Agnieszka SZPARA

Course: Statistics

Field of Study: Finance and Accounting

Form of classes and number of hours: lecture 15 h, laboratory 15 h

Number of ECTS credits: 4

Learning outcomes:

- Student knows the selected discrete probability distributions.
- Student knows the selected continuous probability distributions.
- Student knows the concept of expected value, variance, distribution function.
- Student knows the concept of the two-dimensional distribution.
- Student is able to calculate the expected values, variances, to calculate the cumulative distribution function for the selected discrete and continuous distributions.
- Student is able to calculate the total probability, conditional probability using the classic definition of probability.
- Student is able to calculate the marginal distributions, the correlation coefficient for a discrete two-dimensional random variable.
- Student is prepared to communicate, persuade and defend his/her views in the name of achieving common goals.

Evaluation methods of learning outcomes:
written exam

Subject matter of the classes:

1. Elements of probability.
2. Random variables. Selected discrete and continuous distributions and their characteristics.
3. Normal distribution. Standardisation.
4. Discrete two-dimensional random variable and its characteristics.

References

Books

[1] Montgomery D. C., Runger G. C., *Applied statistics and probability for engineers*.

[2] Pritchett G. D., *Fundamentals of quantitative business methods: business tools and cases in mathematics, descriptive statistics and probability*.

[3] Mendenhall W. M., Sincich T. L., *Statistics for engineering and the sciences*.

Websites

[4] https://goodcalculators.com/probability-calculator/

Table of content

1. The space of elementary events and the general definition of probability

Ω – a set of elementary events, i.e. a set of results that repeat a random experiment any number of times.

A subset $A \subset \Omega$ we call an event,

Ad = $\Omega \setminus A$ we call the event opposite to the event A.

For example:

For the roll of the dice Ω = {1, 2, 3, 4, 5, 6} is the set of possible dots.

An event is, for example, A = {1, 3, 5} saying that an odd number of dots came up. The opposite event is Ad = {2, 4, 6} – an even number of dots was scored.

In fact, in the theory of probability it is said that the events are subsets of A of the Ω set that we can measure somehow. When Ω is finite, the measure of the set A is the number of elements in A. In the case of Ω being infinite, the question of measuring all subsets of the set Ω is very difficult. Therefore, the events A ⊂ Ω are some special subsets for which we are able to determine the measure.

Definition Probability assigns a number P (A) to each event A such that

*Z*0: $0 < P(A) < 1$,

Z1: P (Ω) = 1,

*Z*2: for any sequence (An) of separate events (Ai ∩ Aj = $\emptyset$ $for\ i \neq j$) we have

$$P(nA_n) = \sum_n P(A_n)$$

*Z*2 says that the probability of the sum of disjoint events is equal to the sum of their probabilities. This sum can have an infinite number of components.

Conclusions

W 1 $P(\emptyset) = 0$,

W 2 If $C \subset D$, then $P(D \setminus C) = P(D) - P(C)$,

W 3 $P(A^d) = 1 - P(A)$, ($P(A^d)$ is the probability of the event opposite to *A*),

W 4 for any A and B events $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

1.1. Classic probability

Ω is a finite set, C ⊂ Ω, then | C | denotes the number of elements in the set C.

Classical Probability Definition

$$P(A) = \frac{|A|}{|\Omega|}, A \subset \Omega$$

Exercise 1.

Roll of the dice. Ω = *{*1*,* 2*,* 3*,* 4*,* 5*,* 6*}.* A says it was an even number, B says it was a number greater than 2. We have *A* = *{*2*,* 4*,* 6*}, B* = *{*3*,* 4*,* 5*,* 6*}.*

Classic probability: $P(A) = |A|/|\Omega| = 3/6$*,* $P(B) = |B|/|\Omega| = 4/6$*.*

A ∪ B says that the result is an even number or a number greater than 2. Note that A and B are not disjoint and if Z2 holds, then

$$P(A \cup B) = P(A) + P(B) = \frac{3}{6} + \frac{4}{6} > 1$$

The probability cannot exceed the value 1. W 4 does.

We have $A \cup B = \{2,3,4,5,6\}, A \cap B = \{4,6\}, P(A \cup B) = 5/6\,, P(A \cap B) = 2/6$ and

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$$

Examples of classical probability

1.2. Permutations

The elements of the n-element set Z form a random queue (line up).
Ω is a set of such sequences called permutations of the set Z. The elementary event is a string.

The number of permutations of the n-element set Z is $n! = 1 \cdot 2 \cdots n$*.*

For example

Z = *{a, b, c}* has 3! = 1 · 2 · 3 = 6 permutations/ These are:

abc, acb, bac, bca, cab, cba.

Exercise 2.

A, B, C, D form a random queue. All queues are assumed to be equally probable.

