

Biostatistics

Week V

Ege Ülgen, M.D.

4 November 2021



ACIBADEM
MEHMET ALİ AYDINLAR
ÜNİVERSİTESİ

Probability Density Function (PDF)

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

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

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

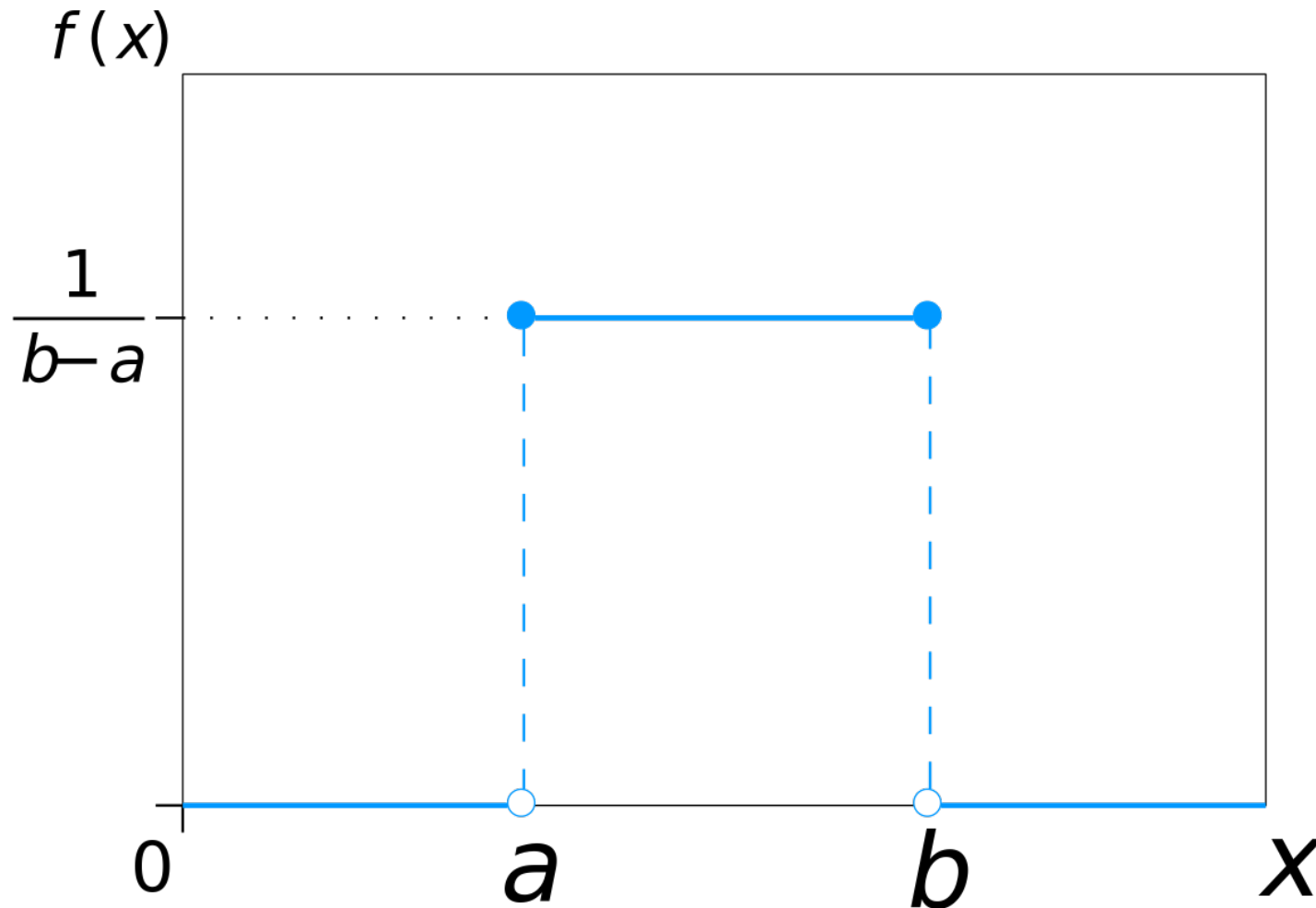
Properties of a Proper PDF (f_X)

