Probability theory

σ-algebra

Definition

Consider a set $\Omega \neq \emptyset$ of elementary random events. This set Ω is called the sample space. Let $\mathcal{A}$ be a nonempty system of subsets of the set Ω such that

a) $\emptyset \in \mathcal{A}$,

b) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$, where A^c is the complement of the set A.

c) if $A_i \in \mathcal{A}$, $i = 1{,}2{,}\ldots$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Then $\mathcal{A}$ is called σ-algebra.

Probability space

Definition

Let $\Omega \neq \emptyset$ and $\mathcal{A}$ be a σ-algebra defined on Ω. Then the probability P is defined as a real function on $\mathcal{A}$, which satisfies

a) $P(\Omega) = 1,\ P(\emptyset) = 0$,

b) $P(A) > 0$ for all $A \in \mathcal{A}$,

c) for all sequences of disjoint events $\{A_i\}_{i=1}^{\infty}$, it holds

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Triple $(\Omega, \mathcal{A}, P)$ is called the probability space.

Terminology

1) $\emptyset$ ... impossible event
2) Ω ... sure event
3) $A \cup B$ ... union of the events A, B (the event, which occurs if and only if at least one from the events A, B occurs)
4) $A \cap B$ ... intersection of the events A, B (the event, which occurs if and only if both the events A and B occur together)
5) $B - A$ ... difference of the events B and A (the event, which occurs if and only if the event B occurs and the event A does not occur)
6) $A \subset B$ ... A is subevent of the event B
7) $A^c = \Omega - A$ ... complement of the event A (the event, which occurs if and only if the event A does not occur)
8) $A \cap B = \emptyset$ ... events A, B are disjoint (they can not occur together)
9) The sequence of disjoint events $\{A_i\}_{i=1}^{\infty}$ such that $\cup_{i=1}^{\infty} A_i = \Omega$ is called a partition of the sample space Ω.

Properties

1) $0 \leq P(A) \leq 1, \quad \forall A \in \mathcal{A},$
2) $A, B \in \mathcal{A}, A \subset B \Rightarrow P(A) \leq P(B),$
3) $P(A^c) = 1 - P(A), \quad \forall A \in \mathcal{A},$
4) $P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad \forall A, B \in \mathcal{A},$
5) $A, B \in \mathcal{A}, A \subset B \Rightarrow P(B - A) = P(B) - P(A),$
6) for all $\{A_i\}_{i=1}^{\infty}$ forming a partition of the sample space Ω, it holds that $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = 1;$

Special probability spaces: classical

Probability space $(\Omega, \mathcal{A}, P)$ is called the classical probability space, if

a) the set Ω is finite and all possible results have the same probability, i.e. denoting $\mathrm{p}_1, \ldots, \mathrm{p}_m$ the probabilities of the individual elementary events, then $\mathrm{p}_1 = \mathrm{p}_2 = \ldots = \mathrm{p}_m = \frac{1}{m}$ (when we have m elementary events),

b) σ-algebra $\mathcal{A}$ is the system of all subsets of the set Ω,

c) probability P of the random event A is equal to

$$P(A) = \frac{m_A}{m},$$

where m_A is the number of results corresponding to the event A and m is the number of all possible results of the random trial.

Special probability spaces: geometrical

Geometrical probability space is the probability space $(\Omega, \mathcal{A}, P)$ such that

a) $\Omega \subset \mathbb{R}^d$ (usually $d = 1, 2, 3$), i.e. the elementary events can be represented by points of an geometrical object,

b) $\mathcal{A} = \mathcal{B}(\Omega)$ is Borel σ-algebra on Ω (i.e. the smallest σ-algebra including all open subsets of Ω, and thus also all closed subsets and their combinations),

c) $P(A) = \frac{\mu^d(A)}{\mu^d(\Omega)}$, where μ^d is d-dimensional Lebesque measure. For our purposes, it is enough to consider $\mu^1(A)$ as the length of A, $\mu^2(A)$ as the area of A and $\mu^3(A)$ the volume of A.

Special probability spaces: general discrete

General discrete probability space is the probability space $(\Omega, \mathcal{A}, P)$ such that

a) $\Omega = \{\omega_1, \omega_2, \ldots\}$,

b) $\mathcal{A}$ is the set of all subsets of Ω,

c) there are given probabilities $P(\omega_i)$ of elementary events ω_i satisfying $\sum_{i=1}^{\infty} P(\omega_i) = 1$. Then the probability of arbitrary event is given by the relation $P(A) = \sum_{\omega_i \in A} P(\omega_i)$.

Special probability spaces: general continuous

General continuous probability space is given by

a) $\Omega = \mathbb{R}$, i.e. all elementary events can be represented by points on real axis,

b) $\mathcal{A} = \mathcal{B}(\mathbb{R})$ is Borel σ-algebra on $\mathbb{R}$,

c) there exists a function f: $\mathbb{R} \to [0, \infty]$ such that $\int_{\mathbb{R}} f(x)dx = 1$ and the probability of arbitrary event $A \in \mathcal{A}$ is uniquely given by

$$P(A) = \int_A f(x)dx.$$

Remark: It is possible to work with more general $\Omega \subset \mathbb{R}^d$, $d = 2, 3, \ldots$, but we do not use it so in this lesson.

