University of Twente

Probability Theory for Engineers

Summary: definitions, properties and more

W. van den Brink
Based on the April 2018 version of reader 642

November 25, 2021

Contents

Chapter 1

Experiment, sample space and probability

1.1 Experiment and sample space

An **probabilistic experiment** or stochastic experiment is an experiment which does not necessarily lead to the same outcome when repeated under the same conditions.

Definition 1
The sample space S of an experiment is the set of all possible outcomes.

Definition 2
An event is a subset of the sample space S.

The empty set $\emptyset$ is called the impossible event.

The set S is the certain event.

Definition 3
The events $A_1, A_2, \ldots, A_n$ or $A_1, A_2, \ldots$ are mutually exclusive if $A_i A_j = \emptyset$ for every possible combination (i, j) for which $i \neq j$.

Definition 4
The sequence of events $\{A_j\}$ is a partition of the event B if the events A_i are mutually exclusive and $B = \cup_i A_i$.

Property 1
Properties of events:

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cup B = A \cup \left(\overline{A} B\right)$
- $B = (AB) \cup \left(\overline{A} B\right)$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

1.2 Symmetric probability spaces

Definition 5 (Probability definition by Laplace)
When the sample space S of an experiment $N(S)$ contains equally likely outcomes and the event A consists of $N(A)$ outcomes, then the probability of an event A, denoted by $P(A)$, equals:

$$P(A) = \frac{N(A)}{N(S)}$$

Definition 6
If S is a finite sample sapce of an experiment and the probabilities $P(A)$ of events A are defined according to Definition 5, the pair (S, P) is called a symmetric probability space.

Property 2
Properties for a symmetric probability space:

- $P(A) \geq 0$ for every event A
- $P(S) = 1$
- $A \subset P \to P(A) \leq P(B)$
- $P\left(\overline{A}\right) = 1 - P(A)$
- If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then $P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P\left(A_i\right)$

1.3 Relative frequency and the empirical law of large numbers

Definition 7
Assume that we have an experiment with sample space S which we can repeat arbitrarily often. If the event A occured $n(A)$ times in total with n repetitions, then we define

$$f_n(A) = \frac{n(A)}{n}$$

as the relative frequency (or frequency quotient) of A in n repetitions.

Empirically, it appears that $\lim_{n\to\infty} f_n(A) = P(A)$: the empirical law of large numbers. This is not strictly true, as we cannot calculate this limit.

1.4 Axioms of Kolmogorov

Definition 8 (Axioms of Kolmogorov)
Consider an experiment with a random non-empty sample space S. A function P which assigns a real number $P(A)$ to every event $A \subset S$, is called a probability or probability measure on S if:

1. $P(A) \geq 0$ for every event A,
2. $P(S) = 1$, and
3. for every countable sequence of mutually exclusive events $A_1, A_2, \ldots, A_n$ or $A_1, A_2, \ldots$:

$$P\left(\cup_i A_i\right) = \sum_i P\left(A_i\right)$$

Property 3
$P(\emptyset) = 0$.

Property 4
$P(\overline{A}) = 1 - P(A)$, for every event A.

Property 5
For two events A and B with $A \subset B$ we have: $P(A) \leq P(B)$.

Property 6
For two events A and B (which are not necessarily mutually exclusive):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Chapter 2

Combinatorial Probability

2.1 Theory and examples

Property 7 (The product rule)
When an experiment consists of performing k partial experiments and the i^{th} partial experiment has n_i possible outcomes, no matter what the results of the partial experiments are, then $n_1 \times n_2 \times \cdots \times n_k$ outcomes of the total experiment are possible.

Property 8 (The permutation rule)
The number of orders or permutations in which k different things can be arranged is $k!$.

Property 9
When $A_1, A_2, \ldots, A_k$ are mutually exclusive events, then:

$$\mathsf{N}\left(\cup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathsf{N}\left(A_i\right)$$

Property 10
We randomly draw k times from a set of n different elements. Then in the following three cases the probability space is symmetric:

1. Draw with replacement, ordered outcomes: $N(S) = n^k$.
2. Draw without replacement, ordered outcomes (variations or permutations of k out of n):

$$\mathsf{N}(S) = n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

3. Draw without replacement, unordered outcomes (combinations of k out of n):

$$N(S) = \binom{n}{k}$$

In case of drawing with replacement, unordered outcomes, the probability space is non-symmetric.

