ERASMUS: Course Probability and Statistics
Part I: Probability - summary

Definition:
The set of all possible outcomes of a random experiment is called the *sample space* of the experiment. The sample space is denoted as X.

A sample space is *discrete* if it consists of a finite or countable infinite set of outcomes.

A sample space is *continuous* if it contains an interval (either finite or infinite) of real numbers.

Definition:
An **event** is a subset of the sample space of a random experiment.

Some of the basic set operations are summarized below in terms of events:

- the *union* of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $A \cup B$.
- The *intersection* of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $A \cap B$.
- The *difference* of two events A and B is the event that consists of all outcomes that are contained in A but are not contained in B. We denote the difference as $A \setminus B$.
- The *complement* of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the component of the event A as A' ($A' = X \setminus A$).

Definition:
Two events, denoted as A and B, such that $A \cap B = \emptyset$ are said to be *mutually exclusive***.**

Definition:
Let X be a sample space and A be any event. *Probability* is a measure that is assigned to each member of a collection of events from a random experiment which satisfies the following properties:

1. $P(A) \geq 0$
2. $P(X) = 1$
3. If $A_1, A_2, \ldots$ are countable mutually exclusive events in X then
$$P(A_1 \cup A_2 \cup \ldots) = \sum_i P(A_i)$$

Properties:

- $P(A') = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A \subseteq B$ then $P(A) \leq P(B)$

For a discrete sample space, the probability of an event *A*, denoted as *P*(A), equals the sum of the probabilities of the outcomes in A: $P(A)=\sum_i p_i$, where A={e_1,e_2,…} and p_i is the probability of {e_i}.

If X is finite and n is a number of elementary events of X, m n is a number of elementary events of A and p_i=P({e_i}) for every i=1,…,n then $P(A)=\frac{m}{n}$.

Example:

The **sample space is defined as space of events – results of the play with cubic die** marked on each of its six faces with a different number.

Then X={ {1}, {2}, {3}, {4}, {5}, {6}} and X is a discrete space of events. All of these numbers have the same appearance within a set of dice. So, $P(\{1\})=\frac{1}{6}$. If A ={{2},{4}, {6}} then $P(A)=\frac{1}{2}$

Definition (elements of combinatorial mathematics):

- *permutation* of the elements is an ordered sequence containing each element from a finite set once, and only once. Number of all permutations of a n-element set is equal to n!
- *k-combination without repetition* of a finite set *S* is a subset of *k* distinct elements of *S.* The number of such combinations of an *n*-element set is equal to the binomial coefficient $\binom{n}{k}=\frac{n!}{k!(n-k)!}$
- *k-combination with repetition* of a finite set *S* is a subset of *k* elements (possible repetition of elements) of *S.* The number of such combinations of an *n*-element set is equal to $\binom{n+k-1}{k}=\frac{(n+k-1)!}{k!(n-1)!}$
- *k-variation without repetition* of a finite set *S* is an ordered sequence of *k* distinct elements of *S.* The number of all such sequences of an *n*-element set is equal to $\frac{n!}{(n-k)!}$
- *k-variation with repetition* of a finite set *S* is an ordered sequence of *k* elements of *S* (possible repetition of elements). The number of all such sequences of an *n*-element set is equal to n^k

Example;

There are 7 white and 3 black balls in the box. Take 3 balls. What is the probability of the event A: 2 balls are white and one is black?

Solution: By #X denote a number of elements of X and by #A, a number of elements of A respectively. Then #X = $\binom{10}{3}=120$ (combination without repetition) and #A = $\binom{7}{2}\cdot\binom{3}{1}=63$ (combination without repetition). Hence, $P(A)=\frac{63}{120}=\frac{21}{40}=0{,}525$

Example;
7 students go to 10 different places in Erasmus program. What is the probability of the event A: no 2 students go to the same place?

