Biostatistics
Week VI

Ege Ülgen, MD, PhD

10 November 2022

ACIBADEM
MEHMET ALİ AYDINLAR
ÜNİVERSİTESİ

Probability Density Function (PDF)

- Probability density function is the probability distribution of a continuous random variable
- It provides the possible values and their associated probabilities

$$f: \mathbb{R} \to [0,1]$$

$$f_X(x) = P(X = x)$$

Properties of a Proper PDF (f_X)

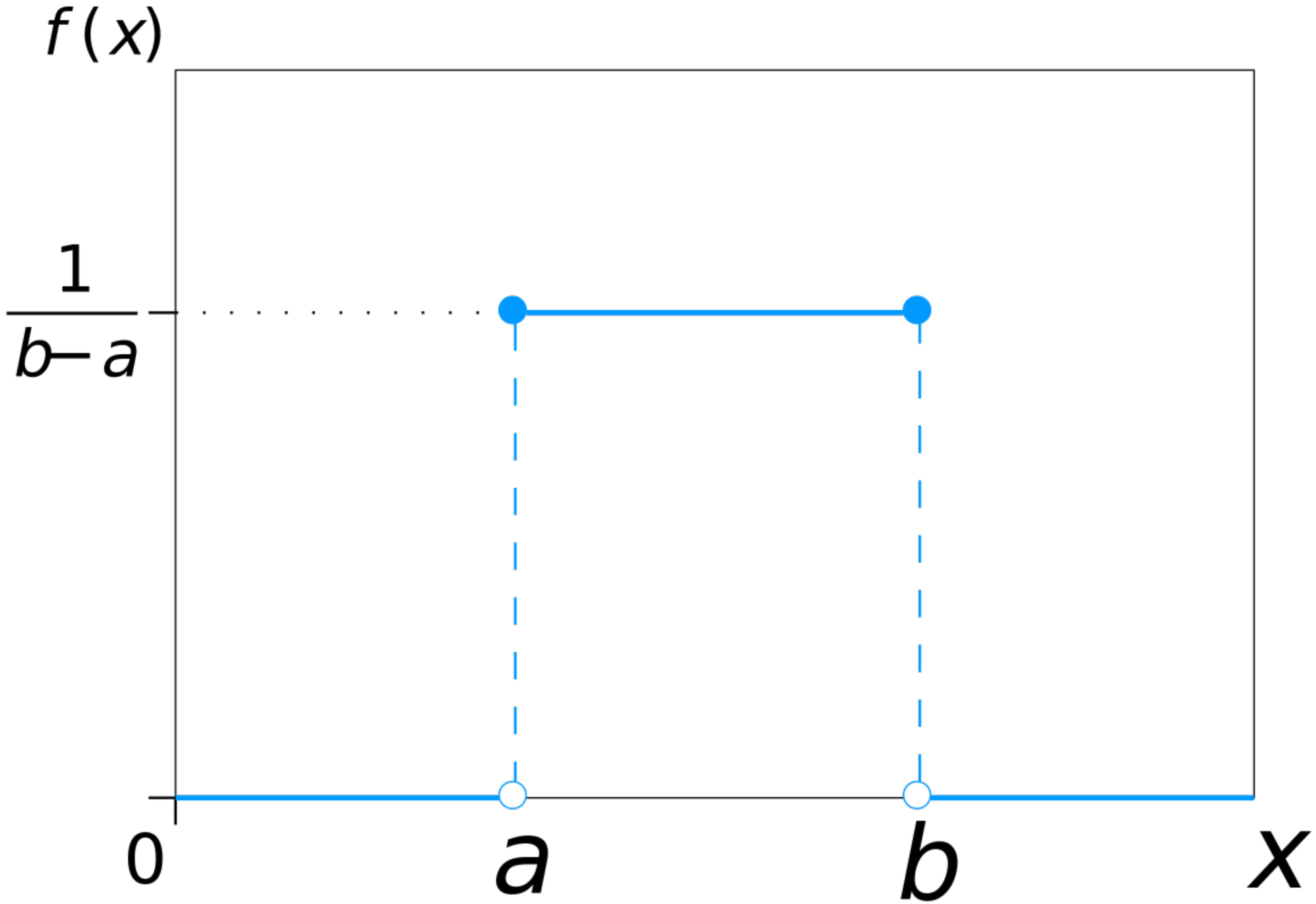
1. f_X is continuous over the given range
2. $0 \leq f_X(x) \leq 1$
3. $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Commonly Used Continuous Distributions

