Random variables I.

Lecturer:
Francesco Dolce

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

Probability and Statistics

BIE-PST, WS 2024/25, Lecture 3

Content

- **Probability theory:**
  - Events, probability, conditional probability, Bayes' Theorem, independence of events.
  - Random variables, distribution function, functions of random variables, characteristics of random variables: expected value, variance, moments, generating function, quantiles, critical values, covariance and correlation, important discrete and continuous distributions.
  - Random vectors, joint and marginal distributions, functions of random vectors, independence of random variables, conditional distribution, conditional expected value, covariance and correlation.
  - Markov's and Chebyshev's inequality, weak law of large numbers, strong law of large numbers, Central limit theorem.
- **Mathematical statistics:**
  - Point estimators, sample mean, sample variance, properties of point estimators, Maximum likelihood method.
  - Interval estimators, hypothesis testing, one-sided vs. two-sided alternatives, linear regression, estimators of regression parameters, testing of linear model.

Recap

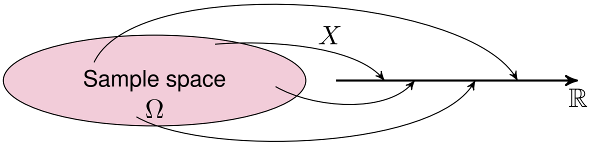
A random experiment is represented using a probability space $(\Omega, \mathcal{F}, \mathrm{P})$:

- Ω is the set of possible outcomes ω.
- $\mathcal{F}$ is a system of subsets of Ω with desirable properties.
- Elements $A \in \mathcal{F}$ are called random events.
- Probability measure P is a function, which assigns real values from 0 to 1 to the random events. It represents the ideal proportion of cases, in which the events occur.

Random variable

For a mathematical processing of a random experiment it is often useful to assign a number to each outcome ω. By this assignment we choose the part of information which is interesting from our point of view.

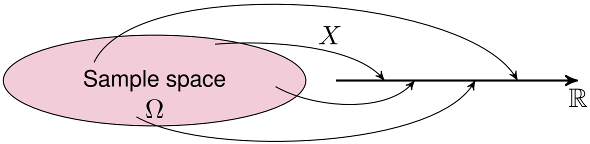
Such assignment can be established in many ways and will be called a **random variable**.



Random variable

For a mathematical processing of a random experiment it is often useful to assign a number to each outcome ω. By this assignment we choose the part of information which is interesting from our point of view.

Such assignment can be established in many ways and will be called a **random variable**.



Examples

- Number of Heads while tossing a coin: $X(\text{Heads}) = 1, X(\text{Tails}) = 0$.
- Number of winnings in the game with: $X(\text{Heads}) = 1, X(\text{Tails}) = -1$.
- How much a player won in a given game at a poker tournament.
- The highest rolled value or n rolls of a die.
- The height of a randomly chosen person.

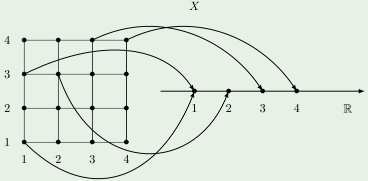
Example – minimum of two rolls of a 4-sided die

Two rolls of a 4-sided die. $\Omega = \{1, 2, 3, 4\}^2$.

Example – minimum of two rolls of a 4-sided die

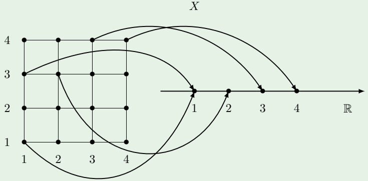
Two rolls of a 4-sided die. $\Omega = \{1, 2, 3, 4\}^2$.

Consider a random variable $X(\omega) = \min\{\omega(1), \omega(2)\}$:



Example – minimum of two rolls of a 4-sided die

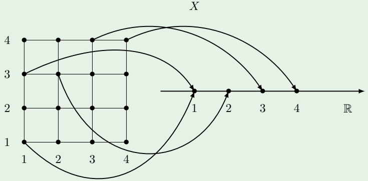
Two rolls of a 4-sided die. $\Omega = \{1,2,3,4\}^2$.
Consider a random variable $X(\omega) = \min\{\omega(1), \omega(2)\}$:



$$\mathrm{P}(X=1) = \mathrm{P}(\{\omega | X(\omega) = 1\})$$
$$= \mathrm{P}(\{(1,1),(1,2),(1,3),(1,4),(2,1),(3,1),(4,1)\}) = \frac{7}{16}.$$

Example – minimum of two rolls of a 4-sided die

Two rolls of a 4-sided die. $\Omega = \{1, 2, 3, 4\}^2$.
Consider a random variable $X(\omega) = \min\{\omega(1), \omega(2)\}$:



$$P(X = 1) = P(\{\omega | X(\omega) = 1\})$$
$$= P(\{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (3, 1), (4, 1)\}) = \frac{7}{16}.$$

Similarly,

$$P(X = 2) = P(\{(2, 2), (2, 3), (2, 4), (3, 2), (4, 2)\}) = \frac{5}{16},$$
$$P(X = 3) = P(\{(3, 3), (3, 4), (4, 3)\}) = \frac{3}{16},$$
$$P(X = 4) = P(\{(4, 4)\}) = \frac{1}{16}.$$

Random variable and its distribution function

Definition

A **random variable** X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a function $X : \Omega \to \mathbb{R}$, assigning to each outcome $\omega \in \Omega$ a number $X(\omega)$, with the property that:

$$\{X \leq x\} \in \mathcal{F}, \ \forall x \in \mathbb{R}.$$

Such a function is said to be **$\mathcal{F}$-measurable**.

- By the notation $\{X \leq x\}$ we mean the set $\{\omega \in \Omega : X(\omega) \leq x\}$.
- The measurability property in fact tells us that $\{X \leq x\}$ is **an event** and allows us to compute $\mathrm{P}(X \leq x), \mathrm{P}(X = x), \mathrm{P}(X \in (a, b))$, etc.

Random variable and its distribution function

Definition

A **random variable** X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a function $X : \Omega \to \mathbb{R}$, assigning to each outcome $\omega \in \Omega$ a number $X(\omega)$, with the property that:

$$\{X \leq x\} \in \mathcal{F}, \; \forall x \in \mathbb{R}.$$

Such a function is said to be **$\mathcal{F}$-measurable**.

- By the notation $\{X \leq x\}$ we mean the set $\{\omega \in \Omega : X(\omega) \leq x\}$.
- The measurability property in fact tells us that $\{X \leq x\}$ is **an event** and allows us to compute $\mathrm{P}(X \leq x)$, $\mathrm{P}(X = x)$, $\mathrm{P}(X \in (a, b))$, etc.
- This condition must be met, but in practice we never verify it.