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

Continuous Uniform Distribution



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

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

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

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

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

- 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

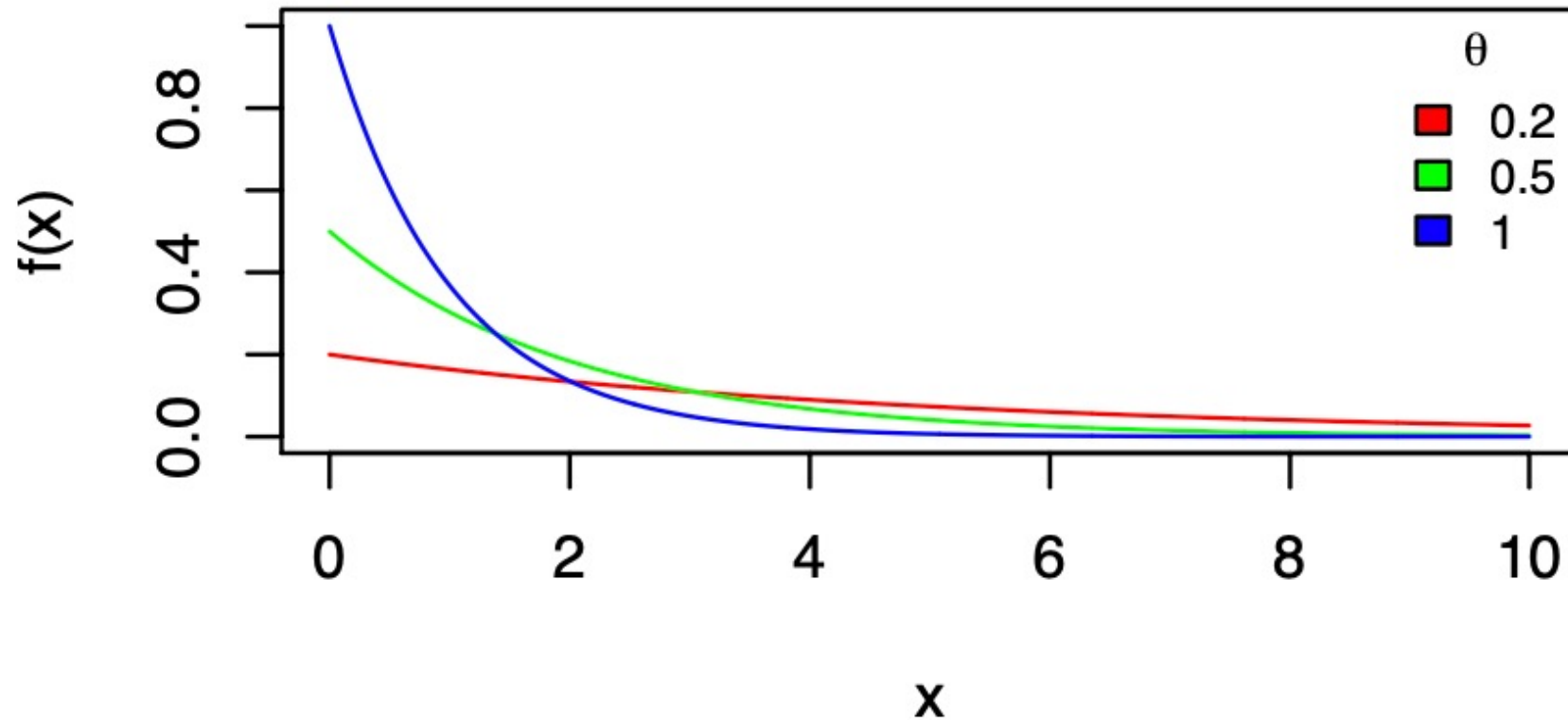
$$P(X = x) = \theta e^{-x\theta} \quad x \in \mathbb{R}^{\geq 0} \quad \text{and} \quad \theta \in \mathbb{R}^+$$

