Outline of the lecture course

Probability Theory II

Summer semester 2016

Preliminary version: April 21, 2016

If you find errors in the outline, please send a short note by email to johannes@math.uni-heidelberg.de

MATHEMATIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
Telefon: +49 6221 54.14.190 – Fax: +49 6221 54.53.31
eMail: johannes@math.uni-heidelberg.de
Webseite zur Vorlesung: www.razbaer.eu/ag-johannes/vl/WT2-SS16/
# Table of contents

1 Preliminaries  
1.1 Basic measure theory ................................................. 1  
1.2 Random variables ....................................................... 2  
1.3 Independence ............................................................. 4  
1.4 Expectation ............................................................... 5  
1.5 Convergence of random variables .................................. 6  
1.6 Conditional expectation .............................................. 6  

2 Stochastic processes  
2.1 Motivating examples .................................................. 9  
2.1.1 The Poisson process ............................................... 9  
2.1.2 Markov chains ...................................................... 10  
2.1.3 Brownian motion .................................................. 10  
2.2 Definition of stochastic processes ................................. 10  
2.3 Probability measures on Polish spaces ............................ 12  
2.4 Adapted stochastic process and stopping times .................. 14
Chapter 1

Preliminaries

This chapter presents elements of the lecture course PROBABILITY THEORY I along the lines of the textbook Klenke [2008], where far more details, examples and further discussions can be found.

1.1 Basic measure theory

In the following, let $\Omega \neq \emptyset$ be a nonempty set and let $\mathcal{A} \subset 2^\Omega$ (power set, set of all subsets of $\Omega$) be a class of subsets of $\Omega$. Later, $\Omega$ will be interpreted as the space of elementary events and $\mathcal{A}$ will be the system of observable events.

§1.1.1 Definition. (a) A pair $(\Omega, \mathcal{A})$ consisting of a nonempty set $\Omega$ and a $\sigma$-algebra $\mathcal{A}$ is called a measurable space. The sets $A \in \mathcal{A}$ are called measurable sets. If $\Omega$ is at most countably infinite and if $\mathcal{A} = 2^\Omega$, then the measurable space $(\Omega, 2^\Omega)$ is called discrete.

(b) A triple $(\Omega, \mathcal{A}, \mu)$ is called a measure space if $(\Omega, \mathcal{A})$ is a measurable space and if $\mu$ is a measure on $\mathcal{A}$.

(c) A measure space $(\Omega, \mathcal{A}, P)$ is called a probability space, if in addition $P(\Omega) = 1$. In this case, the sets $A \in \mathcal{A}$ are called events.

§1.1.2 Remark. Let $\mathcal{A} \subset 2^\Omega$ and let $\mu : \mathcal{A} \to [0, \infty]$ be a set function. We say that $\mu$ is

(a) monotone, if $\mu(A) \leq \mu(B)$ for any two sets $A, B \in \mathcal{A}$ with $A \subset B$.

(b) additive, if $\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i)$ for any choice of finitely many mutually disjoints sets $A_1, \ldots, A_n \in \mathcal{A}$ with $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$. The disjoint union of sets is denoted by the symbol $\bigcup$ which only stresses the fact that the sets involved are mutually disjoint.

(c) $\sigma$-additive, if $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ for any choice of countably many mutually disjoints sets $A_1, A_2, \ldots \in \mathcal{A}$ with $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

$\mathcal{A}$ is called an algebra if (i) $\Omega \in \mathcal{A}$, (ii) $\mathcal{A}$ is closed under complements, and (iii) $\mathcal{A}$ is closed under intersections. Note that, if $\mathcal{A}$ is closed under complements, then we have the equivalences between (i) $\mathcal{A}$ is closed under (countable) unions and (ii) $\mathcal{A}$ is closed under (countable) intersections. An algebra $\mathcal{A}$ is called $\sigma$-algebra, if it is closed under countable intersections. If $\mathcal{A}$ is an algebra and $\mu : \mathcal{A} \to [0, \infty]$ is a set function with $\mu(\emptyset) = 0$, then $\mu$ is called a

(d) content, if $\mu$ is additive,

(e) premeasure, if $\mu$ is $\sigma$-additive,

(f) measure, if $\mu$ is a premeasure and $\mathcal{A}$ is a $\sigma$-Algebra.
A content $\mu$ on an algebra $\mathcal{A}$ is called
\begin{itemize}
  \item[(g)] \textit{finite}, if $\mu(A) < \infty$ for every $A \in \mathcal{A}$,
  \item[(h)] \textit{$\sigma$-finite}, if there is a sequence $\Omega_1, \Omega_2, \ldots \in \mathcal{A}$ such that $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$ and such that $\mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$.
\end{itemize}

\section*{1.2 Random variables}

\subsection*{Example}

(a) For any nonempty set $\Omega$, the classes $\mathcal{A} = \{\emptyset, \Omega\}$ and $\mathcal{A} = 2^\Omega$ are the trivial examples of $\sigma$-algebras.

(b) Let $\mathcal{E} \subset 2^\Omega$. The smallest $\sigma$-algebra $\sigma(\mathcal{E}) = \bigcap\{\mathcal{A} : \mathcal{A} \text{ is } \sigma\text{-algebra and } \mathcal{E} \subset \mathcal{A}\}$ with $\mathcal{E} \subset \sigma(\mathcal{E})$ is called the $\sigma$-algebra \textit{generated by} $\mathcal{E}$ and $\mathcal{E}$ is called a \textit{generator} of $\sigma(\mathcal{E})$.

(c) Let $(\Omega, \tau)$ be a topological space with class of open sets $\tau \subset 2^\Omega$. The $\sigma$-algebra $\mathcal{B}(\Omega)$ that is generated by the open sets is called the \textit{Borel-$\sigma$-algebra} on $\Omega$. The elements $B \in \mathcal{B}(\Omega)$ are called \textit{Borel sets} or \textit{Borel measurable sets}. We write $\mathcal{B} := \mathcal{B}(\mathbb{R})$, $\mathcal{B}^+ := \mathcal{B}(\mathbb{R}^+)$ and $\mathcal{B}^n := \mathcal{B}(\mathbb{R}^n)$ for the Borel-$\sigma$-algebra on $\mathbb{R}$, $\mathbb{R}^+ := [0, \infty)$ and $\mathbb{R}^n$, respectively, equipped with the usual Euclidean distance.

(d) Denote by $1_A(x)$ the indicator function on a set $A$ which takes the value one if $x \in A$ and zero otherwise. Let $\omega \in \Omega$ and $\delta_\omega(A) = 1_A(\omega)$. Then $\delta_\omega$ is a probability measure on any $\sigma$-algebra $\mathcal{A} \subset 2^\Omega$. $\delta_\omega$ is called the \textit{Dirac measure} on the point $\omega$.

(e) Let $\Omega$ be an (at most) countable nonempty set and let $\mathcal{A} = 2^\Omega$. Further let $(p_\omega)_{\omega \in \Omega}$ be non-negative numbers. Then $A \mapsto \mu(A) := \sum_{\omega \in \Omega} p_\omega \delta_\omega(A)$ defines a $\sigma$-finite measure. If $p_\omega = 1$ for every $\omega \in \Omega$, then $\mu$ is called \textit{counting measure} on $\Omega$. If $\Omega$ is finite, then so is $\mu$.

\subsection*{Theorem (Carathéodory)}

Let $\mathcal{A} \subset 2^\Omega$ be an algebra and let $\mu$ be a $\sigma$-finite premeasure on $\mathcal{A}$. There exists a unique measure $\tilde{\mu}$ on $\sigma(\mathcal{A})$ such that $\tilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$. Furthermore, $\tilde{\mu}$ is $\sigma$-finite.

\textit{Proof of Theorem} \(1.1.4\). We refer to Klenke [2008], Theorem 1.41.

\subsection*{Remark}

If $\mu$ is a finite content on an algebra $\mathcal{A}$, then $\sigma$-\textit{continuity at }$\emptyset$, that is, $\mu(A_n) \to \mu(\emptyset)$ as $n \to \infty$ for any sequence $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{A}$ with $\mu(A_n) < \infty$ for some (and then eventually all) $n \in \mathbb{N}$ and $A_n \downarrow \emptyset$ (i.e., $\emptyset \subset A_1 \subset A_2 \subset A_3 \subset \ldots$ and $\cap_{n=1}^{\infty} A_n = \emptyset$), implies $\sigma$-additivity.