Event G says A will overtake B and C. We have | Ω | = 4! = 24.

What queues does the G event consist of?

Maybe A is 1st and then the 3 others are in further places where they can change in 3! = 6 ways. Or A is second and then D is first, and B and C alternate last 2 to 2! ways.

Hence, G consists of eight queues and P (G) = 8/24 = 1/3.

Sequences of n-elements with coordinates from the set of k-elements.

There are such strings k^n*.*

Exercise 3.

We toss the coin n times.

The elementary event is a string $(x1, x2, \ldots, xn)$, where $xi = O$ (obverse) or$xi = R$(reverse).

ThusΩ ∨ 2^n.

Consider, for example, 3 tosses. We have 2^3 = 8 strings

$$(O,O,O),(O,O,R),(O,R,O),(R,O,O),(R,R,R),(R,R,O),(R,O,R),(O,R,R).$$

Event A says that the coin landed on obverse side2 times. Then */A/* = 3 and $P(A) = \frac{3}{8}$.

Exercise 4.

In the case of n-fold roll of the dice, we have 6^n sequences, i.e. elementary events, i.e.. $/\Omega/$ = 6^n.

Consider 2 dice rolls. What is the probability of event A that the the sum of the dice rolls will be 10?

Answer $\Omega \vee 62 = 36. A = \{(4,6),(6,4),(5,5)\}$.

Then $P(A) = \frac{3}{36} = \frac{1}{12}$.

1.3. Combinations

Definition A k-element combination from the n-element set Z(n - k) is called k-element subset of the set Z.

We have $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ *such combinations*

And $0! = 1,1! = 1$.

Exercise 5.

In the group of 12 there are 7 D (girls) and 5 C (boys). We choose 4 people at random. Event A says that only D was drawn. Here Ω is a set of 4-element combinations from the 12-element set.

Thus $|\Omega| = \binom{12}{4} = \frac{12!}{4!\cdot 8!} = 495$

For A we have

$$|A| = \binom{7}{4} = \frac{7!}{4!\,3!} = 35$$

Then $P(A) = \frac{35}{495} = \frac{7}{99}$

Now B says 2D and 2C are drawn. Here $|B| = \binom{7}{2}\binom{5}{2} = 21 \cdot 10 = 210$

Thus $P(B) = \frac{210}{495} = \frac{42}{99}$

Exercise 6.

LOTTO. We draw 6 numbers from the 49-element set. An elementary event is a

randomly selected 6-element subset. Thus $|\Omega| = \binom{49}{9}$

The probability of hitting a winning six is $P(A) = \frac{1}{|\Omega|}$.

A means a randomly drawn elementary event, which gives a full win. We calculate:

$$\binom{49}{6} = \frac{49!}{6! \cdot 43!} = 11 \cdot 3 \cdot 23 \cdot 47 \cdot 8 \cdot 49 = 13983816$$

Thus $P(A) = \frac{1}{13983816}$

1.4. Conditional and total probability. Bayes' Theorem

Suppose that event B happened, and we are interested in what the probability of event A looks like in this light.

The symbol $P(A \vee B)$ denotes the probability of the event A under the condition of event B.

For example, suppose a dice roll results in a number of dots greater than 3 and

B = {4, 5, 6}.

Then the probability of the event *A = {2, 4, 6}* (the number was even) is

$$P(A \cup B) = \frac{2}{3}.$$

This result can be obtained by the formula considered to be the definition *P* (*A*/*B*) :

Definition $P(A/B) = \frac{P(A \cap B)}{P(B)}$

In the dice roll $P(B) = \frac{1}{2}, A \cap B = \{4,6\}, P(A \cap B) = \frac{2}{6}$.

Thus $P(A/B) = \frac{\frac{2}{6}}{\frac{1}{2}} = \frac{2}{3}$

Definition $B1, B2, \ldots, Bn$ is a complete system of events when $P(Bi) > 0 for i = 1,2, \ldots, n$ and the events Bj are disjoint and $B1 \cup B2 \cup \cdots \cup Bn = \Omega$.

1.4.1. Total probability

For any event A and the complete system $B1, B2, \ldots, Bn$ *we have*

$$P(A) = \sum_{i=1}^{n} P(A/B_i)P(B_i)$$

1.4.2. Bayes' Theorem

For any event A and the complete system $B1, B2, \ldots, Bn$ *we have*

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{P(A)}, i = 1,2, \ldots, n.$$

Exercise 7.

There are white and black balls in 3 boxes. First box has 5 white and 5 black balls, second box has 6 white and 4 black balls, third box has 8 white and 2 black balls. We pick a box and then draw a ball from the selected box. The probabilities of the picked box are equal.

A – a white ball was drawn.