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

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

Exponential Distribution

$X \sim \text{Exponential}(\theta)$



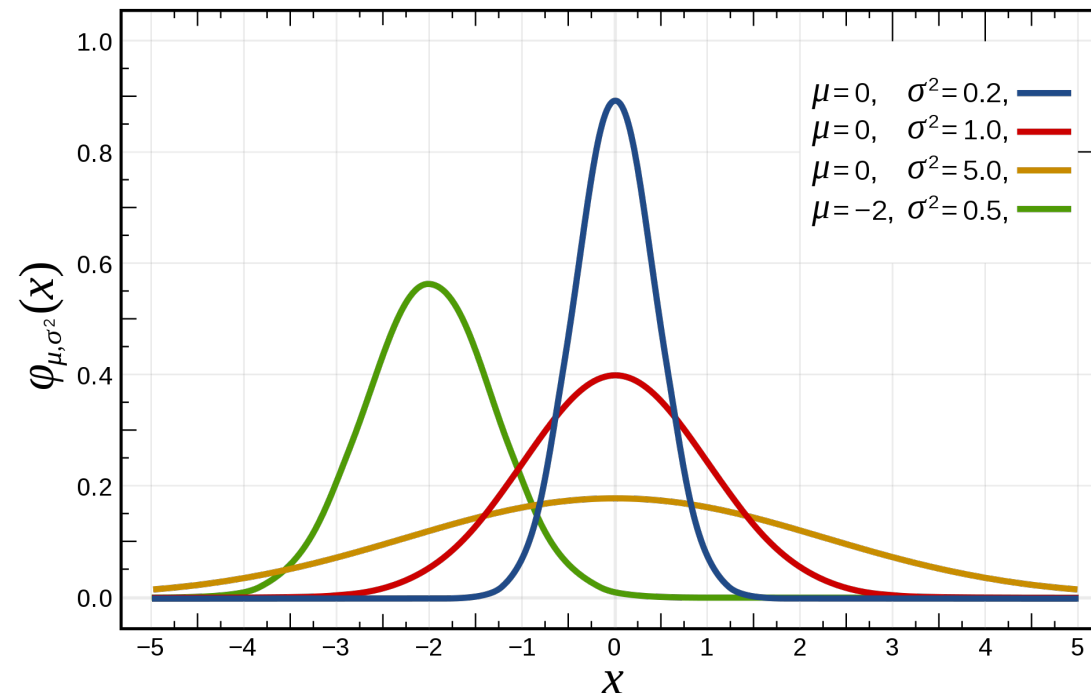
Gaussian/Normal Distribution

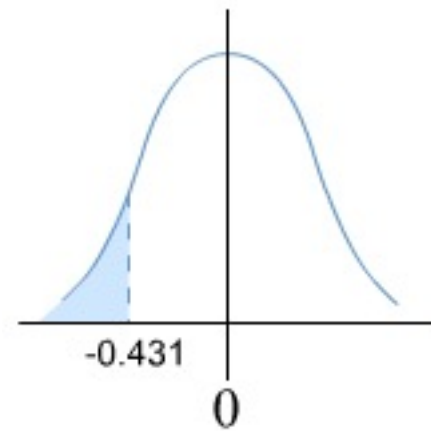
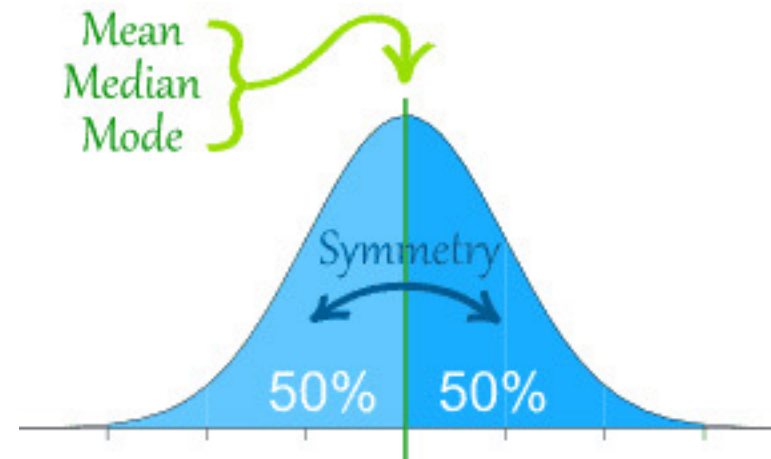
Gaussian/Normal distribution is a continuous probability distribution function where the random variable lies **symmetrically** around a mean (μ) and variance (σ^2)

- $E[X] = \mu$
- $\text{Var}(X) = \sigma^2$
- CDF = $\Phi(x)$ = not available in closed form

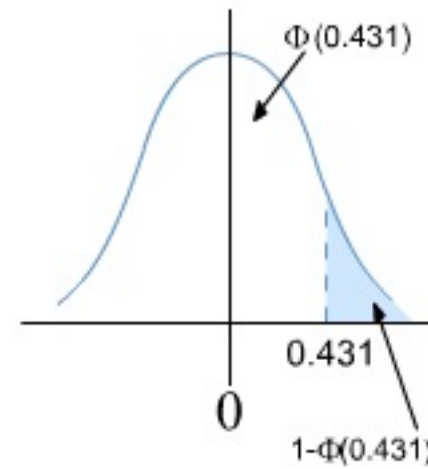
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