2.2 Combinatorics and random variables

Property 11
If we draw n times, at random and without replacement, from a set of N balls, consisting of R red and $N-R$ white balls, the probability of event A_k, that we draw k red (and $n-k$ white) balls, is given by:

$$P(A_k) = \frac{\binom{R}{k} \cdot \binom{N-R}{n-k}}{\binom{N}{n}}$$

The number of drawn red balls is between 0 and R, so $0 \leq k \leq R$.

Similarly the condition for the white balls is $0 \leq n - k \leq N - R$.

Chapter 3

Conditional probability and independence

3.1 Conditional probability

Definition 9
When A and B are events and $P(B) > 0$, then we define

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

as the (conditional) probability of A under condition of B.

Property 12 (General product rule)
For n events $A_1, A_2, \ldots, A_n$ with $n \geq 2$ and $P(A_1 A_2 \ldots A_{n-1}) > 0$ we have:

$$P(A_1 A_2 \ldots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot \cdots \cdot P(A_n|A_1 A_2 \ldots A_{n-1})$$

3.2 Law of total probability and Bayes' Rule

Property 13
If $\{S_i\}$ is a partition of S such that $P(S_i) > 0$ for all i, then for each event A we have:

$$P(A) = P(A|S_1) \cdot P(S_1) + P(A|S_2) \cdot P(S_2) + \cdots = \sum_i P(A|S_i) \cdot P(S_i)$$

Property 14 (Bayes' rule)
If $\{S_i\}$ is a partition of S with $P(S_i) > 0$ for each i, then for each event A with $P(A) > 0$ we have:

$$P(S_k|A) = \frac{P(AS_k)}{P(A)} = \frac{P(A|S_k)P(S_k)}{\sum_i P(A|S_i) \cdot P(S_i)}$$

3.3 Independence of events and random variables

Definition 10
The events A and B are independent when $P(AB) = P(A) \cdot P(B)$.

Definition 11
The events $A_1, A_2, A_3, \ldots$ are independent if for each subsequence $A_{i1}, A_{i2}, \ldots, A_{ik}$ with $k \geq 2$, it is true that

$$P(A_{i1} A_{i2} \ldots A_{ik}) \overset{\text{ind.}}{=} P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots P(A_{ik})$$

Definition 12
A series of experiments is called Bernoulli experiments or trials if

1. Each experiment has two possible outcomes, often denoted with 'Success' and 'Failure',
2. The experiments are independent, and
3. The probability of success is the same for each experiment

The success probability is often denoted by p and the probability of failure with $1-p$.

Property 15 (The binomial formula)
If X is the number of successes in n Bernoulli experiments with success probability p, then:

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

where $k=0,1,2,\ldots,n$

Property 16 (The geometric formula)
If we conduct Bernoulli trials with success probability p until a success occurs and X is the number of required trials, then

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

where $k=1,2,3,\ldots$

X is said to have a geometric distribution with parameter p.

Chapter 4

Discrete random variables

4.1 Random variable

Definition 13
IF S is the sample space of an experiment, then a real function $X : S \to \mathbb{R}$, which assigns a real number $X(s)$ to each outcome $s \in S$, is a random variable.

- We use capitals for random variables.
- $X = x$ means that the variable X attains the real value x. This is called a reliazation or observed value.

Definition 14
The range S_x of a random variable X, defined on a sample space S, is the set of all possible realizations $X(s)$. So $S_x = \{X(s) \mid s \in S\}$.

The range of a variable can be

- Finite, e.g. $S_x = \{1, 2, 3, 4, 5, 6\}$
- Countably infinite, e.g. $S_y = \{1, 2, 3, \dots\}$
- Not countably infinite, e.g. $S_z = [0, \infty)$

Definition 15
A random variable X is discrete if the range S_x is denumerable.

Corollary 1
If X is discrete, S_x has the shape $\{x_1, x_2, \dots, x_n\}$ or $\{x_1, x_2, x_2, \dots\}$.

4.2 The probability function of a discrete random variable

Definition 16
If X is a discrete random variable, then we will call the function that assigns a probability $P(X = x)$ to each $x \in S_x$ the probability function of X.

The probability function is often graphed using a bar graph.

Property 17
For the probability function of a discrete random variable X we have:

1. $P(X = x) \geq 0$ for $x \in S_x$
2. $\sum_{x \in S_x} P(X = x) = 1$

Conversely, any function which satisfies these conditions is a probability function.

4.3 The expectation of a discrete random variable

Definition 17
The expactation or expected value $\mathsf{E}(X)$ of a discrete random variable X is given by

$$\mathsf{E}(X) = \sum_{x \in S_x} xP(X = x)$$

provided that the summation is absolute convergent.

