Probability: Handout

Klaus Pötzelberger
Vienna University of Economics and Business
Institute for Statistics and Mathematics
E-mail: Klaus.Poetzelberger@wu.ac.at

Contents

Chapter 1

Axioms of Probability

1.1 Definitions and Properties

There is a long history of attempts to define probability. Kolmogorov (1933) stated the so-called axioms of probability, i.e. a generally accepted small list of properties of mappings that deserve the name probability. The idea may be illustrated by assuming that a mechanism generates outcome (data) randomly. A potential outcome ω is an element of a general set Ω. The outcome is random, i.e. not certain in an astonishingly hard to define sense, and therefore only the probability (likelihood) that the outcome is in a certain subset $A \subseteq \Omega$ can be specified.

The components for the definition of probability are thus: 1. a set (*space*) Ω, 2. a collection $\mathcal{A}$ of subsets of Ω (if $A \in \mathcal{A}$, then $A \subseteq \Omega$), the set of *events*, and 3. a mapping $P : \mathcal{A} \mapsto [0, 1]$, the *probability*, i.e. for an event $A \in \mathcal{A}$, $P(A)$ is the probability of A.

Definition 1.1 *A subset $\mathcal{A} \subseteq 2^\Omega$ is called* σ-algebra*, if*

1. *$\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$,*
2. *if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$,*
3. *if $A_1, A_2, \ldots$ is a countable sequence with $A_i \in \mathcal{A}$, then $\cup_{i=1}^\infty A_i \in \mathcal{A}$ and $\cap_{i=1}^\infty A_i \in \mathcal{A}$.*

Remark 1.2 *1. 2^Ω denotes the power set of Ω, i.e. $2^\Omega = \{A \mid A \subseteq \Omega\}$. 2^Ω is also denoted by $\mathcal{P}(\Omega)$.*

2. A^c denotes the complement of A, i.e. $A^c = \Omega \setminus A = \{\omega \in \Omega \mid \omega \notin A\}$.

3. If $\mathcal{A} \subseteq 2^\Omega$ satisfies 1. 2. of the definition, but is closed only with respect to finite unions and intersections, it is called an algebra*.*

4. Let $\mathcal{A} \subseteq 2^\Omega$ satisfy 1. and 2. of the definition. Assume that $\mathcal{A}$ is closed with respect to countable unions, then it is closed with respect to countable intersections. Also, if $\mathcal{A}$ is closed with respect to countable intersections, then it is closed with respect to countable unions. This

follows from De Morgan's law: $(A \cap B)^c = A^c \cup B^c$ *and* $(A \cup B)^c = A^c \cap B^c$. *More generally,* $(\cap_{i=1}^{\infty} A_i)^c = \cup_{i=1}^{\infty} A_i^c$ *and* $(\cup_{i=1}^{\infty} A_i)^c = \cap_{i=1}^{\infty} A_i^c$. *Therefore, to check that a collection* $\mathcal{A}$ *of subsets of* Ω *is a* σ*-algebra, it is sufficient to check 1., 2. and either that* $\mathcal{A}$ *is closed w.r.t countable unions or countable intersections.*

Definition 1.3 *Let* $\mathcal{C} \subseteq 2^{\Omega}$. $\sigma(\mathcal{C})$ *is the* σ*-algebra generated by* $\mathcal{C}$*, i.e. the smallest* σ*-algebra on* Ω *containing* $\mathcal{C}$. *That is,* $\sigma(\mathcal{C})$ *is a* σ*-algebra on* Ω, $\mathcal{C} \subseteq \sigma(\mathcal{C})$ *and if* $\mathcal{A}$ *is a* σ*-algebra on* Ω *with* $\mathcal{C} \subseteq \mathcal{A}$, *then* $\sigma(\mathcal{C}) \subseteq \mathcal{A}$.

Example 1.4

1. $\mathcal{A} = \{\emptyset, \Omega\}$ *is a* σ*-algebra on* Ω. *It is called the trivial* σ*-algebra. It is the smallest* σ*-algebra on* Ω.
2. *Let* $A \subseteq \Omega$. *Then* $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$.
3. *Let* $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ *be a partition of* Ω, *i.e.* $\cup_{i=1}^{n} C_i = \Omega$ *and* $C_i \cap C_j = \emptyset$ *for* $i \neq j$. *Then* $\sigma(\mathcal{C}) = \{\cup_{i \in I} C_i \mid I \subseteq \{1, \ldots, n\}, I \neq \emptyset\}$.
4. *Let* $\mathcal{O} \subseteq \Omega$ *be the set of open subsets of* Ω. $\mathcal{B} = \sigma(\mathcal{O})$ *is called the Borel* σ*-algebra on* Ω.

Remark 1.5 *Borel* σ*-algebras* $\mathcal{B}$ *are especially important in probability theory. For instance, if* $\Omega = \mathbb{R}$, *then the Borel* σ*-algebra* $\mathcal{B}$ *is the smallest* σ*-algebra containing all open intervals. However, there are many more Borel sets (elements of* $\mathcal{B}$*) than open intervals. Unions of open intervals are Borel sets, complements of open sets (i.e. closed sets) are Borel sets. Countable unions of open and closed sets and their complements are Borel sets. Countable unions and intersections of these sets are Borel sets.*

Definition 1.6 *A probability measure* P *defined on a* σ*-algebra* $\mathcal{A}$ *is a function:* $P : \mathcal{A} \to [0, 1]$ *that satisfies:*

1. $$P(\Omega) = 1,\ P(\emptyset) = 0. \tag{1.1}$$
2. *For every countable and pairwise disjoint sequence* $(A_n)_{n=1}^{\infty}$ *of events* ($A_n \in \mathcal{A}$ *and* $A_n \cap A_m = \emptyset$ *for* $n \neq m$),
$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n). \tag{1.2}$$

Remark 1.7 *1. A probability measure is also called probability distribution or only distribution or probability.*

2. A probability measure is a special case of a measure: A measure m *is a mapping* $m : \mathcal{A} \to [0, \infty]$ *satisfying* (1.2). *A finite measure* m *is a measure with* $m(\Omega) < \infty$. *Let* Ω *be a general set,* $\mathcal{A}$ *a* σ*-algebra on* Ω. $(\Omega, \mathcal{A})$ *is called measurable space. If* m *is a measure on* $(\Omega, \mathcal{A})$, *then the triple*

$(\Omega, \mathcal{A}, m)$ is called a measure space. In case the measure is a probability measure P, $(\Omega, \mathcal{A}, P)$ is called probability space.
3. A mapping $m : \mathcal{A} \to [0, \infty]$ is called additive, if condition (1.2) is replaced by $m(\bigcup_{n=1}^{k} A_n) = \sum_{n=1}^{k} m(A_n)$ (for finitely many pairwise disjoint events A_n).

Proposition 1.8 *Let P be a probability measure on a σ-algebra $\mathcal{A}$. Then*

1. *P is additive.*
2. *$P(A^c) = 1 - P(A)$ for $A \in \mathcal{A}$.*
3. *If $A, B \in \mathcal{A}$ and $A \subseteq B$, then $P(A) \leq P(B)$.*

Proof. 1. Let $A_1, \ldots, A_k \in \mathcal{A}$ be pairwise disjoint. We define $A_n = \emptyset$ for $n > k$. Then, $(A_n)_{n=1}^{\infty}$ is a sequence of pairwise disjoint events and since $P(\emptyset) = 0$, we have

$$P(\bigcup_{n=1}^{k} A_n) = P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{k} P(A_n).$$

2. Since $\Omega = A \cup A^c$ we have $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$.
3. Let $A \subseteq B$ be events. Since $B = A \cup (B \cap A^c)$ we have

$$P(B) = P(A \cup (B \cap A^c) = P(A) + P(B \cap A^c) \geq P(A). \qquad \square$$

Theorem 1.9 *Let $\mathcal{A}$ be a σ-algebra and suppose that $P : \mathcal{A} \to [0, 1]$ satisfies (1.1) and is additive. Then the following statements are equivalent:*

1. *(1.2), i.e. P is a probability measure.*
2. *If $B_n \in \mathcal{A}$ is increasing, i.e. $B_n \subseteq B_{n+1}$ for all n and $B = \bigcup_{n=1}^{\infty} B_n$, then $P(B_n) \uparrow P(B)$.*
3. *If $C_n \in \mathcal{A}$ is decreasing, i.e. $C_{n+1} \subseteq C_n$ for all n and $C = \bigcap_{n=1}^{\infty} C_n$, then $P(C_n) \downarrow P(C)$.*

Proof. If (B_n) is an increasing sequence of events, if and only if $(C_n) = (B_n^c)$ is a decreasing sequence of events. Since $P(C_n) = 1 - P(B_n)$ in this case, $P(B_n) \uparrow P(B)$ if and only if $P(C_n) \downarrow P(C)$. Therefore, 2. and 3. are equivalent. We have to show that 1. $\Rightarrow$ 2. and 2. $\Rightarrow$ 1.

1. $\Rightarrow$ 2.: Let (B_n) be an increasing sequence of events. We define $A_n = B_n \cap B_{n-1}^c$. (A_n) is then a pairwise disjoint sequence of events, $B_n = \bigcup_{i=1}^{n} A_i$ and $B = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Therefore, since (1.2) holds and P is additive, we have

$$P(B) = \sum_{n=1}^{\infty} P(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_n) = \lim_{n\to\infty} P(B_n).$$

Clearly, $P(B_n)$ is increasing (nondecreasing).

2. $\Rightarrow$ 1.: Let (A_n) be a pairwise disjoint sequence of event. We define $B_n = \bigcup_{i=1}^{n} A_i$ and $B = \bigcup_{n=1}^{\infty} B_n$ $(= \bigcup_{n=1}^{\infty} A_n)$. B_n is increasing and therefore

$$P(\bigcup_{n=1}^{\infty} A_n) = P(B) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{n=1}^{\infty} P(A_n). \qquad \square$$

Remark 1.10 *What is the reason that one has to introduce the concept of a σ-algebra? Would not be simple to take always $\mathcal{A} = 2^{\Omega}$, i.e. all subsets of Ω are events?*
1. First of all there are cases when in fact one can choose 2^{Ω} as the set of events. This is possible, when either Ω is of a simple structure (for instance, when Ω is countable) or when the probability distribution P is simple.
2. It is perhaps astonishing that one cannot define probability distributions with certain desirable properties on the power set 2^{Ω}. For instance, one can prove that if P is the uniform distribution on $\Omega = [0,1]$ (intervals $[a,b] \subseteq [0,1]$ have probability $b-a$), it can be defined on the Borel sets, but not extended to 2^{Ω}.
3. The σ-algebra $\mathcal{A}$ can be considered as a measure of information. If $\mathcal{A}$ is small, there are only few events on which probability statements are possible. If $\mathcal{A}$ has many elements, there are many events that can be distinguished. Let us consider the following example: Assume that the outcome of the random experiment ω is an element of $\Omega = \{0, 1, \ldots, 5\}$. Person A has access to the outcome. Accordingly, $\mathcal{A}$ could be 2^{Ω}, which is also generated by $\{0\}, \{1\}, \ldots, \{5\}$. To person B only $(\omega - 2)^2$ is reported. Person B cannot distinguish between 1 and 3 and between 0 and 4. The appropriate σ-algebra is the one generated by $\{2\}, \{5\}, \{1,3\}, \{0,4\}$.

1.2 Independence

Let a probability space $(\Omega, \mathcal{A}, P)$ be given. Think of $\omega \in \Omega$ as the outcome of a random experiment. Let B be an event with $P(B) > 0$. Then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{1.3}$$

is the probability of A given B, i.e. the likelihood of $\omega \in A$ among all $\omega \in B$. Note that $P(A \mid B) = P(A)$ if and only if $P(A \cap B) = P(A)P(B)$. We call A and B (stochastically) independent.

Definition 1.11 *A collection of events $(A_i)_{i\in I}$ is a collection of independent events if for all finite $J \subseteq I$*

$$P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i). \tag{1.4}$$

Example 1.12 *A card is chosen randomly from a deck of 52 cards, each card chosen with the same probability. Let A be the event that the chosen card is a Queen and B that the card is a heart. Then $P(A) = 1/13, P(B) = 1/4, P(A \cap B) = 1/52 = P(A)P(B)$. Therefore, A and B are independent.*

Example 1.13 *Let $\Omega = \{1,2,3,4\}$, $\mathcal{A}$ the power set of Ω and $P(\{i\}) = 1/4$ for $i = 1,2,3,4$. Let $A = \{1,2\}$, $B = \{1,3\}$, $C = \{2,3\}$. Then $P(A) = P(B) = P(C) = 1/2$ and $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$. Therefore, A, B and C are pairwise independent. However,*

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq \frac{1}{8} = P(A)P(B)P(C).$$

A, B and C are not independent!

Proposition 1.14 *Let a probability space $(\Omega, \mathcal{A}, P)$ be given and $A, B \in \mathcal{A}$.*
1. A and B are independent if and only if the pairs A and B^c, A^c and B, and A^c and B^c are independent.
2. A is independent of A^c if and only if $P(A) = 0$ or $P(A) = 1$.
3. Let $P(B) > 0$. $P(A \mid B) = P(A)$ if and only if A and B are independent.
4. Let $P(B) > 0$. The mapping $P(. \mid B) : \mathcal{A} \to [0,1]$, given by (1.3) is a probability measure on $\mathcal{A}$.

Proof. See exercises.

Proposition 1.15 *Let $A_1, \ldots, A_n$ be events. Then, if $P(A_1 \cap A_1 \cap \cdots \cap A_{n-1}) > 0$*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap A_1 \cap \cdots \cap A_{n-1}).$$

Proof. See exercises.

Theorem 1.16 *(Partition Equation). Let (E_m) be a finite or countable partition of Ω with $P(E_m) > 0$ for all m. Then,*

$$P(A) = \sum_m P(A \mid E_m)P(E_m). \tag{1.5}$$

Proof. We have $P(A \mid E_m)P(E_m) = P(A \cap E_m)$ and $A = \bigcup_m A \cap E_m$. Since the sets $A \cap E_m$ are pairwise disjoint, (1.5) follows. □

Theorem 1.17 *(Bayes' Theorem). Let (E_n) be a finite or countable partition of Ω with $P(E_n) > 0$ for all n. Then, if $P(A) > 0$,*

$$P(E_n \mid A) = \frac{P(A \mid E_n)P(E_n)}{\sum_m P(A \mid E_m)P(E_m)}. \tag{1.6}$$

Proof. The r.h.s. of (1.6) is

$$\frac{P(A \mid E_n)P(E_n)}{P(A)} = \frac{P(A \cap E_n)}{P(A)} = P(E_n \mid A).$$ □

1.3 Exercises

Exercise 1.1 *Let $\Omega = \mathbb{R}$. Show that all one-point sets $A = \{a\}$ are Borel sets. Show that therefore all countable subsets of $\mathbb{R}$ and subsets with countable complement are Borel sets.*

Exercise 1.2 *Let $\mathcal{B}$ be the Borel σ-algebra on $\mathbb{R}$. Show that it is generated also by*

1. *All open intervals.*
2. *All closed subsets of $\mathbb{R}$.*
3. *All closed intervals.*
4. *All half-open intervals $\{(a, b] \mid a < b\}$.*
5. *The collection of intervals $\{(-\infty, b] \mid b \in \mathbb{R}\}$.*
6. *The collection of intervals $\{(-\infty, b] \mid b \in \mathbb{Q}\}$.*

Exercise 1.3 *Let $\mathcal{C}$ be a collection of subsets of Ω. Show that $\sigma(\mathcal{C})$ exists. Hint: Prove that*
1. There is at least one σ-algebra $\mathcal{A}$ with $\mathcal{C} \subseteq \mathcal{A}$.
2. If $\{\mathcal{A}_i\}, i \in I$ is a collection of σ-algebras on Ω, then $\bigcap_{i\in I} \mathcal{A}_i$ is a σ-algebra.
3. $\sigma(\mathcal{C}) = \bigcap\{\mathcal{A} \mid$ is a σ-algebra on Ω with $\mathcal{C} \subseteq \mathcal{A}\}$.

Exercise 1.4 *Let $\mathcal{A}$ be a σ-algebra on Ω and $B \in \mathcal{A}$. Prove that $\mathcal{G} = \{A \cap B \mid A \in \mathcal{A}\}$ is a σ-algebra on B.*

Exercise 1.5 *Let $\Omega = [-1, 1]$, $\mathcal{A} = \{\emptyset, \Omega, [-1, 1/2], (1/2, 1]\}$. Let $X : \Omega \to [0, 1]$ be given by $X(\omega) = \omega^2$. Let $\mathcal{F} = \{X(A) : A \in \mathcal{A}\}$. Check whether $\mathcal{F}$ is a σ-algebra on $[0, 1]$.*

Exercise 1.6 *(Subadditivity). Let $(\Omega, \mathcal{A}, P)$ be a probability space and (A_i) a sequence of events. Show that*

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

for all n and

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

Exercise 1.7 *Let A and B be events. Show that $P(A) + P(B) = P(A \cup B) + P(A \cap B)$.*

Exercise 1.8 *Let A and B be events. Show that if A and B are independent and $A \cap B = \emptyset$, then $P(A) = 0$ or $P(B) = 0$.*

Exercise 1.9 *Prove Proposition 1.14.*

Exercise 1.10 *Prove Proposition 1.15.*

Exercise 1.11 *An insurance company insured an equal number of male and female drivers. In every year a male driver has an accident involving a claim with probability α, a female driver with probability β, with independence over the years. Let A_i be the event that a selected driver makes a claim in year i.*

1. *What is the probability that a selected driver will make a claim this year?*
2. *What is the probability that a selected driver will make a claim in two consecutive years?*
3. *Show that $P(A_2 \mid A_1) \geq P(A_1)$ with equality only if $\alpha = \beta$.*
4. *Find the probability that a claimant is female.*

Chapter 2

Probabilities and Random Variables on Countable Spaces

2.1 Discrete Probabilities

In this chapter Ω is finite or countable. We have $\mathcal{A} = 2^\Omega$. Since Ω is at most countable, P is completely determined by the sequence $p_\omega = P(\{\omega\}), \omega \in \Omega$.