Solution: By #X denote a number of elements of X and by #A, a number of elements of A respectively. Then #X = 10^7 (variation with repetition) and #A = $\frac{10!}{(10-7)!}$. Hence,

$$P(A) = \frac{\# A}{\# X} = 0{,}06048$$

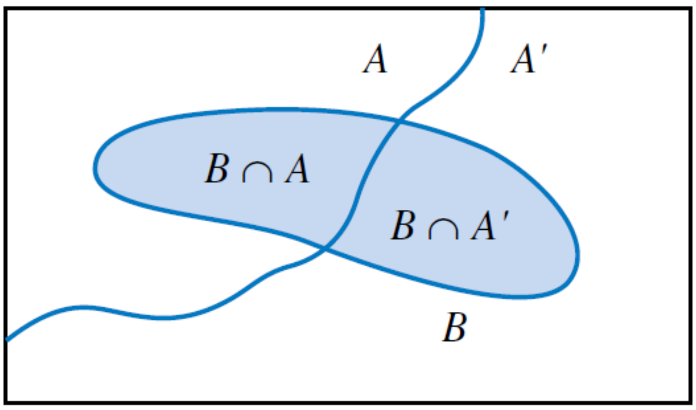
Definition:
The **conditional probability** of an event A given an event B (event B occurs), denoted as P(A/B), is

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ for } P(B) > 0.$$

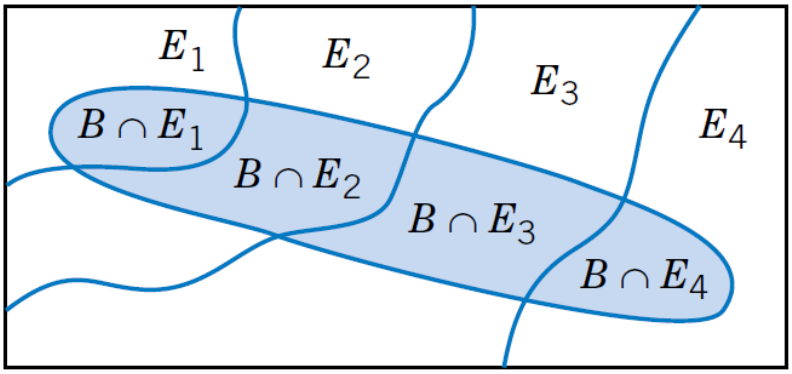
Properties (multiplications rules):

$$P(A \cap B) = P(A/B) \cdot P(B)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_3 / A_1 \cap A_2) \cdot P(A_2 / A_1) \cdot P(A_1)$$

Total probability.



For any events A and B we have $B = (B \cap A) \cup (B \cap A')$. Since A and A' are mutually exclusive, we get: $P(B) = P(B \cap A) + P(B \cap A') = P(B/A) \cdot P(A) + P(B/A') \cdot P(A')$



If $E_1, E_2, \dots, E_n$ is a collection of mutually exclusive events such that $E_1 \cup E_2 \cup \dots \cup E_n = X$, then for any event B we obtain: $P(B) = \sum_{i=1}^{n} P(B/E_i) \cdot P(E_i)$

Bayes' Theorem:

In some examples, we might know one conditional probability but would like to calculate a different one. From the definition of conditional probability we have:

$$P(A \cap B) = P(A/B) \cdot P(B) = P(B \cap A) = P(B/A) \cdot P(A)$$

Now considering the second and last terms in the expression above, we can write

$$P(B/A) = \frac{P(A/B) \cdot P(B)}{P(A)}, \text{ for } P(A) > 0$$

and in general if $E_1, E_2, \dots, E_n$ is a collection of mutually exclusive events such that $E_1 \cup E_2 \cup \dots \cup E_n = X$

$$P(E_j / B) = \frac{P(E_j) \cdot P(B/E_j)}{\sum_{i=1}^{n} P(E_i) \cdot P(B/E_i)}$$

Example;

Students can get to the University, going by metro, train, bus, tram and bicycle:

	metro	train	bus	tram	bicycle
Probability that student will be not late	0,95	0,9	0,8	0,75	0,7
Probability that student uses this kind of transport	0,4	0,2	0,1	0,1	0,2

What is the probability that student will be not late on the University?

Solution: Applying formulas of total probability we define:

- B – event- student is not late,
- A_1- student goes to the University by metro,
- A_2- student goes to the University by train,
- A_3- student goes to the University by bus,
- A_4- student goes to the University by tram,
- A_5- student goes to the University by bicycle.

Hence,

$$P(B) = \sum_{i=1}^{5} P(B/A_i) \cdot P(A_i) = 0{,}95 \cdot 0{,}4 + 0{,}9 \cdot 0{,}2 + 0{,}8 \cdot 0{,}1 + 0{,}75 \cdot 0{,}1 + 0{,}7 \cdot 0{,}2 = 0{,}855$$

Student got to the University and he is not late. What is the probability that student has used a bike?

