Stochastic Processes (Stochastik II)
Outline of the course in winter semester 2016/17

Markus Reiß
Humboldt-Universität zu Berlin

Preliminary version, 9. Februar 2017

Inhaltsverzeichnis

1 Some important processes

1.1 The Poisson process

1.1 Example. We count the number N_t of emissions of a radioactive substance during the time interval $[0,t]$ for $t \in [0,\infty)$. Since radioactivity is genuinely random, $N_t = N_t(\omega)$ is a random variable with values in $\mathbb{N}_0$. We write N_t, $N(t)$, $N_t(\omega)$, $N(t,\omega)$ synonymously.

1.2 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 $\omega \in \Omega$. 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).

1.3 Example. For the radioactive emissions physical modeling assumes that on small time intervals the probability for more than one emission is negligible and the probability for one emission is proportional to the length of the interval. Moreover, the number of emissions during disjoint time intervals is independent and its distribution only depends on the length of the time interval, not on the time points itself. Hence, we model N as a Poisson process in the following sense.

1.4 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 $h \downarrow 0$ and all $t \geqslant 0$;

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

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

1.5 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 $t \geqslant 0$ (setting $\mathrm{Poiss}(0) = \delta_0$);

(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 $t \geqslant 0$ and the law of $(S_1, \ldots, S_n)$ given $\{N_t = n\}$ has for all $n \in \mathbb{N}$ 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 $(S_1, \dots, S_n)$ given $\{N_t = n\}$ for all $n \in \mathbb{N}$, $t > 0$.*

1.6 Remark. Let $U_1, \dots U_n \sim U([0,t])$ i.i.d. and consider their order statistics $U_{(1)}, \dots, U_{(n)}$, i.e. $U_{(1)} = \min_i U_i$, $U_{(2)} = \min(\{U_1, \dots, U_n\} \setminus \{U_{(1)}\})$ etc. Then $(U_{(1)}, \dots, U_{(n)})$ 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. Note that (c) ensures also the existence of a Poisson process because we can construct a probability space with i.i.d. Exp(λ)-distributed random variables $(T_k)_{k\geqslant 1}$, from which we can construct $(S_k)_{k\geqslant 1}$ and thus N.

Proof. We prove the equivalence by a circular 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 $p_0(0) = 1$ (from (iv) with $t = s$) 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 $0 = b_0 \leqslant a_1 < b_1 \leqslant \dots \leqslant a_n < b_n$ and calculate

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{k=1}^{n} \{a_k < S_k \leqslant b_k\}\Big) \\
&= \mathbb{P}\Big(\bigcap_{k=1}^{n-1} \{N_{a_k} - N_{b_{k-1}} = 0, N_{b_k} - N_{a_k} = 1\} \cap \{N_{a_n} - N_{b_{n-1}} = 0, N_{b_n} - N_{a_n} \geqslant 1\}\Big) \\
&\overset{(iii),(iv)}{=} \Big(\prod_{k=1}^{n-1} \mathbb{P}(N_{a_k - b_{k-1}} = 0)\, \mathbb{P}(N_{b_k - a_k} = 1)\Big)\, \mathbb{P}(N_{a_n - b_{n-1}} = 0)\, \mathbb{P}(N_{b_n - a_n} \geqslant 1) \\
&= \Big(\prod_{k=1}^{n-1} \lambda(b_k - a_k) e^{-\lambda(b_k - a_k) - \lambda(a_k - b_{k-1})}\Big) e^{-\lambda(a_n - b_{n-1})}(1 - e^{-\lambda(b_n - a_n)}) \\
&= (e^{-\lambda a_n} - e^{-\lambda b_n}) \lambda^{n-1} \prod_{k=1}^{n-1} (b_k - a_k).
\end{aligned}$$

Taking derivatives with respect to $b_1, \dots, b_n$ we obtain the density of $(S_1, \dots, S_n)$:

$$f^{S_1,\dots,S_n}(b_1, \dots, b_n) = \lambda^n e^{-\lambda b_n} \mathbf{1}(0 \leqslant b_1 \leqslant b_2 \leqslant \dots \leqslant b_n).$$

Consequently, $(T_1, T_2, \dots, T_n) = (S_1, S_2 - S_1, \dots, S_n - S_{n-1})$ has density $\lambda^n e^{-\lambda(x_1+\dots+x_n)} \mathbf{1}(x_1, \dots, x_n \geqslant 0)$ (density transformation). The product density form implies that $T_1, \dots, T_n$ are independent and that 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 $S_n = T_1 + \dots + T_n$ 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 $s_{n+1} \geqslant s_n \geqslant \dots \geqslant s_1 \geqslant s_0 = 0$

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

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

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

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 $K := \sum_{l=1}^n k_l$

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

The last identity follows from

$$\mathbb{P}(N_{t_{l_0}} - N_{t_{l_0-1}} = k_{l_0}) = \frac{\big(\lambda(t_{l_0} - t_{l_0-1})\big)^{k_{l_0}}}{k_{l_0}!} e^{-\lambda(t_{l_0}-t_{l_0-1})}, \quad l_0 = 1, \dots, n,$$

obtained from the previous calculation by summing $\mathbb{P}(\forall l = 1, \dots, n : N_{t_l} - N_{t_{l-1}} = k_l)$ over $k_l \in \mathbb{N}_0$ for all $l \neq l_0$. Hence, $(N_{t_l} - N_{t_{l-1}})_l$ are independent.

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

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

We thus have

$$\mathbb{P}(\forall l : N_{t_l+h} - N_{t_{l-1}+h} = k_l, N_{t+h} = m) = \mathbb{P}(\forall l : N_{t_l} - N_{t_{l-1}} = k_l, N_{t+h} = m)$$

and 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 observe $N_0 = 0$ a.s. due to the existence of a density for S_1) and for $0 < h < 1$

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

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

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

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

□

1.2 Markov chains

1.7 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 $n \in \mathbb{N}$, $t_1 < t_2 < \cdots < t_{n+1}$, $i_1, \dots, i_{n+1} \in S$ with $\mathbb{P}(X_{t_1} = i_1, \dots, X_{t_n} = i_n) > 0$ the Markov property is satisfied:

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \,|\, X_{t_1} = i_1, \dots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} \,|\, X_{t_n} = i_n).$$

1.8 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 $t_1 \leqslant t_2$.

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

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

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 $(X_{t_1}, \ldots, X_{t_n})$ with $n \geqslant 1$, $t_1, \ldots, t_n \in T$ are called finite-dimensional distributions of X. We write $P^X_{t_1,\ldots,t_n} := \mathbb{P}^{(X_{t_1},\ldots,X_{t_n})}$.

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. For instance, defining counting processes via $N_t = \sum_{k=1}^{\infty} \mathbf{1}(S_k < t)$ yields left-continuous sample paths and the left- and right-continuous Poisson processes are versions of each other, but clearly distinguishable.

(b) If X is a version of Y, then X and Y share the same finite-dimensional distributions because countable intersections of sets of measure one have measure one and thus $\mathbb{P}(X_{t_1} = Y_{t_1}, \ldots, X_{t_n} = Y_{t_n}) = 1$. 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) If T is countable, then a version Y of X is also indistinguishable from X because countable intersections of 1-sets are 1-sets. 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:

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

Note that for stochastically continuous processes the finite-dimensional distributions vary continuously in time with respect to convergence in distribution.

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 Example. In analysis it is shown that $\mathbb{R}^d$ with any norm, $(C([a,b];\mathbb{R}), \|\bullet\|_\infty)$, $\ell^p(\mathbb{N})$ and $L^p(\mathbb{R})$, $L^p([a,b])$ for $p \in [1,\infty)$ are separable Banach (complete normed) spaces and thus Polish. The rational numbers $\mathbb{Q}$ with Euclidean distance are not complete, the spaces ℓ^∞, $L^\infty([a,b])$ are examples of non-separable Banach spaces.

2.11 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.

2.12 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. Put $S = \prod_{k \geqslant 1} S_k$ and assume without loss of generality that infinitely many S_k are given (otherwise take the infinite product with one-element spaces S_k, $k \geqslant K+1$). Then $\bigotimes_{k\geqslant 1} \mathfrak{B}_{S_k}$ is the smallest σ-algebra such that the coordinate projections $\pi_i : S \to S_i$, $i \geqslant 1$, are measurable. Analogously, the product topology on S is the coarsest topology such that all π_i are continuous. Consequently, each π_i is in particular $\mathfrak{B}_S$-measurable, which implies $\mathfrak{B}_S \supseteq \bigotimes_{k \geqslant 1} \mathfrak{B}_{S_k}$.

By separability of S_k also S is separable (consider the countable product of countable dense sets), and any open set $O \subseteq S$ is a countable union of open balls in S. Any such ball $B_r(s) = \{t \in S \mid \sum_k 2^{-k}(d_k(s_k,t_k) \wedge 1) < r\}$ can be represented as a countable union of cylinder sets:

$$B_r(s) = \bigcup_{K \in \mathbb{N}} \Big\{ t \in S \,\Big|\, \sum_{k=1}^{K} 2^{-k}(d_k(s_k,t_k) \wedge 1) < r - 2^{-K} \Big\} \in \bigotimes_{k \geqslant 1} \mathfrak{B}_{S_k}.$$

This shows $\mathfrak{B}_S \subseteq \bigotimes_{k \geqslant 1} \mathfrak{B}_{S_k}$. □

2.13 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, see e.g. Elstrodt § III.5.3 and exercises.

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

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

(b) <u>regular</u> (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 : $P(O \setminus B) \leqslant \varepsilon$.

2.15 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 $\bar{B}_\rho(s_n)$ around s_n. Then $S = \bigcup_n \bar{B}_\rho(s_n)$ and σ-continuity implies

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

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

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

Then $K := \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{N_k} \bar{B}_{1/k}(s_n)$ is a closed subset, hence complete. Since for any $\delta > 0$ there is a finite cover of K by balls $\bar{B}_{1/k}(s_n)$ 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,

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

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

2.16 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, ▶ Exercise $\mathcal{D}$ is closed under set differences and countable unions. Altogether we have shown that $\mathcal{D}$ is a σ-algebra containing the closed sets, which implies $\mathcal{D} = \mathfrak{B}_S$, as asserted. □

2.17 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 $I \subseteq J \subseteq T$ the coordinate projection $\pi_{J,I}((s_j)_{j\in J}) = (s_j)_{j\in I}$. Then the finite-dimensional distributions $(P_J^X)_{J\subseteq T \text{ finite}}$ satisfy the following consistency condition:*

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

Proof. We just write with $\bar{X}$ from above:

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

□

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

$$\forall J \subseteq J' \subseteq T \text{ finite, } A \in \mathscr{S}^{\otimes J} :\ \mathbb{P}_J(A) = \mathbb{P}_{J'}(\pi_{J',J}^{-1}(A)).$$

2.19 Theorem (Kolmogorov's consistency/extension theorem; Daniell 1919, Kolmogorov 1933)**.** *Let $(S, \mathfrak{B}_S)$ be a Polish space, $T \neq \varnothing$ an index set and let $(\mathbb{P}_J)$ be a projective family of probability measures for S and T. Then there exists a unique probability measure $\mathbb{P}$ on the product space $(S^T, \mathfrak{B}_S^{\otimes T})$ satisfying*

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

Proof. Let $\mathfrak{A} := \bigcup_{J\subseteq T \text{ finite}} \pi_{T,J}^{-1}(\mathfrak{B}_S^{\otimes J})$ be the algebra (check!) of cylinder sets on S^T, which generates $\mathfrak{B}_S^{\otimes T}$. 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 $J \subseteq T$ and disjoint $A', B' \in \mathfrak{B}_S^{\otimes J}$ with $A = \pi_{T,J}^{-1}(A')$, $B = \pi_{T,J}^{-1}(B')$. 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 $\mathbb{P}(S^T) = \mathbb{P}_J(S^J) = 1$ holds, using any finite $J \subseteq T$. 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 $B_n = \pi^{-1}_{T,J_n}(A_n)$ for some finite $J_n \subseteq T$, $A_n \in \mathfrak{B}_S^{\otimes J_n}$. 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 $K_n' = \bigcap_{l=1}^{n-1} \pi^{-1}_{J_n,J_l}(K_l) \cap K_n$ is compact in S^{J_n} as a closed subset of a compact set and $C_n = \pi^{-1}_{T,J_n}(K_n') = \cap_{l=1}^n \pi^{-1}_{T,J_l}(K_l) \subseteq B_n$ 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 and the decay of $\mathbb{P}(B_n)$ we conclude

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

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 $(\pi_{T,J_1}(y_{n_l^{(1)}}))_{l\geqslant 1}$ converges in K_1', a further subsequence $(n_l^{(2)})$ such that $(\pi_{T,J_2}(y_{n_l^{(2)}}))_{l\geqslant 1}$ converges in K_2' and so on. Along the diagonal sequence $(n_l^{(l)})_{l\geqslant 1}$ then $(\pi_{T,J_m}(y_{n_l^{(l)}}))_{l\geqslant 1}$ converges in K_m' for all $m \geqslant 1$. Hence, $(\pi_{T,\cup_{m\geqslant 1} J_m}(y_{n_l^{(l)}}))_{l\geqslant 1}$ 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). As $C_{n+1} \subseteq C_n$, $n \geqslant 1$, are nested, this implies $z \in \pi_{T,\cup_{m\geqslant 1} J_m}(C_n)$ for all $n \geqslant 1$ and thus $z \in \pi_{T,\cup_{m\geqslant 1} J_m}(\bigcap_{n\geqslant 1} C_n)$. This contradicts $\bigcap_{n\geqslant 1} C_n = \varnothing$ and the claim is proved. □

2.20 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 $\mathbb{P}(\pi^{-1}_{T,\{t_1,\dots,t_n\}}(A)) = \mathbb{P}_{\{t_1,\dots,t_n\}}(A)$ for all $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, $A \in \mathfrak{B}_S^{\otimes n}$. 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

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

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

2.21 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 $B \in \mathfrak{B}_S^{\otimes J}$

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

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 Stochastik I or e.g. Bauer (1991). □

2.22 Example (Markov chains)**.** Let $(S, \mathcal{P}(S))$ be a countable state space. Let an initial distribution $\mu^{(0)}$ on $\mathcal{P}(S)$ and transition probabilities $(p_{i,j})_{i,j\in S}$ be given, that is a stochastic matrix $P = (p_{i,j})_{i,j\in S}$. Then we want to construct a discrete time-homogeneous Markov chain $(X_n,\, n \geqslant 0)$ with $\mathbb{P}^{X_0} = \mu_0$ and $\mathbb{P}(X_n = j \,|\, X_{n-1} = i) = p_{i,j}$ whenever $\mathbb{P}(X_{n-1} = i) > 0$. Note first that $(S, \mathcal{P}(S))$ becomes Polish if we consider the discrete metric $d(s,t) = \mathbf{1}(s \neq t)$ such that $\mathfrak{B}_S = \mathcal{P}(S)$. Consider

$$\mu_n(A) = \sum_{i_0\in S} \cdots \sum_{i_n\in S} \mathbf{1}_A(i_0,\dots,i_n)\mu^{(0)}_{i_0} p_{i_0,i_1}\cdots p_{i_{n-1},i_n}, \quad A \subseteq S^{n+1}.$$

Then, letting $T = \mathbb{N}_0$ and $\mu_{t_1,\dots,t_n}(B) = \mu_{t_n}(\pi^{-1}_{\{1,\dots,t_n\},\{t_1,\dots,t_n\}}(B))$ we see by induction that $(\mu_J)_{J\subseteq T}$ is a projective family iff $\mu_{n+1}(\pi^{-1}_{n+1,n}(A)) = \mu_n(A)$ holds for all $n \geqslant 0$, $A \subseteq S^{n+1}$. The latter is easily verified and by the consistency theorem such a Markov chain always exists.

