Stochastic Processes (Stochastik II)
Outline of the course in winter semester 2013/14

Markus Reiß
Humboldt-Universität zu Berlin

Preliminary version, 16. Januar 2014

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

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

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

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

(iv) (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 (iii),(iv) 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 (iii) of a Poisson process, $\mathbb{E}[N_1] = \lambda$ and (1.1) is the density of $(S_1, \ldots, S_n)$ given $\{N_t = n\}$.

1.4 Remark. Let $U_1, \ldots U_n \sim U([0,t])$ i.i.d. and consider their order statistics $U_{(1)}, \ldots, U_{(n)}$, i.e. $U_{(1)} = \min_i U_i$, $U_{(2)} = \min(\{U_1, \ldots, U_n\} \setminus \{U_{(1)}\})$ etc. Then $(U_{(1)}, \ldots, U_{(n)})$ has exactly density (1.1).

The characterisations give rise to three simple methods to simulate a Poisson process: the definition gives an approximation for small h (forgetting the $o(h)$-term), part (c) 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!).

Proof. We prove the equivalence by a circle argument.

(a)⇒(b) Put $p_n(t) = \mathbb{P}(N_t = n)$. By (i), (ii), (iii) we infer

$$p_0(t+h) = \mathbb{P}(N_t = 0, N_{t+h} - N_t = 0) = p_0(t)(1 - \lambda h + o(h)),$$

which implies

$$p_0'(t) = \lim_{h \downarrow 0} \frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t), \quad t \geqslant 0.$$

In view of $p_0(0) = 1$ we obtain $p_0(t) = e^{-\lambda t}$.

Similarly, we have for $n \geqslant 1$:

$$\begin{aligned} p_n(t+h) &= \mathbb{P}(\{N_{t+h} = n\} \cap (\{N_t \leqslant n-2\} \cup \{N_t = n-1\} \cup \{N_t = n\})) \\ &= \mathbb{P}(N_t \leqslant n-2) o(h) + \mathbb{P}(N_t = n-1)(\lambda h + o(h)) \\ &\quad + \mathbb{P}(N_t = n)(1 - \lambda h + o(h)) \\ &= p_{n-1}(t)\lambda h + p_n(t)(1 - \lambda h) + o(h). \end{aligned}$$

This implies $p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t)$. Using $p_n(0) = 0$ we infer $p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$.

(b)⇒(c) Let $0 = b_0 \leqslant a_1 < b_1 \leqslant \cdots \leqslant a_n < b_n$ and calculate

$$\begin{aligned} &\mathbb{P}\Big(\bigcap_{k=1}^{n} \{a_k \leqslant S_k \leqslant b_k\}\Big) \\ &= \mathbb{P}\Big(\bigcap_{k=1}^{n-1} \{N_{a_k} - N_{b_{k-1}} = 0, N_{b_k} - N_{a_k} = 1\} \cap \{N_{a_n} - N_{b_{n-1}} = 0, N_{b_n} - N_{a_n} \geqslant 1\}\Big) \\ &\overset{(iii),(iv)}{=} \Big(\prod_{k=1}^{n-1} \mathbb{P}(N_{a_k - b_{k-1}} = 0)\, \mathbb{P}(N_{b_k - a_k})\Big)\, \mathbb{P}(N_{a_n - b_{n-1}} = 0)\, \mathbb{P}(N_{b_n - a_n} \geqslant 1) \\ &= \Big(\prod_{k=1}^{n-1} \lambda(b_k - a_k) e^{-\lambda(b_k - a_k) - \lambda(a_k - b_{k-1})}\Big) e^{-\lambda(a_n - b_{n-1})}(1 - e^{-\lambda(b_n - a_n)}) \\ &= (e^{-\lambda a_n} - e^{-\lambda b_n}) \lambda^{n-1} \prod_{k=1}^{n-1} (b_k - a_k) \\ &= \int_{a_1}^{b_1} \int_{a_2 - x_1}^{b_2 - x_1} \cdots \int_{a_n - x_1 - \cdots - x_{n-1}}^{b_n - x_1 - \cdots - x_{n-1}} \lambda^n e^{-\lambda(x_1 + \cdots + x_n)} dx_n \cdots dx_2 dx_1. \end{aligned}$$

Consequently, $(T_1, T_2, \ldots, T_n) = (S_1, S_2 - S_1, \ldots, S_n - S_{n-1})$ has density $\lambda^n e^{-\lambda(x_1 + \cdots + x_n)}$ for $x_i \geqslant 0$. The product density form implies that $T_1, \ldots, T_n$ are independent and each T_i is $\mathrm{Exp}(\lambda)$-distributed.