Conditional probability

Definition

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Consider random events A and B, where $P(B) > 0$. Probability of the event A conditionally on the event B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Conditional probability is probability

Theorem

Let $(\Omega, \mathcal{A}, P)$ be a probability space and B be a random event, where $P(B) > 0$. Then for an arbitrary event $A \in \mathcal{A}$, it holds:

a) $P(A|B) \geq 0$,

b) $P(\Omega|B) = 1$,

c) $P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$ *for all sequences* $\{A_i\}$ *of disjoint events.*

Remark: Interpretation of this theorem is such that the properties of conditional probability are the same as the properties of unconditional probability.

Proof

a) obvious,

b) it follows from the definition that

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1,$$

c) since A_1, $A_2, \ldots$ are disjoint, then $A_1 \cap B, A_2 \cap B, \ldots$ are disjoint, too. Thus

$$\begin{aligned} P(\cup_{n=1}^{\infty} A_n|B) &= \frac{P(\cup_{n=1}^{\infty} A_n \cap B)}{P(B)} = \frac{\sum_{n=1}^{\infty} P(A_n \cap B)}{P(B)} = \\ &= \sum_{n=1}^{\infty} P(A_n|B) \end{aligned}$$

Chain rule

Theorem

For an arbitrary sequence of random events $A_1, A_2, \dots, A_n$, $P(A_1 \cap A_2 \cap \ \dots \ \cap A_{n-1}) > 0$, *it holds*

$$\begin{aligned} P(\cap_{i=1}^{n} A_i) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \ \cap A_2)\dots \\ &\quad \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \end{aligned} \tag{1}$$

Proof

Repeating definition of conditional probability, we get:

$$\begin{aligned} P(\cap_{i=1}^{n-1} A_i \cap A_n) &= P(\cap_{i=1}^{n-1} A_i) P(A_n | \cap_{i=1}^{n-1} A_i) = \\ &= P(\cap_{i=1}^{n-2} A_i) P(A_{n-1} | \cap_{i=1}^{n-2} A_i) P(A_n | \cap_{i=1}^{n-1} A_i) \ldots \\ &= P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \ldots P(A_n | \cap_{i=1}^{n-1} A_i). \end{aligned}$$

Thanks to monotony of probability, we have

$$P(A_1) \geq P(A_1 \cap A_2) \geq \ldots \geq P(A_1 \cap \ldots \cap A_{n-1}) > 0,$$

thus all conditional probabilities in the theorem are well defined.

Law of total probability

Theorem

Let A_1, A_2, … *be a partition of the sample space* Ω, *i.e.*

$$A_i \cap A_j = \emptyset,\ \forall i \neq j \text{ and } \cup_{i=1}^{\infty} A_i = \Omega.$$

Let these random events have the probabilities $P(A_1), P(A_2), \dots,$ *and* $P(A_i) > 0,\ \forall i = 1, 2, \dots$ *Consider an arbitrary random event* B, *for which we know the conditional probabilities*

$$P(B|A_i),\ \forall i = 1, 2, \dots$$

Then

$$P(B) = \sum_{i=1}^{\infty} P(A_i) \cdot P(B|A_i).$$

Proof

$A_1, \ldots, A_n$ form a partition of the sample space Ω

$$\Rightarrow (A_i \ \cap B) \cap (A_j \cap B) = \emptyset \quad \forall i \neq j, \ \cup_{i=1}^{\infty}(A_i \cap B) = B.$$

Then

$$P(B) = P(\cup_{i=1}^{\infty}(A_i \cap B)) = \sum_{i=1}^{\infty} P(A_i \cap B) = \sum_{i=1}^{\infty} P(A_i) \cdot P(B|A_i).$$

Bayes theorem

Theorem

Let A_1, A_2, ... *be a partition of the sample space* Ω. *Let these random events have the probabilities* $P(A_1), P(A_2), \dots,$ *so that* $P(A_i) > 0,\ \forall i = 1, 2, \dots$ *Consider an arbitrary random event* B, *for which we know the conditional probabilities* $P(B|A_i),\ \forall i = 1, 2, \dots$ *Then*

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(A_j) \cdot P(B|A_j)}, \quad i = 1, 2, \dots$$

Proof

Proof

From definition of conditional probability, we have

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}.$$

From Law of total probability, we get

$$P(A_i|B) = \frac{P(A_i \cap B)}{\sum_{j=1}^{\infty} P(A_j) \cdot P(B|A_j)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(A_j) \cdot P(B|A_j)}.$$

Independent events

Definition

Random events A and B are called independent, if it holds

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

Definition

Let $A_1, A_2, \ldots, A_n$ be random events. We call them to be multiple independent, if for an arbitrary sequence of indexes $\{k_1, k_2, \ldots, k_r\} \subset \{1, \ldots, n\},\ r = 2, \ldots, n$, it holds

$$P(A_{k_1} \cap A_{k_2} \cap \ldots \cap A_{k_r}) = P(A_{k_1}) \cdot P(A_{k_2}) \cdot \ldots \cdot P(A_{k_n}).$$

Definition

Let $A_1, \ldots, A_n$ be random events. We call them to be pairwise independent, if the events A_i, A_j are independent for all $i, j = 1, \ldots, n,\ i \neq j$.