- Continuous Uniform Distribution
- Exponential Distribution
- Normal Distribution
- Chi-squared Distribution
- t Distribution
- F Distribution

https://en.wikipedia.org/wiki/Continuous_uniform_distribution

Continuous Uniform Distribution



$a \leq k \leq b, \qquad n = b - a$

PDF $\quad P(X = k) = \frac{1}{n}$

E[X] $\quad \frac{a + b}{2}$

Var(X) $\quad \frac{n^2}{12}$

CDF $\quad P(X \leq k) = \frac{k - a}{n}$

- e.g., an idealized random number generator

Exponential Distribution

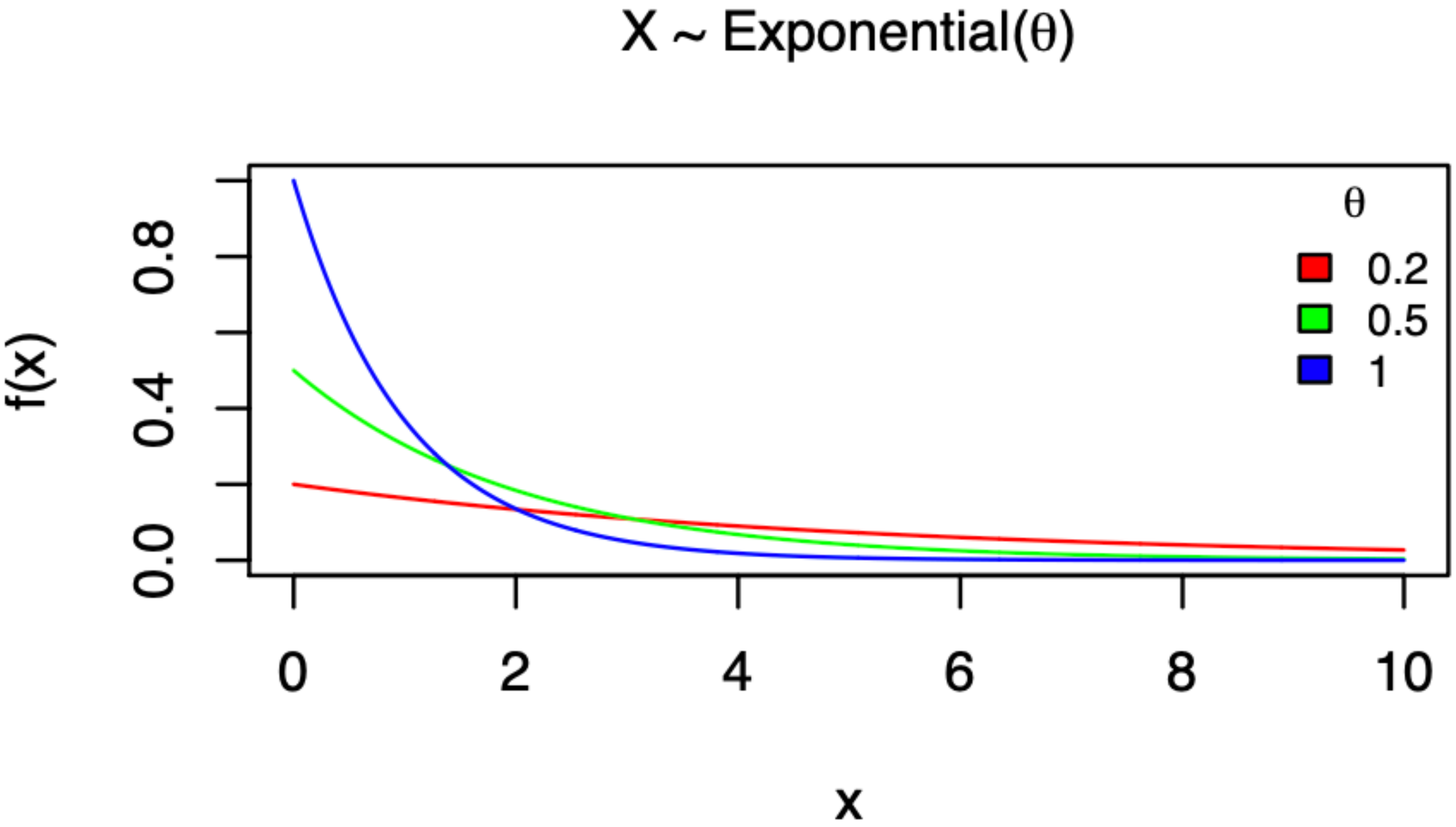
- An exponential random variable takes values ≥ 0
- Often used to model **the time elapsed between events**
- It is widely used in life-time analysis

$$P(X = x) = \theta e^{-x\theta} \quad x \in \mathbb{R}^{\geq 0} \quad and \quad \theta \in \mathbb{R}^{+}$$

$$F_x(X) = 1 - e^{-x\theta}$$

$$E[X] = \frac{1}{\theta}, Var(X) = \frac{1}{\theta^2}$$

Exponential Distribution



Raj S. Important distributions in probability & statistics [Internet]. Medium. 2020 [cited 2021 Oct 11]. Available from: https://medium.com/analytics-vidhya/important-distributions-in-probability-statistics-a868283fa127