Bi – the ball is drawn from the box *i.*

$$P(B1) = P(B2) = P(B3) = 1/3.$$

Total probability

$$P(A) = P(A|B1)P(B1) + P(A|B2)P(B2) + P(A|B3)P(B3) =$$
$$\frac{0.5}{3} + \frac{0.6}{3} + \frac{0.8}{3} = \frac{19}{30}.$$

Suppose that event A happened, i.e. we know that the drawn ball is white. Such information means that the most probable is that it comes from box 3, and the least probable that it comes from box 1. We calculate it according to the Bayes, Theorem.

$$P(B_3/A) = \frac{\frac{0,8}{3}}{\frac{19}{30}} = \frac{8}{19} > P(B_3) = \frac{1}{3}$$
$$P(B_2/A) = \frac{\frac{0,6}{3}}{\frac{19}{30}} = \frac{8}{19} > P(B_2) = \frac{1}{3}$$
$$P(B_1/A) = \frac{\frac{0,5}{3}}{\frac{19}{30}} = \frac{8}{19} > P(B_1) = \frac{1}{3}$$

The probabilities for boxes 1 and 2 fell below their probabilities. For box 3 we have a clear increase in probability of more than 1/3.

2. Random variable

A random variable is a function that assigns a number to elementary events.

Exercise 1.
We roll a symmetrical dice once. The random variable takes the values from the {1,2,3,4,5,6} set.

2.1. Definition of a random variable

The set of possible results of an experiment is called the space of elementary events and is marked with the symbol Ω, while the element of the set Ω, i.e. a single result of the experiment, is called an elementary event and denoted by the symbol ω.

Definition 1. A random variable is a function defined in the space of elementary events Ω, with values in the set of real numbers R.

The values of a random variable are also called realisations.

Exercise 2.
Consider a coin tossed three times. Random variable X represents the number of times the coin landed on the reverse side. Thus, the random variable X is equal to:
0 when $\omega = \{R,R,R\}$
1 when $\omega = \{\{O,R,R\},\{R,O,R\},\{R,R,O\}\}$
2 when $\omega = \{\{R,O,O\},\{O,O,R\},\{O,R,O\}\}$
3 when $\omega = \{O,O,O\}$.

If we have the values of a random variable, then we can compute the probabilities. Since there are all 8 events, using the classical definition of probability we have

$$P(X=0)=\frac{1}{8}, P(X=1)=\frac{3}{8}, P(X=2)=\frac{3}{8}, P(X=3)=\frac{1}{8}$$

Cumulative distribution function of a random variable and its properties

Definition 2. A cumulative $F(x)$ of a random variable X is a function defined on a set of real numbers described by the following formula $F(x)=P(\omega: X(\omega)<x), (short F(x)=P(X<x))$.

We will use the example **Exercise 2**. The cumulative distribution function for a coin tossed three times is the following function:

x	$x \le 0$	$0<x\le 1$	$1<x\le 2$	$2<x\le 3$	x>3
F(x)	0	1/8	4/8	7/8	1

The distribution diagram is shown in the figure below

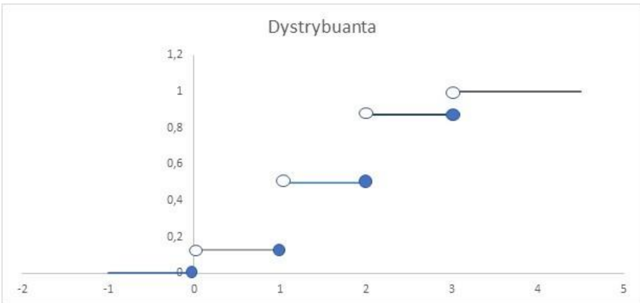


Figure 1.1. Cumulative distribution function for the number of times the coin landed on the reverse side having been tossed three times

The cumulative distribution function of any random variable$X$ has the following properties:

1. The cumulative distribution function $F(x)$ takes values only in the interval [0.1].

2. The cumulative distribution function $F(x)$ is a non-decreasing function.

3. The cumulative distribution factor $F(x)$ is left-hand continuous.

$$4. l \lim_{x\to-\infty} F(x)=0, \lim_{x\to+\infty} F(x)=1.$$

The distributor has the following properties:

$$P(X\ge a)=1-P(X<a)=1-F(a),$$

$$P(a\le X<b)=F(b)-F(a).$$

We divide random variables into discrete (step) and continuous.

2.2. Discrete random variable

Definition 3. Let X be a random variable of the step type. The function p defined in the set $x1, x2, \ldots$ of the equality $p(x_k)=P(X=x_k)=p_k$ is called the probability distribution function or the probability function of the random variable X.

Definition 4. A plot of the probability function in a rectangular coordinate system is a set of points (x_k, p_k).

