Lecture 7.

Examples of probability distributions-continuous variables

prof. dr hab.inż. Katarzyna Zakrzewska

Katedra Elektroniki, AGH

e-mail: zak@agh.edu.pl

http://home.agh.edu.pl/~zak

Outline:

- Definitions of mean and variance for continuous random variables
- Uniform distribution
- Central limit theorem
- Gaussian distribution

MEAN AND VARIANCE OF A CONTINUOUS RANDOM VARIABLE

Definition

Suppose X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \qquad (4\text{-}4)$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

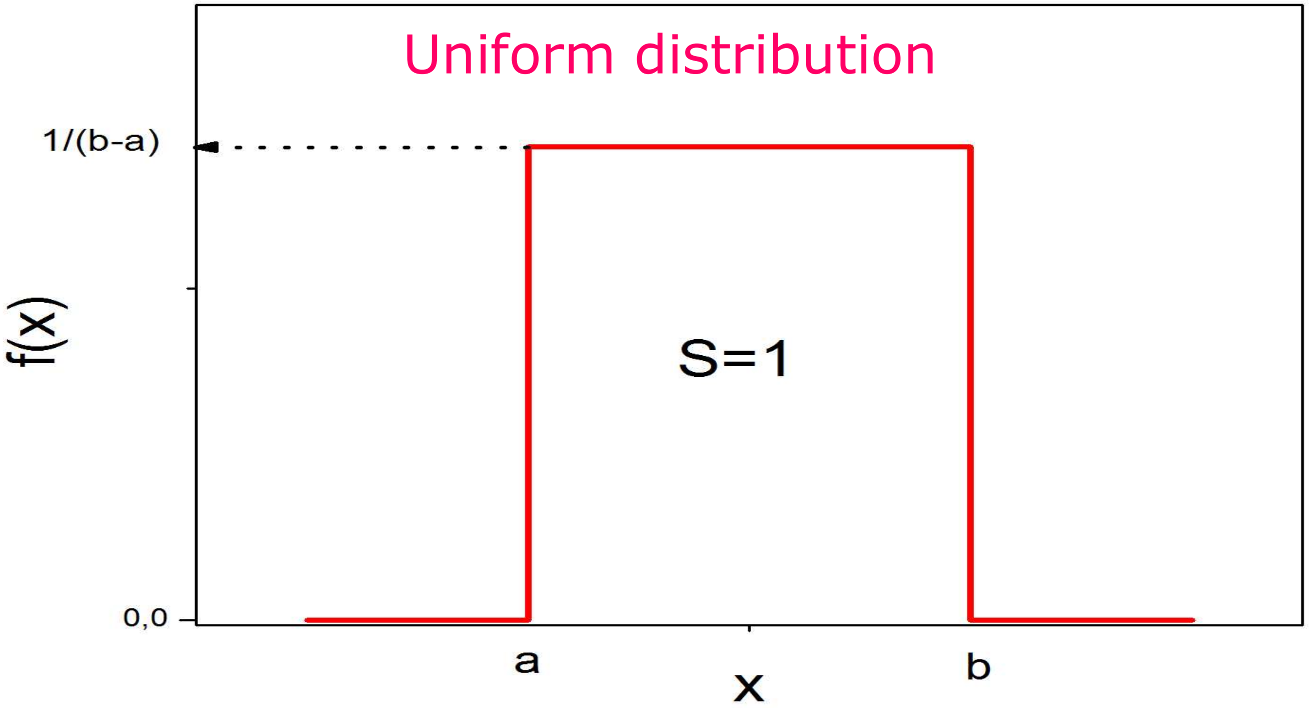
UNIFORM DISTRIBUTION

Definition

A continuous random variable X with probability density function

$$f(x) = 1/(b - a), \qquad a \le x \le b \tag{4-6}$$

is a **continuous uniform random variable.**



UNIFORM DISTRIBUTION

The mean of the continuous uniform random variable X is

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{0.5x^2}{b-a}\Big|_a^b = \frac{(a+b)}{2}$$