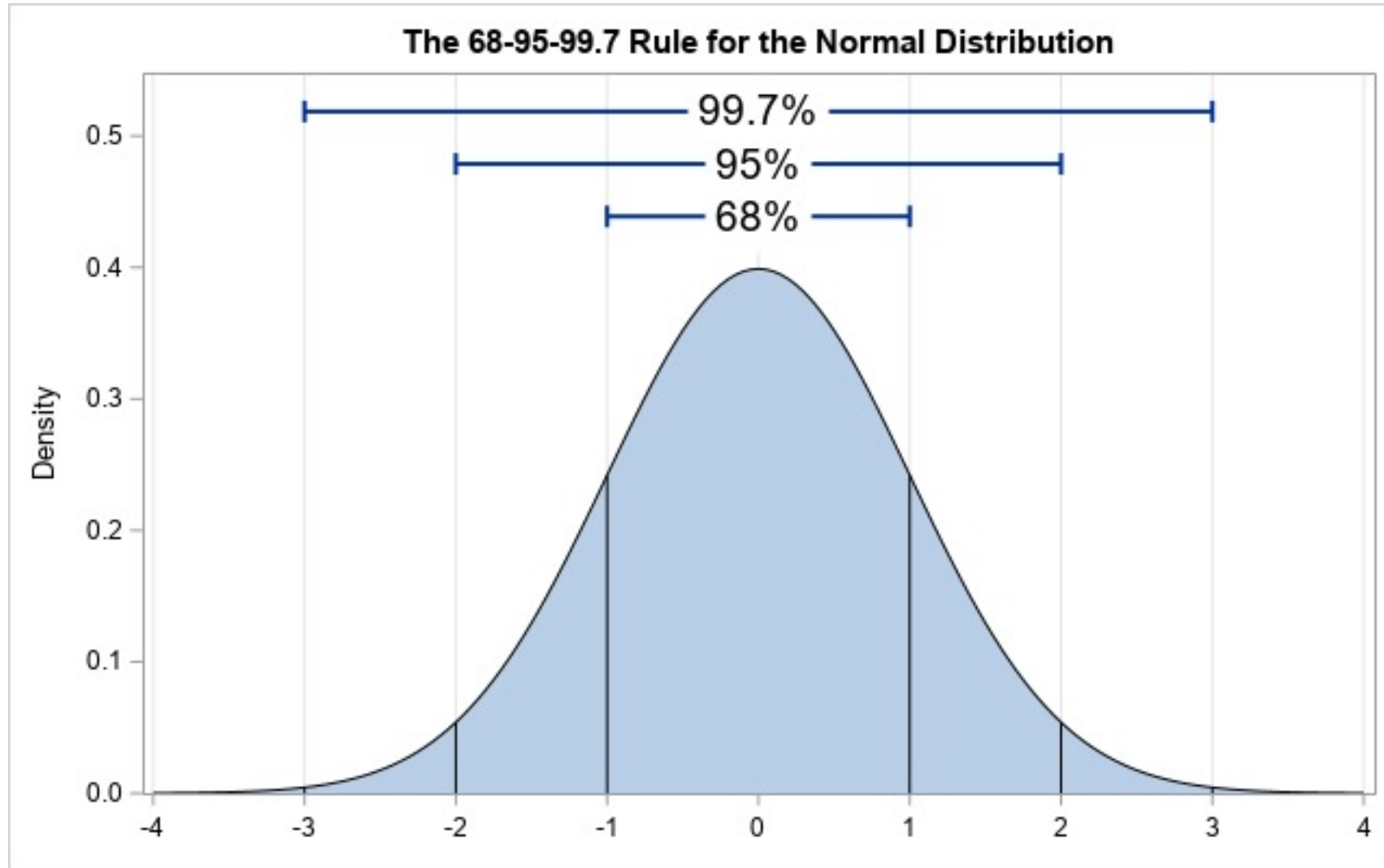
$$X \sim N(\mu, \sigma^2)$$



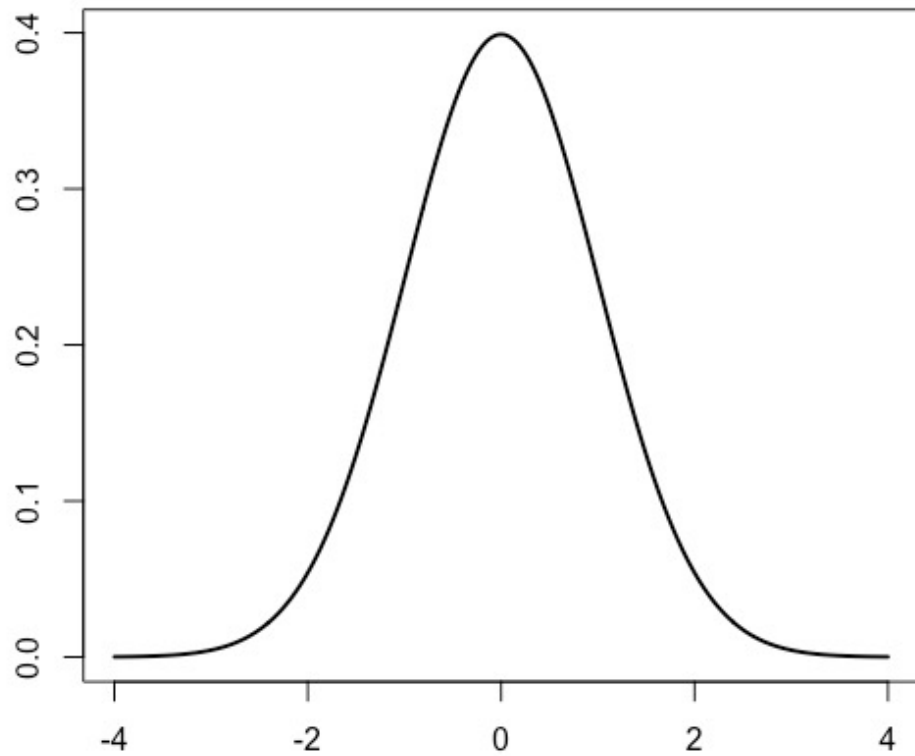


Using
Symmetry





Standard Normal Distribution