According to the probability axioms the probability function of the random variable X has the following properties: for all k, $p_k \geq 0$,

$P(\sum_k p_k)=1$

Exercise 3.

The distribution of a random variable is given:

$$p_1 = P(X = 0) = 0.2,\ p_2 = P(X = 1) = 0.4,\ p_3 = P(X = 3) = 0.2,\ p_4 = P(X =5) = 0.1,$$

$p_5 = P(X = 10) = 0.1.$

Thus, the random variable X takes the values from the set $\{0,1,3,5,10\}$.

Now let us calculate some probabilities:

$P(X < 2) = P(X = 0) + P(X = 1) = p_1 + p_2 = 0.2 + 0.4 = 0.6.$

$P(X \leq 3) = P(X = 0)+P(X = 1)+P(X = 3) = p_1+p_2+p_3 = 0.2+0.4+0.2 = 0.8.$

$P(X < 4) = P(X \leq 3) = 0.8.$

$P(2 \leq X < 6) = P(X = 2) + P(X = 3) + P(X = 5) = p_2 + p_3 + p_4 = 0.7.$

$P(X > 7) = P(X = 10) = p_5 = 0.1.$

For this example, the cumulative distribution function is as follows:

$$F(x) = \begin{cases} 0\ for\ x \leq 0 \\ 0.2\ for\ 0 < x \leq 1 \\ 0.6\ for\ 1 < x \leq 3 \\ 0.8\ for\ 3 < x \leq 5 \\ 0.9\ for\ 5 < x \leq 10 \\ 1\ for\ x > 10 \end{cases}$$

2.3. Characteristics of a random variable

Definition 5. The expected value of a random variable X, taking the values x_1, x_2, ... with the probabilities p_1, p_2, ..., respectively, is called the value

$$EX = \sum_k x_k p_k$$

The expected value of the random variable X in Example 2 is equal to

$$EX = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5$$

The expected value of a random variable of type has the following properties:
1. for any real number a, E (a X) = a E (X),
2. for any real number a, $E(X + a) = E(X) + a$,
3. if $E(X)$ and $E(Y)$ exist, then $E(X + Y) = E(X) + E(Y)$.

Definition 6. The second ordinary moment of the random variable X of the step type, taking the values x_1, x_2, ... respectively with the probabilities p_1, p_2, ..., is called the value

$$EX^2 = \sum_k x_k^2 p_k$$

Definition 7. The variance of a discrete (step) random variable X, taking the values x_1, x_2, ..., with the probabilities p_1, p_2, ..., is called the value

$$Var(X) = \sum_k (x_k - EX)^2 \cdot p_k$$

Caution: The variance of a random variable X is often determined from the formula

$$Var(X) = EX2 - (EX)2.$$

Exercise 4.

We have distribution of random variable X:

$P(X = 1) = 0.1$, $P(X = 2) = 0.2$, $P(X = 3) = 0.3$, $P(X = 6) = 0.4$.

Determine $Var(X)$.

Answer:

$EX = 1 \cdot 0.1 + 2 \cdot 0.2 + 3 \cdot 0.3 + 6 \cdot 0.4 = 0.1 + 0.4 + 0.9 + 2.4 = 3.8.$

$EX^2 = 1^2 \cdot 0.1 + 2^2 \cdot 0.2 + 3^2 \cdot 0.3 + 6^2 \cdot 0.4 = 0.1 + 0.8 + 2.7 + 14.4 = 18.$

$Var(X) = EX^2 - (EX)^2 = 18 - 3.8^2 = 3.56.$

If X is a random variable for which EX^2 exists, then $Var(X)$ exists, that has the following properties:

- $Var(X) \geq 0$,
- $Var(c\,X) = c^2 \cdot Var(X)$,
- $Var(X + a) = Var(X)$.

2.4. Selected discrete distributions of a random variable

Bernoulli distribution and Poisson distribution and their characteristics are presented below. The Bernoulli distribution is one of the simplest discrete distributions. It allows, with a constant probability of success, to calculate the probability of k successes in n trials. The Poisson distribution is an approximation of the binomial distribution for a large sample and a low probability of success. The Poisson distribution is applicable in:

- statistical receipt of goods,
- statistical production quality control,
- describing damage to various types of devices,
- demography,
- describing transport processes,
- number of calls recorded by the control panel,
- the number of accidents occurring to the insured person.

2.4.1. Binomial distribution

Binomial distribution is a common probability distribution that models the probability of obtaining one of two outcomes under a given number of parameters. It summarises the number of tries when each try has the same chance of attaining one specific outcome. The value of a binomial is obtained by multiplying the number of independent tries by the successes.

n – number of fixed trials,
k – number of successes,
p – probability of success in one try,
q – probability of failure in one try.

$$p + q = 1$$

The binomial probability formula is as follows:

$$P(k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

For this distribution $EX = np, Var(X) = npq$.