E is often referred to as the population mean, and therefore denoted with μ, the Greek letter *m* for *mean*. In statistics, we use another mean, the sample mean $\bar{x}$, the average value of observations in a sample, drawn from a (usually large) population. The value of μ is often unknown, and $\bar{x}$ attempts to estimate it.

Property 18
If the probability function is symmetric with respect to $x = c$, then $\mathsf{E}(X) = c$.

4.4 Functions of a discrete random variable; variance

Property 19
If X is a discrete random variable and g a (real) function, then:

$$\mathsf{E}(g(X)) = \sum_{x \in S_x} g(x)P(X = x)$$

(provided that the summation is absolute convergent)

Note that $\mathsf{E}g(X)$ means $\mathsf{E}(g(X))$.

Property 20
If X is a discrete random variable and g and h are real functions, then for real constants $a, b \in \mathbb{R}$ we have:

1. $\mathsf{E}\left(aX + b\right) = a\mathsf{E}(X) + b$
2. $\mathsf{E}\left[ag(X) + bh(X)\right] = a\mathsf{E}g(X) + b\mathsf{E}h(X)$

Definition 18
$E\left(X^k\right)$ is the k^{th} moment of the random variable X $(k = 1, 2, 3, \ldots$.

$$\mathsf{E}\left(X^k\right) = \sum_x x^k P(X = x)$$

Definition 19
The variance of X (notation: $\mathsf{var}(X)$ or σ_x^2 is defined as

$$\mathsf{var}(X) = E\left(X - \mu_x\right)^2$$

According to Property 19, we can compute $\mathsf{var}(X)$ as follows:

$$\mathsf{var}(X) = \sum_x \left(x - \mu_x\right)^2 \cdot P(X = x)$$

However, when we apply Property 20 to express $\mathsf{var}(X)$ in the first and second moment (Definition 18), we find:

$$\begin{aligned}
\mathsf{var}(X) &= \mathsf{E}\left(X - \mu_x\right)^2 \\
&= \mathsf{E}\left(X^2 - 2\mu_x \cdot X + \mu_x^2\right) \\
&= \mathsf{E}\left(X^2\right) - 2\mu_x \cdot \mathsf{E}(X) + \mu_x^2 \\
&= \mathsf{E}\left(X^2\right) - \mu_x^2
\end{aligned}$$

Because of this, it is preferable to use the following definition:

Definition 20
var$(X) =$ "the 2^{nd} moment minus the square of the 1^{st} moment"

Remember that the first moment is $\mathsf{E}(X) = \mu$, the population mean.

Definition 21
The standard deviation of X (notation: σ_X) is the square root of the deviance:

$$\sigma_X = \sqrt{\mathsf{var}(X)}$$

In practice, we use the standard deviation more often, as it has the same unit as X. The variance does not: its unit is the unit of X squared.

Property 21
Properties of variance and standard deviation:

- var$(X) \geq 0$ and $\sigma_X \geq 0$
- var$(X) = \mathsf{E}\left(X^2\right) - \mu_x^2$ (the computational formula)
- If var$(X) > 0$, so if X is not degenerate, we have $\mathsf{E}(X^2) > (\mathsf{E}X)^2$
- var$(aX + b) = a^2 \cdot$ var(X) and $\sigma_{aX+b} = |a| \cdot \sigma_X$

Property 22 (Checyshev's inequality)
For any real number $c > 0$, we have $P\left(|X - \mu_x| \geq c\right) \leq \frac{\mathsf{var}(X)}{c^2}$

4.5 The binomial, hypergeometric and Poisson-distribution

4.5.1 The binomial distribution

Definition 22
X is binomially distributed with parameters n and p, for all $n = 1, 2, \ldots$ and $p \in [0, 1]$, if the probability function of X is given by:

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

where $k = 0, 1, 2, \ldots, n$.

Alternative notations: X is B(n, p)-distributed, or X B(n, p).

Intuition: a binomial distribution decides the probability of success after exactly n trials with probability p.

The probability of the trial succeeding exactly k times requires the trial succeeding k times *and* failing $n - k$ times. The combinatoric finds out how often this is the case.

Corollary 2
If X is B(n, p)-distributed, then expected value and variance are given by:

$$\mathsf{E}(X) = np \qquad \text{and} \qquad \mathsf{var}(X) = np(1 - p)$$

Some special cases of n and p:

- If $p = 1$ ("success guaranteed"), then $P(X = n) = 1$ and $\mathsf{E}(X) = n$: X has a degenerate distribution in n.
- Similarly, if $p = 0$, then $P(X = n) = 0$ and $\mathsf{E}(X) = 0$.
- If $n = 1$, that is, if only one trial is conducted, X is said to have an alternative distribution with success probability p, which is a B$(1, p)$-distribution.