$$\begin{aligned}\mu &= 0 \\ \sigma^2 &= 1 \\ Z &\sim N(0,1)\end{aligned}$$

Standardization

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



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

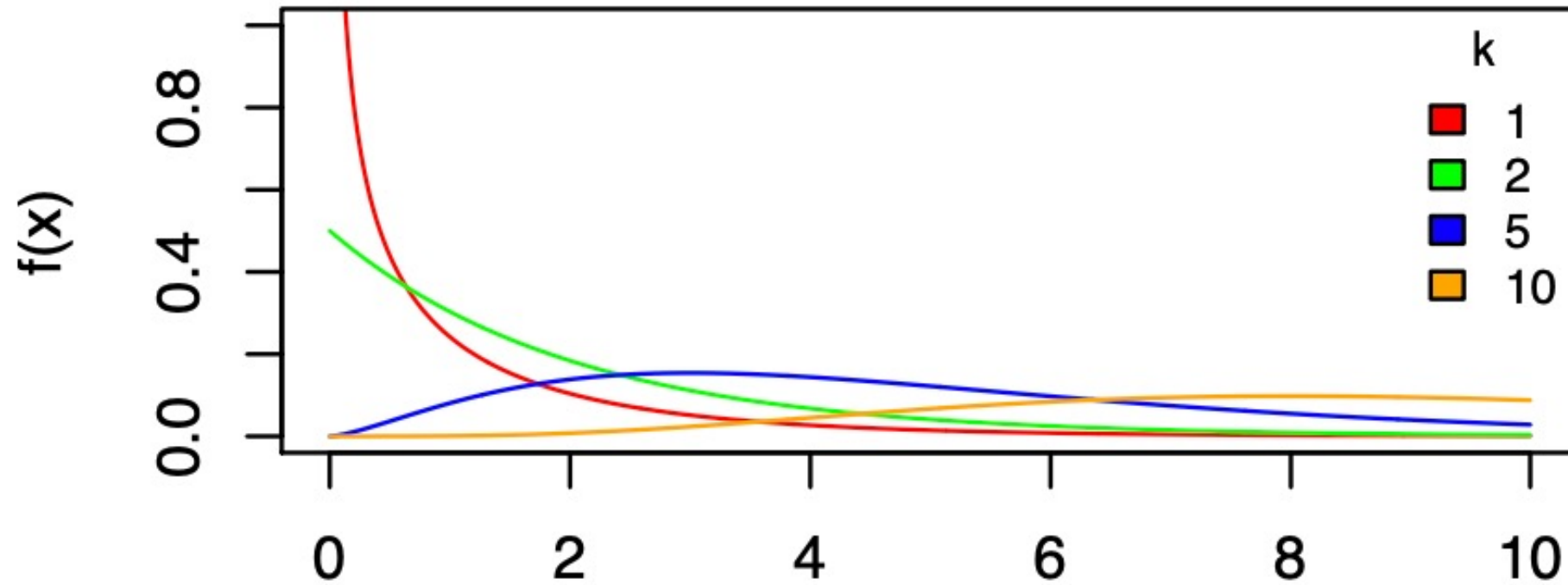
$$X \sim \chi^2(k) \text{ or } X \sim \chi_k^2$$