2.23 Example (Gaussian processes)**.** Let $T \neq \varnothing$ an index set, e.g. $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}^+$ or $\mathbb{R}$, and $\mu : T \to \mathbb{R}$ be any function, $c : T^2 \to \mathbb{R}$ any symmetric, positive semi-definite function, that is $c(t,s) = c(s,t)$ and $\sum_{i,j=1}^n c(t_i,t_j)\alpha_i\alpha_j \geqslant 0$ for all $t,s \in T$, $n \in \mathbb{N}$, $t_1,\dots,t_n \in T$, $\alpha_1,\dots,\alpha_n \in \mathbb{R}$. Then there exists a process $(X_t, t \in T)$ whose finite-dimensional distributions are Gaussian:

$$P^X_{t_1,\dots,t_n} = N\big((\mu(t_1),\dots\mu(t_n))^\top, (c(t_i,t_j))_{i,j=1,\dots,n}\big).$$

This Gaussian process is constructed by the consistency theorem, checking the consistency condition via projections of Gaussian vectors. μ is called expectation function and c covariance function of X.

2.24 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 $\|x - y_x\| = \operatorname{dist}(x, L) := \inf_{y\in L}\|x - y\|$.*

Proof. For $x \in L$ we have $y_x = x$. Otherwise, there is a sequence $(y_n) \subseteq L$ with $\|x - y_n\| \to \operatorname{dist}(x, L)$. Let us show that (y_n) is Cauchy. Note

$$\|y_n - y_m\|^2 = 2\|x - y_m\|^2 + 2\|x - y_n\|^2 - 4\|x - (y_m + y_n)/2\|^2.$$

Since $(y_m + y_n)/2 \in L$ and $\|x - (y_m + y_n)/2\| \leqslant (\|x - y_m\| + \|x - y_n\|)/2$, we see $\lim_{m,n\to\infty}\|x - (y_m + y_n)/2\| = \operatorname{dist}(x, L)$. From the identity above we thus conclude $\lim_{m,n\to\infty}\|y_n - y_m\|^2 = 0$.

By completeness of H and closedness of L, the Cauchy sequence (y_n) has a limit $y_x \in L$. By continuity of the norm, we obtain $\|x - y_x\| = \operatorname{dist}(x, L)$. For another element $z_x \in L$ with this property we obtain $\|x - (y_x + z_x)/2\| < \operatorname{dist}(x, L)$, unless $y_x = z_x$. Hence, $y_x = z_x$ and uniqueness follows. □

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).*

Proof. By definition, $x \in L \Rightarrow P_L x = x$ and $P_L y \in L$ such that $P_L(P_L y) = P_L y$ for all $y \in H$ and (a) holds. For all $x \in H$, $y \in L$ we obtain

$$\|x - P_L x\|^2 \leqslant \|x - (P_L x + y)\|^2 = \|x - P_L x\|^2 + \|y\|^2 - 2\langle x - P_L x, y\rangle.$$

For all $\alpha \in \mathbb{R} \setminus \{0\}$ we therefore find

$$2\langle x - P_L x, \alpha y\rangle \leqslant \|ay\|^2 \Rightarrow 2 \operatorname{sgn}(\alpha)\langle x - P_L x, y\rangle \leqslant |\alpha| \|y\|^2.$$

Letting $\alpha \downarrow 0$ and $\alpha \uparrow 0$, this implies $\langle x - P_L x, y\rangle = 0$ and thus (b). □

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.*

Proof. For (a) it remains to prove uniqueness. Writing $x = y + z$ with $y \in L$, $z \in L^\perp$, we deduce $y - P_L x = (x - P_L x) - z \in L \cap L^\perp = \{0\}$ and thus $y = P_L x$, $z = x - P_L x$. Properties (b) and (c) follow by using the decomposition in (a). □

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

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

3.6 Proposition (Factorisation Lemma)**.** *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 $Y=\varphi(X)$.*

Proof. We argue via measure-theoretic induction. For simple $Y=\sum_{k=1}^n a_k\mathbf{1}_{B_k}$ with $a_k\in\mathbb{R}$ we can assume w.l.o.g. that $B_k\cap B_l=\varnothing$. Then $B_k=Y^{-1}(\{a_k\})$ is in $\sigma(X)$ (since Y is $\sigma(X)$-measurable) and thus $B_k=X^{-1}(A_k)$ for $A_k\in\mathscr{S}$. Therefore $\varphi:=\sum_{k=1}^n a_k\mathbf{1}_{A_k}$ is $\mathscr{S}$-measurable and satisfies $\varphi(X)=Y$.

For $Y\in\mathcal{M}^+(\Omega,\sigma(X))$ we can find simple Y_n with $Y_n\uparrow Y$ and measurable φ_n with $Y_n=\varphi_n(X)$. Let us define inductively $\tilde\varphi_1=\varphi_1$, $\tilde\varphi_{n+1}(x)=\max(\varphi_{n+1}(x),\tilde\varphi_n(x))$. Since $\varphi_{n+1}(X)=Y_{n+1}\geqslant Y_n=\varphi_n(X)$, we have $\varphi_{n+1}(x)\geqslant\varphi_n(x)$ for all $x\in X(\omega)$ and thus $\tilde\varphi_n(X)=\varphi_n(X)=Y$. Moreover, $\tilde\varphi_n$ is measurable and $\tilde\varphi_{n+1}(x)\geqslant\tilde\varphi_n(x)$ holds for all x. Then $\varphi(x)=\lim_{n\to\infty}\tilde\varphi_n(x)\in[0,\infty]$ exists and is measurable. By definition, we infer $\varphi(X)=Y$.

For $Y\in\mathcal{M}(\Omega,\sigma(X))$ write $Y=Y^+-Y_-$ with $Y_+,Y_-\in\mathcal{M}^+(\Omega,\sigma(X))$ and $Y_+=\varphi_+(X)$, $Y_-=\varphi_-(X)$ with measurable φ_+,φ_-. Setting $\varphi(x)=(\varphi_+(x)-\varphi_-(x))\mathbf{1}(\varphi_+(x)<\infty,\varphi_-(x)<\infty)$, we check that φ is measurable and satisfies $\varphi(X)=Y$. □

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 $L^2(\Omega,\mathscr{F},\mathbb{P})$.*

Proof. By definition, we have $\mathscr{L}^2(\Omega,\mathscr{G},\mathbb{P})\subseteq\mathscr{L}^2(\Omega,\mathscr{F},\mathbb{P})$. For $f,g\in\mathscr{L}^2(\Omega,\mathscr{G},\mathbb{P})$ with $f=g$ $\mathbb{P}$-a.s. (i.e., $[f]_{\mathscr{G}}=[g]_{\mathscr{G}}$) we also have $f,g\in\mathscr{L}^2(\Omega,\mathscr{F},\mathbb{P})$ with $f=g$ $\mathbb{P}$-a.s. (i.e., $[f]_{\mathscr{F}}=[g]_{\mathscr{F}}$) and the equivalence classes are embedded in a well defined way. Since $L^2(\Omega,\mathscr{G},\mathbb{P})$ is a complete linear space, its embedding is a complete linear subspace of $L^2(\Omega,\mathscr{F},\mathbb{P})$ and hence also closed. □

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. *For $Y\in L^2(\Omega,\mathscr{F},\mathbb{P})$ and a sub-σ-algebra $\mathscr{G}\subseteq\mathscr{F}$ the conditional expectation satisfies:*

(a) $\mathbb{E}[Y\,|\,\mathscr{G}]\in L^2(\Omega,\mathscr{F},\mathbb{P})$ *has a $\mathscr{G}$-measurable version: there is* $Z\in\mathscr{L}^2(\Omega,\mathscr{G},\mathbb{P})$ *with* $\mathbb{E}[Y\,|\,\mathscr{G}]=Z$ *$\mathbb{P}$-a.s.*

(b) $\mathbb{E}[Y\,|\,\mathscr{G}]=\operatorname{argmin}_{Z\in L^2(\Omega,\mathscr{G},\mathbb{P})}\mathbb{E}[(Y-Z)^2]$;

(c) $\forall\,\alpha\in\mathbb{R},\ Z\in L^2(\Omega,\mathscr{F},\mathbb{P}):\ \mathbb{E}[\alpha Y+Z\,|\,\mathscr{G}]=\alpha\,\mathbb{E}[Y\,|\,\mathscr{G}]+\mathbb{E}[Z\,|\,\mathscr{G}]$ *a.s.;*

(d) $\forall\,Z\in L^2(\Omega,\mathscr{G},\mathbb{P}):\ \mathbb{E}[\mathbb{E}[Y\,|\,\mathscr{G}]Z]=\mathbb{E}[YZ]$;

(e) $Y \geqslant 0$ $\mathbb{P}$*-a.s. implies* $\mathbb{E}[Y \,|\, \mathscr{G}] \geqslant 0$ $\mathbb{P}$*-a.s.*

Proof. Parts (a) and (b) follow immediately from the definition as an orthogonal projection onto the embedding of $L^2(\Omega, \mathscr{G}, \mathbb{P})$ into $L^2(\Omega, \mathscr{F}, \mathbb{P})$. The linearity of the projection implies (c), the orthogonality implies (d). For (e) let $G = \{Z < 0\}$ for a $\mathscr{G}$-measurable version of $\mathbb{E}[Y \,|\, \mathscr{G}]$. Then $G \in \mathscr{G}$ and by (d), $Y \geqslant 0$

$$\mathbb{E}[Z\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G] \geqslant 0.$$

This shows $Z \geqslant 0$ $\mathbb{P}$-a.s. and thus $\mathbb{E}[Y \,|\, \mathscr{G}] \geqslant 0$ $\mathbb{P}$-a.s. □

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

(a) $\mathbb{E}[Y \,|\, \mathscr{G}]$ *has a* $\mathscr{G}$*-measurable version;*

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

Proof. By the previous lemma (a) and (b) hold for $\mathbb{E}[Y \,|\, \mathscr{G}]$, noting $Z = \mathbf{1}_G \in L^2(\Omega, \mathscr{G}, \mathbb{P})$ for $G \in \mathscr{G}$.

Now suppose $Z \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ satisfies (a) and (b). Then using (b) also for $\mathbb{E}[Y \,|\, \mathscr{G}]$ gives

$$\mathbb{E}[(Z - \mathbb{E}[Y \,|\, \mathscr{G}])\mathbf{1}_G] = 0 \text{ for all } G \in \mathscr{G}.$$

Consider the events $G^> = \{Z > \mathbb{E}[Y \,|\, \mathscr{G}]\}$ and $G^< = \{Z < \mathbb{E}[Y \,|\, \mathscr{G}]\}$, which are in $\mathscr{G}$ if $\mathscr{G}$-measurable versions according to (a) are taken. This implies

$$\mathbb{E}[|Z - \mathbb{E}[Y \,|\, \mathscr{G}]|\mathbf{1}_{G^> \cup G^<}] = \mathbb{E}[(Z - \mathbb{E}[Y \,|\, \mathscr{G}])\mathbf{1}_{G^>}] - \mathbb{E}[(Z - \mathbb{E}[Y \,|\, \mathscr{G}])\mathbf{1}_{G^<}] = 0.$$

We deduce $|Z - \mathbb{E}[Y \,|\, \mathscr{G}]| = 0$ $\mathbb{P}$-a.s., that is $Z = \mathbb{E}[Y \,|\, \mathscr{G}]$ $\mathbb{P}$-a.s. □

3.11 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*

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

Proof. First, consider $Y \in \mathcal{M}^+(\Omega, \mathscr{F})$. Then $Y_n := \min(Y, n) \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ and $Y_n \uparrow Y$. For Y_n the conditional expectation is well defined and by monotonicity $\mathbb{E}[Y_{n+1} \,|\, \mathscr{G}] \geqslant \mathbb{E}[Y_n \,|\, \mathscr{G}]$, $n \geqslant 1$, holds $\mathbb{P}$-a.s. Then also the limit $\mathbb{E}[Y \,|\, \mathscr{G}] := \lim_{n\to\infty} \mathbb{E}[Y_n \,|\, \mathscr{G}]$ exists $\mathbb{P}$-a.s. and has a $\mathscr{G}$-measurable version. Monotone convergence implies for $G \in \mathscr{G}$:

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y \,|\, \mathscr{G}]\mathbf{1}_G] &= \mathbb{E}\Big[\lim_{n\to\infty} \mathbb{E}[Y_n \,|\, \mathscr{G}]\mathbf{1}_G\Big] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[Y_n \,|\, \mathscr{G}]\mathbf{1}_G] \\ &= \lim_{n\to\infty} \mathbb{E}[Y_n\mathbf{1}_G] = \mathbb{E}\Big[\lim_{n\to\infty} Y_n\mathbf{1}_G\Big] = \mathbb{E}[Y\mathbf{1}_G].\end{aligned}$$

For $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ write $Y = Y^+ - Y^-$ with $Y^+, Y^- \in \mathcal{M}^+(\Omega, \mathscr{F})$ and set

$$\mathbb{E}[Y \,|\, \mathscr{G}] := (\mathbb{E}[Y^+ \,|\, \mathscr{G}] - \mathbb{E}[Y^- \,|\, \mathscr{G}])\mathbf{1}(\mathbb{E}[Y^+ \,|\, \mathscr{G}] < \infty, \mathbb{E}[Y^- \,|\, \mathscr{G}] < \infty).$$

It is straightforward to check that the asserted properties are satisfied. □

3.12 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, $i \in I$.