\subsection*{Example}

A probability measure $\mathbb{P}$ on the measurable space $(\mathbb{R}^n, \mathcal{B}^n)$ is uniquely determined by the values $\mathbb{P}((-\infty, b])$ (where $(-\infty, b] = \times_{i=1}^{n} (-\infty, b_i]$, $b \in \mathbb{R}^n$). In particular, a probability measure $\mathbb{P}$ on $\mathbb{R}$ is uniquely determined by its \textit{distribution function} $F : \mathbb{R} \to [0, 1]$, $x \mapsto \mathbb{P}((-\infty, x])$.

\section*{1.2 Random variables}

In this section $(\Omega, \mathcal{A})$, $(\mathcal{S}, \mathcal{I})$ and $(\mathcal{S}_i, \mathcal{I}_i)$, $i \in \mathcal{I}$, denote measurable spaces where $\mathcal{I}$ is an arbitrary index set.

\subsection*{Definition}

Let $\Omega$ be a nonempty set and let $X : \Omega \to \mathcal{S}$ be a map.

(a) $X$ is called $\mathcal{A}$-$\mathcal{I}$-measurable (or, briefly, measurable) if $X^{-1}(\mathcal{I}) := \{X^{-1}(S) : S \in \mathcal{I}\} \subset \mathcal{A}$, that is, if $X^{-1}(S) \in \mathcal{A}$ for any $S \in \mathcal{I}$. A measurable map $X : (\Omega, \mathcal{A}) \to \mathbb{R}$ is called $\mathcal{A}$-$\mathcal{B}$-measurable.
§1.2 Random variables

Chapter 1 Preliminaries

(S, ) is called a random variable (r.v.) with values in (S, ). If (S, ) = (R, B) or (S, ) = (R+, B+), then X is called a real or positive random variable, respectively.

(b) The preimage X−1( ) is the smallest σ-algebra on Ω with respect to which X is measurable. We say that σ(X) := X−1( ) is the σ-algebra on Ω that is generated by X.

(c) For any, i ∈ I, let Xi : Ω → Si be an arbitrary map. Then σ(Xi, i ∈ I) := σ(∪i∈I σ(Xi)) = σ(∪i∈I X−1(i)) is called the σ-algebra on Ω that is generated by (Xi, i ∈ I). This is the smallest σ-algebra with respect to which all Xi are measurable.

§1.2.2 Properties. Let I be an arbitrary index set. Consider Si ∈ 2S, i ∈ I, and a map X : Ω → S. Then

(a) X−1(∪i∈I Si) = ∪i∈IX−1(Si), X−1(∩i∈I Si) = ∩i∈IX−1(Si),

(b) X−1( ) is a σ-algebra on Ω and {S ∈ S : X−1(S) ∈ } is a σ-algebra on S.

If E is a class of sets in 2S, then σΩ(X−1(E)) = X−1(σS(E)). ♦

§1.2.3 Examples. (a) The identity map Id : Ω → Ω is - measurable.

(b) If = 2Ω and = {0, S}, then any map X : Ω → S is - measurable.

(c) Let A ⊂ Ω. The indicator function 1A : Ω → {0, 1} is -2(0,1)-measurable, if and only if A ∈ . ♦

For x, y ∈ R we agree on the following notations [x] := max{k ∈ Z : k ≤ x} (integer part), x ∨ y = max(x, y) (maximum), x ∧ y = min(x, y) (minimum), x+ = max(x, 0) (positive part), x− = max(−x, 0) (negative part) and |x| = x− + x+ (modulus).

§1.2.4 Properties. (a) If X, Y are real r.v.’s, then so are X+ := max(X, 0), X− := max(−X, 0), |X| = X+ + X−, X + Y, X − Y, X · Y and X/Y with x/0 := 0 for all x ∈ R. In particular, X+ and |X| is - measurable, respectively.

(b) If X1, X2, . . . are real r.v.’s, then so are supn≥1 Xn, infn≥1 Xn, limn→∞ Xn := infk≥1 supn≥k Xn and limn→∞ Xn := supk≥1 infn≥k Xn.

(c) Let X1, . . . , Xn : Ω → R be maps and define X := (X1, . . . , Xn) : Ω → Rn. Then X is a real r.v. (i.e., - measurable), if and only if each Xi is a real r.v. (i.e., - measurable).

(d) Let E = {Ai ∈ 2Ω, i ∈ I, mutually disjoint and ⋃i∈I Ai = Ω} be a partition of Ω. A map X : Ω → R is σ(E)- measurable, if there exist numbers xi ∈ R, i ∈ I, such that X = ⋃i∈I xi1Ai.

§1.2.5 Definition. (a) A real r.v. X is called simple if there is an n ∈ N and mutually disjoint measurable sets A1, . . . , An ∈ such as well as numbers α1, . . . , αn ∈ R, such that X = ⋃i=1 n αi1Ai.

(b) Assume that X, X1, X2, . . . are maps Ω → R := R ∪ {−∞, +∞} such that X1(ω) ≤ X2(ω) ≤ . . . and limn→∞ Xn(ω) = X(ω) for any ω ∈ Ω. Then we write Xn ↑ X and say that (Xn)n∈N increases (point-wise) to X. Analogously, we write Xn ↓ X if (−Xn)↑ (−X).
\section{Independence}

In the sequel, \((\Omega, \mathcal{A}, \mathbb{P})\) is a probability space, the sets \(A \in \mathcal{A}\) are the events and \(\mathcal{I}\) is an arbitrary index set.

\subsection{Definition}
(a) Let \((A_i)_{i \in \mathcal{I}}\) be an arbitrary family of events. The family \((A_i)_{i \in \mathcal{I}}\) is called independent if for any finite subset \(\mathcal{J} \subset \mathcal{I}\) the product formula holds: \(\mathbb{P}(\bigcap_{j \in \mathcal{J}} A_j) = \prod_{j \in \mathcal{J}} \mathbb{P}(A_j)\).

(b) Let \(E_i \subset \mathcal{A}\) for all \(i \in \mathcal{I}\). The family \((E_i)_{i \in \mathcal{I}}\) is called independent if, for any finite subset \(\mathcal{J} \subset \mathcal{I}\) and any choice of \(E_j \in E_j, j \in \mathcal{J}\), the product formula holds: \(\mathbb{P}(\bigcap_{j \in \mathcal{J}} E_j) = \prod_{j \in \mathcal{J}} \mathbb{P}(E_j)\).

\subsection{Lemma (Borel-Cantelli)}
\begin{itemize}
  \item \(A_1, A_2, \ldots \) be events and define \(A^* := \limsup_{n \to \infty} A_n\).
  \begin{itemize}
    \item If \(\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty\), then \(\mathbb{P}(A^*) = 0\).
    \item If \((A_n)_{n \in \mathbb{N}}\) is independent and \(\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty\), then \(\mathbb{P}(A^*) = 1\).
  \end{itemize}
\end{itemize}

\textbf{Proof of Lemma 1.3.2.} We refer to Klenke [2008], Theorem 2.7.

\subsection{Corollary (Borel’s 0-1 criterion)}
\begin{itemize}
  \item \(A_1, A_2, \ldots \) be independent events and define \(A^* := \limsup_{n \to \infty} A_n\).
  \begin{itemize}
    \item \(\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty\) if and only if \(\mathbb{P}(A^*) = 0\),
    \item \(\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty\) if and only if \(\mathbb{P}(A^*) = 1\).
  \end{itemize}
\end{itemize}

For each \(i \in \mathcal{I}\), let \((S_i, \mathcal{S}_i)\) be a measurable space and let \(X_i : (\Omega, \mathcal{A}) \to (S_i, \mathcal{S}_i)\) be a r.v. with generated \(\sigma\)-algebra \(\sigma(X_i) = X^{-1}(\mathcal{S}_i)\).

\subsection{Definition}
(a) The family \((X_i)_{i \in \mathcal{I}}\) of r.v.’s is called independent if the family \((\sigma(X_i))_{i \in \mathcal{I}}\) of \(\sigma\)-algebras is independent.
(b) Let $\mathcal{E}_i \subset \mathcal{A}$ for all $i \in I$. The family $(\mathcal{E}_i)_{i \in I}$ is called independent if, for any finite subset $J \subset I$ and any choice of $E_j \in \mathcal{E}_j$, $j \in J$, the product formula holds: $\mathbb{P}(\bigcap_{j \in J} E_j) = \prod_{j \in J} \mathbb{P}(E_j)$.

