Continuous Distributions
Part 1

DASC 512

Overview

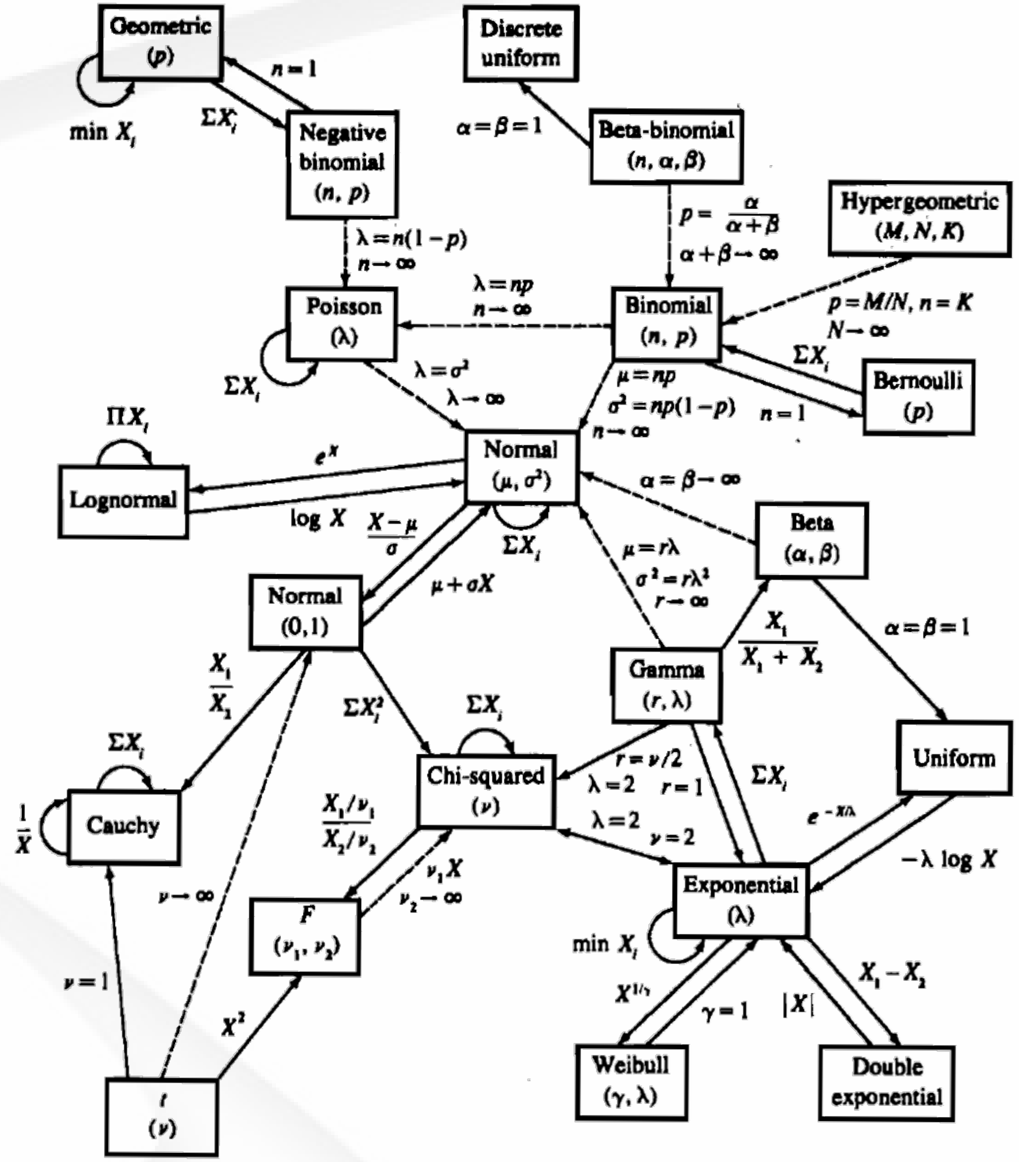
- Normal distribution
- Student's *t* distribution
- Chi Squared distribution
- F distribution
- Lognormal distribution
- Exponential distribution
- Beta distribution
- Uniform distribution
- Other distributions

Normal Distribution

Also known as the “Gaussian distribution” or “bell curve”

The most important family of distributions for this course

Most data problems are either normally distributed or can be transformed to the normal distribution to apply standard techniques



Source: Casella & Berger, *Statistical Inference 2nd Ed.*, 2002.

