Stochastic Processes (Stochastik II)
Outline of the course in winter semester 2009/10

Markus Reiß
Humboldt-Universität zu Berlin

Preliminary version, 9. Februar 2010

Inhaltsverzeichnis

1 Some important processes

1.1 The Poisson process

1.1 Definition. Let $(S_k)_{k\geqslant 1}$ be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ with $0 \leqslant S_1(\omega) \leqslant S_2(\omega) \leqslant \cdots$ for all $k \geqslant 1$, $\omega \in \Omega$. Then $N = (N_t, t \geqslant 0)$ with

$$N_t := \sum_{k\geqslant 1} \mathbf{1}_{\{S_k \leqslant t\}}, \quad t \geqslant 0,$$

is called counting process (Zählprozess) with jump times (Sprungzeiten) (S_k).

1.2 Definition. A counting process N is called Poisson process of intensity $\lambda > 0$ if

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

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

(c) (independent increments) $(N_{t_i} - N_{t_{i-1}})_{1\leqslant i\leqslant n}$ are independent for $0 = t_0 < t_1 < \cdots < t_n$;

(d) (stationary increments) $N_t - N_s \stackrel{d}{=} N_{t-s}$ for all $t \geqslant s \geqslant 0$.

1.3 Theorem. *For a counting process N with jump times (S_k) the following are equivalent:*

(a) N is a Poisson process;

(b) N satisfies conditions (c),(d) of a Poisson process and $N_t \sim \mathrm{Poiss}(\lambda t)$ holds for all $t > 0$;

(c) $T_1 := S_1$, $T_k := S_k - S_{k-1}$, $k \geqslant 2$, are i.i.d. $\mathrm{Exp}(\lambda)$-distributed random variables;

(d) $N_t \sim \mathrm{Poiss}(\lambda t)$ holds for all $t > 0$ and the law of $(S_1, \ldots, S_n)$ given $\{N_t = n\}$ has the density

$$f(x_1, \ldots, x_n) = \tfrac{n!}{t^n} \mathbf{1}_{\{0\leqslant x_1 \leqslant \cdots \leqslant x_n \leqslant t\}}. \tag{1.1}$$

(e) N satisfies condition (c) of a Poisson process, $\mathbb{E}[N_1] = \lambda$ and (1.1) *is the density of $(S_1, \ldots, S_n)$ given $\{N_t = n\}$.*

1.2 Markov chains

1.4 Definition. Let $T = \mathbb{N}_0$ (discrete time) or $T = [0, \infty)$ (continuous time) and S be a countable set (state space). Then random variables $(X_t)_{t\in T}$ with values in $(S, \mathcal{P}(S))$ form a Markov chain if for all $n \in \mathbb{N}$, $t_1 < t_2 < \cdots < t_{n+1}$, $s_1, \ldots, s_{n+1} \in S$ with $\mathbb{P}(X_{t_1} = s_1, \ldots, X_{t_n} = s_n) > 0$ the Markov property is satisfied:

$$\mathbb{P}(X_{t_{n+1}} = s_{n+1} \,|\, X_{t_1} = s_1, \ldots, X_{t_n} = s_n) = \mathbb{P}(X_{t_{n+1}} = s_{n+1} \,|\, X_{t_n} = s_n).$$

1.5 Definition. For a Markov chain X and $t_1 \leqslant t_2$, $i, j \in S$

$$p_{ij}(t_1, t_2) := \mathbb{P}(X_{t_2} = j \,|\, X_{t_1} = i) \text{ (or arbitrary if not well-defined)}$$

defines the transition probability to reach state j at time t_2 from state i at time t_1. 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 $t_1 \leqslant t_2$.

1.6 Proposition. *The transition matrices satisfy the Chapman-Kolmogorov equation*

$$\forall t_1 \leqslant t_2 \leqslant t_3 : P(t_1, t_3) = P(t_1, t_2) P(t_2, t_3) \text{ (matrix multiplikation).}$$

In the time-homogeneous case this gives the semigroup property

$$\forall t, s \in T : P(t + s) = P(t) P(s),$$

in particular $P(n) = P(1)^n$ *for* $n \in \mathbb{N}$.

2 General theory of stochastic processes

2.1 Basic notions

2.1 Definition. A family $X = (X_t,\, t \in T)$ of random variables on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called stochastic process. We call X time-discrete if $T = \mathbb{N}_0$ and time-continuous if $T = \mathbb{R}_0^+ = [0, \infty)$. If all X_t take values in $(S, \mathscr{S})$, then $(S, \mathscr{S})$ is the state space (Zustandsraum) of X. For each fixed $\omega \in \Omega$ the mapping $t \mapsto X_t(\omega)$ is called sample path (Pfad), trajectory (Trajektorie) or Realisation (Realisierung) of X.

2.2 Lemma. *For a stochastic process* $(X_t,\, t \in T)$ *with state space* $(S, \mathscr{S})$ *the mapping* $\bar{X} : \Omega \to S^T$ *with* $\bar{X}(\omega)(t) := X_t(\omega)$ *is a* $(S^T, \mathscr{S}^{\otimes T})$*-valued random variable.*

2.3 Definition. Given a stochastic process $(X_t,\, t \in T)$, the laws of the random vectors $(X_{t_1}, \ldots, X_{t_n})$ with $n \geqslant 1$, $t_1, \ldots, t_n \in T$ are called finite-dimensional distributions of X. We write $P_{t_1, \ldots, t_n} := \mathbb{P}^{(X_{t_1}, \ldots, X_{t_n})}$.

2.4 Lemma. *Let* $(X_t,\, t \in T)$ *be a stochastic process with state space* $(S, \mathscr{S})$ *and denote by* $\pi_{J,I} : S^J \to S^I$ *for* $I \subseteq J$ *the coordinate projection. Then the finite-dimensional distributions satisfy the following consistency condition:*

$$\forall\, I \subseteq J \subseteq T \text{ with } I, J \text{ finite } \forall A \in \mathscr{S}^{\otimes I} :\; P_J(\pi_{J,I}^{-1}(A)) = P_I(A). \tag{2.1}$$

2.5 Definition. Two processes $(X_t,\, t \in T)$, $(Y_t,\, t \in T)$ on $(\Omega, \mathscr{F}, \mathbb{P})$ are called

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

(b) versions or modifications (Versionen, Modifikationen) of each other if we have $\forall\, t \in T:\ \mathbb{P}(X_t = Y_t) = 1$.

2.6 Definition. A process $(X_t, t \geqslant 0)$ is called continuous if all sample paths are continuous. It is called stochastically continuous, if $t_n \to t$ always implies $X_{t_n} \xrightarrow{\mathbb{P}} X_t$ (convergence in probability).

2.2 Polish spaces and Kolmogorov's consistency theorem

2.7 Definition. A metric space (S, d) is called Polish space if it is separable and complete. More generally, a separable topological space which is metrizable with a complete metric is called Polish. Canonically, it is equipped with its Borel σ-algebra $\mathfrak{B}_S$, generated by the open sets.