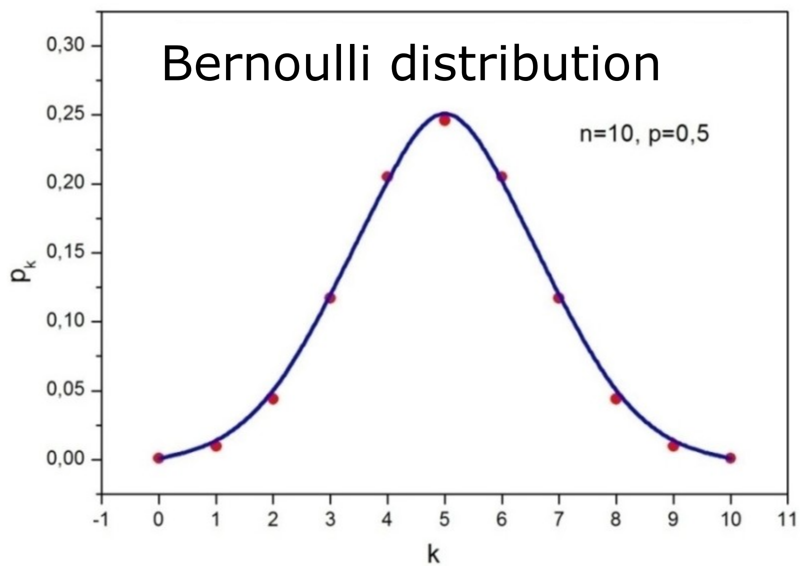
The variance of X is

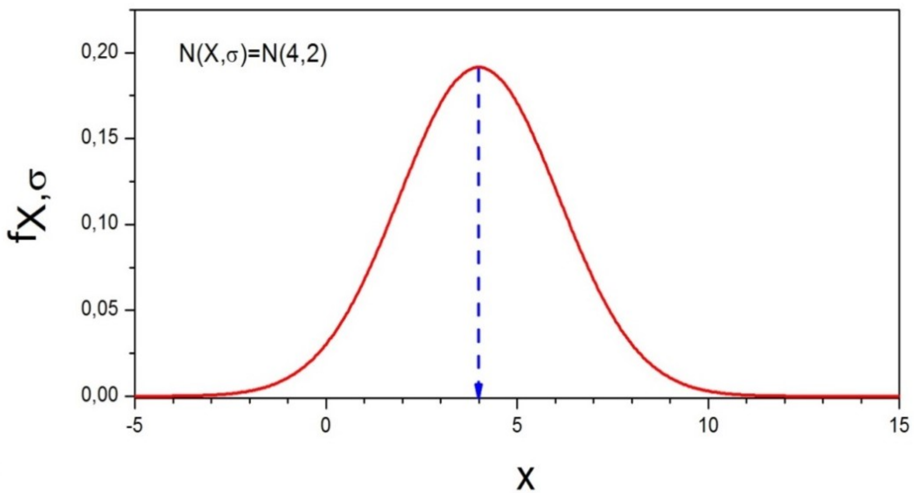
$$V(X) = \int_a^b \frac{\left(x - \left(\frac{a+b}{2}\right)\right)^2}{b-a}\,dx = \frac{\left(x - \frac{a+b}{2}\right)^3}{3(b-a)}\Bigg|_a^b = \frac{(b-a)^2}{12}$$

$$f(x) = \frac{1}{b-a} \qquad \mu = EX = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

Central limit theorem

Limiting case (normal distribution)





The most widely used model for the distribution of random variable is a **normal distribution**.

Central limit theorem formulated in 1733 by De Moivre

Whenever a random experiment is replicated, the random variable that equals the average (or total) result over the replicas tends to have a normal distribution as the number of replicas becomes large.

Central limit theorem

Example: toss of a die

Probability mass function for a single toss (population):

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

Uniform distribution; the mean value 3.5

$$\mu = E(x) = \sum x \cdot p(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3.5$$

Variance:

$$\sigma^2 = \frac{1}{N} \sum (x-\mu)^2 = \frac{1}{6} \cdot \left[(1-3.5)^2 + (2-3.5)^2 + \ldots + \cdot (6-3.5)^2 \right]$$

$$\cong 2.92$$

Central limit theorem

For a double toss (n=2) we are interested in the distribution of the arithmetic average.