  It follows that $P(x = 1) = p$ and $P(X = 0) = 1 - p$, so:

$$
\begin{aligned}
\mathsf{E}(X) &= \sum_x xP(X=x) \\
&= 1\cdot p + 0\cdot(1-p) \\
&= p
\end{aligned}
$$

and:

$$
\begin{aligned}
\mathsf{E}\left(X^2\right) &= \sum_x x^2 P(X=x) \\
&= 1^2\cdot p + 0^2\cdot(1-p) \\
&= p
\end{aligned}
$$

We find: $\mathsf{var}(X) = \mathsf{E}(X^2) - (\mathsf{E}X)^2 = p(1-p)$, the variance of a $\mathsf{B}(1,p)$-distribution.

So-called cumulative binomial tables can be found at the end of the reader. They contain probabilities of the shape $P(X \leq c) = \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}$. Use the following formulas to calculate desired probabilities:

- $P(X \leq k)$: given
- $P(X = k) = P(X \leq k) - P(X \leq (k-1))$

4.5.2 The hypergeometric distribution

Definition 23
X is hypergeometrically distributed (with parameters N, R and n) if

$$
P(X=k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{m}}
$$

where $k = 0, 1, 2, \ldots n$.

Intuition: given a dichotomous population of N, where R have a certain people, draw n randomly without replacement. What is the probability that k have this property?

Property 23
For a hypergeometric distribution with parameters N, R, and n, in general:

$$
\mathsf{E}(X) = np \text{ and } \mathsf{var}(X) = np(1-p)\cdot\frac{N-n}{N-1}, \text{ where } p = \frac{R}{N}
$$

Property 24
For relatively large R and $N-R$ and relatively small n, the hypergeometric distribution with parameters N, R and n can be approximated by a $\mathsf{B}(n, \frac{R}{N})$-distribution.

4.5.3 The geometric distribution

Definition 24
X has a geometric distribution with parameter $p \in (0,1]$ if

$$
P(X=k) = (1-p)^{k-1} p, \text{ where } k = 1, 2, \ldots
$$

Intuition: what is the probability of success after k trials with probability p?

4.5.4 The Poisson distribution

Definition 25
X has a Poisson distribution with parameter $\mu > 0$ if

$$P(X = k) = \frac{\mu^k e^{-\mu}}{k!}, \text{ for } k = 0, 1, 2, \dots$$

Intuition: given something which occurs on average μ times in a given time period, what is the probability it will occur k times?

Property 25
IF X has a B(n, p)-distribution with "large n and small p", then X has approximately a Poisson distribution with parameter $\mu = np$.

Rule of Thumb 1
Apply Property 25 when $n > 25$ and $np < 10$ or $n(1-p) < 10$.

4.5.5 Summary: common distributions and their characteristics

Distribution	**Probability function**	**E**(X)	**var**(X)	**Example**
Homogeneous on $1, 2, \dots, N$	$P(X = x) = \frac{1}{N}$ $x = 1, 2, \dots N$	$\frac{N+1}{2}$	N/A	Result of one roll of a dice
Alternative (p)	$P(X = 1) = p$ $P(X = 0) = 1 - p$	p	$p(1-p)$	Dice result is 6 ($X = 1$) or not ($X = 0$)
Binomial B(n, p)	$P(X = x) = \binom{n}{x} p^x \, (1-p)^{n-x}$ $x = 0, 1, 2, \dots, n$	np	$np(1-p)$	Number of sixes in 30 rolls of a dice
Geometric (p)	$P(X = x) = (1-p)^{x-1} \, p$ $x = 1, 2, 3, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	Number of rolls of a dice until 6 occurs
Poisson (μ)	$P(X = x) = \frac{\mu^x e^{-\mu}}{x!}$ $x = 0, 1, \dots$	μ	μ	Number of clients that enter an office in 10 minutes
Hypergeometric $(R$, N, $n)$	$P(X = x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}$ $x = 0, 1, \dots, n$	np, $p = \frac{R}{N}$	$np(1-p) \cdot \frac{N-n}{N-1}$	Number of girls if we choose 5 from a group of 10 boys and 12 girls.

