Lecture 5.

Random variable and distribution of probability

prof. dr hab.inż. Katarzyna Zakrzewska
Katedra Elektroniki, AGH
e-mail: zak@agh.edu.pl
http://home.agh.edu.pl/~zak

Outline:

- Concept of random variable
- Quantitative description of random variables
- Examples of probability distributions

The concept of random variable

Random variable is a function X, that attributes a real value x to a certain results of a random experiment.

$$\Omega = \{e_1, e_2, \ldots\}$$

$$X : \Omega \rightarrow R$$

$$X(e_i) = x_i \in R$$

Examples:

1) Coin toss: event 'head' takes a value of 1; event 'tails' - 0.
2) Products: event 'failure' - 0, well-performing – 1
3) Dice: '1' – 1, '2' – 2 etc.…
4) Interval [a, b]– a choice of a point of a coordinate 'x' is attributed a value, e.g. $\sin^2(3x+17)$ etc. ….

The concept of random variable

Random variable

Discrete

When the values of random variable X are isolated points on an number line

- **Toss of a coin**
- **Transmission errors**
- **Faulty elements on a production line**
- **A number of connections coming in 5 minutes**

Continuous

When the values of random variable cover all points of an interval

- **Electrical current, I**
- **Temperature, T**
- **Pressure, p**

Quantitative description of random variables

- Probability distributions and **probability mass functions** (for discrete random variables)
- **Probability density functions** (for continuous variables)
- Cumulative distribution function (distribution function for discrete and continuous variables)
- Characteristic quantities (expected value, variance, quantiles, etc.)

Distribution of random variable

Distribution of random variable (probability distribution for discrete variables) is a set of pairs (x_i, p_i) where x_i is a value of random variable X and p_i is a probability, that a random variable X will take a value x_i

Example 4.1

Probability mass function for a single toss of coin.
Event corresponding to heads is attributed $x_1=1$; tails means $x_2=0$.

$$x_1 = 1 \quad p(X = 1) = p(x_1) = \frac{1}{2}$$

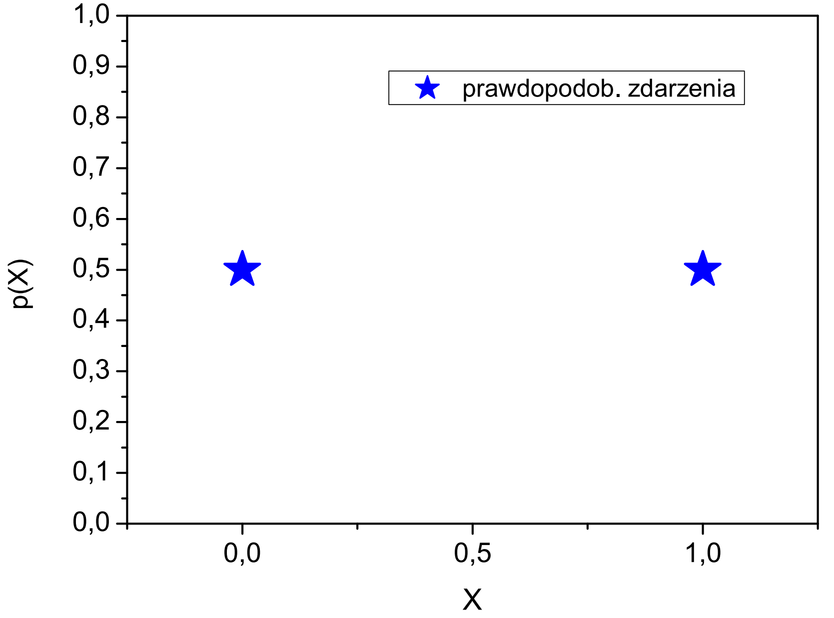
$$x_2 = 0 \quad p(X = 0) = p(x_2) = \frac{1}{2}$$

Distribution of random variable

Example 4.1 cont.

Probability mass function for a single toss of coin is given by a set of the following pairs:

$$\{(1, \frac{1}{2}), (0, \frac{1}{2})\}$$



Random variable when discrete entails probability distribution also discrete.

Probability density function

Probability function is introduced for continuous variables; it is related to probability in the following way:

$$f(x)dx \equiv P(x \le X < x + dx)$$

Properties of probability density function:

$$1.\ \ f(x) \ge 0$$

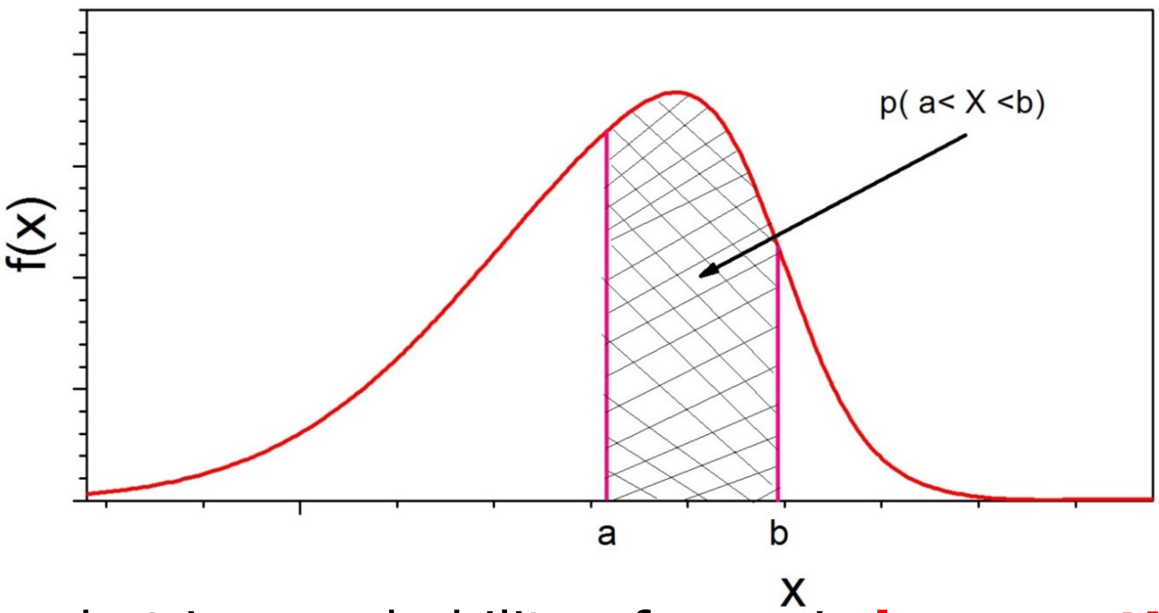
$$2.\ \ f(x)\ is\ normalized\ \ \int_{-\infty}^{+\infty} f(x)dx = 1$$

3. f(x) has a measure of 1/x

Probability density function

Directly from a definition of probability density function f(x) we get a formula of calculating the probability that the random variable will assume a value within an interval of [a,b]:

$$P(a < X < b) = \int_a^b f(x)dx$$

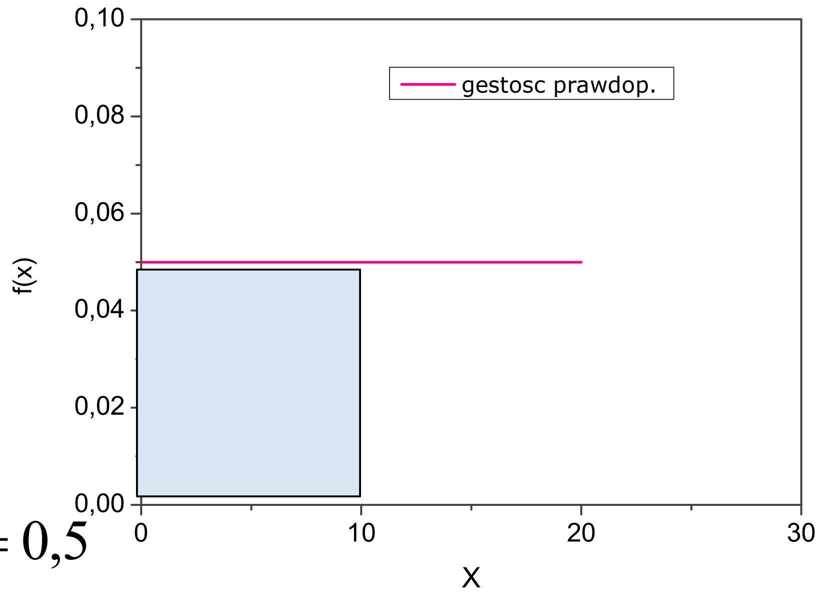


Question: what is a probability of x=a is **incorrect!!!**

Probability density function

Example 4.2

Let the continuous random variable X denote the current measured in a thin copper wire in mA. Assume that the range of X is [0, 20 mA], and assume that the probability density function of X is f(x)=0,05 for 0≤x≤20. What is the probability that a current measured is less than 10 mA.



$$P(0 \le X < 10) = \int_0^{10} f(x)dx = \int_0^{10} 0{,}05\, dx = 0{,}5$$

Quantitative description of random variables

- Cumulative distribution function (CDF) F(x) is a probability of an event that the random variable X will assume a value smaller than or equal to x (at most x)