Solution:

By Bayes' Theorem we have: $P(A_5 / B) = \frac{P(B/A_5) \cdot P(A_5)}{P(B)} = \frac{0{,}7 \cdot 0{,}2}{0{,}855} \approx 0{,}16$

Definition:

The events A, B are *independent* if and only if $P(A \cap B) = P(A) \cdot P(B)$.

In general: the events $E_1, E_2, \dots, E_n$ are *independent* if and only if

$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \dots \cdot P(E_n)$

Random experiment – the outcomes are elements of the sample space S (probability space). We assign a real number X(s) for each outcome $s \in S$.

Definition:

A function $X: s \to X(s)$ is a *random variable* if

- it is a real valued function of S;
- for every $x_0 \in R$ the set $\{s: X(s)<x_0\}$ is an event.

$P(X \in A) = P[\{s: X(s) \in A\}]$ - probability that X takes a value from $A \subseteq R$ – probability of the event $\{s: X(s) \in A\}$

The probability distribution of a random variable X is a description of the probabilities associated with the possible values of X.

Definition:

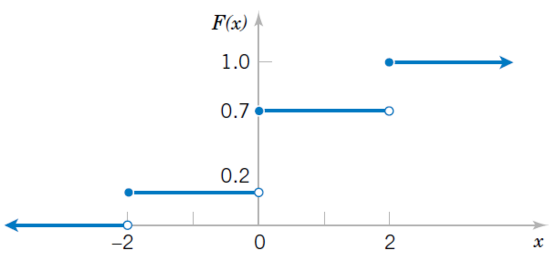
A real function $F_X(x_0) = P(X<x_0)$ is called a *distribution function* of a random variable X

Properties:

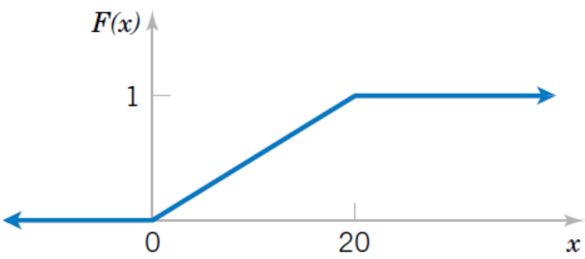
1. $0 \le F_X(x_0) \le 1$
2. $x \le y$ then $F_X(x) \le F_X(y)$
3. $P(a \le x < b) = F(b) – F(a)$

Definition:

X is called a *discrete* random variable if the set of values is finite or countable. If the distribution function of a random variable X is continuous, then X is called a *continuous* random variable



a distribution function of a discrete random variable



a distribution function of a continuous random variable

Example:

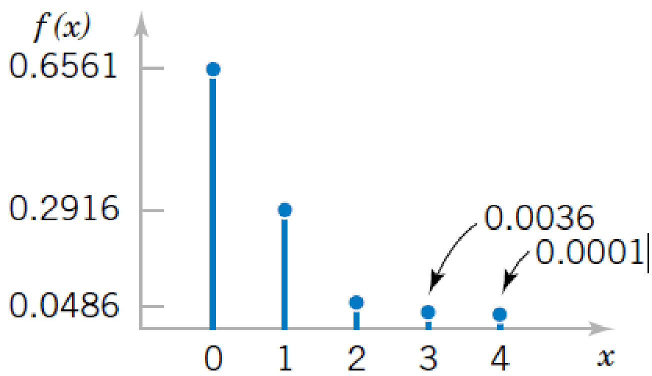
describng the process of rolling a fair dice and the possible outcomes. For the set {1, 2, 3, 4, 5, 6} as the sample space, defining the random variable X as the number rolled.

Definition:

For a discrete random variable we define a *probability mass function* by $f(x) = P(X = x)$.