2.8 Lemma. *Let $S_1, \ldots, S_n$ be Polish spaces, then the Borel σ-algebra of the product satisfies $\mathfrak{B}_{S_1 \times \cdots \times S_n} = \mathfrak{B}_{S_1} \otimes \cdots \otimes \mathfrak{B}_{S_n}$.*

2.9 Definition. A probability measure $\mathbb{P}$ on a metric space $(S, \mathfrak{B}_S)$ is called

(a) tight (straff) if $\forall \varepsilon > 0\, \exists K \subseteq S$ compact : $P(K) \geqslant 1 - \varepsilon$,

(b) regular (regulär) if $\forall \varepsilon > 0$, $B \in \mathfrak{B}_S\, \exists K \subseteq B$ compact : $P(B \setminus K) \leqslant \varepsilon$ and $\forall \varepsilon > 0$, $B \in \mathfrak{B}_S\, \exists O \supseteq B$ open : $P(O \setminus B) \leqslant \varepsilon$.

2.10 Proposition. *Every probability measure on a Polish space is tight.*

2.11 Theorem (Ulam, 1939)**.** *Every probability measure on a Polish space is regular.*

2.12 Definition. Let $I \neq \varnothing$ be an index set and $(S, \mathscr{S})$ be a measurable set. Let for each finite subset $J \subseteq I$ a probability measure $\mathbb{P}_J$ on the product space $(S^J, \mathscr{S}^{\otimes J})$ be given. Then $(\mathbb{P}_J)_{J \subseteq I \text{ finite}}$ is called projective family if the following consistency condition is satisfied:

$$\forall J \subseteq J' \subseteq I \text{ finite}, B \in \mathscr{S}^{\otimes J}:\ \mathbb{P}_J(B) = \mathbb{P}_{J'}(\pi_{J',J}^{-1}(B)),$$

where $\pi_{J',J} : S^{J'} \to S^J$ denotes the coordinate projection.

2.13 Theorem (Kolmogorov's consistency theorem)**.** *Let $(S, \mathfrak{B}_S)$ be a Polish space, I an index set and let $(\mathbb{P}_J)$ be a projective family for S and I. Then there exists a unique probability measure $\mathbb{P}$ on the product space $(S^I, \mathscr{S}^{\otimes I})$ satisfying*

$$\forall J \subseteq I \text{ finite}, B \in \mathscr{S}^{\otimes J}:\ \mathbb{P}_J(B) = \mathbb{P}(\pi_{I,J}^{-1}(B)).$$

2.14 Corollary. *For any Polish state space $(S, \mathfrak{B}_S)$ and index set $T \neq \varnothing$ there exists to a prescribed projective family $(\mathbb{P}_J)$ a stochastic process $(X_t, t \in T)$ whose finite-dimensional distributions are given by $(\mathbb{P}_J)$.*

2.15 Corollary. *For any family $(\mathbb{P}_i)_{i \in I}$ of probability measures on $(S, \mathscr{S})$ there exists the product measure $\bigotimes_{i \in I} \mathbb{P}_i$ on $(S^I, \mathscr{S}^{\otimes I})$. In particular, a family $(X_i)_{i \in I}$ of independent random variables with prescribed laws $\mathbb{P}^{X_i}$ exists. [Proof only for S Polish]*

3 The conditional expectation

3.1 Orthogonal projections

3.1 Proposition. *Let L be a closed linear subspace of the Hilbert space H. Then for each $x \in H$ there is a unique $y_x \in L$ with $\|x - y_x\| = \operatorname{dist}_L(x) := \inf_{y \in L}\|x - y\|$.*

3.2 Definition. For a closed linear subspace L of the Hilbert space H the orthogonal projection $P_L : H \to L$ onto L is defined by $P_L(x) = y_x$ with y_x from the previous proposition.

3.3 Lemma. *We have:*

(a) $P_L \circ P_L = P_L$ *(projection property);*

(b) $\forall\, x \in H : \ (x - P_L x) \in L^\perp$ *(orthogonality).*

3.4 Corollary. *We have:*

(a) Each $x \in H$ can be decomposed uniquely as $x = P_L x + (x - P_L x)$ in the sum of an element of L and an element of $L^\perp$;

(b) P_L is selfadjoint: $\langle P_L x, y\rangle = \langle x, P_L y\rangle$;

(c) P_L is linear.

3.2 Construction and properties

3.5 Definition. For a random variable X on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in $(S, \mathscr{S})$ we introduce the σ-algebra (!) $\sigma(X) := \{X^{-1}(A) \,|\, A \in \mathscr{S}\} \subseteq \mathscr{F}$. For a given probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we set

$$\begin{aligned}
\mathcal{M} := \mathcal{M}(\Omega, \mathscr{F}) &:= \{X : \Omega \to \mathbb{R} \text{ measurable}\};\\
\mathcal{M}^+ := \mathcal{M}^+(\Omega, \mathscr{F}) &:= \{X : \Omega \to [0, \infty] \text{ measurable}\};\\
\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P}) &:= \{X \in \mathcal{M}(\Omega, \mathscr{F}) \,|\, \mathbb{E}[|X|^p] < \infty\};\\
L^p := L^p(\Omega, \mathscr{F}, \mathbb{P}) &:= \{[X] \,|\, X \in \mathcal{L}^p(\Omega, \mathscr{F}, \mathbb{P})\}\\
&\qquad \text{where } [X] := \{Y \in \mathcal{M}(\Omega, \mathscr{F}) \,|\, \mathbb{P}(X = Y) = 1\}.
\end{aligned}$$

3.6 Proposition. *Let X be a $(S, \mathscr{S})$-valued and Y a real-valued random variable. Then Y is $\sigma(X)$-measurable if and only if there is a $(\mathscr{S}, \mathfrak{B}_\mathbb{R})$-measurable function $\varphi : S \to \mathbb{R}$ such that $Y = \varphi(X)$.*

3.7 Lemma. *Let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$. Then $L^2(\Omega, \mathscr{G}, \mathbb{P})$ is embedded as closed linear subspace in the Hilbert space $L^2(\Omega, \mathscr{F}, \mathbb{P})$.*

3.8 Definition. Let X be a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$. Then for $Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ the conditional expectation (bedingte Erwartung) of Y given X is defined as the $L^2(\Omega, \mathscr{F}, \mathbb{P})$-orthogonal projection of Y onto $L^2(\Omega, \sigma(X), \mathbb{P})$: $\mathbb{E}[Y \,|\, X] := P_{L^2(\Omega, \sigma(X), \mathbb{P})} Y$. If φ is the measurable function such that $\mathbb{E}[Y \,|\, X] = \varphi(X)$ a.s., we write $\mathbb{E}[Y \,|\, X = x] := \varphi(x)$ (conditional expected value, bedingter Erwartungswert).