3.13 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.*

Proof.

(a) $\Omega \in \mathscr{G}$ implies $\mathbb{E}[\mathbb{E}[Y \,|\, \mathscr{G}]\mathbf{1}_\Omega] = \mathbb{E}[Y\mathbf{1}_\Omega]$ and thus the claim.

(b) Y satisfies the defining properties of $\mathbb{E}[Y \,|\, \mathscr{G}]$.

(c) The right-hand side satisfies the defining properties of $\mathbb{E}[\alpha Y + Z \,|\, \mathscr{G}]$, using linearity for the expectation.

(d) Let $G = \{\mathbb{E}[Y \,|\, \mathscr{G}] < 0\} \in \mathscr{G}$ for a $\mathscr{G}$-measurable version of the conditional expectation. $Y \geqslant 0$ a.s. therefore implies

$$\mathbb{E}[\mathbb{E}[Y \,|\, \mathscr{G}]\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G] \geqslant 0 \Rightarrow \mathbb{P}(G) = 0.$$

(e) Using (c), (d), we infer $\mathbb{E}[Y_n \,|\, \mathscr{G}] \uparrow U$ for some $U \in \mathcal{M}^+(\Omega, \mathscr{G})$. Monotone convergence for expectations shows

$$\forall G \in \mathscr{G}: \ \mathbb{E}[U\mathbf{1}_G] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[Y_n \,|\, \mathscr{G}]\mathbf{1}_G] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbf{1}_G] = \mathbb{E}[Y\mathbf{1}_G].$$

We conclude $U = \mathbb{E}[Y \,|\, \mathscr{G}]$.

(f) ▶ Exercise

(g) ▶ Exercise

(h) The left-hand side satisfies the defining properties of $\mathbb{E}[Y \,|\, \mathscr{H}]$. Remark also that for $Y \in L^2$ this is just the composition of orthogonal projections.

(i) For $Z = \mathbf{1}_{G'}$, $G' \in \mathscr{G}$, the right-hand side satisfies the defining properties of $\mathbb{E}[ZY \,|\, \mathscr{G}]$. Use measure-theoretic induction to extend the results to simple $\mathscr{G}$-measurable functions Z, to $Y, Z \in \mathscr{M}^+(\Omega, \mathscr{G})$ and finally to $Y, Z \in \mathcal{M}(\Omega, \mathscr{G})$ with $ZY \in L^1$.

(j) By independence of Y and $\mathbf{1}_G$ for $G \in \mathscr{G}$ and by Fubini's theorem we see that $\mathbb{E}[Y]$ satisfies the defining properties of $\mathbb{E}[Y \,|\, \mathscr{G}]$.

□

3.14 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 $\mathscr{F}$.*

Proof. ▶ Exercise □

3.15 Definition. For a sub-σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ define the conditional probability of $A \in \mathscr{F}$ given $\mathscr{G}$ as

$$\mathbb{P}(A \,|\, \mathscr{G}) = \mathbb{E}[\mathbf{1}_A \,|\, \mathscr{G}].$$

3.16 Remark. $\mathbb{P}(A \,|\, \mathscr{G})$ is only $\mathbb{P}$-a.s. defined. For fixed pairwise disjoint $A_n \in \mathscr{F}$, $n \geqslant 1$, we have the σ-additivity $\mathbb{P}(\bigcup_{n \geqslant 1} A_n \,|\, \mathscr{G}) = \sum_{n \geqslant 1} \mathbb{P}(A_n \,|\, \mathscr{G})$ $\mathbb{P}$-a.s., but there might be no version such that $A \mapsto \mathbb{P}(A \,|\, \mathscr{G})$ is indeed a probability measure on all of $\mathscr{F}$. For Polish spaces such a version, a so-called regular conditional probability or Markov kernel, always exists, cf. Klenke. In the case of densities, the conditional density gives a constructive way to define conditional probabilities, see exercises.

3.17 Lemma. *For a sub-σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ and $A \in \mathscr{F}$, $B \in \mathscr{G}$ we have*

$$\mathbb{P}(A \cap B) = \int_B \mathbb{P}(A \,|\, \mathscr{G}) \, dP.$$

In the case $\mathbb{P}(A) > 0$ this implies the following general Bayes formula

$$\mathbb{P}(B \,|\, A) = \frac{\int_B \mathbb{P}(A \,|\, \mathscr{G}) \, d\mathbb{P}}{\int_\Omega \mathbb{P}(A \,|\, \mathscr{G}) \, d\mathbb{P}}.$$

Proof. ▶ Exercise □

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 Remark. For martingales (M_n) the expectation is constant: $\mathbb{E}[M_n] = \mathbb{E}[\mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}]] = \mathbb{E}[M_{n-1}] = \cdots = \mathbb{E}[M_0]$. Similarly, for submartingales the expectation is increasing and for supermartingales the expectation is decreasing.

Suppose (M_n) is even an L^2-martingale (i.e., $M_n \in L^2$ for all $n \geqslant 0$). Then the martingale differences $\Delta_n M := M_{n+1} - M_n$, $\Delta_m M := M_{m+1} - M_m$ for $m < n$ are uncorrelated:

$$\begin{aligned}\mathbb{E}[\Delta_m M \Delta_n M] &= \mathbb{E}\Big[\mathbb{E}[\Delta_m M \Delta_n M \,|\, \mathscr{F}_{m+1}]\Big] = \mathbb{E}\Big[\Delta_m M\, \mathbb{E}[\Delta_n M \,|\, \mathscr{F}_{m+1}]\Big] \\ &= \mathbb{E}\Big[\Delta_m M\, \mathbb{E}\big[\mathbb{E}[\Delta_n M \,|\, \mathscr{F}_n] \,\big|\, \mathscr{F}_{m+1}\big]\Big] = 0\end{aligned}$$

since $\mathscr{F}_{m+1} \subseteq \mathscr{F}_n$ and $\mathbb{E}[\Delta_n M \,|\, \mathscr{F}_n] = 0$ by the martingale property.

4.4 Example.

(a) Let $(X_k)_{k\geqslant 1}$ be independent random variables with $X_k \in L^1$, $\mathbb{E}[X_k] = 0$, $k \geqslant 1$. Put $S_0 = 0$, $S_n = \sum_{k=1}^n X_k$. For $n \geqslant 1$ we have $\mathscr{F}_n^S = \sigma(X_1, \dots, X_n)$ and $(S_n)_{n\geqslant 0}$ is a martingale with respect to its natural filtration. Similarly, $(S_n)_{n\geqslant 0}$ is a submartingale if $\mathbb{E}[X_k] \geqslant 0$ for all $k \geqslant 1$ and a supermartingale if $\mathbb{E}[X_k] \leqslant 0$ for all $k \geqslant 1$.

(b) Let $(X_k)_{k\geqslant 1}$ be independent random variables with $X_k \in L^1$, $\mathbb{E}[X_k] = 1$, $k \geqslant 1$. Put $P_0 = 1$, $P_n = \prod_{k=1}^n X_k$. For $n \geqslant 1$ we have $\mathscr{F}_n^P \subseteq \sigma(X_1, \dots, X_n)$ and $(P_n)_{n\geqslant 0}$ is a martingale with respect to its natural filtration. In Stochastics I we have seen that for $\mathbb{P}(X_k = 3/2) = \mathbb{P}(X_k = 1/2) = 1/2$ we have $P_n \to 0$ $\mathbb{P}$-a.s. although $P_n \geqslant 0$ and $\mathbb{E}[P_n] = 1$ for all $n \geqslant 0$. This can be interpreted as a fair game with initial capital $P_0 = 1$ where the stake in each round is half of the capital.

(c) Let $X \in L^1$ and $(\mathscr{F}_n)$ be any filtration, then $M_n := \mathbb{E}[X \,|\, \mathscr{F}_n]$ defines a martingale with respect to $(\mathscr{F}_n)$.

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

4.6 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 $(X{\bullet}M)_0 := 0$, $(X{\bullet}M)_n := \sum_{k=1}^n X_k(M_k - M_{k-1})$.

4.7 Remark. Any predictable process is adapted.

4.8 Lemma. *For a bounded (i.e., $\|X_n\|_\infty < \infty$) predictable process (X_n) and a martingale (M_n) (or just predictable (X_n) and $X_n \in L^p, M_n \in L^q$ with $\frac{1}{p}+\frac{1}{q} = 1$ for all n) $((X{\bullet}M)_n)_{n\geqslant 0}$ is again a martingale.*

Proof. If (X_n) is bounded, we have

$$\mathbb{E}[|(X{\bullet}M)_n|] \leqslant \sum_{k=1}^n \mathbb{E}[|X_k||M_k - M_{k-1}|] \leqslant \sum_{k=1}^n \|X_k\|_\infty (\|M_k\|_{L^1} + \|M_{k-1}\|_{L^1}) < \infty$$

and thus $(X{\bullet}M)_n \in L^1$. In the case $X_n \in L^p, M_n \in L^q$ conclude $(X{\bullet}M)_n \in L^1$ via Hölder inequality.

By definition and adaptedness of X, M it follows that $(X{\bullet}M)$ is adapted. By predictability of X and martingale property of M we conclude

$$\begin{aligned}\mathbb{E}[(X{\bullet}M)_{n+1} - (X{\bullet}M)_n \,|\, \mathscr{F}_n] &= \mathbb{E}[X_{n+1}(M_{n+1} - M_n) \,|\, \mathscr{F}_n] \\ &= X_{n+1}\, \mathbb{E}[M_{n+1} - M_n \,|\, \mathscr{F}_n] = 0.\end{aligned}$$

□

4.9 Remark. Interpreting the martingale as a fair game in the sense that $M_n - M_{n-1}$ is the payoff in round n for a stake of 1 Euro, the result can be interpreted as saying that under a predictable *investment* strategy of X_n Euros in round n a fair game remains fair. Obviously, this need not be the case if X_n is not predictable, knowing future outcomes of M.

A look at the proof shows also that for a submartingale M and a bounded, predictable X with $X_n \geqslant 0$ a.s., $n \geqslant 1$, the process $(X{\bullet}M)$ is again a submartingale.

4.10 Lemma. *Let (M_n) be a (sub-)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).*

Proof. By Jensen's inequality $\mathbb{E}[\varphi(M_{n+1}) \,|\, \mathscr{F}_n] \geqslant \varphi(\mathbb{E}[M_{n+1} \,|\, \mathscr{F}_n]) \geqslant \varphi(M_n)$ holds $\mathbb{P}$-a.s. □

4.11 Theorem (Doob decomposition)**.** *Given a submartingale (X_n), there exists a martingale (M_n) and a predictable increasing (i.e., $A_{n+1} \geqslant A_n$ a.s.) 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 $A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} \,|\, \mathscr{F}_{k-1}]$.

Proof. The process $A_n := \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} \,|\, \mathscr{F}_{k-1}]$, $n \geqslant 1$, is by definition predictable, in L^1 and increasing (for this use that X is a submartingale). Now define $M_n := X_n - X_0 - A_n$, $n \geqslant 1$, $M_0 := 0$. Then M_n is in L^1 and adapted since X_n, A_n are so. Moreover, $\mathbb{E}[M_{n+1} - M_n \,|\, \mathscr{F}_n] = \mathbb{E}[(X_{n+1} - X_n) - (A_{n+1} - A_n) \,|\, \mathscr{F}_n] = 0$ holds and A, M give a Doob decomposition of X.

To prove uniqueness, suppose $X_n = X_0 + M'_n + A'_n$ is another Doob decomposition of X. Then $M_n - M'_n = A'_n - A_n$, $n \geqslant 1$, holds as well as $M_0 - M'_0 = A'_0 - A_0 = 0$. This shows that $(M_n - M'_n)_{n\geqslant 0}$ is a predictable martingale starting in zero. Yet, a predictable martingale is easily shown to be a.s. constant and thus $M_n - M'_n = 0$ a.s. This shows $M_n = M'_n$, $A_n = A'_n$ a.s. □

4.12 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 $\langle M\rangle_n$.

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

Proof. Using the definition and martingale property of M we conclude

$$\begin{aligned}\langle M\rangle_{n+1} - \langle M\rangle_n &= \mathbb{E}[M_{n+1}^2 - M_n^2 \,|\, \mathscr{F}_n]\\ &= \mathbb{E}[(M_{n+1} - M_n)^2 + 2M_n(M_{n+1} - M_n) \,|\, \mathscr{F}_n]\\ &= \mathbb{E}[(M_{n+1} - M_n)^2 \,|\, \mathscr{F}_n].\end{aligned}$$

With $\langle M\rangle_0 = 0$ the claim follows. □

4.14 Example.

(a) Let $(X_k)_{k\geqslant 1}$ be independent random variables with $X_k \in L^2$, $\mathbb{E}[X_k] = 0$, $k \geqslant 1$. Put $S_0 = 0$, $S_n = \sum_{k=1}^n X_k$. Then $(S_n)_{n\geqslant 0}$ is an L^2-martingale with respect to its natural filtration with quadratic variation

$$\langle S\rangle_n = \sum_{k=1}^n \mathbb{E}[X_k^2 \,|\, \mathscr{F}_{k-1}] = \sum_{k=1}^n \mathbb{E}[X_k^2] = \mathrm{Var}(S_n).$$

In particular, the quadratic variation of S is deterministic because S has independent increments.