(c)⇒(d) We find $\mathbb{P}(N_t = 0) = \mathbb{P}(S_1 > t) = e^{-\lambda t}$ and

$$\mathbb{P}(N_t = n) = \mathbb{P}(N_t \geqslant n) - \mathbb{P}(N_t \geqslant n+1) = \mathbb{P}(S_n \leqslant t) - \mathbb{P}(S_{n+1} \leqslant t).$$

Since $S_n = T_1 + \cdots + T_n$ is $\Gamma(\lambda, n)$-distributed, we obtain

$$\mathbb{P}(N_t = n) = \int_0^t \Big(\frac{\lambda^n x^{n-1}}{(n-1)!} - \frac{\lambda^{n+1}x^n}{n!}\Big)e^{-\lambda x}dx = \frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

and we conclude $N_t \sim \mathrm{Poiss}(\lambda t)$. By density transformation the joint density of $(S_1, \dots, S_{n+1})$ is for $s_{n+1} \geqslant s_n \geqslant \cdots \geqslant s_1 \geqslant s_0 = 0$

$$f^{S_1,\dots,S_{n+1}}(s_1, \dots, s_{n+1}) = \prod_{k=1}^{n+1} \lambda e^{-\lambda(s_k - s_{k-1})} = \lambda^{n+1}e^{-\lambda s_{n+1}}.$$

Noting $\{N_t = n\} = \{S_n \leqslant t, S_{n+1} > t\}$ we consider $0 \leqslant a_1 < b_1 \leqslant \cdots \leqslant a_n < b_n \leqslant t$ and obtain the conditional law via

$$\begin{aligned}
&\mathbb{P}(S_1 \in [a_1, b_1], \dots, S_n \in [a_n, b_n] \,|\, N_t = n)\\
&= \frac{\mathbb{P}(S_1 \in [a_1, b_1], \dots, S_n \in [a_n, b_n], S_{n+1} > t)}{\frac{(\lambda t)^n}{n!}e^{-\lambda t}}\\
&= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{n!}{t^n}\mathbf{1}(0 \leqslant s_1 \leqslant \cdots \leqslant s_n \leqslant t)\, ds_n \cdots ds_1,
\end{aligned}$$

which identifies the integrand as the conditional density.

(d)⇒(e) $\mathbb{E}[N_1] = \lambda$ is direct from the assumption. For $0 = t_0 < t_1 < \cdots t_n = t$ and $k_1, \dots, k_n \in \mathbb{N}_0$ consider with $K := \sum_{l=1}^n k_l$

$$\begin{aligned}
&\mathbb{P}(\forall l = 1, \dots, n:\ N_{t_l} - N_{t_{l-1}} = k_l)\\
&= \mathbb{P}(N_{t_n} = K)\,\mathbb{P}(\forall l = 1, \dots, n:\ N_{t_l} - N_{t_{l-1}} = k_l \,|\, N_t = K)\\
&= \frac{(\lambda t)^K}{K!}e^{-\lambda t}\,\mathbb{P}(S_{k_1} \leqslant t_1 < S_{k_1+1}, \dots, S_K \leqslant t_n < S_{K+1} \,|\, N_t = K)\\
&= \frac{(\lambda t)^K}{K!}e^{-\lambda t}\frac{K!}{t^K}\prod_{l=1}^n \frac{(t_l - t_{l-1})^{k_l}}{k_l!}\\
&= \prod_{l=1}^n \mathbb{P}(N_{t_l} - N_{t_{l-1}} = k_l).
\end{aligned}$$

Hence, $(N_{t_l} - N_{t_{l-1}})_l$ are independent.