More generally, for a sub-σ-algebra $\mathscr{G}$ the conditional expectation of $Y \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ given $\mathscr{G}$ is defined as $\mathbb{E}[Y \,|\, \mathscr{G}] = P_{L^2(\Omega, \mathscr{G}, \mathbb{P})} Y$.

3.9 Lemma. *$\mathbb{E}[Y\,|\,\mathscr{G}]$ is as element of L^2 uniquely determined by the following properties:*

(a) *$\mathbb{E}[Y\,|\,\mathscr{G}]$ is $\mathscr{G}$-measurable (modulo null sets);*

(b) $\forall\, G \in \mathscr{G}:\ \mathbb{E}[\mathbb{E}[Y\,|\,\mathscr{G}]\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G]$.

3.10 Theorem. *Let $Y \in \mathcal{M}^+(\Omega, \mathscr{F})$ or $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$. Then there is a $\mathbb{P}$-a.s. unique element $\mathbb{E}[Y\,|\,\mathscr{G}]$ in $\mathcal{M}^+(\Omega, \mathscr{G})$ and $L^1(\Omega, \mathscr{G}, \mathbb{P})$, respectively, such that*

$$\forall\, G \in \mathscr{G}:\ \mathbb{E}[\mathbb{E}[Y\,|\,\mathscr{G}]\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G].$$

3.11 Definition. For $Y \in \mathcal{M}^+(\Omega, \mathscr{F})$ or $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and a sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$ the <u>general conditional expectation</u> of Y given $\mathscr{G}$ is defined as $\mathbb{E}[Y\,|\,\mathscr{G}]$ from the preceding theorem. We put $\mathbb{E}[Y\,|\,(X_i)_{i\in I}] := \mathbb{E}[Y\,|\,\sigma(X_i,\ i \in I)]$ for random variables X_i, $i \in I$.

3.12 Proposition. *Let $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$. Then:*

(a) $\mathbb{E}[\mathbb{E}[Y\,|\,\mathscr{G}]] = \mathbb{E}[Y]$;

(b) *Y $\mathscr{G}$-measurable $\Rightarrow \mathbb{E}[Y\,|\,\mathscr{G}] = Y$ a.s.;*

(c) *$\alpha \in \mathbb{R}$, $Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$: $\mathbb{E}[\alpha Y + Z\,|\,\mathscr{G}] = \alpha\,\mathbb{E}[Y\,|\,\mathscr{G}] + \mathbb{E}[Z\,|\,\mathscr{G}]$ a.s.;*

(d) *$Y \geqslant 0$ a.s. $\Rightarrow \mathbb{E}[Y\,|\,\mathscr{G}] \geqslant 0$ a.s.;*

(e) *$Y_n \in \mathcal{M}^+(\Omega, \mathscr{F})$, $Y_n \uparrow Y$ a.s. $\Rightarrow \mathbb{E}[Y_n\,|\,\mathscr{G}] \uparrow \mathbb{E}[Y\,|\,\mathscr{G}]$ a.s. (monotone convergence);*

(f) *$Y_n \in \mathcal{M}^+(\Omega, \mathscr{F}) \Rightarrow \mathbb{E}[\liminf_n Y_n\,|\,\mathscr{G}] \leqslant \liminf_n \mathbb{E}[Y_n\,|\,\mathscr{G}]$ a.s. (Fatou's Lemma);*

(g) *$Y_n \in \mathcal{M}(\Omega, \mathscr{F})$, $Y_n \to Y$, $|Y_n| \leqslant Z$ with $Z \in L^1(\Omega, \mathscr{F}, \mathbb{P})$: $\mathbb{E}[Y_n\,|\,\mathscr{G}] \to \mathbb{E}[Y\,|\,\mathscr{G}]$ a.s. (dominated convergence);*

(h) *$\mathscr{H} \subseteq \mathscr{G} \Rightarrow \mathbb{E}[\mathbb{E}[Y\,|\,\mathscr{G}]\,|\,\mathscr{H}] = \mathbb{E}[Y\,|\,\mathscr{H}]$ a.s. (projection/tower property);*

(i) *Z $\mathscr{G}$-measurable, $ZY \in L^1$: $\mathbb{E}[ZY\,|\,\mathscr{G}] = Z\,\mathbb{E}[Y\,|\,\mathscr{G}]$ a.s.;*

(j) *Y independent of $\mathscr{G}$: $\mathbb{E}[Y\,|\,\mathscr{G}] = \mathbb{E}[Y]$ a.s.*

3.13 Proposition (Jensen's Inequality)**.** *If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and Y, $\varphi(Y)$ are in L^1, then $\varphi(\mathbb{E}[Y\,|\,\mathscr{G}]) \leqslant \mathbb{E}[\varphi(Y)\,|\,\mathscr{G}]$ holds for any sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$.*

4 Martingale theory

4.1 Martingales, sub- and supermartingales

4.1 Definition. A sequence $(\mathscr{F}_n)_{n\geqslant 0}$ of sub-σ-algebras of $\mathscr{F}$ is called filtration if $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$, $n \geqslant 0$, holds. $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_n))$ is called filtered probability space.

4.2 Definition. A sequence $(M_n)_{n\geqslant 0}$ of random variables on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_n))$ forms a martingale (submartingale, supermartingale) if:

(a) $M_n \in L^1$, $n \geqslant 0$;

(b) M_n is $\mathscr{F}_n$-measurable, $n \geqslant 0$ (adapted);

(c) $\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] = M_n$ (resp. $\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] \geqslant M_n$ for submartingale, resp. $\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] \leqslant M_n$ for supermartingale).

If $\mathscr{F}_n = \sigma(M_0, \dots, M_n)$ holds, then $(\mathscr{F}_n)$ is the natural filtration of M, notation $(\mathscr{F}_n^M)$.

4.3 Definition. A martingale (M_n) is closable (abschließbar), if there exists an $X \in L^1$ with $M_n = \mathbb{E}[X \mid \mathscr{F}_n]$, $n \geqslant 0$.

4.4 Definition. A process $(X_n)_{n\geqslant 1}$ is predictable (vorhersehbar) (w.r.t. $(\mathscr{F}_n)$) if each X_n is $\mathscr{F}_{n-1}$-measurable. For a predictable process (X_n) and a martingale (or more general: adapted process) (M_n) the martingale transform (or discrete stochastic integral) $((X{\bullet}M)_n)_{n\geqslant 0}$ is defined by $(X{\bullet}M)_0 := 0$, $(X{\bullet}M)_n := \sum_{k=1}^n X_k(M_k - M_{k-1})$.

4.5 Lemma. *For a bounded predictable (X_n) and a martingale (M_n) (or just predictable (X_n) and $X_n, M_n \in L^2$ for all n) $((X{\bullet}M)_n)_{n\geqslant 0}$ is again a martingale.*

4.6 Lemma. *Let (M_n) be a martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ convex with $\varphi(M_n) \in L^1$, $n \geqslant 0$. Then $\varphi(M_n)$ is a submartingale. In particular, (M_n^2) is a submartingale for an L^2-martingale (M_n).*

4.7 Theorem (Doob decomposition)**.** *Given a submartingale (X_n), there exists a martingale (M_n) and a predictable increasing process (A_n) such that*

$$X_n = X_0 + M_n + A_n, \quad n \geqslant 1; \qquad M_0 = A_0 = 0.$$

This decomposition is a.s. unique and $A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} \mid \mathscr{F}_{k-1}]$.