Independency of complements

Theorem

Let A, B be independent random events. Then the pairs of events (A, B^c), (A^c, B), (A^c, B^c) are independent.

Proof

$$\begin{aligned} P(A^c \cap B) &= P(B - A) = P(B - [A \cap B]) = P(B) - P(A \cap B) = \\ &= P(B) - P(B) \cdot P(A) = P(B) \cdot (1 - P(A)) = \\ &= P(B) \cdot P(A^c). \end{aligned}$$

Proof of independency of the events A, B^c is analogous.
If the events A, B are independent, then also the events A, B^c are independent, and so alsoA^c, B^c are independent.

Definition of random variable

Definition

Let $(\Omega, \mathcal{A}, P)$ be a probability space. The real function X defined on Ω is called the random variable, if X is measurable mapping
$X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$, i.e.

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$$

for an arbitrary Borel set $B \in \mathcal{B}$.

Notation and properties

Notation:

1. Random variables are denoted by capitals, i.e. $X, Y, Z \ldots$
2. Their values are denoted by small letters $x, y, z \ldots$
3. Instead of $\{\omega \in \Omega : X(\omega) \in B\}$ we write $\{X \in B\}$, especially instead of $\{\omega \in \Omega : X(\omega) \leq x\}$ we write $\{X \leq x\}$.

Properties: Sum, product, ratio, minimum, maximum etc. of random variables are again random variables.

Distribution function

Definition

Let X be a random variable. Its distribution function is a real function F defined as

$$F(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}), \quad x \in \mathbb{R}.$$

Basic properties of the distribution function:

The distribution function $F(x)$ of a random variable X is

1. nondecreasing, i.e. for arbitrary $a, b \in \mathbb{R}, a \leq b$, it holds that $F(a) \leq F(b)$,
2. right continuous in an arbitrary point $x \in \mathbb{R}$,
3. $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1.$

Discrete random variable

Definition

The random variable X is called *discrete* (or we say that it has a discrete distribution), if there exists a finite or countably infinite sequence of real numbers $\{x_n\}$ with corresponding sequence of non-negative numbers $\{p_n\} = P(X = x_n)$ such that $\sum_{n=1}^{\infty} p_n = 1$.

Distribution function of the discrete random variables X is of the form

$$F(x) = P(X \leq x) = \sum_{\{n: x_n \leq x\}} P(X = x_n) = \sum_{\{n: x_n \leq x\}} p_n$$

and it holds that

$$P(a < X \leq b) = F(b) - F(a) = \sum_{\{n: a < x_n \leq b\}} P(X = x_n) = \sum_{\{n: a < x_n \leq b\}} p_n$$

for arbitrary real numbers a, b, where $a \leq b$.

Absolutly continuous random variable

Definition

The random variable X is called *absolutely continuous* (or we say that it has an absolutely continuous distribution), if there exists a non-negative inegrable function f such that

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad x \in (-\infty, \infty).$$

The function f is called *probability density*.

Basic properties of the density f:

1. $f(x) = \frac{d}{dx}F(x)$ a.s.,
2. $\int_{-\infty}^{\infty} f(x)dx = 1$,
3. $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$
   for arbitrary real numbers a, b, where $a \leq b$.

Generalisation

Definition

A measure is defined as a set function on $(\Omega, \mathcal{A})$, i.e.

1. $\mu : \mathcal{A} \to [0, \infty]$,
2. $\mu(\emptyset) = 0$,
3. if $A_n \in \mathcal{A}, n \geq 1$ are disjoint, then $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

If $\mu(\Omega) = 1$, we call it *probability measure*.

Definition

Each random variable X and Borel set $B \in \mathcal{B}$ may be connected with a probability measure on $(\mathbb{R}, \mathcal{B})$,

$$\mu_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}),$$

which is called *probability distribution* of the random variables X.

Generalisation

- Denoting $B = (-\infty, x]$, we get

$$\mu_X(B) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = F(x),$$

i.e. the distribution function.

- Denoting $B = (a, b]; \; -\infty < a \leq b < \infty$, we get

$$P(X \in (a, b]) = F(b) - F(a) = \mu_X((a, b]).$$

- Thus for all Borel sets, it holds that

$$P(X \in B) = \mu_X(B) = \int_B 1 d\mu_X(x) = \int_B 1 dF(x), \quad \forall B \in \mathcal{B}.$$

Expected value

Definition

Let X be a random variable defined on a probability space $(\Omega, \mathcal{A}, P)$. *Expected value* $\mathbb{E}X$ of the random variable X is

$$\mathbb{E}X = \int_{-\infty}^{\infty} x dF(x),$$

if the integral exists.