Chapter 5

Two or more discrete variables

5.1 Joint probability functions

Definition 26
If a pair (X, Y) of discrete random variables are defined on the same probability space and range $S_x \times S_y$, then for $(x, y) \in S_x \times S_y$:

$P(X = x$ and $Y = y)$ is the joint probability function of X and Y.

Property 26
For each joint probability function $P(X = x$ and $Y = y)$ of X and Y we have:

- $P(X = x$ and $Y = y \geq 0$
- $\sum_{x\in S_x}\sum_{y\in S_y} P(x = x$ and $Y = y) = 1$

Property 27 (Marginal probability functions $P(X = x)$ and $P(Y = y)$)

$$P(X=x)=\sum_{y\in S_y} P(X=x \text{ and } Y=y) \text{ and } P(Y=y)=\sum_{x\in S_x} P(X=x \text{ and } Y=y)$$

5.2 Conditional distributions

Definition 27
If X and Y are discrete random variables, then the conditional probability function of X, given $Y = y$, is defined by

$$P\left(X=x \mid Y=y\right)=\frac{P\left(X=x \text{ and } Y=y\right)}{P(Y=y)}, \text{ for } x\in S_x$$

Property 28
The conditional expectation X, given $Y = y$, is

$$\mathsf{E}(X \mid Y)=\sum_{x\in S_x} x\cdot P(X=x \mid Y=y)$$

Definition 28
If X and Y are discrete random variables, then the conditional expectation $\mathsf{E}(X \mid Y)$ is a random variable, that attains values with $\mathsf{E}(X \mid Y = y)$ with probability $P(Y = y)$.

Property 29

$$\mathsf{E}\left[\mathsf{E}\left(X \mid Y\right)\right]=\mathsf{E}\left(X\right)$$

Proof.

$$\begin{aligned}
\mathsf{E}\left[\mathsf{E}\left(X \mid Y\right)\right] &= \sum_{y} \mathsf{E}\left(X \mid Y = y\right) \cdot P(Y = y) \\
&= \sum_{y} \left(\sum_{x} x \cdot P\left(X = x \mid Y = y\right) \right) \cdot P\left(Y = y\right) \\
&= \sum_{X} \sum_{y} x \cdot \frac{P\left(X = x \text{ and } Y = y\right)}{P\left(Y = y\right)} \cdot P\left(Y = y\right) \\
&= \sum_{x} x \cdot \sum_{y} P\left(X = x \text{ and } Y = y\right) \\
&= \sum_{x} x P\left(X = x\right) \\
&= \mathsf{E}(X)
\end{aligned}$$

□

5.3 Independent random variables

Definition 29
Two discrete random variables X and Y are independent if

$$\forall_{(x,y) \in S_x \times S_y} : P(X = x \text{ and } Y = y) = P(X = x) \cdot P(Y = y)$$

This equality is known as the product rule for independent variables, and can be extended to n discrete random variables $X_1, X_2, \ldots, X_n$.

5.4 Functions of discrete random variables

Property 30 (Convolution sum)
If X and Y are independent discrete random variables, with integer numbered ranges, then:

$$P(X + Y = n) = \sum_{k \in S_x} P(X = k) \cdot P(Y = n - k)$$

Property 31
For a pair (X, Y) of discrete random variables we have:

$$\mathsf{E}g(X, Y) = \sum_{x \in S_x} \sum_{y \in S_y} g(x, y) P(X = x \text{ and } Y = y)$$

Property 32
For (discrete) random variables X and Y we have:

- $\mathsf{E}(X + Y) = \mathsf{E}(X) + \mathsf{E}(Y)$
- If X and Y are independent, then $\mathsf{E}(XY) = \mathsf{E}(X) \cdot \mathsf{E}(Y)$

Property 33

$$\mathsf{E}(X_1 + X_2 + \cdots + X_n) = \mathsf{E}(X_1) + \mathsf{E}(X_2) + \cdots + \mathsf{E}(X_n)$$

Property 34
If $X_1, X_2, \ldots, X_n$ are independent, then so are $g(X_1, \ldots, X_{n-1})$ and X_n.