Exercise 5.

We toss a coin 5 times. Calculate the probability that the coin lands on the obverse side:

a) Twice
b) At least 4 times
c) At most 2 times

$$n = 5, p = q = 0.5,$$

Answer:

a) $P(k = 2) = \binom{5}{2} \cdot 0.5^2 \cdot 0.5^3$

b) $P(k \geq 4) = P(k = 4) + P(k = 5) = \binom{5}{4} \cdot 0{,}5^4 \cdot 0{,}5^1 + \binom{5}{5} \cdot 0.5^0 \cdot 0.5^0$

c) $P(k \leq 2) = P(k = 2) + P(k = 1) + P(k = 0) =$

$$\binom{5}{2} \cdot 0{,}5^2 \cdot 0{,}5^3 + \binom{5}{1} \cdot 0{,}5^1 \cdot 0{,}5^4 + \binom{5}{0} \cdot 0{,}5^0 \cdot 0{,}5^5$$

I this example $EX = 5 \cdot 0.5 = 2.5, Var(X) = 5 \cdot 0.5 \cdot 0.5 = 1.25$

2.4.2. Poisson Distribution

In statistics, a Poisson distribution is a probability distribution that can be used to show how many times an event is likely to occur within a specified period of times. In other words, it is a count distribution. Poisson distributions are often used to understand independent events that occur at a constant rate within a given interval of time. It was named after French mathematician Siméon Denis Poisson.

The Formula for the Poisson Distribution Is

$$P(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!},$$

- e is Euler's number (e = 2.71828...),
- k is the number of occurrences,
- $k!$ is the factorial of k,
- $\lambda = n \cdot p$.

In this distribution $EX = \lambda, Var(X) = \lambda$

Exercise 6.

The probability that a new car will break down after 3 years of use is 0.05. 100 car were sold. Count the probability of breakdown of:
a) 2 cars,
b) At least 2 cars,
c) At most 2 cars.

Answer

$$n = 100, p = 0.05, \lambda = 100 \cdot 0.05 = 5$$

a) $P(k = 2) = e^{-5} \cdot \frac{5^2}{2!}$

b) $P(k \geq 2) = 1 - P(k < 2) = 1 - P(k \leq 1) = 1 - \left(e^{-5} \cdot \frac{5^1}{1!} + e^{-5} \cdot \frac{5^0}{0!}\right)$

c) $P(k \leq 2) = P(k = 2) + P(k = 1) + P(k = 0) =$

$$e^{-5} \cdot \frac{5^2}{2!} + e^{-5} \cdot \frac{5^1}{1!} + e^{-5} \cdot \frac{5^0}{0!}$$

$$EX = 5, Var(X) = 5$$

2.5. Continuous random variable

Continuous random variables take values from a certain numerical range (a, b). *Sometimes even from the compartment (-∞, ∞).*

Thus, the number of possible values for such a variable is uncountable.

It follows that the probability that a continuous random variable will take a value equal to a specific number equals zero.

Only the probability that this variable will take a value belonging to some sub-interval of the interval (a, b) can be different from zero.

For this reason, the diagram cannot be used to describe the probability distribution of a continuous random variable.

Its counterpart is the probability density function.

2.5.1. Probability density function

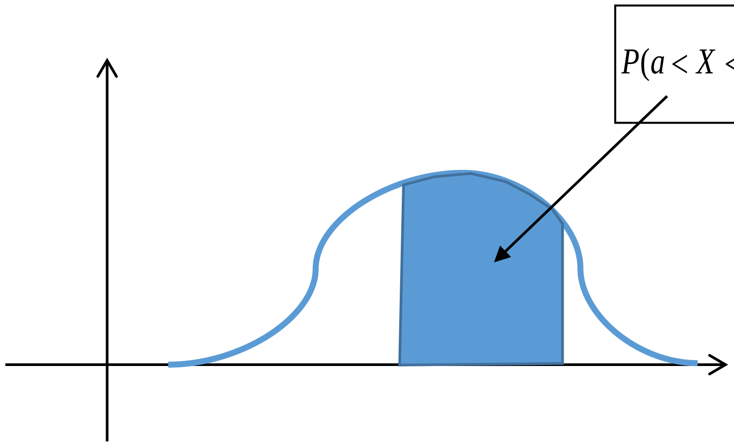
Each function $f(x)$ with the following properties:

f(*x*) ≥ 0 for each value of $x$belonging to its domain;

The area between the graph of the function $f(x)$and the abscissa axis equals 1, $\int_{-\infty}^{\infty} f(x)dx = 1$,

is the probability density function of some random variable.