Expected value of discrete and continuous random variables

- Let X be a discrete random variable with the values x_1, x_2, x_3,... Then its expected value $\mathbb{E}X$ is of the form

$$\mathbb{E}X = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i),$$

  if the sequence converges.

- Let X be an absolutely continuous random variable with the density f. Then its expected value is of the form

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

  if the integral exists.

Basic properties of the expected value

1. $\mathbb{E}a = a$,
2. $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$,
3. $X_1 \leq X \leq X_2$ *a.s.* $\Rightarrow \mathbb{E}X_1 \leq \mathbb{E}X \leq \mathbb{E}X_2$,
4. $X \geq 0$ *a.s.* $\Rightarrow \mathbb{E}X \geq 0$

Expected value of a function of random variable

Theorem

Let X be a random variable defined on probability space $(\Omega, \mathcal{A}, P)$ and let $\phi : \mathbb{R} \to \mathbb{R}$. Then

$$\mathbb{E}\phi(X) = \int_{-\infty}^{\infty} \phi(x) dF_X(x),$$

if the integral exists.

Expected value of a function of random variable

1. For a random variable X having the discrete distribution with the values x_1, x_2, $x_3, \ldots$, it holds that

$$\mathbb{E}\phi(X) = \sum_{i=1}^{\infty} \phi(x_i) \cdot P(X = x_i),$$

if both sides of the equation exist.

2. For a random variable X having the absolutely continuous distribution with the density f, it holds that

$$\mathbb{E}\phi(X) = \int_{-\infty}^{\infty} \phi(x) f(x) dx,$$

if both sides of the equation exist.

Moments, variance and covariance

Definition

Let X be a random variable.
$\mathbb{E}X^n$ is called the n-th moment of the random variable X,
$\mathbb{E}(X - \mathbb{E}X)^n$ is called the n-th central moment of the random variable X,
$\mathbb{E}|X - \mathbb{E}X|$ is called the absolute moment of the random variable X.

Definition

The second central moment is called the variance and it is denoted by $varX = \mathbb{E}(X - \mathbb{E}X)^2$.

Definition

Let X, Y be the random variables such that $\mathbb{E}X^2 < \infty$ and $\mathbb{E}Y^2 < \infty$. Then their covariance is defined as

$$cov(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Remark: $cov(X, X) = var(X)$.

Properties of the variance and covariance

1. Let X be a random variable. Then $varX = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
2. Let c be a constant. Then $var\ c = 0$.
3. Let X be a random variable and a be a real number. Then $var(aX) = a^2 varX$.
4. Let X be a random variable and c be a constant. Then $var(X + c) = varX$.
5. Let X be a random variable with finite expected value and finite non-zero variance. Let
$$Z = \frac{X - \mathbb{E}X}{\sqrt{varX}}.$$
Then $\mathbb{E}Z = 0$ and $varZ = 1$.
6. For random variables X, Y it holds that $var(X + Y) = varX + varY + 2cov(X, Y)$.
7. For random variables X, Y it holds that $cov(X, Y) = \mathbb{E}(XY)$ - $\mathbb{E}X\mathbb{E}Y$.

Chebyshev inequality

Theorem

Let X be a random variable with finite variance. Then for an arbitrary $\varepsilon > 0$, it holds that

$$P[|X - \mathbb{E}X| \geq \varepsilon] \leq \frac{varX}{\varepsilon^2}.$$

Proof

Consider a random variable $Y = X - \mathbb{E}X$ with the distribution function F. Then

$$\begin{aligned} varX &= \mathbb{E}Y^2 = & \int_{-\infty}^{\infty} x^2 dF(x) \geq \int_{|x|\geq\varepsilon} x^2 dF(x) \geq \\ &\geq & \varepsilon^2 \int_{|x|\geq\varepsilon} dF(x) = \varepsilon^2 P[|Y| \geq \varepsilon]. \end{aligned}$$

Alternative distribution ($X \sim Alt(p)$)

- X takes only the values 0 and 1 subsequently with probabilities $1 - p$ and p.
- The number p is called the parameter of the alternative distribution, $0 < p < 1$.
- The distribution function is of the form

$$F(x) = \begin{cases} 0 & \text{for } x \; < 0 \\ 1 - p & \text{for } 0 \leq x \; < 1 \\ 1 & \text{for } x \; \geq 1 \end{cases}$$

- The expected value $\mathbb{E}X = p$ and the variance $varX = p(1 - p)$.

Binomial distribution ($X \sim Binom(n,p)$)

- X takes the values $k = 0, 1, 2, \dots, n$.
- It is uniquely given by two parameters $n \in \mathbb{N}$ and $p \in (0,1)$.
- Probabilities $P(X = k)$ are of the form

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0, 1, \dots, n.$$

- The distribution function is

$$F(x) = \begin{cases} 0 & x < 0 \\ \sum_{0 \le k \le x} \binom{n}{k} p^k (1-p)^{n-k} & 0 \le x < n \\ 1 & x \ge n. \end{cases}$$

- The expected value $\mathbb{E}X = np$ and the variance $varX = np(1-p)$.