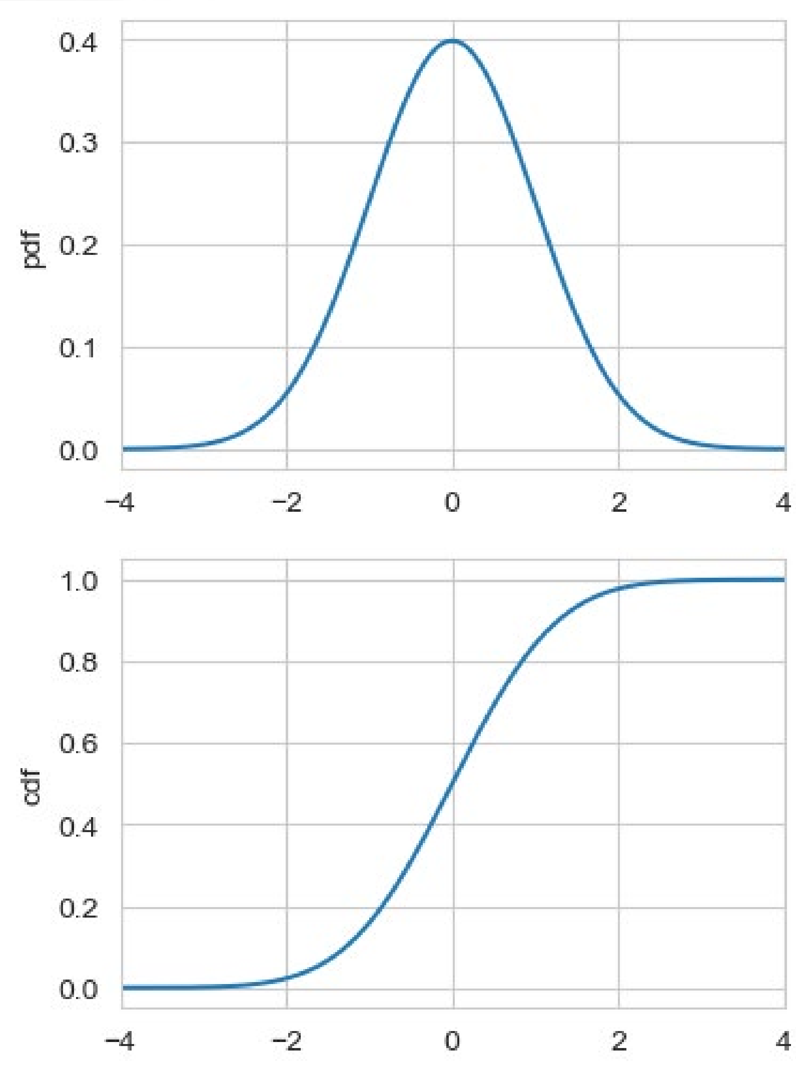
Normal Distribution

$$X \sim \mathrm{N}(\mu, \sigma)$$

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

$$\mu = \mu$$

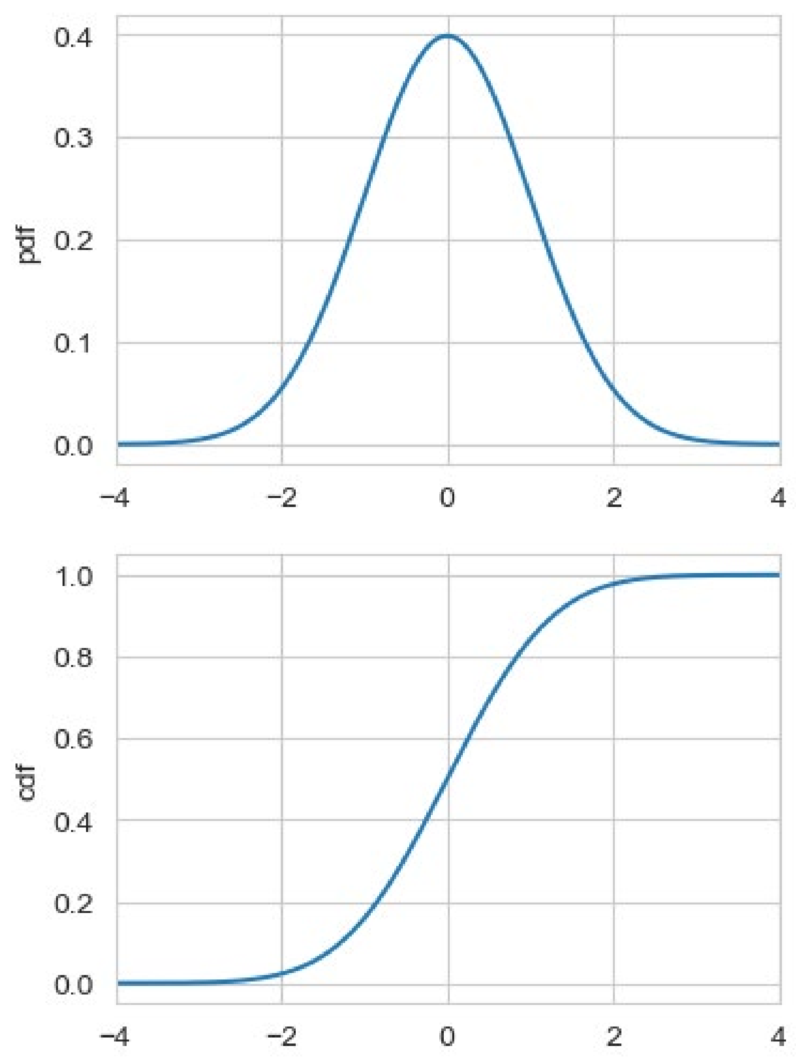
$$\sigma^2 = \sigma^2$$



Normal Distribution

In SciPy:

scipy.stats.norm(x, loc=μ, scale=σ)



Let’s talk about beer…

- The early 1900s was a time of scientific explosion
  - Industries hired scientists to study and perfect processes
  - Most scientific endeavors benefit from statistical process control
- Brewing beer is applied organic chemistry
  - A chemist had published a paper revealing Guiness trade secrets