Random variable and its distribution function

Definition

A **random variable** X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a function $X : \Omega \to \mathbb{R}$, assigning to each outcome $\omega \in \Omega$ a number $X(\omega)$, with the property that:

$$\{X \leq x\} \in \mathcal{F}, \ \forall x \in \mathbb{R}.$$

Such a function is said to be **$\mathcal{F}$-measurable**.

- By the notation $\{X \leq x\}$ we mean the set $\{\omega \in \Omega : X(\omega) \leq x\}$.
- The measurability property in fact tells us that $\{X \leq x\}$ is **an event** and allows us to compute $\mathrm{P}(X \leq x), \mathrm{P}(X = x), \mathrm{P}(X \in (a, b))$, etc.
- This condition must be met, but in practice we never verify it.

The **probability distribution** of a random variable is given by its distribution function:

Definition

The **distribution function** of a random variable X is a function $F : \mathbb{R} \to [0, 1]$ defined as

$$F(x) = \mathrm{P}(X \leq x).$$

Random variable and its distribution function

There are various types of random variables.

- Some can take only isolated values (e.g., 0 or 1 for Heads and Tails of a coin toss, $1, \dots, 6$ for a die roll).

Random variable and its distribution function

There are various types of random variables.

- Some can take only isolated values (e.g., 0 or 1 for Heads and Tails of a coin toss, $1, \ldots, 6$ for a die roll).
- Some can take values from a continuous interval (e.g., weight of a newborn, time spent waiting for a bus, . . .).

Random variable and its distribution function

There are various types of random variables.

- Some can take only isolated values (e.g., 0 or 1 for Heads and Tails of a coin toss, $1, \ldots, 6$ for a die roll).
- Some can take values from a continuous interval (e.g., weight of a newborn, time spent waiting for a bus, …).

This divides the variables into **discrete** and **continuous**.

Random variable and its distribution function

There are various types of random variables.

- Some can take only isolated values (e.g., 0 or 1 for Heads and Tails of a coin toss, $1, \ldots, 6$ for a die roll).
- Some can take values from a continuous interval (e.g., weight of a newborn, time spent waiting for a bus, ...).

This divides the variables into **discrete** and **continuous**.

For discrete random variables, we will be interested in probabilities of the singular values, whereas for continuous we will work with probabilities of intervals.

Random variable and its distribution function

There are various types of random variables.

- Some can take only isolated values (e.g., 0 or 1 for Heads and Tails of a coin toss, $1, \ldots, 6$ for a die roll).
- Some can take values from a continuous interval (e.g., weight of a newborn, time spent waiting for a bus, ...).

This divides the variables into **discrete** and **continuous**.

For discrete random variables, we will be interested in probabilities of the singular values, whereas for continuous we will work with probabilities of intervals.

Regardless of the type, the distribution function gives us a full description of the random variable.

For any real number x, we can answer the question: "what is the probability that the random variable will be less than or equal to x"?

This allows us to answer questions about any equalities and inequalities.

Properties of the distribution function

Theorem

The distribution function F of a random variable X has following properties:

- **i)** F *is non-decreasing:* *if* $x < y$, *then* $F(x) \leq F(y)$
- **ii)** F *"starts at* 0 *and ends at* 1*":* $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$
- **iii)** F *is right continuous:* $\lim_{y \to x^+} F(y) = F(x)$

Properties of the distribution function

Theorem

The distribution function F of a random variable X has following properties:

- **i)** *F is non-decreasing:* *if* $x < y$, *then* $F(x) \leq F(y)$
- **ii)** *F "starts at 0 and ends at 1":* $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$
- **iii)** *F is right continuous:* $\lim_{y \to x^+} F(y) = F(x)$

Proof

- **i)** Recall the notation $\{X \leq x\} = \{\omega \in \Omega \colon X(\omega) \leq x\}$. Consider the disjoint partition

$$\{X \leq y\} = \{X \leq x\} \cup \{x < X \leq y\},$$

therefore $F(y) = \mathrm{P}(X \leq y) = \mathrm{P}(X \leq x) + \mathrm{P}(x < X \leq y) \quad \geq \quad \mathrm{P}(X \leq x) = F(x)$.

- **ii)** For simplicity we only sketch the proof by means of a sequence of events $B_n = \{X \leq -n\}$. For $n \to \infty$ it is decreasing in the sense of inclusion with the intersection equal to $\emptyset$, i.e., $B_n \searrow \emptyset$. From the continuity of probability theorem we have $\mathrm{P}(B_n) \to \mathrm{P}(\emptyset) = 0$. For the proof of the second statement it is enough to consider a sequence $A_n = \{X \leq n\} \nearrow \Omega$ and from the same theorem we have $\mathrm{P}(A_n) \to \mathrm{P}(\Omega) = 1$.
- **iii)** Similarly as **ii)** (see bibliography).

□

Properties of the distribution function

By means of the distribution function it is possible to express some important properties.

Lemma

Let F be a distribution function of a random variable X, then it holds that:

i) $\mathrm{P}(X > x) = 1 - F(x)$,

Properties of the distribution function

By means of the distribution function it is possible to express some important properties.

Lemma

Let F be a distribution function of a random variable X, then it holds that:

- **i)** $\mathrm{P}(X > x) = 1 - F(x)$,
- **ii)** $\mathrm{P}(X \in (x, y]) = \mathrm{P}(x < X \leq y) = F(y) - F(x)$,

Properties of the distribution function

By means of the distribution function it is possible to express some important properties.

Lemma

Let F be a distribution function of a random variable X, then it holds that:

i) $\mathrm{P}(X > x) = 1 - F(x)$,

ii) $\mathrm{P}(X \in (x, y]) = \mathrm{P}(x < X \leq y) = F(y) - F(x)$,

iii) $\mathrm{P}(X < x) = \lim_{y \to x^-} F(y)$,

Properties of the distribution function

By means of the distribution function it is possible to express some important properties.

Lemma

Let F be a distribution function of a random variable X, then it holds that:

- **i)** $\mathrm{P}(X > x) = 1 - F(x)$,
- **ii)** $\mathrm{P}(X \in (x, y]) = \mathrm{P}(x < X \leq y) = F(y) - F(x)$,
- **iii)** $\mathrm{P}(X < x) = \lim\limits_{y \to x^-} F(y)$,
- **iv)** $\mathrm{P}(X = x) = F(x) - \lim\limits_{y \to x^-} F(y)$.

Properties of the distribution function

By means of the distribution function it is possible to express some important properties.