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

$$E[X] = k, \text{Var}(X) = 2k$$

Chi-squared Distribution

$$X \sim \chi_k^2$$



(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

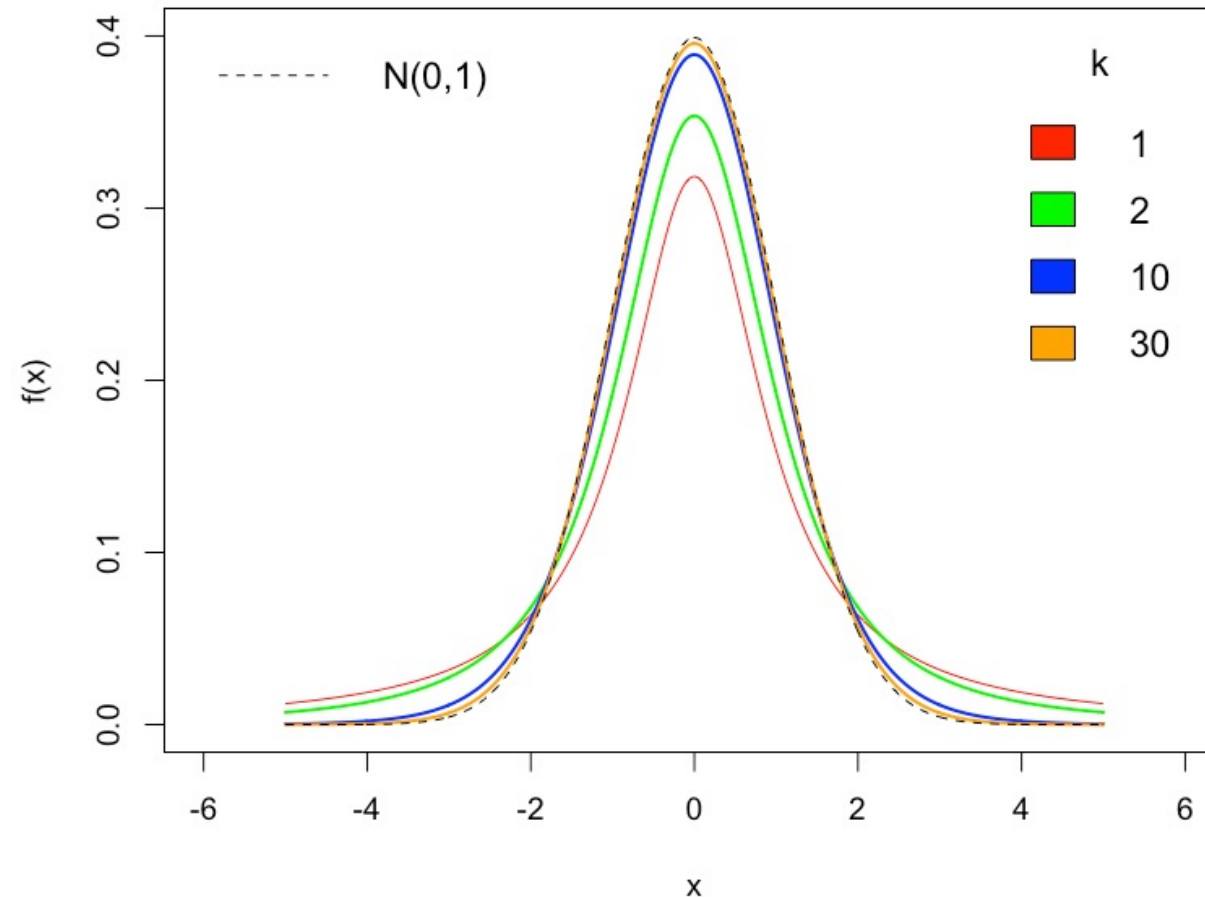
For a RV $X \sim N(0,1)$ and another RV $Y \sim \chi_k^2$, $Z = \frac{X}{\sqrt{Y/k}} \sim t_k$.