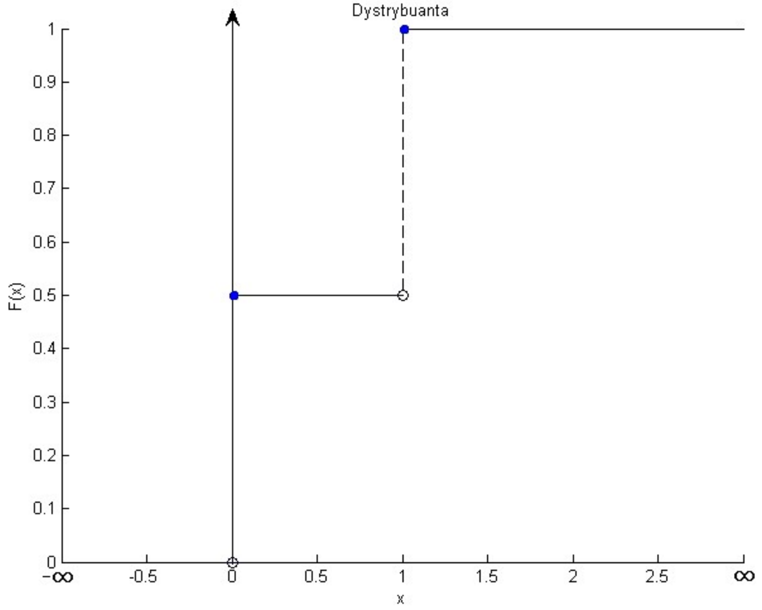
$$F(x) = P(X \leq x)$$

Example 4.1 cont.

CDF of coin toss:

$$F(x=0) = P(X \leq 0) = \frac{1}{2}$$

$$F(x=1) = P(X \leq 1) = 1$$



Properties of CDF

1. $0 \le F(x) \le 1$

2. $F(-\infty) = 0$

3. $F(+\infty) = 1$

4. $x \le y \;\Rightarrow\; F(x) \le F(y)$

non-decreasing function

5. F(x) has no unit

6. $f(x) = \dfrac{dF(x)}{dx}$

Relationship between cumulative distribution function and probability density (for continuous variable)

CDF of discrete variable

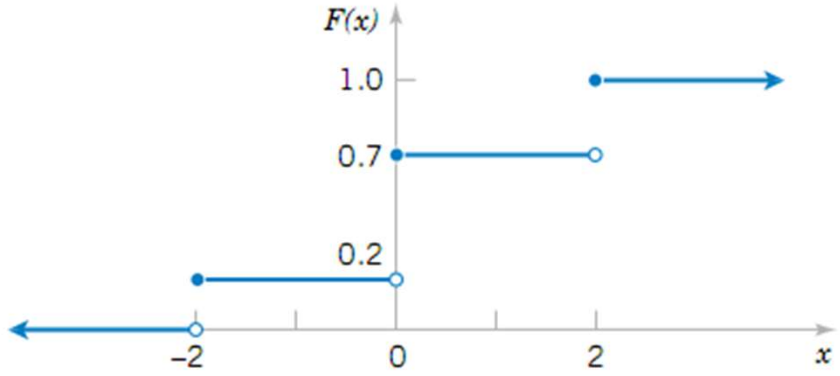
$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

$f(x_i)$ – probability mass function

Example 4.3

Determine probability mass function of X from the following cumulative distribution function F(x)

$$F(x) = \begin{cases} 0 & \text{for } x < -2 \\ 0.2 & \text{for } -2 \le x < 0 \\ 0.7 & \text{for } 0 \le x < 2 \\ 1 & \text{for } 2 \le x \end{cases}$$



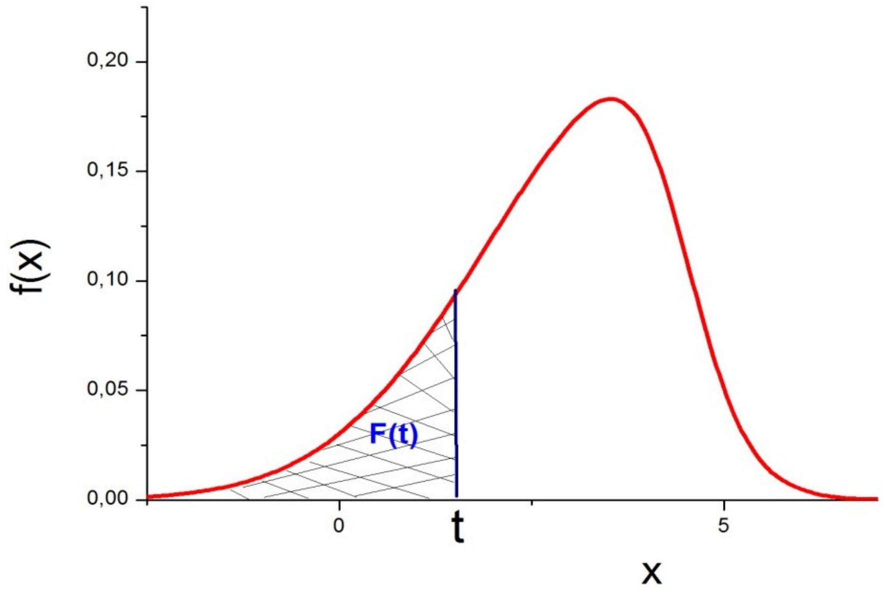
From the plot, the only points to receive f(x)≠0 are -2, 0, 2.

$$f(-2) = 0.2 - 0 = 0.2 \qquad f(0) = 0.7 - 0.2 = 0.5 \qquad f(2) = 1.0 - 0.7 = 0.3$$

CDF for continuous variable

$$F(t) = P(X \le t) = \int_{-\infty}^{t} f(x)\,dx$$

Cumulative distribution function F(t) of continuous variable is a non-decreasing continuous function and can be calculated as an area under density probability function f(x) over an interval from - ∞ to t.



Numerical descriptors

Parameters of

Position

- Quantile (e.g. median, quartile)
- Mode
- Expected value (average)

Dispersion

- Variance (standard deviation)
- Range

Numerical descriptors

Quantile x_q represents a value of random variable for which the cumulative distribution function takes a value of q.

$$F(x_q) = P(X \le x_q) = \int_{-\infty}^{x_q} f(u)\,du = q$$

Median i.e. $x_{0.5}$ is the most frequently used quantile.

In example 4.2 current I=10 mA is a median of distribution.

Example 4.4

For a discrete distribution : 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 median is 22 (middle value or arithmetic average of two middle values)

Numerical descriptors

Mode represents the most frequently occurring value of random variable (x at which probability distribution attains a maximum)

Unimodal distribution has one mode (multimodal distributions – more than one mode)

In example 4.4: x_k= 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 mode equals to 21 (which appears 3 times, i.e., the most frequently)

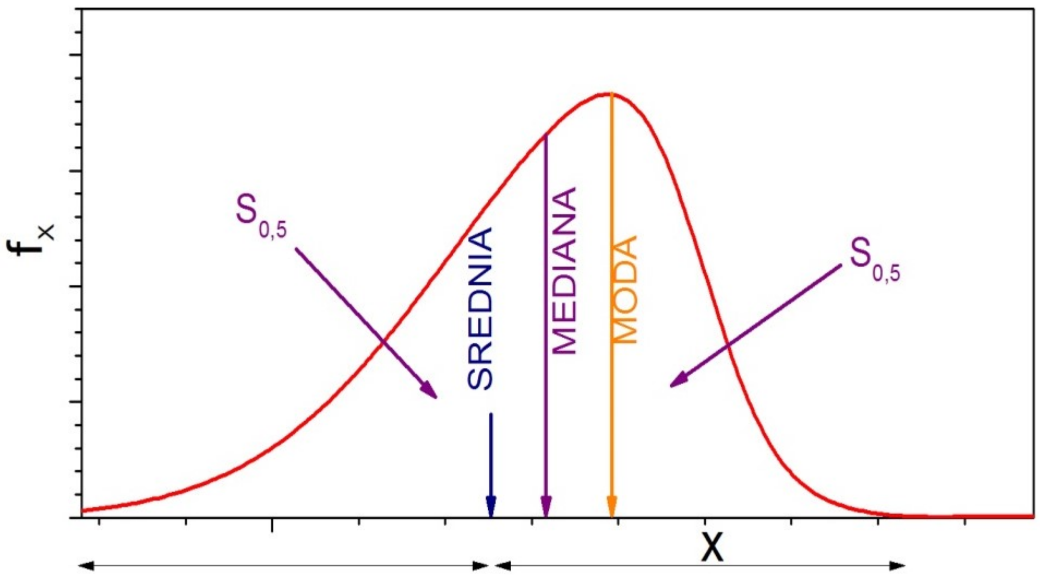
Average value

Arithmetic average:

x_i - belongs to a set of n – elements

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In example 4.4: x_i = 19, 21, 21, 21, 22, 22, 23, 25, 26, 27,
the arithmetic average is 22.7



Arithmetic average

Many elements having the same value, we divide the set into classes containing n_k identical elements

Example 4.5

x_k	n_k	f_k
10.2	1	0.0357
12.3	4	0.1429
12.4	2	0.0714
13.4	8	0.2857
16.4	4	0.1429
17.5	3	0.1071
19.3	1	0.0357
21.4	2	0.0714
22.4	2	0.0714
25.2	1	0.0357
Sum	28	

$$\bar{x} = \frac{\sum_{k=1}^{p} n_k x_k}{n} = \sum_{k=1}^{p} f_k x_k$$

where: $f_k = \dfrac{n_k}{n}, p - number \quad of \quad classes \ (p \le n)$

Normalization condition $\sum_k f_k = 1$

$$\bar{x} = x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n =$$

$$= 10.2 \cdot 0.04 + 12.3 \cdot 0.14 + \ldots + 25.2 \cdot 0.04$$

$$\bar{x} = 15.77$$

Moments of distribution functions

Moment of the order k with respect to x_0

$$m_k(x_0) \equiv \sum_i (x_i - x_0)^k \, p(x_i)$$ for discrete variables

$$m_k(x_0) \equiv \int (x - x_0)^k \, f(x)\, dx$$ for continuous variables

The most important are the moments calculated with respect to $x_0=0$ (m_k) and $X_0=m_1$ the first moment (m_1 is called the expected value) – these are central moments μ_k.