Lemma

Let F be a distribution function of a random variable X, then it holds that:

i) $\mathrm{P}(X > x) = 1 - F(x)$,

ii) $\mathrm{P}(X \in (x, y]) = \mathrm{P}(x < X \leq y) = F(y) - F(x)$,

iii) $\mathrm{P}(X < x) = \lim_{y \to x^-} F(y)$,

iv) $\mathrm{P}(X = x) = F(x) - \lim_{y \to x^-} F(y)$.

Proof

i) $\Omega = \{X > x\} \cup \{X \leq x\}$ is a disjoint partition. Therefore $\mathrm{P}(\{X > x\}) = \mathrm{P}(\{X \leq x\}^c)$.

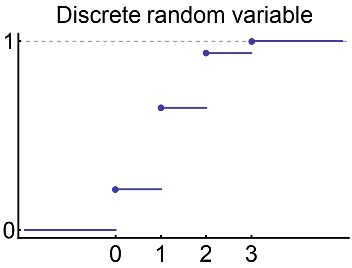
ii) See proof of **i)** of the previous theorem.

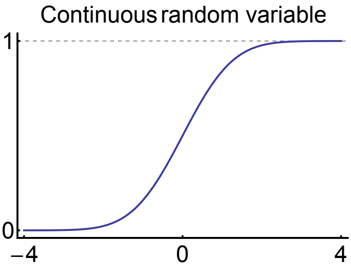
iii) See bibliography. Idea of the proof using a non-decreasing sequence and continuity of probability:

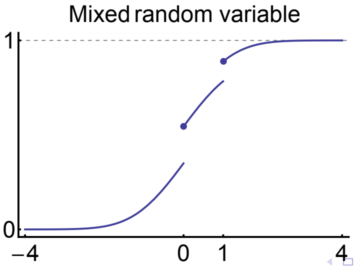
$$\{X \leq x - 1/n\} \nearrow \{X < x\} \qquad \Rightarrow \qquad F(x - 1/n) = \mathrm{P}(X \leq x - 1/n) \to \mathrm{P}(X < x).$$

iv) $\{X \leq x\} = \{X < x\} \cup \{X = x\}$ is a disjoint partition. Therefore
$\mathrm{P}(X = x) = \mathrm{P}(X \leq x) - \mathrm{P}(X < x)$. □

Types of random variables and their distribution functions







Discrete random variables

Definition

A random variable X is called **discrete** if it takes only values from some countable set $\{x_1, x_2, \dots\}$.

Discrete random variables

Definition

A random variable X is called **discrete** if it takes only values from some countable set $\{x_1, x_2, \dots\}$.

Probabilities of the values of a discrete random variable X are given by

$$\mathrm{P}(X = x_k), \qquad k = 1, 2, \dots$$

The probabilities $\mathrm{P}(X = x_k)$ can be viewed as a function of x and are sometimes called a **probability function**, or a **probability mass function** or a **discrete density** of the variable X.

Discrete random variables

Definition

A random variable X is called **discrete** if it takes only values from some countable set $\{x_1, x_2, \dots\}$.

Probabilities of the values of a discrete random variable X are given by

$$\mathrm{P}(X = x_k), \qquad k = 1, 2, \dots$$

The probabilities $\mathrm{P}(X = x_k)$ can be viewed as a function of x and are sometimes called a **probability function**, or a **probability mass function** or a **discrete density** of the variable X.

The **distribution function** of a discrete random variable has the form

$$F_X(x) = \mathrm{P}(X \le x) = \sum_{\text{all } x_k \le x} \mathrm{P}(X = x_k).$$

Discrete random variables

Definition

A random variable X is called **discrete** if it takes only values from some countable set $\{x_1, x_2, \dots\}$.

Probabilities of the values of a discrete random variable X are given by

$$\mathrm{P}(X = x_k), \qquad k = 1, 2, \dots$$

The probabilities $\mathrm{P}(X = x_k)$ can be viewed as a function of x and are sometimes called a **probability function**, or a **probability mass function** or a **discrete density** of the variable X.

The **distribution function** of a discrete random variable has the form

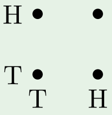
$$F_X(x) = \mathrm{P}(X \leq x) = \sum_{\text{all } x_k \leq x} \mathrm{P}(X = x_k).$$

From this it follows that $F_X(x)$ has jumps at points x_k and it is constant elsewhere. The size of the jump at point x_k is equal to $\mathrm{P}(X = x_k)$.

Example of a discrete random variable

Example – toss with two coins

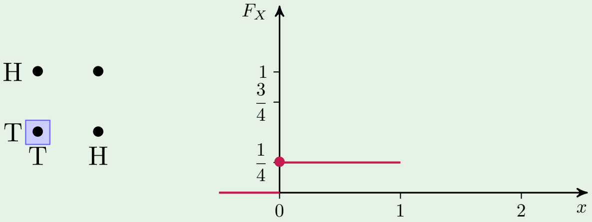
The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}), (\mathrm{H},\mathrm{T}), (\mathrm{T},\mathrm{H}), (\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads. The distribution function is $F_X(x) = \mathrm{P}(X \le x)$:



Example of a discrete random variable

Example – toss with two coins

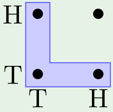
The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads. The distribution function is $F_X(x) = \mathrm{P}(X \le x)$:

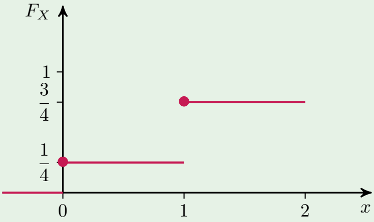


Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads. The distribution function is $F_X(x) = \mathrm{P}(X \leq x)$:

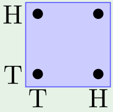


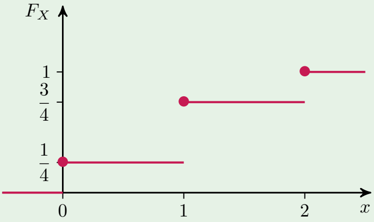


Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads. The distribution function is $F_X(x) = \mathrm{P}(X \leq x)$:

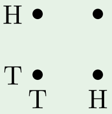


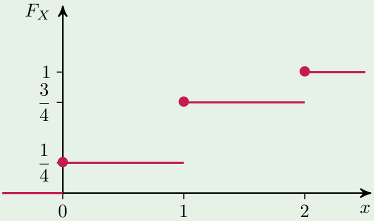


Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}), (\mathrm{H},\mathrm{T}), (\mathrm{T},\mathrm{H}), (\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads. The distribution function is $F_X(x) = \mathrm{P}(X \le x)$:





The distribution function $F_X = \mathrm{P}(X \le x)$ is given by

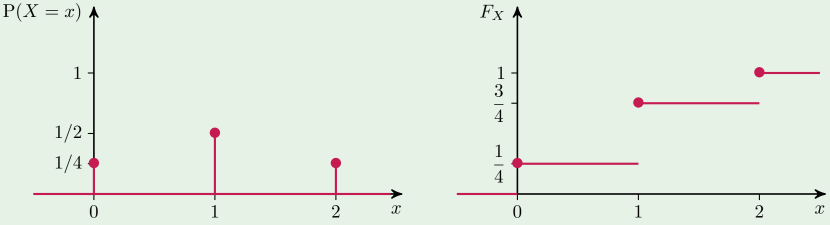
$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 & \mathrm{P}(\emptyset) \\ 1/4 & \text{for } 0 \le x < 1 & \mathrm{P}(\{(\mathrm{T},\mathrm{T})\}) \\ 3/4 & \text{for } 1 \le x < 2 & \mathrm{P}(\{(\mathrm{T},\mathrm{T}), (\mathrm{H},\mathrm{T}), (\mathrm{T},\mathrm{H})\}) \\ 1 & \text{for } 2 \le x & \mathrm{P}(\Omega). \end{cases}$$

Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}), (\mathrm{H},\mathrm{T}), (\mathrm{T},\mathrm{H}), (\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads.

Draw the probabilities of the values and the distribution function.

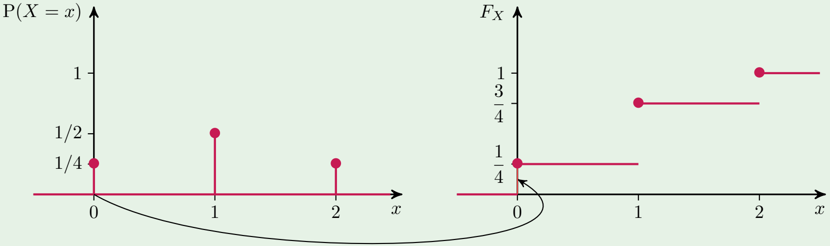


Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads.

Draw the probabilities of the values and the distribution function.

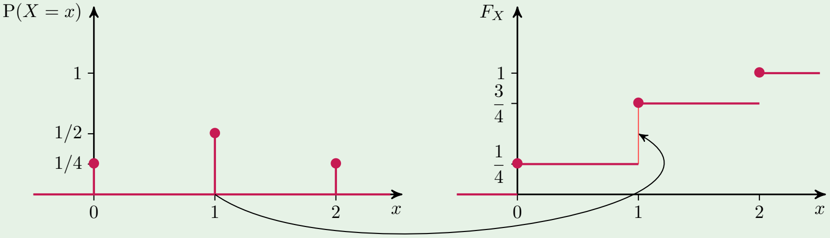


Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}), (\mathrm{H},\mathrm{T}), (\mathrm{T},\mathrm{H}), (\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads.

Draw the probabilities of the values and the distribution function.

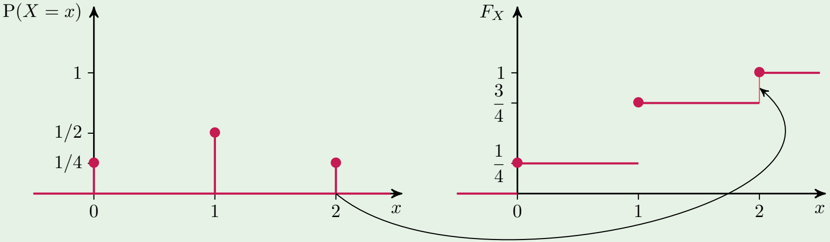


Example of a discrete random variable

Example – toss with two coins

The sample space is $\Omega = \{(\mathrm{H},\mathrm{H}), (\mathrm{H},\mathrm{T}), (\mathrm{T},\mathrm{H}), (\mathrm{T},\mathrm{T})\}$. Let the random variable X give the number of Heads.

Draw the probabilities of the values and the distribution function.



Relation between probabilities and distribution functions

When assigning probabilities to the values x_k, the **normalization condition** must hold:

$$\sum_{\text{all } x_k} \mathrm{P}(X = x_k) = 1.$$

Relation between probabilities and distribution functions

When assigning probabilities to the values x_k, the **normalization condition** must hold:

$$\sum_{\text{all } x_k} \mathrm{P}(X = x_k) = 1.$$

Generally, for computing $\mathrm{P}(X \in B)$, with $B \subset \mathbb{R}$, it is enough to know the probabilities of the possible values X: $\mathrm{P}(\{X \in B\}) = \sum_{x_k \in B} \mathrm{P}(X = x_k)$.

Relation between probabilities and distribution functions

When assigning probabilities to the values x_k, the **normalization condition** must hold:

$$\sum_{\text{all } x_k} \mathrm{P}(X = x_k) = 1.$$

Generally, for computing $\mathrm{P}(X \in B)$, with $B \subset \mathbb{R}$, it is enough to know the probabilities of the possible values X: $\quad \mathrm{P}(\{X \in B\}) = \sum_{x_k \in B} \mathrm{P}(X = x_k)$.

The **distribution of X** can be equivalently given by F_X or by the probabilities. Considering that $\mathrm{P}(X = x_k) = F_X(x_k) - F_X(x_{k-1})$ (we are considering an increasing ordering $x_1 < x_2 < x_3 < \dots$), the knowledge of the distribution function is equivalent to the knowledge of the probabilities $\mathrm{P}(X = x_k)$.

Relation between probabilities and distribution functions

When assigning probabilities to the values x_k, the **normalization condition** must hold:

$$\sum_{\text{all } x_k} \mathrm{P}(X = x_k) = 1.$$

Generally, for computing $\mathrm{P}(X \in B)$, with $B \subset \mathbb{R}$, it is enough to know the probabilities of the possible values X: $\mathrm{P}(\{X \in B\}) = \sum_{x_k \in B} \mathrm{P}(X = x_k)$.

The **distribution of X** can be equivalently given by F_X or by the probabilities. Considering that $\mathrm{P}(X = x_k) = F_X(x_k) - F_X(x_{k-1})$ (we are considering an increasing ordering $x_1 < x_2 < x_3 < \dots$), the knowledge of the distribution function is equivalent to the knowledge of the probabilities $\mathrm{P}(X = x_k)$.

Computation of the probabilities $\mathrm{P}(X = x_k)$:
Collect all ω for which $X(\omega) = x$ and sum their probabilities.