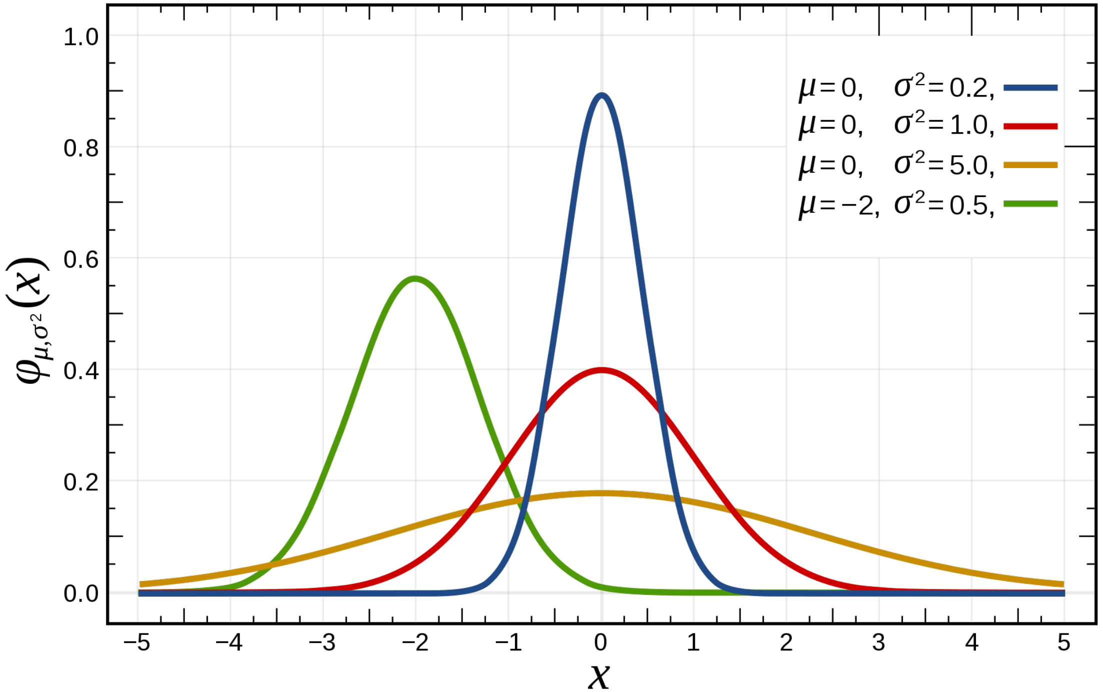
Gaussian/Normal Distribution

Gaussian/Normal distribution is a continuous probability distribution function where the random variable lies **symmetrically** around a mean (μ) and variance (σ²)

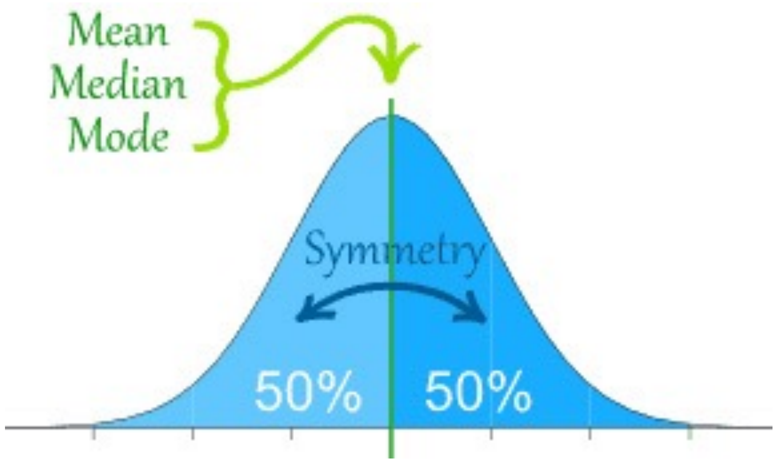
- E[X] = μ
- Var(X) = σ^2
- CDF = Φ(x) = not available in closed form