4.8 Definition. The predictable process (A_n) in the Doob decomposition of (X_n) is called compensator of (X_n). For an L^2-martingale (M_n) the compensator of (M_n^2) is called quadratic variation of (M_n), denoted by $\langle M \rangle_n$.

4.9 Lemma. *We have $\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}[(M_k - M_{k-1})^2 \mid \mathscr{F}_{k-1}]$, $n \geqslant 1$.*

4.2 Stopping times

4.10 Definition. A map $\tau : \Omega \to \{0, 1, \ldots, +\infty\}$ is called stopping time (Stoppzeit) with respect to a filtration $(\mathscr{F}_n)$ if $\{\tau = n\} \in \mathscr{F}_n$ holds for all $n \geqslant 0$.

4.11 Lemma. *Every deterministic time $\tau = n_0$ is stopping time. For stopping times σ and τ also $\sigma \wedge \tau$, $\sigma \vee \tau$ and $\sigma + \tau$ are stopping times.*

4.12 Theorem (Optional Stopping)**.** *Let (M_n) be a (sub/super-)martingale and τ a stopping time. Then the stopped process $(M_n^\tau) = (M_{n\wedge\tau})$ is again a (sub/super-)martingale.*

4.13 Definition. For a stopping time τ the σ-algebra of τ-history (τ-Vergangenheit) is defined by $\mathscr{F}_\tau := \{A \in \mathscr{F} \,|\, \forall n \geqslant 0 : \; A \cap \{\tau \leqslant n\} \in \mathscr{F}_n\}$.

4.14 Lemma. *$\mathscr{F}_\tau$ is a σ-Algebra and τ is $\mathscr{F}_\tau$-measurable.*

4.15 Lemma. *For stopping times σ and τ with $\sigma \leqslant \tau$ we have $\mathscr{F}_\sigma \subseteq \mathscr{F}_\tau$.*

4.16 Lemma. *For an adapted process (X_n) and a finite stopping time τ the random variable X_τ is $\mathscr{F}_\tau$-measurable.*

4.17 Theorem (Optional Sampling)**.** *Let (M_n) be a martingale (submartingale) and σ, τ bounded stopping times with $\sigma \leqslant \tau$. Then $\mathbb{E}[M_\tau \,|\, \mathscr{F}_\sigma] = M_\sigma$ (resp. $\mathbb{E}[M_\tau \,|\, \mathscr{F}_\sigma] \geqslant M_\sigma$) holds.*

4.18 Corollary. *Let (M_n) be a martingale and τ a finite stopping time. Then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$ holds under one of the following conditions:*

(a) τ is bounded;

(b) $(M_{\tau\wedge n})_{n\geqslant 0}$ is uniformly bounded;

(c) $\mathbb{E}[\tau] < \infty$ and $(\mathbb{E}[|M_{n+1} - M_n| \,|\, \mathscr{F}_n])_{n\geqslant 0}$ is uniformly bounded.

4.19 Corollary (Wald's Identity)**.** *Let $(X_k)_{k\geqslant 1}$ be $(\mathscr{F}_k)$-adapted random variables such that $\sup_k \mathbb{E}[|X_k|] < \infty$, $\mathbb{E}[X_k] = \mu \in \mathbb{R}$ and X_k is independent of $\mathscr{F}_{k-1}$, $k \geqslant 1$. Then for $S_n := \sum_{k=1}^n X_k$, $S_0 = 0$ and every $(\mathscr{F}_k)$-stopping time τ with $\mathbb{E}[\tau] < \infty$ we have $\mathbb{E}[S_\tau] = \mu\,\mathbb{E}[\tau]$.*

4.3 Martingale inequalities and convergence

4.20 Proposition (Maximal inequality)**.** *Any martingale (M_n) satisfies*

$$\forall \alpha > 0 : \; \mathbb{P}\Big(\sup_{0\leqslant k\leqslant n} |M_k| \geqslant \alpha\Big) \leqslant \tfrac{1}{\alpha}\,\mathbb{E}[|M_n|], \quad n \geqslant 0.$$

4.21 Theorem (Doob's L^p-inequality)**.** *An L^p-martingale (M_n) (i.e. $M_n \in L^p$ for all n) with $p > 1$ satisfies*

$$\Big\| \max_{1\leqslant k\leqslant n} |M_k| \Big\|_{L^p} \leqslant \frac{p}{p-1} \|M_n\|_{L^p}.$$

4.22 Definition. The number of upcrossings (aufsteigende Überquerungen) on an interval $[a,b]$ by a process (X_k) until time n is defined by $U_n^{[a,b]} := \sup\{k \geqslant 1 \,|\, \tau_k \leqslant n\}$, where inductively $\tau_0 := 0$, $\sigma_{k+1} := \inf\{\ell \geqslant \tau_k \,|\, X_\ell \leqslant a\}$, $\tau_{k+1} := \inf\{\ell \geqslant \sigma_k \,|\, X_\ell \geqslant b\}$.

4.23 Proposition (Upcrossing Inequality)**.** *The upcrossings of a submartingale (X_n) satisfy $\mathbb{E}[U_n^{[a,b]}] \leqslant \frac{1}{b-a}\,\mathbb{E}[(M_n - a) \vee 0]$.*

4.24 Theorem (First martingale convergence theorem)**.** *Let (M_n) be a (sub-/super-)martingale with $\sup_n \mathbb{E}[|M_n|] < \infty$. Then $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. and M_∞ is in L^1.*

4.25 Corollary. *Each non-negative supermartingale converges a.s.*

4.26 Proposition. *Let (M_n) be an L^2-martingale. Then $\lim_{n\to\infty} M_n(\omega)$ exists for $\mathbb{P}$-almost all ω, for which $\lim_{n\to\infty}\langle M\rangle_n(\omega) < \infty$ holds.*

4.27 Corollary (Strong law of large numbers for L^2-martingales)**.** *An L^2-martingale (M_n) satisfies for any $\alpha > 1/2$*

$$\lim_{n\to\infty} \frac{M_n(\omega)}{(\langle M\rangle_n(\omega))^\alpha} = 0$$

for $\mathbb{P}$-almost all ω, for which $\lim_{n\to\infty}\langle M\rangle_n(\omega)$ is infinite.