Expected value

Symbols: m_1, E(X), μ, $\overline{x}$, $\hat{x}$

$$E(X) = \sum_i x_i p_i$$ for discrete variables

$$E(X) \equiv \int x \, f(x) \, dx$$ for continuous variables

Properties of E(X)

E(X) is a linear operator, i.e.:

1. $$E(\sum_i C_i X_i) = \sum_i C_i E(X_i)$$

In a consequence:

$$E(C) = C$$

$$E(CX) = CE(X)$$

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

2. For independent variables $X_1, X_2, \dots X_n$

$$E(\prod_i X_i) = \prod_i E(X_i)$$

Variables are independent when:

$$f(X_1, X_2, \dots, X_n) = f_1(X_1) f_2(X_2) \cdot \dots \cdot f_n(X_n)$$

Properties of E(X)

3. For a function of X; Y= Y(X) the expected value E(Y) can be found on the basis of distribution of variable X without necessity of looking for distribution of f(y)

$$E(Y) = \sum_i y(x_i) p_i$$ for discrete variables

$$E(Y) \equiv \int y(x) \, f(x) \, dx$$ for continuous variables

Any moment $m_k(x_0)$ can be treated as an expected value of a function $Y(X)=(X-x_0)^k$

$$m_k(x_0) \equiv \int (x - x_0)^k f(x) \, dx = E((x - x_0)^k)$$

Variance

VARIANCE (dispersion) symbols: $\sigma^2(X)$, var(X), V(X), D(X).
Standard deviation σ(x)

$$\sigma^2(X) \equiv \sum_i p_i (x_i - E(X))^2$$ for discrete variables

$$\sigma^2(X) \equiv \int f(x)(x - E(X)^2\, dx$$ for continuous variables

Variance (or the standard deviation) is a measure of scatter of random variables around the expected value.

$$\sigma^2(X) = E(X^2) - E^2(X)$$

Properties of $\sigma^2(X)$

Variance can be calculated using expected values only:

1. $$\sigma^2(X) = E(X^2) - E^2(X)$$

In a consequence we get:

$$\sigma^2(C) = 0$$

$$\sigma^2(CX) = C^2 \sigma^2(X)$$

$$\sigma^2(C_1X + C_2) = C_1^2 \sigma^2(X)$$

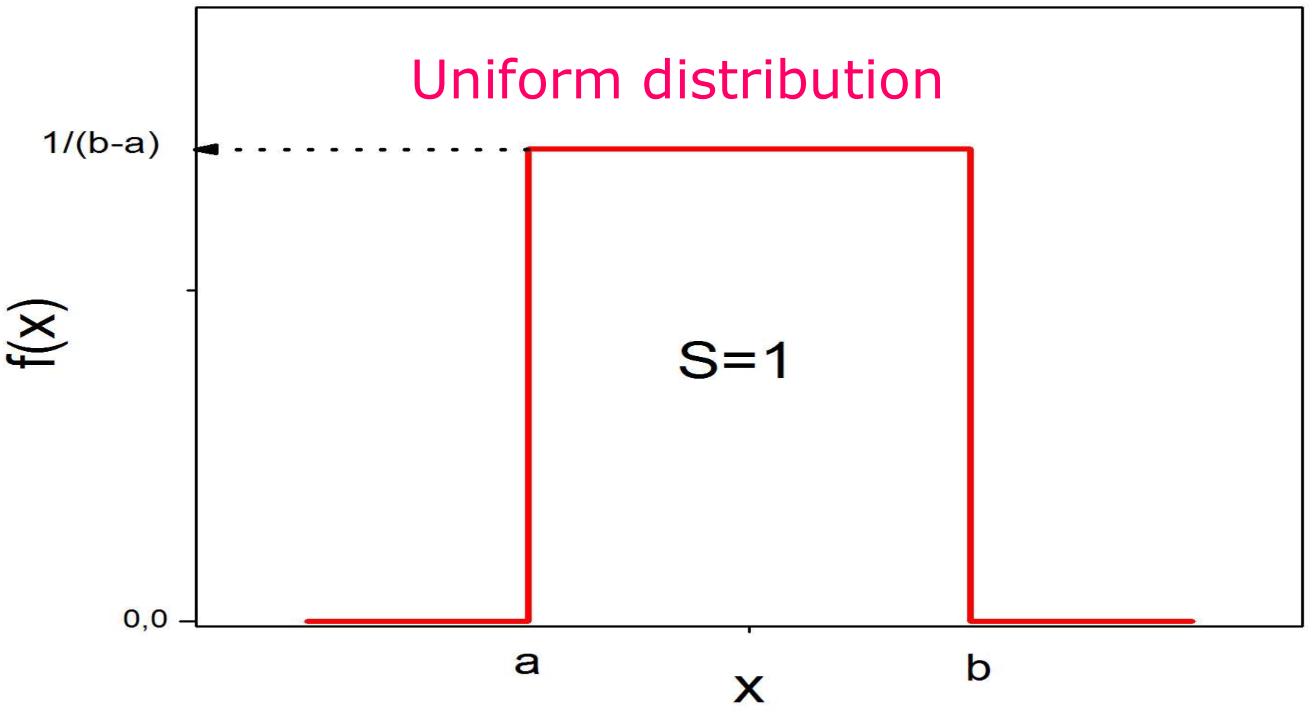
2. For independent variables $X_1, X_2, \ldots X_n$

$$\sigma^2\left(\sum_i C_i X_i\right) = \sum_i C_i^2 \sigma^2(X)$$

UNIFORM DISTRIBUTION

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

$$a \le x \le b$$



Czebyszew inequality

Interpretation of variance results from Czebyszew theorem:

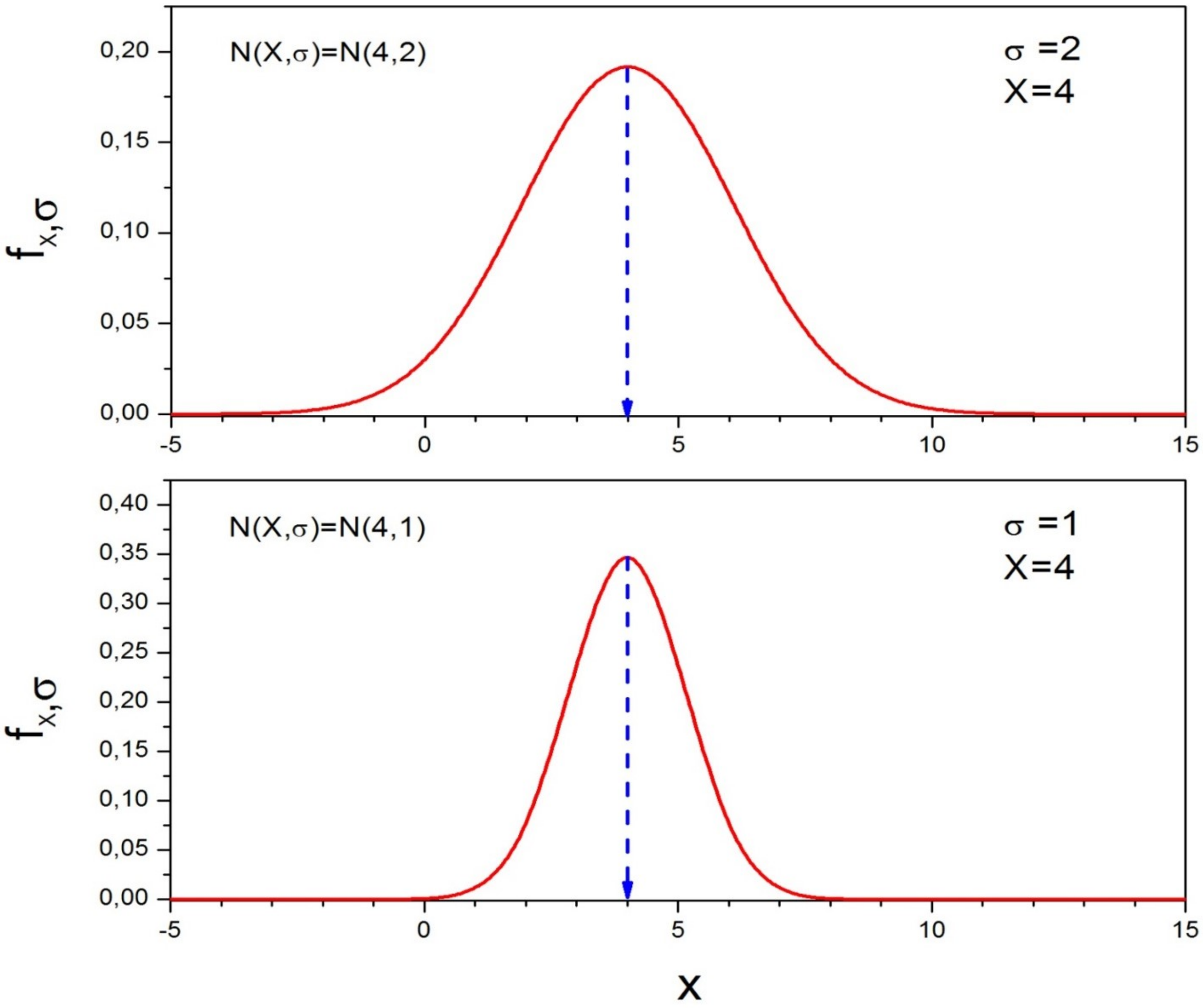
$$P\left(\left|X - E(X)\right| \geq a.\sigma(X)\right) \leq \frac{1}{a^2}$$