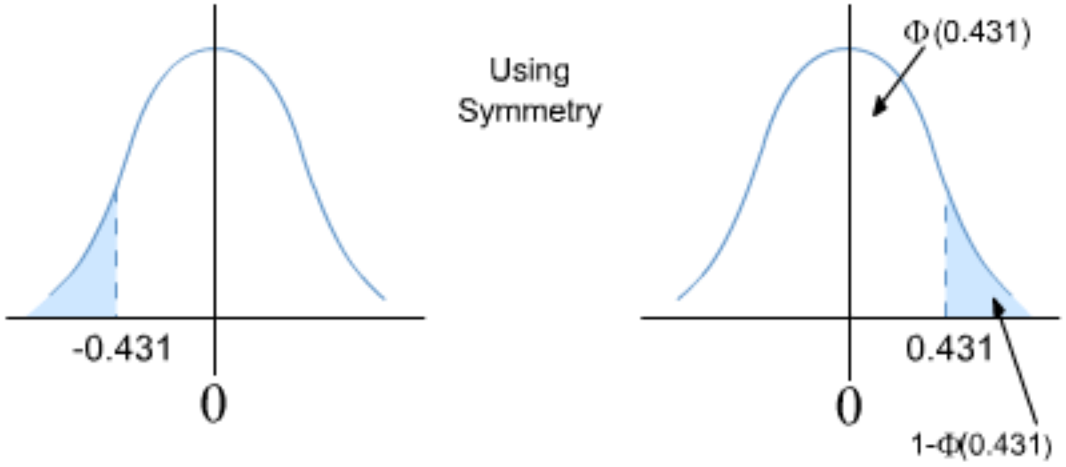
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\boldsymbol{X \sim N(\mu, \sigma^2)}$$