§1.3.5 **Property.** Let $\mathcal{K}$ be an arbitrary set and $\mathcal{I}_k$, $k \in \mathcal{K}$, arbitrary mutually disjoint index sets. Define $I = \bigcup_{k \in \mathcal{K}} \mathcal{I}_k$. If the family $(X_i)_{i \in I}$ of r.v.’s is independent, then the family of $\sigma$-algebras $(\sigma(X_j, j \in \mathcal{I}_k))_{k \in \mathcal{K}}$ is independent. □

§1.3.6 **Definition.** Let $X_1, X_2, \ldots$ be r.v.’s. The $\sigma$-algebra $\bigcap_{n \geq 1} \sigma(X_i, i \geq n)$ is called the tail $\sigma$-algebra and its elements are called tail events.

§1.3.7 **Example.** $\{\omega : \sum_{n \geq 1} X_n(\omega) \text{ is convergent}\}$ is an tail event.

§1.3.8 **Theorem** (Kolmogorov’s 0-1 Law). The tail events of a sequence $(X_n)_{n \in \mathbb{N}}$ of independent r.v.’s have probability 0 or 1.

*Proof of Theorem* §1.3.8. We refer to Klenke [2008], Theorem 2.37. □

### 1.4 Expectation

§1.4.1 **Definition.** We denote by $\mathcal{M} := \mathcal{M}(\Omega, \mathcal{A})$ the set of all real r.v.’s defined on the measurable space $(\Omega, \mathcal{A})$. By $\mathcal{M}^+ := \mathcal{M}^+(\Omega, \mathcal{A}) \subset \mathcal{M}$ the subset of all positive r.v.’s. Given a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ the expectation is the unique functional $E : \mathcal{M}^+ \to [0, \infty]$ satisfying

(a) $E(aX_1 + X_2) = aE(X_1) + E(X_2)$ for all $X_1, X_2 \in \mathcal{M}^+$ and $a \in \mathbb{R}^+$;

(b) Assume $X, X_1, X_2, \ldots \in \mathcal{M}^+$ such that $X_n \uparrow X$ then $EX_n \uparrow EX$;

(c) $E1_A = \mathbb{P}(A)$ for all $A \in \mathcal{A}$.

The expectation of $X \in \mathcal{M}$ is defined by $E(X) := E(X^+) - E(X^-)$, if $E(X^+) < \infty$ or $E(X^-) < \infty$. Given $||X||_p := (E(|X|^p)^{1/p}$, $p \in [1, \infty)$, and $||X||_\infty := \inf\{c : \mathbb{P}(X > c) = 0\}$ for $p \in [1, \infty]$ set $L_p(\Omega, \mathcal{A}, P) := \{X \in \mathcal{M}(\Omega, \mathcal{A}) : ||X||_p < \infty\}$ and $L_\infty := L_\infty(\Omega, \mathcal{A}, P) := \{[X] : X \in L_\infty(\Omega, \mathcal{A}, P)\}$ where $[X] := \{Y \in \mathcal{M}(\Omega, \mathcal{A}) : \mathbb{P}(X = Y) = 1\}$. □

§1.4.2 **Remark.** $L_1$ is the domain of definition of the expectation $E$, that is, $E : L_1 \to \mathbb{R}$. The vector space $L_p$, equipped with the norm $||.||_p$ is a Banach space and in case $p = 2$ it is a Hilbert space with norm $||.||_2$ induced by the inner product $\langle X, Y \rangle_2 := E(XY)$. □

§1.4.3 **Properties.**

(a) For r.v.’s $X, Y \in L_1$ we have the equivalences between (i) $E(1_A X) \leq E(1_A Y)$ for all $A \in \mathcal{A}$ and (ii) $\mathbb{P}(X \leq Y) = 1$. In particular, $E(1_A X) = E(1_A Y)$ holds for all $A \in \mathcal{A}$ if and only if $\mathbb{P}(X = Y) = 1$.

(b) (Fatou’s lemma) Assume $X_1, X_2, \ldots \in \mathcal{M}^+$, then $\mathbb{E}(\lim \inf_{n \to \infty} X_n) \leq \lim \inf_{n \to \infty} E(X_n)$.

(c) (Dominated convergence) Assume $X, X_1, X_2, \ldots \in \mathcal{M}$ such that $\lim_{n \to \infty} |X_n(\omega) - X(\omega)| = 0$ for all $\omega \in \Omega$. If there exists $Y \in L_1$ with $\sup_{n \geq 1} |X_n| \leq Y$, then we have $\lim_{n \to \infty} E|X_n - X| = 0$ which in turn implies $X \in L_1$ and $\lim_{n \to \infty} E|X_n - EX| = 0$.

(d) (Hölder’s inequality) For $X, Y \in \mathcal{M}$ holds $E|XY| \leq ||X||_p ||Y||_q$ with $p^{-1} + q^{-1} = 1$.

(e) (Cauchy-Schwarz inequality) For $X, Y \in \mathcal{M}$ holds $E|XY| \leq \sqrt{E(X^2)} \sqrt{E(Y^2)}$ and $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}$. □
1.5 Convergence of random variables

In the sequel we assume r.v.’s \( X_1, X_2, \ldots \in \mathcal{M}(\Omega, \mathcal{A}) \) and a probability measure \( P \) on \((\Omega, \mathcal{A})\).

§1.5.1 Definition. (a) Let \( C := \{ \omega \in \Omega : \lim_{n \to \infty} X_n(\omega) \text{ exists and is finite} \} \). The sequence \((X_n)_{n \geq 1} \) converges almost surely (a.s.), if \( P(C) = 1 \). We write \( X_n \xrightarrow{a.s.} X \).

(b) The sequence \((X_n)_{n \geq 1} \) converges in probability, if \( \lim_{n \to \infty} P(|X_n - X| \geq \varepsilon) = 0 \) for all \( \varepsilon > 0 \). We write \( X_n \xrightarrow{p} X \) or briefly, \( X_n \xrightarrow{P} X \).

(c) The sequence \((X_n)_{n \in \mathbb{N}} \) converges in distribution, if \( \mathbb{E}(f(X_n)) \xrightarrow{n \to \infty} \mathbb{E}(f(X)) \) for any continuous and bounded function \( f : \mathbb{R} \to \mathbb{R} \). We write \( X_n \xrightarrow{d} X \) in distribution, or briefly, \( X_n \xrightarrow{d} X \).

(d) The sequence \((X_n)_{n \in \mathbb{N}} \) converges in \( L_p \), if \( \lim_{n \to \infty} \mathbb{E}|X_n - X|^p = 0 \). We write \( X_n \xrightarrow{L_p} X \) in \( L_p \), or briefly, \( X_n \xrightarrow{L_p} X \).

§1.5.2 Remark. In (a) the set \( C = \bigcap_{k \geq 1} \bigcup_{n \geq 1} \bigcup_{i \geq 1} \{ |X_{n+i}(\omega) - X_n(\omega)| < 1/k \} \) is measurable. Moreover, if \( P(C) = 1 \) then there exists a r.v. \( X \in \mathcal{M} \) such that \( P(\lim_{n \to \infty} X_n = X) = 1 \) where \( X = \lim \sup_{n \to \infty} X_n \) noting that \( X(\omega) = \lim \sup_{n \to \infty} X_n(\omega) \) for \( \omega \in C \).

§1.5.3 Properties. (a) We have \( X_n \xrightarrow{a.s.} X \ if \ and \ only \ if \ \sup_{m \geq n} |X_n - X_m| \xrightarrow{n \to \infty} 0 \ in \ \mathbb{P} if \ and \ only \ if \ \sup_{j \geq n} |X_j - X| \xrightarrow{n \to \infty} 0 \ in \ \mathbb{P} \) if and only if for all \( \varepsilon, \delta > 0, \exists \mathbb{N}(\varepsilon, \delta) \in \mathbb{N}, \forall n \geq N(\varepsilon, \delta), P(\bigcap_{j \geq n} \{ |X_j - X| \leq \varepsilon \}) \geq 1 - \delta \).

(b) If \( X_n \xrightarrow{a.s.} X \), then \( X_n \xrightarrow{P} X \).

(c) If \( X_n \xrightarrow{a.s.} X \), then \( g(X_n) \xrightarrow{a.s.} g(X) \) for any continuous function \( g \).