Theorem 2.1 *Let (p_ω) be a sequence of real numbers. There exists a unique probability measure P on $(\Omega, 2^\Omega)$ such that $P(\{\omega\}) = p_\omega$ if and only if $p_\omega \geq 0$ (for all ω) and $\sum_\omega p_\omega = 1$. In this case for $A \subseteq \Omega$,*

$$P(A) = \sum_{\omega \in A} p_\omega.$$

Example 2.2 *(Uniform distribution.) If Ω is finite ($|\Omega| < \infty$), we define the uniform distribution by choosing $p_\omega = 1/|\Omega|$. Then for $A \subseteq \Omega$,*

$$P(A) = \frac{|A|}{|\Omega|}. \tag{2.1}$$

The uniform distribution is sometimes called the method of Laplace. The ratio (2.1) is referred to as number of favourable cases ($\omega \in A$) over number of possible cases ($\omega \in \Omega$).

Example 2.3 *(Hypergeometric distribution $H(N, M, n)$.) Consider an urn containing two sets of objects, N red balls and M blue balls. n balls are drawn without replacement, each ball has the same probability of being drawn. The hypergeometric distribution is the probability distribution of the number X of red balls drawn: Let $N \geq 1, M \geq 1$, $1 \leq n \leq N + M$ and $0 \leq k \leq n$. What is the probability that k of the n drawn balls are red? Let $\Omega = \{0, 1, \ldots, n\}$.*

First we specify an auxiliary random experiment: We assume that the balls in the urn are labelled, with labels 1 up to N corresponding to "red" and $N+1$ up to $N+M$ corresponding to "blue".

Define $\tilde{\Omega}$ as the set of subsets of $\{1, 2, \ldots, M+N\}$ of size n. The auxiliary experiment picks subsets A of size n uniformly from $\tilde{\Omega}$. The number of red elements of A is then hypergeometrically distributed.

The number of subsets of $\{1, \ldots, N, N+1, \ldots N+M\}$ of size n is

$$\binom{N+M}{n},$$

the proportion with $|A \cap \{1, \ldots, N\}| = k$ and $|A \cap \{N+1, \ldots, N+M\}| = n-k$ is

$$P(X = k) = \frac{\binom{N}{k}\binom{M}{n-k}}{\binom{N+M}{n}}.$$

Example 2.4 *(Binomial distribution $B(n,p)$.) Again, two types of objects are sampled (1 for "success" and 0 for "failure"), but this time with replacement. The proportion of 1's is p, with $0 \leq p \leq 1$. The probability of drawing a 1 is thus p. X is the number of successes among the n draws. What is $P(X = k)$ (for $0 \leq k \leq n$)? The range of X is $\Omega = \{0, 1, \ldots, n\}$.*

First, we specify the data generating process. If the experiment is repeated n times, this leads to a sequence $\omega = (\omega_1, \ldots, \omega_n) \in \{0,1\}^n$.

Let $x = x(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n$. Then

$$P(\{\omega\}) = p^x (1-p)^{n-x}.$$

There are

$$\binom{n}{k}$$

elements of ω with $x(\omega) = k$ and thus

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2.2}$$

for $0 \leq k \leq n$ and $P(X = k) = 0$ else.

Example 2.5 *(Poisson distribution $P(\lambda)$.) The Poisson distribution is defined on $\Omega = \mathbb{N}_0 = \{0, 1, \ldots\}$. $\lambda > 0$ is the parameter and*

$$P(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}. \tag{2.3}$$

Example 2.6 *(Geometric distribution $G(\theta)$.) The geometric distribution is defined on $\Omega = \mathbb{N} = \{1, 2, \ldots\}$. $\theta \in (0,1)$ is the parameter and*

$$P(\{n\}) = (1-\theta)\theta^{n-1}. \tag{2.4}$$

2.2 Random Variables on Countable Spaces

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(T, \mathcal{C})$ a measurable space. Furthermore, let $X : \Omega \to T$ be a mapping. We are interested in probabilities $P(X \in C)$ for $C \in \mathcal{C}$. Since $\{X \in C\}$ is short for $X^{-1}(C) = \{\omega \in \Omega \mid X(\omega) \in C\}$, the mapping X has to be measurable in the sense of the following definition:

Definition 2.7 *Let $(\Omega, \mathcal{A})$ and $(T, \mathcal{C})$ be two measurable spaces and $X : \Omega \to T$ a mapping. X is measurable (w.r.t. $(\Omega, \mathcal{A})$ and $(T, \mathcal{C})$) if $X^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$. If $(\Omega, \mathcal{A}, P)$ is (even) a probability space, we call X random variable.*

$X : (\Omega, \mathcal{A}) \to (T, \mathcal{C})$ is short for $X : \Omega \to T$ and it is measurable w.r.t. $(\Omega, \mathcal{A})$ and $(T, \mathcal{C})$.

If $\mathcal{A} = 2^\Omega$, then any mapping to a measurable space is measurable. For general spaces this is not the case.

Theorem 2.8 *Let $X : \Omega \to T$ be a mapping and assume that T is endowed with a σ-algebra $\mathcal{C}$. Let*

$$\sigma(X) = \{X^{-1}(C) \mid C \in \mathcal{C}\}. \tag{2.5}$$

$\sigma(X)$ is a σ-algebra on Ω and it is the smallest σ-algebra $\mathcal{A}$ s.t. $X : (\Omega, \mathcal{A}) \to (T, \mathcal{C})$.

Proof. If $\sigma(X)$ is a σ-algebra on Ω, it is, by its construction, automatically the smallest σ-algebra that makes X measurable. Therefore, we only have to prove that it is a σ-algebra. We have

1. $\emptyset = X^{-1}(\emptyset)$ and $\Omega = X^{-1}(T)$, therefore $\emptyset, \Omega \in \sigma(X)$.
2. Let $A \in \sigma(X)$, i.e. $A = X^{-1}(C)$ for a $C \in \mathcal{C}$. Since $C^c \in \mathcal{C}$ and
$$A^c = X^{-1}(C)^c = X^{-1}(C^c),$$
$A^c \in \sigma(X)$ follows.
3. Let (A_i) be a sequence with $A_i \in \sigma(X)$. There are $C_i \in \mathcal{C}$ s.t. $A_i = X^{-1}(C_i)$. Since $\bigcup_i C_i \in \mathcal{C}$ and
$$\bigcup_i A_i = \bigcup_i X^{-1}(C_i) = X^{-1}(\bigcup_i C_i),$$
$\bigcup_i A_i \in \sigma(X)$. □

Note that if $(\Omega, \mathcal{A})$ and $(T, \mathcal{C})$ are measurable spaces, then $X : \Omega \to T$ is measurable if and only if $\sigma(X) \subseteq \mathcal{A}$.

Let us assume that Ω is countable and $\mathcal{A} = 2^\Omega$. Let $X : \Omega \to T$. We may assume that T is also countable. Let $\mathcal{C} = 2^T$. X is measurable. We have for $j \in T$

$$p_j^X = P(X = j) = P^X(\{j\}) = P(\{\omega \mid X(\omega) = j) = \sum_{\omega : X(\omega) = j} p_\omega. \tag{2.6}$$

Definition 2.9 *Let X be real-valued on a countable space Ω. The expectation of X is defined as*

$$E(X) = \sum_{\omega \in \Omega} X(\omega) p_\omega, \tag{2.7}$$

if the sum makes sense: 1. If $\sum_{\omega\in\Omega} |X(\omega)|p_\omega$ is finite, then X "has an expectation". We call X integrable.
2. If $\sum_{\omega\in\Omega, X(\omega)<0} |X(\omega)|p_\omega < \infty$ and $\sum_{\omega\in\Omega, X(\omega)>0} X(\omega)p_\omega = \infty$, then X has expectation ∞.
3. If $\sum_{\omega\in\Omega, X(\omega)>0} X(\omega)p_\omega < \infty$ and $\sum_{\omega\in\Omega, X(\omega)<0} |X(\omega)|p_\omega = \infty$, then X has expectation $-\infty$.

Example 2.10 *(Binomial distribution.) Let $\Omega = \{0,1\}^n$, $x(\omega) = \omega_1 + \cdots + \omega_n$ for $\omega = (\omega_1, \ldots, \omega_n) \in \{0,1\}^n$. Furthermore, if $p_\omega = p^x(1-p)^{n-x}$ and $X : \Omega \to \{0, 1, \ldots, n\}$ with $X(\omega) = x$, then P^X is the binomial distribution.*

To compute the expectation of X, note that $X = X_1 + \cdots + X_n$ with $X_i(\omega) = \omega_i$. Thus

$$\mathbb{E}(X_i) = 1 \times P(\omega_i = 1) + 0 \times P(\omega_i = 0) = 1 \times p + 0 \times (1-p) = p.$$

Therefore,

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np.$$

Another possibility is a direct computation:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=0}^{n} i\, p_i \\
&= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{j=0}^{n-1} \frac{n!}{j!(n-1-j)!} p^{j+1} (1-p)^{n-1-j} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\
&= np.
\end{aligned}$$

Example 2.11 *(Poisson distribution.) Let X have a Poisson distribution with parameter $\lambda > 0$. Then*

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} i\, p_i$$

$$
\begin{aligned}
&= \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} \\
&= \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} e^{-\lambda} \\
&= \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} e^{-\lambda} \\
&= \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} \\
&= \lambda.
\end{aligned}
$$

The following example should discuss the role of the σ-algebra:

Example 2.12 *Assume we have a market consisting of an asset S and a bank account B. The prices today are 100 (the asset) and 1 (the bank account). Prices next week are random. Let $\theta > 1$, for instance $\theta = 1.1$. We assume that the future price S of the asset is either $100\,\theta^{-1}, 100$ or $100\,\theta$ and that B, the future price of the bank account, is either 1 or θ.*

Let $\Omega = \{100\,\theta^{-1}, 100, 100\,\theta\} \times \{1, \theta\}$ and $\mathcal{A} = 2^{\Omega}$. Elements $\omega \in \Omega$ are pairs of possible prices of the asset and possible values of the bank account. We will observe both S and B and thus for all subsets of Ω probability statements are feasible.

Now assume there is an investor who will observe not both S and B, but only the discounted price of the asset $\bar{S} = S/B$. $\bar{S}$ is a mapping $\bar{S} : \Omega \to T = \{100\,\theta^{-2}, 100\,\theta^{-1}, 100, 100\,\theta\}$. Let $\mathcal{C} = 2^T$. Given S and B we can compute $\bar{S}$, it is measurable w.r.t. $(\Omega, \mathcal{A})$ and $(T, \mathcal{C})$. But given $\bar{S}$, we are not always able to compute S and B. We are not able to distinguish between the pairs $(100\,\theta^{-1}, 1)$ and $(100, \theta)$ and between the pairs $(100, 1)$ and $(100\,\theta, \theta)$. Therefore, $\sigma(\bar{S})$ is generated by the partition

$$\{(100\,\theta^{-1}, 1), (100, \theta)\}, \{(100, 1), (100\,\theta, \theta)\}, \{(100\,\theta^{-1}, \theta)\}, \{(100\,\theta, 1)\}.$$

Note that $\mathcal{C}$ is a σ-algebra on T, whereas $\sigma(\bar{S})$ is a σ-algebra on Ω.

2.3 Exercises

Exercise 2.1 *Show that (2.2), (2.3) and (2.4) define probability distributions on $\{0, 1, \ldots, n\}$, $\mathbb{N}_0$ and $\mathbb{N}$ resp.*

Exercise 2.2 *Show that the Poisson distribution is the limit of the binomial distribution for $p = \lambda/n$ and $n \to \infty$.*

Exercise 2.3 *An exam is passed successfully with probability p, ($0 < p < 1$). In case of failure, the exam can be retaken. Let X be the number of times the exam is taken until the first success. Show that X is geometrically distributed.*

Exercise 2.4 *Compute the variance of the binomial and the Poisson distribution and the expectation of the geometric distribution. Recall that the variance σ^2 is the expectation of $(X - \mathbb{E}(X))^2$. It can be computed as $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.*

Exercise 2.5 *Let X have a probability distribution on $\mathbb{N}_0$. Prove that*

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} P(X > i).$$

Exercise 2.6 *Let X and Y be independent geometric random variables with parameter θ. Compute $P(X = Y)$ and $P(X > Y)$.*

Exercise 2.7 *Let X and Y be independent random variables, X Poisson and Y geometrically distributed. Compute $P(X = Y)$.*

Chapter 3

Probabilities on $\mathbb{R}$

3.1 Distribution Functions

In this chapter we have $\Omega = \mathbb{R}$. $\mathbb{R}$ is endowed with the Borel σ-algebra $\mathcal{B}$. Recall that $\mathcal{B}$ is generated by the open sets (it also generated by the open intervals, by the closed intervals, by half-open intervals). P is a probability measure on $(\mathbb{R}, \mathcal{B})$.

Definition 3.1 *The distribution function of P is defined (for $x \in \mathbb{R}$) by*

$$F(x) = P((-\infty, x]). \tag{3.1}$$

The distribution function (c.d.f.) F is a mapping $F : \mathbb{R} \to [0, 1]$.

Theorem 3.2 *1. The distribution function F characterizes the probability measure P: Two probability measures with the same c.d.f. are equal.*
2. A function $F : \mathbb{R} \to \mathbb{R}$ is the distribution function of a probability measure on $(\mathbb{R}, \mathcal{B})$ if and only if

(a) F is non-decreasing, $F(x) \leq F(y)$ for $x \leq y$,

(b) F is right-continuous ($F(x) = \lim_{y \downarrow x} F(y)$),

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

Proof. We prove only the "easy half" of part 2:

If $x \leq y$, then, since $(-\infty, x] \subseteq (-\infty, y]$, we have

$$F(x) = P((-\infty, x]) \leq P((-\infty, y]) = F(y).$$

Let $y_n > x$, $\lim_{n \to \infty} y_n = x$ and $y_{n+1} \leq y_n$. Then

$$(-\infty, y_n] \downarrow (-\infty, x]$$

and therefore,

$$F(y_n) = P((-\infty, y_n]) \downarrow P((-\infty, x]) = F(x).$$

Let $y_n \uparrow \infty$. Since $(-\infty, y_n] \uparrow (-\infty, \infty) = \mathbb{R}$, $F(y_n) \to 1 = P(\mathbb{R})$ follows. Similarly, let $y_n \downarrow -\infty$. Then $(-\infty, y_n] \downarrow \emptyset$ and $F(y_n) \to 0 = P(\emptyset)$. □

Note that a c.d.f. is not necessarily continuous!

Theorem 3.3 *Let F be the c.d.f. of the probability measure P on $(\mathbb{R}, \mathcal{B})$. Let $F(x-)$ denote the left limit of F at x.*

1. *$P((x, y]) = F(y) - F(x)$ for $x < y$.*
2. *$P([x, y]) = F(y) - F(x-)$ for $x \leq y$.*
3. *$P([x, y)) = F(y-) - F(x-)$ for $x < y$.*
4. *$P((x, y)) = F(y-) - F(x)$ for $x < y$.*
5. *$P(\{x\}) = F(x) - F(x-)$.*

Proof. Exercise.

A non-decreasing, right continuous function F with corresponding limits for $x \to \pm\infty$ defines a probability measure on $(\mathbb{R}, \mathcal{B})$. A useful method for constructing such functions is the idea of a density. Let f be a nonnegative integrable function such that

$$\int_{-\infty}^{\infty} f(y)dy = 1. \tag{3.2}$$

Let

$$F(x) = \int_{-\infty}^{x} f(y)dy. \tag{3.3}$$

F is the c.d.f. of a probability measure P. Here one has to be careful: Up to and including this chapter, the integral in (3.2) is the so-called Riemann integral. It is defined in a first step for continuous functions on bounded and closed intervals and then extended to piecewise continuous functions on $\mathbb{R}$. We assume that f is "nice enough" s.t. (3.2) makes sense. The next chapter is devoted to a new concept of integral, the Lebesgue integral, which allows a mathematically rigorous treatment.

3.2 Distributions

Example 3.4 *(Uniform distribution $\mathbb{U}(a, b)$). Let $a < b$. The density of the uniform distribution on the interval (a, b) is*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b), \\ 0 & \text{if } x \notin (a, b). \end{cases} \tag{3.4}$$

Example 3.5 *(Gamma distribution $\Gamma(\alpha,\beta)$). Let $\alpha,\beta>0$. The density f of the gamma distribution is $f(x)=0$ for $x\le 0$ and*

$$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \qquad \text{if } x>0. \tag{3.5}$$

$\Gamma(.)$ is the gamma function, defined by

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx.$$

It is defined for all $\alpha>0$ and satisfies $\Gamma(1)=1$, $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$. Thus $\Gamma(n+1)=n!$ for $n\in\mathbb{N}_0$.