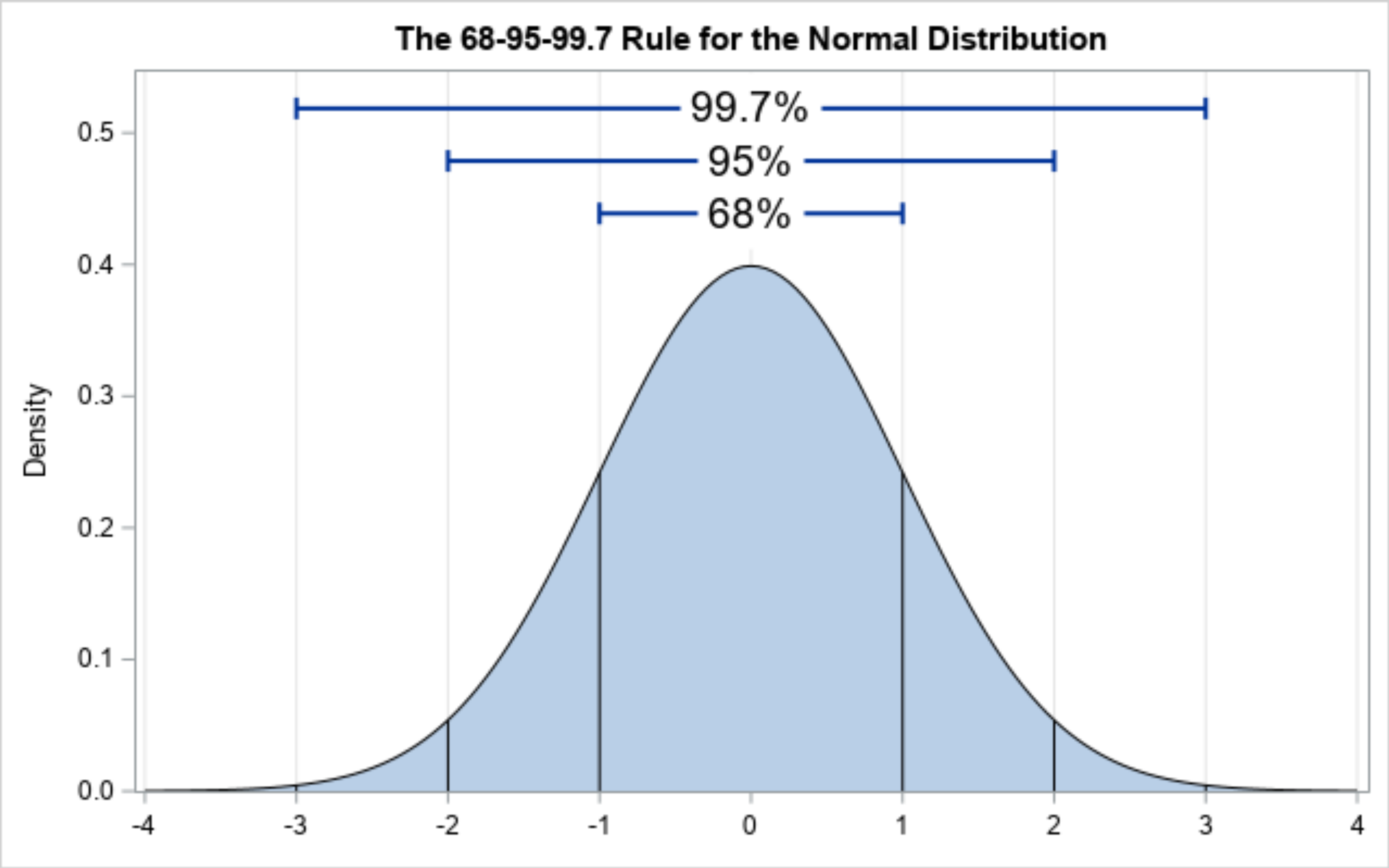


Normal distribution [Internet]. [cited 2021 Oct 11]. Available from: https://www.mathsisfun.com/data/standard-normal-distribution.html

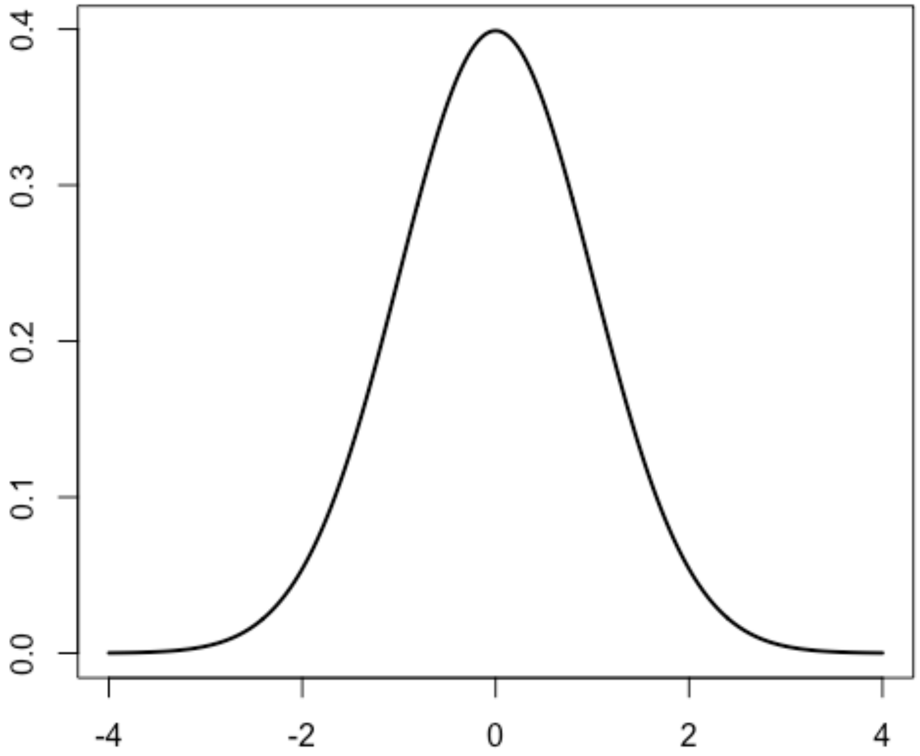




Mudugandla SY. 10 normality tests in python(Step-by-step guide 2020) [Internet]. Medium. 2020 [cited 2021 Oct 11]. Available from: https://towardsdatascience.com/normality-tests-in-python-31e04aa4f411



Standard Normal Distribution



$$\mu = 0$$
$$\sigma^2 = 1$$
$$Z \sim N(0,1)$$

Standardization

$$\text{If } X \sim N(\mu, \sigma^2), \quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Normal distribution [Internet]. [cited 2021 Oct 11]. Available from: https://www.mathsisfun.com/data/standard-normal-distribution.html