Result	Av.	Result	Av.	Result	Av.	Result	Av.	Result	Av.	Result	Av.
1|1	1	2|1	1.5	3|1	2	4|1	2.5	5|1	3	6|1	3.5
1|2	1.5	2|2	2	3|2	2.5	4|2	3	5|2	3.5	6|2	4
1|3	2	2|3	2.5	3|3	3	4|3	3.5	5|3	4	6|3	4.5
1|4	2.5	2|4	3	3|4	3.5	4|4	4	5|4	4.5	6|4	5
1|5	3	2|5	3.5	3|5	4	4|5	4.5	5|5	5	6|5	5.5
1|6	3.5	2|6	4	3|6	4.5	4|6	5	5|6	5.5	6|6	6

Central limit theorem

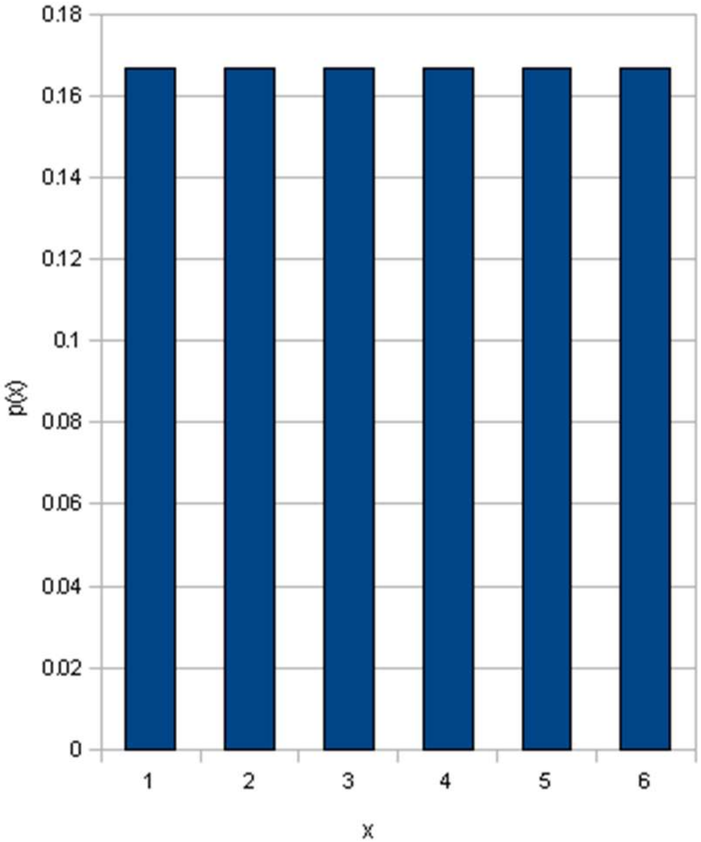
Probability mass function of the arithmetic average

$\bar{x}$	$p(\bar{x})$	$\bar{x}$	$p(\bar{x})$
1	1/36	4	5/36
1.5	2/36	4.5	4/36
2	3/36	5	3/36
2.5	4/36	5.5	2/36
3	5/36	6	1/36
3.5	6/36		

Arithmetic average of two samples is a **statistics.** At the same time it represents an **estimator of the expected value.** What kind of estimator is it?