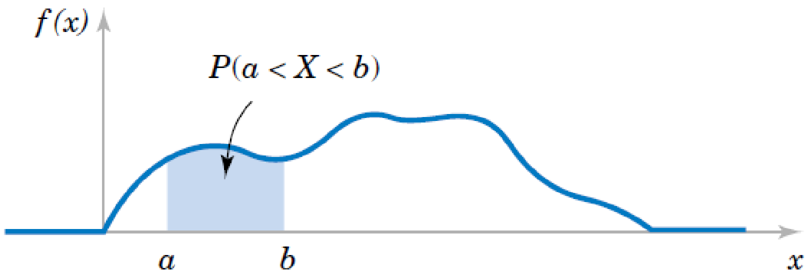
Then:

- $f(x) \ge 0$
- $\sum_x f(x) = 1$



Definition:
For a continuous random variable we define a *probability density function* by $f(x) = \frac{dF_X}{dx}$. Then:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \leq x \leq b) = \int_a^b f(x)dx$
- $F_X(x) = P(X < x) = \int_{-\infty}^{x} f(u)du$



The probability distribution of a random variable is often characterised by a small number of parameters, which also have a practical interpretation. For example, it is often enough to know what its "average value" is. This is captured by the mathematical concept of expected value of a random variable, denoted EX. Once the "average value" is known, one could then ask how far from this average value the values of X typically are, a question that is answered by the variance and standard deviation of a random variable.

Definition:
The *mean* (*expected value*) (measure of central tendency) of a random variable X is defined by

- $EX = \sum_x x \cdot P(X = x)$ if X is a discrete random variable
- $EX = \int_{-\infty}^{\infty} xf(x)dx$ if X is a continuous random variable

Definition:
The *variance* (measure of variability, dispersion) of a random variable X is defined by

- $VarX = \sum_x x^2 \cdot P(X = x) - (EX)^2$ if X is a discrete random variable
- $VarX = \int_{-\infty}^{\infty} x^2 f(x)dx - (EX)^2$ if X is a continuous random variable

$\sigma = \sqrt{VarX}$ is called a *standard deviation* of a random variable X.

Examples of discrete random variable:

- discrete uniform distribution:

x_i	x_1	...	x_n
p_i	$\frac{1}{n}$	...	$\frac{1}{n}$

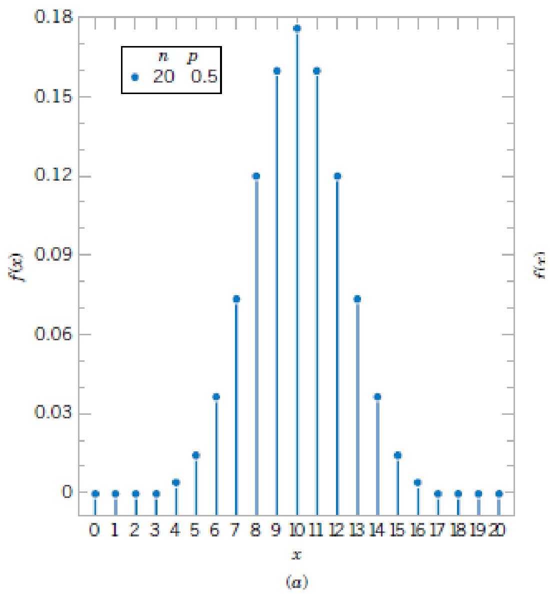
- zero-one distribution (EX = p, VarX = p(1-p)):

x_i	0	1
p_i	1-p	p

Bernoulli trials (binomial distribution)

A trial with only two possible outcomes is used so frequently as a building block of a random experiment.. It is usually assumed that the trials that constitute the random experiment are independent. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the probability of a success in each trial is constant.

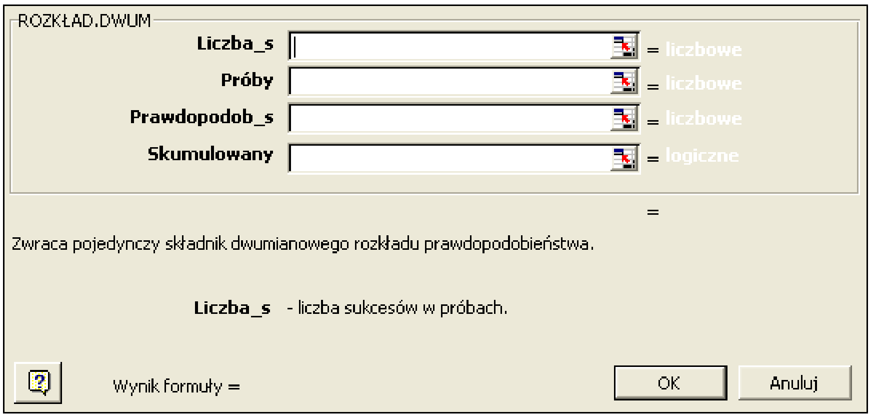
- the trials are independent
- each trial results in only two possible outcomes, labeled as "success'' and "failure''
- the probability of a success in each trial, denoted as p, remains constant