(e)⇒(a) For $0 = t_0 < t_1 < \cdots t_n = t$ and $k_1, \dots, k_n \in \mathbb{N}_0$, $h > 0$, $m \geqslant k_1 + \cdots + k_n$ note the shift invariance

$$\begin{aligned}
&\mathbb{P}(\forall l = 1, \dots, n:\ N_{t_l+h} - N_{t_{l-1}+h} = k_l \,|\, N_{t+h} = m)\\
&= \frac{m!}{(t+h)^m}\prod_{l=1}^m \frac{(t_l + h - (t_{l-1} + h))^{k_l}}{k_l!}\\
&= \mathbb{P}(\forall l = 1, \dots, n:\ N_{t_l} - N_{t_{l-1}} = k_l \,|\, N_{t+h} = m)
\end{aligned}$$

Summing up over all $m \geqslant k_1 + \cdots + k_n$ yields identity in law:

$$(N_{t_1+h} - N_{t_0+h}, \dots, N_{t_n+h} - N_{t_{n-1}+h}) \stackrel{d}{=} (N_{t_1} - N_{t_0}, \dots, N_{t_n} - N_{t_{n-1}}).$$

This gives (iv) (put $n=1$) and for $0<h<1$

$$\mathbb{P}(N_h=0)=\sum_{k=0}^{\infty}\mathbb{P}(N_1=k)\,\mathbb{P}(N_1-N_h=k\,|\,N_1=k)=\sum_{k=0}^{\infty}\mathbb{P}(N_1=k)(1-h)^k.$$

Because of $\sum_{k=0}^{\infty}\mathbb{P}(N_1=k)k=\mathbb{E}[N_1]=\lambda<\infty$ the function $p(h):=\mathbb{P}(N_h=0)$ is differentiable in $[0,1]$ with $p'(0)=-\lambda$. We conclude

$$\mathbb{P}(N_h=0)=\mathbb{P}(N_0=0)-\lambda h+o(h)=1-\lambda h+o(h).$$

By a similar argument, $\mathbb{P}(N_h=1)$ equals

$$\sum_{k=1}^{\infty}\mathbb{P}(N_1=k)\,\mathbb{P}(N_1-N_h=k-1\,|\,N_1=k)=\sum_{k=1}^{\infty}\mathbb{P}(N_1=k)k(1-h)^{k-1},$$

and this implies $\mathbb{P}(N_h=1)=\lambda h+o(h)$.

□

1.2 Markov chains

1.5 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 $n\in\mathbb{N}$, $t_1<t_2<\dots<t_{n+1}$, $s_1,\dots,s_{n+1}\in S$ with $\mathbb{P}(X_{t_1}=s_1,\dots,X_{t_n}=s_n)>0$ the Markov property is satisfied:

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

1.6 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 $t_1\leqslant t_2$.

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

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

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* $n\in\mathbb{N}$.

Proof. By definition we obtain

$$\begin{aligned}
P(t_1,t_3)_{ij} &= \mathbb{P}(X_{t_3}=j \,|\, X_{t_1}=i) \\
&= \sum_{k\in S} \mathbb{P}(X_{t_3}=j, X_{t_2}=k \,|\, X_{t_1}=i) \\
&= \sum_{k\in S} \mathbb{P}(X_{t_3}=j \,|\, X_{t_1}=i, X_{t_2}=k)\, \mathbb{P}(X_{t_2}=k \,|\, X_{t_1}=i) \\
&\overset{\text{Markov}}{=} \sum_{k\in S} \mathbb{P}(X_{t_3}=j \,|\, X_{t_2}=k)\, \mathbb{P}(X_{t_2}=k \,|\, X_{t_1}=i) \\
&= \sum_{k\in S} P(t_2,t_3)_{kj} P(t_1,t_2)_{ik} \\
&= (P(t_1,t_2)P(t_2,t_3))_{ij}.
\end{aligned}$$

For time-homogeneous Markov chains this reduces to $P(t_3-t_1) = P(t_2-t_1)P(t_3-t_2)$ and substituting $t = t_2 - t_1$, $s = t_3 - t_2$ yields the assertion. □

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

Proof. We have to show measurability. Since $\mathscr{S}^{\otimes T}$ is generated by the projections $\pi_t : S^T \to S$, $t\in T$, onto the t-th coordinate, $\bar{X}$ is measurable if all compositions $\pi_t \circ \bar{X} : \Omega \to S$ are measurable, but by definition $\pi_t\circ\bar{X} = X_t$, $t\in T$, are measurable as random variables. □

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

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 Remarks.