4.28 Definition. A family $(X_i)_{i\in I}$ of random variables is uniformly integrable (gleichgradig integrierbar) if

$$\lim_{R\to\infty} \sup_{i\in I} \mathbb{E}[|X_i| \mathbf{1}_{\{|X_i|>R\}}] = 0.$$

4.29 Lemma.

(a) *If $(X_i)_{i\in I}$ is uniformly integrable, then $(X_i)_{i\in I}$ is L^1-bounded: $\sup_{i\in I}\mathbb{E}[|X_i|] < \infty$.*

(b) *If $(X_i)_{i\in I}$ is L^p-bounded ($\sup_{i\in I}\mathbb{E}[|X_i|^p] < \infty$) for some $p > 1$, then $(X_i)_{i\in I}$ is uniformly integrable.*

(c) *If $|X_i| \leqslant Y$ holds for all $i \in I$ and some $Y \in L^1$, then $(X_i)_{i\in I}$ is uniformly integrable.*

4.30 Theorem (Vitali)**.** *Let $(X_n)_{n\geqslant 0}$ be random variables with $X_n \xrightarrow{\mathbb{P}} X$ (in probability). Then the following statements are equivalent:*

(a) *$(X_n)_{n\geqslant 0}$ is uniformly integrable;*

(b) *$X_n \to X$ in L^1;*

(c) *$\mathbb{E}[|X_n|] \to \mathbb{E}[|X|] < \infty$.*

4.31 Theorem (Second martingale convergence theorem)**.**

(a) *If* (M_n) *is a uniformly integrable martingale, then* (M_n) *converges a.s. and in* L^1 *to some* $M_\infty \in L^1$. (M_n) *is closable with* $M_n = \mathbb{E}[M_\infty \mid \mathscr{F}_n]$.

(b) *If* (M_n) *is a closable martingale, with* $M_n = \mathbb{E}[M \mid \mathscr{F}_n]$ *say, then* (M_n) *is uniformly integrable and (a) holds with* $M_\infty = \mathbb{E}[M \mid \mathcal{F}_\infty]$ *where* $\mathscr{F}_\infty = \sigma(\mathscr{F}_n, n \geqslant 1)$.

4.32 Corollary. *Let* $p > 1$. *Every* L^p*-bounded martingale* (M_n) *(i.e.* $\sup_n \mathbb{E}[|M_n|^p] < \infty$*) converges for* $n \to \infty$ *a.s. and in* L^p*, hence also in* L^1.

4.33 Definition. A process $(M_{-n})_{n \geqslant 0}$ is called backward martingale (Rückwärtsmartingal) with respect to $(\mathscr{F}_{-n})_{n \geqslant 0}$ with $\mathscr{F}_{-n-1} \subseteq \mathscr{F}_{-n}$ if $M_{-n} \in L^1$, M_{-n} $\mathscr{F}_{-n}$-measurable and $\mathbb{E}[M_{-n} \mid \mathscr{F}_{-n-1}] = M_{-n-1}$ hold for all $n \geqslant 0$.

4.34 Theorem. *Every backward martingale* $(M_{-n})_{n \geqslant 0}$ *converges for* $n \to \infty$ *a.s. and in* L^1.

4.35 Corollary. *(Kolmogorov's strong law of large numbers) For i.i.d. random variables* $(X_k)_{k \geqslant 1}$ *in* L^1 *we have*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{a.s.\ and\ L^1} \mathbb{E}[X_1].$$

4.4 The Radon-Nikodym theorem

4.36 Definition. Let μ and ν be measures on the measurable space $(\Omega, \mathscr{F})$. Then μ is absolutely continuous (absolutstetig) with respect to ν, notation $\mu \ll \nu$, if $\forall A \in \mathscr{F} : \nu(A) = 0 \Rightarrow \mu(A) = 0$. μ and ν are equivalent (äquivalent), notation $\mu \sim \nu$, if $\mu \ll \nu$ and $\nu \ll \mu$. If there is an $A \in \mathscr{F}$ with $\nu(A) = 0$ and $\mu(A^C) = 0$, then μ and ν are singular (singulär), notation $\mu \perp \nu$.

4.37 Theorem (Radon-Nikodym). *Let* ν *be a* σ*-finite measure and* μ *a finite measure with* $\mu \ll \nu$*, then there is an* $f \in L^1(\Omega, \mathscr{F}, \nu)$ *such that*

$$\mu(A) = \int_A f \, d\nu \text{ for all } A \in \mathscr{F}.$$

4.38 Definition. The function f in the Radon-Nikodym theorem is called Radon-Nikodym derivative, density or likelihood function of μ with respect to ν, notation $f = \frac{d\mu}{d\nu}$.

4.39 Theorem (Kakutani). *Let* $(X_k)_{k \geqslant 1}$ *be independent random variables with* $X_k \geqslant 0$ *and* $\mathbb{E}[X_k] = 1$. *Then* $M_n := \prod_{k=1}^n X_k$, $M_0 = 1$ *is a non-negative martingale converging a.s. to some* M_∞. *The following statements are equivalent:*

(a) $\mathbb{E}[M_\infty] = 1$;

(b) $M_n \to M_\infty$ *in* L^1;

(c) (M_n) *is uniformly integrable;*

(d) $\prod_{k=1}^{\infty} a_k > 0$, *where* $a_k := \mathbb{E}[X_k^{1/2}] \in (0, 1]$;

(e) $\sum_{k=1}^{\infty}(1-a_k)<\infty$.

If one (then all) statement fails to hold, then $M_\infty = 0$ holds a.s. (Kakutani's dichotomy).

5 Markov chains: recurrence and transience

In this section $(X_n,\, n \geqslant 0)$ always denotes a time-homogeneous Markov chain with state space $(S, \mathcal{S})$, realized as coordinate process on $\Omega = S^{\mathbb{N}_0}$ with σ-algebra $\mathscr{F} = \mathcal{S}^{\otimes \mathbb{N}_0}$, filtration $\mathscr{F}_n = \mathscr{F}_n^X$ and measure $\mathbb{P}_\mu$, where μ denotes the initial distribution. We write short $\mathbb{P}_x := \mathbb{P}_{\delta_x}$.

5.1 Definition. For $n \geqslant 0$ the shift operator $\vartheta_n : \Omega \to \Omega$ is given by $\vartheta_n((s_k)_{k\geqslant 0}) = (s_{k+n})_{k\geqslant 0}$.

5.2 Theorem. *Let $Y \in \mathcal{M}^+(\Omega, \mathscr{F})$ and τ be a finite $(\mathscr{F}_n)$-stopping time. Then the strong Markov property holds:*

$$\mathbb{E}_\mu[Y \circ \vartheta_\tau \,|\, \mathscr{F}_\tau] = \mathbb{E}_{X_\tau}[Y] \quad \mathbb{P}_\mu\text{-a.s.}$$

5.3 Definition. For $y \in S$, $k \in \mathbb{N}$ introduce the k^{th} time of return to y recursively by $T_y^k := \inf\{n > T_y^{k-1} \,|\, X_n = y\}$ and $T_y^0 := 0$. Put $T_y := T_y^1$ and $\rho_{xy} := \mathbb{P}_x(T_y < \infty)$ for $x \in S$.