(b) For an L^2-martingale M and predictable, bounded X the martingale transform $X{\bullet}M$ can be checked to be an L^2-martingale. Its quadratic variation is

$$\begin{aligned}\langle X{\bullet}M\rangle_n &= \sum_{k=1}^n \mathbb{E}[X_k^2(M_k - M_{k-1})^2 \,|\, \mathscr{F}_{k-1}]\\ &= \sum_{k=1}^n X_k^2(\langle M\rangle_k - \langle M\rangle_{k-1}) = (X^2{\bullet}\langle M\rangle)_n.\end{aligned}$$

So, the quadratic variation of $X{\bullet}M$ is again represented as a discrete stochastic integral with 'integrand' X^2 and 'integrator' $\langle M\rangle$.

4.2 Stopping times

4.15 Definition. A map $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called stopping time (Stoppzeit) with respect to a filtration $(\mathscr{F}_n)$ if $\{\tau = n\} \in \mathscr{F}_n$ holds for all $n \geqslant 0$.

4.16 Lemma.

(a) *A stopping time is an* $([0,\infty], \mathfrak{B}_{[0,\infty]})$*-valued random variable.*

(b) *A map* $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ *is a stopping time if and only if* $\{\tau \leqslant n\} \in \mathscr{F}_n$ *for all* $n \geqslant 0$.

(c) *Every deterministic time* $\tau(\omega) = n_0$ *is a stopping time.*

(d) *For stopping times* σ *and* τ *also* $\sigma \wedge \tau$, $\sigma \vee \tau$ *and* $\sigma + \tau$ *are stopping times.*

Proof.

(a) For $m \in \mathbb{N}_0$ we have $\tau^{-1}(\{m\}) = \{\tau = m\} \in \mathscr{F}_m \subseteq \mathscr{F}$ and thus also $\tau^{-1}(\{\infty\}) = \tau^{-1}(\mathbb{N}_0)^{\complement} \in \mathscr{F}$. For any $B \in \mathfrak{B}_{[0,\infty]}$ we conclude $\tau^{-1}(B) = \bigcup_{m \in B \cap (\mathbb{N}_0 \cup \{\infty\})} \tau^{-1}(\{m\}) \in \mathscr{F}$.

(b) If $\{\tau = k\} \in \mathscr{F}_k$ for all $k \geqslant 0$, then $\{\tau \leqslant n\} = \bigcup_{k=0}^n \{\tau = k\} \in \mathscr{F}_n$ for all $n \geqslant 0$. Conversely, if $\{\tau \leqslant k\} \in \mathscr{F}_k$ for all $k \geqslant 0$, then $\{\tau = n\} = \{\tau \leqslant n\} \setminus \{\tau \leqslant n-1\} \in \mathscr{F}_n$ for all $n \geqslant 1$ and $\{\tau \leqslant 0\} = \{\tau = 0\}$ trivially.

(c) We have $\{\tau = n\} = \varnothing \in \mathscr{F}_n$ for all $n \neq n_0$ and also $\{\tau = n_0\} = \Omega \in \mathscr{F}_{n_0}$.

(d) Use part (b) to see $\{\sigma \wedge \tau \leqslant n\} = \{\sigma \leqslant n\} \cup \{\tau \leqslant n\} \in \mathscr{F}_n$ and $\{\sigma \vee \tau \leqslant n\} = \{\sigma \leqslant n\} \cap \{\tau \leqslant n\} \in \mathscr{F}_n$. Moreover, $\{\sigma + \tau = n\} = \bigcup_{k=0}^n \{\sigma = k\} \cap \{\tau = n-k\} \in \mathscr{F}_n$ holds.

□

4.17 Example. Let $(X_n)_{n \geqslant 0}$ be an $(\mathscr{F}_n)$-adapted real-valued process and $B \in \mathfrak{B}_{\mathbb{R}}$. Then the entrance time into B

$$\tau_B := \inf\{n \geqslant 0 \,|\, X_n \in B\} \text{ with } \inf \varnothing := \infty$$

is an $(\mathscr{F}_n)$-stopping time: $\{\tau_B \leqslant n\} = \bigcup_{k=0}^n \{X_k \in B\} \in \mathscr{F}_n$. For $k \in \mathbb{N}$ also $\tau_{B,k} = \inf\{n \geqslant k \,|\, X_{n-k} \in B\} = \tau_B + k$ is a stopping time, but usually $\tau_{B,-k} = \inf\{n \geqslant 0 \,|\, X_{n+k} \in B\}$ is *not* a stopping time.

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

Proof. Note first $\mathbb{E}[|M_{\tau \wedge n}|] \leqslant \mathbb{E}[\sum_{k=0}^n |M_k|] < \infty$. Put $C_n := \mathbf{1}(\tau \geqslant n)$, $n \geqslant 1$. Then (C_n) is a bounded predictable process, noting $\{\tau \geqslant n\} = \{\tau \leqslant n-1\}^{\complement} \in \mathscr{F}_{n-1}$, and satisfies

$$(C \bullet M)_n = \sum_{k=1}^n C_k (M_k - M_{k-1}) = \sum_{k=1}^{\tau \wedge n} (M_k - M_{k-1}) = M_{\tau \wedge n} - M_0.$$

If M is a martingale, then the result on the martingale transform shows that $(C \bullet M)$ is a martingale and thus also M^τ. Since $C_n \geqslant 0$, we can deduce in the submartingale case that $C \bullet M$ is also a submartingale and so is M^τ. For a supermartingale M consider the submartingale $-M$ and conclude from that. □

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

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

Proof. The axioms of a σ-algebra are checked directly. For all $n \geqslant 0$ the event $\{\tau = m\} \cap \{\tau \leqslant n\}$ is either empty (if $n < m$) or equal to $\{\tau = m\} \in \mathscr{F}_m \subseteq \mathscr{F}_n$ (if $n \geqslant m$). This shows $\{\tau = m\} \in \mathscr{F}_\tau$ and as above for $\mathscr{F}$ we conclude that τ is $\mathscr{F}_\tau$-measurable. □

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

Proof. For $A \in \mathscr{F}_\sigma$ we have $A \cap \{\sigma \leqslant n\} \in \mathscr{F}_n$ for all $n \geqslant 0$. Since $\{\tau \leqslant n\} \in \mathscr{F}_n$ holds by the stopping time property, we conclude from $\sigma \leqslant \tau$ that

$$A \cap \{\tau \leqslant n\} = A \cap \{\sigma \leqslant n\} \cap \{\tau \leqslant n\} \in \mathscr{F}_n.$$

Hence, $A \in \mathscr{F}_\tau$. □

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

Proof. For any $B \in \mathfrak{B}_\mathbb{R}$ we have to check $\{\omega \in \Omega \mid X_{\tau(\omega)}(\omega) \in B\} \in \mathscr{F}_\tau$, which is equivalent to

$$\forall n \geqslant 0 : \; \{X_\tau \in B\} \cap \{\tau \leqslant n\} \in \mathscr{F}_n.$$

Writing $\{X_\tau \in B\} \cap \{\tau \leqslant n\} = \bigcup_{k=0}^n \{X_\tau \in B, \tau = k\} = \bigcup_{k=0}^n \{X_k \in B\} \cap \{\tau = k\}$ we see that $\{X_k \in B\}, \{\tau = k\} \in \mathscr{F}_k \subseteq \mathscr{F}_n$ for $k \leqslant n$ indeed implies $\{X_\tau \in B\} \in \mathscr{F}_\tau$. □

4.23 Theorem (Optional Sampling)**.** *Let (M_n) be a martingale (submartingale) and σ, τ bounded stopping times with $\sigma \leqslant \tau$ (i.e., $\exists C \in \mathbb{N} \, \forall \omega : \sigma(\omega) \leqslant \tau(\omega) \leqslant C$). Then $\mathbb{E}[M_\tau \mid \mathscr{F}_\sigma] = M_\sigma$ (resp. $\mathbb{E}[M_\tau \mid \mathscr{F}_\sigma] \geqslant M_\sigma$) holds.*

Proof. (martingale case) Because of $\tau \leqslant C$ we see from above $M_\tau = M_{\tau \wedge C} \in L^1$. We have to show $\mathbb{E}[M_\sigma \mathbf{1}_A] = \mathbb{E}[M_\tau \mathbf{1}_A]$ for all $A \in \mathscr{F}_\sigma$. Putting $\rho := \sigma \mathbf{1}_A + \tau \mathbf{1}_{A^\complement}$, we have

$$\{\rho = n\} = (A \cap \{\sigma = n\}) \cup (A^\complement \cap \{\tau = n\}) \in \mathscr{F}_n,$$

using $A^\complement \in \mathscr{F}_\sigma \subseteq \mathscr{F}_\tau$. Hence, ρ is a bounded stopping time as well and $\mathbb{E}[M_\rho] = \mathbb{E}[M_0] = \mathbb{E}[M_\tau]$ implies $\mathbb{E}[M_\rho - M_\tau] = 0$. By definition of ρ, we have $M_\rho - M_\tau = (M_\sigma - M_\tau)\mathbf{1}_A$ and $\mathbb{E}[M_\sigma \mathbf{1}_A] = \mathbb{E}[M_\tau \mathbf{1}_A]$ follows. □

4.24 Proposition. *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 dominated ($|M_{\tau\wedge n}|\leqslant Y$ for all n and some $Y\in L^1$);*

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

Proof.

(a) There is some $C\in\mathbb{N}$ with $\tau\leqslant C$. Optional stopping therefore yields $\mathbb{E}[M_\tau]=\mathbb{E}[M_{\tau\wedge C}]=\mathbb{E}[M_0]$.

(b) We have $M_{\tau\wedge n}\to M_\tau$ as $n\to\infty$ and dominated convergence together with (a) implies

$$\mathbb{E}[M_\tau]=\lim_{n\to\infty}\mathbb{E}[M_{\tau\wedge n}]=\lim_{n\to\infty}\mathbb{E}[M_0]=\mathbb{E}[M_0].$$

(c) We have $|M_{\tau\wedge n}-M_0|\leqslant\sum_{k=1}^{\infty}|M_k-M_{k-1}|\mathbf{1}(k\leqslant\tau)=:Z$. Using $\mathbb{E}[|M_k-M_{k-1}|\mathbf{1}(k\leqslant\tau)\,|\,\mathscr{F}_{k-1}]\leqslant C\mathbf{1}(k\leqslant\tau)$ by assumption and $\mathscr{F}_{k-1}$-measurability of $\{k\leqslant\tau\}$, we obtain

$$\begin{aligned}\mathbb{E}[Z]&=\sum_{k=1}^{\infty}\mathbb{E}\Big[\mathbb{E}\big[|M_k-M_{k-1}|\mathbf{1}(k\leqslant\tau)\,\big|\,\mathscr{F}_{k-1}\big]\Big]\\&\leqslant\sum_{k=1}^{\infty}C\,\mathbb{P}(k\leqslant\tau)=C\,\mathbb{E}[\tau]<\infty.\end{aligned}$$

With $Y=Z+|M_0|\in L^1$ we thus have $|M_{\tau\wedge n}|\leqslant Y$ for all n and applying (b) yields the assertion.

□

4.25 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 $\mathbb{E}[S_\tau]=\mu\,\mathbb{E}[\tau]$.*

4.26 Remark. If additionally all X_k are in L^2 and $\mathbb{E}[X_k]=0$, $\mathrm{Var}(X_k)=\sigma^2$, $k\geqslant 1$, holds, then also the second Wald identity is valid: $\mathrm{Var}(S_\tau)=\mathbb{E}[\tau]\sigma^2$; see Bauer, Satz 17.7.

Proof. $M_n=S_n-n\mu$, $n\geqslant 0$, forms an $(\mathscr{F}_n)$-martingale due to $\mathbb{E}[M_{n+1}-M_n\,|\,\mathscr{F}_n]=\mathbb{E}[X_{n+1}-\mu]=0$. Moreover,

$$\mathbb{E}[|M_{n+1}-M_n|\,|\,\mathscr{F}_n]=\mathbb{E}[|X_{n+1}-\mu|]\leqslant\sup_{k\geqslant 1}\mathbb{E}[|X_k|]+\mu$$

holds such that by part (c) of the preceding proposition $\mathbb{E}[M_\tau]=0$. This gives $\mathbb{E}[S_\tau]=\mu\,\mathbb{E}[\tau]$. □

4.27 Example.

(a) Let (X_k) be i.i.d., $X_k \in L^1$, and τ be an $(\mathscr{F}_n^X)$-stopping time. Then Wald's identity applies and gives $\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\,\mathbb{E}[X_1]$. In the case $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = 1/2$ (*toin cossing game*) one can show that $\tau = \inf\{n \geqslant 0 \,|\, S_n = 1\}$ is an almost surely finite stopping time (compare Stochastics I). Then $S_\tau = 1$ a.s., but $\mathbb{E}[X_1] = 0$ such that we conclude $\mathbb{E}[\tau] = \infty$.

(b) Let (X_k) be i.i.d., $X_k \in L^1$, and τ be a random time independent of (X_k). Then for the filtration $\mathscr{F}_n = \sigma(\tau, X_1, \dots, X_n)$ τ is an $(\mathscr{F}_n)$-stopping time, Wald's identity applies and gives $\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\,\mathbb{E}[X_1]$ (cf. the compound Poisson case). This is an example of a filtration which is not natural for X.

4.28 Example (random walk)**.** Let $(X_k)_{k\geqslant 1}$ be independent with $\mathbb{P}(X_k = 1) = p$, $\mathbb{P}(X_k = -1) = q = 1 - p$ for $p \in (0,1)$ and define $S_0 = 0$, $S_n = \sum_{k=1}^n X_k$, $n \geqslant 1$. Then $(S_n)_{n\geqslant 0}$ defines a simple random walk. Let $a < 0 < b$ with $a, b \in \mathbb{Z}$ be given and consider the $(\mathscr{F}_n^X)$-stopping time

$$\tau = \inf\{n \geqslant 0 \,|\, S_n \in \{a, b\}\}.$$

Let $R = |a| + b$ and observe that $(S_{mR} - S_{(m-1)R})_{m\geqslant 1}$ are i.i.d. and $\mathbb{P}(S_{mR} - S_{(m-1)R} = R) > 0$. Hence, the time $\sigma = \inf\{m \geqslant 1 \,|\, S_{mR} - S_{(m-1)R} = R\}$ is geometrically distributed and has finite expectation. From $\tau \leqslant R\sigma$ we conclude $\mathbb{E}[\tau] < \infty$.