(d) \( X_n \xrightarrow{P} X \ if \ and \ only \ if \ \lim_{n \to \infty} \sup_{j \geq n} \mathbb{P}(|X_j - X_n| \geq \varepsilon) = 0 \ for \ all \ \varepsilon > 0 \ if \ and \ only \ if \ any \ sub-sequence \ of \ (X_n)_{n \in \mathbb{N}} \ contains \ a \ sub-sequence \ converging \ to \ X \ a.s.. \)

(e) If \( X_n \xrightarrow{P} X \), then \( g(X_n) \xrightarrow{P} g(X) \) for any continuous function \( g \).

(f) \( X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X \Leftarrow X_n \xrightarrow{L_p} X \) and \( X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X \)

1.6 Conditional expectation

In the sequel \((\Omega, \mathcal{A}, \mathbb{P})\) is a probability space and \( \mathcal{F} \) is a sub-\(\sigma\)-algebra of \( \mathcal{A} \).

§1.6.1 Theorem. If \( X \in \mathcal{M}^+(\Omega, \mathcal{A}) \) or \( X \in L_1(\Omega, \mathcal{A}, \mathbb{P}) \) then there exists \( Y \in \mathcal{M}^+(\Omega, \mathcal{F}) \) or \( Y \in L_1(\Omega, \mathcal{F}, \mathbb{P}) \), respectively, such that \( \mathbb{E}(X 1_F) = \mathbb{E}(Y 1_F) \) for all \( F \in \mathcal{F} \), moreover \( Y \) is unique up to equality a.s..

Proof of Theorem §1.6.1. We refer to Klenke [2008], Theorem 8.12.

§1.6.2 Definition. For \( X \in \mathcal{M}^+(\Omega, \mathcal{A}) \) or \( X \in L_1(\Omega, \mathcal{A}, \mathbb{P}) \) each version \( Y \) as in Theorem §1.6.1 is called conditional expectation (bedingte Erwartung) of \( X \) given \( \mathcal{F} \), symbolically \( \mathbb{E}(X|\mathcal{F}) := Y \). For \( A \in \mathcal{A} \), \( \mathbb{P}(A|\mathcal{F}) := \mathbb{E}(1_A|\mathcal{F}) \) is called a conditional probability of \( A \) given the \( \sigma \)-algebra \( \mathcal{F} \). Given r.v.’s \( X_i, i \in I \), we set \( \mathbb{E}(X|(X_i)_{i \in I}) := \mathbb{E}(X|\sigma(X_i, i \in I)) \).
§1.6.3 Remark. Employing Proposition §1.2.7 there exists a $\mathcal{B}-\mathcal{B}$-measurable function $f$ such that $E(Y|X) = f(X)$ a.s.. Therewith, we write $E(Y|X = x) := f(x)$ (conditional expected value, bedingter Erwartungswert). Since conditional expectations are defined only up to equality a.s., all (in)equalities with conditional expectations are understood as (in)equalities a.s., even if we do not say so explicitly.

§1.6.4 Properties. Let $\mathcal{G} \subset \mathcal{F} \subset \mathcal{A}$ be $\sigma$-algebras and let $X, Y \in L_1(\Omega, \mathcal{A}, \mathbb{P})$. Then:

(a) (Linearity) $E(\lambda X + Y|\mathcal{F}) = \lambda E(X|\mathcal{F}) + E(Y|\mathcal{F})$.

(b) (Monotonicity) If $X \geq Y$ a.s., then $E(X|\mathcal{F}) \geq E(Y|\mathcal{F})$.

(c) If $E(|XY|) < \infty$ and $Y$ is measurable with respect to $\mathcal{F}$, then $E(XY|\mathcal{F}) = YE(X|\mathcal{F})$ and $E(Y|\mathcal{F}) = E(Y|Y) = Y$.

(d) (Tower property) $E(E(X|\mathcal{F})|\mathcal{G}) = E(E(X|\mathcal{G})|\mathcal{F}) = E(X|\mathcal{F})$.

(e) (Triangle inequality) $E(|X| | \mathcal{F}) \geq |E(X|\mathcal{F})|$.

(f) (Independence) If $\sigma(X)$ and $\mathcal{F}$ are independent, then $E(X|\mathcal{F}) = E(X)$.

(g) If $\mathbb{P}(A) \in \{0, 1\}$ for any $A \in \mathcal{F}$, then $E(X|\mathcal{F}) = E(X)$.

(h) (Jensen’s inequality) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and let $\varphi(Y)$ be an element of $L_1(\Omega, \mathcal{A}, \mathbb{P})$. Then $\varphi(E(Y|\mathcal{F})) \leq E(\varphi(Y)|\mathcal{F})$.

(i) (Dominated convergence) Assume $Y \in L_1(\mathbb{P})$, $Y \geq 0$ and $(X_n)_{n \in \mathbb{N}}$ is a sequence of r.v.’s with $|X_n| \leq Y$ for $n \in \mathbb{N}$ and such that $X_n \xrightarrow{a.s.} X$. Then $\lim_{n \to \infty} E(X_n|\mathcal{F}) = E(X|\mathcal{F})$ a.s. and in $L_1(\mathbb{P})$.

§1.6.5 Proposition. Let $(\mathbb{H}, \langle \cdot, \cdot \rangle_\mathbb{H})$ be a Hilbert space equipped with induced norm $\|\cdot\|_\mathbb{H}$ and let $\mathcal{U}$ be a closed linear subspace of $\mathbb{H}$. For each $x \in \mathbb{H}$ there exists a unique element $u_x \in \mathcal{U}$ with $\|x - u_x\|_\mathbb{H} = \inf_{u \in \mathcal{U}} \|x - u\|_\mathbb{H}$.

§1.6.6 Definition. For a closed subspace $\mathcal{U}$ of the Hilbert space $(\mathbb{H}, \langle \cdot, \cdot \rangle_\mathbb{H})$ the orthogonal projection $\Pi_\mathcal{U} : \mathbb{H} \to \mathcal{U}$ is defined by $\Pi_\mathcal{U}(x) = u_x$ with $u_x$ as in Proposition §1.6.5.

§1.6.7 Properties. Let $\mathcal{U}^\perp$ be the orthogonal complement of $\mathcal{U}$ in $\mathbb{H}$. Then:

(a) (projection property) $\Pi_\mathcal{U} \circ \Pi_{\mathcal{U}^\perp} = \Pi_{\mathcal{U}^\perp}$;

(b) (orthogonality) $x - \Pi_{\mathcal{U}^\perp} x \in \mathcal{U}^\perp$ for each $x \in \mathbb{H}$;

(c) each $x \in \mathbb{H}$ can be decomposed uniquely as $x = \Pi_{\mathcal{U}} x + (x - \Pi_{\mathcal{U}} x)$ in the orthogonal sum of an element of $\mathcal{U}$ and an element of $\mathcal{U}^\perp$;

(d) $\Pi_{\mathcal{U}}$ is selfadjoint: $\langle \Pi_{\mathcal{U}} x, y \rangle_\mathbb{H} = \langle x, \Pi_{\mathcal{U}} y \rangle_\mathbb{H}$;

(e) $\Pi_{\mathcal{U}}$ is linear.

§1.6.8 Lemma. Let $\mathcal{F}$ be a sub-$\sigma$-algebra of $\mathcal{A}$. Then $L_2(\Omega, \mathcal{F}, \mathbb{P})$ is embedded as closed linear subspace in the Hilbert space $L_2(\Omega, \mathcal{A}, \mathbb{P})$.

§1.6.9 Corollary. Let $\mathcal{F} \subset \mathcal{A}$ be a sub-$\sigma$-algebra and let $X \in L_2(\Omega, \mathcal{A}, \mathbb{P})$ be a r.v.. Then $E(X|\mathcal{F})$ is the orthogonal projection of $X$ on $L_2(\Omega, \mathcal{F}, \mathbb{P})$. That is, for any $Y \in L_2(\Omega, \mathcal{F}, \mathbb{P})$, $\|X - Y\|_2^2 = E[(X - Y)^2] \geq E[(X - E(X|\mathcal{F}))^2] = \|X - E(X|\mathcal{F})\|_2^2$ with equality if and only if $Y = E(X|\mathcal{F})$. 

Probability theory II
§1.6.10 Example. Let $X, Y \in L^1(\mathbb{P})$ be independent. Then $\mathbb{E}(X + Y | Y) = \mathbb{E}(X | Y) + \mathbb{E}(Y | Y) = \mathbb{E}(X) + Y$.