  - The board barred publications mentioning beer or including a surname
- William Sealy Gosset, the head brewer for Guinness, discovered that sample means of normally distributed data had a specific distribution that was not quite normal. He published it under the pseudonym “Student”.

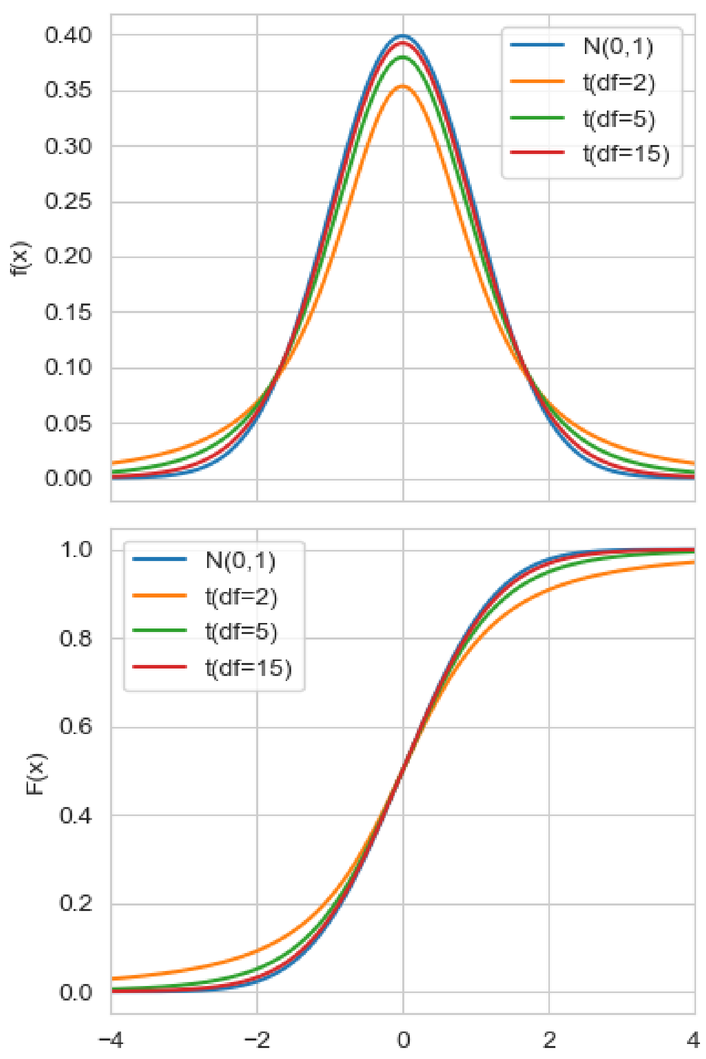
Student's *t*-distribution

The *t* distribution with $\nu = n - 1$ degrees of freedom corrects for randomness in the observed variance of a sample of size n. For large samples ($n \to \infty$) this converges to the standard normal distribution.