(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 $(X_t, t \geqslant 0)$ and $(Y_t, t \geqslant 0)$ are real-valued stochastic processes with right-continuous sample paths. Then they are indistinguishable already if they are versions of each other. ▶ Exercise

2.7 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).

2.8 Remark. Every continuous process is stochastically continuous since almost sure convergence implies stochastic convergence. On the other hand, the Poisson process is stochastically continuous, but obviously not continuous:

$$\forall \varepsilon \in (0,1): \lim_{t_n \to t} \mathbb{P}(|N_t - N_{t_n}| > \varepsilon) = \lim_{t_n \to t} (1 - e^{-\lambda |t - t_n|}) = 0.$$

2.2 Polish spaces and Kolmogorov's consistency theorem

2.9 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.

2.10 Definition. For finitely or countably many metric spaces (S_k, d_k) the product space $\prod_k S_k$ is canonically equipped with the product metric $d((s_k),(t_k)) := \sum_k 2^{-k}(d_k(s_k, t_k) \wedge 1)$, which generates the product topology, in which a vector/sequence converges iff all coordinates converge.

2.11 Lemma. *Let S_k, $k \geqslant 1$, be Polish spaces, then the Borel σ-algebra of the product satisfies $\mathfrak{B}_{\prod_{k\geqslant 1} S_k} = \bigotimes_{k\geqslant 1} \mathfrak{B}_{S_k}$.*

Proof. $\bigotimes_{k\geqslant 1} \mathfrak{B}_{S_k}$ is the smallest σ-algebra such that the coordinate projections $\pi_i : \prod_{k\geqslant 1} S_k \to S_i$, $i \geqslant 1$, are measurable. Analogously, the product topology on $\prod_{k\geqslant 1} S_k$ is the coarsest topology such that all π_i are continuous. Consequently, each π_i is in particular $\mathfrak{B}_{\prod_{k\geqslant 1} S_k}$-measurable, which implies $\mathfrak{B}_{\prod_{k\geqslant 1} S_k} \supseteq \bigotimes_{k\geqslant 1} \mathfrak{B}_{S_k}$.

By separability, any open set $O \subseteq \prod_{k\geqslant 1} S_k$ is a countable union of open sets of the form $\bigcap_{i=1}^N \pi_i^{-1}(O_i)$ with $N \in \mathbb{N}$ and $O_i \in S_i$ open, which are elements of $\bigotimes_{k\geqslant 1} \mathfrak{B}_{S_k}$. This shows $\mathfrak{B}_{\prod_{k\geqslant 1} S_k} \subseteq \bigotimes_{k\geqslant 1} \mathfrak{B}_{S_k}$. □

2.12 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.

2.13 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 : $P(O \setminus B) \leqslant \varepsilon$.

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

Proof. Let $(s_n)_{n \geqslant 1}$ be a dense sequence in S and consider for any radius $\rho > 0$ the closed balls $B_\rho(s_n)$ around s_n. Then $S = \bigcup_n B_\rho(s_n)$ and σ-continuity implies

$$\lim_{N \to \infty} \mathbb{P}\Big(\bigcup_{n=1}^{N} B_\rho(s_n)\Big) = 1.$$

Now select for $\varepsilon > 0$ and every $\rho = 1/k$ an index N_k such that

$$\mathbb{P}\Big(\bigcup_{n=1}^{N_k} B_{1/k}(s_n)\Big) \geqslant 1 - \varepsilon 2^{-k}.$$