5.4 Proposition. *For $k \in \mathbb{N}$ and $x, y \in S$ we have $P_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}$.*

5.5 Definition. A state $y \in S$ is called recurrent (rekurrent) if $\rho_{yy} = 1$ and transient (transient) if $\rho_{yy} < 1$.

5.6 Definition. By $N_y := \sum_{n\geqslant 1} \mathbf{1}_{\{X_n = y\}}$ we denote the number of visits to state y.

5.7 Proposition.

(a) If a state y is transient, then $\mathbb{E}_x[N_y] = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty$ holds for all $x \in S$.

(b) A state y is recurrent if and only if $\mathbb{E}_y[N_y] = \infty$ holds.

5.8 Proposition. *Let $x \in S$ be recurrent and $\rho_{xy} > 0$ for some $y \in S$. Then y is recurrent and $\rho_{yx} = 1$.*

5.9 Definition. A set $C \subseteq S$ of states is closed (abgeschlossen) if $\rho_{xy} = 0$ holds for all $x \in C$, $y \in S \setminus C$. A set $D \subseteq S$ is irreducible (irreduzibel) if $\rho_{xy} > 0$ holds for all $x, y \in D$. If S is irreducible, then the Markov chain is called irreducible.

5.10 Proposition. *For an irreducible Markov chain on a finite state space S all states are recurrent.*

6 Ergodic theory

6.1 Stationary and ergodic processes

6.1 Definition. A stochastic process $(X_t,\, t \in T)$ with $T \in \{\mathbb{N}_0, \mathbb{Z}, \mathbb{R}^+, \mathbb{R}\}$ is stationary (stationär) if $(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+s}, \dots, X_{t_n+s})$ holds for all $n \geqslant 1$, $t_1, \cdots, t_n \in T$ and $s \in T$.

6.2 Definition. For a time-homogeneous Markov chain $(X_n,\, n \geqslant 0)$ an initial distribution μ is invariant if $\mathbb{P}_\mu(X_1 = i) = \mathbb{P}_\mu(X_0 = i) = \mu(\{i\})$ holds for all $i \in S$.

6.3 Lemma. *A time-homogeneous Markov chain with invariant initial distribution is stationary.*

6.4 Definition. A measurable map $T : \Omega \to \Omega$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called measure-preserving (maßerhaltend) if $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$ holds for all $A \in \mathscr{F}$.

6.5 Lemma.

(a) *Every S-valued stationary process $(X_n,\, n \geqslant 0)$ induces a measure-preserving transformation T on $(S^{\mathbb{N}_0}, \mathcal{S}^{\otimes \mathbb{N}_0}, \mathbb{P}^X)$ via*

$$T((x_0, x_1, x_2, \dots)) = (x_1, x_2, \cdots) \text{ (left shift).}$$

(b) *For a random variable Y and a measure-preserving map T on $(\Omega, \mathscr{F}, \mathbb{P})$ the process $X_n(\omega) := Y(T^n(\omega))$, $n \geqslant 0$, ($T^0 := \mathrm{Id}$) is stationary.*

6.6 Definition. A event A is (almost) invariant with respect to a measure-preserving map T on $(\Omega, \mathscr{F}, \mathbb{P})$ if $\mathbb{P}(T^{-1}(A) \Delta A) = 0$ holds. The σ-algebra (!) of all (almost) invariant events is denoted by $\mathscr{I}_T$. T is ergodic if $\mathscr{I}_T$ is trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ holds for all $A \in \mathscr{I}_T$.

6.7 Lemma. *Let $\mathscr{I}_T$ be the invariant σ-algebra with respect to some measure-preserving transformation T on $(\Omega, \mathscr{F}, \mathbb{P})$. Then:*

(a) *A (real-valued) random variable Y is $\mathscr{I}_T$-measurable if and only if it is $\mathbb{P}$-a.s. invariant, i.e. $\mathbb{P}(Y \circ T = Y) = 1$. In particular, T is ergodic if and only if each $\mathbb{P}$-a.s. invariant and bounded random variable is $\mathbb{P}$-a.s. constant.*

(b) *For each invariant event $A \in \mathscr{I}_T$ there exists a strictly invariant event B (i.e. with $T^{-1}(B) = B$ exactly) such that $\mathbb{P}(A \Delta B) = 0$.*

6.2 Ergodic theorems

6.8 Lemma (Maximal ergodic lemma)**.** *Let $X \in L^1$ and T be measure-preserving on $(\Omega, \mathscr{F}, \mathbb{P})$. Denoting $S_n := \sum_{i=0}^{n-1} X \circ T^i$, $S_0 := 0$ and $M_n := \max\{S_0, \dots, S_n\}$, we have $\mathbb{E}[X \mathbf{1}_{\{M_n > 0\}}] \geqslant 0$.*

6.9 Theorem (Birkhoff's ergodic theorem)**.** *Let $X \in L^1$ and T be measure-preserving on $(\Omega, \mathscr{F}, \mathbb{P})$. Then:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = \mathbb{E}[X \mid \mathscr{I}_T] \qquad \mathbb{P}\text{-a.s. and in } L^1.$$

If T is even ergodic, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = \mathbb{E}[X] \qquad \mathbb{P}\text{-a.s. and in } L^1.$$

6.10 Theorem (von Neumann's ergodic theorem)**.** *For $X \in L^p$, $p \geqslant 1$, and measure-preserving T on $(\Omega, \mathscr{F}, \mathbb{P})$we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = \mathbb{E}[X \mid \mathscr{I}_T] \qquad \mathbb{P}\text{-a.s. and in } L^p.$$

6.11 Corollary. *Let $(X_n,\, n \geqslant 0)$ be an ergodic process in L^1 (i.e. $X_n \in L^1$ and the associated left shift on $(S^{\mathbb{N}_0}, \mathcal{S}^{\mathbb{N}_0}, \mathbb{P}^X)$ is ergodic). Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X_i = \mathbb{E}[X_1] \qquad \mathbb{P}\text{-a.s. and in } L^1.$$

In particular, Kolmogorov's strong law of large number for (X_n) in L^1 follows.

6.3 The structure of the invariant measures

6.12 Definition. Let $T : \Omega \to \Omega$ be measurable on $(\Omega, \mathscr{F})$. Each probability measure μ on $\mathscr{F}$ with $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathscr{F}$ is called invariant with respect to T. If T is even ergodic on $(\Omega, \mathscr{F}, \mu)$, then also μ is called ergodic. The set of all invariant probability measures with respect to T is denoted by $\mathscr{M}_T$.

6.13 Lemma. *$\mathscr{M}_T$ is convex.*

6.14 Proposition. *Any two distinct ergodic measures are singular.*

6.15 Theorem. *The ergodic measures are exactly the extremal points of the convex set $\mathscr{M}_T$.*

6.16 Corollary. *If T possesses exactly one invariant probability measure, then this measure is ergodic.*

6.4 Application to Markov chains

6.17 Definition. A recurrent state $x \in S$ is called positive-recurrent if $\mathbb{E}_x[T_x] < \infty$, otherwise it is called null-recurrent.