Relation between probabilities and distribution functions

When assigning probabilities to the values x_k, the **normalization condition** must hold:

$$\sum_{\text{all } x_k} \mathrm{P}(X = x_k) = 1.$$

Generally, for computing $\mathrm{P}(X \in B)$, with $B \subset \mathbb{R}$, it is enough to know the probabilities of the possible values X: $\mathrm{P}(\{X \in B\}) = \sum_{x_k \in B} \mathrm{P}(X = x_k)$.

The **distribution of X** can be equivalently given by F_X or by the probabilities. Considering that $\mathrm{P}(X = x_k) = F_X(x_k) - F_X(x_{k-1})$ (we are considering an increasing ordering $x_1 < x_2 < x_3 < \dots$), the knowledge of the distribution function is equivalent to the knowledge of the probabilities $\mathrm{P}(X = x_k)$.

Computation of the probabilities $\mathrm{P}(X = x_k)$:
Collect all ω for which $X(\omega) = x$ and sum their probabilities.

Computation of the distribution function $F_X(x) = \mathrm{P}(X \leq x_k)$:
Collect all ω for which $X(\omega) \leq x$ and sum their probabilities.

Examples of discrete random variables

Remark

A random variable X can be discrete even if the sample space itself is not discrete.

Examples of discrete random variables

Remark

A random variable X can be discrete even if the sample space itself is not discrete.

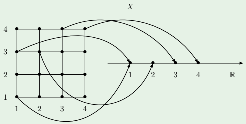
Example

Let us throw darts at a target $T \subset \mathbb{R}^2$.
The target can be divided into parts (often concentric annulus), denoted as T_1, T_2, T_3, T_4, T_5.
We can consider a discrete random variable X denoting the points obtained from one throw, for example

$$X(\omega) = \begin{cases} 10 & \text{for } \omega \in T_5 \\ 5 & \text{for } \omega \in T_4 \\ i & \text{for } \omega \in T_i,\ i = 1, 2, 3 \end{cases}$$

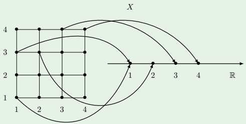
Example – minimum of two rolls of a 4-sided die (continuation)

$X = \min\{1^{\text{st}} \text{ roll}, 2^{\text{nd}} \text{ roll}\}$:

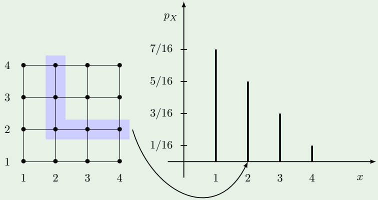


Example – minimum of two rolls of a 4-sided die (continuation)

$X = \min\{1^{\text{st}}\text{ roll}, 2^{\text{nd}}\text{ roll}\}$:

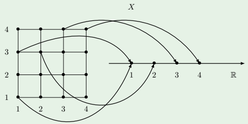


Probabilities:

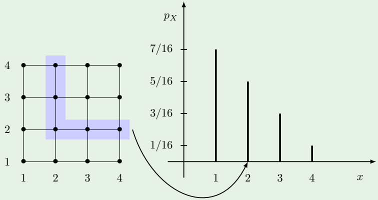


Example – minimum of two rolls of a 4-sided die (continuation)

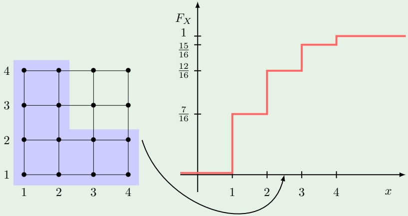
$X = \min\{1^{\text{st}}\text{ roll}, 2^{\text{nd}}\text{ roll}\}$:



Probabilities:



Distribution function:



Important discrete probability distributions

Example

(Will be studied later)

- **Bernoulli** (Alternating) distribution with a parameter $p \in [0, 1]$, $X \sim \text{Be}(p)$:
  (*One toss of an unbalanced coin.*)

$$\mathrm{P}(X = 1) = p, \qquad \mathrm{P}(X = 0) = 1 - p.$$

- **Binomial** distribution with parameter $p \in [0, 1]$, $X \sim \text{Binom}(n, p)$:
  (*Number of Heads in n tosses of an unbalanced coin.*)

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- **Geometric** distribution with a parameter $p \in (0, 1)$, $X \sim \text{Geom}(p)$:
  (*Number of tosses of an unbalanced coin until the first Heads appear.*)

$$\mathrm{P}(X = k) = (1 - p)^{k-1} p$$

- **Poisson** distribution with a parameter $\lambda > 0$, $X \sim \text{Poisson}(\lambda)$:
  (*Limit of the Binomial distribution for $n \to \infty$.*)

$$\mathrm{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Continuous random variables – motivation

In some situations, a random variable can take **uncountably** many possible values.

This arises when dealing with continuous models – measuring time, height, coordinates, etc.

Continuous random variables – motivation

In some situations, a random variable can take **uncountably** many possible values.

This arises when dealing with continuous models – measuring time, height, coordinates, etc.

We cannot assign a positive probability $\mathrm{P}(X = x)$ to each value, because then the probabilities of the uncountable many values would sum up to infinity.

Therefore we regard each singular value as having zero probability (intuitively, it is, e.g., infinitely improbable having to wait for the bus for exactly $3:00:00\ldots$ minutes).

Instead, we need a way to measure the probability of intervals.

Recall the Romeo and Juliet problem, where each of them arrives at a random time point in an one-hour window, evenly chosen.

Continuous random variables – motivation

In some situations, a random variable can take **uncountably** many possible values.

This arises when dealing with continuous models – measuring time, height, coordinates, etc.

We cannot assign a positive probability $\mathrm{P}(X = x)$ to each value, because then the probabilities of the uncountable many values would sum up to infinity.

Therefore we regard each singular value as having zero probability (intuitively, it is, e.g., infinitely improbable having to wait for the bus for exactly $3 : 00 : 00\ldots$ minutes).

Instead, we need a way to measure the probability of intervals.

Recall the Romeo and Juliet problem, where each of them arrives at a random time point in an one-hour window, evenly chosen.

Often we need to introduce an uneven distribution of values.

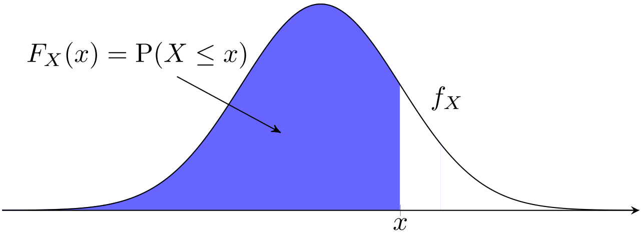
Continuous random variables

Definition

A random variable X is called (absolutely) **continuous**, if there exists a **non-negative** function $f_X : \mathbb{R} \to [0, +\infty)$ such that for all $x \in \mathbb{R}$ the distribution function F_X can be expressed as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t.$$

The function f_X is called the **probability density** of the random variable X.