Then $K := \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{N_k} B_{1/k}(s_n)$ is a closed subset, hence complete. Since for any $\delta > 0$ there is a finite cover of K by balls $B_{1/k}(s_n)$ of diameter less than δ (K is *totally bounded*), any sequence in K has a subsequence which is Cauchy. By completeness, the Cauchy sequence converges and K is compact. By construction,

$$\mathbb{P}(S \setminus K) = \mathbb{P}\Big(\bigcup_{k=1}^{\infty} \bigcap_{n=1}^{N_k} B_{1/k}(s_n)^{\complement}\Big) \leqslant \sum_{k=1}^{\infty} \varepsilon 2^{-k} = \varepsilon$$

holds. Since $\varepsilon > 0$ was arbitrary, this shows tightness. □

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

Proof. We consider the family of Borel sets

$$\mathcal{D} := \Big\{ B \in \mathfrak{B}_S \,\Big|\, P(B) = \sup_{K \subseteq B \text{ compact}} P(K) = \inf_{O \supseteq B \text{ open}} P(O) \Big\}.$$

Note first $S \in \mathcal{D}$ because S is open and $\mathbb{P}$ is tight by the preceding theorem.

Now consider any closed set $F \subseteq S$. By tightness, for any $\varepsilon > 0$ there is a compact set K_ε with $\mathbb{P}(K_\varepsilon) \geqslant 1 - \varepsilon$. Then $F \cap K_\varepsilon \subseteq F$ is compact with

$$\mathbb{P}(F \setminus (F \cap K_\varepsilon)) \leqslant \mathbb{P}(K_\varepsilon^{\complement}) \leqslant \varepsilon.$$

This shows $\mathbb{P}(F) = \sup_K \mathbb{P}(K)$ with $K \subseteq F$ compact. The open sets $O_n := \{s \in S \mid \inf_{x \in F} d(s, x) < 1/n\}$ satisfy $F = \bigcap_{n \geqslant 1} O_n$. By σ-continuity, we infer $\mathbb{P}(F) = \inf_{N \geqslant 1} \mathbb{P}(\bigcap_{n=1}^{N} O_n)$. Since finite intersections of open sets are open, we have shown the second regularity property and thus $F \in \mathcal{D}$.

Furthermore, $\mathcal{D}$ is closed under set differences and countable unions ($\mathcal{D}$ is a *σ-ring*) ► Exercise .

Altogether we have shown that $\mathcal{D}$ is a σ-algebra containing the closed sets, which implies $\mathcal{D} = \mathfrak{B}_S$, as asserted. □

2.16 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 $\pi_{J,I}((s_j)_{j\in J}) = (s_j)_{j\in I}$. Then the finite-dimensional distributions satisfy the following consistency condition:*

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

Proof. We just write

$$\begin{aligned} P_I(A) &= \mathbb{P}((X_t)_{t\in I} \in A) = \mathbb{P}(\bar{X} \in \pi_{T,I}^{-1}(A)) \\ &= \mathbb{P}(\bar{X} \in (\pi_{J,I} \circ \pi_{T,J})^{-1}(A)) = \mathbb{P}((X_t)_{t\in J} \in \pi_{J,I}^{-1}(A)) \\ &= P_J(\pi_{J,I}^{-1}(A)). \end{aligned}$$

□

2.17 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:

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

2.18 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 of probability measures for S and I. Then there exists a unique probability measure $\mathbb{P}$ on the product space $(S^I, \mathfrak{B}_S^{\otimes I})$ satisfying*

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

Proof. Let $\mathfrak{A} := \bigcup_{J\subseteq I \text{ finite}} \pi_{I,J}^{-1}(\mathfrak{B}_S^{\otimes J})$ be the algebra (check!) of cylinder sets on S^I, which generates $\mathfrak{B}_S^{\otimes I}$. Since $\mathfrak{A}$ is $\cap$-stable, $\mathbb{P}$ is uniquely determined by its values on the cylinder sets.