Calculation of the expected value

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{n} k\tbinom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^{k} (1-p)^{n-k-1} \\
&= np(p+(1-p))^{n-1} = np.
\end{aligned}$$

Calculation of the variance
For the calculation of the variance we use the relation

$$varX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X(X-1) + \mathbb{E}X - (\mathbb{E}X)^2.$$

The calculation of the first term

$$\mathbb{E}X(X-1) = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} = ... = n(n-1)p^2$$

is analogous to that one for expected value. In this way, we obtain the variance

$$varX = np(1-p).$$

.

Poisson distribution ($X \sim Po(\lambda)$)

- X takes the values $k = 0, 1, 2, \dots$
- It is uniquely given by the parameter $\lambda > 0$.
- Probabilities $P(X = k)$ are of the form

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \text{ for } k = 0, 1, \dots$$

- The distribution function is

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \sum_{0 \leq j \leq x} e^{-\lambda}\frac{\lambda^j}{j!} & \text{for } 0 \leq x < \infty. \end{cases}$$

- The expected value and the variance are $\mathbb{E}X = varX = \lambda$ (the calculation is analogous to the previous one).

Relation between binomial and Poisson distributions
Consider the random variable $X \sim Binom(n, p)$, where $n \to \infty$, $p \to 0$, while $np = \lambda$. Then

$$P(X = k) = \frac{n(n-1)\dots(n-k+1)}{k!} p^k (1 - \frac{\lambda}{n})^{n-k} \xrightarrow[n\to\infty, p\to 0]{} \frac{\lambda^k}{k!} e^{-\lambda},$$

so we obtain Poisson distribution.

Geometrical distribution ($X \sim Ge(p)$)

- X takes the values $k = 0, 1, 2, \dots$
- It is uniquely given by the parameter $p \in (0, 1)$.
- Probabilities $P(X = k)$ are of the form

$$P(X = k) = p(1 - p)^k \text{ for } k = 0, 1, \dots$$

- The distribution function is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \sum_{0 \le k \le x} p(1-p)^k & \text{for } x \ge 0. \end{cases}$$

- Using relations for geometrical sequences, we obtain the expected value $\mathbb{E}X = \frac{1-p}{p}$ and the variance $varX = \frac{1-p}{p^2}$.

Uniform distribution in $[a, b]$ $(X \sim Ro(a, b))$

- X takes the values from the interval $[a, b]$ ($a, b \in \mathbb{R}$ are the parameters).
- It is given by the density

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & x < a, \quad x > b. \end{cases}$$

- The distribution function is

$$F(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x \geq b. \end{cases}$$

- The expected value and the variance are

$$\mathbb{E}X = \frac{a+b}{2}, varX = \frac{1}{12}(b-a)^2.$$

Exponential distribution ($X \sim Exp(\lambda)$)

- X takes the values from the interval $(0, \infty)$.
- It is given by the density with parameter λ:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- The distribution function is

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-\lambda x} & x > 0. \end{cases}$$

- Using per partes, we obtain the expected value

$$\mathbb{E}X = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Further,

$$\mathbb{E}X^2 = \int_0^\infty x^2\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

so the variance is

$$varX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{1}{\lambda^2}.$$

Exponential distribution ($X \sim Exp(\lambda)$)

Properties of the exponential distribution:

1. *It has no memory:*
   For the random variable X with exponential distribution, it holds that

   $$P(X > x + y | X > y) = P(X > x) \quad \forall x > 0, y > 0,$$

   since $P(X > x + y | X > y)$ can be rewritten (using the definition of conditional probability) to the form

   $$\frac{P(X > x + y)}{P(X > y)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x}.$$

2. *Connection with Poisson distribution:*
   If the random variable X describing the time of waiting for an event has the distribution $Exp(\lambda)$, then the random variable Y describing the number of that events in the time interval of the length T has the distribution $Po(\lambda T)$.

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

- X takes the values from $\mathbb{R}$.
- It is uniquely determined by the parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
- It is given by the density

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

- The distribution function is

$$F(x) = \frac{1}{\sqrt{2\pi\ \sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt, \quad -\infty < x < \infty.$$

- The expected value is $\mathbb{E}X = \mu$ and the variance is $varX = \sigma^2$.

Standard normal distribution ($X \sim N(0,1)$)

- X again takes the values from $\mathbb{R}$.
- It is given by the density

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty.$$

- The distribution function is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt, \; -\infty < x < \infty.$$

- The expected value is $\mathbb{E}X = 0$ and the variance is $varX = 1$.
- The values of Φ can be found in statistical tables.
- Thanks to the symetry of the function $\Phi(x) = 1 - \Phi(-x)$, the values of Φ are often tabulated only for non-negative x.