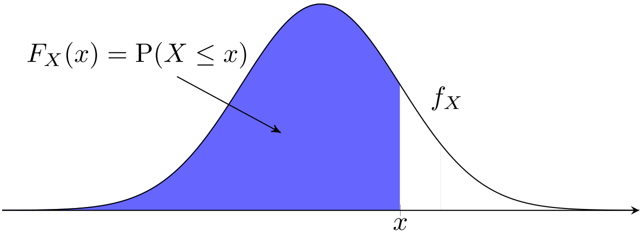
Continuous random variables

Definition

A random variable X is called (absolutely) **continuous**, if there exists a **non-negative** function $f_X : \mathbb{R} \to [0, +\infty)$ such that for all $x \in \mathbb{R}$ the distribution function F_X can be expressed as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t.$$

The function f_X is called the **probability density** of the random variable X.



The distribution function of a continuous random variable is continuous.

Properties of continuous random variables

Theorem

Let f_X be a density of a continuous random variable X. Then it holds that

- **i)** $\displaystyle\int_{-\infty}^{+\infty} f_X(t)\mathrm{d}t = 1$ *(**normalization condition**),*
- **ii)** $\mathrm{P}(X = x) = 0$ *for all* $x \in \mathbb{R}$*,*
- **iii)** $f_X(t) = \dfrac{\mathrm{d}F_X}{\mathrm{d}t}(t)$ *at points where the derivative exists,*
- **iv)** $\mathrm{P}(a < X \leq b) = \displaystyle\int_a^b f_X(t)\,\mathrm{d}t = F_X(b) - F_X(a)$*,*
- **v)** $\mathrm{P}(X \in B) = \displaystyle\int_B f_X(t)\mathrm{d}t$ *for all B in the Borel σ-algebra on $\mathbb{R}$, i.e., for all "common" sets.*

Properties of continuous random variables

Theorem

Let f_X be a density of a continuous random variable X. Then it holds that

i) $\int_{-\infty}^{+\infty} f_X(t)\mathrm{d}t = 1$ *(**normalization condition**),*

ii) $\mathrm{P}(X = x) = 0$ *for all* $x \in \mathbb{R}$,

iii) $f_X(t) = \frac{\mathrm{d}F_X}{\mathrm{d}t}(t)$ *at points where the derivative exists,*

iv) $\mathrm{P}(a < X \leq b) = \int_a^b f_X(t)\,\mathrm{d}t = F_X(b) - F_X(a)$,

v) $\mathrm{P}(X \in B) = \int_B f_X(t)\mathrm{d}t$ *for all* B *in the Borel* σ*-algebra on* $\mathbb{R}$*, i.e., for all "common" sets.*

Consequences:

- $\mathrm{P}(X \leq x) = \mathrm{P}(X < x)$ – from **ii)**
- $f_X(t)\mathrm{d}t \approx \mathrm{P}(t < X < t + \mathrm{d}t)$ for $\mathrm{d}t << 1$ – from **iv)**

Properties of continuous random variables

Proof

i) $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = \lim_{x\to+\infty} F_X(x) = 1.$

ii) Using the continuity of the distribution function and the previous theorem:
$\mathrm{P}(X = x) = F(x) - \lim_{y\to x-} F(y) = 0.$

iii) It follows from the properties of derivatives and integrals (first fundamental Theorem of calculus).

iv) $\displaystyle\mathrm{P}(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f_X(t)\mathrm{d}t - \int_{-\infty}^{a} f_X(t)\mathrm{d}t = \int_{a}^{b} f_X(t)\mathrm{d}t.$

(second fundamental Theorem of calculus – Newton's formula)

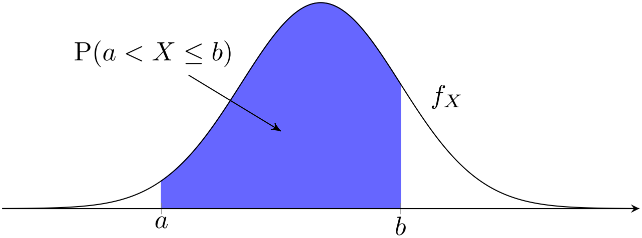
v) From the properties of the Lebesgue integral – advanced, see bibliography.

□

Relation between density and probability

Now we recall and illustrate the important property of the probability density:

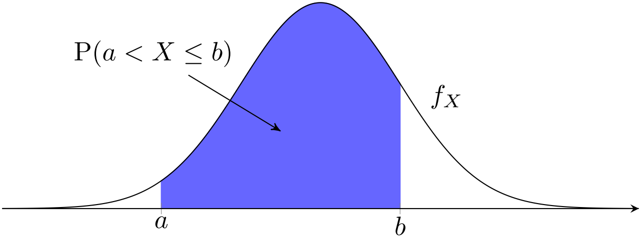
$$\mathrm{P}(a < X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x = \Big[F(x)\Big]_a^b = F(b) - F(a).$$



Relation between density and probability

Now we recall and illustrate the important property of the probability density:

$$\mathrm{P}(a < X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x = \Big[F(x)\Big]_a^b = F(b) - F(a).$$



Note that when dealing with continuous random variables, it does not matter whether the inequalities are strict or non-strict.

$$\mathrm{P}(a < X \leq b) = \mathrm{P}(a \leq X \leq b) = \mathrm{P}(a \leq X < b) = \mathrm{P}(a < X < b).$$

Romeo, Juliet and the uniform distribution

Example – uniform distribution of Romeo's arrival

Denote the time when Romeo arrives at the meeting point as a random variable X. Suppose that X has the **uniform distribution** on the interval $[0, 1]$, meaning that its density is constant on this interval and zero elsewhere.

$$f_X(x) = \begin{cases} c & \text{for } x \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$

Determine the value of c, so that f truly forms a density of a random variable.

Romeo, Juliet and the uniform distribution

Example – uniform distribution of Romeo's arrival

Denote the time when Romeo arrives at the meeting point as a random variable X. Suppose that X has the **uniform distribution** on the interval $[0, 1]$, meaning that its density is constant on this interval and zero elsewhere.

$$f_X(x) = \begin{cases} c & \text{for } x \in [0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

Determine the value of c, so that f truly forms a density of a random variable.