The existence of $\mathbb{P}$ follows from Caratheodory's extension theorem if $\mathbb{P}$ on $\mathfrak{A}$, as defined in the theorem, is a premeasure. The consistency of (P_J) ensures that $\mathbb{P}$ is well-defined on $\mathfrak{A}$ and additive: for disjoint $A, B \in \mathfrak{A}$ there are a finite $J \subseteq I$ and disjoint $A', B' \in \mathfrak{B}_S^{\otimes J}$ with $A = \pi_{I,J}^{-1}(A')$, $B = \pi_{I,J}^{-1}(B')$. Since $\mathbb{P}_J$ is a probability measure and standard set operations commute with taking preimages, we conclude

$$\mathbb{P}(A \cup B) = \mathbb{P}_J(A' \cup B') = \mathbb{P}_J(A') + \mathbb{P}_J(B') = \mathbb{P}(A) + \mathbb{P}(B).$$

Trivially, also $\mathbb{P}(S^I) = \mathbb{P}_J(S^J) = 1$ holds, using any finite $J \subseteq I$. It remains to show that $\mathbb{P}$ is σ-additive on $\mathfrak{A}$, which is (under finite additivity) equivalent to $\mathbb{P}(B_n) \to 0$ for any sequence $B_n \downarrow \varnothing$ of sets $B_n \in \mathfrak{A}$ (σ-continuity at $\varnothing$).

We can write $B_n = \pi_{I,J_n}^{-1}(A_n)$ for some finite $J_n \subseteq I$, $A_n \in \mathfrak{B}_S^{\otimes J_n}$. Without loss of generality we shall assume $J_n \subseteq J_{n+1}$ for all n. Now let $K_n \subseteq A_n$ be compact with $\mathbb{P}_{J_n}(A_n \setminus K_n) \leqslant \varepsilon 2^{-n}$ by Ulam's Theorem. Then $K_n' = \bigcap_{l=1}^{n-1} \pi_{J_n,J_l}^{-1}(K_l) \cap K_n$ is compact in S^{J_n} as a closed subset of a compact set and $C_n = \pi_{I,J_n}^{-1}(K_n') = \cap_{l=1}^n \pi_{I,J_l}^{-1}(K_l) \subseteq B_n$ satisfies also $C_n \downarrow \varnothing$. Below we prove that there is already an $n_0 \in \mathbb{N}$ with $C_{n_0} = \varnothing$. From this we conclude

$$\limsup_{n\to\infty} P(B_n) \leqslant \mathbb{P}(B_{n_0}) = \mathbb{P}(B_{n_0} \setminus C_{n_0}) \leqslant \sum_{l=1}^{n_0} \mathbb{P}_{J_l}(A_l \setminus K_l) \leqslant \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows $\mathbb{P}(B_n) \to 0$, as desired.

We prove the claim via reductio ad absurdum, assuming that for all $n \in \mathbb{N}$ there is a $y_n \in C_n$. Since K_n' is compact in S^{J_n}, we can find a subsequence $(n_l^{(1)})$, such that $(\pi_{I,J_1}(y_{n_l^{(1)}}))_{l\geqslant 1}$ converges in K_1', a further subsequence $(n_l^{(2)})$ such that $(\pi_{I,J_2}(y_{n_l^{(2)}}))_{l\geqslant 1}$ converges in K_2' and so on. Along the diagonal sequence $(n_l^{(l)})_{l\geqslant 1}$ then $(\pi_{I,J_m}(y_{n_l^{(l)}}))_{l\geqslant 1}$ converges in K_m' for all $m \geqslant 1$. Hence, $(\pi_{I,\cup_{m\geqslant 1} J_m}(y_{n_l^{(l)}}))_{l\geqslant 1}$ converges in the product topology (metric) to some $z \in S^{\cup_{m\geqslant 1} J_m}$ (note: $\bigcup_{m\geqslant 1} J_m$ is countable). Because $C_{n+1} \subseteq C_n$, $n \geqslant 1$, are nested, this implies $z \in \pi_{I,\cup_{m\geqslant 1} J_m}(C_n)$ for all $n \geqslant 1$ and thus $z \in \pi_{I,\cup_{m\geqslant 1} J_m}(\bigcap_{n\geqslant 1} C_n)$. This contradicts $\bigcap_{n\geqslant 1} C_n = \varnothing$ and the claim is proved. □