$$X \sim t(\nu)$$

$$\mu = 0, \qquad \nu > 1$$

$$\sigma^2 = \frac{\nu}{\nu - 2}, \qquad \nu > 2$$

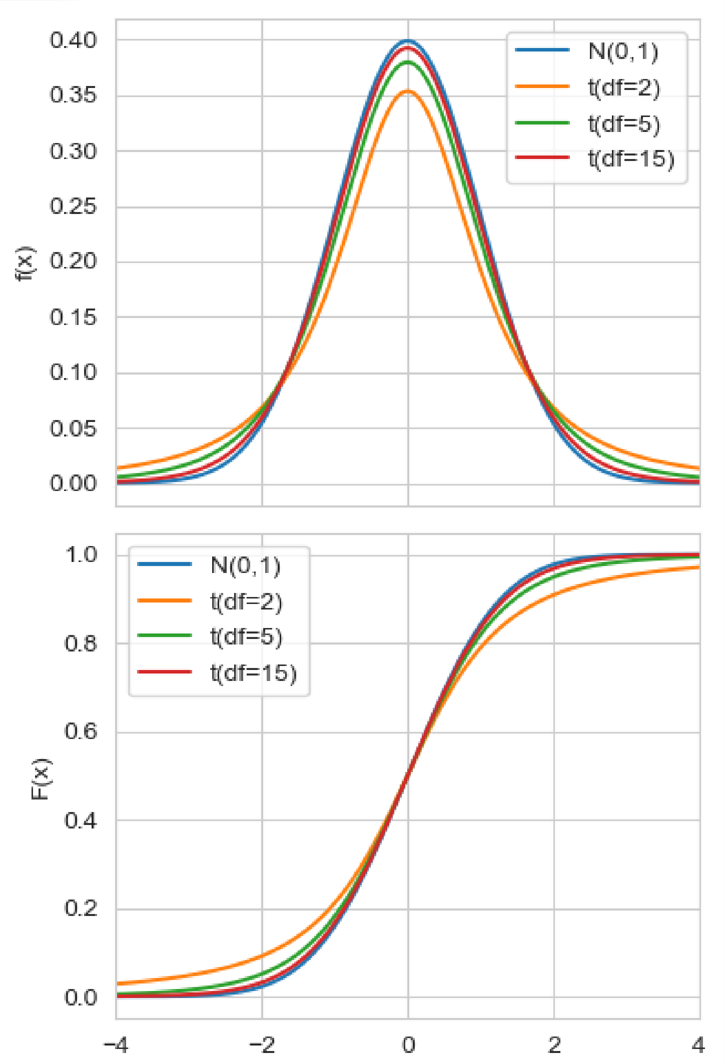


Student's *t*-distribution

In SciPy:

scipy.stats.t(x, df=ν, loc=μ, scale=σ)