α is called the shape parameter and β the scale parameter. If $\alpha=1$, the gamma distribution $\Gamma(1,\beta)$ is called the exponential distribution. It has the density

$$f(x)=\begin{cases}\beta\, e^{-\beta x} & \text{if } x>0,\\ 0 & \text{if } x\le 0.\end{cases} \tag{3.6}$$

Example 3.6 *(Normal distribution $N(\mu,\sigma^2)$). Let $\mu\in\mathbb{R}$ and $\sigma^2>0$. The normal distribution (with expectation μ and variance σ^2) has the density*

$$\phi_{\mu,\sigma^2}(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.7}$$

A special case is the standard normal distribution. Here $\mu=0$ and $\sigma^2=1$. The c.d.f. is denoted by Φ_{μ,σ^2} and Φ in the special case of the $N(0,1)$ distribution.

Example 3.7 *(Pareto distribution). Let $\alpha>0$. The Pareto distribution has the density*

$$f(x)=\begin{cases}\frac{\alpha}{x^{\alpha+1}} & \text{if } x>1,\\ 0 & \text{if } x\le 1.\end{cases} \tag{3.8}$$

and the c.d.f.

$$F(x)=\begin{cases}1-\frac{1}{x^\alpha} & \text{if } x>1,\\ 0 & \text{if } x\le 1.\end{cases} \tag{3.9}$$

Example 3.8 *(Cauchy distribution). The density of the Cauchy distribution is*

$$f(x)=\frac{1}{\pi}\frac{1}{1+x^2}. \tag{3.10}$$

Example 3.9 *Let $F(x)=I_{[a,\infty)}(x)$, i.e.*

$$F(x)=\begin{cases}1 & \text{if } x\ge a,\\ 0 & \text{if } x<a.\end{cases} \tag{3.11}$$

F has the properties (Theorem 3.2) to guarantee that it is the c.d.f. of a probability distribution. Since F is constant except in $x=a$ and has a jump of size 1 in $x=a$, we conclude that $P(\{a\})=1$ and $P(B)=0$ for all B with $a\notin B$. The distribution is called the Dirac distribution in a (sometimes also called the one-point distribution in a).

Example 3.10 *Let*

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1/4 & \text{if } 0 \leq x < 1, \\ x/2 & \text{if } 1 \leq x < 2, \\ 1 & \text{if } x \geq 2. \end{cases} \tag{3.12}$$

Again, F is the c.d.f. of a probability distribution on $\mathbb{R}$.

1. *F is constant on $[2, \infty)$, on $(0, 1)$ and on $(-\infty, 0)$, therefore $P((-\infty, 0)) = P((0, 1)) = P((2, \infty)) = 0$ and therefore $P(\{0\} \cup [1, 2]) = 1$.*

2. *F has jumps both of size $1/4$ at $x = 0$ and at $x = 1$, therefore $P(\{0\}) = P(\{1\}) = 1/4$.*

3. *F is linear on $(1, 2)$, where it has a derivative $F'(x) = 1/2$.*

P has a discrete component: The points 0 and 1 have a strictly positive probability. On $(1, 2)$, P is uniform. Therefore, if a random variable X has this c.d.f., then X is uniform on $(1, 2)$ (with probability $1/2$), it is 0 with probability $1/4$ and it is 1, again with probability $1/4$.

3.3 Exercises

Exercise 3.1 *Prove Theorem (3.3).*

Exercise 3.2 *Prove that the densities of the uniform, the exponential and the Pareto distribution are in fact densities, i.e. that they integrate to 1.*

Exercise 3.3 *Let*

$$F(x) = \begin{cases} 0 & \text{if } x < -1, \\ (x+1)^2/2 & \text{if } -1 \leq x < 0, \\ \sqrt{x+2}/2 & \text{if } 0 \leq x < 2, \\ 1 & \text{if } x \geq 2. \end{cases} \tag{3.13}$$

1. *Show that F is a c.d.f. on $\mathbb{R}$.*

2. *Identify all points a with $P(\{a\}) > 0$.*

3. *Find the smallest interval $[a, b]$ with $P([a, b]) = 1$.*

4. *Compute $P([-1, 0])$ and $P([0, 2])$.*

Exercise 3.4 *Compute the distribution function of the uniform distribution.*

Exercise 3.5 *Compute the distribution function of the exponential distribution.*

Exercise 3.6 *Let F be the distribution function of a probability measure P. Show that if F is continuous, $P(\{x\}) = 0$ for all one-point sets $\{x\}$ and $P(C) = 0$ for all countable sets C. Explain why F is continuous if P has a density. It is sufficient to assume that the density is bounded.*

Exercise 3.7 *Let $0 < p < 1$ and a distribution function F be given by*

$$F(x) = \sum_{k=1}^{\infty} (1-p)p^{k-1} I_{[1/k^2,\infty)}(x).$$

Find the probability of $A = [0, 1/5]$.

Exercise 3.8 *Let the random variable X have distribution function*

$$F(x) = \frac{1}{4}1_{[0,\infty)} + \frac{1}{2}1_{[1,\infty)} + \frac{1}{4}1_{[2,\infty)}.$$

Compute the expectation of X.

Chapter 4

Random Variables and Integration with respect to a Probability Measure

4.1 Random Variables

Let $X : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$. Recall that this is short for X being a function $X : \Omega \to F$ and measurable, i.e. $X^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{F}$. If a probability measure P is specified on Ω, we call a measurable X random variable.

Theorem 4.1 *Let $\mathcal{C} \subseteq \mathcal{F}$ such that $\mathcal{F} = \sigma(\mathcal{C})$, i.e. $\mathcal{F}$ is generated by $\mathcal{C}$. A function $X : \Omega \to F$ is measurable w.r.t. $\mathcal{A}$ and $\mathcal{F}$, if and only if $X^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$.*

Proof. Let $\mathcal{F}_0 = \{C \in \mathcal{F} \mid X^{-1}(C) \in \mathcal{A}\}$. Clearly, $\mathcal{C} \subseteq \mathcal{F}_0 \subseteq \mathcal{F}$. We show that $\mathcal{F}_0$ is a σ-algebra. Then, since $\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{C}$, $\mathcal{F} = \mathcal{F}_0$ follows. Furthermore, for all $C \in \mathcal{F}$ $(=\mathcal{F}_0)$, $X^{-1}(C) \in \mathcal{A}$ and X is therefore measurable w.r.t. $\mathcal{A}$ and $\mathcal{F}$.

To show that $\mathcal{F}_0$ is a σ-algebra, we have to check that the defining properties of a σ-algebra hold.

1. $X^{-1}(\emptyset) = \emptyset$ and $X^{-1}(F) = \Omega$.

   X^{-1} commutes with intersections, unions and taking complements, for instance $X^{-1}(C^c) = X^{-1}(C)^c$, $X^{-1}(\bigcup C_i) = \bigcup X^{-1}(C_i)$ and $X^{-1}(\bigcap C_i) = \bigcap X^{-1}(C_i)$. Therefore

2. If $C \in \mathcal{F}_0$, then, since $\mathcal{A}$ is closed w.r.t. complements and $X^{-1}(C^c) = X^{-1}(C)^c$, $C^c \in \mathcal{F}_0$.

3. Let $A_i \in \mathcal{F}_0$. Again, since $\mathcal{A}$ is closed w.r.t. countable unions and $X^{-1}(\bigcup_{i=1}^{\infty} C_i) = \bigcup_{i=1}^{\infty} X^{-1}(C_i)$, it follows that $\bigcup_{i=1}^{\infty} C_i \in \mathcal{F}_0$. □

The significance of the Theorem 4.1 is that in order to show that a mapping $X : \Omega \to F$ is measurable, one does not have to check that all preimages $X^{-1}(C)$ of $C \in \mathcal{F}$ are in $\mathcal{A}$. It is

sufficient to show this for a typically small subset of $\mathcal{F}$. The Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ is generated by intervals $(-\infty, a]$. Therefore, a real-valued X is measurable, if for all $a \in \mathbb{R}$, $\{\omega \mid X(\omega) \leq a\}$ is in $\mathcal{A}$. $\mathcal{B}$ is also generated by open intervals $(-\infty, a)$. Therefore, X is measurable, if all $a \in \mathbb{R}$, $\{\omega \mid X(\omega) < a\} \in \mathcal{A}$.

The following theorem summarizes properties of measurable functions. Recall that

1. The indicator function I_A (for $A \subseteq \Omega$) is defined by

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

2. For a sequence (X_n) of mappings $X_n : \Omega \to \mathbb{R}$,

$$\begin{aligned} \limsup_{n\to\infty} X_n(\omega) &= \lim_{n\to\infty} \sup_{m\geq n} X_m(\omega) = \inf_{n\to\infty} \sup_{m\geq n} X_m(\omega), \\ \liminf_{n\to\infty} X_n(\omega) &= \lim_{n\to\infty} \inf_{m\geq n} X_m(\omega) = \sup_{n\to\infty} \inf_{m\geq n} X_m(\omega). \end{aligned}$$

Theorem 4.2 *Let $(\Omega, \mathcal{A})$ and $(F, \mathcal{F})$ be measurable spaces and $X, X_n : \Omega \to F$.*

1. *Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. X is measurable if and only if for all $a \in \mathbb{R}$, $\{X \leq a\} \in \mathcal{A}$ or if for all $a \in \mathbb{R}$, $\{X < a\} \in \mathcal{A}$.*

2. *Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. If X_n is measurable for all n, then $\inf X_n$, $\sup X_n$, $\liminf X_n$ and $\limsup X_n$ are measurable.*

3. *Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. If X_n is measurable for all n and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$, then X is measurable.*

4. *Let $X : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ and $Y : (F, \mathcal{F}) \to (G, \mathcal{G})$, then $Y \circ X : (\Omega, \mathcal{A}) \to (G, \mathcal{G})$.*

5. *If $\mathcal{A}$ is the Borel σ-algebra on Ω and $\mathcal{F}$ is the Borel σ-algebra on F, then every continuous function $X : \Omega \to F$ is measurable.*

6. *Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. Let $X_1, \ldots, X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ and $f : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}, \mathcal{B})$. Then $f(X_1, \ldots, X_n) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$.*

7. *Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ and X, Y be real-valued and measurable. Then, $X + Y$, XY, X/Y (if $Y \neq 0$), $X \wedge Y$, $X \vee Y$ are measurable.*

8. *An indicator function I_A is measurable if and only if $A \in \mathcal{A}$.*

Proof. Statement *1.* is a special case of Theorem 4.1, since $\mathcal{B}$ is generated by both the closed and the open intervals.

2. Let (X_n) be a sequence of measurable functions $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$. Note that $\inf X_n < a$ if and only if there exists an n s.t. $X_n < a$. Therefore

$$\{\inf X_n < a\} = \bigcup_{n=1}^{\infty} \{X_n < a\} \in \mathcal{A}.$$

Similarly, $\sup X_n \leq a$ if and only if for all n $X_n \leq a$. Hence

$$\{\sup X_n \leq a\} = \bigcap_{n=1}^{\infty} \{X_n \leq a\} \in \mathcal{A}.$$

$$\begin{aligned} \liminf_{n\to\infty} X_n &= \sup_n \inf_{k\geq n} X_k, \\ \limsup_{n\to\infty} X_n &= \inf_n \sup_{k\geq n} X_k, \end{aligned}$$

implies that both $\liminf_{n\to\infty} X_n$ and $\limsup_{n\to\infty} X_n$ are measurable.

3. If $X_n \to X$, then $X = \lim_{n\to\infty} X_n = \limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n$ is measurable.

4. Let $B \in \mathcal{G}$. We have $(Y \circ X)^{-1}(B) = X^{-1}(Y^{-1}(B))$. $Y^{-1}(B) \in \mathcal{F}$ since Y is measurable. Then $(X^{-1}(Y^{-1}(B)) \in \mathcal{A}$, since X is measurable.

5. The Borel σ-algebras are generated by the open sets. If X is continuous and $O \subseteq F$ is open, then $X^{-1}(O) \subseteq \Omega$ is open. Again apply Theorem 4.1.

6. This is a special case of 4. with $X = (X_1, \ldots, X_n)$ and $Y = f$. Note that

$$X^{-1}([a_1, b_1] \times \cdots \times [a_n, b_n]) = X_1^{-1}([a_1, b_1]) \cap \cdots \cap X_n^{-1}([a_n, b_n]).$$

7. The functions $f(x, y) = x + y$, $f(x, y) = xy$, $f(x, y) = x/y$, $f(x, y) = x \wedge y$, $f(x, y) = x \vee y$ are continuous on their domains.

8. Note that for $B \in \mathcal{B}$,

$$(I_A)^{-1}(B) = \begin{cases} \Omega & \text{if } 1 \in B \text{ and } 0 \in B, \\ A & \text{if } 1 \in B \text{ and } 0 \notin B, \\ A^c & \text{if } 1 \notin B \text{ and } 0 \in B, \\ \emptyset & \text{if } 1 \notin B \text{ and } 0 \notin B. \end{cases}$$

□

4.2 Expectation

Let $(\Omega, \mathcal{A}, P)$ be a probability space. In this section X, X_n, Y are always measurable real-valued random variables. We want to define the expectation (or the integral) of X. The integral that we define in this section is called Lebesgue-integral. We will see that it is a concept that can be applied to measurable functions, not only to continuous functions as is the case with the Riemann-integral.

Definition 4.3 *A r.v. X is called simple, if it can be written as*

$$X = \sum_{i=1}^{m} a_i I_{A_i} \tag{4.1}$$

with $a_i \in \mathbb{R}$ and $A_i \in \mathcal{A}$.

A r.v. is simple if it is measurable and has only a finite number of values. Its representation (4.1) is not unique, it is unique only if all the a_i's are different. Every real-valued r.v. X can be approximated by simple functions, i.e. there is a sequence of simple functions converging to X.

Proposition 4.4 *Let X be a nonnegative real-valued r.v. There exists a sequence (Y_n) of nonnegative simple functions s.t. $Y_n \uparrow X$.*

Proof. Define

$$Y_n(\omega) = \begin{cases} k2^{-n} & \text{if } k2^{-n} \le X(\omega) < (k+1)\,2^{-n}, k \le n\,2^n - 1, \\ n & \text{if } X(\omega) \ge n. \end{cases}$$

To approximate a not necessarily nonnegative r.v. X by simple functions, write X as $X = X^+ - X^-$, with $X^+ = X \vee 0$ the positive part and $X^- = -(X \wedge 0)$ the negative part of X and approximate X^+ and X^-.

The expectation $\mathbb{E}(.)$ maps random variables to $\mathbb{R}$, is linear $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$, for reals α, β, and $\mathbb{E}(I_A) = P(A)$. Therefore we define

Definition 4.5 *Let X be a real-valued r.v. on $(\Omega, \mathcal{A}, P)$.*

1. *If $X = \sum_{i=1}^m a_i I_{A_i}$ simple,*
$$\mathbb{E}(X) = \sum_{i=1}^{m} a_i P(A_i). \tag{4.2}$$
2. *If $X \ge 0$,*
$$\mathbb{E}(X) = \sup_{Y\, simple, 0 \le Y \le X} \mathbb{E}(Y). \tag{4.3}$$
3. *If $X = X^+ - X^-$, $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$,*
$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-). \tag{4.4}$$

The expectation $\mathbb{E}(X)$ can be written as

$$\int X dP$$

or as

$$\int X(\omega) dP(\omega), \qquad \int X(\omega) P(d\omega),$$

if appropriate. Note that, since $|X| = X^+ + X^-$, both $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ are finite iff $\mathbb{E}(|X|)$ is finite. It is convenient to stick to the following:

X is called integrable (has an expectation) if $\mathbb{E}(|X|)$ is finite. X admits an expectation if it is either integrable or if $\mathbb{E}(X^+) = \infty$ and $\mathbb{E}(X^-) < \infty$ or $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) = \infty$. The set of all integrable real-valued random variable is denoted by $\mathcal{L}^1$. Note that all almost surely (a.s.) bounded random variables are integrable. A random variable X is a.s. bounded, if there exist a $c \geq 0$ s.t. $P(|X| > c) = 0$.

Theorem 4.6 *1. Let $X \in \mathcal{L}^1$ and $X \geq 0$. Let (Y_n) be a sequence of nonnegative simple functions s.t. $Y_n \uparrow X$. Then $\mathbb{E}(Y_n) \to \mathbb{E}(X)$.*

2. $\mathcal{L}^1$ is a linear space (vector space), i.e. if $X, Y \in \mathcal{L}^1$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha X + \beta Y \in \mathcal{L}^1$.