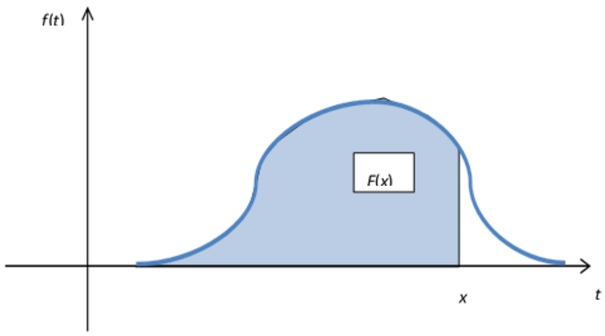
The relationship between the probability density function and the probability that the values of the random variable X fall within the interval (a, b).



2.5.2. Cumulative distribution function of a continuous random variable

When considering a discrete random variable the definition of the given distribution function remains valid. However, now, as shown in the previous figure, the condition $F(x) = P(X < x)$ for $x \in R$ means that the value of the distribution function $F(x)$ is the area between the graph of the density function $f(x)$ and the axis OX in the interval $(-\infty, x)$.

Formally, it is mathematically written as $F(x) = \int_{-\infty}^{x} f(t)dt$



The cumulative distribution function of a continuous random variable is a function:

- continuous,
- non-decreasing $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$,
- as for a step random variable: $P(a \leq X < b) = F(b) - F(a)$.

The expected value of a continuous random variable X and the variance are calculated as follows:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$Var(X) = \int_{-\infty}^{\infty} \left(x - E(x)\right)^2 f(x)dx$$

For any X and Y random variables and any constant $a \in R$, the following properties of the expected value are true:

- *E*(*a*) = *a*,
- *E*(*aX*) = *aE*(*X*),
- *E*(*X*±*Y*) = *E*(*X*) ± *E*(*Y*).

3. Normal distribution

The best known and most widely used continuous distribution is the normal distribution, which describes many natural phenomena. It is defined by two values – the mean and the standard deviation. Some of its uses are: ROE, People Growth, Inflation Rate, Sales Revenue. Thanks to the Central Limit Theorem, most of the phenomena can be approximated by a normal distribution.

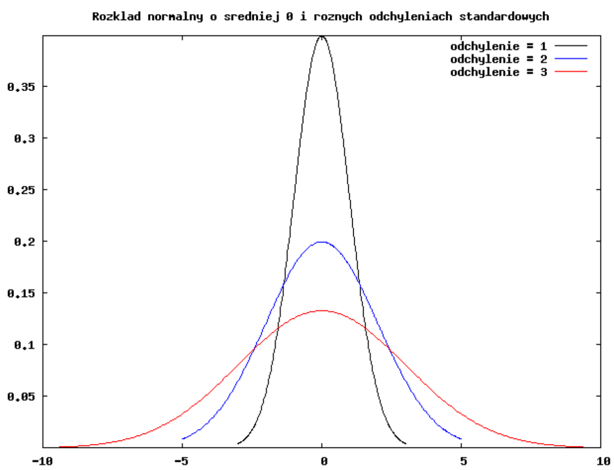
Normally distributed random variables are of particular importance in the statistical analysis.

The normal distribution is described by the following probability density function: $f(x) = \frac{1}{\sqrt{2\sigma^2}} exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$

The plot of this function is called the Gaussian curve.

The normal distribution density function depends only on two parameters:
- Expected value μ,
- Standard deviation σ.

The line $x = \mu$ is the symmetry axis of the normal distribution density function plot. The standard deviation affects the shape of the graph of this function: the larger it is, the flatter the graph



The fact that the random variable X has a normal distribution with the expected value μ and standard deviation σ is written symbolically in the form of

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

if $X \sim N(\mu,\sigma)$, then:

- 68% value of variable X is in (μ - σ, μ + σ);
- 95,5% value of variable X is in (μ - 2σ, μ + 2σ),
- 99,7% value of variable X is in (μ - 3σ, μ + 3σ).

The last of the above-mentioned rules is called the three-sigma rule.

Exercise 7.
As a result of the research, it was found that the height X of cob settling of a certain variety of maize is a random variable with a normal distribution with an expected value of 55 cm and a standard deviation of 6 cm. Calculate the probability of that

- $X > 60$,
- $40 < X < 70$.