- μ and σ here refer to the location and scale family of *t* distributions



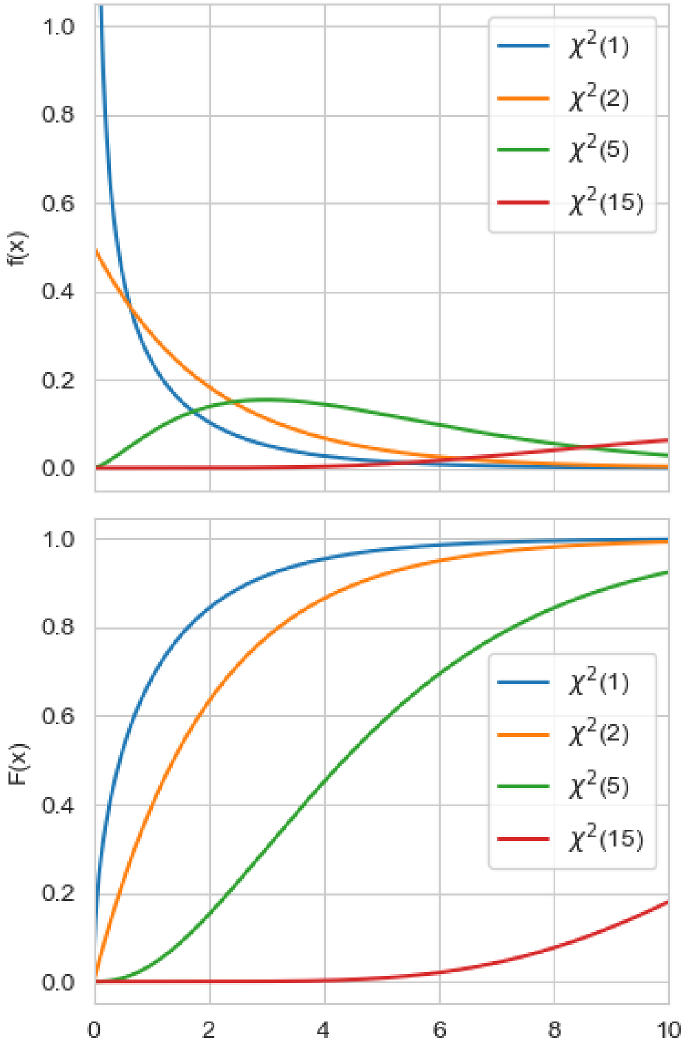
χ^2 (Chi-Squared) Distribution

The sum of the squares of *k* independent random variables $Z_1, Z_2, \dots, Z_k \sim N(0,1)$ is itself a random variable distributed χ^2 with *k* degrees of freedom.

$$\sum_{i=1}^{k} Z^2 = X \sim \chi^2(k)$$

$$\mu = k$$

$$\sigma^2 = 2k$$



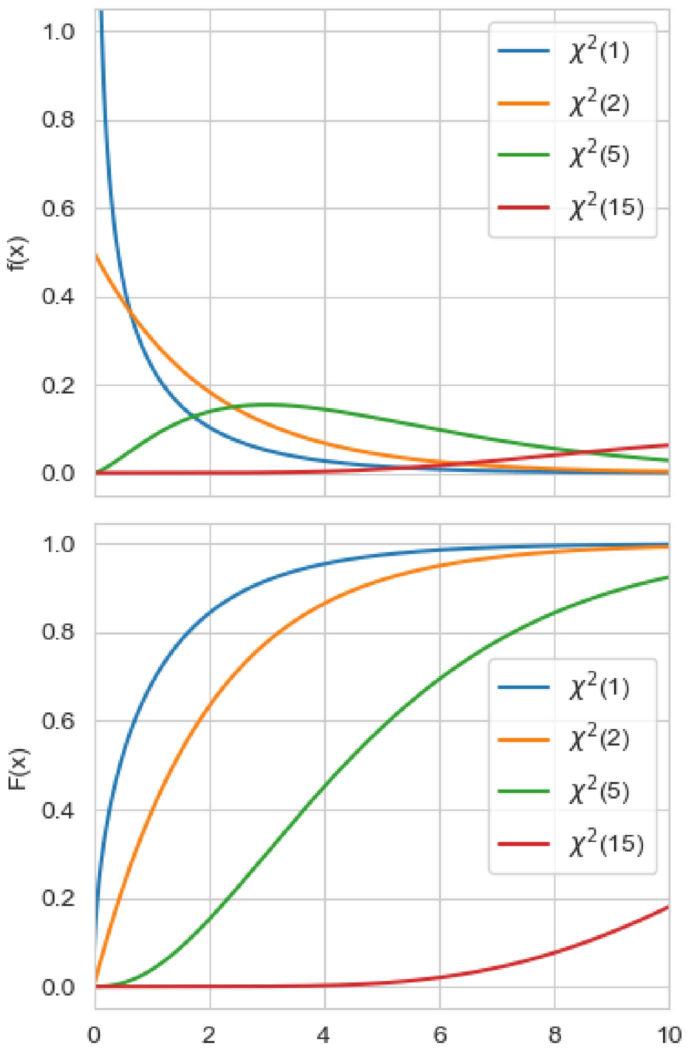
χ^2 (Chi-Squared) Distribution

Let $Z_1, \ldots, Z_k$ be distributed *N*(0,1). Let $X_i = \sigma Z_i$.