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

Proof. By Kolmogorov's consistency theorem construct the probability measure $\mathbb{P}$ on $(S^T, \mathfrak{B}_S^{\otimes T})$ which satisfies $\mathbb{P}(\pi_{T,\{t_1,\dots,t_n\}}^{-1}(A)) = \mathbb{P}_{\{t_1,\dots,t_n\}}(A)$ for all $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, $A \in \mathfrak{B}_S^n$. Define X to be the coordinate process on $(S^T, \mathfrak{B}_S^{\otimes T}, \mathbb{P})$ via $X_t((s_\tau)_{\tau\in T}) := s_t$. Then X_t is measurable for every $t \in T$ and

$$\mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in A) = \mathbb{P}(\pi_{T,\{t_1,\dots,t_n\}}^{-1}(A)) = \mathbb{P}_{\{t_1,\dots,t_n\}}(A)$$

for all $A \in \mathfrak{B}_S^{\otimes n}$. □

2.20 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 for $(S, \mathscr{S})$ Polish: for finite product measures the consistency condition holds because for all $B \in \mathfrak{B}_S^{\otimes J}$

$$\Big(\bigotimes_{j\in J'} \mathbb{P}_j\Big)(\pi_{J',J}^{-1}(B)) = \Big(\bigotimes_{j\in J} \mathbb{P}_j\Big)(B)\bullet\Big(\bigotimes_{j\in J'\setminus J} \mathbb{P}_j\Big)(S^{J'\setminus J}) = \Big(\bigotimes_{j\in J} \mathbb{P}_j\Big)(B).$$

Define $X_i : S^I \to S$ by $X_i((s_j)_{j\in I}) := s_i$. Then the assertion follows from the preceding corollary. For general measure spaces $(S, \mathscr{S})$ the proof is similar to that of Kolmogorov's consistency theorem, see e.g. Bauer (1991). □

2.21 Remark. Kolmogorov's consistency theorem does not hold for general measure spaces $(S, \mathcal{S})$, cf. the counterexample by Sparre Andersen, Jessen (1948). The Ionescu-Tulcea Theorem, however, shows the existence of the probability measure on general measure spaces under a Markovian dependence structure, see e.g. Klenke (2008).

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 general conditional expectation 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) $\mathbb{E}[M_{n+1} \,|\, \mathscr{F}_n] = M_n$ (resp. $\mathbb{E}[M_{n+1} \,|\, \mathscr{F}_n] \geqslant M_n$ for submartingale, resp. $\mathbb{E}[M_{n+1} \,|\, \mathscr{F}_n] \leqslant M_n$ 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 $M_n = \mathbb{E}[X \,|\, \mathscr{F}_n]$, $n \geqslant 0$.

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 $A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} \,|\, \mathscr{F}_{k-1}]$.

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

4.2 Stopping times

4.10 Definition. A map $\tau : \Omega \to \{0, 1, \dots, +\infty\}$ 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 $U_n^{[a,b]} := \sup\{k \geqslant 1 \mid \tau_k \leqslant n\}$, where inductively $\tau_0 := 0$, $\sigma_{k+1} := \inf\{\ell \geqslant \tau_k \mid X_\ell \leqslant a\}$, $\tau_{k+1} := \inf\{\ell \geqslant \sigma_k \mid X_\ell \geqslant b\}$.

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*

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

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 $(X_{t_1}, \ldots, X_{t_n}) \overset{d}{=} (X_{t_1+s}, \ldots, X_{t_n+s})$ 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

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

(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 $M_n := \max\{S_0, \ldots, S_n\}$, 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* $(\mathbb{R}^{\mathbb{N}_0}, \mathfrak{B}_{\mathbb{R}}^{\otimes \mathbb{N}_0}, \mathbb{P}^X)$ *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 Theorem. *An irreducible Markov chain has at most one invariant initial distribution μ. If it exists, it satisfies $\mu(\{y\}) > 0$ for all $y \in S$ and the Markov chain is ergodic under μ.*

6.20 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 invariant initial distribution μ from Theorem 6.18.*

6.21 Theorem. *If an irreducible Markov chain $(X_n,\, n \geqslant 0)$ has an invariant initial distribution μ, then it is ergodic, 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.*