3. $\mathbb{E}$ is positive, i.e. if $X \in \mathcal{L}^1$ and $X \geq 0$, then $\mathbb{E}(X) \geq 0$. If $0 \leq X \leq Y$ and $Y \in \mathcal{L}^1$, then $X \in \mathcal{L}^1$. Furthermore, $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

4. If $X = Y$ a.s. ($\{\omega \mid X(\omega) \neq Y(\omega)\}$ has probability 0), then $\mathbb{E}(X) = \mathbb{E}(Y)$.

Proof. We only sketch the proof that $\mathcal{L}^1$ is a vector space. Let $X, Y \in \mathcal{L}^1$ and $\alpha, \beta \in \mathbb{R}$. We have

$$|\alpha X + \beta Y| \leq |\alpha|\,|X| + |\beta|\,|Y|.$$

We have to show that $|\alpha|\,|X| + |\beta|\,|Y|$ is integrable. Let (U_n) and (V_n) be sequences of nonnegative simple functions s.t. $U_n \uparrow |X|$ and $V_n \uparrow |Y|$. Define $Z_n = |\alpha|\,U_n + |\beta|\,V_n$. Each Z_n is simple and nonnegative (both U_n and V_n have only a finite number of values!). $Z_n \uparrow |\alpha|\,|X| + |\beta|\,|Y|$ and

$$\mathbb{E}(Z_n) = |\alpha|\ \mathbb{E}(U_n) + |\beta|\,\mathbb{E}(V_n) \leq |\alpha|\ \mathbb{E}(|X|) + |\beta|\,\mathbb{E}(|Y|) < \infty. \qquad \square$$

4.3 Properties

First, we discuss the continuity of taking expectation, i.e. can the operations of taking limits and expectations be interchanged? Is it true that if (X_n) is a sequence of integrable functions converging a.s. to X, that $\mathbb{E}(X_n) \to \mathbb{E}(X)$? Since the answer is no, see the example below, we have to state additional properties for the sequence (X_n) that imply $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Example 4.7 *Let P be the uniform distribution on $\Omega = (0, 1)$. Let $X_n = a_n I_{(0,1/n)}$. We have $X_n = a_n$ with probability $1/n$ and $X_n = 0$ with probability $1 - 1/n$. The expectation is $\mathbb{E}(X_n) = a_n/n$.*

For $n \to \infty$, $X_n \to 0$, because for all $\omega \in (0, 1)$, $X_n(\omega) = 0$ if $1/n \leq \omega$. By choosing appropriate sequences (a_n), we can have all possible nonnegative limits for $(\mathbb{E}(X_n))$. For instance, if $a_n = n$, then $\mathbb{E}(X_n) = 1 \to 1 \neq 0 = \mathbb{E}(0)$. For $a_n = n^2$, $\mathbb{E}(X_n) \to \infty$.

Theorem 4.8 *1. (Monotone convergence theorem). Let $X_n \geq 0$ and $X_n \uparrow X$. Then $\mathbb{E}(X_n) \to \mathbb{E}(X)$ (even if $\mathbb{E}(X) = \infty$).*

2. (Fatou's lemma). Let $Y \in \mathcal{L}^1$ and let $X_n \geq Y$ a.s. for all n. Then

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n). \tag{4.5}$$

In particular, if all X_n's are nonnegative, (4.5) holds.

3. (Lebesgue's dominated convergence theorem). Let $X_n \to X$ a.s. and $|X_n| \leq Y \in \mathcal{L}^1$ for all n. Then $X_n, X \in \mathcal{L}^1$ and $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Theorem 4.9 *Let (X_n) be a sequence of real-valued random variables.*

1. If the X_n's are all nonnegative, then

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n). \tag{4.6}$$

2. If $\sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty$, then $\sum_{n=1}^{\infty} X_n$ converges a.s. and (4.6) holds.

Proof. 1. $\sum_{n=1}^{\infty} X_n$ is the limit of the partial sums $T_m = \sum_{n=1}^{m} X_n$. If all X_n are nonnegative, (T_m) is increasing and we may apply the monotone convergence theorem.

2. Define $S_m = \sum_{n=1}^{m} |X_n|$. Again, (S_m) is increasing and has a limit S in $[0, \infty]$. Again, the monotone convergence theorem implies that

$$\mathbb{E}(S) = \sum_{n=1}^{\infty} \mathbb{E}(|X_n|),$$

and is therefore finite. Therefore, S is also finite a.s. and it is integrable. Since

$$|\sum_{n=1}^{m} X_n| \leq S$$

for all m, the limit $\sum_{n=1}^{\infty} X_n$ exists, is dominated by S and we may apply Lebesgue's dominated convergence theorem to derive (4.6). □

A real-valued random variable X is called square-integrable, if $\mathbb{E}(X^2) < \infty$. The set of square-integrable random variables is denoted by $\mathcal{L}^2$.

Theorem 4.10 *1. $\mathcal{L}^2$ is a linear space, i.e. if $X, Y \in \mathcal{L}^2$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha X + \beta Y \in \mathcal{L}^2$.*

2. (Cauchy-Schwarz inequality). If $X, Y \in \mathcal{L}^2$, then $XY \in \mathcal{L}^1$ and

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \tag{4.7}$$

3. If $X \in \mathcal{L}^2$, then $X \in \mathcal{L}^1$ and $\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)$.

Proof. 1. Let $X, Y \in \mathcal{L}^2$ and $\alpha, \beta \in \mathbb{R}$. Note that for all $u, v \in \mathbb{R}$, $(u+v)^2 \leq 2u^2 + 2v^2$. Therefore

$$(\alpha X + \beta Y)^2 \leq 2\alpha^2 X^2 + 2\beta^2 Y^2$$

and

$$\mathbb{E}\left((\alpha X + \beta Y)^2\right) \leq \mathbb{E}\left(2\alpha^2 X^2 + 2\beta^2 Y^2\right) = 2\alpha^2 \mathbb{E}(X^2) + 2\beta^2 \mathbb{E}(Y^2) < \infty.$$

Therefore, $\alpha X + \beta Y \in \mathcal{L}^2$.

2. Let X and Y be square-integrable. Since

$$|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2,$$

we have

$$\mathbb{E}(|XY|) \leq \frac{1}{2}\mathbb{E}(X^2) + \frac{1}{2}\mathbb{E}(Y^2) < \infty.$$

To prove the Cauchy-Schwarz inequality: If $\mathbb{E}(Y^2) = 0$, then $Y = 0$ holds with probability 1 and therefore $XY = 0$, again with probability 1 and thus $\mathbb{E}(XY) = 0$.

Now assume that $\mathbb{E}(Y^2) > 0$. Let $t = \mathbb{E}(XY)/\mathbb{E}(Y^2)$. Then

$$0 \leq \mathbb{E}\left((X - tY)^2\right) = \mathbb{E}(X^2) - 2t\mathbb{E}(XY) + t^2\mathbb{E}(Y^2) = \mathbb{E}(X^2) - \mathbb{E}(XY)^2/\mathbb{E}(Y^2).$$

Therefore,

$$\mathbb{E}(XY)^2/\mathbb{E}(Y^2) \leq \mathbb{E}(X^2)$$

and (4.7) follow.

3. Choose $Y = 1$ and apply the Cauchy-Schwarz inequality. □

The variance of $X \in \mathcal{L}^2$ is defined as

$$\mathbb{V}(X) = \sigma_X^2 = \mathbb{E}((X - \mathbb{E}(X))^2). \tag{4.8}$$

Theorem 4.11 *Let X be a real-valued random variable.*

1. Let $h : \mathbb{R} \to \mathbb{R}$ be measurable and nonnegative. Then for all $a > 0$,

$$P(h(X) \geq a) \leq \frac{\mathbb{E}(h(X))}{a}. \tag{4.9}$$

2. (Markov's inequality). For $a > 0$ and $X \in \mathcal{L}^1$,

$$P(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}. \tag{4.10}$$

3. (Chebyshev's inequality). For $a > 0$ and $X \in \mathcal{L}^2$,

$$P(|X| \geq a) \leq \frac{\mathbb{E}(X^2)}{a^2} \tag{4.11}$$

and

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{\sigma_X^2}{a^2}. \tag{4.12}$$

Proof. Let $A = \{\omega \mid h(X(\omega)) \geq a\}$. Then

$$\mathbb{E}(h(X)) \geq \mathbb{E}(h(X)I_A) \geq \mathbb{E}(aI_A) = aP(A).$$

Taking $h(x) = |x|$ gives Markov's inequality. To prove Chebyshev's inequalities, note that, for instance $|X| \geq a$ is equivalent to $X^2 \geq a^2$. Now apply Markov's inequality to X^2 and a^2. Similarly, $|X - \mathbb{E}(X)| \geq a$ if and only if $(X - \mathbb{E}(X))^2 \geq a^2$. □

Let $X : (\Omega, \mathcal{A}, P) \to (F, \mathcal{F})$ be measurable. It defines a probability measure P^X on $(F, \mathcal{F})$ by $P^X(B) = P(X^{-1}(B))$. Now let $h : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ and $Y = h \circ X$, i.e. $Y(\omega) = h(X(\omega))$. Then $Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$. Y defines a probability measure P^Y on $(\mathbb{R}, \mathcal{B})$. For instance, let X be integrable with expectation $\mu = \mathbb{E}(X)$ and $h(x) = (x - \mu)^2$. Then $Y = (X - \mu)^2$, the expectation of Y is the variance of X. Which of the probability measures P, P^X or P^Y is to be used to compute this expectation? It does not matter!

Theorem 4.12 *(Expectation rule). Let X be a r.v. on $(\Omega, \mathcal{A}, P)$ with values in $(F, \mathcal{F})$ and $h : (F, \mathcal{F}) \to \mathbb{R}, \mathcal{B})$. We have $h(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ if and only if $h \in \mathcal{L}^1(F, \mathcal{F}, P^X)$. Furthermore,*

$$\mathbb{E}(h(X)) = \int h(X(\omega))dP(\omega) = \int h(x)dP^X(x).$$

4.4 Lebesgue Measure and Densities

Let a probability space $(\Omega, \mathcal{A}, P)$ and a real-valued random variable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ be given. In Chapter 3 we have introduced a first version of the concept of a density of X. For f nonnegative with $\int_{-\infty}^{\infty} f(x)dx = 1$, we defined the c.d.f. F by $F(x) = \int_{-\infty}^{x} f(y)dy$. We stated that regularity conditions have to hold such that these integrals make sense. Here we are more precise. First, we introduce the Lebesgue measure.

Definition 4.13 *Let u_n denote the uniform distribution on $(n, n + 1]$. We define the Lebesgue measure $m : \mathcal{B} \to [0, \infty]$ by*

$$m(A) = \sum_{n \in \mathbb{Z}} u_n(A). \tag{4.13}$$

We remark that by this definition, m maps the Borel sets to $\overline{\mathbb{R}}_+ = [0, \infty]$. It is countably additive, i.e. if (A_n) is a sequence of pairwise disjoint Borel sets, then

$$m(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} m(A_n).$$

Such a countably additive mapping is called a measure. Probability measures P are finite measures with $P(\Omega) = 1$. Measures are not necessarily finite, for instance $m([0, \infty)) = \infty$.

Lebesgue measure is the only measure which gives $m((a, b]) = b - a$ for all intervals $(a, b]$ $(a < b)$. It can be viewed as an extension of the uniform distribution to $\mathbb{R}$. Another characterization is the following: Let μ be a measure such that for all intervals $(a, b]$ and all $t \in \mathbb{R}$, $\mu((a+t, b+t]) = \mu((a, b])$. Then there exists a constant c s.t. $\mu = cm$.

Let f be a measurable function $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$. The integral $\int f dm$ w.r.t. Lebesgue measure is defined exactly as in the case of a probability measure. In a first step, it is defined for simple functions $f = \sum_{i=1}^{n} a_i I_{A_i}$ as

$$\int f dm = \sum_{i=1}^{n} a_i m(A_i).$$

Then it is extended to nonnegative measurable functions f as

$$\int f dm = \sup_{0 \le g \le f, g \, simple} \int g dm.$$

Finally, if $\int f dm = \int f^+ dm < \infty$ and $\int f^- dm = \int f^+ dm < \infty$,

$$\int f dm = \int f^+ dm - \int f^- dm.$$

Note that bounded functions are not necessarily integrable w.r.t. the Lebesgue measure. For instance, the constant $f(x) = 1$ has $\int f dm = \infty$. We often write $\int f(x) dx$ for $\int f(x) dm(x)$.

Definition 4.14 *A density on $\mathbb{R}$ is a nonnegative measurable function $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ with $\int f dm = 1$. It defines a probability measure P on $(\mathbb{R}, \mathcal{B})$ by*

$$P(A) = \int I_A f dm, \qquad A \in \mathcal{B}. \tag{4.14}$$

If a probability measure is defined by a density it is called *absolutely continuous* with respect to Lebesgue measure.

Proposition 4.15 *If a probability measure on $(\mathbb{R}, \mathcal{B})$ has a density, i.e. P is defined by (4.14)), then its density is unique a.e. That is, if f and g are densities of P then*

$$m(\{x \mid f(x) \neq g(x)\}) = 0.$$

4.5 Exercises

Exercise 4.1 *Let $X : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ be measurable. Define $\sigma(X) = \{A \subseteq \Omega \mid$ there exist a $B \in \mathcal{F}$ s.t. $A = X^{-1}(B)\}$. Show that $\sigma(X)$ is a sub-σ-algebra of $\mathcal{A}$. It is the smallest σ-algebra on Ω which makes X measurable. It is called the σ-algebra generated by X.*

Exercise 4.2 *Let a probability space $(\Omega, \mathcal{A}, P)$ be given. $N \subseteq \Omega$ is called a null set, if there exists a $A \in \mathcal{A}$ s.t. $N \subseteq A$ and $P(A) = 0$. Null sets are not necessarily in $\mathcal{A}$, a fact, which can produce problems in proof. This problem can be resolved: Let*

$$\mathcal{A}' = \{A \cup N \mid A \in \mathcal{A}, N \text{ a null set}\}.$$

Prove that

1. *$\mathcal{A}'$ is again a σ-algebra, the smallest containing $\mathcal{A}$ and all null sets.*
2. *P extends uniquely to a probability measure on $\mathcal{A}'$ by defining $P(A \cup N) = P(A)$.*

Exercise 4.3 *Let X be integrable on $(\Omega, \mathcal{A}, P)$. Let (A_n) be a sequence of events s.t. $P(A_n) \to 0$. Prove that $\mathbb{E}(XI_{A_n}) \to 0$. Note that $P(A_n) \to 0$ does not imply $I_{A_n} \to 0$ a.s.*

Exercise 4.4 *Let X be a random variable on $(\Omega, \mathcal{A}, P)$ with $X \geq 0$ and $\mathbb{E}(X) = 1$. Define $Q(A) = \mathbb{E}(XI_A)$.*

1. *Prove that Q is a probability measure on $(\Omega, \mathcal{A})$.*
2. *Prove that $P(A) = 0$ implies $Q(A) = 0$.*
3. *Let $P(X > 0) = 1$. Let $Y = 1/X$. Show that $P(A) = \mathbb{E}^Q(YI_A)$ and that P and Q have the same null sets.*

Q is called absolutely continuous with respect to P if $P(A) = 0$ implies $Q(A) = 0$. If Q is absolutely continuous with respect to P and P is absolutely continuous with respect to Q then P and Q are called equivalent.

Exercise 4.5 *Compute the expectation and the variance of the gamma distribution.*

Exercise 4.6 *Compute the expectation and the variance of the normal distribution.*

Exercise 4.7 *Let X have a Cauchy distribution. The Cauchy distribution is defined by its density*

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \qquad x \in \mathbb{R}.$$

Prove that X is not integrable, i.e. $\mathbb{E}(|X|) = \infty$.

Exercise 4.8 *Let $X \sim N(0,1)$. Prove that $P(X \geq x) \leq \phi(x)/x$ for $x > 0$. Compare with the Chebyshev inequality.*

Exercise 4.9 *(Jensen's inequality). Let f be a measurable and convex function defined on $C \subseteq \mathbb{R}$. Let X be a random variable with values in C. Assume that both X and $f \circ X$ are integrable. Prove that $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$.*

Exercise 4.10 *Suppose $\mathbb{E}(X) = 3$ and $\mathbb{E}(|X-3|) = 1$. Give a reasonable upper bound of $P(X \leq -2 \text{ or } X \geq 8)$. Let furthermore the variance be known to be $\sigma^2 = 2$. Improve the estimate of $P(X \leq -2 \text{ or } X \geq 8)$.*

Chapter 5

Probability Distributions on $\mathbb{R}^n$

5.1 Independent Random Variables

Let $X : (E, \mathcal{E}, P) \to (G, \mathcal{G})$ and $Y : (F, \mathcal{F}, Q) \to (H, \mathcal{H})$ be two random variables defined on two different probability spaces. We want to construct a probability space on which both random variables are defined, i.e. on which (X, Y) is defined. Moreover, X and Y should be independent and we have to discuss how to compute expectations on the product space.

Recall that we have defined independence of events.

Definition 5.1 *Let a probability space $(\Omega, \mathcal{A}, P)$ be given. Sub σ-algebras $(\mathcal{A}_i)_{i\in I}$ are independent, if for all finite $J \subseteq I$ and all $A_i \in \mathcal{A}_i$,*

$$P(\cap_{i\in J} A_i) = \prod_{i\in J} P(A_i).$$

Random variables $X_i : (\Omega, \mathcal{A}, P) \to (E_i, \mathcal{E}_i), i \in I$ are independent, if the σ-algebras $X_i^{-1}(\mathcal{E}_i)$ are independent, i.e. if for all finite $J \subseteq I$ and $B_i \in \mathcal{E}_i$,

$$P(X_i \in B_i \text{ for all } i \in J) = \prod_{i\in J} P(X_i \in B_i).$$

Theorem 5.2 *Let X and Y be random variables defined on $(\Omega, \mathcal{A}, P)$ with values in $(E, \mathcal{E})$ and $(F, \mathcal{F})$ resp. X and Y are independent if and only if any one of the following conditions holds.*

1. *$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all $A \in \mathcal{E}, B \in \mathcal{F}$.*
2. *$f(X)$ and $g(Y)$ are independent for all measurable f and g.*
3. *$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$ for all bounded measurable or positive measurable functions f and g.*
4. *If $\mathcal{E}$ and $\mathcal{F}$ are the Borel σ-algebras on E and F, then $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$ for all bounded and continuous functions f and g.*

Remark and sketch of proof. $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ is the definition of the independence of X and Y. This equation, with probabilities written as expectations of indicator functions, reads as

$$\mathbb{E}(I_{A\times B}(X,Y)) = \mathbb{E}(I_A(X))\mathbb{E}(I_B(Y)).$$

by linearity, the equation holds for simple functions. Positive measurable functions are approximated by simple functions and finally integrable functions are written as the difference of its positive and its negative part.