Consider the case $p \neq 1/2$ first. Then $((q/p)^{S_n})_{n\geqslant 0}$ forms an $(\mathscr{F}_n^X)$-martingale due to

$$\mathbb{E}[(q/p)^{S_{n+1}} \,|\, \mathscr{F}_n^X] = (q/p)^{S_n}\,\mathbb{E}[(q/p)^{X_{n+1}}] = (q/p)^{S_n}(p(q/p) + q(p/q)) = (q/p)^{S_n}.$$

Because of $|S_{\tau\wedge n}| \leqslant |a| \vee b$ the $S_{\tau\wedge n}$, $n \geqslant 1$, are dominated and by part (b) of the above proposition we infer $\mathbb{E}[(q/p)^{S_\tau}] = \mathbb{E}[(q/p)^{S_0}] = 1$. This gives the two equations

$$\mathbb{P}(S_\tau = a)(q/p)^a + \mathbb{P}(S_\tau = b)(q/p)^b = 1, \quad \mathbb{P}(S_\tau = a) + \mathbb{P}(S_\tau = b) = 1.$$

We can thus calculate the probability whether S_n first hits a or b:

$$\mathbb{P}(S_\tau = a) = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}, \quad \mathbb{P}(S_\tau = b) = \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}.$$

Using $\mathbb{E}[\tau] < \infty$, Wald's identity gives $\mathbb{E}[S_\tau] = (p - q)\,\mathbb{E}[\tau]$ and we can solve for $\mathbb{E}[\tau]$:

$$\mathbb{E}[\tau] = \frac{a((q/p)^b - 1) + b(1 - (q/p)^a)}{(p - q)((q/p)^b - (q/p)^a)}.$$

As a special case note that for $p > q$ and $a \downarrow -\infty$ we find $\mathbb{E}[\tau] \uparrow \frac{b}{p-q}$. Using a monotone convergence argument, one can indeed show that this is the expectation for the one-sided stopping time $\inf\{n \geqslant 0 \,|\, X_n = b\}$.

For the symmetric simple random walk with $p = q = 1/2$ Wald's identity yields directly $\mathbb{E}[S_\tau] = 0$ and thus $\mathbb{P}(S_\tau = a) = \frac{b}{|a|+b}$, $\mathbb{P}(S_\tau = b) = \frac{|a|}{|a|+b}$. The second Wald identity then gives $\mathrm{Var}(S_\tau) = \mathbb{E}[\tau]$ and thus $\mathbb{E}[\tau] = |a|b$. Note that here $\mathbb{E}[\tau] \uparrow \infty$ as $a \downarrow -\infty$.

4.3 Martingale inequalities and convergence

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

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

Proof. Put $\tau := \inf\{n \geqslant 0 \,|\, |M_n| \geqslant \alpha\}$. Then τ is a stopping time with $\{\max_{0\leqslant k\leqslant n}|M_k| \geqslant \alpha\} = \{\tau \leqslant n\}$. We obtain:

$$\begin{aligned}\mathbb{P}\Big(\max_{0\leqslant k\leqslant n}|M_k| \geqslant \alpha\Big) &= \mathbb{E}[\mathbf{1}(\tau \leqslant n)] \leqslant \mathbb{E}\Big[\frac{|M_{\tau\wedge n}|}{\alpha}\mathbf{1}(\tau \leqslant n)\Big] \\ &\leqslant \frac{1}{\alpha}\,\mathbb{E}[|M_{\tau\wedge n}|] \leqslant \frac{1}{\alpha}\,\mathbb{E}[|M_n|],\end{aligned}$$

where the last bound follows by optional sampling of the submartingale $(|M_n|)$.
□

4.30 Remark. For submartingales (M_n) the same proof applied to M_n instead of $|M_n|$ gives

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

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

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

Proof. Write $M_n^* = \max_{1\leqslant k\leqslant n}|M_k|$. Then by partial integration (cf. Stochastik I), by the same bound for the stopping time $\tau = \inf\{n \geqslant 0 \,|\, |M_n| \geqslant x\}$ as in the maximal inequality and by Tonelli's theorem we obtain

$$\begin{aligned}\tfrac{1}{p}\,\mathbb{E}[(M_n^*)^p] &= \int_0^\infty x^{p-1}\,\mathbb{P}(M_n^* \geqslant x)\,dx \leqslant \int_0^\infty x^{p-2}\,\mathbb{E}[|M_{n\wedge\tau}|\mathbf{1}(\tau \leqslant x)]\,dx \\ &= \mathbb{E}\Big[|M_{n\wedge\tau}|\int_0^\infty x^{p-2}\mathbf{1}(M_n^* \geqslant x)\,dx\Big] = \tfrac{1}{p-1}\,\mathbb{E}[|M_{n\wedge\tau}|(M_n^*)^{p-1}].\end{aligned}$$

By Hölder inequality for $p^{-1} + q^{-1} = 1$ and optional sampling we thus have

$$\mathbb{E}[(M_n^*)^p] \leqslant \tfrac{p}{p-1}\,\mathbb{E}[|M_{n\wedge\tau}|^p]^{1/p}\,\mathbb{E}[(M_n^*)^p]^{1/q} \leqslant \tfrac{p}{p-1}\,\mathbb{E}[|M_n|^p]^{1/p}\,\mathbb{E}[(M_n^*)^p]^{(p-1)/p}.$$

Dividing by $\mathbb{E}[(M_n^*)^p]^{(p-1)/p}$, the assertion follows. □

4.32 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 $\tau_0 := 0$, $\sigma_{k+1} := \inf\{\ell \geqslant \tau_k \,|\, X_\ell \leqslant a\}$, $\tau_{k+1} := \inf\{\ell \geqslant \sigma_{k+1} \,|\, X_\ell \geqslant b\}$.

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

Proof. $Y_n = (X_n - a)_+$, $n \geqslant 0$, forms by Jensen's inequality a submartingale. Since the upcrossings by (Y_n) of $[0, b-a]$ equals $U_n^{[a,b]}$, we can and shall assume in the sequel $a = 0$ and $X_n \geqslant 0$ without loss of generality (w.l.o.g.).

By definition of the upcrossing stopping times a telescoping sum argument yields

$$\mathbb{E}[X_n] = \mathbb{E}[X_{\sigma_1 \wedge n}] + \sum_{k=1}^{n} \mathbb{E}[X_{\tau_k \wedge n} - X_{\sigma_k \wedge n}] + \sum_{k=1}^{n} \mathbb{E}[X_{\sigma_{k+1} \wedge n} - X_{\tau_k \wedge n}].$$

Since (X_n) is a non-negative submartingale by assumption, optional sampling shows that all summands in this decomposition are non-negative. From

$$\sum_{k=1}^{n} (X_{\tau_k \wedge n} - X_{\sigma_k \wedge n}) = \sum_{k=1}^{U_n^{[0,b]}} (X_{\tau_k} - X_{\sigma_k}) \geqslant b U_n^{[0,b]}$$

we thus infer $\mathbb{E}[X_n] \geqslant b\, \mathbb{E}[U_n^{[0,b]}]$, the upcrossing inequality for $a = 0$. □

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

4.35 Remark. If (M_n) is a submartingale, it is L^1-bounded already if $\sup_n \mathbb{E}[(M_n)_+]$ is finite because

$$\mathbb{E}[(M_n)_-] = \mathbb{E}[(M_n)_+] - \mathbb{E}[M_n] \leqslant \mathbb{E}[(M_n)_+] - \mathbb{E}[M_0]$$

holds. Let us also emphasize that (M_n) need not converge in L^1 to M_∞.

Proof in the submartingale case. By monotonicity $U_n^{[a,b]}$ converges to some $U^{[a,b]} \in \mathbb{N}_0 \cup \{\infty\}$ as $n \to \infty$ and by monotone convergence and the upcrossing inequality

$$\mathbb{E}[U^{[a,b]}] \leqslant \frac{1}{b-a} \lim_{n\to\infty} \mathbb{E}[(M_n - a)_+] \leqslant \frac{1}{b-a} \Big(\sup_n \mathbb{E}[|M_n|] + a \Big) < \infty.$$

This shows $\mathbb{P}(U^{[a,b]} = \infty) = 0$. For $a < b$ and the event

$$\Lambda_{a,b} = \Big\{ \limsup_{n\to\infty} M_n \geqslant b,\, \liminf_{n\to\infty} M_n \leqslant a \Big\}$$

this implies $\mathbb{P}(\Lambda_{a,b}) = 0$ and thus

$$\mathbb{P}\Big(\limsup_{n\to\infty} M_n > \liminf_{n\to\infty} M_n \Big) = \mathbb{P}\Big(\bigcup_{a<b,a,b\in\mathbb{Q}} \Lambda_{a,b} \Big) = 0.$$

Hence, (M_n) converges $\mathbb{P}$-almost surely to some M_∞ with values in $\mathbb{R} \cup \{\pm\infty\}$. Fatou's Lemma gives

$$\mathbb{E}[|M_\infty|] = \mathbb{E}\Big[\liminf_{n\to\infty} |M_n| \Big] \leqslant \liminf_{n\to\infty} \mathbb{E}[|M_n|] < \infty.$$

This shows that M_∞ is a.s. finite and in L^1. □

4.36 Corollary. *Each non-negative supermartingale* (M_n) *converges* $\mathbb{P}$*-a.s. to some* M_∞ *with* $\mathbb{E}[M_\infty] \leqslant \lim_{n\to\infty} \mathbb{E}[M_n] = \inf_n \mathbb{E}[M_n]$.

Proof. The decay of $n \mapsto \mathbb{E}[M_n]$ and $M_n \geqslant 0$ show $\sup_n \mathbb{E}[|M_n|] = \mathbb{E}[M_0] < \infty$. By the first martingale convergence theorem we obtain $M_n \to M_\infty$ $\mathbb{P}$-a.s. By Fatou's Lemma

$$\mathbb{E}[M_\infty] \leqslant \liminf_{n\to\infty} \mathbb{E}[M_n] = \lim_{n\to\infty} \mathbb{E}[M_n] = \inf_n \mathbb{E}[M_n]$$

follows. □

4.37 Example (A fair game where you lose in the long run)**.** Consider the multiplicative martingale $K_0 := 1$, $K_n = \prod_{i=1}^n R_i$, $n \geqslant 1$, with independent random variables (R_i), satisfying $\mathbb{P}(R_i = 3/2) = \mathbb{P}(R_i = 1/2) = 1/2$. Then the strong law of large numbers applied to $\frac{1}{n}\log(K_n)$ shows $K_n \to K_\infty = 0$ $\mathbb{P}$-a.s., although $\mathbb{E}[K_n] = 1$ for all $n \geqslant 0$.

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

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

4.39 Lemma.

(a) $(X_i)_{i\in I}$ *is uniformly integrable if and only if* $(X_i)_{i\in I}$ *is* L^1*-bounded and* $\forall \varepsilon > 0\, \exists \delta > 0 : \ \mathbb{P}(A) < \delta \Rightarrow \sup_{i\in I} \mathbb{E}[|X_i|\mathbf{1}_A] < \varepsilon$.

(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.*

Proof. For (a) see ► Exercise . For (b) note

$$\mathbb{E}[|X_i|\mathbf{1}_{\{|X_i|>R\}}] \leqslant \mathbb{E}[|X_i|(|X_i|/R)^{p-1}] = R^{-(p-1)}\|X_i\|_{L^p}^p$$

where the right-hand side tends to zero for $R \to \infty$, uniformly over i. Part (c) follows directly from $|X_i|\mathbf{1}_{\{|X_i|>R\}} \leqslant Y\mathbf{1}(Y > R)$ and dominated convergence. □

4.40 Theorem (Vitali)**.** *Let* $(X_n)_{n\geqslant 0}$ *be random variables in* L^1 *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) $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|] < \infty$.

Proof. To show (a)⇒(b) we can assume w.l.o.g. that $X_n \to X$ $\mathbb{P}$-a.s.: if (X_n) does not converge to X in L^1, then there is a subsequence (n_k) and $\varepsilon > 0$ such that $\|X_{n_k} - X\| \geqslant \varepsilon$ for all k and by Stochastik I a subsubsequence (n_{k_l}) such that $X_{n_{k_l}} \to X$ $\mathbb{P}$-a.s., for which, however, we now prove $X_{n_{k_l}} \to X$ in L^1.

Since (X_n) is L^1-bounded, Fatou's Lemma shows $\mathbb{E}[|X|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|X_n|] < \infty$. For $\varepsilon > 0$ choose by uniform integrability some $R > 0$ with

$$\sup_n \mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] + \mathbb{E}[|X|\mathbf{1}(|X| > R)] < \frac{\varepsilon}{2}.$$

Put $\varphi_R(x) = -R \vee (x \wedge R)$. By dominated convergence there is $n_0 \in \mathbb{N}$ with $\mathbb{E}[|\varphi_R(X_n) - \varphi_R(X)|] < \varepsilon/2$ for all $n \geqslant n_0$. Consequently,

$$\mathbb{E}[|X_n - X|] \leqslant \mathbb{E}[|\varphi_R(X_n) - \varphi_R(X)|] + \mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] + \mathbb{E}[|X|\mathbf{1}(|X| > R)] < \varepsilon$$

holds for all $n \geqslant n_0$. Since $\varepsilon > 0$ was arbitrary, (b) follows.

The implication (b)⇒(c) follows immediately by the continuity of the norm.

For (c)⇒(a) put $\psi_R(x) = |x|$ for $|x| \leqslant R - 1$, $\psi_R(x) = 0$ for $|x| \geqslant R$ and interpolate linearly on $[-R, -R+1]$ and $[R-1, R]$. Then ψ_R is continuous and satisfies $\psi_R(x) \leqslant |x|\mathbf{1}(|x| \leqslant R)$ such that for any n, R

$$\mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] \leqslant \mathbb{E}[|X_n|] - \mathbb{E}[\psi_R(X_n)].$$

Since ψ_R is bounded and continuous and $X_n \to X$ in distribution, we have $\mathbb{E}[\psi_R(X_n)] \to \mathbb{E}[\psi_R(X)]$. Letting first $n \to \infty$ and then $R \to \infty$ we thus have

$$\lim_{R\to\infty} \limsup_{n\to\infty} \mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] = 0.$$

By the decay of the expectation as $R \uparrow \infty$, there is an $n_0 \in \mathbb{N}$ such that $\sup_{n \geqslant n_0} \mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] \to 0$ as $R \to \infty$. Since $\mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] \to 0$ as $R \to \infty$ for each $n \leqslant n_0 - 1$, we also have $\max_{n \leqslant n_0 - 1} \mathbb{E}[|X_n|\mathbf{1}(|X_n| > R)] \to 0$ as $R \to \infty$ and uniform integrability follows. □

4.41 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 $M_n = \mathbb{E}[M_\infty \,|\, \mathscr{F}_n]$.*

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

Proof.