Chi-squared Distribution

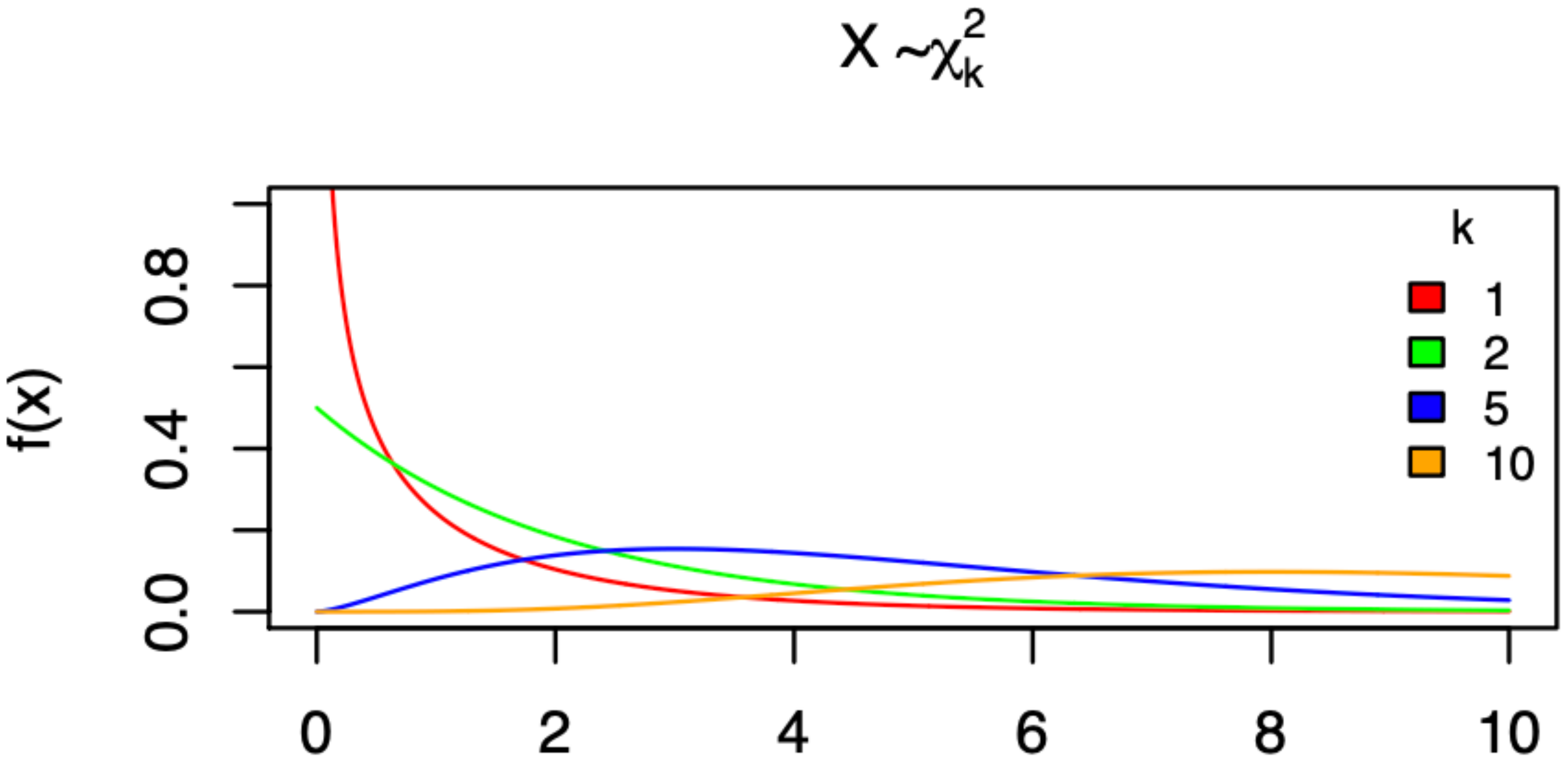
- If Y_i are *k* i.i.d. standard normal RVs
- $X = \sum_{i=1}^{k} Y_i^2$ is chi-squared distributed with degree-of-freedom *k*

$$X \sim \chi^2(k) \text{ or } X \sim \chi_k^2$$

$$P(X = x) = \frac{e^{-\frac{x}{2}} x^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, \quad x \in \mathbb{R}^{\geq 0}$$

$$E[X] = k, Var(X) = 2k$$

Chi-squared Distribution



(Student’s) t Distribution

- Used when estimating the mean of a normally distributed population in situations where **the sample size is small** and **population standard deviation is unknown**
- The t-distribution is **symmetric and bell-shaped**, like the normal distribution, but has **heavier tails**, meaning that it is more prone to producing values that fall far from its mean