Note that the theorem implies that, in case X and Y are independent and $f(X)$ and $g(Y)$ integrable, then $f(X)g(Y)$ is also integrable. Here we do not have to assume that $f(X)$ and $g(Y)$ are square-integrable. □

Let probability spaces $(E, \mathcal{E}, P)$ and $(F, \mathcal{F}, Q)$ be given. To construct a probability R on $E \times F$ with the property $R(A \times B) = P(A)Q(B)$, we first have to define a σ-algebra on $E \times F$.

Denote by $\mathcal{E} \times \mathcal{F}$ the set of rectangles on $E \times F$, i.e.

$$\mathcal{E} \times \mathcal{F} = \{A \times B \mid A \in \mathcal{E}, B \in \mathcal{F}\}.$$

Note that $\mathcal{E} \times \mathcal{F}$ is not a σ-algebra, for instance the complement of a rectangle $A \times B$ is typically not a rectangle, but a union of rectangles,

$$(A \times B)^c = A^c \times B \ \cup \ A \times B^c \ \cup \ A^c \times B^c.$$

Definition 5.3 *The product σ-algebra $\mathcal{E} \otimes \mathcal{F}$ is the σ-algebra on $E \times F$ generated by $\mathcal{E} \times \mathcal{F}$.*

Theorem 5.4 *(Tonelli-Fubini). Let $(E, \mathcal{E}, P)$ and $(F, \mathcal{F}, Q)$ be probability spaces.*
a) Define $R(A \times B) = P(A)Q(B)$ for $A \in \mathcal{E}$, $B \in \mathcal{F}$. R extends uniquely to a probability measure $P \otimes Q$ on $\mathcal{E} \times \mathcal{F}$ (it is called the product measure).
b) If f is measurable and integrable (or positive) w.r.t. $\mathcal{E} \otimes \mathcal{F}$, then the functions

$$\begin{aligned} x &\mapsto \int f(x,y)Q(dy), \\ y &\mapsto \int f(x,y)P(dx) \end{aligned}$$

are measurable and

$$\begin{aligned} \int f\, dP \otimes Q &= \int \left\{ \int f(x,y)Q(dy) \right\} P(dx) \\ &= \int \left\{ \int f(x,y)P(dx) \right\} Q(dy). \end{aligned}$$

Remark 5.5 *Let X and Y be random variables on $(\Omega, \mathcal{A}, P)$ with values in $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively. The pair (X, Y) is a random variable with values in $(E \times F, \mathcal{E} \otimes \mathcal{F})$. X and Y are independent if and only if*

$$P^{(X,Y)} = P^X \otimes P^Y.$$

5.2 Joint, Marginal and Conditional Distributions

Recall that the Borel σ-algebra $\mathcal{B}^n$ on $\mathbb{R}^n$ is generated by the open subsets of $\mathbb{R}^n$. It is also the n-fold product σ-algebra $\mathcal{B}^n = \mathcal{B} \otimes \cdots \otimes \mathcal{B}$ of the Borel σ-algebra on $\mathbb{R}$.

Definition 5.6 *Lebesgue measure m_n on $(\mathbb{R}^n, \mathcal{B}^n)$ is the n-fold product measure $m_n = m \otimes \cdots \otimes m$ of the Lebesgue measure m on $\mathbb{R}$. It is the unique measure on $(\mathbb{R}^n, \mathcal{B}^n)$ satisfying $m_n((a_1, b_1] \times \cdots \times (a_n, b_n]) = \prod_{i=1}^n (b_i - a_i)$.*

Definition 5.7 *A probability measure P on $(\mathbb{R}^n, \mathcal{B}^n)$ has a density f w.r.t. Lebesgue measure, if for $A \in \mathcal{B}^n$,*

$$P(A) = \int I_A \, f \, dm_n.$$

If P is defined by a density w.r.t. m_n, P is called absolutely continuous (w.r.t. m_n).

Theorem 5.8 *A measurable function f is the density of a probability measure P on $(\mathbb{R}^n, \mathcal{B}^n)$ if and only if $f \geq 0$ and $\int f dm_n = 1$. If a density of P exists, it is unique, i.e. if f and g are densities of P then*

$$m_n(\{x \mid f(x) \neq g(x)\}) = 0.$$

Theorem 5.9 *Let $X = (Y, Z) \in \mathbb{R}^2$ be absolutely continuous with density $f(y, z)$. Then*
1. Both Y and Z are absolutely continuous with densities f_Y and f_Z given by

$$\begin{aligned} f_Y(y) &= \int f(y, z) \, dz \\ f_Z(z) &= \int f(y, z) \, dy. \end{aligned}$$

2. Y and Z are independent if and only if $f(y, z) = f_Y(y) f_Z(z)$ a.e.
3. Define $f(z \mid y)$ by

$$f(z \mid y) = \begin{cases} \frac{f(y,z)}{f_Y(y)} & \text{if } f_Y(y) \neq 0, \\ 0 & \text{else.} \end{cases} \tag{5.1}$$

$f(. \mid y)$ is a density on $\mathbb{R}$, called the conditional density of Z given $Y = y$.

Proof. We prove that $f_Y(y) = \int f(y, z) dz$ is the density of Y, i.e. for all $A \in \mathcal{B}$,

$$P(Y \in A) = \int I_A(y) f_Y(y) \, dy.$$

We have

$$\begin{aligned} \int I_A(y) f_Y(y) \, dy &= \int I_A(y) \int f(y, z) \, dz \, dy \\ &= \int \int I_A(y) f(y, z) \, dy \, dz \\ &= \int \int I_{A \times \mathbb{R}}(y, z) f(y, z) \, dy \, dz \\ &= P(Y \in A, Z \in \mathbb{R}) = P(Y \in A). \end{aligned}$$

We have applied the Theorem of Tonelli-Fubini.

That $f(y,z) = f_Y(y)f_Z(z)$ implies the independence of Y and Z is left as an easy exercise. To show that the independence implies the product-form of $f(y,z)$ is harder and we do not prove it.

$f(y \mid z)$ is a density, since it is nonnegative and

$$\int f(y \mid z)\, dy = \int \frac{f(y,z)}{f_Z(z)}\, dy = \frac{1}{f_Z(z)} \int f(y,z)\, dy = \frac{f_Z(z)}{f_Z(z)} = 1. \qquad \square$$

Definition 5.10 *Let X and Y be square-integrable $\mathbb{R}$-valued random variables with expectations μ_X and μ_Y and variances σ_X^2 and σ_Y^2. The covariance of X and Y is*

$$\mathbb{C}ov(X,Y) = \mathbb{E}((X-\mu_X)(Y-\mu_Y)). \tag{5.2}$$

If both σ_X^2 and σ_Y^2 are strictly positive, we define the correlation coefficient of X and Y as

$$\rho(X,Y) = \frac{\mathbb{C}ov(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}. \tag{5.3}$$

The covariance and the correlation are measures of independence of X and Y. Since

$$\mathbb{C}ov(X,Y) = \mathbb{E}((X-\mu_X)(Y-\mu_Y)) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

we have $\mathbb{C}ov(X,Y) = 0$ iff $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Compare Theorem 5.2. X and Y are independent if this relationship holds for all integrable functions of X and Y, not for linear functions only. Therefore, independence of X and Y implies $\mathbb{C}ov(X,Y) = 0$ (X and Y are uncorrelated), but the converse is true only for special distributions.

The Theorem of Cauchy-Schwarz implies that $-1 \le \rho \le 1$, see the exercises.

Example 5.11 *Let X be uniformly distributed on $[-1,1]$ and $P(Z=1) = P(Z=-1) = 1/2$, independent of X. Let $Y = XZ$. Note that both X and Y are uniformly distributed on $[-1,1]$ and $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0$. Therefore*

$$\mathbb{C}ov(X,Y) = \mathbb{E}(XY) = \mathbb{E}(X^2Z) = \mathbb{E}(X^2)\mathbb{E}(Z) = 0.$$

X and Y are thus uncorrelated. They are not independent. We have $|X| = |Y|$. Furthermore, $|X|$ is uniformly distributed on $[0,1]$. With $f(X) = |X|$, $g(Y) = |Y|$ we get $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(X^2) = 1/3$, but $\mathbb{E}(f(X))\mathbb{E}(g(Y)) = \mathbb{E}(|X|)^2 = 1/4$.

Definition 5.12 *Let $X = (X_1, \ldots, X_n)$ be $\mathbb{R}^n$-valued and square-integrable (the variables X_i are square-integrable). The covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^n$ is defined by $\sigma_{ij} = \mathbb{C}ov(X_i, X_j)$.*

Proposition 5.13 *Let $X = (X_1, \ldots, X_n)$ be $\mathbb{R}^n$-valued and square-integrable with covariance matrix Σ. Then*

1. Σ is symmetric and positive semi-definite, i.a. for all $a_1, \ldots, a_n \in \mathbb{R}$,

$$\sum_{i,j=1}^{n} a_i a_j \sigma_{ij} \geq 0. \tag{5.4}$$

2. Let A be a $m \times n$ matrix and $Y = AX$. The covariance matrix Σ_Y of Y is

$$\Sigma_Y = A\Sigma A^T, \tag{5.5}$$

A^T being the transpose of A.

Proof. Let $Z = a_1 X_1 + \cdots + a_n X_n$ and $W = b_1 X_1 + \cdots + b_n X_n$. The covariance of Z and W is

$$\mathbb{C}ov(W, Z) = \sum_{i,j=1}^{n} a_i b_j \sigma_{ij}. \tag{5.6}$$

Apply this to the case $a_1 = b_1, \ldots, a_n = b_n$ to see that (5.4) is the variance of X, which cannot be strictly negative. Apply (5.6) to the rows of A and derive (5.5). □

We remark that if a covariance matrix is not positive definite, then there is at least one linear combination $Z = a_1 X_1 + \cdots + a_n X_n$ with variance 0. In this case, one of the random variables $X_1, \ldots, X_n$ is an affine function of the remaining ones.

5.3 Transformations

Let X be an $\mathbb{R}^n$-valued random variable with density f_X. Let $Y = g(X)$, with g a mapping from the range of X to $\mathbb{R}^n$. How does the distribution of Y look like? Does it have a density f_Y and if, how can it be computed?

Typically, probability distributions are neither absolutely continuous nor have a countable support. However, these two types of distributions are important.

Let us begin with $n = 1$. Let X be a real-valued random variable with density f_X and assume that $Y = g(X)$ with g continuous and strictly increasing. Let $h = g^{-1}$ denote the inverse of g. The c.d.f. of Y is then

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq h(y)) \\ &= F_X(h(y)). \end{aligned}$$

If g is continuous and strictly decreasing, then

$$F_Y(y) = P(Y \leq y)$$

$$
\begin{aligned}
&= P(g(X) \le y) \\
&= P(X \ge h(y)) \\
&= 1 - P(X < h(y)) \\
&= 1 - F_X(h(y)-).
\end{aligned}
$$

If the density f_X is continuous, then $F'_X = f_X$ and

$$f_Y(y) = F'_Y(y) = f_X(h(x))h'(x)$$

if g and hence h is increasing and

$$f_Y(y) = -F'_Y(y) = -f_X(h(x))h'(x)$$

if h is decreasing. In both cases

$$f_Y(y) = f_X(h(x))|h'(x)|. \tag{5.7}$$

Theorem 5.14 *Let X have a continuous density f_X and $Y = g(X)$ with g continuously differentiable and strictly monotone with inverse h. Then Y has a density f_Y, given by* (5.7).

Example 5.15 *Let $X \sim \mathbb{U}(0,1)$ and $Y = -\log X$. Then $Y > 0$, $f_X(x) = 1$ if $0 < x < 1$ and $f_X(x) = 0$ else. Since $g(x) = -\log x$, we have $h(y) = e^{-y}$ and $|h'(y)| = e^{-y}$. Thus*

$$f_Y(y) = \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{else} \end{cases},$$

Y is exponentially distributed with parameter 1.

Example 5.16 *Let X have an invertible c.d.f. F. Let $Y = F(X)$. Then $Y \in [0,1]$ and for $y \in [0,1]$ we have*

$$F_Y(y) = P(Y \le y) = P(F(X) \le y) = P(X \le F^{-1}(y)) = F(F^{-1}(y)) = y.$$

$F(X)$ is uniformly distributed on $[0,1]$.

Example 5.17 *Let X be uniformly distributed on $(0,1)$ and $Y = F^{-1}(X)$ with F an invertible c.d.f. Then*

$$P(Y \le y) = P(F^{-1}(X) \le y) = P(X \le F(y)) = F(y).$$

Example 5.18 *If the transformation is not invertible, Theorem 5.14 cannot be applied directly. Often the c.d.f. of Y can be computed and its derivative gives the density. Let $X \sim N(0,1)$ and $Y = X^2$. We have $Y \ge 0$. For $y \ge 0$ we get*

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}).$$

Its derivative is

$$\begin{aligned} f_Y(y) &= \phi(\sqrt{y})(\sqrt{y})' - \phi(-\sqrt{y})(-\sqrt{y})' \\ &= 2\frac{1}{\sqrt{2\pi}}e^{-y/2}\frac{y^{-1/2}}{2} \\ &= \frac{y^{1/2-1}}{\sqrt{2}\Gamma(\pi)}e^{-y/2}. \end{aligned}$$

For $y < 0$ the density is 0. This is the density of the $\Gamma(1/2, 1/2)$-distribution. It is also called the χ^2-distribution with 1 degree of freedom.

An application of Jacobi's Transformation Formula shows how the densities of random vectors transform. Recall that the Jacobian $J_h(y)$ of a transformation $h : \mathbb{R}^n \to \mathbb{R}^n$ is the matrix of the partial derivatives.

Theorem 5.19 *Let $X = (X_1, \ldots, X_n)^T$ have a density f_X. Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable, injective, with $det(J_g(x)) \neq 0$ for all x. Denote the inverse of g by h. Then $Y = g(X)$ has density*

$$f_Y(y) = \begin{cases} f_X(h(y))|det(J_h(y))| & \text{if } y \text{ is in the range of } g, \\ 0 & \text{otherwise .} \end{cases} \tag{5.8}$$

Example 5.20 *Let $X = (X_1, X_2)$ with X_1 and X_2 independent and standard normal, $\mathbb{E}(X_i) = 0, \mathbb{V}ar(X_i) = 1$. Let $Y = (Y_1, Y_2)$ with $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Then*

$$\begin{aligned} g(x_1, x_2) &= \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix}, \\ h(y_1, y_2) &= \begin{pmatrix} \frac{y_1+y_2}{2} \\ \frac{y_1-y_2}{2} \end{pmatrix}, \\ J_h(y_1, y_2) &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \\ |det(J_h(y_1, y_2)| &= |\frac{1}{2} \times \frac{1}{2} - \frac{1}{2} \times (-\frac{1}{2})| = |\frac{1}{2}|. \end{aligned}$$

Therefore,

$$\begin{aligned} f_Y(y_1, y_2) &= \frac{1}{2}\phi((y_1 + y_2)/2)\,\phi(y_1 - y_2)/2) \\ &= \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-(y_1+y_2)/2)^2/2}\frac{1}{\sqrt{2\pi}}e^{-(y_1-y_2)/2)^2/2} \\ &= \frac{1}{4\pi}e^{-y_1^2/4-y_2^2/4} \\ &= \frac{1}{\sqrt{2\pi 2}}e^{-y_1^2/4} \times \frac{1}{\sqrt{2\pi 2}}e^{-y_2^2/4} \\ &= \phi_{0,2}(y_1)\phi_{0,2}(y_2). \end{aligned}$$

Y_1 and Y_2 are again independent and normally distributed, with expectation 0 and variance 2.

Example 5.21 *Let X and Y be independent $\mathbb{R}$-valued with densities f_X and f_Y. We want to compute the density of $U = X + Y$. We cannot apply Theorem* 5.19*, since the mapping $(X, Y) \to x + y$ is not invertible (in fact it is a mapping from $\mathbb{R}^2$ to $\mathbb{R}$). We define $V = Y$ and in a first step compute the joint density of (U, V). We have $(X, Y) = (U - V, V)$ with Jacobian* 1*. Therefore,*

$$f_{(U,V)}(u, v) = f_X(u - v) f_Y(v).$$

f_U is then

$$f_U(u) = \int f_X(u - v) f_Y(v) dv. \tag{5.9}$$

The density f_U is called the convolution *of f_X and f_Y.*

Example 5.22 *Let $X \sim \Gamma(1, 1)$ and $Y \sim \Gamma(\alpha, 1)$ and $U = X + Y$. The density of U is (for $u > 0$)*

$$\begin{aligned} f_U(u) &= \int f_X(u - v) f_Y(v) dv \\ &= \int I_{\{u-v>0\}} e^{-(u-v)} I_{\{v>0\}} \frac{v^{\alpha-1}}{\Gamma(\alpha)} e^{-v}\, dv \\ &= e^{-u} \frac{1}{\Gamma(\alpha)} \int_0^u v^{\alpha-1}\, dv \\ &= e^{-u} \frac{u^\alpha}{\Gamma(\alpha)\alpha} = e^{-u} \frac{u^{(\alpha+1)-1}}{\Gamma(\alpha + 1)}. \end{aligned}$$

$X + Y$ is $\Gamma(\alpha + 1, 1)$-distributed.