Theorem:

Probability of the random variable X to be shifted from the expected value E(X) by a-times standard deviation is smaller or equal to $1/a^2$

This theorem is valid for all distributions that have a variance and the expected value. Number a is any positive real value.

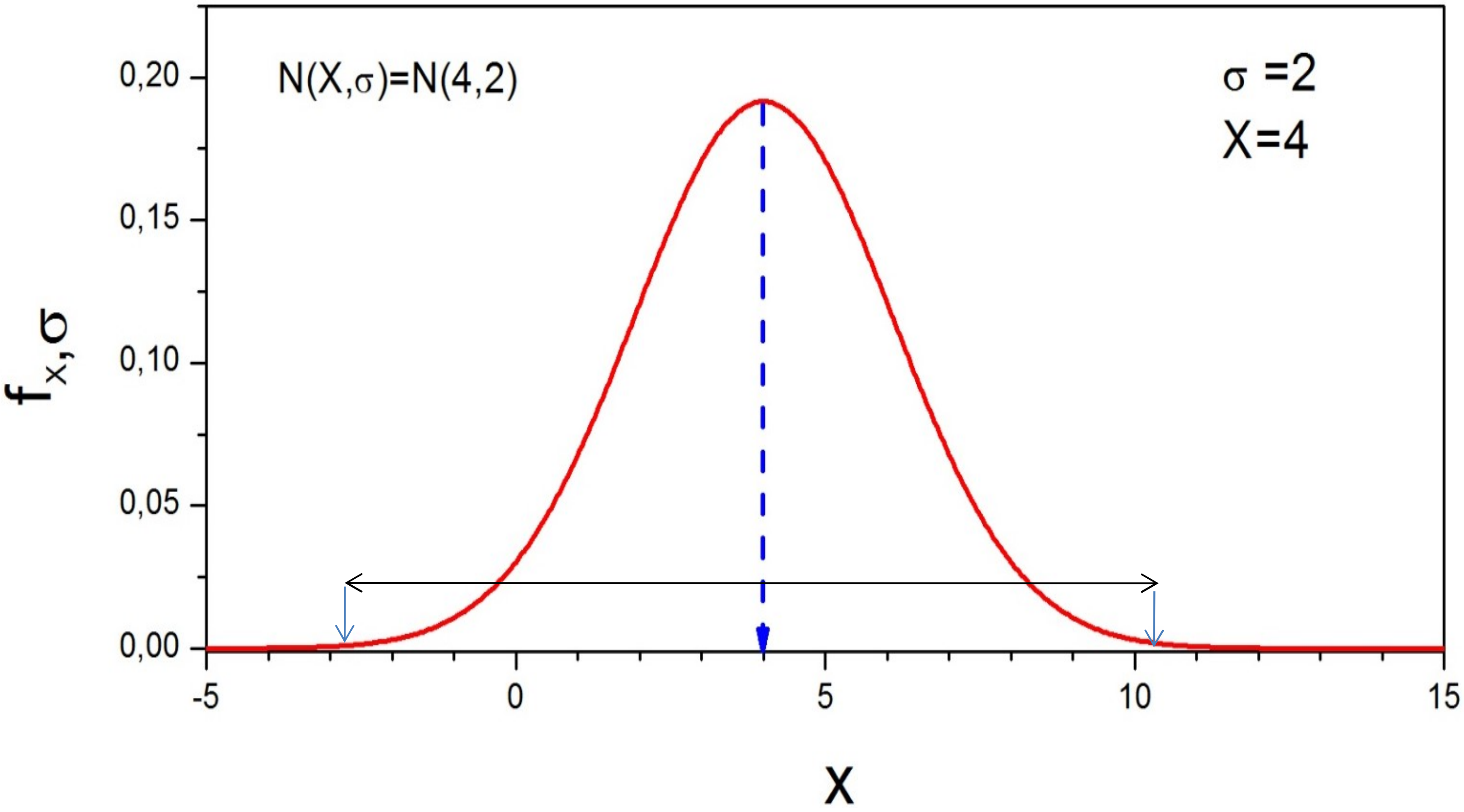
Variance as a measure of data scatter



Big scatter of data

Smaller scatter of data

Range as a measure of scatter



RANGE = $x_{max} - x_{min}$

Practical ways of calculating variance

Variance of n-element sample:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \qquad s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$$

$$\overline{x} - average$$

Variance of N-element population :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2$$

$$\mu - expected \quad value$$

Practical ways of calculating standard deviation

Standard deviation of sample (or: standard uncertainty):

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Standard deviation (population):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2}$$

Examples of probability distributions – discrete variables

Two-point distribution (zero-one), e.g. coin toss, head = failure x=0, tail = success x=1, p – probability of success, its distribution:

x_i	0	1
p_i	1-p	p

Binomial (Bernoulli)

$$p_k = \binom{n}{k} \cdot p^k (1-p)^{n-k} \ , \ k = 0,1,\dots,n$$

where 0<p<1; X={0, 1, 2, … k} k – number of successes when n-times sampled with replacement

For k=1 two-point distribution

Binomial distribution - assumptions

Random experiment consists of *n* Bernoulli trials :

1. **Each trial is independent of others.**
2. **Each trial can have only two results: „success" and „failure" (binary!).**
3. **Probability of success *p* is constant.**

Probability p_k of an event that random variable X will be equal to the number of k-successes at n trials.

$$p_k = \binom{n}{k} \cdot p^k (1-p)^{n-k} \quad , \; k = 0, 1, \ldots, n$$

Pascal's triangle

Symbol $\binom{n}{k} = \frac{n\,!}{(n-k)!k\,!}$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}$$

Newton's binomial

$n = 0$ $\binom{0}{0} = 1$

$n = 1$ $\binom{1}{0} = 1$ $\binom{1}{1} = 1$

$n = 2$ $\binom{2}{0} = 1$ $\binom{2}{1} = 2$ $\binom{2}{2} = 1$

Pascal's triangle

1	n = 0
1 + 1	n = 1
1 2 1	n = 2
1 3 3 1	n = 3
1 4 6 4 1	n = 4
1 5 10 10 5 1	n = 5
1 6 15 20 15 6 1	n = 6

Bernoulli distribution

Example 4.6

Probability that in a company the daily use of water will not exceed a certain level is p=3/4. We monitor a use of water for 6 days.

Calculate a probability the daily use of water will not exceed the set-up limit in 0, 1, 2, …, 6 consecutive days, respectively.