From the normalization condition we know that the area under the graph of the density needs to be equal to one. Therefore the density needs to integrate to one:

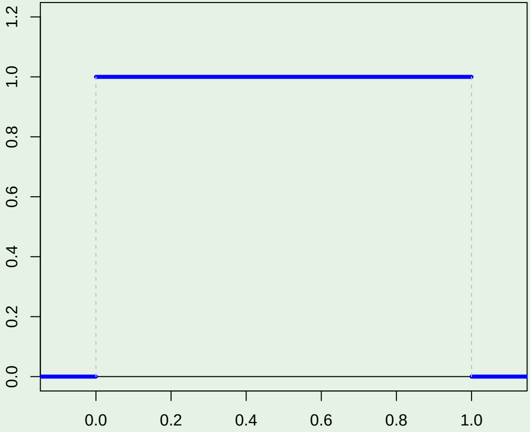
$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 c \cdot dx = [c \cdot x]_0^1 = c \cdot 1 - c \cdot 0 = c = 1.$$

The constant c has to be equal to one.

Romeo, Juliet and the uniform distribution

Example – uniform distribution of Romeo's arrival (continued)

Density of the continuous uniform distribution on the interval $[0, 1]$:



Romeo, Juliet and the uniform distribution

Example – uniform distribution of Romeo's arrival (continued)

What is the probability that Romeo arrives between 12:15 and 12:45?

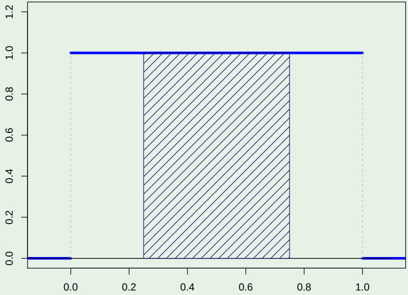
Romeo, Juliet and the uniform distribution

Example – uniform distribution of Romeo's arrival (continued)

What is the probability that Romeo arrives between 12:15 and 12:45?

Probabilities concerning intervals are obtained as the corresponding area under the density:

$$\int_{1/4}^{3/4} 1dx = [x]_{1/4}^{3/4} = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}.$$

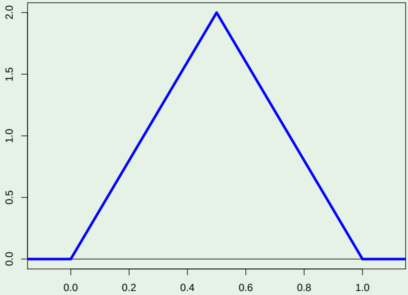


Romeo, Juliet and a non-uniform distribution

Example – non-uniform distribution of Juliet's arrival

Suppose that Juliet arrives at the meeting point according to a non-uniform distribution with density:

$$f_X(x) = \begin{cases} 4x & \text{for } x \in [0, 1/2] \\ 4 - 4x & \text{for } x \in [1/2, 1] \\ 0 & \text{otherwise.} \end{cases}$$



What is the probability that Juliet arrives between 12:15 and 12:45?

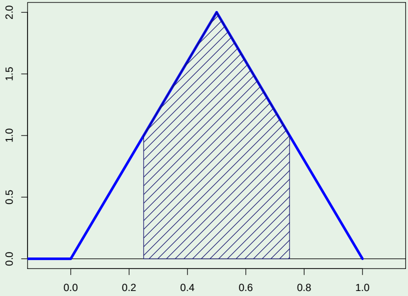
Romeo, Juliet a non-uniform distribution

Example – non-uniform distribution of Juliet's arrival (continued)

What is the probability that Juliet arrives between 12:15 and 12:45?

Probabilities concerning intervals are obtained as the corresponding area under the density:

$$\int_{1/4}^{3/4} f(x)dx = \cdots = \frac{3}{4}.$$



Note that when the distribution of the arrivals is not uniform, the probability that they will meet cannot be obtained using the geometric approach as before.

Functions of random variables

For a random variable X with a known distribution, we are often interested in the distribution of values somehow calculated from the values of X, say $Y = g(X)$.

Functions of random variables

For a random variable X with a known distribution, we are often interested in the distribution of values somehow calculated from the values of X, say $Y = g(X)$.

Example – linear transformation

Let X be a random temperature in degrees Celsius.
Then $Y = 1.8X + 32$ corresponds to the temperature in degrees Fahrenheit.

Functions of random variables

For a random variable X with a known distribution, we are often interested in the distribution of values somehow calculated from the values of X, say $Y = g(X)$.

Example – linear transformation

Let X be a random temperature in degrees Celsius.
Then $Y = 1.8X + 32$ corresponds to the temperature in degrees Fahrenheit.

In the case of a discrete random variables the situation is relatively easy.

- $g(X)$ is always a random variable.
- The distribution of the random variable $g(X)$ is always discrete.

Functions of random variables

For a random variable X with a known distribution, we are often interested in the distribution of values somehow calculated from the values of X, say $Y = g(X)$.

Example – linear transformation

Let X be a random temperature in degrees Celsius.
Then $Y = 1.8X + 32$ corresponds to the temperature in degrees Fahrenheit.

In the case of a discrete random variables the situation is relatively easy.

- $g(X)$ is always a random variable.
- The distribution of the random variable $g(X)$ is always discrete.

If X is a continuous random variable, the following complications arise:

- It can happen that $g(X)$ is not a random variable.
  (Therefore the assumption of **measurability** of g is needed.)

Functions of random variables

For a random variable X with a known distribution, we are often interested in the distribution of values somehow calculated from the values of X, say $Y = g(X)$.

Example – linear transformation

Let X be a random temperature in degrees Celsius.
Then $Y = 1.8X + 32$ corresponds to the temperature in degrees Fahrenheit.

In the case of a discrete random variables the situation is relatively easy.

- $g(X)$ is always a random variable.
- The distribution of the random variable $g(X)$ is always discrete.

If X is a continuous random variable, the following complications arise:

- It can happen that $g(X)$ is not a random variable.
  (Therefore the assumption of **measurability** of g is needed.)
- The distribution of a random variable $g(X)$ can be discrete, continuous or mixed.