Example 5.23 *Let X and Y be independent and standard normal. We want to compute the density of $U = X/Y$. Again, we cannot apply Theorem* 5.19 *directly. Note that X/Y has the same distribution as $X/|Y|$. Let $V = |Y|$. The density of V is $2\phi(v)$ on $v > 0$. We have $(X, |Y|) = (UV, V)$ with Jacobian matrix*

$$\begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix}$$

and Jacobian determinant v. Therefore,

$$f_{(U,V)}(u, v) = \frac{1}{\sqrt{2\pi}} e^{-u^2 v^2/2} \frac{2}{\sqrt{2\pi}} e^{-v^2/2} v = \frac{v}{\pi} e^{-(u^2+1)v^2/2}.$$

Therefore

$$f_U(u) = \int_0^\infty \frac{v}{\pi} e^{-(u^2+1)v^2/2} = \frac{1}{\pi} \frac{1}{1 + u^2}.$$

X/Y has a Cauchy distribution. The Cauchy distribution is the t-distribution with 1 *degree of freedom.*

5.4 Gaussian Distribution

Let $X_1, \ldots, X_m$ be independent and $X_i \sim N(0,1)$. The density of $X = (X_1, \ldots, X_m)^T$ is the product of the densities of the marginal distributions. For $x = (x_1, x_2, \ldots, x_m)^T$ it is

$$\begin{aligned} f(x_1, \ldots, x_m) &= \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-x_2^2/2} \ldots \frac{1}{\sqrt{2\pi}} e^{-x_m^2/2} \\ &= \frac{1}{(2\pi)^{m/2}} e^{-(x_1^2+x_2^2+\cdots+x_m^2)/2} \\ &= \frac{1}{(2\pi)^{m/2}} e^{-x'x/2}. \end{aligned}$$

expectation and covariance matrix of X are $\mu = 0$ and $\Sigma = I_m$, with $0 = (0, \ldots, 0)^T$ and I_m the $m \times m$-unit matrix. We write $X \sim \mathrm{N}(0, I_m)$.

$X = (X_1, \ldots, X_m)^T$ is normally distributed (Gaussian), if for a matrix A, a vector $\mu = (\mu_1, \ldots, \mu_m)'$ and $Z \sim N(0, I_m)$,

$$X = AZ + \mu.$$

Then

$$\mathbb{E}(X) = \mu, \qquad \mathbb{V}ar(X) = AA^T =: \Sigma.$$

We write $X \sim \mathrm{N}(\mu, \Sigma)$. If A is invertible, then X has a density

$$\phi_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{m/2} \det(\Sigma)^{1/2}} e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2}. \tag{5.10}$$

Properties of the Gaussian distribution:

- Independence: If $X \sim \mathrm{N}(\mu, \Sigma)$. The components of X are independent if and only if they are uncorrelated, i.e. if Σ is diagonal.
- Linear combinations: If $X \sim N(\mu, \Sigma)$ and M a $d \times m$-matrix of rank $d \leq m$, then $MX \sim \mathrm{N}(M\mu, M\Sigma M^T)$. In particular, if $w = (w_1, \ldots, w_m)^T$ is a vector, then $w'X \sim \mathrm{N}(w^T\mu, w^T\Sigma w)$.

  Let $w = e_i$ be the i-th unit vector, then

  $$\begin{aligned} X_i &= e_i^T X \\ \mathbb{E}(X_i) &= e_i^T \mu = \mu_i \\ \mathbb{V}(X_i) &= e_i^T \Sigma e_i = \sigma_{ii}. \end{aligned}$$

  The components X_i are normally distributed, $X_i \sim \mathrm{N}(\mu_i, \sigma_{ii})$.

  Let $\Sigma = UDU^T$ be the spectral decomposition of Σ with U orthogonal and D diagonal. Then $U^T\Sigma U = U^T U D U^T U = D$: The components of $U^T X$ are independent.

- Conditional distribution: Let $X = (X_1, \ldots, X_m)^T$ be normally distributed and $1 \leq k < m$. The conditional distribution of $(X_1, \ldots, X_k)^T$ given $(X_{k+1}, \ldots, X_m)^T$ is normal.

  Let us consider the bivariate case only. Let X and Y be jointly Gaussian, with expectations μ_X, μ_Y, variances σ_X^2, σ_Y^2 and correlation ρ. Then

$$Y \mid (X = x) \sim \mathrm{N}(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)). \tag{5.11}$$

Example 5.24 *Three assets are traded in a market. Let X_i denote the gain of asset i. We assume that the gains (X_1, X_2, X_3) are Gaussian with expectations $\mu_1 = 7$, $\mu_2 = 10$, $\mu_3 = 0$ and covariance matrix*

$$\Sigma = \begin{pmatrix} 10 & 0 & 3 \\ 0 & 15 & 5 \\ 3 & 5 & 5 \end{pmatrix}.$$

A portfolio consists of 100 shares of asset 1 and asset 2 and has gain $G = 100X_1 + 100X_2$. We want to compute the distribution of the gain of the portfolio, given that $X_3 = -5$.

Let

$$G = 100X_1 + 100X_2.$$

We have

$$\begin{pmatrix} G \\ X_3 \end{pmatrix} = \begin{pmatrix} 100 & 100 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}.$$

$(G, X_3)'$ is normally distributed with expectation

$$\mu = \begin{pmatrix} 100 \times 7 + 100 \times 10 \\ 0 \end{pmatrix} = \begin{pmatrix} 1700 \\ 0 \end{pmatrix}$$

and covariance matrix

$$\begin{pmatrix} 100 & 100 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 10 & 0 & 3 \\ 0 & 15 & 5 \\ 3 & 5 & 5 \end{pmatrix} \begin{pmatrix} 100 & 0 \\ 100 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 250000 & 800 \\ 800 & 5 \end{pmatrix}.$$

The correlation of G and X_3 is

$$\rho = \frac{800}{\sqrt{250000 \times 5}} = 0.716.$$

Thus,

$$G \mid X_3 \sim N(a + bX_3, \sigma_G^2(1 - \rho^2)),$$

with

$$b = \frac{\mathbb{C}ov[G, X_3]}{\sigma_{X_3}^2} = \frac{800}{5} = 160,$$

$$a = \mu_G - b\mu_{X_3} = 1700 - 160 \times 0 = 1700,$$

$$\sigma_G^2(1-\rho^2) = 250000(1 - 0.716^2) = 122000.$$

In particular, $a + b(-5) = 1700 - 160 \times 5 = 900$.

$$G \mid (X_3 = -5) \sim N(900, 122000).$$

Example 5.25 *Let* $(X_1, X_2) \sim f(x_2, x_2)$*, with* $f(x_1, x_2) = 2\phi(x_1)\phi(x_2)$*, if* x_1 *and* x_2 *have the same sign (both positive or both negative) and* $f(x_1, x_2) = 0$ *else. We show that* X_1 *is normally distributed and by symmetry* X_2 *is then also normally distributed. However, the vector* (X_1, X_2) *is not. Let* $x_1 > 0$*. Then*

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x_1, x_2)dx_2 &= \int_0^{\infty} 2\phi(x_1)\phi(x_2)dx_2 \\
&= 2\phi(x_1)\int_0^{\infty} \phi(x_2)dx_2 \\
&= 2\phi(x_1)1/2 \\
&= \phi(x_1).
\end{aligned}$$

If $x_1 < 0$ *then*

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x_1, x_2)dx_2 &= \int_{-\infty}^{0} 2\phi(x_1)\phi(x_2)dx_2 \\
&= 2\phi(x_1)\int_{-\infty}^{0} \phi(x_2)dx_2 \\
&= 2\phi(x_1)1/2 \\
&= \phi(x_1).
\end{aligned}$$

Thus $X_1 \sim N(0, 1)$.

5.5 Exercises

Exercise 5.1 *Let* X *and* Y *be independent real-valued with densities* f_X *and* f_Y*. Let* $Y > 0$ *a.s. Compute the density of* $U = XY$ *and the conditional density of* U *given* Y*.*

Exercise 5.2 *Let* X *and* Y *be independent,* $X \sim \Gamma(1, 1)$ *and* $Y \sim \Gamma(\alpha, 1)$*. Prove that* $X + Y \sim \Gamma(\alpha + 1, 1)$.

Exercise 5.3 *Let* X *and* Y *be independent and* $X, Y \sim \mathbb{U}([0, 1])$*. Compute the density of* $X + Y$*.*

Exercise 5.4 *Let* X *be exponentially distributed. Prove that* $P(X > s + t \mid X > s) = P(X > t)$ *for* $s, t \geq 0$*. The exponential distribution is without memory!*

Exercise 5.5 *Prove that for invertible A, (5.10) is the density of $X \sim N(\mu, \Sigma)$, if $X = AZ + \mu$, $\Sigma = AA^T$ and $Z \sim N(0, I_m)$.*

Exercise 5.6 *Let Z be a real-valued random variable and $g, h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ measurable functions, either both increasing or both decreasing. Assume that X and Y are square-integrable. Prove that $X = g(Z)$ and $Y = h(Z)$ are positively correlated, i.e. $\mathbb{C}ov(X, Y) \geq 0$.*

Exercise 5.7 *Let $X = (X_1, X_2, X_3)^T \sim N(\mu, \Sigma)$ with $\mu = (0, 5, -2)^T$ and*

$$\Sigma = \begin{pmatrix} 8.1 & -1.7 & -6.3 \\ -1.7 & 0.4 & 1.3 \\ -6.3 & 1.3 & 6.2 \end{pmatrix}.$$

Compute the distribution of $(X_1 + X_2 + X_3, X_3)^T$ and the conditional distribution of $X_1 + X_2 + X_3 \mid (X_3 = 0)$.

Exercise 5.8 *Prove that $\rho \in [-1, 1]$, where ρ is the correlation of X and Y.*

Exercise 5.9 *Let X and Y be square-integrable. Find constants a, b, s.t. X and $aX + bY$ are uncorrelated.*

Exercise 5.10 *Let X and Y be square-integrable and uncorrelated. Let $\rho \in [-1, 1]$. Find constants a, b, s.t. the correlation of X and $aX + bY$ is ρ.*

Exercise 5.11 *Let $X \sim N(0, 1)$, $P(Z = 1) = P(Z = -1) = 1/2$, independent of X. Let $Y = XZ$. Prove that Y is again standard normal, but (X, Y) is not normal.*

Exercise 5.12 *Let X be exponentially distributed with parameter b, i.e. its density is $f(x) = be^{-bx}$ for $x > 0$ and $f(x) = 0$ else. Compute the density of $Y = \sqrt{X}$.*

Exercise 5.13 *Let X be uniformly distributed on $[0, 1]$. Compute the density of $Y = X/(1 + X)$.*

Exercise 5.14 *Let X be a random variable on $\mathbb{R}$ with distribution function F. Let $Y = X^+$ be the positive part of X. Derive the distribution function of Y.*

Exercise 5.15 *Let X and Y be square-integrable random variables, both with expectations 0, $\sigma^2(X) = 1$, $\sigma^2(Y) = 4$ and correlation coefficient $\rho(X, Y) = 1/4$. Let $U = 3X$ and $V = -5Y$. Compute the covariance and the correlation coefficient of (U, V).*

Chapter 6

Characteristic Functions

6.1 Definition and Properties

Recall that for two vectors $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ the inner product (also called the scalar product) is

$$\langle x, y \rangle = x_1\, y_1 + x_2\, y_2 + \cdots + x_n\, y_n.$$

Furthermore, let $i = \sqrt{-1}$ be the imaginary unit. The exponential e^{is} can be written as $e^{is} = \cos(s) + i \sin(s)$.

Definition 6.1 *Let μ be a probability measure on $\mathbb{R}^n$. The characteristic function of $\hat{\mu} : \mathbb{R}^n \to \mathbb{C}$ is*

$$\hat{\mu}(s) = \int e^{i\langle s,x\rangle} \mu(dx) = \int \cos(\langle s, x\rangle)\mu(dx) + i \int \sin(\langle s, x\rangle)\mu(dx). \tag{6.1}$$

The characteristic function (c.f.) is also called the Fourier transform of μ. If $X \sim \mu$ we denote $\hat{\mu}$ by φ_X.

Theorem 6.2 *(Uniqueness Theorem). The characteristic function characterizes the distribution: If two probability measures have the same characteristic function, then they are the same.*

Theorem 6.3 *The c.f. $\hat{\mu}$ of a probability measure μ is continuous, bounded ($|\hat{\mu}(s)| \leq 1$) with $\hat{\mu}(0) = 1$.*

Proof. The function $s \mapsto e^{i\langle s,x\rangle}$ is continuous and bounded, $|e^{i\langle s,x\rangle}| = 1$. The theorem of dominated convergence implies that for all sequences (s_n) with $s_n \to s$,

$$\lim_{n\to\infty} \hat{\mu}(s_n) = \lim_{n\to\infty} \int e^{i\langle s_n,x\rangle} \mu(dx) = \int \lim_{n\to\infty} e^{i\langle s_n,x\rangle} \mu(dx) = \int e^{i\langle s,x\rangle} \mu(dx) = \hat{\mu}(s).$$

$\hat{\mu}(0) = 1$ holds since $e^{i\langle 0,x\rangle} = 1$. Finally,

$$\begin{aligned}
|\hat{\mu}(s)|^2 &= \left(\int \cos(\langle u,x\rangle)\mu(dx)\right)^2 + \left(\int \sin(\langle u,x\rangle)\mu(dx)\right)^2 \\
&\leq \int \cos(\langle u,x\rangle)^2\mu(dx) + \int \sin(\langle u,x\rangle)^2\mu(dx) \\
&= \int \left(\cos(\langle u,x\rangle)^2 + \sin(\langle u,x\rangle)^2\right)\mu(dx) \\
&= \int 1\mu(dx) = 1.
\end{aligned}$$

□

Example 6.4 *Let $X \sim B(n,p)$. Then*

$$\begin{aligned}
\varphi_X(s) &= \sum_{k=0}^{n} e^{isk}\binom{n}{k}p^k\,(1-p)^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}\left(e^{is}p\right)^k\,(1-p)^{n-k} = \left(pe^{is}+1-p\right)^n.
\end{aligned}$$

Example 6.5 *Let $X \sim P(\lambda)$. Then*

$$\begin{aligned}
\varphi_X(s) &= \sum_{k=0}^{\infty} e^{isk}e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=0}^{\infty} e^{-\lambda}\frac{(e^{is}\lambda)^k}{k!} \\
&= e^{-\lambda(1-e^{is})}.
\end{aligned}$$

Example 6.6 *Let $X \sim N(\mu,\sigma^2)$. First, we compute the c.f. for the standard normal distribution. We have*

$$\varphi_X(s) = \int \cos(sx)\phi(x)\,dx + i\int \sin(sx)\phi(x)\,dx.$$

Since ϕ is symmetric, $\phi(-x) = \phi(x)$ and $\sin(-x) = -\sin(x)$, the imaginary part of the c.f. is 0. Theorem 6.8 implies that φ_X is differentiable. We have

$$\varphi_X'(s) = \int (\cos(sx))'\,\phi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int -\sin(sx)xe^{-x^2/2}\,dx.$$

Integration by parts gives

$$\varphi_X'(s) = -\frac{1}{\sqrt{2\pi}}\int s\cos(sx)xe^{-x^2/2}\,dx = -s\varphi_X(s).$$

To solve this differential equation, i.e.

$$\frac{\varphi_X'(s)}{\varphi_X(s)} = -s,$$

note that the l.h.s. is the derivative of $\log \varphi_X(s)$. *Thus,*

$$\log \varphi_X(s) = -\frac{s^2}{2} + c_1$$

and

$$\varphi_X(s) = c_2 e^{-s^2/2}.$$

Finally, $\varphi_X(0) = 1$ *implies*

$$\varphi_X(s) = e^{-s^2/2}.$$

If $X \sim N(\mu, \sigma^2)$, *then* $X = \mu + \sigma Z$ *with* $Z \sim N(0,1)$. *Therefore,*

$$\varphi_X(s) = \mathbb{E}(e^{is(\mu+\sigma Z)} = e^{is\mu}\mathbb{E}(e^{is\sigma Z}) = e^{is\mu}\varphi_Z(s\sigma) = e^{is\mu - \sigma^2 s^2/2}.$$

Example 6.7 *Let* X *be real-valued with c.f.* φ_X *and* $Y = a + bX$. *Then*

$$\varphi_Y(s) = e^{isa}\varphi_X(bs).$$

More generally, let X *be* $\mathbb{R}^n$*-valued,* $Y = a + BX$, *where* $a \in \mathbb{R}^m$ *and* B *and* $m \times n$ *matrix. Then for* $s \in \mathbb{R}^m$,

$$\varphi_Y(s) = e^{i\langle s,a\rangle}\varphi_X(B^T s).$$

Theorem 6.8 *Let* X *be a real-valued random variable. If* $\mathbb{E}(|X|^k) < \infty$, *then* φ_X *is* k*-times continuously differentiable and*

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k). \tag{6.2}$$

Proof. We give the proof for $k = 1$ only. Note that if we are allowed to interchange taking the derivative and taking expectation, then

$$\varphi_X'(0) = \lim_{s\to 0} \frac{\varphi_X(s) - 1}{s} = \lim_{s\to 0} \mathbb{E}\left(\frac{e^{isX} - 1}{s}\right) = \mathbb{E}\left(\lim_{s\to 0} \frac{e^{isX} - 1}{s}\right) = \mathbb{E}(iX).$$

We have

$$\left|\frac{e^{isX} - 1}{s}\right| \le |X|.$$

Therefore, both the real part and the imaginary part of the difference ratio are dominated by $|X|$. If X is integrable, we may apply Lebesgue's dominated convergence theorem. □

6.2 Sums of Random Variables and the Central Limit Theorem

Theorem 6.9 *Let* $X = (X_1, \ldots, X_n)$ *be* $\mathbb{R}^n$*-valued.*

1. *The* $\mathbb{R}$*-valued random variables* $X_1, \ldots, X_n$ *are independent if and only if for all* $s = (s_1, \ldots, s_n)$

$$\varphi_X(s) = \prod_{k=1}^{n} \varphi_{X_k}(s_k). \tag{6.3}$$

2. Let $Y = u_1X_1 + u_2X_2 + \cdots + u_nX_n$, with $X_1, \ldots, X_n$ independent. The c.f. of Y is

$$\varphi_{u_1X_1+\cdots+u_nX_n}(s) = \prod_{k=1}^{n} \varphi_{X_k}(s\,u_k). \tag{6.4}$$

3. Let the random variables be additionally identically distributed. Then

$$\varphi_{X_1+\cdots+X_n}(s) = \varphi_{X_1}(s)^n. \tag{6.5}$$

Example 6.10 *Let $X = (X_1, \ldots, X_n)$ be Gaussian with expectation μ and covariance matrix Σ. To compute the c.f. of X, let $s = (s_1, \ldots, s_n)$. $\langle s, X\rangle = s_1X_1 + \cdots + s_nX_n$ is again normal with expectation $s^T\mu$ and variance $s^T\Sigma s$. Therefore,*

$$\varphi_X(s) = \varphi_{s_1X_1+\cdots+s_nX_n}(1) = e^{is^T\mu - s^T\Sigma s/2}.$$

Example 6.11 *Let X and Y be independent Poisson distributed with parameter λ_1 and λ_2 resp. The c.f. of $X + Y$ is*

$$\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s) = e^{-\lambda_1(1-e^{is})}\, e^{-\lambda_2(1-e^{is})} = e^{-(\lambda_1+\lambda_2)(1-e^{is})},$$

$X + Y$ therefore again Poisson with parameter $\lambda_1 + \lambda_2$.