6.18 Theorem. *Suppose $x \in S$ is positive-recurrent and set*

$$\mu(\{y\}) := \frac{\mathbb{E}_x[\sum_{n=0}^{T_x-1} \mathbf{1}_{\{X_n=y\}}]}{\mathbb{E}_x[T_x]} = \frac{\sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y,\, T_x > n)}{\mathbb{E}_x[T_x]}, \quad y \in S.$$

Then μ is an invariant initial distribution.

6.19 Corollary. *If $(X_n,\, n \geqslant 0)$ is an irreducible Markov chain with some positive-recurrent state x, then it is an ergodic process under the preceding invariant initial distribution μ.*

6.20 Theorem. *If an irreducible Markov chain $(X_n,\, n \geqslant 0)$ has an invariant initial distribution μ, then all its states are positive-recurrent and $\mu(\{y\}) = 1/\mathbb{E}_y[T_y]$, $y \in S$, holds.*

7 Weak convergence

7.1 Fundamental properties

Throughout $(S, \mathfrak{B}_S)$ denotes a metric space with Borel σ-algebra. The space of all bounded continuous and real-valued functions on S is denoted by $C_b(S)$.

7.1 Definition. Probability measures $\mathbb{P}_n$ converge weakly (schwach) to a probability measure $\mathbb{P}$ on $(S, \mathfrak{B}_S)$ if

$$\forall f \in C_b(S): \lim_{n\to\infty} \int_S f\, d\,\mathbb{P}_n = \int_S f\, d\,\mathbb{P}$$

holds, notation $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$. $(S, \mathfrak{B}_S)$-valued random variables X_n converge in distribution (or in law, in Verteilung) to some random variable X if $\mathbb{P}^{X_n} \xrightarrow{w} \mathbb{P}^X$ holds, i.e.

$$\forall f \in C_b(S): \lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

Notation $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{d} \mathbb{P}^X$.

7.2 Proposition. *For $(S, \mathfrak{B}_S)$-valued random variables $d(X_n, X) \xrightarrow{\mathbb{P}} 0$ (in probability) implies $X_n \xrightarrow{d} X$.*

7.3 Theorem (Portmanteau)**.** *For probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$, $\mathbb{P}$ on $(S, \mathfrak{B}_S)$ the following are equivalent:*

(a) $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$;

(b) $\forall\, U \subseteq S$ *open*: $\liminf_{n\to\infty} \mathbb{P}_n(U) \geqslant \mathbb{P}(U)$;

(c) $\forall\, F \subseteq S$ *closed*: $\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F)$;

(d) $\forall\, A \in \mathfrak{B}_S$ *with* $\mathbb{P}(\partial A) = 0$: $\lim_{n\to\infty} \mathbb{P}_n(A) = \mathbb{P}(A)$.

7.4 Theorem (Continuous mapping)**.** *If $g: S \to T$ is continuous, T another metric space, then: $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$.*

7.5 Proposition. *$\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ is already valid if $\int f\, d\mathbb{P}_n \to \int f\, d\mathbb{P}$ holds for all bounded, Lipschitz-continuous functions f.*

7.6 Lemma. *(Slutsky) We have for $(S, \mathfrak{B}_s)$-valued random variables (X_n), (Y_n)*

$$X_n \xrightarrow{d} X,\ d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0 \Rightarrow Y_n \xrightarrow{d} X.$$

7.7 Corollary. *If real-valued random variables satisfy $Y_n \xrightarrow{d} a$, $a \in \mathbb{R}$, and $X_n \xrightarrow{d} X$, then $(X_n, Y_n) \xrightarrow{d} (X, a)$, in particular $X_n Y_n \xrightarrow{d} aX$, $X_n + Y_n \xrightarrow{d} X + a$.*

7.2 Tightness

7.8 Definition. A family $(\mathbb{P}_i)_{i\in I}$ of probability measures on $(S, \mathfrak{B}_S)$ is called (weakly) relatively compact if each sequence $(\mathbb{P}_{i_k})_{k\geqslant 1}$ has a weakly convergent subsequence. The family $(\mathbb{P}_i)_{i\in I}$ is (uniformly) tight (straff) if for any $\varepsilon > 0$ there is a compact set $K_\varepsilon \subseteq S$ such that $\mathbb{P}_i(K_\varepsilon) \geqslant 1 - \varepsilon$ for all $i \in I$.

7.9 Theorem. *Any relatively compact family of probability measures on a separable metric space is tight.*

7.10 Theorem (Prohorov). *Any tight family of probability measures on a Polish space is relatively compact.*

7.11 Corollary (Prohorov). *On a Polish space a family of probability measures is relatively compact if and only if it is tight.*

7.3 Weak convergence on $C([0,T])$, $C(\mathbb{R}^+)$

In the sequel C stands for $C([0,T])$ or $C(\mathbb{R}^+)$, equipped with the supremum norm and the uniform convergence on compact sets, respectively.

7.12 Theorem. *A sequence $(\mathbb{P}_n)$ of probability measures on $\mathfrak{B}_C$ converges weakly to $\mathbb{P}$ if and only if all finite-dimensional distributions $\mathbb{P}_n(\pi^{-1}_{\{t_1,\dots,t_m\}}(\bullet))$ converge weakly to $\mathbb{P}(\pi^{-1}_{\{t_1,\dots,t_m\}}(\bullet))$ and $(\mathbb{P}_n)$ is tight.*

7.13 Definition. For $f \in C([0,T])$ and $\delta > 0$ the modulus of continuity (Stetigkeitsmodul) is defined as

$$\omega_\delta(f) := \max\{|f(s) - f(t)| \,|\, s, t \in [0,T],\ |s-t| \leqslant \delta\}.$$

7.14 Theorem (Arzelà-Ascoli). *A subset $A \subseteq C([0,T])$ is relatively compact if*

(a) $\sup_{f\in A} |f(0)| < \infty$ *and*

(b) $\lim_{\delta\to 0} \sup_{f\in A} \omega_\delta(f) = 0$ *(uniform integrability).*

7.15 Corollary. *A sequence $(\mathbb{P}_n)_{n\geqslant 1}$ of probability measures on $\mathfrak{B}_{C([0,T])}$ is tight if and only if*

(a) $\lim_{R\to\infty} \sup_n \mathbb{P}_n(\{|f(0)| > R\}) = 0$ *and*

(b) $\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}_n(\{\omega_\delta(f)\geqslant\varepsilon\})=0$ *for all* $\varepsilon>0$.

7.16 Lemma. *A sequence* $(\mathbb{P}_n)_{n\geqslant 1}$ *of probability measures on* $\mathfrak{B}_{C([0,T])}$ *is already tight if*

(a) $\lim_{R\to\infty}\sup_n\mathbb{P}_n(\{|f(0)|>R\})=0$ *and*

(b') $\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{t\in[0,T-\delta]}\delta^{-1}\,\mathbb{P}_n(\{\max_{s\in[t,t+\delta]}|f(s)-f(t)|\geqslant\varepsilon\})=0$ *for all* $\varepsilon>0$.