Answer

a) According to the definition of cumulative distribution, we have

P ($X > 60$) = 1 – F(60)

b) P ($40 < X < 70$) = F(70) – F(40)

However, the distribution function of the random variable $X \sim N(55, 6)$ has a rather complicated form

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

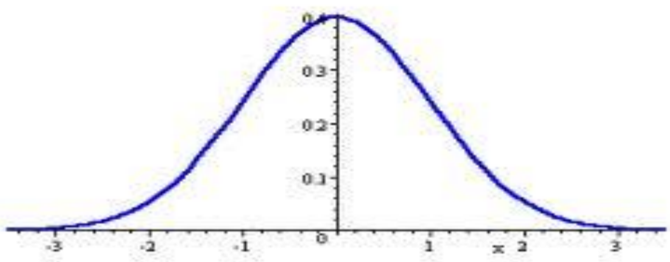
$$f(t) = \frac{1}{\sqrt{2 \cdot 6^2}} exp\left[\frac{-(t-55)^2}{2 \cdot 6^2}\right]$$

You can use the appropriate tables of the standard normal distribution.

For this, however, the random variable X must be standardised.

3.1. Random variable with standard normal distribution.

We say that a random variable$X$ with a normal probability distribution has a standard normal distribution if its expected value is zero and its standard deviation is equal to one, that is,
$X \sim N(0, 1)$.



3.1.1. Standardisation

If $X \sim N(\mu, \sigma)$, then random variable $\frac{X-\mu}{\sigma}$ has a standard normal distribution.

Exercise 7. (continuation)

$X \sim N(55, 6)$, then $\frac{X-55}{6}$ N (0, 1)

- P (X > 60) = 1 – P(X < 60)

We carry out standardisation

$$P(X < 60) = P\left(\frac{X-55}{6} < \frac{60-55}{6}\right) = P\left(\frac{X-55}{6} < \frac{5}{6}\right) \approx F(0.83)$$

Where F is the standard normal distribution distribution. We read from the tables

F (0.83) = 0.7967

Thus P (X > 60) = 1 – 0.7967 = 0.2033.

- P (40 < X < 70)

By doing likewise, we get

$$P(40 < X < 70) = P\left(\frac{40-55}{6} < \frac{X-55}{6} < \frac{70-55}{6}\right) = P\left(-2.5 < \frac{X-55}{6} < 2.5\right)$$
$$= F(2.5) - F(-2.5)$$

Due to the symmetry of the standard normal distribution density function with respect to the OY axis, we have:

$F(-\alpha) = 1 - F(\alpha)$ for each real number $\alpha > 0$.

Thus $F(-2.5) = 1 - F(2.5)$

We read from the tables the distribution function of the normal distribution of the standard value

$F(2.5) = 0.9938$

We get:

$$P(40 < X < 70) = F(2.5) - (1 - F(2.5)) = 0.9938 - 1 + 0.9938 = 0.9876$$

4. Two dimensional discrete random variable and its characteristics

Let X and Y be random variables defined not necessarily on the same probabilistic space. A pair of random variables X, Y are called a two-dimensional random variable or a two-dimensional random vector, and X and Y are their coordinates.

Two-dimensional random variable (X, Y) that takes a finite or countable number of values (x_i, y_k), each with a probability.

$P(X = x_i, Y = y_k) = p_{ik}$ for $i, k \in N$, Wherein $\sum_i \sum_k p_{ik} = 1$, we call a step-type two-dimensional random variable.

The function which assigns appropriate probabilities p_{ik} to the values (x_i, y_k), is called the probability function of a two-dimensional random variable (X, Y).

The $F(x, y)$ cumulative distributor of a two-dimensional step random variable is given by the formula

$F(x, y) = \sum_{x_i<x} \sum_{y_k<y} p_{ik}$.

If a two-dimensional random variable (X, Y) takes a finite number of values, it is convenient to put the values of the probability function in the following table

X/Y	y_1	y_2	...	y_b
x_1	p_{11}	p_{12}		p_{1b}
x_2	p_{21}	p_{22}		p_{2b}
...				
x_a	p_{a1}	p_{a2}		p_{ab}

We say that the probability distribution of a two-dimensional step random variable has been determined when its distribution function or probability function is known.

The boundary distribution of a random variable X in the distribution of a two-dimensional discrete random variable (X, Y) is defined as:

$p_{i.} = P(X = xi) = \sum_k p_{ik}$ dla $i \in N$.

Similarly, we define the boundary distribution of a random variable Y :

$p_{.k} = P(Y = y_k) = \sum_i p_{ik}$ dla $k \in N$.

X/Y	y_1	y_2	...	y_b	$p_{i.}$
x_1	p_{11}	p_{12}	...	p_{1b}	$p_{1.}$
x_2	p_{21}	p_{22}	...	p_{2b}	$p_{2.}$
...	...	...	...	...	...
x_a	p_{a1}	p_{a2}	...	p_{ab}	$p_{i.}$
$p_{.k}$	$p_{.1}$	$p_{.2}$	...	$p_{.k}$	1

Let g be a continuous real function of two real variables and let (X, Y) be a two-dimensional random variable. Then $Z = g(X, Y)$ is a random variable. In particular, the random variables are sums and products of the random variables.