(a) The first martingale convergence theorem together with Vitali's theorem ensures L^1-convergence. Furthermore, for any $n \geqslant m \geqslant 1$ and $A \in \mathscr{F}_m$ we have $\mathbb{E}[M_m \mathbf{1}_A] = \mathbb{E}[M_n \mathbf{1}_A]$ and $\|M_n - M_\infty\|_{L^1} \to 0$ implies $\mathbb{E}[|M_n - M_\infty|\mathbf{1}_A] \to 0$. This shows $\mathbb{E}[M_\infty \mathbf{1}_A] = \mathbb{E}[M_m \mathbf{1}_A]$, hence $M_m = \mathbb{E}[M_\infty \,|\, \mathscr{F}_m]$.

(b) Assume first $M \geqslant 0$. Then $0 \leqslant M_n = \mathbb{E}[M \mid \mathscr{F}_n]$ holds a.s. From $\mathbb{E}[M_n] = \mathbb{E}[M] < \infty$ we conclude that $M_\infty = \lim_{n\to\infty} M_n$ exists a.s. (first martingale convergence theorem). Fatou's Lemma yields $\mathbb{E}[M_\infty] \leqslant \mathbb{E}[M_n] = \mathbb{E}[M]$. On the other hand, dominated convergence and Jensen's inequality applied to $x \mapsto (x-R)_+$ give for any $R > 0$

$$\begin{aligned}\mathbb{E}[M_\infty \wedge R] &= \lim_{n\to\infty} \mathbb{E}[M_n \wedge R] = \lim_{n\to\infty} (\mathbb{E}[M_n] - \mathbb{E}[(M_n - R)_+]) \\ &\geqslant \mathbb{E}[M] - \mathbb{E}[\mathbb{E}[(M-R)_+ \mid \mathscr{F}_n]] = \mathbb{E}[M \wedge R].\end{aligned}$$

This implies $\mathbb{E}[M_\infty] \geqslant \lim_{R\to\infty} \mathbb{E}[M \wedge R] = \mathbb{E}[M]$, whence $\mathbb{E}[M_\infty] = \mathbb{E}[M] = \mathbb{E}[M_n]$. By Vitali's Theorem, we infer that (M_n) is uniformly integrable and from part (a) that $M_n = \mathbb{E}[M_\infty \mid \mathscr{F}_n]$.

The limit M_∞ is $\mathscr{F}_\infty$-measurable and satisfies for any $A \in \mathscr{F}_n$, $n \geqslant 1$:

$$\mathbb{E}[M_\infty \mathbf{1}_A] = \mathbb{E}[M_n \mathbf{1}_A] = \mathbb{E}[\mathbb{E}[M \mid \mathscr{F}_n]\mathbf{1}_A] = \mathbb{E}[M\mathbf{1}_A].$$

Hence, the probability measures(!) $\mathbb{Q}_1(A) := \mathbb{E}[M_\infty \mathbf{1}_A]/\mathbb{E}[M]$ and $\mathbb{Q}_2(A) := \mathbb{E}[M\mathbf{1}_A]/\mathbb{E}[M]$, $A \in \mathscr{F}_\infty$, coincide on $\bigcup_{m\geqslant 1} \mathscr{F}_m$. As the latter is an $\cap$-stable generator of $\mathscr{F}_\infty$, $\mathbb{Q}_1$ and $\mathbb{Q}_2$ agree everywhere. By definition, this means $M_\infty = \mathbb{E}[M \mid \mathscr{F}_\infty]$ a.s. This gives the result for $M \geqslant 0$ and for general M consider $M_n^+ := \mathbb{E}[M^+ \mid \mathscr{F}_n]$, $M_n^- := \mathbb{E}[M^- \mid \mathscr{F}_n]$ separately.

□

4.42 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.*

Proof. ▶ Exercise □

4.43 Example. The random harmonic sum $S_n = \sum_{k=1}^n \frac{\varepsilon_k}{k}$ with (ε_k) i.i.d., $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$ forms an L^2-bounded martingale. Hence $S_n \to S_\infty$ holds a.s. and in L^2. More generally, for $(a_k) \in \ell^2$ deterministic we have that $S_n = \sum_{k=1}^n \varepsilon_k a_k$ is an L^2-bounded martingale converging in L^2 and $\mathbb{P}$-a.s., e.g. $a_k = k^{-\alpha}$ is eligible for any $\alpha > 1/2$.

In the case of random (A_k) we could choose $A_k = \varepsilon_k/k$ and $S_n = \sum_{k=1}^n \varepsilon_k A_k = \sum_{k=1}^n k^{-1}$ diverges. If (A_k) is predictable and $M_n = \sum_{k=1}^n \varepsilon_k$, then $S_n = (A \bullet M)_n$ is an L^2-martingale and $\langle S \rangle_n = (A^2 \bullet \langle M \rangle)_n = \sum_{k=1}^n A_k^2$. Using $\mathbb{E}[S_n^2] = \mathbb{E}[\langle S \rangle_n] = \sum_{k=1}^n \mathbb{E}[A_k^2]$, we conclude that (S_n) is an L^2-bounded martingale if and only if $\sum_{k=1}^\infty \mathbb{E}[A_k^2] < \infty$.

4.44 Definition. For $A, B \in \mathscr{F}$ we write $A \subseteq B$ $\mathbb{P}$-a.s. if $\mathbb{P}(A \setminus B) = 0$.

4.45 Proposition. *Let (M_n) be an L^2-martingale. Then:*

$$\Big\{ \lim_{n\to\infty} \langle M \rangle_n < \infty \Big\} \subseteq \Big\{ \lim_{n\to\infty} M_n \text{ exists} \Big\} \quad \mathbb{P}\text{-a.s.}$$

4.46 Remark. If the increments of (M_n) are uniformly bounded then also the relation '$\supseteq$' holds $\mathbb{P}$-a.s., see Williams 12.13.

4.47 Example. The martingale $S_n = (A\bullet M)_n$ from the previous example converges for $\mathbb{P}$-almost all ω such that $\sum_{k=1}^\infty A_k(\omega)^2$ is finite.

Proof. W.l.o.g. assume $M_0 = 0$. $(\langle M\rangle_n)$ is a predictable process and $\tau_k := \inf\{n \geqslant 0 \,|\, \langle M\rangle_{n+1} > k\}$ are stopping times for each $k \in \mathbb{N}$. Then also $\langle M\rangle_n^{\tau_k} = \langle M\rangle_{\tau_k\wedge n}$ is predictable:

$$\{\langle M\rangle_{\tau_k\wedge n} \in B\} = \bigcup_{l=0}^{n-1}\{\tau_k = l,\, \langle M\rangle_l \in B\} \cup \{\tau_k \geqslant n,\, \langle M\rangle_n \in B\} \in \mathscr{F}_{n-1}$$

holds for $n \geqslant 1$, $B \in \mathfrak{B}_{\mathbb{R}}$. Because of $(M_n^{\tau_k})^2 - \langle M\rangle_n^{\tau_k} = (M^2 - \langle M\rangle)_n^{\tau_k}$ we have $\langle M^{\tau_k}\rangle = \langle M\rangle^{\tau_k}$. Hence, $\mathbb{E}[(M_n^{\tau_k})^2] = \mathbb{E}[\langle M\rangle_n^{\tau_k}] \leqslant k$ holds for all $n \geqslant 0$ and M^{τ_k} is an L^2-bounded and thus also L^1-bounded martingale. This shows that for all k $\lim_{n\to\infty} M_n^{\tau_k}$ exists $\mathbb{P}$-a.s. Now $\{\exists k \geqslant 1 : \tau_k = \infty\} = \{\lim_{n\to\infty}\langle M\rangle_n < \infty\}$ implies the assertion. ☐

4.48 Lemma. *Let $(X_n)_{n\geqslant 0}$ and $(A_n)_{n\geqslant 0}$ be processes, (A_n) non-negative and increasing with $A_n \uparrow A_\infty \in \mathbb{R}\cup\{\infty\}$. Then*

$$\{A_\infty = \infty\} \cap \Big\{\lim_{n\to\infty}((1+A)^{-1}\bullet X)_n \text{ exists in } \mathbb{R}\Big\} \subseteq \Big\{\lim_{n\to\infty}\frac{X_n}{A_n} = 0\Big\}.$$

Proof. Put $a_n = 1+A_n(\omega)$, $c_n = (X_n - X_{n-1})(\omega)$. Then for all ω in the left-hand side $a_n \uparrow \infty$ holds and $d_N := \sum_{n=1}^N c_n/a_n$ converges as $N \to \infty$. Kronecker's Lemma says that then $\frac{1}{a_N}\sum_{n=1}^N c_n \to 0$. This means $\frac{X_N(\omega)-X_0(\omega)}{1+A_N(\omega)} \to 0$ and thus $(X_N/A_N)(\omega) \to 0$, as asserted.

To verify Kronecker's Lemma put $b_n = a_n - a_{n-1}$ ($a_0 := 0$). Then partial summation gives for any $1 \leqslant k < N$

$$\begin{aligned}\frac{1}{a_N}\sum_{n=1}^N c_n &= \frac{1}{a_N}\sum_{n=1}^N a_n(d_n - d_{n-1}) = \sum_{n=1}^N \frac{b_n}{a_N}(d_N - d_{n-1})\\ &\leqslant \Big|\sum_{n=1}^k \frac{b_n}{a_N}(d_N - d_{n-1})\Big| + \max_{k\leqslant n\leqslant N}|d_N - d_n|.\end{aligned}$$

Letting first $N \to \infty$, then $k \to \infty$ the right-hand side tends to zero. ☐

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

$$\Big\{\lim_{n\to\infty}\langle M\rangle_n = \infty\Big\} \subseteq \Big\{\lim_{n\to\infty}\frac{M_n}{\langle M\rangle_n^\alpha} = 0\Big\} \quad \mathbb{P}\text{-a.s.}$$

Proof. Consider $X_n = ((1+\langle M\rangle^\alpha)^{-1}\bullet M)_n$. Then (X_n) is an L^2-martingale with

$$\langle X\rangle_n = ((1+\langle M\rangle^\alpha)^{-2}\bullet\langle M\rangle)_n = \sum_{k=1}^n \frac{\langle M\rangle_k - \langle M\rangle_{k-1}}{(1+\langle M\rangle_k^\alpha)^2} \leqslant \sum_{k=1}^n \int_{\langle M\rangle_{k-1}}^{\langle M\rangle_k}(1+t^\alpha)^{-2}dt.$$

This shows $\lim_{n\to\infty}\langle X\rangle_n \leqslant \int_0^\infty (1+t^\alpha)^{-2}dt < \infty$ $\mathbb{P}$-a.s. The above proposition implies that (X_n) converges $\mathbb{P}$-a.s. and the lemma with $A_n = \langle M\rangle_n^\alpha$ yields the result. ☐

4.50 Example. Let $(X_k)_{k\geqslant 1}$ be independent L^2-random variables with $\mathbb{E}[X_k]=\mu_k$, $\operatorname{Var}(X_k)=\sigma_k^2$. Then $S_n=\sum_{k=1}^n (X_k-\mu_k)$ is an L^2-martingale with $\langle S\rangle_n=\operatorname{Var}(S_n)=\sum_{k=1}^n\sigma_k^2$. If $\operatorname{Var}(S_n)\to\infty$, then we conclude

$$\forall \alpha>1/2:\ \frac{S_n-\mathbb{E}[S_n]}{\operatorname{Var}(S_n)^\alpha}\to 0\quad \mathbb{P}\text{-a.s.}$$

4.51 Remark. The limiting case $\alpha=1/2$ gives rise to a martingale central limit theorem. If (M_n) is an L^2-martingale with $M_0=0$, $\langle M\rangle_n\to\infty$, then

$$\frac{M_n}{\langle M\rangle_n^{1/2}}\xrightarrow{d} N(0,1)$$

holds, provided $\langle M\rangle_n/\mathbb{E}[M_n^2]\to 1$ and the conditional Lindeberg condition is satisfied:

$$\forall\varepsilon>0:\ \frac{1}{\mathbb{E}[M_n^2]}\sum_{k=1}^n\mathbb{E}\Big[(M_k-M_{k-1})^2\mathbf{1}\big((M_k-M_{k-1})^2\geqslant\varepsilon^2\,\mathbb{E}[M_n^2]\big)\,\Big|\,\mathscr{F}_{k-1}\Big]\xrightarrow{\mathbb{P}}0,$$

cf. Shiryaev, Thm. VII.8.4 or Luschgy, Satz 5.31.

4.52 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 $\mathbb{E}[M_{-n}\,|\,\mathscr{F}_{-n-1}]=M_{-n-1}$ hold for all $n\geqslant 0$.