Tightness on $\mathfrak{B}_{C(\mathbb{R}^+)}$ *follows if conditions (a), (b') are satisfied for all* $T>0$.

7.17 Theorem. *Let* $(X_n(t),\,0\leqslant t\leqslant T)$*,* $n\geqslant 1$*, be continuous processes. Then their laws* $\mathbb{P}^{X_n}$ *are tight on* $C([0,T])$ *if*

(a) $\lim_{R\to\infty}\sup_n\mathbb{P}(\{|X_n(0)|>R\})=0$ *and*

(b") $\exists\alpha,\,\beta>0,\,K>0\;\forall\,n\geqslant 1,\,s,t\in[0,T]:\;\mathbb{E}[|X_s^{(n)}-X_t^{(n)}|^\alpha]\leqslant K|s-t|^{1+\beta}$.

8 Invariance principle and the empirical process

8.1 Invariance principle and Brownian motion

8.1 Definition. A process $(B_t,\,t\geqslant 0)$ is called <u>Brownian motion</u> (Brownsche Bewegung) if

(a) $B_0=0$ and $B_t\sim N(0,t)$, $t>0$, holds;

(b) the increments are stationary and independent: for $0\leqslant t_0<t_1<\dots<t_m$ we have $(B_{t_1}-B_{t_0},\dots,B_{t_m}-B_{t_{m-1}})\sim N(0,\operatorname{diag}(t_1-t_0,\dots,t_m-t_{m-1}))$;

(c) B has continuous sample paths.

8.2 Lemma. *Suppose* $(X_k)_{k\geqslant 1}$ *are i.i.d.,* $X_k\in L^2$*,* $\mathbb{E}[X_k]=0$*,* $\operatorname{Var}(X_k)=1$*. Consider* $S_n:=\sum_{k=1}^n X_k$*,* $S_0=0$ *and the rescaled, linearly interpolated random walk*

$$Y_n(t):=\frac{1}{\sqrt{n}}S_{\lfloor nt\rfloor}+\frac{nt-\lfloor nt\rfloor}{\sqrt{n}}X_{\lfloor nt\rfloor+1},\quad t\in[0,1].$$

Then the finite-dimensional distributions of Y_n *converge to those of a Brownian motion.*

8.3 Lemma. *In the setting of the preceding lemma we have for any* $\lambda\geqslant\sqrt{2}$ $N\in\mathbb{N}$

$$\mathbb{P}\Big(\max_{1\leqslant n\leqslant N}|S_n|\geqslant\lambda\sqrt{N}\Big)\leqslant\mathbb{P}\Big(|S_N|\geqslant(\lambda-\sqrt{2})\sqrt{N}\Big).$$

8.4 Theorem (Donsker, functional CLT)**.** *In the setting of the preceding lemmata we have* $Y^{(n)}\xrightarrow{d}B$ *with a Brownian motion* $(B_t,\,0\leqslant t\leqslant 1)$ *and convergence in distribution on* $(C([0,1]),\mathfrak{B}_{C([0,1])})$.

8.5 Corollary. *Brownian motion exists.*

8.6 Proposition (Reflection principle)**.** *Let* $(X_k)_{k\geqslant 1}$ *be a sequence of i.i.d. random variables in* L^2 *with* $\mathbb{E}[X_k]=0$*,* $\mathbb{E}[X_k^2]=1$*. Set* $S_n:=\sum_{k=1}^n X_k$*,* $M_n:=\frac{1}{\sqrt{n}}\max_{1\leqslant i\leqslant n}S_i$*. Then* $M_n\xrightarrow{d}|B_1|$ *follows with* $B_1\sim N(0,1)$*. Also for the Brownian motion* B *we have:* $\max_{0\leqslant t\leqslant 1}B_t\stackrel{d}{=}|B_1|$.

8.2 Empirical process and Brownian bridge

8.7 Definition. For i.i.d. real-valued random variables $X_1, \dots, X_n$ with distribution function F the (random) function

$$F_n(x) := \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leqslant x\}}, \quad x \in \mathbb{R},$$

is called empirical distribution function (empirische Verteilungsfunktion). The associated empirical process (empirischer Prozess) is given by

$$G_n(x) := \sqrt{n}(F_n(x) - F(x)), \quad x \in \mathbb{R}\,.$$

8.8 Lemma. *For each $x \in \mathbb{R}$ we have $G_n(x) \xrightarrow{d} N(0, F(x)(1 - F(x)))$ as $n \to \infty$.*

8.9 Definition. The Brownian bridge $(X_t, t \in [0,1])$ is the(!) centered and continuous Gaussian process with $\mathrm{Cov}(X_s, X_t) = s(1-t)$ for $0 \leqslant s \leqslant t \leqslant 1$.

8.10 Theorem (Donsker)**.** *Let $X_1, \dots, X_n$ be independent $U([0,1])$-distributed random variables. Consider the linear interpolation $\tilde{F}_n : [0,1] \to [0,1]$ of F_n satisfying $\tilde{F}_n(X_i) = F_n(X_i)$, $i = 1, \dots, n$, $\tilde{F}_n(0) = 0$, $\tilde{F}_n(1) = 1$ and the associated empirical process $\tilde{G}_n = \sqrt{n}(\tilde{F}_n - F)$. Then we have convergence of $\tilde{G}_n$ to a Brownian bridge B in distribution on $C([0,1])$: $\tilde{G}_n \xrightarrow{d} B$.*

8.11 Corollary (Kolmogorov-Smirnov)**.** *Let $X_1, \dots, X_n$ be i.i.d. random variables with continuous distribution function F_0 and $T_n := \sqrt{n}\sup_{x\in\mathbb{R}}|F_n(x) - F_0(x)|$. Then $T_n \xrightarrow{d} \max_{0\leqslant t\leqslant 1}|B(t)|$ holds with a Brownian bridge B.*
(The latter has a so-called Kolomogorov distribution: $\mathbb{P}(\max_{0\leqslant t\leqslant 1}|B(t)| \leqslant x) = \sum_{j\in\mathbb{Z}}(-1)^j e^{-2j^2x^2}$, $x > 0$.)

8.12 Corollary. *Let $\alpha \in (0,1)$ and let $X_1, \dots, X_n$ be i.i.d. random variables with continuous distribution function F which is continuously differentiable in a neighbourhood of $q_\alpha := F^{-1}(\alpha)$ with $f(q_\alpha) := F'(q_\alpha) > 0$. Then the empirical quantile $\hat{q}_\alpha^n := \tilde{F}_n^{-1}(\alpha)$ satisfies $\sqrt{n}(\hat{q}_\alpha^n - q_\alpha) \xrightarrow{d} N(0, \frac{\alpha(1-\alpha)}{f^2(q_\alpha)})$.*