Example 6.12 *Let $X = X_1 + \cdots + X_n$ with $X_1, \ldots, X_n$ i.i.d. and Bernoulli distributed, i.e. X has a binomial distribution with parameter n and p. The expectation of X_1 is p and its variance is $p(1-p)$. Let*

$$Y_n = \frac{X_1 + \cdots + X_n - np}{\sqrt{n}}.$$

Y_n has c.f.

$$\varphi_{Y_n}(s) = \left(pe^{is(1-p)/\sqrt{n}} + (1-p)e^{-isp/\sqrt{n}}\right)^n.$$

In the following $O(n^{-3/2})$ are terms of order $n^{-3/2}$. We have

$$\begin{aligned} pe^{is(1-p)/\sqrt{n}} + (1-p)e^{-isp/\sqrt{n}} &= p\left(1 + \frac{is}{\sqrt{n}}(1-p) + \frac{(is)^2}{2n}(1-p)^2 + O(n^{-3/2})\right) \\ &+ (1-p)\left(1 - \frac{is}{\sqrt{n}}p + \frac{(is)^2}{2n}p^2 + O(n^{-3/2})\right) \\ &= 1 - \frac{p(1-p)}{2n}s^2 + O(n^{-3/2}). \end{aligned}$$

Therefore, for $n \to \infty$,

$$\varphi_{Y_n}(s) = \left(1 - \frac{p(1-p)}{2n}s^2 + O(n^{-3/2})\right)^n \to e^{-p(1-p)s^2/2}.$$

The c.f. of Y_n converges to the c.f. of the normal distribution with the same expectation 0 and the same variance $p(1-p)$. The distribution of Y_n converges in a certain sense to the normal distribution. To make this statement precise, we define

Definition 6.13 *Let μ_n and μ be probability measures on $\mathbb{R}^n$. The sequence (μ_n) converges weakly to μ if for all bounded and continuous functions $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu. \tag{6.6}$$

Theorem 6.14 *(Lévy's Continuity Theorem). Let μ_n be probability measures on $\mathbb{R}^n$ with c.f. $\hat{\mu}$.*

1. *If (μ_n) converges weakly to a probability measure μ, then $\hat{\mu}_n(s) \to \hat{\mu}(s)$ for all $s \in \mathbb{R}^n$.*
2. *If $(\hat{\mu}_n(s))$ converges to a function $f(s)$ for all $s \in \mathbb{R}^n$ and if f is continuous at 0, then f is the characteristic function of a probability measure μ and (μ_n) converges weakly to μ.*

The generalization of Example 6.12 is the Central Limit Theorem (CLT): If (X_k) are i.i.d. and independent with finite expectation μ and finite variance σ^2. Let

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}}.$$

Then, given regularity conditions, the distribution of Y_n converges weakly to the normal distribution with expectation 0 and variance σ^2. A proof expands the c.f. of Y_n into a quadratic polynomial and a remainder term. The regularity conditions imply that, exactly as in the example, the remainder term is of order smaller than $1/n$. This is the case, for instance, if the c.f. of Y_n is three times differentiable. The existence of a third moment guarantees the existence of a third derivative, see Theorem 6.8.

Theorem 6.15 *(Central Limit Theorem). Let (X_k) be a sequence of i.i.d. random variables with variance $\sigma^2 > 0$. Then the distribution of*

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\sigma}$$

converges weakly to the standard normal distribution.

Let us remark that weak convergence implies the convergence of the c.d.f. at all points, where the c.d.f. of the limit distribution is continuous. Since Φ is continuous, we have

$$\lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\sigma} \leq x\right) = \Phi(x)$$

for all $x \in \mathbb{R}$.

6.3 Exercises

Exercise 6.1 *Compute the characteristic function of the uniform distribution on $[-a, a]$.*

Exercise 6.2 *Compute the characteristic function of the geometric distribution distribution. Compute the* $\mathbb{E}(X)$ *by means of the c.f.*

Exercise 6.3 $\hat{\mu}(s) = e^{-|s|}$ *is the c.f. of a real-valued random variable* X*. Prove that* X *is not integrable. Remark:* $e^{-|s|}$ *is the c.f. of the Cauchy distribution.*

Exercise 6.4 *Let* $c > 0$ *and* $0 < \alpha \leq 2$*.* $\hat{\mu}(s) = e^{-c|s|^\alpha}$ *is the c.f. of a real-valued random variable* X*. Special cases are the Cauchy distribution for* $\alpha = 1$ *and the normal distribution for* $\alpha = 2$*.*

1. *Prove that* X *is not integrable for* $\alpha \leq 1$ *and not square-integrable for* $1 < \alpha < 2$*.*
2. *Prove that if* $X_1, \ldots, X_n$ *are i.i.d. with c.f.* $\hat{\mu}(s) = e^{-c|s|^\alpha}$*, then* $(X_1 + \cdots + X_n)/n^{1/\alpha}$ *and* X_1 *have the same distribution.*

Exercise 6.5 *The c.f. of the gamma distribution with parameter* α *and* β *is*

$$\hat{\mu}(s) = \left(\frac{\beta^2 + is\beta}{\beta^2 + s^2}\right)^\alpha.$$

Prove that X *and* Y *independent,* $X \sim \Gamma(\alpha_1, \beta)$ *and* $Y \sim \Gamma(\alpha_2, \beta)$*, then* $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$*. It is sufficient to give a proof for* $\beta = 1$*.*

Exercise 6.6 *Let* X *be a real-valued random variable. Show that* φ_X *is real (i.e. the imaginary part of* $\varphi_X(s)$ *is* 0 *for all* s*) if and only if* X *and* $-X$ *have the same distribution. In particular, if* X *and* Y *are i.i.d. then* $X - Y$ *has a symmetric distribution.*

Exercise 6.7 *Let* X *and* Y *be independent,* $X \sim B(n, p)$*,* $Y \sim B(m, p)$*. Using characteristic functions compute the distribution of* $X + Y$*.*

Exercise 6.8 *Prove that* $\varphi_{a+bX}(s) = e^{isa}\varphi_X(bs)$*.*

Exercise 6.9 *Let* $X_1, X_2, \ldots$ *be i.i.d. and* $N \sim P(\lambda)$ *independent of the* $X_i's$*. Let* $Y = \sum_{i=1}^N X_i$ *with* $Y = 0$ *if* $N = 0$*. Prove that*

$$\varphi_Y(s) = e^{\lambda(\varphi_{X_1}(s)-1)}.$$

Compute the expectation and the variance of Y *in terms of* $\mathbb{E}(X)$ *and* σ_X^2*.*

Exercise 6.10 *Let* $X_n \sim P(n)$*. Prove that the distribution of* $(X_n - n)/\sqrt{n}$ *converges weakly to the standard normal distribution.*

Chapter 7

Conditional Expectation

As a motivation for the concept of the conditional expectation, consider the following problem of predicting a random variable Y: Let a probability space $(\Omega, \mathcal{F}, P)$ and a square integrable random variable $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be given.

To predict Y without further information, a real number $c = \hat{Y}$ has to chosen, s.t. $\mathbb{E}((Y-c)^2)$ is a small as possible. We know that the solution is the expectation $c = \mathbb{E}(Y)$. With this choice $\mathbb{E}((Y-c)^2)$ is then the variance of Y.

Now assume that the prediction of Y may be based on information provided by a random variable X, i.e. one has to choose a function g and predict Y by $\hat{Y} = g(X)$. If (X, Y) has a joint distribution and if a conditional distribution of Y given $X = x$ can be defined, then $g(x)$ is the expectation of this conditional distribution and $\hat{Y} = g(X)$. Note that $g(X)$ is a random variable, since X is random.

Theorem 7.1 *(Causality Theorem). Let X be an $\mathbb{R}^n$-valued random variable on the measurable space $(\Omega, \mathcal{F})$. Let Y be an $\mathbb{R}$ valued random variable. Y is $\sigma(X)$-measurable if and only if a measurable $g : \mathbb{R}^n \to \mathbb{R}$ exists such that $Y = g(X)$.*

Proof. Only one direction has to be shown. Let us recall that $\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{B}^n\}$. As usual, we first prove the theorem for the special case that Y is simple, i.e. $Y = \sum_{i=1}^k c_i I_{A_i}$ with $A_i \in \sigma(X)$ and $(A_1, \ldots, A_k)$ a partition. Therefore, $Y = \sum_{i=1}^k c_i I_{X^{-1}(B_i)}$ with $B_i \in \mathcal{B}^n$ and $(B_1, \ldots, B_k)$ a partition. Then, if we define g by $g(x) = c_i$ if $x \in B_i$, we have $Y = g(X)$.

In the general case, $Y = \lim_{m\to\infty} Y_m$ with Y_m simple and $\sigma(X)$-measurable. Therefore, $Y_m = g_m(X)$. If we define $g(x) = \limsup_{m\to\infty} g_m(x)$, then g is measurable and $Y = g(X)$. □

Given the *Causality Theorem* the problem of predicting Y may be generalized. Let a probability space $(\Omega, \mathcal{F}, P)$, a sub σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and a square integrable random variable Y be given. Find the $\mathcal{G}$-measurable random variable $\hat{Y}$ that minimizes $\mathbb{E}((Y-\hat{Y})^2)$. Note that $L^2(\mathcal{G})$, the set of $\mathcal{G}$-measurable random variables, is a subspace of $L^2(\mathcal{F})$ and $\hat{Y}$ is the projection of Y onto this

subspace. Projections have (in this case) the property that $Y - \hat{Y}$ is orthogonal to $L^2(\mathcal{G})$, i.e. for all $Z \in L^2(\mathcal{G})$, $\mathbb{E}(Z(Y - \hat{Y})) = 0$, which may be written as

$$\mathbb{E}(Z\hat{Y}) = \mathbb{E}(ZY).$$

The random variable $\hat{Y}$ is called the conditional expectation of Y given $\mathcal{G}$ and denoted by $\mathbb{E}(Y \mid \mathcal{G})$. It is uniquely defined by two properteis: $\mathbb{E}(Y \mid \mathcal{G})$ is $\mathcal{G}$-measurable and $\mathbb{E}(Z\hat{Y}) = \mathbb{E}(ZY)$ holds for all $Z \in L^2(\mathcal{G})$.

Definition 7.2 *Let a probability space $(\Omega, \mathcal{F}, P)$, a sub σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and a random variable Y be given. A random variable $\mathbb{E}(Y \mid \mathcal{G})$ is called the conditional expectation of Y given $\mathcal{G}$ if it satisfies*

$$\mathbb{E}(Y \mid \mathcal{G}) \qquad \text{is } \mathcal{G}\text{-measurable,} \tag{7.1}$$

$$\mathbb{E}(Z\mathbb{E}(Y \mid \mathcal{G})) = \mathbb{E}(ZY) \text{ for all bounded and } \mathcal{G}\text{-measurable } Z. \tag{7.2}$$

Theorem 7.3 *Let a probability space $(\Omega, \mathcal{F}, P)$, a sub σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and a random variable Y be given. If Y is integrable, then $\mathbb{E}(Y \mid \mathcal{G})$ exists and is a.s. unique in the sense that if any other r.v. $\hat{Y}$ satisfies* (7.1) *and* (7.2)*, then $\hat{Y} = \mathbb{E}(Y \mid \mathcal{G})$ a.s.*

Theorem 7.4 *(Properties). Let a probability space $(\Omega, \mathcal{F}, P)$, a sub σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and integrable random variables Y, Y_1, Y_2 be given.*

1. *$\mathbb{E}(Y \mid \mathcal{G}) = Y$ if and only if Y is $\mathcal{G}$-measurable.*
2. *$\mathbb{E}(Y \mid \mathcal{G}) = \mathbb{E}(Y)$ if Y is independent of $\mathcal{G}$.*
3. *$\mathbb{E}(\alpha Y_1 + \beta Y_2 \mid \mathcal{G}) = \alpha\mathbb{E}(Y_1 \mid \mathcal{G}) + \beta\mathbb{E}(Y_2 \mid \mathcal{G})$. $\alpha, \beta \in \mathbb{R}$.*
4. *If $Y \geq 0$, then $\mathbb{E}(Y \mid \mathcal{G}) \geq 0$.*
5. *$\mathbb{E}(\mathbb{E}(Y \mid \mathcal{G})) = \mathbb{E}(Y)$.*
6. *If $\mathcal{H} \subseteq \mathcal{G}$ is a sub σ-algebra, then $\mathbb{E}(\mathbb{E}(Y \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(Y \mid \mathcal{H})$ (Tower property).*
7. *If Z is bounded and $\mathcal{G}$-measurable, then $\mathbb{E}(ZY \mid \mathcal{G}) = Z\mathbb{E}(Y \mid \mathcal{G})$.*
8. *If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}(Y \mid \mathcal{G}) = \mathbb{E}(Y)$.*
9. *If $Y = g(X, Z)$, Z independent of X, then $\mathbb{E}(Y \mid \sigma(X)) = h(X)$, with $h(x) = \mathbb{E}(g(x, Z))$.*
10. *If f is measurable and convex, then $\mathbb{E}(f(Y) \mid \mathcal{G}) \geq f(\mathbb{E}(Y \mid \mathcal{G}))$.*
11. *Let Y be square integrable. Then* (7.2) *holds for all square integrable and $\mathcal{G}$-measurable Z.*

12. Let $\hat{Y}$ be $\mathcal{G}$-measurable. If $\mathbb{E}(YI_A) = \mathbb{E}(\hat{Y}I_A)$ for all $A \in \mathcal{G}$, then $\hat{Y} = \mathbb{E}(Y \mid \mathcal{G})$.

Proof.