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

Proof. Denote by $U_{-n}^{[a,b]}$ the upcrossings on $[a,b]$ of $(M_{-k},k=0,\dots,n)$. The upcrossing inequality gives $\mathbb{E}[U_{-n}^{[a,b]}]\leqslant\mathbb{E}[(M_0-a)_+]/(b-a)$ because it relies on the martingale property for finitely many time indices only. With $U_{-n}^{[a,b]}\uparrow U^{[a,b]}$ as $n\uparrow\infty$ monotone convergence shows $\mathbb{E}[U^{[a,b]}]\leqslant\mathbb{E}[(M_0-a)_+]/(b-a)<\infty$ for any $a<b$. Now, $U^{[a,b]}$ counts the upcrossings on $[a,b]$ of $(M_{-n},n\geqslant 0)$. As in the first martingale convergence theorem this implies $\mathbb{P}$-a.s. convergence $M_{-n}\to M_{-\infty}$. Now $M_{-n}=\mathbb{E}[M_0\,|\,\mathscr{F}_n]$, $n\geqslant 0$, holds and the same argument as for the second martingale convergence theorem shows that $(M_{-n},n\geqslant 0)$ is uniformly integrable. Vitali's Theorem thus implies $M_{-n}\to M_{-\infty}$ in L^1. □

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

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

Proof. Put $S_0=0$, $S_n=\sum_{k=1}^n X_k$, $n\geqslant 1$ and $\mathscr{F}_{-n}=\sigma(S_k,k\geqslant n)$, $n\geqslant 0$. Then $\mathscr{F}_{-n-1}\subseteq\mathscr{F}_{-n}$ holds for $n\geqslant 0$ and S_n is $\mathscr{F}_{-n}$-measurable. From

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

and the fact that $(X_k)_{k=1,\dots,n}$ has the same law as $(X_{\pi(k)})_{k=1,\dots,n}$ for any permutation π of $\{1,\dots,n\}$ ((X_k) are *exchangeable*), while S_m for $m \geqslant n$ is invariant under each π, we conclude that $\mathbb{E}[X_k \,|\, \mathscr{F}_{-n}]$ does not depend on k and equals S_n/n for $k = 1,\dots,n$. By definition $M_{-n} := \mathbb{E}[X_1 \,|\, \mathscr{F}_{-n}]$ forms a backward martingale. We conclude that $S_n/n = M_{-n} \to M_{-\infty}$ converges $\mathbb{P}$-a.s. and in L^1. By Kolmogorov's 0-1 law the limit $M_{-\infty}$ must be $\mathbb{P}$-a.s. constant. From $\mathbb{E}[M_{-n}] \to \mathbb{E}[M_{-\infty}]$ and $\mathbb{E}[M_{-n}] = \mathbb{E}[X_1]$ we infer $M_{-\infty} = \mathbb{E}[X_1]$ $\mathbb{P}$-a.s. □

4.4 The Radon-Nikodym theorem

4.55 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 $\mu \perp \nu$.

4.56 Example.

(a) A probability measure $\mathbb{P}$ on $\mathfrak{B}_{\mathbb{R}^d}$ with Lebesgue density f is absolutely continuous with respect to Lebesgue measure λ_d: $\lambda_d(A) = 0 \Rightarrow \mathbb{P}(A) = \int_A f(x)\lambda_d(dx) = 0$. More generally, any measure μ with a ν-density f, i.e. $\mu(A) = \int_A f d\nu$, satisfies $\mu \ll \nu$. If $f > 0$ ν-almost everywhere holds, then $\mu(A) = 0$ implies $\int f \mathbf{1}_A d\nu = 0$ and thus $\nu(A) = 0$. This shows that also $\nu \ll \mu$ and both measures are equivalent.

(b) The measures λ_1 and δ_0 on $\mathfrak{B}_{\mathbb{R}}$ are singular: $\lambda_1(\{0\}) = 0$, $\delta_0(\{0\}^C) = 0$.

4.57 Lemma. *A finite measure μ is absolutely continuous with respect to ν if and only if*

$$\forall \varepsilon \geqslant 0 \,\exists \delta > 0 \,\forall A \in \mathscr{F} :\; \nu(A) < \delta \Rightarrow \mu(A) < \varepsilon.$$

4.58 Remark.

(a) It is necessary to ask for finite μ as the following counterexample shows: take ν the Lebesgue measure on $((0,1], \mathfrak{B}_{(0,1]})$ and $\mu(dx) = x^{-1}dx$. Then $\mu \ll \nu$ holds, while $\nu((0,\delta)) = \delta$ and $\mu((0,\delta)) = \infty$ hold for all $\delta \in (0,1)$.

(b) If F is the distribution function of a probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, absolutely continuous with respect to Lebesgue measure, then the lemma says that there is for any $\varepsilon > 0$ a $\delta > 0$ such that

$$\forall a_1 \leqslant b_1 \leqslant \dots \leqslant a_n \leqslant b_n :\; \sum_{i=1}^n (b_i - a_i) < \delta \Rightarrow \sum_{i=1}^n (F(b_i) - F(a_i)) < \varepsilon.$$

In real analysis one says that the function F is *absolutely continuous* and thus *weakly differentiable* with the Radon-Nikodym derivative f. Note that for Cantor measure $\mathbb{P}$ the distribution function F is continuous, but not absolutely continuous in that sense.

Proof. To prove '$\Leftarrow$' suppose $\nu(A)=0$. Then $\nu(A)<\delta$ holds for all $\delta>0$ and thus $\mu(A)<\varepsilon$ for all $\varepsilon>0$. This implies $\mu(A)=0$, hence $\mu \ll \nu$.

The implication '$\Rightarrow$' is shown by contradiction. Assume there are $A_n \in \mathscr{F}$ and $\varepsilon>0$ with $\nu(A_n)\leqslant 2^{-n}$ and $\mu(A_n)\geqslant\varepsilon$. Consider the event $A_\infty = \bigcap_{m\geqslant 1}\bigcup_{n\geqslant m} A_n$ that infinitely many events A_n occur. Then for all $n_0\in\mathbb{N}$

$$\nu(A_\infty)=\int \limsup_{n\to\infty} \mathbf{1}_{A_n}d\nu \leqslant \int \sum_{n>n_0} \mathbf{1}_{A_n}d\nu = \sum_{n>n_0}\nu(A_n)=2^{-n_0}$$

holds, which for $n_0\to\infty$ yields $\nu(A_\infty)=0$ (for $\nu=\mathbb{P}$ this is Borel-Cantelli). On the other hand, Fatou's Lemma gives

$$\mu(A_\infty)=\int \limsup_{n\to\infty}(1-1_{A_n^C})\,d\mu \geqslant \mu(\Omega)-\liminf_{n\to\infty}\int \mathbf{1}_{A_n^C}d\mu = \limsup_{n\to\infty}\mu(A_n)\geqslant\varepsilon.$$

This contradicts $\mu \ll \nu$. □

4.59 Theorem (Radon-Nikodym Theorem (Lebesgue 1910, Radon 1913, Nikodym 1930))**.** *Let ν be a σ-finite measure and μ a measure on $\mathscr{F}$ with $\mu\ll\nu$, then there is an $f\in\mathcal{M}^+(\Omega,\mathscr{F})$ such that*

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

4.60 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 $f=\frac{d\mu}{d\nu}$.

Proof. We give the proof in the case μ finite, $\nu=\mathbb{P}$ and $\mathscr{F}=\sigma(F_n,n\geqslant 1)$ for some $F_n\subseteq\Omega$ ($\mathscr{F}$ *separable*, e.g. Borel σ-algebra of a Polish space); see Williams for the general case.

Put $\mathscr{F}_n=\sigma(F_1,\dots,F_n)$. Then $\mathscr{F}_n$ consists of finitely many events only (intersections of the F_i and finite unions). In particular, there are finitely many atoms $A_1^{(n)},\dots,A_{r_n}^{(n)}\in\mathscr{F}_n$ with $\bigcup_{i=1}^{r_n}A_i^{(n)}=\Omega$, $A_i^{(n)}\cap A_j^{(n)}=\varnothing$ for $i\neq j$ and $\mathscr{F}_n=\sigma(A_1^{(n)},\dots,A_{r_n}^{(n)})$. We set

$$M_n:=\sum_{i=1}^{r_n}\Big(\frac{\mu(A_i^{(n)})}{\mathbb{P}(A_i^{(n)})}\mathbf{1}(\mathbb{P}(A_i^{(n)})>0)\Big)\mathbf{1}_{A_i^{(n)}}.$$

Then $M_n\in\mathcal{M}^+(\Omega,\mathscr{F}_n)$ and for $F\in\mathscr{F}_n$ we have $\int_F M_n d\,\mathbb{P}=\sum_i\mu(A_i^{(n)}\cap F)=\mu(F)$. This shows $M_n=\frac{d\mu|_{\mathscr{F}_n}}{d\,\mathbb{P}|_{\mathscr{F}_n}}$ and (M_n) is a martingale because $\mathbb{E}[M_n\mathbf{1}_F]=\mu(F)=\mathbb{E}[M_{n+1}\mathbf{1}_F]$ for $F\in\mathscr{F}_n$ means $\mathbb{E}[M_{n+1}\,|\,\mathscr{F}_n]=M_n$.

From the above lemma for any $\varepsilon>0$ there is a $\delta>0$ such that $\mathbb{P}(M_n>R)<\delta$ implies $\mu(M_n>R)<\varepsilon$. By Markov's inequality this holds in particular for $R>\mu(\Omega)/\delta$: $\mathbb{P}(M_n>R)\leqslant\frac{\mathbb{E}[M_n]}{R}<\delta$. Then $\sup_n\mathbb{E}[M_n\mathbf{1}(M_n>R)]=\sup_n\mu(M_n>R)<\varepsilon$ holds. We conclude that (M_n) is uniformly integrable and hence forms a closable martingale.

Define the finite measure $\rho(A) = \int_A M_\infty d\,\mathbb{P}$ on $\mathscr{F}$. For $A \in \mathscr{F}_n$ the closable martingale property gives

$$\rho(A) = \mathbb{E}[\mathbf{1}_A\,\mathbb{E}[M_\infty \,|\, \mathscr{F}_n]] = \mathbb{E}[\mathbf{1}_A M_n] = \mu(A).$$

Consequently ρ and μ coincide on the $\cap$-stable generator $\bigcup_{n\geqslant 1}\mathscr{F}_n$ of $\mathscr{F}$ and have the same mass $\rho(\Omega) = \mu(\Omega)$. The uniqueness theorem for finite measures thus gives $\rho = \mu$ on $\mathscr{F}$ and the theorem follows with $f = M_\infty$. □

4.61 Corollary (Lebesgue decomposition)**.** *For σ-finite measures μ, ν on $(\Omega, \mathscr{F})$ we can decompose $\mu = \mu_1 + \mu_2$ with σ-finite measures $\mu_1 \ll \nu$ and $\mu_2 \perp \nu$.*

4.62 Remark. It is easy to see that this decomposition is unique unless $\nu = 0$.

Proof. Put $\rho = \mu + \nu$. Since $\nu \ll \rho$ and ρ is σ-finite, there is a Radon-Nikodym derivative $f = \frac{d\nu}{d\rho}$. Set $\mu_1(A) := \mu(A \cap \{f > 0\})$, $\mu_2(A) := \mu(A \cap \{f = 0\})$, $A \in \mathscr{F}$. Then $\mu = \mu_1 + \mu_2$ holds and

$$\nu(\{f = 0\}) = \int_{\{f=0\}} f\,d\rho = 0, \quad \mu_2(\{f = 0\}^C) = \mu(\varnothing) = 0 \Rightarrow \nu \perp \mu_2.$$

For $A \in \mathscr{F}$ with $\nu(A) = 0$, on the other hand, we have

$$\int_A f\,d\rho = 0 \Rightarrow \int_{A\cap\{f>0\}} f\,d\mu = 0 \Rightarrow \mu(A \cap \{f > 0\}) = 0,$$

which implies $\mu_1 \ll \nu$. □

4.63 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).

Proof. First note that $a_k \leqslant \mathbb{E}[X_k]^{1/2} = 1$ follows from Jensen's inequality and $a_k > 0$ from $\mathbb{P}(X_k > 0) > 0$ due to $\mathbb{E}[X_k] = 1$. The equivalence (a) $\iff$ (b) $\iff$ (c) is due to Vitali's Theorem. The equivalence (d) $\iff$ (e) is shown in analysis (consider $\log(\prod_k a_k)$).

(a)$\Rightarrow$(d): Define $N_n := \prod_{k=1}^n a_k^{-1} X_k^{1/2}$, $n \geqslant 1$, $N_0 = 1$. Then (N_n) is a non-negative L^2-martingale with $M_n^{1/2} / \prod_{k=1}^n a_k = N_n \to N_\infty$ $\mathbb{P}$-a.s. for some

$N_\infty \in L^1$. Since the nonnegative martingale (M_n) satisfies $M_n \to M_\infty$ $\mathbb{P}$-a.s. with $\mathbb{E}[M_\infty] = 1$ by (a) and the product converges always, we have

$$M_\infty = N_\infty^2 \prod_{k=1}^\infty a_k \ \mathbb{P}\text{-a.s.} \Rightarrow \mathbb{E}\Big[N_\infty^2 \prod_{k=1}^\infty a_k\Big] = 1 \Rightarrow \prod_{k=1}^\infty a_k > 0.$$

(d)⇒(c): the martingale (N_n) above satisfies under (d) $\mathbb{E}[N_n^2] = \mathbb{E}[M_n]/\prod_{k=1}^n a_k^2 \leqslant \prod_{k=1}^\infty a_k^{-2} < \infty$. Monotone convergence, Doob's L^2-inequality and $a_k \leqslant 1$ thus give

$$\mathbb{E}\Big[\sup_{n\geqslant 0}|M_n|\Big] \leqslant \mathbb{E}\Big[\sup_{n\geqslant 0} N_n^2\Big] = \lim_{K\to\infty} \mathbb{E}\Big[\max_{0\leqslant n\leqslant K} N_n^2\Big] \leqslant 4 \lim_{K\to\infty} \mathbb{E}[N_K^2] < \infty.$$

This shows that $(M_n)_{n\geqslant 0}$ is dominated by $\sup_{n\geqslant 0}|M_n| \in L^1$ and thus uniformly integrable.

If (a)-(e) do not hold, then $\prod_{k=1}^\infty a_k = 0$ and the above calculation shows $\mathbb{E}[M_\infty] = \mathbb{E}[N_\infty^2 \prod_{k=1}^\infty a_k] = 0$ and thus $M_\infty = 0$ $\mathbb{P}$-a.s. □

4.64 Definition. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures with densities $p = \frac{d\mathbb{P}}{d\mu}$, $q = \frac{d\mathbb{Q}}{d\mu}$ for some *dominating* measure μ (e.g. $\mu = \mathbb{P} + \mathbb{Q}$). Then their Hellinger distance is defined as

$$H(\mathbb{P}, \mathbb{Q}) := \Big(\int (\sqrt{p} - \sqrt{q})^2 d\mu\Big)^{1/2} = \|\sqrt{p} - \sqrt{q}\|_{L^2(\mu)}.$$

4.65 Remark. One can show that this definition does not depend on the dominating measure μ. H defines a metric on the set of all probability measures on $(\Omega, \mathscr{F})$.