Data:

$$p = \frac{3}{4} \qquad q = \frac{1}{4} \qquad N = 6 \qquad k = 0, 1, \ldots, 6$$

Bernoulli distribution

$k = 0$ $\quad P(k = 0) = \binom{6}{0} \cdot \left(\frac{3}{4}\right)^0 \cdot \left(\frac{1}{4}\right)^6$

$k = 1$ $\quad P(k = 1) = \binom{6}{1} \cdot \left(\frac{3}{4}\right)^1 \cdot \left(\frac{1}{4}\right)^5$

$k = 2$ $\quad P(k = 2) = \binom{6}{2} \cdot \left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^4$

$k = 3$ $\quad P(k = 3) = \binom{6}{3} \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^3$

$k = 4$ $\quad P(k = 4) = \binom{6}{4} \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^2$

$k = 5$ $\quad P(k = 5) = \binom{6}{5} \cdot \left(\frac{3}{4}\right)^5 \cdot \left(\frac{1}{4}\right)^1$

$k = 6$ $\quad P(k = 6) = \binom{6}{6} \cdot \left(\frac{3}{4}\right)^6 \cdot \left(\frac{1}{4}\right)^0$

Bernoulli distribution

$$k=0 \qquad P(0)=1\cdot 1\cdot \frac{1}{4^6}\cong 0.00024$$

$$k=1 \qquad P(1)=6\cdot \frac{3}{4}\cdot \frac{1}{4^5}=\frac{6\cdot 3}{4^6}=18\cdot P(0)\cong 0.004$$

$$k=2 \qquad P(2)=15\cdot \left(\frac{3}{4}\right)^2\cdot \frac{1}{4^4}=\frac{15\cdot 9}{4^6}=135\cdot P(0)\cong 0.033$$

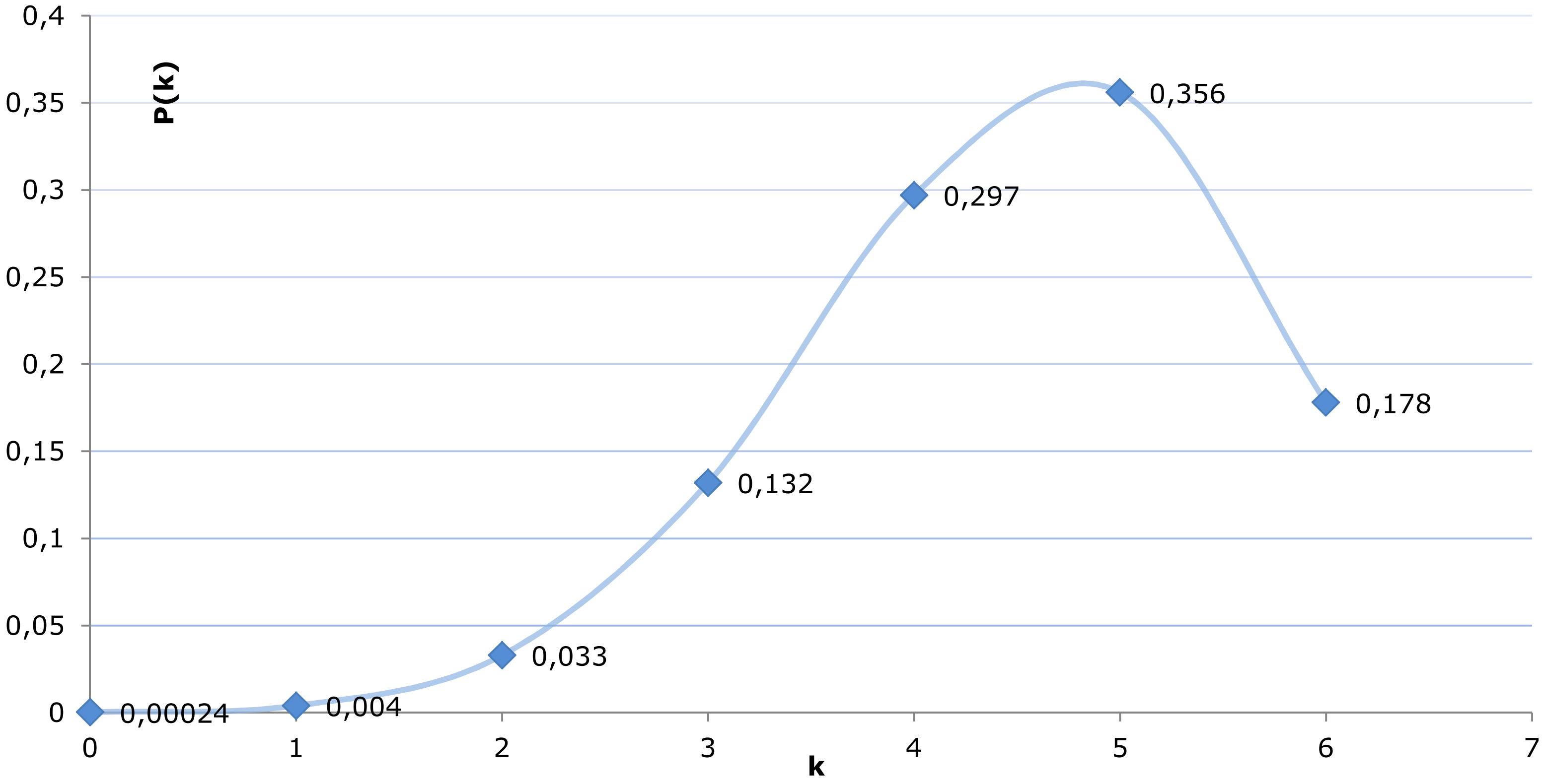
$$k=3 \qquad P(3)=20\cdot \left(\frac{3}{4}\right)^3\cdot \frac{1}{4^3}=\frac{20\cdot 9\cdot 3}{4^6}=540\cdot P(0)\cong 0.132$$

$$k=4 \qquad P(4)=15\cdot \left(\frac{3}{4}\right)^4\cdot \frac{1}{4^2}=\frac{15\cdot 9\cdot 9}{4^6}=1215\cdot P(0)\cong 0.297$$

$$k=5 \qquad P(5)=6\cdot \left(\frac{3}{4}\right)^5\cdot \frac{1}{4^1}=\frac{6\cdot 9\cdot 9\cdot 3}{4^6}=1458\cdot P(0)\cong 0.356$$

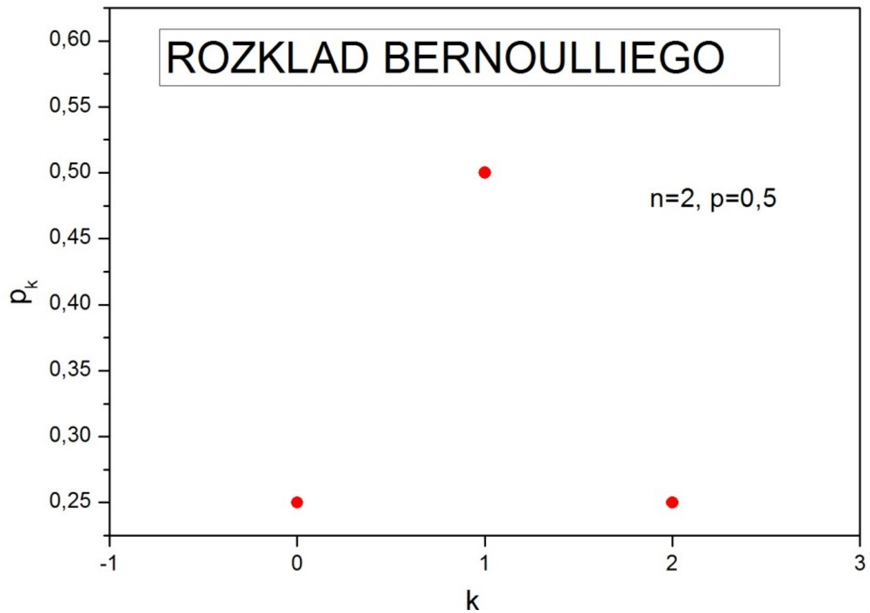
$$k=6 \qquad P(6)=1\cdot \left(\frac{3}{4}\right)^6\cdot \frac{1}{4^0}=\frac{9\cdot 9\cdot 9}{4^6}=729\cdot P(0)\cong 0.178$$