1. Trivial.

2. We have to show that the constant function $\mathbb{E}(Y)$ is the conditional expectation of Y given $\mathcal{G}$. Constant functions are measurable w.r.t. all σ-algebras. To check (7.2), let Z be bounded and $\mathcal{G}$-measurable. Then Z and Y are independent and therefore,

$$\mathbb{E}(ZY) = \mathbb{E}(Z)\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y)Z).$$

3. We have to show that $\alpha\mathbb{E}(Y_1 \mid \mathcal{G}) + \beta\mathbb{E}(Y_2 \mid \mathcal{G})$ is the conditional expectation of $\alpha Y_1 + \beta Y_2$ given $\mathcal{G}$. It is obviously $\mathcal{G}$-measurable. To check (7.2), let Z be bounded and $\mathcal{G}$-measurable. Then

$$\mathbb{E}(Z(\alpha Y_1 + \beta Y_2)) = \alpha\mathbb{E}(ZY_1) + \beta\mathbb{E}(ZY_2)$$

$$= \alpha\mathbb{E}(Z\mathbb{E}(Y_1 \mid \mathcal{G})) + \beta\mathbb{E}(Z\mathbb{E}(Y_2 \mid \mathcal{G})) = \mathbb{E}(Z(\alpha\mathbb{E}(Y_1 \mid \mathcal{G}) + \beta\mathbb{E}(Y_2 \mid \mathcal{G}))).$$

4. Let A denote the event $\{\mathbb{E}(Y \mid \mathcal{G}) < 0\}$. I_A is bounded and $\mathcal{G}$-measurable. $P(A) > 0$ is not possible, since otherwise

$$0 > \mathbb{E}(I_A\mathbb{E}(Y \mid \mathcal{G})) = \mathbb{E}(I_AY) \geq 0.$$

5. Let, in (7.2), $Z = 1$.

6. We have to prove that $\mathbb{E}(\mathbb{E}(Y \mid \mathcal{G}) \mid \mathcal{H})$ is the conditional expectation of Y given $\mathcal{H}$. It is obviously $\mathcal{H}$-measurable. Let Z be bounded and $\mathcal{H}$-measurable. Since $\mathcal{H} \subseteq \mathcal{G}$, it is also $\mathcal{G}$-measurable. Therefore

$$\mathbb{E}(Z\mathbb{E}(\mathbb{E}(Y \mid \mathcal{G}) \mid \mathcal{H})) = \mathbb{E}(Z\mathbb{E}(Y \mid \mathcal{G})) = \mathbb{E}(ZY).$$

7. We have to prove that $Z\mathbb{E}(Y \mid \mathcal{G})$ is the conditional expectation of ZY given $\mathcal{G}$. It is obviously $\mathcal{G}$-measurable. Let U be bounded and $\mathcal{G}$-measurable. Then UZ is bounded and $\mathcal{G}$-measurable. Therefore,

$$\mathbb{E}(U(Z\mathbb{E}(Y \mid \mathcal{G}))) = \mathbb{E}(UZ\mathbb{E}(Y \mid \mathcal{G})) = \mathbb{E}(UZY).$$

8. If $\mathcal{G} = \{\emptyset, \Omega\}$, then only the constant functions are measurable. From 5, it follows that this constant is $\mathbb{E}(Y)$.

9. Let $h(x) = \mathbb{E}(g(x, Z))$. The Theorem of Tonelli-Fubini implies that h is measurable (w.r.t. $\sigma(X)$) and $h(X)$ is integrable. Let Z be bounded and measurable w.r.t. $\sigma(X)$. The Causality

Theorem implies that $Z = u(X)$ for a bounded and measurable function u. To show that $h(X)$ is the conditional expectation of Y given $\sigma(X)$, note that

$$\mathbb{E}(ZY) = \mathbb{E}(u(X)g(X,Z)) \quad \text{and} \quad \mathbb{E}(Zh(X)) = \mathbb{E}(u(X)h(X)),$$

and the two expectations are the same, again by the Theorem of Tonelli-Fubini.

10. No proof.

11. No proof.

12. No proof.

□

$\mathbb{E}(Y \mid \sigma(X))$ is abbreviated by $\mathbb{E}(Y \mid X)$.

Example 7.5 *Let the σ-algebra $\mathcal{G}$ be generated by the partition $(B_1, \ldots, B_n)$. Any $\mathcal{G}$-measurable function is a linear combination of the indicator functions $I_{B_1}, \ldots, I_{B_n}$. Therefore, $\mathbb{E}(Y \mid \mathcal{G}) = \sum_{i=1}^n c_i I_{B_i}$. To identify the numbers c_k, let $Z = I_{B_k}$. Z is bounded and $\mathcal{G}$-measurable. From* (7.2) *we get*

$$\mathbb{E}(ZY) = \mathbb{E}(Z\mathbb{E}(Y \mid \mathcal{G})),$$

i.e.

$$\mathbb{E}(I_{B_k}Y) = \mathbb{E}(I_{B_k}\sum_{i=1}^n c_i I_{B_i}) = \sum_{i=1}^n c_i \mathbb{E}(I_{B_k}I_{B_i}) = c_k\mathbb{E}(I_{B_k}) = c_k P(B_k),$$

and therefore

$$c_k = \frac{\mathbb{E}(I_{B_k}Y)}{P(B_k)}.$$

7.1 Exercises

Exercise 7.1 *Let (X,Y) be bivariate Gaussian. Compute $\mathbb{E}(Y \mid X)$ and $\mathbb{E}(Y^2 \mid X)$.*

Exercise 7.2 *Let Y be square integrable. Prove that $\mathbb{E}(Y \mid \mathcal{G})$ and $Y - \mathbb{E}(Y \mid \mathcal{G})$ are uncorrelated.*

Exercise 7.3 *Let Y be square integrable. Prove that*

$$\sigma^2 = \mathbb{E}((Y - \mathbb{E}(Y \mid \mathcal{G}))^2) + \mathbb{E}((\mathbb{E}(Y \mid \mathcal{G}) - \mathbb{E}(Y))^2).$$

Conclude that $\mathbb{E}(Y \mid \mathcal{G})$ is also square integrable.

Exercise 7.4 *Let $X_1, \ldots, X_n$ be i.i.d. and integrable. Let $S = X_1 + X_2 + \cdots + X_n$. Find $\mathbb{E}(X_1 \mid S)$.*

Exercise 7.5 *Let $X \sim \mathbb{U}([-1,1])$. Compute $\mathbb{E}(|X| \mid X)$ and $\mathbb{E}(X \mid |X|)$. Compute also $\mathbb{E}(X \mid |X|)$ for $X \sim \mathbb{U}([-1,2])$ and for $X \sim f$, with f a density.*

Exercise 7.6 *Let S_t, $t = 0, 1, 2$ denote the value of an asset at time t. Assume that $S_1 = S_0 e^{\mu+\sigma X_1}$ and $S_2 = S_0 e^{2\mu+\sigma(X_1+X_2)}$, with $\sigma > 0$, S_0, X_1, X_2 independent and both X_1 and X_2 Gaussian with expectation 0 and variance 1. Compute $\mathbb{E}(S_2 \mid S_1)$.*

Exercise 7.7 *Show that if $|Y| \leq c$, then $|\mathbb{E}(Y \mid \mathcal{G})| \leq c$.*

Exercise 7.8 *Let Y be square integrable, $\mathbb{E}(Y \mid X) = X$ and $\mathbb{E}(Y^2 \mid X) = X^2$. Show $Y = X$ a.s.*

Exercise 7.9 *Let $X \sim P(\lambda)$ (Poisson). Let, conditional on $X = x$, $Y \sim B(x, p)$. Compute $\mathbb{E}(Y \mid X)$ and $\mathbb{E}(X \mid Y)$.*

Chapter 8

Appendix

8.1 Preliminaries and Notation

A. Sets.

1. Set Operations. Let Ω denote a set and A, B, A_i subsets of Ω.

$A \cup B,$	$\bigcup_{i=1}^{\infty} A_i$	union
$A \cap B,$	$\bigcap_{i=1}^{\infty} A_i$	intersection
A^c		complement
$A \backslash B$	$= A \cap B^c$	difference
$\mathcal{P}(\Omega)$	$= \{A \mid A \subseteq \Omega\}$	power set
$\emptyset$		empty set

$\mathcal{P}(\Omega)$ is also denoted by 2^{Ω}.

(A_i) is *increasing*, if $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$.

(A_i) is *decreasing*, if $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$.

(A_i) is *monotone*, if it is increasing or decreasing.

De Morgan's Laws: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$.

2. Cartesian Products. $A \times B$, the Cartesian product of A and B is

$$A \times B = \{(x, y) \mid x \in A,\, y \in B\}.$$

$A^2 = A \times A$, $A^n = A \times A \times \cdots \times A$ (n-times).

$x \in A_1 \times A_2 \times \cdots \times A_n$, then $x = (x_1, \ldots, x_n)$, $x_i \in A_i$ are the *components* of x.

3. Countable Sets. An infinite set A (i.e. a set with infinitely many elements) is *countable* (denumerable), if it can be written as a sequence, $A = \{a_1, a_2, \ldots\}$. More precisely, A is countable,

if there exists a function f of $\mathbb{N}$ onto A.

- $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ are countable, $\mathbb{R}$ is uncountable.
- If all sets A_n are countable, then $A = \bigcup_{n=1}^{\infty} A_n$ is countable.
- If A and B are countable, then $A \times B$ is countable.
- Let $p_i, i \in I$ be positive, $p_i > 0$. If $\sum_{i \in I} p_i < \infty$, then I is countable (or finite).

B. Functions.

1. Definitions. Let X, Y be nonempty sets. A *function* (mapping) $f : X \to Y$ is a set of pairs $(x, f(x))$ s.t. for all $x \in X$ there exists exactly one $f(x)$. X is the *domain* of f, Y the *codomain* of f, $f(X) = \{f(x) \mid x \in X\}$ the *range* of f.

f is *injective* (one to one, 1-1) if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

f is *surjective* (onto) if for all $y \in Y$ there exists (at least one) $x \in X$ with $y = f(x)$.

f is *bijective* if it is injective and surjective.

2. Preimages. Let $f : X \to Y$, $A \subseteq X$, $B \subseteq Y$.

$$\begin{aligned} f(A) &= \{f(x) \mid x \in A\} && \text{image of } A \\ f^{-1}(B) &= \{x \mid f(x) \in B\} && \text{preimage of } B \end{aligned}$$

Facts:

$$\begin{aligned} f^{-1}(B_1 \cup B_2) &= f^{-1}(B_1) \cup f^{-1}(B_2) \\ f^{-1}(B_1 \cap B_2) &= f^{-1}(B_1) \cap f^{-1}(B_2) \\ f^{-1}(B^c) &= (f^{-1}(B))^c \\ f(A_1 \cup A_2) &= f(A_1) \cup f(A_2) \\ f(A_1 \cap A_2) &\subseteq f(A_1) \cap f(A_2) \\ f(A_1 \cap A_2) &= f(A_1) \cap f(A_2) \quad \text{if } f \text{ is injective} \\ f(f^{-1}(B)) &= f(X) \cap B \\ A &\subseteq f^{-1}(f(A)) \end{aligned}$$

Example. Let $X = Y = \mathbb{R}$, $f(x) = x^2$. If $A_1 = [0, \infty)$, $A_2 = (-\infty, 0]$, then $f(A_1) = f(A_2) = [0, \infty)$, $f(A_1 \cap A_2) = f(\{0\}) = \{0\}$. Furthermore, $f^{-1}(f(A_1)) = \mathbb{R} \neq A_1$. □

3. Simple Functions. Let $\Omega \neq \emptyset$. The *indicator* function I_A of a subset A of Ω is defined as $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ if $x \notin A$.

A function $f : \Omega \to \mathbb{R}$ is called *simple*, if $f(\Omega)$ is finite. There exists a canonical representation of simple functions: Let $f(\Omega) = \{y_1, \ldots, y_n\}$ have n elements. Let $A_i = f^{-1}(\{y_i\})$. Then

$$f = \sum_{i=1}^{n} y_i I_{A_i}.$$

C. Real Numbers.

1. Order. $\mathbb{R}$ is an *ordered* set, $x \leq y$. Let $A \subseteq \mathbb{R}$. An *upper bound* of A is a real number y s.t. for all $x \in A$, $x \leq y$. A *lower bound* of A is a real number y s.t. for all $x \in A$, $y \leq x$.

- If A is bounded from above, it has an upper bound. The set of upper bounds contains a smallest element y, the least upper bound of A, called the *supremum* of A, $y = \sup A$. If A is not bounded from above, $\sup A = \infty$.
- If A is bounded from below, it has a lower bound. The set of lower bounds contains a greatest element y, called the *infimum* of A, $y = \inf A$. If A is not bounded from below, $\inf A = -\infty$.
- If A has a maximal element it is called the *maximum*, $\max A$. In that case, $\sup A = \max A$.
- If A has a minimal element it is called the *minimum*, $\min A$. In that case, $\inf A = \min A$.

Example. Let $A = [0, 1)$. $\max A$ does not exist. $\sup A = 1$. $\min A$ exists and $\min A = \inf A = 0$.

Let $A = (-2, \infty)$. $\max A$ and $\min A$ do not exist, $\sup A = \infty$, $\inf A = -2$.

2. Convergence. Let $x \in A$. A *neighborhood* A of x is a set s.t. there exists an open interval (a, b) with $x \in (a, b)$ and $(a, b) \subseteq A$. An *open set* is a union of open intervals. Complements of open sets are called *closed*.

Example. $(0, 1)$ is open. $(2, \infty)$ is open. $[1, 2]$ is closed ($[1, 2] = (-\infty, 1) \cup (2, \infty))^c$). $(0, 1]$ is neither open nor closed. $\{1/n \mid n \in \mathbb{N}\}$ is neither open nor closed. $\{1/n \mid n \in \mathbb{N}\} \cup \{0\}$ is closed. □

A *sequence* (x_n) is a function with domain $\mathbb{N}$. Let $x_n \in \mathbb{R}$. The sequence (x_n) converges to a *limit* x, if every neighborhood of x contains all but finitely many x_n. $x_n \to x$, $\lim_{n\to\infty} x_n = x$.

- Every bounded increasing sequence has a limit.
- Every bounded monotone sequence has a limit.
- An increasing sequence converges (if it is bounded) or diverges to ∞.
- Every bounded sequence has a converging subsequence.
- If F is closed and $x_n \in F$ for all n, then $\lim x_n \in F$.

Limits of subsequences are called *accumulation points.*

$$\begin{array}{ll} \limsup\limits_{n\to\infty} x_n & = \lim_{n\to\infty} \sup_{m\geq n} x_m \qquad \textit{limes superior} \\ \liminf\limits_{n\to\infty} x_n & = \lim_{n\to\infty} \inf_{m\geq n} x_m \qquad \textit{limes inferior} \end{array}$$

If (x_n) is bounded from above, $\limsup_{n\to\infty} x_n < \infty$. Furthermore, $\limsup_{n\to\infty} x_n = \inf_{n\geq 1} \sup_{m\geq n} x_m$. $\limsup x_n$ is the greatest accumulation point of (x_n).

If (x_n) is bounded from below, $\liminf_{n\to\infty} x_n > -\infty$. Furthermore, $\liminf_{n\to\infty} x_n = \sup_{n\geq 1} \inf_{m\geq n} x_m$. $\liminf x_n$ is the smallest accumulation point of (x_n).

Example. If $x_n = (-1)^n$, then $\limsup_{n\to\infty} x_n = 1$, $\liminf_{n\to\infty} x_n = -1$. If $x_n = (-1)^n n/(n+1)$, then $\limsup_{n\to\infty} x_n = 1$, $\liminf_{n\to\infty} x_n = -1$. In both cases, 1 and -1 are the only accumulation points. □

3. Convergence in $\mathbb{R}^m$. Let $x, y \in \mathbb{R}^m$, $x = (x_1, \ldots, x_m)$, $y = (y_1, \ldots, y_m)$. The norm of x is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_m^2}.$$

A sequence (x_n) $(\in \mathbb{R}^m)$ converges to y $(\in \mathbb{R}^m)$ if $\|x_n - y\| \to 0$. $x_n \to y$, $y = \lim_{n\to\infty} x_n$.

$y = \lim_{n\to\infty} x_n$ is equivalent to the convergence of all components of (x_n) to the components of y.

4. Continuity. Let $\Omega \subseteq \mathbb{R}^m$ and $f : \Omega \to \mathbb{R}$. f is continuous at x, if $x_n \to x$ implies $f(x_n) \to f(x)$. f is continuous, if it is continuous at all $x \in \Omega$. A function $f : \Omega \to \mathbb{R}^k$ is continuous, if all its components $f_1, \ldots, f_k$ are continuous.

- $f : \Omega \to \mathbb{R}^k$ is continuous, iff the preimages of all open sets $U \subseteq \mathbb{R}^k$ are open.
- $f : \Omega \to \mathbb{R}^k$ is continuous, iff the preimages of all closed sets $F \subseteq \mathbb{R}^k$ are closed.
- $f : \Omega \to \mathbb{R}$ is continuous, iff the preimages of all open intervals are open.
- $f : \Omega \to \mathbb{R}$ is continuous, iff the preimages of all closed intervals are closed.

5. Convergence of Functions. Let $f_n, f : \Omega \to \mathbb{R}^m$. (f_n) *converges pointwise* to f, if for all $x \in \Omega$, $f_n(x) \to f(x)$.

(f_n) *converges uniformly* to f, if

$$\lim_{n\to\infty} \sup_{x\in\Omega} \|f_n(x) - f(x)\| = 0.$$

Example. Let $\Omega = [0, 1)$, $f_n(x) = x^n$, $f(x) = 0$. Then $f_n \to f$ pointwise, but not uniformly. □

6. Landau Symbols. Let sequences (a_n) and (b_n) be given. (a_n) is $O(b_n)$, if there exists a constant C s.t. for all n, $|a_n| \leq C|b_n|$. (a_n) is $o(b_n)$, if $\lim_{n\to\infty} a_n/b_n = 0$.

D. Complex Numbers.

Complex numbers z are typically represented as $z = x + iy$, with $x, y \in \mathbb{R}$. i is the imaginary unit $i = \sqrt{-1}$. x is called the real part and y the imaginary part of z. The absolute value is $|z| = \sqrt{x^2 + y^2}$. $\bar{z} = x - iy$ is the complex conjugate of $z = x + iy$.

Sometimes it is useful to write z in polar form as $z = re^{i\varphi}$. r is the absolute value of z, $r = |z|$, also called the modulus of z. φ is called the argument (also the phase or the angle). We have

$$e^{i\varphi} = \cos\varphi + i\sin\varphi.$$

$e^{i\varphi}$ may be represented as a point on the unit circle in $\mathbb{R}^2$.

Bibliography

[1] Jacod, J. and Protter P. (2004). Probability Essentials. 2nd edition. Springer. ISBN 3-540-43871-8.