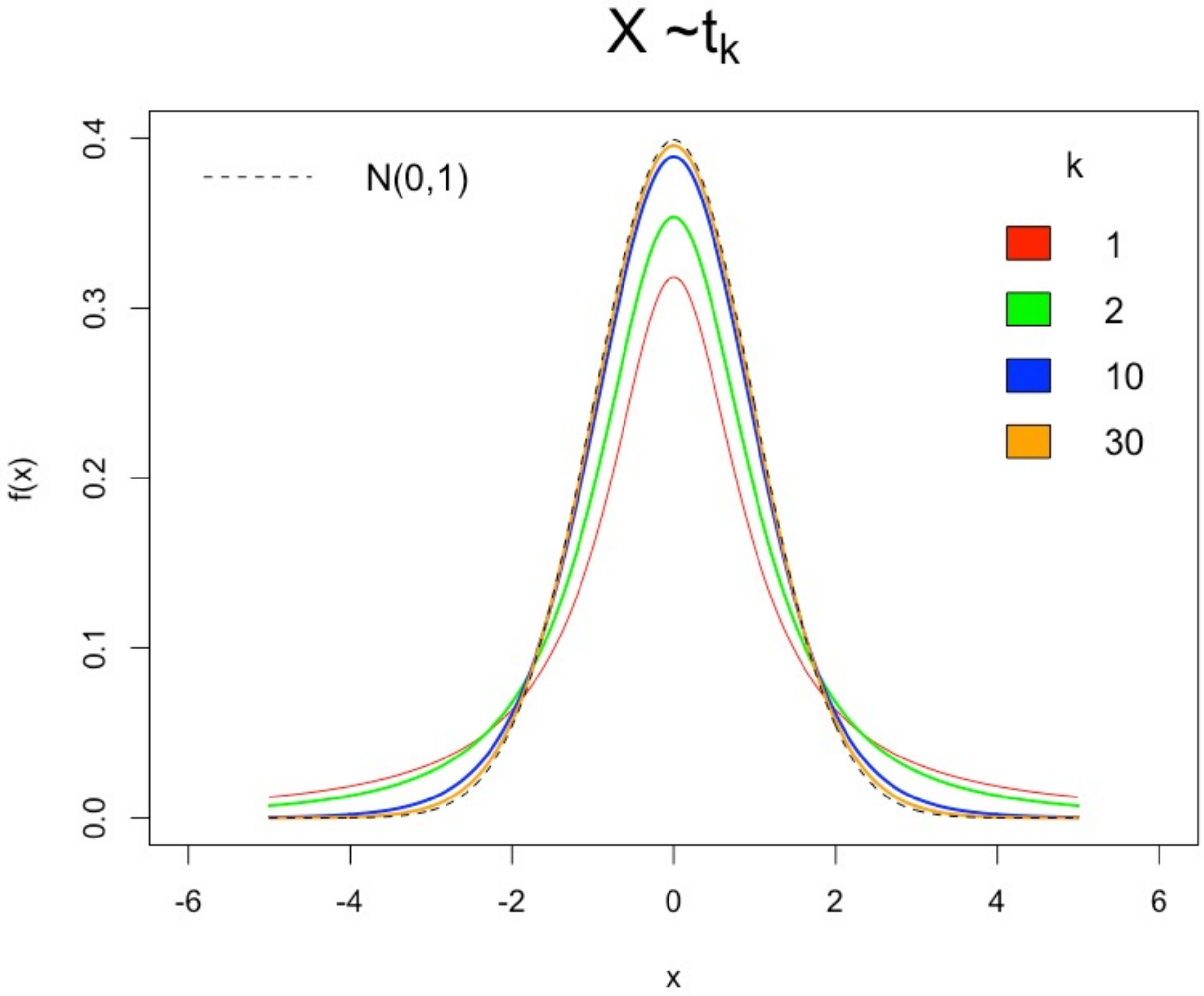
t Distribution

For a RV $X \sim N(0,1)$ and another RV $Y \sim \chi^2_k$, $Z = \frac{X}{\sqrt{Y/k}} \sim t_k$.

$$P(X = x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi\Gamma(\frac{k}{2})}}(1 + \frac{x^2}{k})^{-\frac{k+1}{2}}$$

$E[X] = 0$, $Var(X) = \frac{k}{k-2}$

(Student's) t Distribution



- As k gets larger (as a rule of thumb, $k \geq 30$), t distribution approximates standard normal distribution

F Distribution

- F distribution is a continuous probability distribution that **arises frequently as the null distribution of a test statistic**, most notably in the analysis of variance (ANOVA)