Transformation of the variables with normal distribution

Theorem

1. *If X has standard normal distribution and $Y = \mu + \sigma X$, then Y has normal distribution with the parameters μ and σ^2.*
2. *If X has normal distribution with parameters μ, σ^2 and if $Y = a + bX$, then Y has again normal distribution with parameters $a + b\mu$ and $b^2\sigma^2$.*
3. *Let X, Y be random variables, $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ and $cov(X, Y) = 0$. Then $Z = X + Y$ has the distribution $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Recall that...

... sum, product, ratio, minimum, maximum etc. of random variables are again random variables.

Further, if X is a random variable, then

$$Y = \varphi(X)$$

is also a random variable for any $\varphi : \mathbb{R} \to \mathbb{R}$.

Function of a random variable

Theorem

Let X be a random variable with distribution function F and let $\varphi : \mathbb{R} \to \mathbb{R}$. Denote $Y = \varphi(X)$ and G its distribution function. Then

$$G(y) = \int_{\{x;\varphi(x)\le y\}} dF(x), \quad \forall y \in \mathbb{R}.$$

Especially, if F is discrete $\{x_n, p_n\}$, then

$$G(y) = \sum_{\{x_n;\varphi(x_n)\le y\}} p_n, \quad \forall y \in \mathbb{R}$$

and if it is absolutely continuous with the density f, then

$$G(y) = \int_{\{x;\varphi(x)\le y\}} f(x)\ dx, \quad \forall y \in \mathbb{R}.$$

Proof

Proof

Denote $B_y = \{x; \varphi(x) \leq y\}$. Then

$$G(y) = P(Y \leq y) = P(\varphi(X) \leq y) = P(X \in B_y) = \int_{\{x;\varphi(x)\leq y\}} dF(x).$$

Sum of random variables - convolution

Consider two independent (mathematical definition - see below) random variables X and Y with distribution functions $F(x)$ and $G(y)$, respectively. The aim is to obtain the distribution function of $Z = X + Y$. Let $H(z)$ be the distribution function of the random variable Z. Then

$$\begin{aligned} H(z) &= \int\int_{x+y\leq z} dF(x)dG(y) = \int_{-\infty}^{\infty} F(z-y)dG(y) = \\ &= \int_{-\infty}^{\infty} G(z-x)dF(x). \end{aligned}$$

Definition

The probability distribution given by the distribution function $H(z)$ is called *the convolution of the distributions with distribution functions* $F(x)$ and $G(y)$ and H is called *the convolution of distribution functions* F and G.

Convolution is denoted as $H = F * G$.

convolution of discrete random variables

Theorem

Let F, G be discrete distribution functions with corresponding probabilities $\{p_n\}, \{q_n\}$, i.e.

$$F(x) = \sum_{0 \leq n \leq x} p_n, \quad G(y) = \sum_{0 \leq n \leq y} q_n.$$

Let $H = F * G$. Then H is discrete distribution function given by

$$H(z) = \sum_{0 \leq n \leq z} h_n, \text{ where } h_n = \sum_{k=0}^{n} p_k \, q_{n-k}.$$

convolution of continuous random variables

Theorem

Let X and Y be independent random variables with absolutely continuous distribution functions F(x) and G(y), respectively, and corresponding densities $f(x)$ a $g(y)$, respectively. Then $H = F * G$ is absolutely continuous and for its density $h(z)$ (i.e. for the density of the random variable $Z = X + Y$) it holds that

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_{-\infty}^{\infty} f(z-y)g(y)dy. \tag{2}$$

Remark

The function $h(z)$ defined by (2) is called the convolution of the densities $f(x)$ and $g(y)$ and denoted as $h = f * g$. In order to verify the property of the density, we know from (2) that $h(z) \geq 0$ and

$$\begin{aligned} \textstyle\int_{-\infty}^{\infty} h(z)dz &= \textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-y)g(y)dydx = \\ &= \textstyle\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x-y)dx\right) g(y)dy = \int_{-\infty}^{\infty} g(y)dy = 1. \end{aligned}$$

Basic convolutions

- **Convolution of two binomial distributions**
  Let X and Y be independent random variables, $X \sim Binom(n_1, p)$ and $Y \sim Binom(n_2, p)$. Then the distribution of the random variable $Z = X + Y$ is $Binom(n_1 + n_2, p)$.
- **Convolution of two Poisson distributions**
  Let $X \sim Po(\lambda_1)$ and $Y \sim Po(\lambda_2)$ be independent. Then the distribution of the random variable $Z = X + Y$ is $Po(\lambda_1 + \lambda_2)$.