$$\sum_{i=1}^{k} (Z_i - \bar{Z})^2 \sim \chi^2_{k-1}$$

$$\frac{\sum_{i=1}^{k} (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{k-1}$$

$$\sigma^2 \sim \frac{\sum_{i=1}^{k} (X_i - \bar{X})^2}{\chi^2_{k-1}}$$

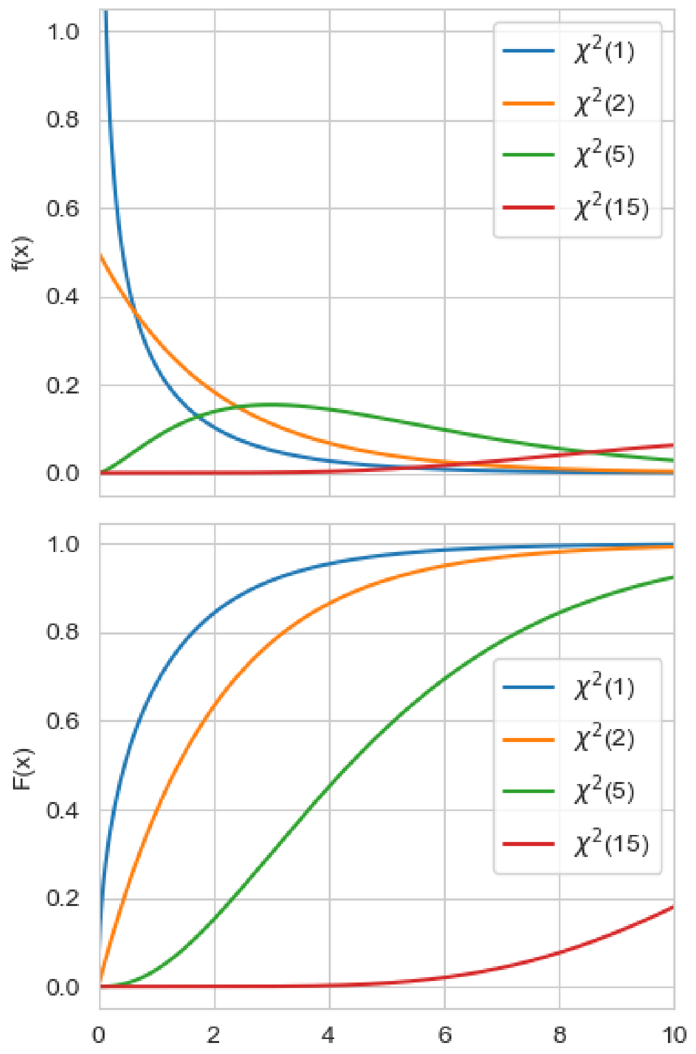


χ^2 (Chi-Squared) Distribution

In SciPy:

scipy.stats.chi2(x, df=k, loc=μ, scale=σ)

- μ and σ here refer to the location and scale family of χ^2 distributions (not commonly used)