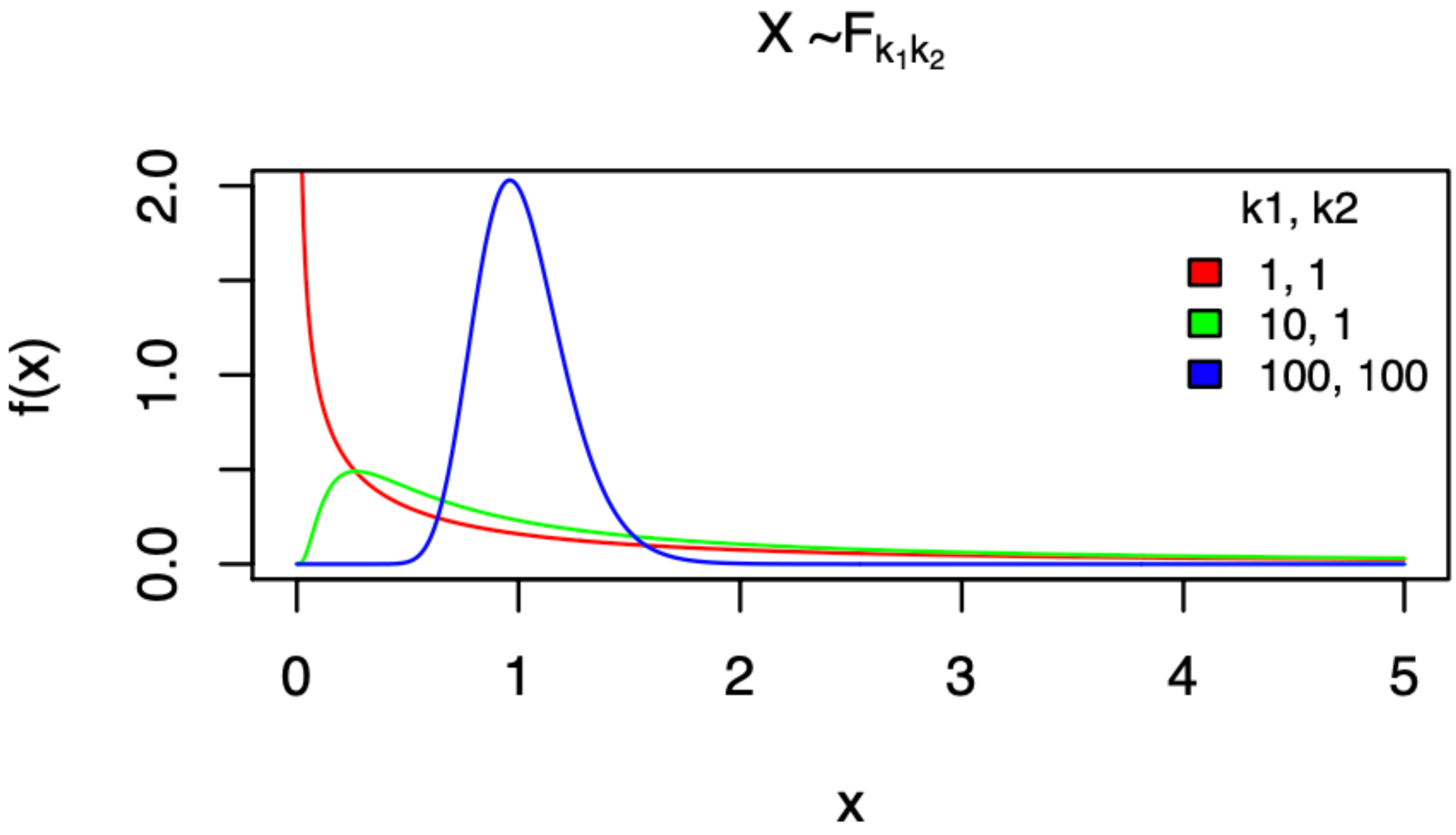
- An F random variable takes values ≥ 0
- If X and Y and are two independent chi-squared random variables with degree-of-freedom parameters k_1 and k_2, then
- $Z = \frac{X/k_1}{Y/k_2}$ is said to have F distribution with parameters k_1 and k_2

F Distribution

$$P(X = x) = \frac{\sqrt{\frac{(k_1 x)^{k_1} k_2{}^{k_2}}{(k_1 x + k_2)^{k_1 + k_2}}}}{x B(\frac{k_1}{2}, \frac{k_2}{2})}, \quad x \in \mathbb{R}^{\geq 0}$$

$$E[X] = \frac{k_2}{k_2 - 2}, Var(X) = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$$

F Distribution



Brief Summary

- Commonly used continuous distributions include:
  - Continuous Uniform Distribution
  - Exponential Distribution
  - Normal Distribution
  - Chi-squared Distribution
  - t Distribution
  - F Distribution