If (X, Y) is a random variable of a step type, then the expected value of a random variable $Z = g(X, Y)$ is given by the formula

$$E\big(g(X, Y)\big) = \sum_i \sum_k g(x_i, y_k) p_{ik}:$$

For $Z = g(X, Y)$= X or $Z = g(X, Y) = Y$

We obtain

$$EX = \sum_i x_i p_{i.}\,, EX^2 = \sum_i x_i^2 p_{i.}, Var(X) = EX^2 - (EX)^2$$

$$EY = \sum_k y_k p_{.k}\,, EY^2 = \sum_k y_k^2 p_{.k}, Var(Y) = EY^2 - (EY)^2$$

If X and Y are random variables, then their covariance is a number defined by one of the formulas:

$$cov(X,Y) = E([X - E(X)][Y - E(Y)])$$
$$cov(X,Y) = E(XY) - E(X)E(Y).$$

If random variables X i $Y$are independent, then $cov(X,Y) = 0$.

The correlation coefficient of the random variables X and Y is determined by the formula:

$$r(x,y) = \frac{cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Properties of the correlation coefficient:

- $1 \le r(x,y) \le 1$
- If X and $Y$are independent, then $r(x,y) = 0$
- $|r(x,y)| = 1 \Leftrightarrow P(Y = aX + b) = 1$

The variables for which the correlation coefficient equals 0 are said to be uncorrelated.

Exercise 8.

In the table we have a distribution of two-dimensional random variable. Find a correlation coefficient.

X/Y	**1**	**2**
0	0,3	0,3
1	0,2	0,2

First, we have to find marginal probabilities of variable X and Y.

So we have to add up probabilities in rows and in columns.

X/Y	**1**	**2**	**sum**
0	0,3	0,3	0,6
1	0,2	0,2	0,4
Sum	0,5	0,5	1

So:

x_i	0	1
$p_{i.}$	0,6	0,4

And:

y_i	1	2
$p_{.k}$	0,5	0,5

Then:

- $EX = 0 \cdot 0.6 + 1 \cdot 0.4 = 0.4$
- $EX^2 = 0^2 \cdot 0.6 + 1^2 \cdot 0.4 = 0.4$
- $Var(X) = 0.4 - 0.4^2 = 0.24$
- $EY = 1 \cdot 0.5 + 2 \cdot 0.5 = 1.5$
- $EY^2 = 1^2 \cdot 0.5 + 2^2 \cdot 0.5 = 2.5$
- $Var(Y) = 2.5 - 1.5^2 = 0.25$
- $EXY = 0 \cdot 1 \cdot 0.3 + 0 \cdot 2 \cdot 0.3 + 1 \cdot 1 \cdot 0.2 + 1 \cdot 2 \cdot 0.2 = 0.6$
- $cov(X, Y) = 0.6 - 0.4 \cdot 1.5 = 0$
- $r(x, y) = \frac{0}{\sqrt{0.24}\sqrt{0.25}} = 0$

Variables $X$and Y are independent.

Probability and statistics are two branches of mathematics that are highly related, so they are often studied together since statistical analysis often involve the use of probability distributions. There are also many aspects of probability theory that involve topics that are almost entirely mathematical.

Sample tasks

1. We throw the ball 20 times into a basket. What is the probability that there will be:
   - 7 successful throws,
   - At most 10 successful throws,
   - At least 5 successful throws.

2. The probability that a new car will break down after 3 years of use is 0.05

100 cars were sold. What is the probability that there will be:

- 5 broken cars,
- 10 broken cars,
- More the 3 broken cars,
- Less than 7 broken cars.

3. There are machines of type A and machines of type B in the plant. They produce 40% and 60% of the items, respectively, with machines of type A producing 2% of defective items and machines of type B producing 4% of defective items. One item was taken randomly. Calculate the probability that this item is:
   a) defective,
   b) manufactured by a type A machine if known to be defective.

4. We have a distribution of a random variable X.

x_i	**-2**	**0**	**1**	**2**
p_i	0.1	0.2	0.4	A

Calculate:

- A
- $P(X < 1.5)$
- $P(X > 0.2)$
- EX
- $Var(X)$
- $EZ, when\ Z = 3X - 2.$

5. We have two-dimensional random Variable taken from a table. What is a value of correlation coefficient?

X/Y	1	2	3	4	6
-1	0.1	0.1	0	0.3	0.15
1	0.05	0.15	0.1	0.05	0

Examples of questions:

1. Define a random variable.
2. What is it and how do we calculate the variance of a random variable?
3. When the correlation coefficient equals 0?

Lecturer:
Agnieszka Szpara, PhD

e-mail address: agnieszka.szpara@pwsz.nysa.pl