§1.6.11 Theorem. Let $p \in [1, \infty]$ and $\mathcal{F} \subset \mathcal{A}$ be a sub-$\sigma$-algebra. Then the linear map $L^p(\Omega, \mathcal{A}, \mathbb{P}) \to L^p(\Omega, \mathcal{F}, \mathbb{P})$, $X \mapsto \mathbb{E}(X | \mathcal{F})$, is a contraction (that is, $\|\mathbb{E}(X | \mathcal{F})\|_p \leq \|X\|_p$) and thus bounded and continuous. Hence, for $X, X_1, X_2, \ldots \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ with $\|X_n - X\|_p \xrightarrow{n \to \infty} 0$ we have $\|\mathbb{E}(X_n | \mathcal{F}) - \mathbb{E}(X | \mathcal{F})\|_p \xrightarrow{n \to \infty} 0$.

§1.6.12 Definition. A family $(X_i)_{i \in I}$ of r.v.’s in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ with arbitrary index set $I$ is called uniformly integrable if $\inf_{a \in [0, \infty)} \sup_{i \in I} \mathbb{E}(1_{\{|X_i| > a\}} | X_i) = 0$ which is satisfied in case that $\sup_{i \in I} |X_i| \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

§1.6.13 Corollary. Let $(X_i)_{i \in I}$ be uniformly integrable in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ and let $(\mathcal{F}_j, j \in J)$ be a family of sub-$\sigma$-algebras of $\mathcal{A}$. Define $X_{i,j} := \mathbb{E}(X_i | \mathcal{F}_j)$. Then $(X_{i,j})_{i \in I, j \in J}$ is uniformly integrable in $L^1(\Omega, \mathcal{A}, \mathbb{P})$. In particular, for $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ the family $\{\mathbb{E}(X | \mathcal{F}) : \mathcal{F} \text{ is sub-$\sigma$-algebra of $\mathcal{A}$}\}$ of r.v.’s in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ is uniformly integrable.
Chapter 2

Stochastic processes

2.1 Motivating examples

2.1.1 The Poisson process

§2.1.1 Definition. Let \((S_k)_{k \in \mathbb{N}}\) be positive r.v.’s on a probability space \((\Omega, \mathcal{A}, \mathbb{P})\) with \(0 \leq S_1(\omega) \leq S_2(\omega) \leq \ldots\) for any \(\omega \in \Omega\). The family \(N = (N_t)_{t \geq 0}\) of \(\mathbb{N}_0\)-valued r.v.’s given by \(N_t := \sum_{k=1}^{\infty} \mathbbm{1}_{(S_k \leq t)}\), \(t \geq 0\), is called counting process (Zählprozess) with jump times (Sprungzeiten) \((S_k)_{k \in \mathbb{N}}\).

§2.1.2 Definition. A counting process \((N_t)_{t \geq 0}\) is called Poisson process of intensity \(\lambda > 0\) if

(a) \(\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)\) as \(h \downarrow 0\);

(b) \(\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h)\) as \(h \downarrow 0\);

(c) (independent increments) \((N_{t_1} - N_{t_{i-1}})_{i=1}^{n}\) are independent for any numbers \(0 = t_0 < t_1 < \ldots < t_n\) in \(\mathbb{R}^+\);

(d) (stationary increments) \(N_t - N_s \overset{d}{=} N_{t-s}\) for all numbers \(t \geq s \geq 0\) in \(\mathbb{R}^+\).

§2.1.3 Theorem. For a counting process \(N = (N_t)_{t \geq 0}\) with jump times \((S_k)_{k \in \mathbb{N}}\) we have the equivalences between:

(a) \(N\) is a Poisson process;

(b) \(N\) satisfies the conditions (iii), (iv) in the Definition §2.1.2 of a Poisson (\(\mathcal{P}\)oi) process and \(N_t \sim \mathcal{P}oi(\lambda t)\) holds for all \(t > 0\);

(c) (waiting times) The r.v.’s \(T_1 := S_1\) and \(T_k := S_k - S_{k-1}, k = 2, 3, \ldots\), are independent and identically \(\mathcal{E}xp(\lambda)\)-distributed;

(d) \(N_t \sim \mathcal{P}oi(\lambda t)\) holds for all \(t > 0\) and the conditional distribution of \((S_1, \ldots, S_n)\) given \(\{N_t = n\}\) has the density

\[
f(x_1, \ldots, x_n) = \frac{n!}{t^n} \mathbb{1}_{\{0 \leq x_1 \leq \ldots \leq x_n \leq t\}}.
\]

(2.1)

(e) \(N\) satisfies the condition (c) in the Definition §2.1.2 of a Poisson process, \(\mathbb{E}(N_1) = \lambda\) and (2.1) is the conditional density of \((S_1, \ldots, S_n)\) given \(\{N_t = n\}\).

Proof of Theorem §2.1.3 is given in the lecture course.

§2.1.4 Remark. Let \((U_i)_{i=1}^{n}\) be independent and identically \(\mathcal{U}([0, t])\)-distributed r.v.’s and let \((U_{(i)})_{i=1}^{n}\) be their order statistics where \(U_{(1)} = \min\{U_i\}_{i=1}^{n}\) and \(U_{(k+1)} = \min\{U_i\}_{i=1}^{n} \setminus \{U_{(i)}\}_{i=1}^{k}\), \(k = 2, \ldots, n\). Then the joint density of \((U_{(i)})_{i=1}^{n}\) is given exactly by (2.1). The characterisations give rise to three simple methods to simulate a Poisson process: the definition §2.1.2 gives an approximation for small \(h\) (forgetting the \(o(h)\)-term), part (c) in §2.1.3 just uses exponentially
distributed inter-arrival times $T_k$ and part (d) uses the value at a specified right-end point and then uses the uniform order statistics as jump times in-between (write down the details!).

2.1.2 Markov chains

§2.1.5 Definition. Let $T = \mathbb{N}_0$ (discrete time) or $T = [0, \infty)$ (continuous time), let $S$ be a (at most) countable nonempty set (state space) and let $\mathcal{F} = 2^S$. A family $(X_t)_{t \in T}$ of $S$-valued r.v.’s forms a Markov chain if for all $n \in \mathbb{N}$, all $t_1 < t_2 < \ldots < t_n < t$ in $T$ and all $s_1, \ldots, s_n, s$ in $S$ with $\mathbb{P}(X_{t_1} = s_1, \ldots, X_{t_n} = s_n) > 0$ the Markov property is satisfied: $\mathbb{P}(X_t = s|X_{t_1} = s_1, \ldots, X_{t_n} = s_n) = \mathbb{P}(X_t = s|X_{t_n} = s_n)$. For a Markov chain $(X_t)_{t \in T}$ and $t_1 \leq t_2$ in $T$, $i, j \in S$ the transition probability to reach state $j$ at time $t_2$ from state $i$ at time $t_1$ is defined by $p_{ij}(t_1, t_2) := \mathbb{P}(X_{t_2} = j|X_{t_1} = i)$ (or arbitrary if not well-defined). The transition matrix is given by $P(t_1, t_2) := (p_{ij}(t_1, t_2))_{i,j \in S}$. The transition matrix and the Markov chain are called time-homogeneous if $P(t_1, t_2) = P(0, t_2 - t_1) =: P(t_2 - t_1)$ holds for all $t_1 \leq t_2$.

§2.1.6 Proposition. The transition matrices satisfy the Chapman-Kolmogorov equation, that is, for any $t_1 \leq t_2 \leq t_3$ in $T$, $P(t_1, t_3) = P(t_1, t_2)P(t_2, t_3)$ (matrix multiplication). In the time-homogeneous case this gives the semigroup property $P(t_1 + t_2) = P(t_1)P(t_2)$ for all $t_1, t_2 \in T$, and in particular $P(n) = P(1)^n$ for $n \in \mathbb{N}$.

Proof of Proposition §2.1.6 is given in the lecture course.

2.1.3 Brownian motion

§2.1.7 Definition. A family $(W_t)_{t \geq 0}$ of real r.v.’s is called a Brownian motion if

(a) $W_0 = 0$ a.s.;
(b) (independent increments) $(W_t - W_{t-1})_{t=1}^n$ are independent for any numbers $0 = t_0 < t_1 < \ldots < t_n$ in $\mathbb{R}^+$;
(c) (stationary increments) $W_t - W_s \overset{d}{=} W_{t-s} \sim \mathcal{N}(0, t-s)$ for all numbers $0 \leq s < t$ in $\mathbb{R}^+$;
(d) $t \mapsto W_t$ is continuous a.s..