Basic convolutions

- **Convolution of two uniform distributions**

Let

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

and

$$g(y) = \begin{cases} \frac{1}{d-c} & \text{for } c \le y \le d \\ 0 & \text{otherwise.} \end{cases}$$

For $d - c \ge b - a$ it holds that

$$h(z) = \begin{cases} 0 & \text{for } z \le a + c \text{ or } b + d \le z \\ \frac{z-(a+c)}{(b-a)(d-c)} & \text{for } a + c \le z \le b + c \\ \frac{1}{d-c} & \text{for } b + c \le z \le a + d \\ \frac{(b+d)-z}{(b-a)(d-c)} & \text{for} a + d \le z \le b + d. \end{cases}$$

Basic convolutions

- **Convolution of two normal distributions**
  Let X, Y be independent random variables, $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. Then the distribution of $Z = X + Y$ is $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
- **Convolution of two exponential distributions**
  If X, Y are independent random variables with the same exponential distributions with parameter $\lambda > 0$, then the density of the random variable $Z = X + Y$ is

$$h(z) = \begin{cases} \lambda^2 z \exp\{-z\lambda\} & z > 0, \\ \\ 0 & z \leq 0. \end{cases}$$

Random vector

Definition

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Consider random variables X_1, X_2, …, X_n defined on this space. Then the vector $\mathbb{X} = (X_1, \ldots, X_n)^T$ is called *random vector*.

Remark:

Random vector is thus a mapping from Ω to $\mathbb{R}^n$. The values of the random vector may be interpreted as points in the n-dimensional space.

Distribution function of random vector

Definition

Let $\mathbb{X} = (X_1, \dots, X_n)^T$ be a random vector defined on a probability space $(\Omega, \mathcal{A}, P)$. (Joint) distribution function $F_{\mathbb{X}}$ of the random vector $\mathbb{X}$ is the real function of n variables defined on $\mathbb{R}^n$ as

$$\begin{aligned} F_{\mathbb{X}}(x_1, \dots, x_n) &= P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n) = \\ &= P(\cap_{i=1}^{n} \{\omega : X_i(\omega) \le x_i\}), \\ & \quad -\infty < x_i < \infty, \ i = 1, \dots, n. \end{aligned}$$

1. $F_{\mathbb{X}}(x_1, \ldots, x_n)$ is nondecreasing function in each variable while the values of the remaining values are fixed.
2. $F_{\mathbb{X}}(x_1, \ldots, x_n)$ is right continuous in each variable.
3. $\lim_{x_i \to -\infty} F_{\mathbb{X}}(x_1, \ldots, x_n) = 0, \ i = 1, \ldots, n,$
   where the remaining values x_j $(j = 1, \ldots, n, \, j \neq i)$ are fixed.
4. $\lim_{x_1, \ldots, x_n \to \infty} F_{\mathbb{X}}(x_1, \ldots, x_n) = 1.$

Discrete random vector

Definition

The random vector $\mathbb{X}$ has *discrete distribution*, if there exists a sequence $\{\mathbf{x}_k\}_{k=1}^{\infty}$, $\mathbf{x}_k \in \mathbb{R}^n$, and corresponding sequence $\{p_k\}_{k=1}^{\infty}$ of positive numbers such that

$$\sum_{k=1}^{\infty} p_k = 1, \qquad \text{where } p_k = P(\mathbb{X} = \mathbf{x}_k) = P(\{\omega \in \Omega : \mathbb{X}(\omega) = \mathbf{x}_k\}).$$

Distribution function of the discrete random vector $\mathbb{X}$ is of the form

$$F_{\mathbb{X}}(\mathbf{x}) = \sum_{\{k : \mathbf{x}_k \leq \mathbf{x}\}} p_k, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

where $\mathbf{x}_k \leq \mathbf{x}$ is considered in each variable, i.e. $x_k^i \leq x^i$ for all $i = 1, \ldots, n$.

Continuous random vector

Definition

The random vector $\mathbb{X} = (X_1, \ldots, X_n)^T$ has *absolutely continuous distribution*, if there exists a non-negative function $f_{\mathbb{X}}$ of n real variables such that

$$F_{\mathbb{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\mathbb{X}}(t_1, \ldots, t_n) dt_1, \ldots, dt_n,$$

where the function $f_{\mathbb{X}}$ is called the probability density of the random vector $\mathbb{X}$ or joint density of the random variables X_1, ..., X_n.

Generalisation

Remark

As in the case of random variables, we can generalise the random vector using probability measure. However, it is enough for our purposes to consider the random vectors of discrete and continuous type, respectively.

Marginal distribution

Definition

Distribution (distribution function, probability density, respectively) of a random vector $(X_1, \ldots, X_k)^T$, which is subvector of the random vector $\mathbb{X} = (X_1, \ldots, X_n)^T$, is called marginal distribution (distribution function, probability density, respectively).