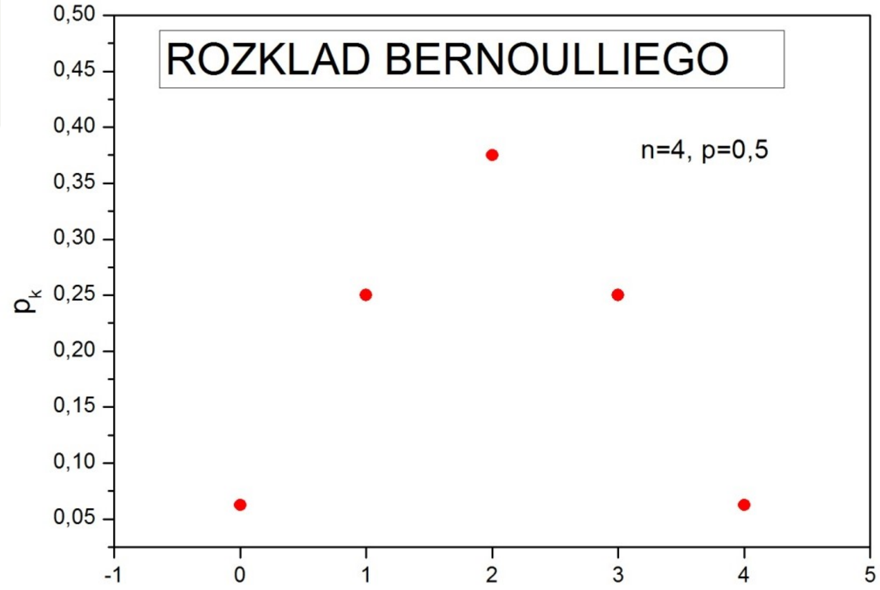
Bernoulli distribution

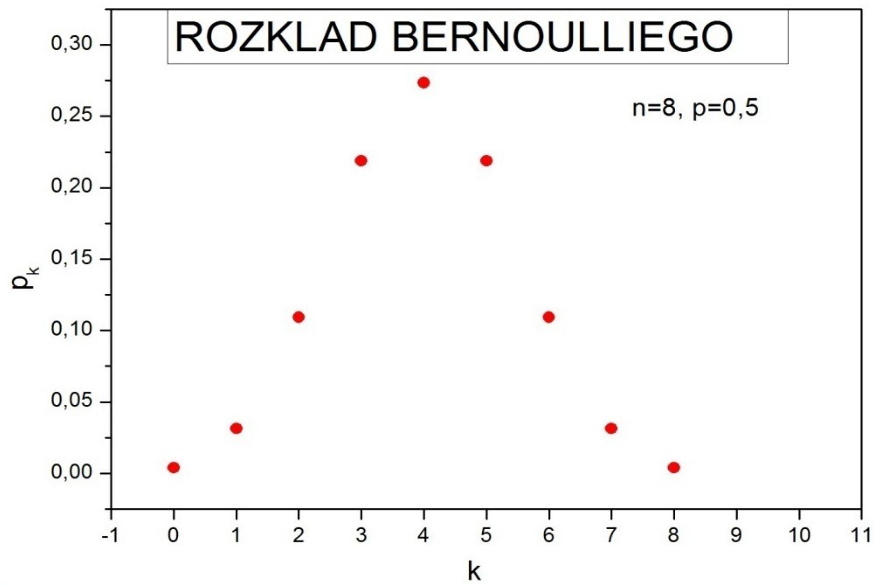


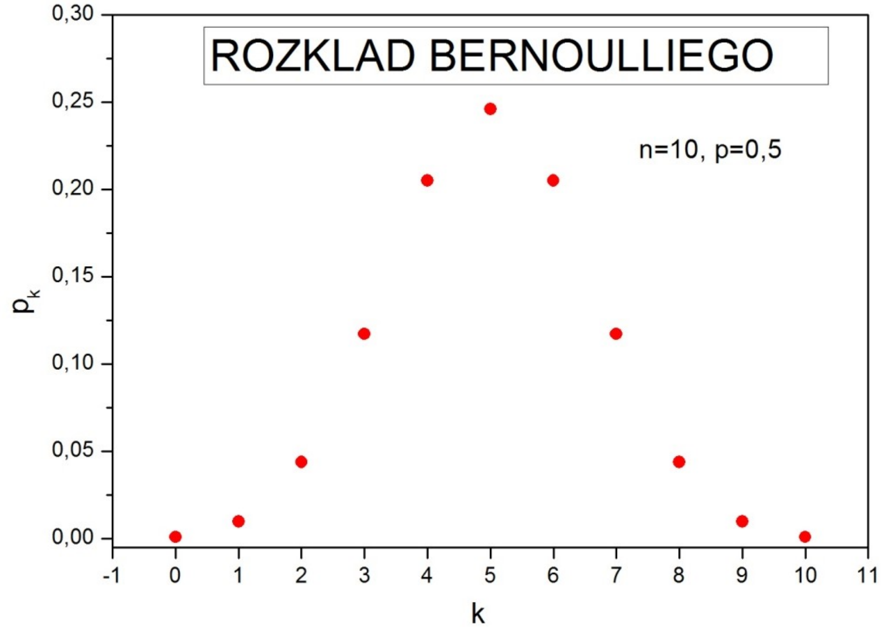
Maximum for k=5

Bernoulli distribution

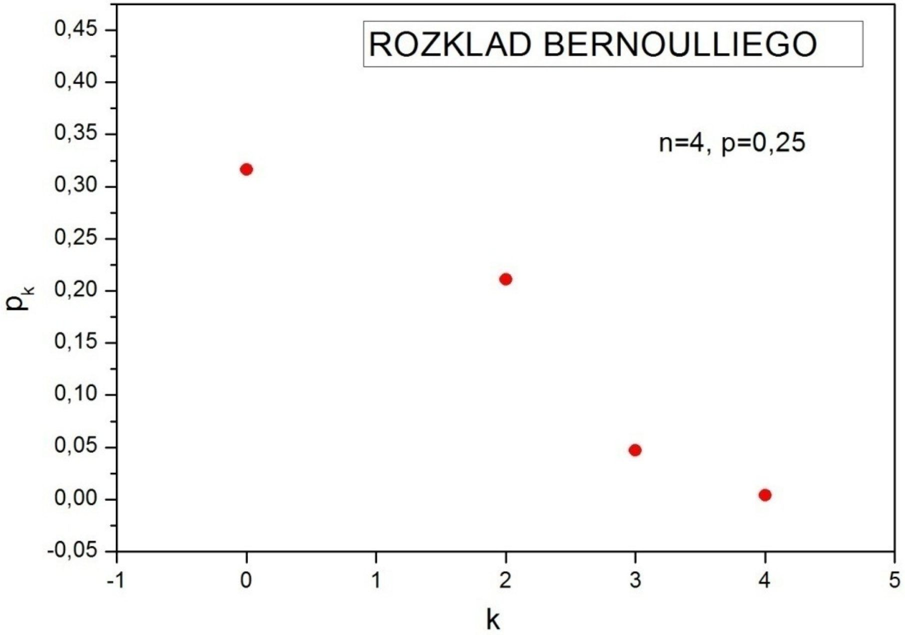


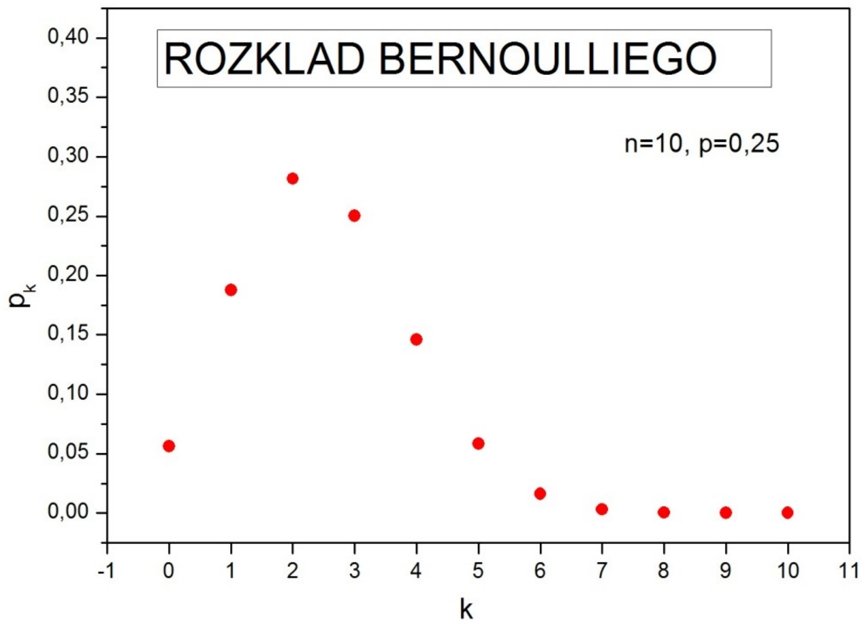






Bernoulli distribution





Expected value

$$E(X) = \mu = np$$

Variance

$$V(X) = \sigma^2 = np(1-p)$$

Errors in transmission

Example 4.7

Digital channel of information transfer is prone to errors in single bits. Assume that the probability of single bit error is $p=0.1$

Consecutive errors in transmissions are independent. Let X denote the random variable, of values equal to the number of bits in error, in a sequence of 4 bits.

E - bit error, O - no error

OEOE corresponds to $X=2$; for EEOO - $X=2$ (order does not matter)

Errors in transmission

Example 4.7 cd

For X=2 we get the following results:

{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE}

What is a probability of P(X=2), i.e., two bits will be sent with error?

Events are independent, thus

$P(EEOO)=P(E)P(E)P(O)P(O)=(0.1)^2\ (0.9)^2\ = 0.0081$

Events are mutually exhaustive and have the same probability, hence

$P(X=2)=6\ P(EEOO)= 6\ (0.1)^2\ (0.9)^2\ = 6\ (0.0081)=0.0486$

Errors in transmission

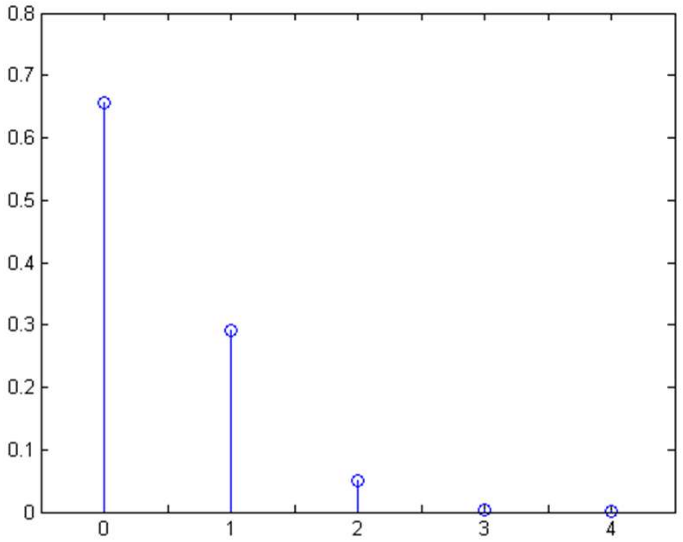
Example 4.7 continued

$$\binom{4}{2} = \frac{4!}{(2)!2!} = 6$$

Therefore, $P(X=2)=6\ (0.1)^2\ (0.9)^2$ is given by Bernoulli distribution

$$P(X = x) = \binom{4}{x} \cdot p^x (1-p)^{4-x} \ , \ x = 0,1,2,3,4,\ p = 0.1$$

P(X = 0) = 0,6561
P(X = 1) = 0,2916
P(X = 2) = 0,0486
P(X = 3) = 0,0036
P(X = 4) = 0,0001



Poisson's distribution

We introduce a parameter λ=pn (E(X) = λ)

$$P(X = x) = \binom{n}{x} \cdot p^x (1-p)^{n-x} = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