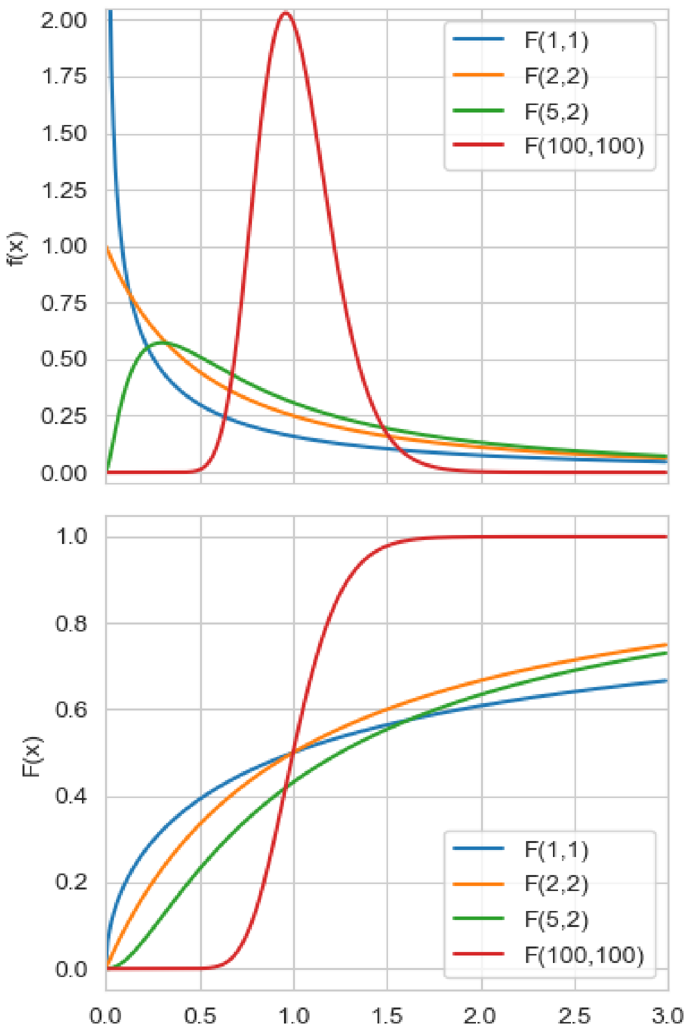
F Distribution

The ratio of two independent χ^2-distributed random variables divided by their degrees of freedom is a random variable that is *F* distributed.

$$\left(\frac{\chi^2_{r_1}}{r_1}\right)\bigg/\left(\frac{\chi^2_{r_2}}{r_2}\right) = F(r_1, r_2)$$

$$\mu = \frac{r_2}{r_2 - 2}, r_2 > 2$$

$$\sigma^2 = 2\left(\frac{r_2}{r_2 - 2}\right)^2 \frac{r_1 + r_2 - 2}{r_1(r_2 - 4)}, r_2 > 4$$

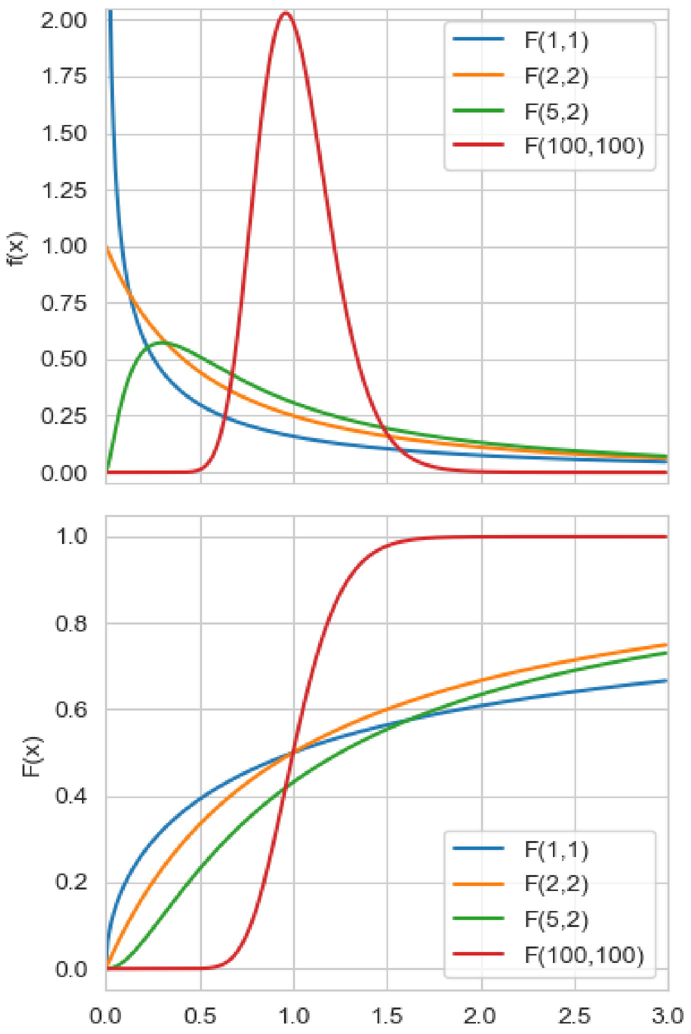


F Distribution

In SciPy:

scipy.stats.f(x, dfn=r_1, dfd=r_2, loc=μ, scale=σ)

- μ and σ here refer to the location and scale family of *F* distributions (not commonly used)



Resources

Wikipedia

- https://en.Wikipedia.org

SciPy.Stats Reference

- https://docs.scipy.org/doc/scipy/reference/stats.html

For deep theory, the STAT 601/602 textbook

- Casella, G., & Berger, R. L. (2002). *Statistical inference*. Cengage Learning.

Recap

- Normal distribution
- Student's *t* distribution
- Chi Squared distribution
- F distribution
- Lognormal distribution
- Exponential distribution
- Beta distribution
- Uniform distribution
- Other distributions