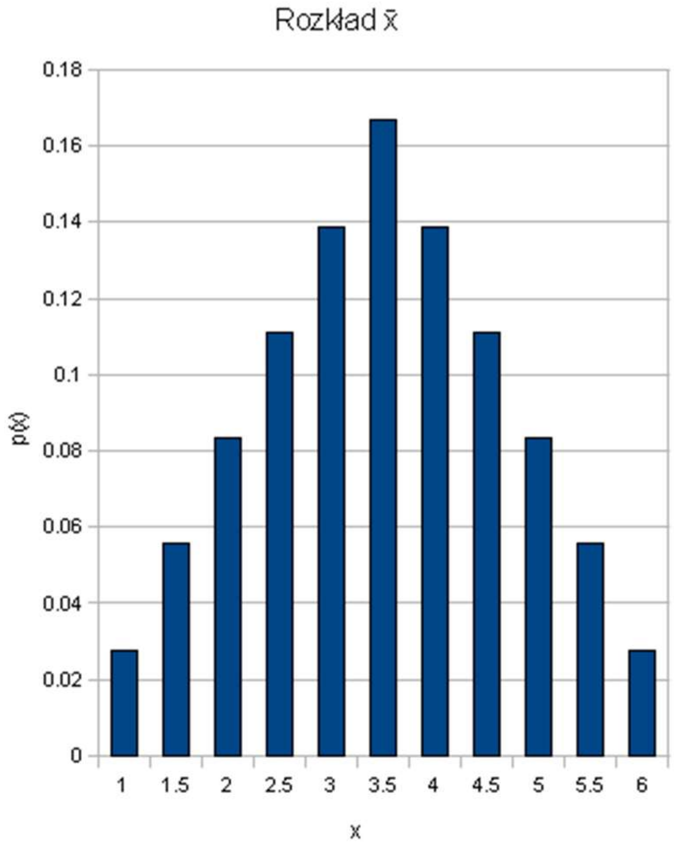
Central limit theorem

Distribution of random variable x



$$\mu_x = 3.5 \qquad \sigma_x^2 = 2.92$$

Distribution of the arithmetic average of two samples

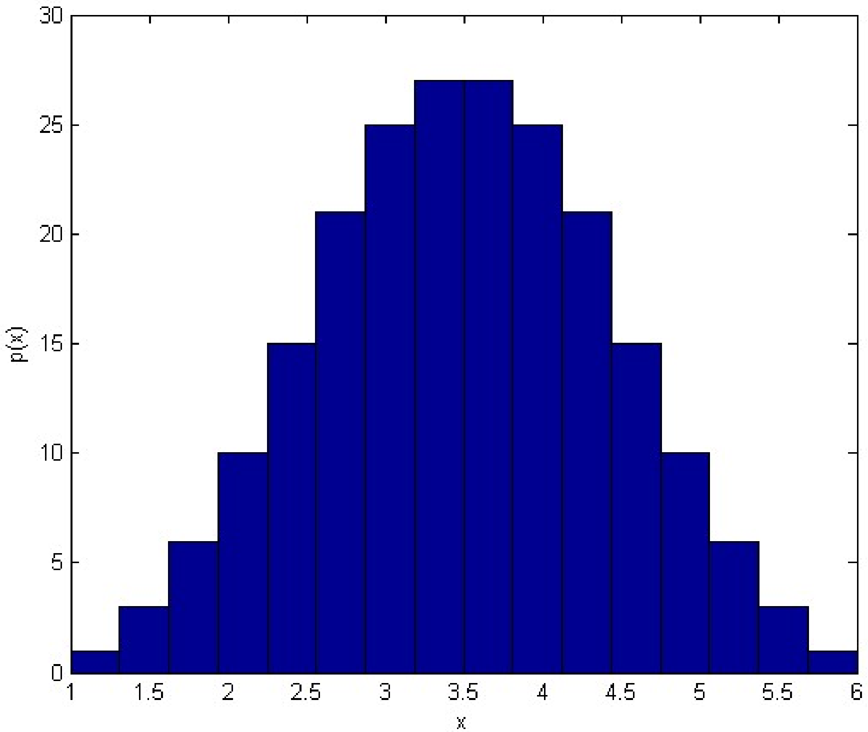


$$\mu_x = 3.5 \quad \sigma_{\bar{x}}^2 = 1.46 = \frac{\sigma_x^2}{2}$$

Central limit theorem

For n=3 distribution of the arithmetic average tends to normal distribution (Gaussian).

Result	Av.
1 | 1 | 1	1
1 | 1 | 2	1.33
1 | 1 | 3	1.66
1 | 1 | 4	2
1 | 1 | 5	2.33
1 | 1 | 6	2.66
2 | 1 | 1	1.33
2 | 1 | 2	1.66
2 | 1 | 3	2
2 | 1 | 4	2.33
2 | 1 | 5	2.66
2 | 1 | 6	3
3 | 1 | 1	1.66
3 | 1 | 2	2
…	…
…	…



$$\mu_x = 3.5 \quad \sigma_{\bar{x}}^2 = 0.97 = \frac{\sigma_x^2}{3}$$

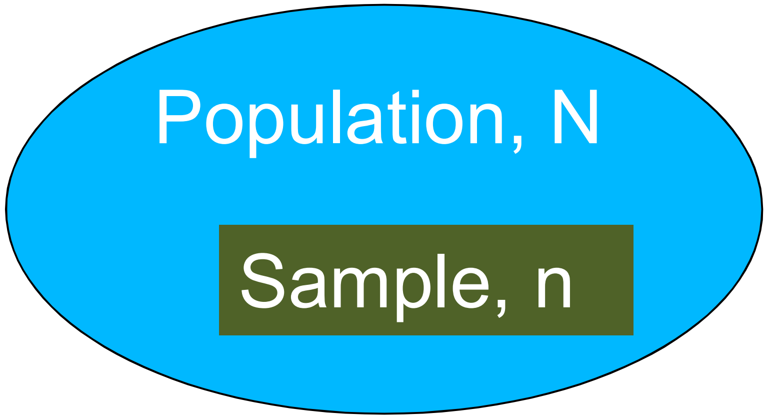
Population and sample

Conclusions from the example:

1. Arithmetic average $\overline{x}$ of sample has approximately normal distribution

2. Variance is:

$$\sigma_{\overline{x}}^2 = \frac{\sigma_x^2}{n}$$



Normal distribution (Gaussian)

A random variable X with probability density function f(x):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu^2}{2\sigma^2} \right] \text{, } where\, -\infty < x < +\infty$$

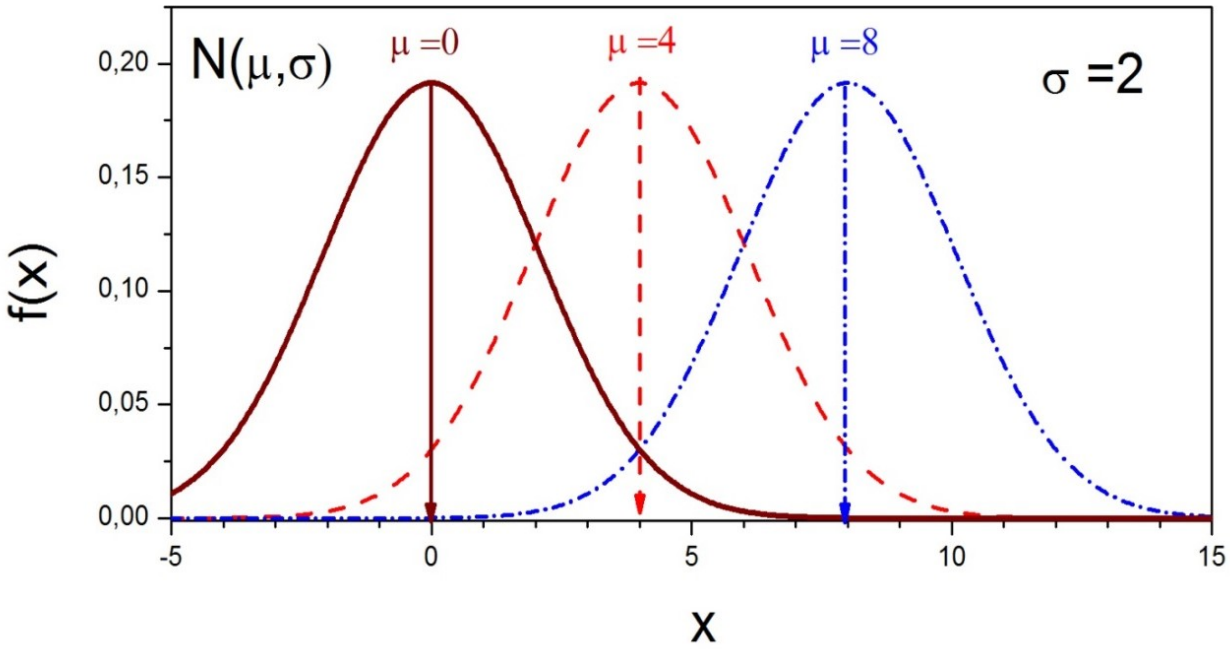
is a **normal random variable** with two parameters:

$$-\infty < \mu < +\infty, \quad \sigma > 1$$

We can show that $E(X)=\mu$ and $V(X)=\sigma^2$

Notation $N(\mu,\sigma)$ is used to denote this distribution

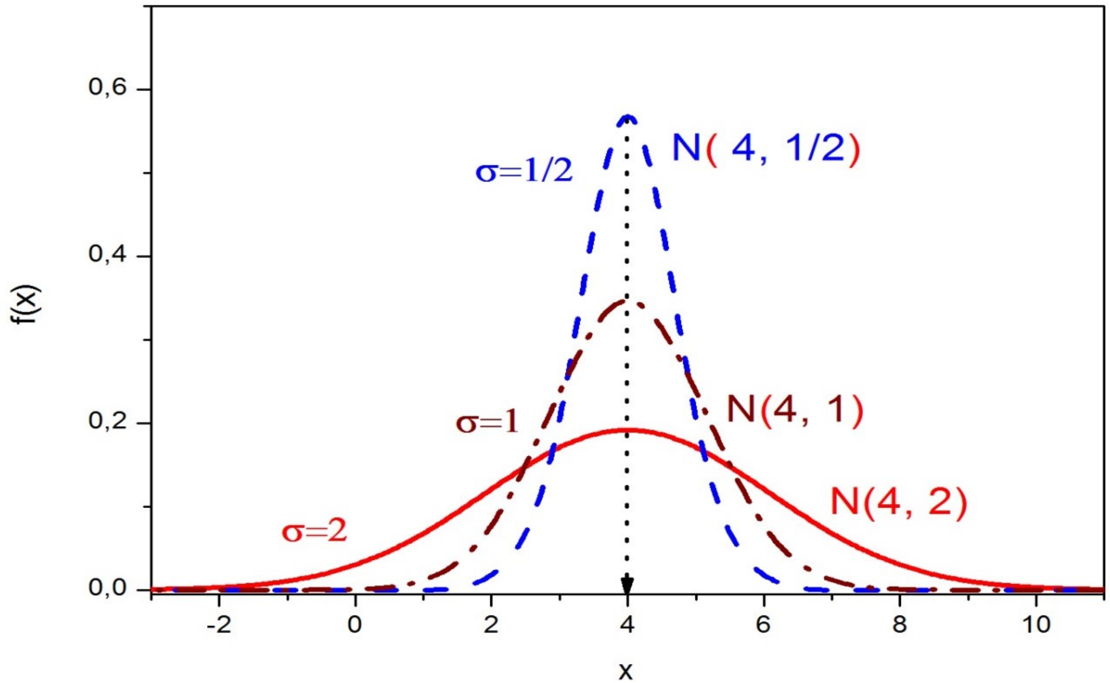
Normal distribution (Gaussian)



Expected value, maximum of density probability (mode) and median overlap ($x=\mu$). Symmetric curve (Gaussian curve is bell shaped).

Variance is a measure of the width of distribution. At $x=+\sigma$ and $x=-\sigma$ there are the inflection points of $N(0, \sigma)$.

Normal distribution (Gaussian)



Is used in experimental physics and describes distribution of random errors. Standard deviation σ is a measure of random uncertainty. Measurements with larger σ correspond to bigger scatter of data around the average value and thus have **less precision.**

Standard normal distribution

A normal random variable Z with probability density N(z):

$$N(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2} \right] \text{, } where \text{ - } \infty < z < +\infty$$

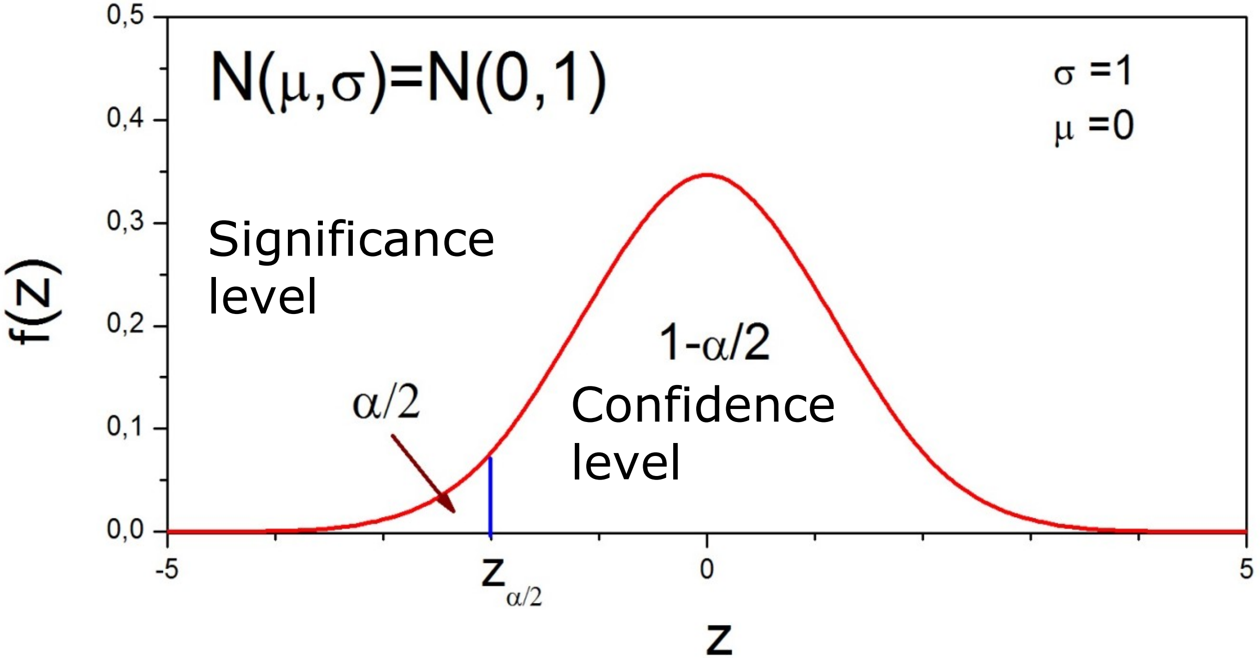
is called a **standard normal random variable**