Property 35
If $X_1, X_2, \ldots, X_n$ are independent and X_i has a Poisson distribution with parameter μ_i for each $i = 1, 2, \ldots, n$, then:

$$\sum_{i=1}^{n} X_i \text{ has a Poisson distribution with parameter } \mu = \sum_{i=1}^{n} \mu_i$$

5.5 Corollation

Definition 30
The covariance of two random variables X and Y is defined as

$$\mathsf{cov}(X,Y) = \mathsf{E}(X - \mu_x)(Y - \mu_y)$$

Property 36
Properties of the covariance:

- $\mathsf{cov}(X,Y) = \mathsf{E}(XY) - \mu_x\mu_y$
- If X and Y are independent, then $\mathsf{cov}(X,Y) = 0$

Property 37
Properties of the covariance:

- $\mathsf{cov}(X,X) = \mathsf{var}(X)$
- $\mathsf{cov}(X,Y) = \mathsf{cov}(Y,X)$
- $\mathsf{cov}(aX+b,Y) = a \cdot \mathsf{cov}(X,Y)$, for $a \in \mathbb{R}$ and $b \in \mathbb{R}$
- $\mathsf{cov}(X+Y,Z) = \mathsf{cov}(X,Z) + \mathsf{cov}(Y,Z)$
- $\mathsf{var}(X+Y) = \mathsf{var}(X) + \mathsf{var}(Y) + 2\mathsf{cov}(X,Y)$ and

  $\mathsf{var}(X-Y) = \mathsf{var}(X) + \mathsf{var}(Y) - 2\mathsf{cov}(X,Y)$ and

Definition 31
The correlation coefficient $\rho(X,Y)$ of two random variables X and Y is defined by

$$\rho(X,Y) = \frac{\mathsf{cov}(X,Y)}{\sigma_x\sigma_y}$$

Property 38
Properties of a correlation coefficient:

- $\rho(aX+b,Y) = \begin{cases} \rho(X,Y) & \text{if } a > 0 \\ -\rho(X,Y) & \text{if } a < 0 \end{cases}$
- $-1 \leq \rho(X,Y)$
- If $Y = aX + b$, then $\rho(X,Y) = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{if } a < 0 \end{cases}.$

  And reversely, if $\rho(X,Y) = 1$, then $Y = aX + b$, with $a > 0$,

  and if $\rho(X,Y) = -1$, then $Y = aX + b$, with $a < 0$

Property 39
Properties for the variance of a sum of variables:

- $\mathsf{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathsf{var}(X_i) + \sum\sum_{i \neq j} \mathsf{cov}\left(X_i, X_j\right)$
- If $X_1, \ldots, X_n$ are independent, then $\mathsf{var}\left(\sum_{i=1}^{n}\right) = \sum_{i=1}^{n} \mathsf{var}\left(X_i\right)$

5.6 The weak law of large numbers

Property 40
If $X_1, X_2, \ldots$ are independent and all have the same distribution with expactation μ and variance σ^2, then for the mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and every constant $c > 0$ we have:

$$\lim_{n\to\infty} P\left(|\overline{X}_n| \geq c\right) = 0$$

In words: $\overline{X}_n$ converges in probability to μ.

Chapter 6

Continuous random variables

6.1 Density function, expectation and variance of a continuous variable

Definition 32
The density function of a continuous random variable X is a non-negative function f, such that $P(a \leq X \leq b) = \int_a^b f(x)dx$.

Property 41
f is a density function if

1. $f(x) \geq 0$ and
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

i.e. the probability is never negative, and the total probability is 100 %.

Definition 33
The expectation (expected value) of a continuous random variable X is

$$\mathsf{E}(x) = \inf_{-\infty}^{\infty} x f(x)dx$$

provided that this integral is absolute convergent: $\int_{-\infty}^{\infty} |x| \cdot f(x)dx < \infty$.

Property 42
For every real valued function g we have $\mathsf{E}g(x) = \int_{-\infty}^{\infty} g(x)f(x)dx$.

6.2 Distribution function

Definition 34
The function F, defined by $F(x) = P(X \leq x)$ with $x \in \mathbb{R}$, is the (cumulative) distribution function (cdf) of the random variable X.

Property 43
For any distribution function $F(x)$ of a random variable X, we have:

1. F is non-decreasing (if $x_2 > x_1$, then $F(x_2) \geq F(x_1)$.
2. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$.
3. F is continuous from the right ($\lim_{h \to 0^+} F(x+h) = F(x)$.

Property 44
For any distribution function $F(x)$ of a random variable X, we have:

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

2. $P(X > x) = 1 - F(x)$
3. $P(X < x) = \lim_{h \to 0^+} F(x-h) = F(x)$
4. $P(X = x) = F(x) - P(X < x)$

Definition 35
A random variable X is continuous if the distribution function F of X is a continuous function.