Function of a discrete random variable

Lemma – function of discrete random variable

Consider a function $g\colon \mathbb{R} \to \mathbb{R}$ and a ***discrete*** *random variable X, and define the function of the random variable $g(X)$ by $g(X)(\omega) = g(X(\omega))$ for all $\omega \in \Omega$.*

Function of a discrete random variable

Lemma – function of discrete random variable

Consider a function $g\colon \mathbb{R} \to \mathbb{R}$ and a ***discrete*** *random variable X, and define the function of the random variable $g(X)$ by $g(X)(\omega) = g(X(\omega))$ for all $\omega \in \Omega$.*

Then $g(X)$ is a discrete random variable with probabilities of the values

$$\mathrm{P}(g(X) = y) = \sum_{x_k : g(x_k) = y} \mathrm{P}(X = x_k).$$

Function of a discrete random variable

Lemma – function of discrete random variable

*Consider a function $g\colon \mathbb{R} \to \mathbb{R}$ and a **discrete** random variable X, and define the function of the random variable $g(X)$ by $g(X)(\omega) = g(X(\omega))$ for all $\omega \in \Omega$.*

Then $g(X)$ is a discrete random variable with probabilities of the values

$$\mathrm{P}(g(X) = y) = \sum_{x_k : g(x_k) = y} \mathrm{P}(X = x_k).$$

Proof

The probabilities of the values of $g(X)$ can be obtained from the (countable) disjoint partition

$$\{g(X) = y\} = \bigcup_{x_k : g(x_k) = y} \{X = x_k\}.$$

□

Functions of random variables

Lemma – function of a general random variable

Consider a ***measurable*** *function* $g : \mathbb{R} \to \mathbb{R}$ *and an* ***arbitrary*** *random variable* X *and define the function of random the variable* $g(X)$ *as* $g(X)(\omega) = g(X(\omega))$ *for all* $\omega \in \Omega$.

Functions of random variables

Lemma – function of a general random variable

Consider a ***measurable*** *function* $g: \mathbb{R} \to \mathbb{R}$ *and an* ***arbitrary*** *random variable* X *and define the function of random the variable* $g(X)$ *as* $g(X)(\omega) = g(X(\omega))$ *for all* $\omega \in \Omega$*. Then the function* $g(X)$ *of the random variable* X *is a random variable.*

Functions of random variables

Lemma – function of a general random variable

Consider a ***measurable*** *function* $g : \mathbb{R} \to \mathbb{R}$ *and an* ***arbitrary*** *random variable* X *and define the function of random the variable* $g(X)$ *as* $g(X)(\omega) = g(X(\omega))$ *for all* $\omega \in \Omega$*. Then the function* $g(X)$ *of the random variable* X *is a random variable.*

Note: g is measurable if the set $\{x \in \mathbb{R} : g(x) \leq y\}$ belongs to the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ for all $y \in \mathbb{R}$.

Functions of random variables

Lemma – function of a general random variable

*Consider a **measurable** function $g\colon \mathbb{R} \to \mathbb{R}$ and an **arbitrary** random variable X and define the function of random the variable $g(X)$ as $g(X)(\omega) = g(X(\omega))$ for all $\omega \in \Omega$. Then the function $g(X)$ of the random variable X is a random variable.*

Note: g is measurable if the set $\{x \in \mathbb{R}\colon g(x) \leq y\}$ belongs to the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ for all $y \in \mathbb{R}$.

Proof

The proof that $g(X)$ is a random variable consists in verifying the measurability of $Y = g(X)$, i.e., that $\{Y \leq y\}$ is an event for all y:

$$\{g(X) \leq y\} = \{\omega \in \Omega\colon g(X(\omega)) \leq y\} \in \mathcal{F}, \ \forall y \in \mathbb{R}.$$

A detailed proof can be found in the bibliography. □

Functions of random variables

Remark

Generally for a distribution function $F_Y(y)$ of a random variable $Y = g(X)$ it holds that

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}\left(\{\omega \in \Omega\colon g(X(\omega)) \leq y\}\right).$$

Functions of random variables

Remark

Generally for a distribution function $F_Y(y)$ of a random variable $Y = g(X)$ it holds that

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}\left(\{\omega \in \Omega\colon g(X(\omega)) \leq y\}\right).$$

If Y is continuous we obtain f_Y as the derivative of $F_Y(y)$ with respect to y.

Functions of random variables

Remark

Generally for a distribution function $F_Y(y)$ of a random variable $Y = g(X)$ it holds that

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}\left(\{\omega \in \Omega\colon g(X(\omega)) \leq y\}\right).$$

If Y is continuous we obtain f_Y as the derivative of $F_Y(y)$ with respect to y.

Possible simplification:

- If the inverse g^{-1} of g exists and is increasing, then it holds that

$$F_Y(y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}\left(X \leq g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right).$$

Functions of random variables

Remark

Generally for a distribution function $F_Y(y)$ of a random variable $Y = g(X)$ it holds that

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}\left(\{\omega \in \Omega \colon g(X(\omega)) \leq y\}\right).$$

If Y is continuous we obtain f_Y as the derivative of $F_Y(y)$ with respect to y.

Possible simplification:

- If the inverse g^{-1} of g exists and is increasing, then it holds that

$$F_Y(y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}\left(X \leq g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right).$$

- If g is strictly monotone, then g^{-1} is differentiable and $Y = g(X)$ is continuous with

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}.$$

✓ Proofs and more information can be found in bibliography.

Recap

- A **random variable** X is a measurable function, which assigns real values to the outcomes of a random experiment.

Recap

- A **random variable** X is a measurable function, which assigns real values to the outcomes of a random experiment.
- The **distribution** of X gives the information of the probabilities of its values and is uniquely given by the **distribution function**:

$$F_X(x) = \mathrm{P}(X \leq x).$$

- There are two major types of random variables:
  - **discrete**, taking only countably many possible values;
  - **continuous**, taking values from an interval.

Recap

- A **random variable** X is a measurable function, which assigns real values to the outcomes of a random experiment.
- The **distribution** of X gives the information of the probabilities of its values and is uniquely given by the **distribution function**:

$$F_X(x) = \mathrm{P}(X \leq x).$$

- There are two major types of random variables:
  - **discrete**, taking only countably many possible values;
  - **continuous**, taking values from an interval.
- The distribution can be given by:
  - for discrete distributions by the **probabilities** of possible values $\mathrm{P}(X = x_k)$.
  - for continuous distributions by the **density** f_X for which

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

Recap

- A **random variable** X is a measurable function, which assigns real values to the outcomes of a random experiment.
- The **distribution** of X gives the information of the probabilities of its values and is uniquely given by the **distribution function**:

$$F_X(x) = \mathrm{P}(X \leq x).$$

- There are two major types of random variables:
  - **discrete**, taking only countably many possible values;
  - **continuous**, taking values from an interval.
- The distribution can be given by:
  - for discrete distributions by the **probabilities** of possible values $\mathrm{P}(X = x_k)$.
  - for continuous distributions by the **density** f_X for which

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

- The generalized inverse of the distribution function is called the **quantile function** and can be used for simulations.