N(0,1) $\quad E(Z) = 0, \quad V(Z) = 1$

Definition of standard normal variable

$$Z = \frac{X - \mu}{\sigma}$$

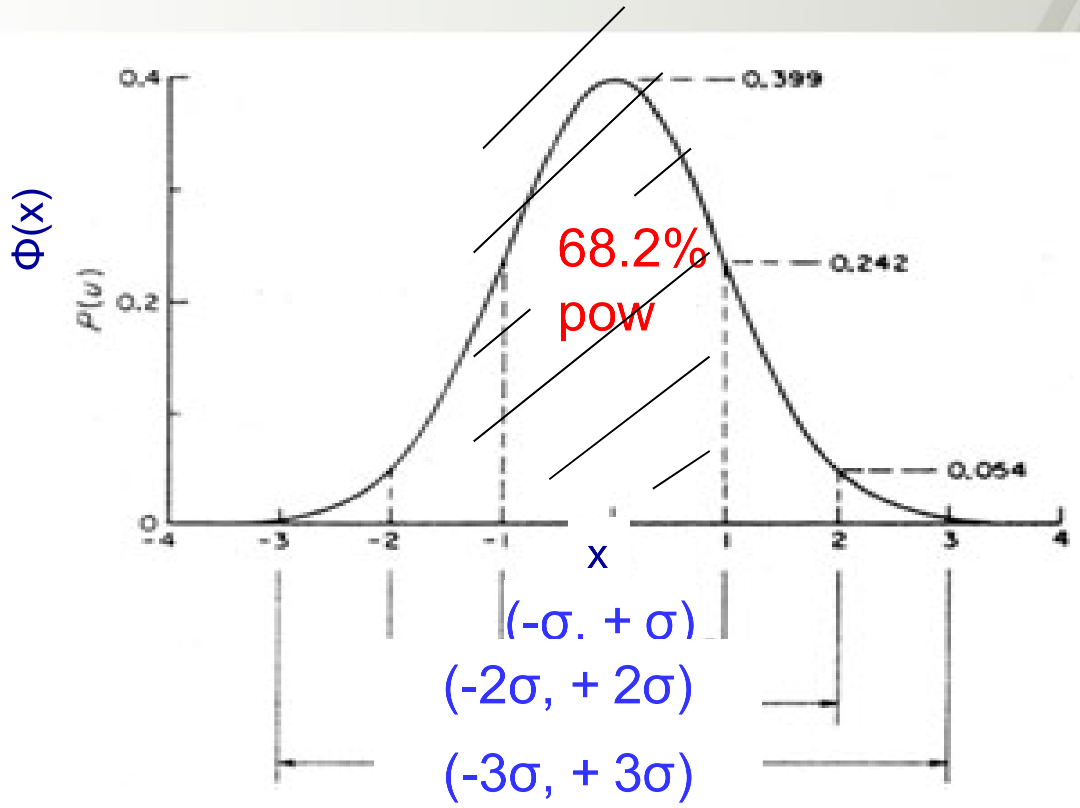
Standard normal distribution



Advantages of standardization:

- Tables of values of probability density and CDF can be constructed for N(0,1). A new variable of the N(μ,σ) distribution can be created by a simple transformation $X = \sigma * Z + \mu$
- By standardization we shift all original random variables to the region close to zero and we rescale the x-axis. The unit changes to standard deviation. Therefore, we can compare different distributions.

Calculations of probability (Gaussian distribution)



P(μ-σ <X< μ+σ) = 0,6827 (about 2/3 of results)
P(μ-2σ <X< μ+2σ) = 0,9545
P(μ-2σ <X< μ+2σ) = 0,9973 (almost all)