§2.1.8 Remark. Questions:

(i) Existence?
(ii) $W := (W_t)_{t \geq 0}$ r.v. on which space?
(iii) For which functions $f$ is $f(W)$ a r.v.? (e.g. $f(W) = \sup_{0 \leq t \leq 1} W_t$)

Importance of the Brownian motion:

- If $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}(X_i) = 0$ and $\mathbb{V}(X_i) = \sigma^2 < \infty$ then $W$ is a “limit” of $S_n^t = \frac{1}{\sigma \sqrt{t}} \sum_{1 \leq i \leq n} X_i$ (Donsker’s theorem).
- $W$ is a central element in stochastic differential equations $X_t = \int_0^t \sigma(X_s) dW_s + \int_0^t b(X_s) ds$. How to define the first integral? (“Ito integral”)

2.2 Definition of stochastic processes

§2.2.1 Definition. A family $X = (X_t)_{t \in T}$ of r.v.’s on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called stochastic process. We call $X$ time-discrete if $T \subset \mathbb{Z}$ and time-continuous if $(a, b) \subset \mathbb{R}$.
the probability measure on $\sigma$ forces us to specify a $X$.

Remark (on §2.2.9 Definition) finite distributions of the r.v.’s does not exists an explicit formula. Therefore, we are interested in a characterisation exploiting Remark §2.2.7 $X$. For this reason, we identify $X = (X_t)_{t \in \mathbb{T}}$ as a r.v. with values in $S^T$ which forces us to specify a $\sigma$-algebra on $S^T$.

Definition. Let $(S_i, \mathcal{A}_i), i \in I$, be an arbitrary family of measurable spaces.

(a) The set $X_{i \in I} S_i$ of maps $(s_i)_{i \in I} : I \rightarrow \bigcup_{i \in I} S_i$ such that $s_i \in S_i$ for all $i \in I$ is called product space. For $J \subset I$, let $S_J := \bigcap_{j \in J} S_j$. If, in particular, all the $S_j$ are equal, say $S_i = S$, then we write $X_{i \in I} S_i = S^I$.

(b) If $j \in I$, then $\Pi_j : S_I \rightarrow S_J, (s_i)_{i \in I} \mapsto s_j$ denotes the jth coordinate map. More generally, for $J \subset K \subset I$, the restricted map $\Pi^K_J : S_K \rightarrow S_J, (s_k)_{k \in K} \mapsto (s_j)_{j \in J}$ are called canonical projection. In particular, we write $\Pi_J := \Pi^K_J$.

(c) The product-$\sigma$-algebra $\mathcal{A}_J := \bigotimes_{i \in I} \mathcal{A}_i$ is the smallest $\sigma$-algebra on the product space $S_I$ such that for every $j \in I$ the coordinate map $\Pi_j : S_I \rightarrow S_J$ is measurable with respect to $\mathcal{A}_I$, that is, $\mathcal{A}_J = \bigotimes_{i \in I} \mathcal{A}_i = \sigma(\Pi_i, i \in I) := \bigvee_{i \in I} \Pi_i^{-1}(\mathcal{A}_i)$. For $J \subset I$, let $\mathcal{A}_J = \bigotimes_{j \in J} \mathcal{A}_j$. If $(S_i, \mathcal{A}_i) = (S, \mathcal{A})$ for all $i \in I$, then we also write $\bigotimes_{i \in I} \mathcal{A}_i = \mathcal{A}^\otimes$.

Lemma. For a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ with state space $(S, \mathcal{A})$ the mapping $X : \Omega \rightarrow S^T, \omega \mapsto (X_t(\omega))_{t \in \mathbb{T}}$ is a $(S^T, \mathcal{A}^\otimes)$-valued r.v.

Proof of Lemma §2.2.4 is given in the lecture course.

Remark. Later on, we shall also consider smaller function spaces than $S^T$, e.g. $C(\mathbb{R}^+)$ instead of $\mathbb{R}^\mathbb{R}^+$.

Definition. The distribution $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ of a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(S^T, \mathcal{A}^\otimes)$ is the image probability measure of $\mathbb{P}$ under the map $X$.

Remark. The distribution of a stochastic process is often complicate and generally there does not exists an explicit formula. Therefore, we are interested in a characterisation exploiting the distributions of the r.v.’s $X_t$.

Definition. Let $X = (X_t)_{t \in \mathbb{T}}$ be a stochastic process with distribution $\mathbb{P}_X$. For any finite $\mathcal{T} \subset \mathbb{T}$ let $\mathbb{P}_{X, T} := \mathbb{P}_{\Pi_T \circ X}$ be the distribution of the r.v. $(X_t)_{t \in \mathbb{T}} = \Pi_T \circ X$. The family $\{\mathbb{P}_{X, T}, \mathcal{T} \subset \mathbb{T} \text{ finite}\}$ is called family of the finite-dimensional distributions of $X$ or $\mathbb{P}_X$.

Definition. A family $\{\mathbb{P}_J, J \subset I \text{ finite}\}$ of probability measures is called consistent on $(S_T, \mathcal{A}_T)$ if for any finite $J \subset K \subset I$ the canonical projection $\Pi^K_J$ as in §2.2.3 (c) and the probability measure $\mathbb{P}_J$ and $\mathbb{P}_K$ on $(S_J, \mathcal{A}_J)$ and $(S_K, \mathcal{A}_K)$, respectively, satisfy $\mathbb{P}_J = \mathbb{P}_K \circ (\Pi^K_J)^{-1}$.

Remark. Let $\mathbb{P}_X$ be the distribution of a stochastic process $X$ on $(S^T, \mathcal{A}^\otimes)$ then its family $\{\mathbb{P}_{X, T}, \mathcal{T} \subset \mathbb{T} \text{ finite}\}$ of finite-dimensional distributions is consistent. Indeed, for $J \subset \mathbb{T}$
\(\mathcal{K} \subset \mathcal{T}\) finite, \(P_X^T = P_X \circ \Pi_J^{-1} = P_X \circ (\Pi_J^k \circ \Pi_K)^{-1} = P_X \circ (\Pi_K)^{-1} \circ (\Pi_J^k)^{-1} = P_X^K \circ (\Pi_J^k)^{-1}\).

\section*{Chapter 2 Stochastic processes}

\subsection*{2.3 Probability measures on Polish spaces}

\section*{2.3 Probability measures on Polish spaces}

\section*{Definition 2.3 Probability measures on Polish spaces}

Two processes \((X_t)_{t \in T}, (Y_t)_{t \in T}\) on \((\Omega, \mathcal{A}, P)\) are called

(a) \textit{indistinguishable} (ununterscheidbar) if \(P(\forall t \in T : X_t = Y_t) = 1\);

(b) \textit{versions or modifications} (Versionen, Modifikationen) of each other, if \(P(X_t = Y_t) = 1\) for all \(t \in T\).

\section*{Remark 2.3.2}

(a) Obviously, indistinguishable processes are versions of each other. The converse is in general false.

(b) If \(X\) is a version of \(Y\), then \(X\) and \(Y\) share the same finite-dimensional distributions. Processes with the same finite-dimensional distributions need not even be defined on the same probability space and will in general not be versions of each other.

(c) Suppose \((X_t)_{t \in \mathbb{R}^+}\) and \((Y_t)_{t \in \mathbb{R}^+}\) are real-valued stochastic processes with right-continuous sample paths. Then they are indistinguishable already if they are versions of each other.

\section*{Definition 2.3.1}

A stochastic processes \((X_t)_{t \in \mathbb{R}^+}\) is called \textit{continuous} if all sample paths are continuous. It is called \textit{stochastically continuous}, if \(t_n \xrightarrow{n \to \infty} t\) always implies \(X_{t_n} \xrightarrow{P} X_t\) (convergence in probability).

\section*{Remark 2.3.12}

Every continuous stochastic process is stochastically continuous since a.s. convergence implies convergence in probability. On the other hand, the Poisson process is obviously not continuous but stochastically continuous, since \(\lim_{t_n \to t} P(|N_t - N_{t_n}| > \varepsilon) = \lim_{t_n \to t} (1 - e^{-\lambda(t-t_n)}) = 0\) for all \(\varepsilon \in (0, 1)\).

\section*{2.3 Probability measures on Polish spaces}

\section*{Definition 2.3.1}

A metric space \((S, d)\) is called \textit{Polish space} if it is \textit{separable} and \textit{complete}. More generally, a separable completely metrisable topological space is called \textit{Polish}. Canonically, it is equipped with its Borel-\(\sigma\)-algebra \(\mathcal{B}(S)\) generated by the open sets.