Calculation of marginal distribution: discrete case

If the random vector $\mathbb{X} = (X_1, \ldots, X_n)^T$ has discrete distribution with joint probabilities $P(X_1 = .,\ \ldots,\ X_{i-1} = .,\ X_i = .,\ X_{i+1} = .,\ \ldots,\ X_n = .)$, where the random variables X_l have values $x_{l,1}, \ldots, x_{l,k_l}$ for $l = 1, \ldots, n$, then the marginal probabilities are

$$P(X_i = x) = \sum_{j_1=1}^{k_1} \ldots \sum_{j_{i-1}=1}^{k_{i-1}} \sum_{j_{i+1}=1}^{k_{i+1}} \ldots \sum_{j_n=1}^{k_n} P(X_1 = x_{1,j_1},\ \ldots,\ X_{i-1} = x_{i-1,j_{i-1}},$$
$$X_i = x,\ X_{i+1} = x_{i+1,j_{i+1}},\ \ldots,\ X_n = x_{n,j_n}).$$

Calculation of marginal distribution: continuous case

If the random vector $\mathbb{X} = (X_1, \dots, X_n)^T$ is continuous with joint density $f_{\mathbb{X}}$, then marginal density of the random variable X_i is $(n-1)$-dimensional integral

$$f_{X_i}(x) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f_{\mathbb{X}}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) dx_1, \dots, dx_{i-1} dx_{i+1}, \dots, dx_n.$$

Characteristics of random vector

Consider the random vector $\mathbb{X} = (X_1, \ldots, X_n)^T$.

1. Vector of expected values

$$\mathbb{E}\mathbb{X} = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n)^T.$$

2. Variance matrix $var\mathbb{X}$ with elements

$$cov(X_i, X_j) = \mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j), \quad 1 \leq i, j \leq n.$$

3. Correlation matrix $corr\mathbb{X}$ with elements

$$corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{varX_i}\sqrt{varX_j}}, \quad 1 \leq i, j \leq n.$$

Remark

For the correlation, it holds that

$$-1 \leq corr(X, Y) \leq 1.$$

Independent random variables

Definition

The random variables X_1, X_2 ..., X_n are called mutually independent if

$$P(\cap_{j=1}^{r}\{\omega : X_{i_j}(\omega) < x_{i_j}\}) = \Pi_{j=1}^{r} P(\{\omega : X_{i_j}(\omega) < x_{i_j}\})$$

$$\forall\{i_1, i_2, \ldots, i_r\} \subset \{1, 2, \ldots, n\}, 1 \leq r \leq n, \forall x_{i_j} \in \mathbb{R}.$$

Verifying of independency of random variables

Theorem

1. *Let* $\mathbb{X} = (X_1, X_2 \ldots, X_n)^T$ *be a discrete random vector. The random variables* $X_1, X_2 \ldots, X_n$ *are mutually independent if and only if it holds that*

$$P(X_1 = x_1^{(i)}, \ldots, X_n = x_n^{(i)}) = \Pi_{j=1}^{n} P(X_j = x_j^{(i)})$$

for all $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}),\ i = 1, 2, \ldots,$ *which* $\mathbb{X}$ *can take.*

2. *Let* $\mathbb{X} = (X_1, X_2 \ldots, X_n)^T$ *be a continuous random vector. The random variables* $X_1, X_2 \ldots, X_n$ *are mutually independent if and only if it holds that*

$$f_{\mathbb{X}}(x_1, x_2 \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \ldots f_{X_n}(x_n),\ \forall (x_1, x_2 \ldots, x_n) \in \mathbb{R}^n.$$

Independency and covariance

Theorem

Let X and Y be independent random variables with finite expected values, then

1. $\mathrm{E}XY = (\mathrm{E}X)(\mathrm{E}Y)$.
2. *Moreover, if* $\mathrm{E}X^2 < \infty$ *and* $\mathrm{E}Y^2 < \infty$, *then* $cov(X, Y) = 0$.

Remark

If $cov(X, Y) = 0$, then we say that the random variables are non-correlated. However, it does not imply the independency!

Convergence of random variables

Definition

Consider a sequence of random variables X$_1$, X$_2$, X$_3$, ... and a random variable X. Let these random variables be defined on the same probability space $(\Omega, \mathcal{A}, P)$.

We say that X$_n$ converges to X *almost surely*, if

$$P\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1.$$

If for all $\varepsilon > 0$ it holds that

$$\lim_{n\to\infty} P\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} = 0,$$

then we say that X$_n$ converges to X *in probability*.

Theorem

Convergence almost surely $\Rightarrow$ *convergence in probability.*

Laws of large numbers

Theorem

Weak law of large numbers:
Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random variables with the same expected values μ and the same variances $\sigma^2 < \infty$. Then for $n \to \infty$ it holds that

$$\frac{1}{n}(X_1 + X_2 + \ldots + X_n) \to \mu$$

in probability.

Theorem

Strong law of large numbers:
Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent identically distributed random variables with finite expected value $\mathrm{E}X_1 = \mu$. Then for $n \to \infty$ it holds that

$$\frac{1}{n}(X_1 + X_2 + \ldots + X_n) \to \mu$$

both in probability and almost surely.

Central limit theorem

Theorem

Let X_1, X_2, … be independent identically distributed random variables with expected value μ and finite variance σ^2. Denote

$$Z_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} \quad n = 1, 2, \ldots$$

and $F_n(x)$ the distribution function of Z_n. Then $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for all $-\infty < x < \infty$, where $\Phi(x)$ is the distribution function of $N(0,1)$.

Remark

Central limit theorem (CLT) has many versions. The introduced one is called Lévy-Lindeberg CLT.