Property 45
For a continuous random variable X with density function f and (cumulative) distribution function F we have:

1. $P(X = x) = 0$, for $x \in \mathbb{R}$
2. $P(X \in [a, b]) = \int_a^b f(x)dx = F(b) - F(a)$
3. $F(x) = \int_{-\infty}^{x} f(u)du$
4. $f(x) = \frac{d}{dx}F(x)$
5. If the density function $f(x)$ of X is symmetric about $x = c$, then $\mathsf{E}(X) = c$ (provided that $\mathsf{E}(X)$ exists)

6.3 The uniform, exponential and standard normal distributions

6.3.1 The uniform distribution on the interval $[a, b]$

Definition 36
The random variable X has a uniform distribution on the interval $[a, b]$ if

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{for } x \notin [a, b] \end{cases}$$

Notation: $X \sim U(a, b)$.

Property 46
The expectation and variance of the uniform distribution on $[a, b]$ are:

1. $\mathsf{E}(X) = \frac{a+b}{2}$
2. $\mathsf{var}(X) = \frac{(b-a)^2}{12}$

6.3.2 The exponential distribution with parameter λ

Definition 37
The random variable X has an exponential distribution parameter $\lambda > 0$ if

$$f(x) = \begin{cases} \lambda e^{-lambdax} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Notation: $X \sim \mathsf{Exp}(\lambda)$

Property 47
If X is an exponentially distributed variable with parameter λ, then:

1. $P(X > x) = e^{-\lambda x}$, for $x \geq 0$
2. $F(X) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$
3. $\mathsf{E}(X) = \frac{1}{\lambda}$
4. $\mathsf{var}(X) = \frac{1}{\lambda^2}$

6.3.3 The standard normal distribution

Definition 38
The continuous random variable Z has a standard normal distribution if

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \text{ where } z \in \mathbb{R}$$

Notation: $Z \sim N(0, 1)$.

Furthermore, $\mu = 0$ and $\sigma^2 = 1$.

6.4 Functions of a continuous normal variable

Property 48
If the continuous random variable X has a density function f_x, then for $Y = aX + b$, with $a \neq 0$:

$$f_Y(y) = \frac{1}{|a|} f_x \left(\frac{y-b}{a} \right)$$

Property 49
If X has a uniform distribution on $(0, 1)$, then $Y = -\frac{\ln(X)}{\lambda}$ has exponential distribution with parameter $\lambda > 0$.

6.5 The normal distribution $N(\mu, \sigma^2)$

Definition 39
The random variable X has a normal distribution with parameters μ and σ^2 if the density function of X is defined by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ with } x \in \mathbb{R}$$

Short notation: $X \sim N(\mu, \sigma^2)$

Property 50
If $X \sim N(\mu, \sigma^2)$, and $Z \sim N(0, 1)$, then:

1. $\sigma Z + \mu \sim N(\mu, \sigma^2)$
2. The z-score $\frac{X-\mu}{\sigma} \sim N(0, 1)$
3. E$(X) = \mu$
4. var$(X) = \sigma^2$

Property 51
For a $N(\mu, \sigma^2)$-distributed random variable X, we have:

$Y = aX + b$ is $N(a\mu + b, a^2\sigma^2)$-distributed (for all $a \neq 0$ and $b \in \mathbb{R}$).

6.6 Overview of frequently used continuous distributions

Distribution	**Density function**	**E**(X)	**var**(X)
Uniform $U(a, b)$	$f(x) = \frac{1}{b-a}, a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential Exp(λ)	$f(x) = \lambda e^{-\lambda x}, x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal $N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in \mathbb{R}$	μ	σ^2

The following relations between distributions can be given:

- Standardisation: if $X \sim N\left(\mu, \sigma^2\right)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- Link between the standard normal and normal distribution: if $Z \sim N(0,1)$, then $X = \sigma Z + \mu \sim N\left(\mu, \sigma^2\right)$
- An exponential distribution can be simulated with random numbers between 0 and 1: if $X \sim U(0,1)$, then $Y = -\frac{\ln(X)}{\lambda} \sim \text{Exp}(\lambda)$
- Relation between $U(0,1)$ and $U(a,b)$: if $X \sim U(0,1)$, then $Y = (b-a)X + a \sim U(a,b)$

Chapter 7

Two or more continuous variables

7.1 Independence

Definition 40
The random variables X and Y are independent if for each pair of sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ we have:

$$P(X \in A \wedge Y \in B) = P(X \in A) \cdot P(Y \in B)$$

7.2 The convolution integral

Property 52 (The convolution integral)
If X and Y are independent continuous variables, we have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