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

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

(Student's) t Distribution

$$X \sim t_k$$



- 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 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

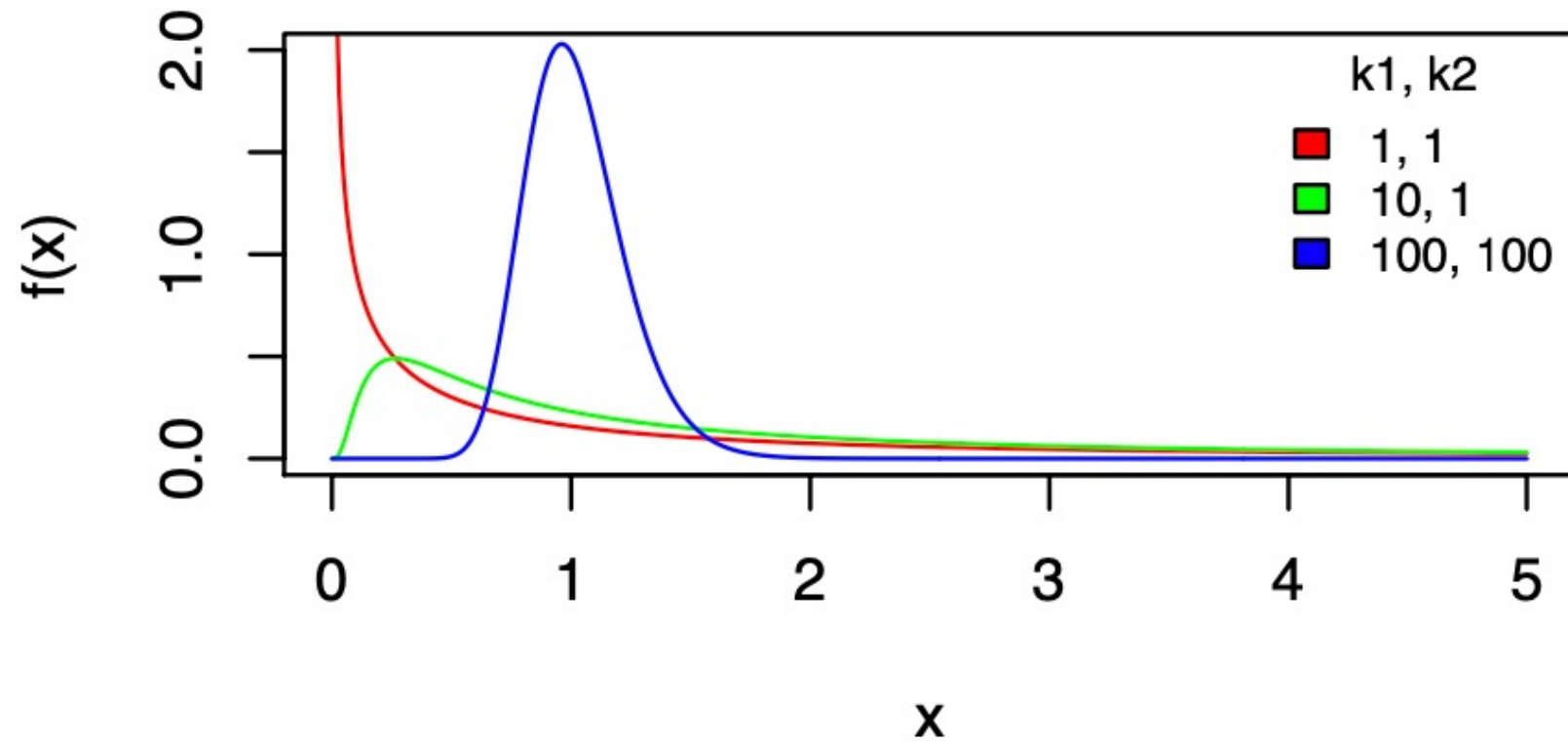
F Distribution

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

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

F Distribution

$$X \sim F_{k_1 k_2}$$



Brief Summary

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