\section*{Remark 2.3.2}

Let \((\Omega, \tau)\) be a topological space. For \(A \subset \Omega\) we denote by \(\overline{A}\) the closure of \(A\), by \(A^o\) the interior and by \(\partial A\) the boundary of \(A\). A set \(A \subset \Omega\) is called \textit{dense} if \(\overline{A} = \Omega\). A set \(A \subset \Omega\) is called \textit{compact} if each open cover \(U \subset \tau\) of \(A\) (that is, \(A \subset \cup\{U; U \in \mathcal{U}\}\)) has a finite subcover; that is, a finite \(U' \subset \mathcal{U}\) with \(A \subset \cup\{U; U \in U'\}\). Compact sets are closed. A set \(A \subset \Omega\) is called \textit{relatively compact} if \(\overline{A}\) is compact. On the other hand, \(A\) is called \textit{sequentially compact} (respectively \textit{relatively sequentially compact}) if any sequence \((\omega_n)_{n \in \mathbb{N}}\) with values in \(A\) has a subsequence \((\omega_{n_k})_{k \in \mathbb{N}}\) that converges to some \(\omega \in A\) (respectively \(\omega \in \overline{A}\)).

\(\Omega, \tau\) is called \textit{metrisable} if there exists a metric \(d\) on \(\Omega\) such that \(\tau\) is induced by the open balls \(B_\varepsilon(x) = \{\omega \in \Omega : d(x, \omega) < \varepsilon\}\). In metrisable spaces, the notions compact and sequentially compact coincide. A metric \(d\) on \(\Omega\) is called \textit{complete} if any Cauchy sequence with respect to \(d\) converges in \(\Omega\). \(\Omega, \tau\) is called \textit{completely metrisable} if there exists a complete metric on \(\Omega\) that induces \(\tau\). A metrisable space \((\Omega, \tau)\) is called \textit{separable} if there exists a countable dense subset of \(\Omega\). Separability in metrisable spaces is equivalent to the existence of a countable basis of the topology; that is, a countable set \(\mathcal{U} \subset \tau\) with \(A = \bigcup\{U; U \subset A, U \in \mathcal{U}\}\) for all \(A \in \tau\).
Two measurable spaces \((\Omega_1, \mathcal{B}_1), (\Omega_2, \mathcal{B}_2)\) with Borel-\(\sigma\)-algebra \(\mathcal{B}_1, \mathcal{B}_2\), respectively, are called \textit{Borel-isomorphic}, if there exists a bijective map \(g : \Omega_1 \to \Omega_2\), such that \(g\) and \(g^{-1}\) are measurable. In particular, each Polish space is Borel-isomorphic to a Borel subset of \([0, 1]\).

Two topological spaces \((\Omega_1, \tau_1), (\Omega_2, \tau_2)\) are called \textit{homeomorphic} if there exists a bijective map \(g : \Omega_1 \to \Omega_2\) such that \(g\) and \(g^{-1}\) are continuous. Therewith, each Polish space is homeomorphic to a subset of \([0, 1]^\mathbb{N}\), equipped with its product topology.

\section*{§2.3.3 Examples.} \(\mathbb{R}, \mathbb{R}^n, \ell_p \subset \mathbb{R}^n\) and \(L_p([0, 1])\) equipped with their usual distance are Polish spaces.

\section*{§2.3.4 Definition.} Let \((S_i, d_i), i \in \mathcal{I} \subset \mathbb{N}\), be a finite or countable family of metric spaces. The \textit{product space} \(X_{i \in \mathcal{I}} S_i\) is canonically equipped with the \textit{product metric} \(d((s_i)_{i \in \mathcal{I}}, (s'_i)_{i \in \mathcal{I}}) := \sum_{i \in \mathcal{I}} 2^{-i}(d_i(s_i, s'_i) \wedge 1)\) generating the product topology on \(X_{i \in \mathcal{I}} S_i\), in which a vector/sequence converges if and only if all coordinates converge, that is, \(d(s^{(n)}, s)^{n \to \infty} 0 \iff d_i(s_i^{(n)}, s_i)^{n \to \infty} 0\) for all \(i \in \mathcal{I}\).

\section*{§2.3.5 Lemma.} Let \((S_n, d_n), n \in \mathbb{N}\), be a family of Polish spaces, then the Borel-\(\sigma\)-Algebra \(\mathcal{B}(\bigotimes_{n \in \mathbb{N}} S_n)\) on the product space \(\bigotimes_{n \in \mathbb{N}} S_n\) equals the product Borel-\(\sigma\)-algebra \(\prod_{n \in \mathbb{N}} \mathcal{B}(S_n)\).

\section*{Proof of Lemma §2.3.5} is given in the lecture course.

\section*{§2.3.6 Remark.} The \(\supseteq\)-relation holds for all topological spaces and products of any cardinality with the same proof. The \(\subseteq\)-property can already fail for the product of two topological (non-Polish) spaces.

\section*{§2.3.7 Definition.} Let \((S, d)\) be a metric space equipped with its Borel-\(\sigma\)-algebra \(\mathcal{B}(S)\). A probability measure \(\mathbb{P}\) on \((S, \mathcal{B}(S))\) is called

- \(\texttt{tight} \) (straff) if for all \(\varepsilon > 0\) there is a compact set \(K\) such that \(\mathbb{P}(K) \geq 1 - \varepsilon\),
- \(\texttt{regular} \) (regular) if \(B \in \mathcal{B}(S)\) and \(\varepsilon > 0\) then there exist a compact set \(K\) and an open set \(O\) such that \(K \subset B \subset O\) and \(\mathbb{P}(O \setminus K) \leq \varepsilon\).

A family \(\mathcal{P}\) of probability measures on \((S, \mathcal{B}(S))\) is called (uniformly) tight, if for all \(\varepsilon > 0\) there is a compact set \(K\) such that \(\mathbb{P}(K) \geq 1 - \varepsilon\) for all \(\mathbb{P} \in \mathcal{P}\).

\section*{§2.3.8 Remark.} Considering a probability measure \(\mathbb{P}\) on a metric space \(S\) we have the equivalences between

\begin{itemize}
\item \(\text{i}\) \(\mathbb{P}\) is tight and \(\text{ii}\) \(\mathbb{P}(B) = \sup\{\mathbb{P}(K) : K \subseteq B \text{ compact}\}\) for all \(B \in \mathcal{B}(S)\),
\item \(\text{i}'\) \(\mathbb{P}\) is regular and \(\text{ii}'\) \(\sup\{\mathbb{P}(K) : K \subseteq B \text{ compact}\} = \mathbb{P}(B) = \inf\{\mathbb{P}(O) : O \supseteq B \text{ open}\}\) for all \(B \in \mathcal{B}(S)\).
\end{itemize}

\section*{§2.3.9 Proposition (Ulam (1939)).} Every probability measure on a Polish space is tight.

\section*{Proof of Proposition §2.3.9} is given in the lecture course.

\section*{§2.3.10 Theorem.} Every probability measure on a Polish space is regular.

\section*{Proof of Theorem §2.3.10} is given in the lecture course.

\section*{§2.3.11 Theorem (Kolmogorov’s consistency theorem).} Let \(\mathcal{I}\) be an arbitrary index set and let \((S_i, \mathcal{B}_i), i \in \mathcal{I}\) be Polish spaces, \(i \in \mathcal{I}\). Let \(\{\mathbb{P}_j, J \subset \mathcal{I} \text{ finite}\}\) be a consistent family of probability measures on the product space \((S_\mathcal{I}, \mathcal{B}_\mathcal{I})\) as in §2.2.9. Then there exists a unique probability measure \(\mathbb{P}\) on \((S_\mathcal{I}, \mathcal{B}_\mathcal{I})\) having \(\{\mathbb{P}_j, J \subset \mathcal{I} \text{ finite}\}\) as family of finite dimensional distributions, that is, \(\mathbb{P}_j = \mathbb{P} \circ \Pi_j^{-1}\) for any \(J \subset \mathcal{I} \text{ finite}\).
2.4 Adapted stochastic process and stopping times

In the sequel, the index set \( T \) is a subset of \( \mathbb{R} \), \( X = (X_t)_{t \in T} \) is a stochastic process on a probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) with state space \( (\mathcal{S}, \mathcal{B}) \) and image probability measure \( \mathbb{P}_X \) on \( (\mathcal{S}^T, \mathcal{B}^\otimes T) \).