7.3 The sum of independent and normally distributed variables

Property 53
If $X_i \sim N\left(\mu, \sigma^2\right)$ for $i = 1, 2, \ldots, n$, and $X_1, X_2, \ldots, X_n$ are independent, and $S_n = \sum_{i=1}^n X_i$, then:

1. $\mathsf{E}\left(S_n\right) = \sum_{i=1}^n \mu_i$
2. $\mathsf{var}\left(S_n\right) = \sum_{i=1}^n \sigma_i^2$
3. $S_n \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$

Property 54
If $X_i \sim N\left(\mu, \sigma^2\right)$ for $i = 1, 2, \ldots, n$, and $X_1, X_2, \ldots, X_n$ are independent, then we have

1. if $S_n = \sum_{i=1}^n X_i$, then $S_n \sim N\left(n\mu, n\sigma^2\right)$ and
2. if $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, then $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

7.4 The Central Limit Theorem

Property 55 (The Central Limit Theorem)
If $X_1, X_2, \ldots$ is a sequence of independent, identically distributed variables, with expectation μ and variance $\sigma^2 > 0$, then for $S_n = \sum_{i=1}^n X_i$ we have:

$$\lim_{n \to \infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) = \Phi(z)$$

where Φ is the standard normal distribution function.

Consequence. If n is "sufficiently large", then:

- $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ is approximately $N(0,1)$-distributed.
- S_n is approximately $N\left(n\mu, n\sigma^2\right)$-distributed.
- $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is approximately $N\left(\mu, \frac{\sigma^2}{n}\right)$-distributed.

Rule of Thumb 2
When $n \geq 25$, we can use the Central Limit Theorem for normal approximation.

Property 56 (Consequence of the CLT: normal approximation of the binomial distribution)
If $X \sim B(n,p)$, then, for sufficiently large n, X is approximately $N\left(np, np(1-p)\right)$

Rule of Thumb 3
To approximate binomial probabilities:

- Use the Central Limit Theorem when $n \geq 25$
- Use the Poisson-approximation with $\mu = np$, if $np \leq 10$ or $n(1-p) \leq 10$
- Use the normal approximation according to the CLT with $\mu = np$ and $\sigma^2 = np(1-p)$ if $np > 5$ and $n(1-p) > 5$

Chapter 8

Waiting times

8.1 Waiting time distributions and the lack of memory property

Definition 41
The distribution of a random variable X has the lack of memory property on its range S_x, if for all $t, s \in S_x$:

$$P(X > t + s \mid X > s) = P(X > t)$$

Property 57
For a continuous random variable X with range $S_x = [0, \infty)$ the following statements are equivalent:

1. X is exponentially distributed parameter λ
2. $P(X > t) = e^{-\lambda t}$, for $t \geq 0$
3. The distribution of X has the lack of memory property on S_x and $\mathsf{E}(X) = \frac{1}{\lambda}$

Property 58
For a discrete random variable X with range $S_X = \{1, 2, \dots\}$ the following statements are equivalent:

1. X is geometrically distributed with parameter p
2. $P(X > n) = (1 - p)^n$, for $n = 0, 1, 2, \dots$
3. The distribution of X has the lack of memory property on S_X and $p = P(X = 1)$

8.2 Summation of independent waiting times

Definition 42
X has an Erlang distribution with parameters n and λ, if

$$f_X(x) = \frac{\lambda\,(\lambda x)^{n-1}\,e^{-\lambda x}}{(n-1)!} \text{ for } x \geq 0 \text{ and } f_X(x) = 0 \text{ for } x < 0$$

Notation: $X \sim \text{Erlang}(n, \lambda)$

Property 59
If $X_1, X_2, \dots$ are independent and exponentially distributed with parameter λ, then:

$$S_n = \sum_{i=1}^{n} X_i \sim \text{Erlang}(n, \lambda)$$

Appendix A

Formula sheet Probability Theory for BIT and TCS in module 4

Distribution	**E** (X)	**var** (X)
Geometric	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometric	$n \cdot \frac{R}{N}$	$n \cdot \frac{R}{N} \cdot \frac{N-R}{N} \cdot \frac{N-n}{N-1}$
Poisson $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$, $x = 0, 1, 2, \ldots$	μ	μ
Uniform on (a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlang $f_X(x) = \frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!}$, $x \geq 0$	$\frac{n}{\delta}$	$\frac{n}{\lambda^2}$

$$\mathsf{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathsf{var}\,(X_i) + \sum_{i \neq j}\sum \mathsf{cov}\,(X_i, X_j)$$