The random variable X that equals the number of trials that result in a success has a binomial random variable with parameters $0 < p < 1$ and $n = 1, 2, 3, \ldots$. The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad EX = np, \; VarX = np(1-p)$$

EXCEL: **ROZKŁAD.DWUM - binomial distribution**
Liczba_s – number of success
Próby – number of trials
Prawdopodob_s – probability of success
Skumulowany – 1: distribution function, 0: mass function



Example;

Each sample of water has a chance 10% of containing a particular organic pollutant. Find the probability that in the next 18 samples

- exactly 2 samples contain the pollutant
- at least 4 samples contain the pollutant

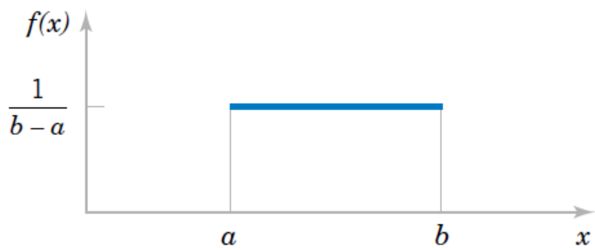
Solution: Denote p = 0,1, n = 18

$$f(x_1) = \binom{18}{2} \cdot 0.1^2 \cdot 0.9^{16}; \; f(x_2) = \sum_{k=4}^{18} \binom{18}{k} \cdot 0.1^k \cdot 0.9^{18-k}$$

A uniform continuous distribution

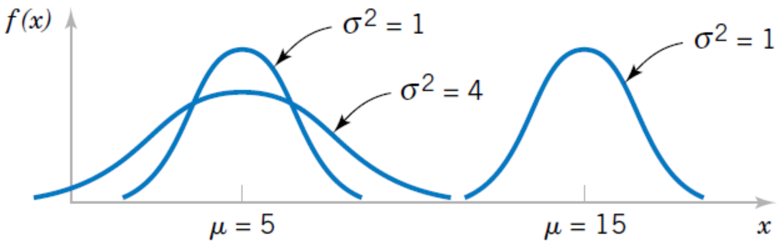
$$f(x) = \frac{1}{b-a} \quad x \in [a,b]$$

$$EX = \frac{a+b}{2}, \quad Var = \frac{(b-a)^2}{12}$$



Normal distribution

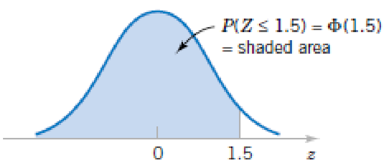
The most widely used model for the distribution of a random variable is a *normal distribution*. Whenever a random experiment is replicated, the random variable that equals the average (or total) result over the replicates tends to have a normal distribution as the number of replicates becomes large.



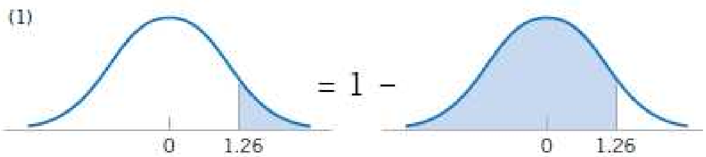
A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

is a *normal random variable* with parameters μ (-∞ < μ < ∞) and σ (σ>0). Also, EX = μ and VarX = σ^2. The notation **N(μ, σ^2)** is used to denote the distribution.



and



If X is a normal random variable , EX = μ and VarX = σ^2, the random variable

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

is a normal random variable with E(Z) = 0 and VarZ = 1. That is, Z is *a standard normal random variable*.

EXCEL: **ROZKŁAD.MORMALNY - normal distribution N(μ, σ^2)**
X - argument
średnia **(μ)** – mean; odchylenie_std **(σ)**– standard deviation
Skumulowany – 1: distribution function, 0: density function