2.4.1 Definition. A family \( \mathcal{F} = (\mathcal{F}_t)_{t \in T} \) of \( \sigma \)-algebras with \( \mathcal{F}_t \subset \mathcal{A} \), \( t \in T \), is called a filtration if \( \mathcal{F}_s \subset \mathcal{F}_t \) for all \( s, t \in T \) with \( s \leq t \). \( (\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}) \) is called filtered probability space.

2.4.2 Definition. A stochastic process \( X = (X_t)_{t \in T} \) is called adapted to the filtration \( \mathcal{F} = (\mathcal{F}_t)_{t \in T} \) if \( X_t \in \mathcal{F}_t \)-measurable for all \( t \in T \). If \( \mathcal{F}_t = \sigma(X_s, s \leq t) \) for all \( t \in T \), then we denote by \( \mathcal{F}^X = \sigma(X) \) the natural filtration generated by \( X \).

2.4.3 Remark. Clearly, a stochastic process is always adapted to the natural filtration it generates. The natural filtration is the smallest filtration to which the process is adapted.

2.4.4 Definition. A stochastic process \( X = (X_n)_{n \in \mathbb{N}_0} \) is called predictable (or previsible) with respect to a filtration \( \mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{N}_0} \) if \( X_0 \) is constant and if, for every \( n \in \mathbb{N} \), \( X_n \) is \( \mathcal{F}_{n-1} \)-measurable.

2.4.5 Definition. A r.v. \( \tau \) with values in \( T \cup \{ \infty \} \) is called a stopping time (with respect to the filtration \( \mathcal{F} \)) if for any \( t \in T \), \( \{ \tau \leq t \} \in \mathcal{F}_t \).

2.4.6 Proposition. Let \( T \) be countable, \( \tau \) is a stopping time if and only if \( \{ \tau = t \} \in \mathcal{F}_t \) for all \( t \in T \).

Proof of Proposition §2.4.6 is left as an exercise.

2.4.7 Examples. (a) Let \( t_0 \in T \), then \( \tau \equiv t_0 \) (constant) is a stopping time where \( \sigma(\tau) = \{ \emptyset, \Omega \} \).
2.4 Adapted stochastic process and stopping times

Chapter 2 Stochastic processes

(b) Let \( X = (X_n)_{n \in \mathbb{N}_0} \) be a stochastic process adapted to a filtration \( \mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0} \). For \( S \in \mathcal{F} \) we call **waiting time** the first time that \( X \) is in \( S \), that is,

\[
\tau_S(\omega) := \begin{cases} 
\inf \{ n \in \mathbb{N}_0 : X_n(\omega) \in S \}, & \text{if } \omega \in \bigcup_{n \in \mathbb{N}_0} X_n^{-1}(S), \\
\infty, & \text{otherwise}
\end{cases}
\]

Then \( \tau_S \) is a stopping time with respect to \( \mathcal{F} \). Note that \( \tau_0 \equiv \infty \) and \( \tau_S \equiv 0 \). \( \square \)

§2.4.8 **Lemma.** Let \( \tau \) and \( \sigma \) be stopping times. Then

(a) \( \tau \vee \sigma \) and \( \tau \wedge \sigma \) are stopping times.

(b) If \( \tau, \sigma \geq 0 \), then \( \tau + \sigma \) is also a stopping time.

(c) If \( s \in \mathbb{R}^+ \), then \( \tau + s \) is a stopping time. However, in general, \( \tau - s \) is not.

**Proof of Lemma §2.4.8** is left as an exercise. \( \square \)

§2.4.9 **Remark.** We note that (a) and (c) are properties we would expect of stopping times. With (a), the interpretation is clear. For (c), note that \( \tau - s \) peeks into the future by \( s \) time units (in fact, \( \{ \tau - s \leq t \} \in \mathcal{F}_{t+s} \)), while \( \tau + s \) looks back \( s \) time units. For stopping times, however, only retrospection is allowed. \( \square \)

§2.4.10 **Example.** Let \( X = (X_n)_{n \in \mathbb{N}_0} \) be a stochastic process adapted to a filtration \( (\mathcal{F}_n)_{n \in \mathbb{N}_0} \). For \( S_1, S_2 \in \mathcal{F} \) let \( \tau_{S_1} \) and \( \tau_{S_2} \) be waiting times as in §2.4.7 (b), then (i) \( \tau_{S_1 \cup S_2} = \tau_{S_1} \wedge \tau_{S_2} \), (ii) \( \tau_{S_1 \cap S_2} = \tau_{S_1} \vee \tau_{S_2} \) and (iii) if \( S_1 \subset S_2 \), then \( \tau_{S_1} \geq \tau_{S_2} \). \( \square \)

§2.4.11 **Definition.** Let \( \tau \) be a stopping time. Then

\[
\mathcal{F}_\tau := \{ A \in \mathcal{F} : A \cap \{ \tau \leq t \} \in \mathcal{F}_t \text{ for any } t \in \mathbb{T} \}
\]

is called the **\( \sigma \)-algebra of \( \tau \)-past.** \( \square \)

§2.4.12 **Example.** If \( \tau \equiv t_0 \) is a constant stopping time at \( t_0 \in \mathbb{T} \), then \( \mathcal{F}_\tau = \mathcal{F}_{t_0} \). \( \square \)

§2.4.13 **Lemma.** If \( \tau \) and \( \sigma \) are stopping times then (i) \( \mathcal{F}_{\tau \wedge \sigma} = \mathcal{F}_\tau \cap \mathcal{F}_\sigma \), (ii) \( \mathcal{F}_{\tau \vee \sigma} = \mathcal{F}_\tau \vee \mathcal{F}_\sigma \) and (iii) if \( \tau \leq \sigma \), then \( \mathcal{F}_\tau \subset \mathcal{F}_\sigma \).

**Proof of Lemma §2.4.13** is given in the lecture course. \( \square \)

§2.4.14 **Definition.** If \( \tau < \infty \) is a stopping time, then we define \( X_\tau(\omega) := X_\tau(\omega)(\omega) \). \( \square \)

§2.4.15 **Lemma.** Let \( \mathbb{T} \) be countable, let \( X \) be adapted and let \( \tau < \infty \) be a stopping time. Then \( X_\tau(\omega) \) is measurable with respect to \( \mathcal{F}_\tau \). In particular, \( \tau \) is \( \mathcal{F}_\tau \)-measurable.

**Proof of Lemma §2.4.15** is given in the lecture course. \( \square \)

§2.4.16 **Remark.** For uncountable \( \mathbb{T} \) and for fixed \( \omega \), in general, the map \( \mathbb{T} \rightarrow S, t \mapsto X_t(\omega) \) is not measurable; hence neither is the composition \( X_\tau \) always measurable. Here one needs assumptions on the regularity of the paths \( t \mapsto X_t(\omega) \); for example, right continuity. \( \square \)

§2.4.17 **Corollary.** Let \( \mathbb{T} \) be countable, let \( X \) be adapted and let \( (\tau_s)_{s \in \mathbb{T}} \) be a family of stopping times with \( \tau_s \leq \tau_t < \infty, s, t \in \mathbb{T}, s \leq t \). Then the process \( (X_n)_{n \in \mathbb{T}} \) is adapted to the filtration \( (\mathcal{F}_{\tau_s})_{s \in \mathbb{T}} \). In particular, \( (X_{\tau_x} : x \in \mathbb{T}) \) is adapted to both filtration \( (\mathcal{F}_{\tau_{\tau_x}})_{x \in \mathbb{T}} \) and \( (\mathcal{F}_t)_{t \in \mathbb{T}} \).

**Proof of Corollary §2.4.17** is given in the lecture course. \( \square \)
§2.4.18 **Definition.** Let $\mathbb{T}$ be countable, let $(X_t)_{t \in \mathbb{T}}$ be adapted and let $\tau$ be a stopping time. We define the *stopped process* $X^\tau = (X^\tau_t)_{t \in \mathbb{T}}$ by $X^\tau_t = X^\tau_{t \wedge \tau}$ for any $t \in \mathbb{T}$ which is adapted to both filtration $\mathcal{F}^\tau = (\mathcal{F}^\tau_t)_{t \in \mathbb{T}} = (\mathcal{F}^\tau_{t \wedge \tau})_{t \in \mathbb{T}}$ and $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$.  

\qed
Bibliography