Let us assume that n increases while p decreases, but λ=pn remains constant. Bernoulli distribution changes to Poisson's distribution.

$$\lim_{n\to\infty} P(X = x) = \lim_{n\to\infty} \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}$$

Poisson's distribution

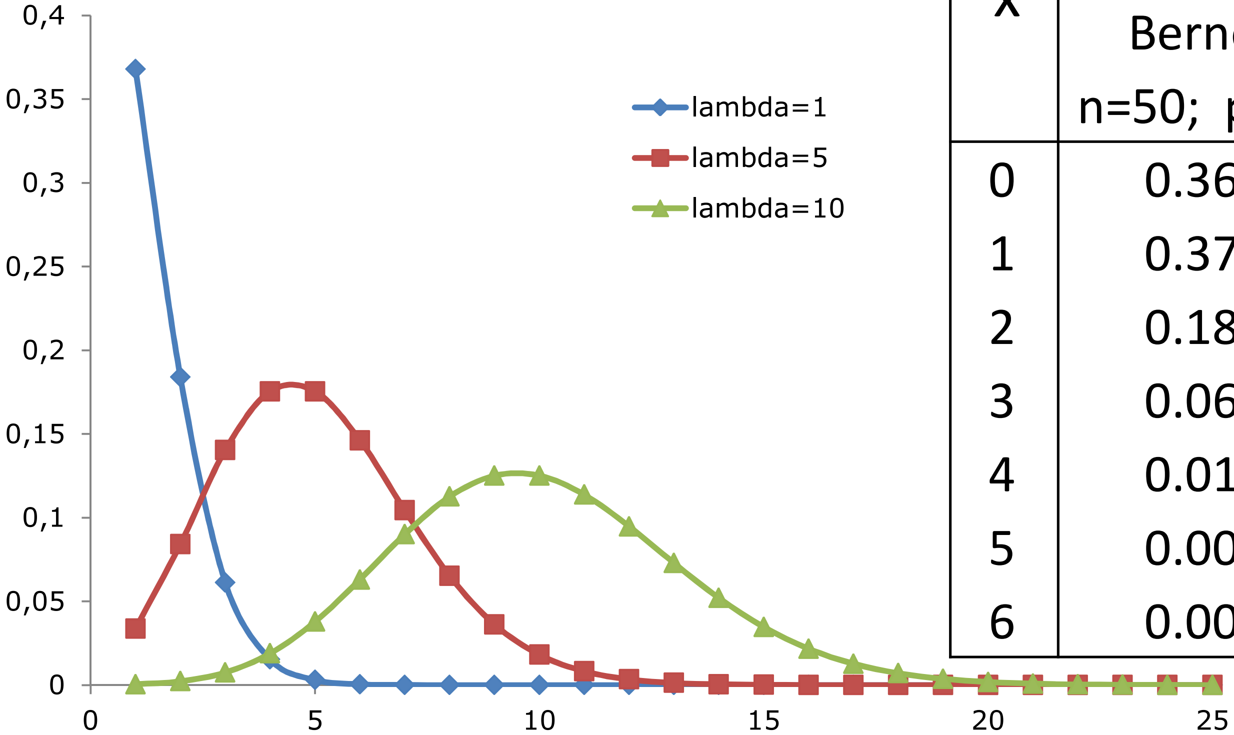
It is one of the rare cases where expected value equals to variance:

$$E(X) = np = \lambda$$

Why?

$$V(X) = \sigma^2 = \lim_{n \to \infty, p \to 0} (np - np^2) = np = \lambda$$

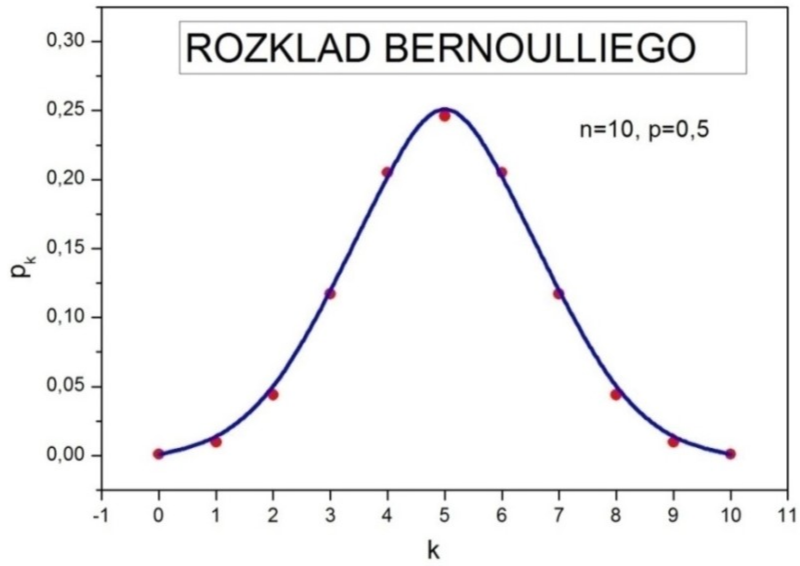
Poisson's distribution



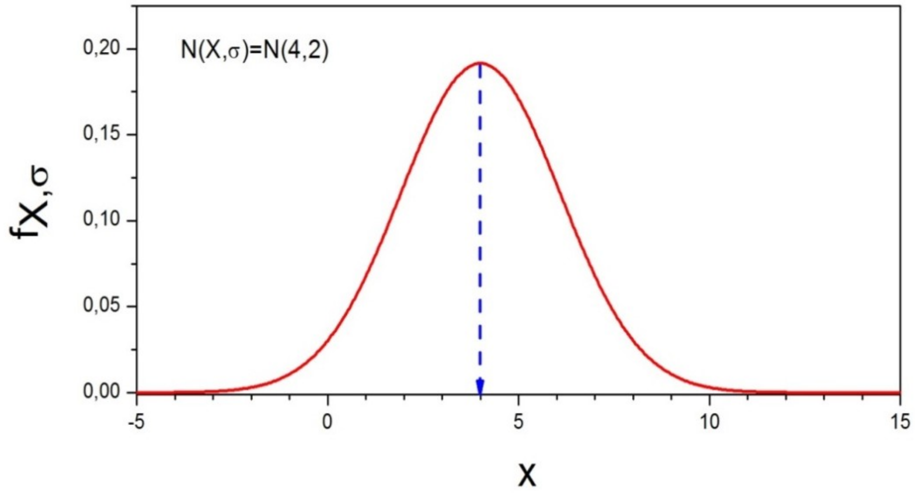
x	p(X)	
	Bernoulli n=50; p=0.02	Poisson: λ=1
0	0.364	0.368
1	0.372	0.368
2	0.186	0.184
3	0.061	0.061
4	0.014	0.015
5	0.003	0.003
6	0.000	0.001

(x- integer, infinite; $x \geq 0$) For big n Bernoulli distribution resembles Poisson's distribution

Normal distribution (Gaussian)



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.

Normal distribution (Gaussian)

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

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

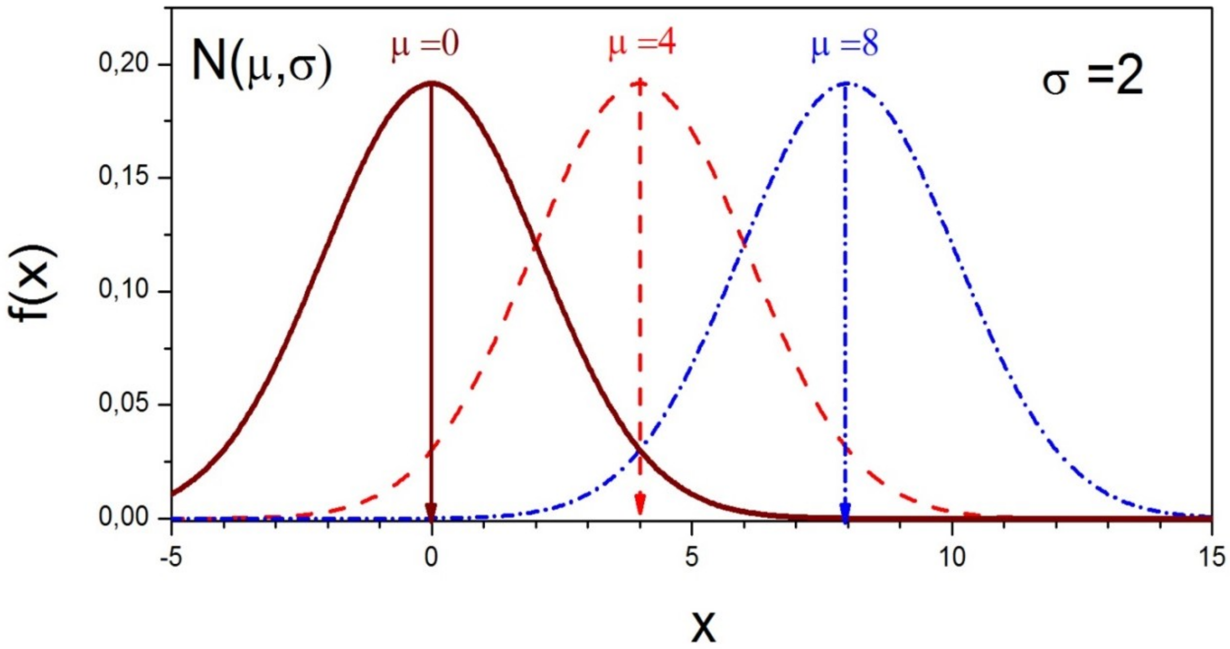
is a **normal random variable** with two parameters:

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

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

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

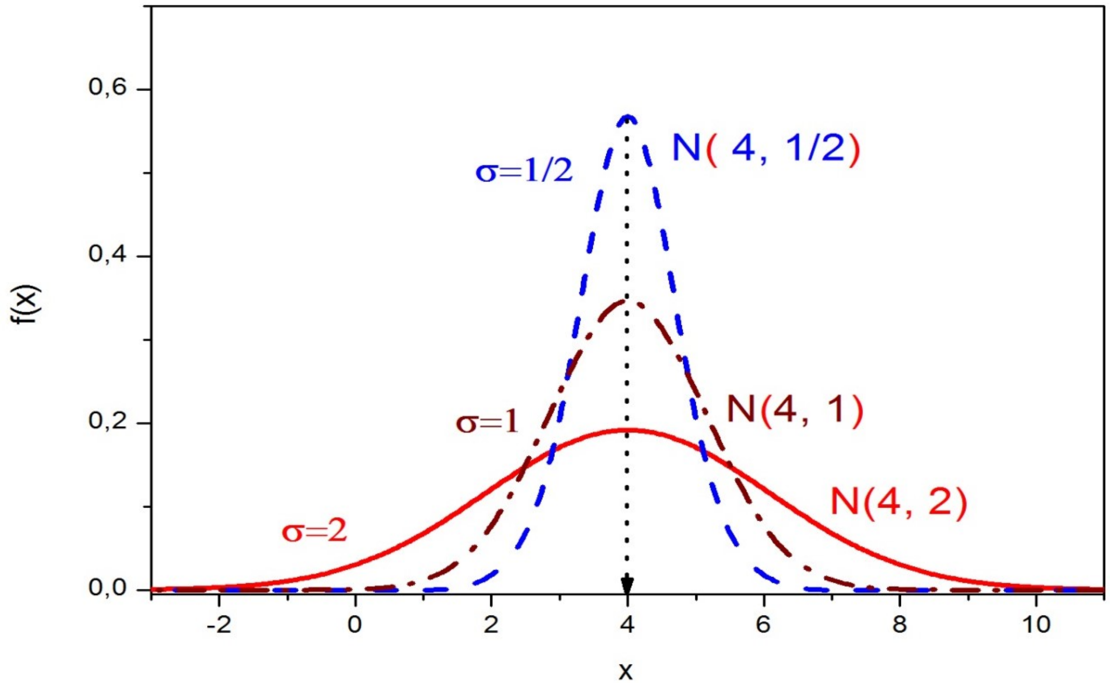
Normal distribution (Gaussian)



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):

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

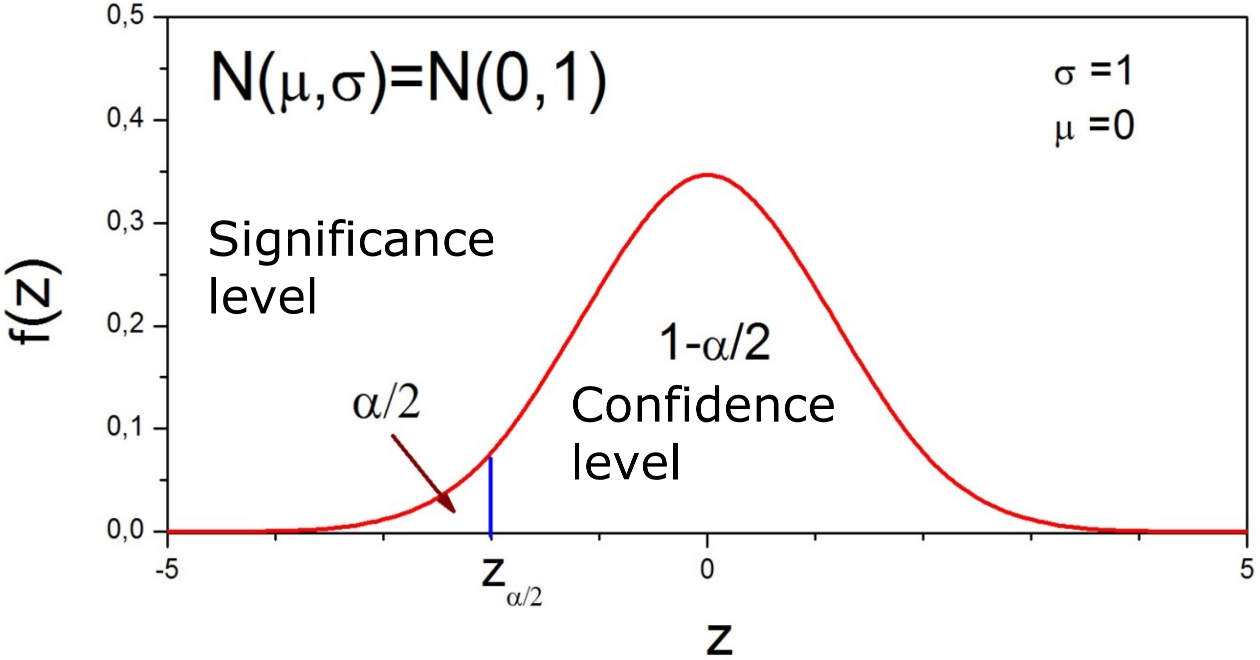
is called a **standard normal random variable**

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

Definition of standard normal variable

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

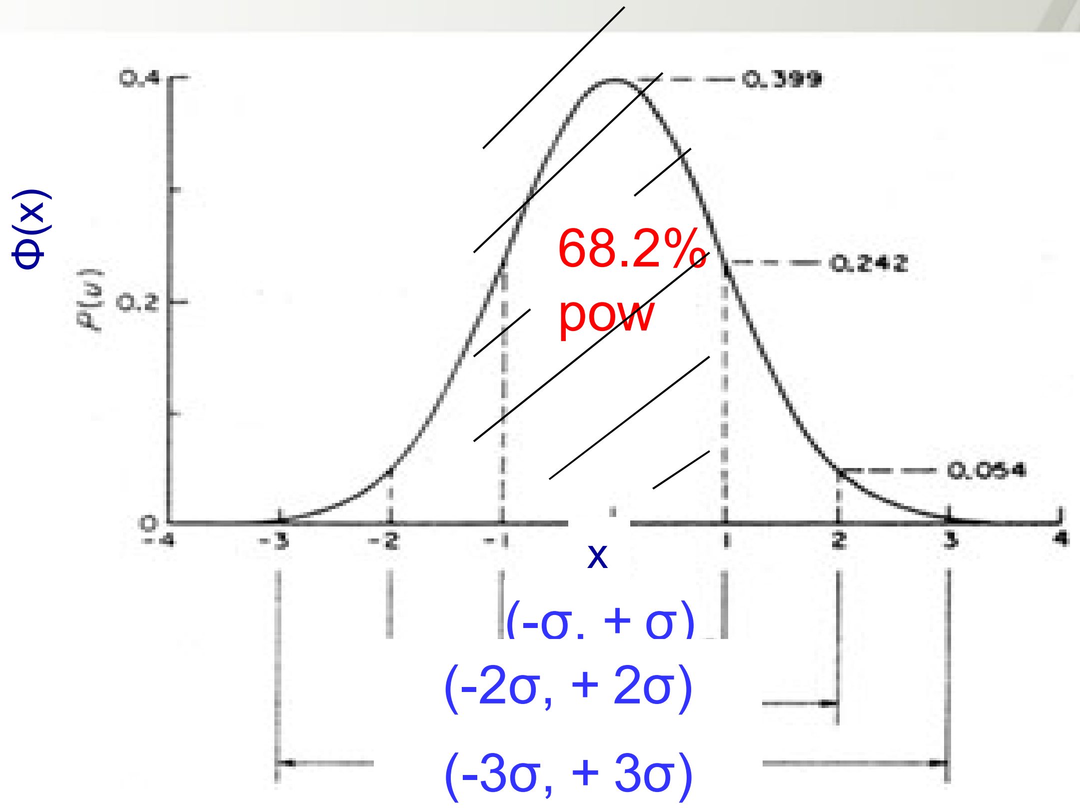
Standard normal distribution



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 distribution.

Calculations of probability (Gaussian distribution)



$P(\mu - \sigma < X < \mu + \sigma) = 0{,}6827$ (about 2/3 of results)
$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0{,}9545$
$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0{,}9973$ (almost all)