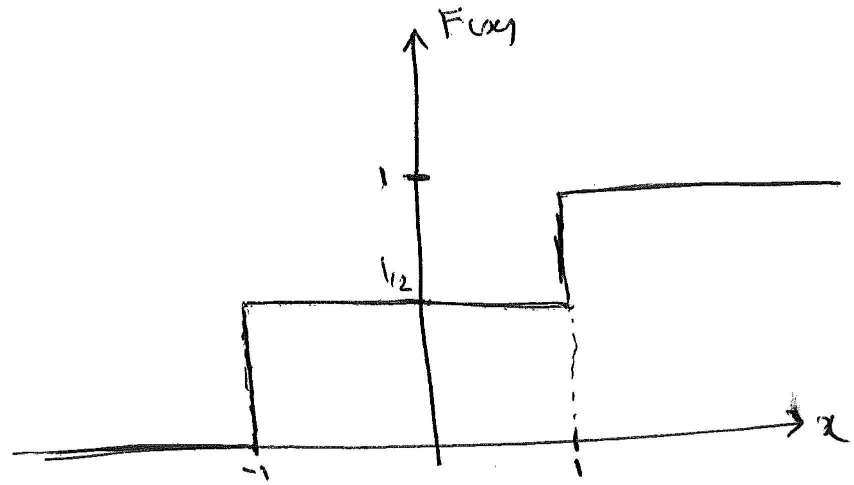
EX.

In case of flipping a coin:

$$X(s) = \begin{cases} 1 & s = H \\ -1 & s = T \end{cases}$$



CDF plot:



Probability density function is related to CDF by the following equation:

$$\frac{dF(x)}{dx} = p(x) \qquad -\infty < x < +\infty$$

$$F(x) = \int_{-\infty}^{x} p(u)\,du$$

Probability of X between x_1 and x_2 $(x_2 > x_1)$

$$\overbrace{P(X \leq x_2)}^{F(x_2)} = \overbrace{P(X \leq x_1)}^{F(x_1)} + P(x_1 < X \leq x_2)$$

$$F(x_2) - F(x_1) = P(x_1 < X \leq x_2)$$

* To drive the Chi distribution equation, let's start with a simple case and go on from there:

$X, Y \equiv$ random variables

1) $$Y = aX + b \qquad a, b: \text{constants}$$

$F_X(x) \equiv \text{CDF}$ for X

$F_Y(y) \equiv \text{CDF}$ for Y

$$F_Y(y) = P(Y \leqslant y) = P(aX + b \leqslant y) = P\left(X \leqslant \frac{y-b}{a}\right)$$

$$F_Y(y) = \int_{-\infty}^{\frac{y-b}{a}} P_X(x)\,dx = F_X\left(\frac{y-b}{a}\right)$$

$$\Rightarrow \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right)$$

$$\Rightarrow P_Y(y) = \frac{1}{a} P_X\left(\frac{y-b}{a}\right)$$

2) We now will calculate for a more complex case:

$$Y = aX^2 + b$$

$$F_Y(y) = P(Y \leqslant y) = P(aX^2 + b \leqslant y)$$

$$= P\left(|X| \leqslant \sqrt{\frac{y-b}{a}}\right)$$

$$= \underbrace{\int_{-\sqrt{\frac{y-b}{a}}}^{-\infty} p(x)\,dx}_{-\int_{-\infty}^{-\sqrt{\frac{y-b}{a}}} p(x)\,dx} + \int_{-\infty}^{\sqrt{\frac{y-b}{a}}} p(x)\,dx$$

$$= -F_X\left(-\sqrt{\frac{y-b}{a}}\right) + F_X\left(\sqrt{\frac{y-b}{a}}\right)$$

$$\Rightarrow \frac{dF_Y(y)}{dy} = P_Y(y) = \frac{1/a}{2\sqrt{\frac{y-b}{a}}} P_X\left(-\sqrt{\frac{y-b}{a}}\right) + \frac{1/a}{2\sqrt{\frac{y-b}{a}}} P_X\left(\sqrt{\frac{y-b}{a}}\right)$$

** Finally in order to obtain the solution for chi distribution we consider $Y = X^2$.

which means : $a = 1 \quad , \quad b = 0$

$$\Rightarrow P_y(y) = \frac{1}{2\sqrt{y}} P_x(-\sqrt{y}) + \frac{1}{2\sqrt{y}} P_x(\sqrt{y})$$

Now if we assume the general form of a Gaussian distribution to be:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

by assuming $a = 0$ (without losing generality), we obtain:

$$P_x(\sqrt{y}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y/2\sigma^2} \quad , \quad P(-\sqrt{y}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y/2\sigma^2}$$

Then: $$P_y(y) = \frac{1}{\sqrt{2\pi y\sigma^2}} e^{-y/2\sigma^2}$$

Finally the characteristic Function would be:

$$F_y(y) = \int_0^y P_y(U)\,dU = \int_0^y \frac{1}{\sqrt{2\pi U\sigma^2}} e^{-U/2\sigma^2} dU$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^y \frac{1}{\sqrt{U}} e^{-U/2\sigma^2} dU$$

$$\Rightarrow \varphi(i\omega) = \frac{1}{(1 - 2i\omega\sigma^2)^{1/2}}$$ ✓

The inverse transform of the obtained characteristic Function leads to probability density function:

$$P_y(y) = \frac{1}{\sqrt{2}\,\sigma\,\underbrace{\sqrt{\pi}}_{\Gamma(\frac{1}{2})}}\, y^{-1}\, e^{-y/2\sigma^2}$$

** Since for a system with $\underline{n}$ degrees of freedom we have:

$$\varphi(i\omega) = \frac{1}{(1-2i\omega\,\sigma^2)^{n/2}}$$

(obtained in an identical way to the 1 degree of freedom system, just assume: $Y = \sum_{i=1}^{n} X_i^2$)

thus:

$$P_y(y) = \frac{1}{\sigma^n\, 2^{n/2}\, \Gamma(n/2)}\, y^{\frac{n}{2}-1}\, e^{-y/2\sigma^2}$$

which is called a chi-square pdf with n degrees of freedom.

END of Mathematical Background

A practical example of F-distribution is given in the following

*Variance Ratio:

we could use F-distribution when estimates of the variance for two independent samples from normal distributions

standard deviations $\sim$ $S_1^2 = \sum_{i=1}^{m} \frac{(x_i - \bar{x})^2}{m-1}$, $S_2^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}$

have been made.

Here S_1^2 and S_2^2 are estimates of σ_1^2 and σ_2^2. That is to say $(m-1)S_1^2/\sigma_1^2$ and $(n-1)S_2^2/\sigma_2^2$ are distributed according to the chi-square distribution with $m-1$ and $n-1$ degrees of freedom respectively.

In this case the quantity

$$F = \frac{S_1^2}{\sigma_1^2} \cdot \frac{\sigma_2^2}{S_2^2}$$

is distributed according to the F-distribution with $m-1$ and $n-1$ degrees of freedom.

** If the true variances of the two populations are indeed the same, then the variance ratio S_1^2/S_2^2 has

F- distribution.

~~$\alpha\alpha$~~ We would reject the null hypotheses at the α confidence level if the F-ratio is less than $F_{(1-\alpha/2),(m-1),(n-1)}$ or greater than $F_{(\alpha/2),(m-1),(n-1)}$

where ~~α~~ $F_{\alpha,m,n}$ is defined by:

$$\int_0^{F_{\alpha,m,n}} f(F;m,n)\,dF = 1-\alpha$$

~~which~~
where α would be the probability content of the distribution above the value $F_{\alpha,m-1,n-1}$.