4.66 Lemma. *The formula $H^2(\mathbb{P}, \mathbb{Q}) = 2(1 - \int \sqrt{pq}\, d\mu)$ holds and if $\mathbb{Q} \ll \mathbb{P}$ then $H^2(\mathbb{P}, \mathbb{Q}) = 2(1 - \mathbb{E}_\mathbb{P}[(\frac{d\mathbb{Q}}{d\mathbb{P}})^{1/2}])$.*

Proof. The first identity follows from the binomial formula:

$$H^2(\mathbb{P}, \mathbb{Q}) = \int \Big(p - 2\sqrt{pq} + q\Big)\, d\mu = 1 - 2\int \sqrt{pq}\, d\mu + 1.$$

For $\mathbb{Q} \ll \mathbb{P}$ consider $\mu = \mathbb{P}$ and note $q = \frac{d\mathbb{Q}}{d\mathbb{P}}$, $p = 1$. □

4.67 Corollary. *Let $(\mathbb{P}_n)_{n\geqslant 1}$, $(\mathbb{Q}_n)_{n\geqslant 1}$ be two sequences of probability measures on $(\Omega, \mathscr{F})$ with $\mathbb{Q}_n \ll \mathbb{P}_n$, $n \geqslant 1$. Then for the product measures on $(\Omega^\mathbb{N}, \mathscr{F}^{\otimes \mathbb{N}})$ we have*

$$\bigotimes_{n=1}^\infty \mathbb{Q}_n \ll \bigotimes_{n=1}^\infty \mathbb{P}_n \iff \sum_{n=1}^\infty H^2(\mathbb{P}_n, \mathbb{Q}_n) < \infty.$$

Otherwise, we have singularity $\bigotimes_{n=1}^\infty \mathbb{Q}_n \perp \bigotimes_{n=1}^\infty \mathbb{P}_n$ (Kakutani's dichotomy).

Proof. Denote by $X_n : \Omega^\mathbb{N} \to \Omega$ the projection on the nth coordinate and recall that under $\mathbb{P} = \bigotimes_{n=1}^\infty \mathbb{P}_n$ the (X_n) are independent random variables with laws $\mathbb{P}_n$. Introduce

$$\Lambda_n(\omega) = \frac{d(\mathbb{Q}_1 \otimes \cdots \otimes \mathbb{Q}_n)}{d(\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n)}(\omega_1, \ldots, \omega_n) = \prod_{k=1}^n \frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(\omega_k), \quad \omega \in \Omega^\mathbb{N}.$$

Then $\Lambda_n = \prod_{k=1}^n \frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(X_k)$ forms a martingale with respect to $\mathscr{F}_n = \sigma(X_1,\dots,X_n)$ under $\mathbb{P}$ as in Kakutani's Theorem. Observing

$$a_n = \mathbb{E}_{\mathbb{P}}\Big[\Big(\frac{d\,\mathbb{Q}_n}{d\,\mathbb{P}_n}(X_n)\Big)^{1/2}\Big] = \mathbb{E}_{\mathbb{P}_n}\Big[\Big(\frac{d\,\mathbb{Q}_n}{d\,\mathbb{P}_n}\Big)^{1/2}\Big] = 1 - \tfrac{1}{2}H^2(\mathbb{P}_n,\mathbb{Q}_n),$$

Condition (e) of Kakutani's Theorem is satisfied if and only if $\sum_{n=1}^\infty H^2(\mathbb{P}_n,\mathbb{Q}_n) < \infty$. In that case (Λ_n) is a closable martingale with $\Lambda_n = \mathbb{E}[\Lambda_\infty \,|\, \mathscr{F}_n]$. Set

$$\mathbb{Q}(A) := \int_A \Lambda_\infty \, d\,\mathbb{P}, \quad A \in \mathscr{F}^{\otimes \mathbb{N}}.$$

Then $\mathbb{Q}$ is a probability measure (note $\mathbb{E}_{\mathbb{P}}[\Lambda_\infty] = 1$) and for $A \in \mathscr{F}_n$, i.e. $A = (X_1,\dots,X_n)^{-1}(B)$ for some $B \in \mathscr{F}^{\otimes n}$, we have

$$\begin{aligned}\mathbb{Q}(A) &= \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[\Lambda_\infty|\,\mathscr{F}_n]\,\mathbf{1}_A] = \mathbb{E}_{\mathbb{P}}[\Lambda_n \mathbf{1}_B(X_1,\dots,X_n)]\\ &= \mathbb{E}_{\bigotimes_{k=1}^n \mathbb{P}_k}\Big[\frac{d(\mathbb{Q}_1\otimes\cdots\otimes\mathbb{Q}_n)}{d(\mathbb{P}_1\otimes\cdots\times\mathbb{P}_n)}\mathbf{1}_B\Big] = (\mathbb{Q}_1\otimes\cdots\otimes\mathbb{Q}_n)(B).\end{aligned}$$

Consequently, $\mathbb{Q}$ satisfies the definition of the product measure $\bigotimes_{n=1}^\infty \mathbb{Q}_n$ and by uniqueness ($\bigcup_n \mathscr{F}_n$ is an $\cap$-stable generator of $\mathscr{F}^{\otimes\mathbb{N}}$) $\mathbb{Q} = \bigotimes_{n=1}^\infty \mathbb{Q}_n$ follows. With Λ_∞ we have exhibited its density with respect to $\mathbb{P}$ such that $\mathbb{Q} \ll \mathbb{P}$ holds.

In the case $\sum_{n=1}^\infty H^2(\mathbb{P}_n,\mathbb{Q}_n) = \infty$ Kakutani's Theorem gives $\Lambda_\infty = 0$ $\mathbb{P}$-a.s. Consider

$$\Lambda_n' = \prod_{k=1}^n \Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(X_k)\Big)^{-1}\mathbf{1}\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(X_k) > 0\Big).$$

Then (Λ_n') forms a non-negative supermartingale with respect to $\mathbb{Q}$ (from above) and $(\mathscr{F}_n)$ due to the independence of (X_k) and

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\Big[\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(X_k)\Big)^{-1}\mathbf{1}\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(X_k) > 0\Big)\Big] &= \mathbb{E}_{\mathbb{P}_k}\Big[\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}\Big)^{-1}\mathbf{1}\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k} > 0\Big)\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}\Big]\\ &= \mathbb{P}_k\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k} > 0\Big).\end{aligned}$$

Hence, $\Lambda_n' \to \Lambda_\infty'$ holds $\mathbb{Q}$-a.s. with some $\Lambda_\infty' \in L^1(\mathbb{Q})$. Because of

$$\mathbb{Q}\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}(X_k) = 0\Big) = \mathbb{Q}_k\Big(\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k} = 0\Big) = \int_{\{\frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k} = 0\}} \frac{d\,\mathbb{Q}_k}{d\,\mathbb{P}_k}\, d\,\mathbb{P}_k = 0,$$

we have $\Lambda_n'\Lambda_n = 1$ $\mathbb{Q}$-a.s. This implies $\mathbb{Q}(\Lambda_n \to 0) = \mathbb{Q}(\Lambda_n' \to \infty) = 0$. Together with $\Lambda_\infty = 0$ $\mathbb{P}$-a.s., i.e. $\mathbb{P}(\Lambda_n \to 0) = 1$, this shows $\mathbb{P} \perp \mathbb{Q}$. □

4.68 Remark. Suppose $\sum_{n=1}^\infty H^2(\mathbb{P}_n,\mathbb{Q}_n) = \infty$ and even equivalence $\mathbb{P}_n \sim \mathbb{Q}_n$, $n \geqslant 1$. As the proof shows, we then have by symmetry between $\mathbb{Q}_n$ and $\mathbb{P}_n$ that $\Lambda_n \to 0$ $\mathbb{P}$-a.s. and $\Lambda_n \to \infty$ $\mathbb{Q}$-a.s. The *likelihood process* (Λ_n) is therefore a good criterion to discriminate between $\mathbb{Q}$ and $\mathbb{P}$ (it gives even rise to optimal tests, cf. the Neyman-Pearson Lemma in statistics).

4.69 Example. We want to test the null hypothesis $H_0 : \vartheta = \vartheta_0$ against the alternative $H_1 : \vartheta = \vartheta_1$ for two parameters ϑ_0, ϑ_1 from independent observations $X_1, \dots, X_n$. We assume that the laws of X_n under the null hypothesis $\mathbb{P}_{\vartheta_0}$ and the alternative $\mathbb{P}_{\vartheta_1}$ are equivalent. Mathematically, we consider the situation of the corollary with $\mathbb{P}_n = \mathbb{P}_{\vartheta_0}^{X_n}$, $\mathbb{Q}_n = \mathbb{P}_{\vartheta_1}^{X_n}$ and $\mathbb{P}_n \sim \mathbb{Q}_n$.

The likelihood-ratio test rejects H_0 (in favour of H_1) if $\Lambda_n := \prod_{k=1}^n \frac{d\mathbb{Q}_k}{d\mathbb{P}_k}(X_k) > \kappa$ holds for some fixed deterministic *critical value* $\kappa > 0$ and accepts H_0 for $\Lambda_n \leqslant \kappa$. Suppose $\sum_{n=1}^\infty H^2(\mathbb{P}_n, \mathbb{Q}_n) = \infty$, e.g. in the canonical case that $(X_k)_{k\geqslant 0}$ are i.i.d. under $\mathbb{P}_{\vartheta_0}$ and under $\mathbb{P}_{\vartheta_1}$, but with different distributions. Then the probability for an error of the first kind satisfies as the sample size n tends to infinity (apply Fatou's Lemma to $1 - \mathbf{1}(\Lambda_n > \kappa)$):

$$\lim_{n\to\infty} \mathbb{P}_{\vartheta_0}(\Lambda_n > \kappa) \leqslant \mathbb{E}_{\vartheta_0}[\limsup_{n\to\infty} \mathbf{1}(\Lambda_n > \kappa)] \leqslant \mathbb{P}_{\vartheta_0}(\Lambda_\infty > 0) = 0.$$

Similarly, we have for the error probability of the second kind:

$$\lim_{n\to\infty} \mathbb{P}_{\vartheta_1}(\Lambda_n \leqslant \kappa) \leqslant \mathbb{P}_{\vartheta_1}(\Lambda_\infty < \infty) = 0.$$

We say that the likelihood-ratio test is *asymptotically consistent*.

A concrete example is the *Gauß test* where $X_n \sim N(\vartheta_0, 1)$ under H_0 and $X_n \sim N(\vartheta_1, 1)$ under H_1 with $\vartheta_0, \vartheta_1 \in \mathbb{R}$ and different. Then we have

$$\Lambda_n = \prod_{k=1}^n \frac{(2\pi)^{-1/2} e^{-(X_k - \vartheta_1)^2/2}}{(2\pi)^{-1/2} e^{-(X_k - \vartheta_0)^2/2}} = \exp\Big(\frac{1}{2}\sum_{k=1}^n (X_k - \vartheta_0)^2 - \frac{1}{2}\sum_{k=1}^n (X_k - \vartheta_1)^2\Big).$$

The null hypothesis is therefore rejected if the difference of deviations $\sum_{k=1}^n (X_k - \vartheta_0)^2 - \sum_{k=1}^n (X_k - \vartheta_1)^2$ is larger than $2\log(\kappa)$.

Let us add a specific example with $\sum_{n=1}^\infty H^2(\mathbb{P}_n, \mathbb{Q}_n) < \infty$. Suppose $X_n \sim N(\vartheta_0, \sigma_n^2)$ under H_0 and $X_n \sim N(\vartheta_1, \sigma_n^2)$ under H_1 for some sequence (σ_n). Then the same calculation shows that the test rejects if the difference of weighted deviations $\sum_{k=1}^n \sigma_k^{-2}(X_k - \vartheta_0)^2 - \sum_{k=1}^n \sigma_k^{-2}(X_k - \vartheta_1)^2$ is larger than $2\log(\kappa)$. For the Hellinger distance we use Lebesgue measure as dominating measure and find

$$\begin{aligned} H^2(N(\vartheta_0, \sigma^2), N(\vartheta_1, \sigma^2)) &= 2 - 2\int \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\vartheta_0)^2}{4\sigma^2} - \frac{(x-\vartheta_1)^2}{4\sigma^2}\Big)\,dx \\ &= 2 - 2e^{-(\vartheta_1 - \vartheta_0)^2/(8\sigma^2)}. \end{aligned}$$

The equivalence (d) $\iff$ (e) in Kakutani's Theorem therefore shows $\sum_{n=1}^\infty H^2(\mathbb{P}_n, \mathbb{Q}_n) < \infty$ if and only if

$$\prod_{k=1}^\infty e^{-(\vartheta_1 - \vartheta_0)^2/(8\sigma_k^2)} > 0 \iff \sum_{k=1}^\infty \sigma_k^{-2} < \infty.$$

From the corollary we can conclude that in the case $\sum_{k=1}^\infty \sigma_k^{-2} < \infty$, e.g $\sigma_k = k$, the laws of $(X_k)_{k\geqslant 1}$ are equivalent under H_0 and H_1. This means that any test, based on the entire sequence $(X_k)_{k\geqslant 1}$, which accepts H_0 with positive probability under H_0 will also accept H_0 with positive probability under H_1 and vice versa. Null hypothesis and alternative cannot be separated perfectly and there exist no asymptotically consistent tests.

5 Ergodic theory

5.1 Stationary and ergodic processes

5.1 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 $(X_{t_1}, \dots, X_{t_n}) \overset{d}{=} (X_{t_1+s}, \dots, X_{t_n+s})$ holds for all $n \geqslant 1$, $t_1, \cdots, t_n \in T$ and $s \in T$.

5.2 Remark. Stationarity does not mean that the random variables (X_t) are invariant under time shifts (sample paths are still usually non-constant!), but that the finite-dimensional distributions are invariant under time shifts.

5.3 Example.

(a) $X_t = A\sin(\omega t + U)$, $t \geqslant 0$, with $A, \omega \in \mathbb{R}$ and $U \sim U([0, 2\pi])$ (periodic signal with random phase). Then $X_t = \mathrm{Re}(Z_t)$ holds with $Z_t = Ae^{i(\omega t+U)}$ and

$$(Z_{t_1+s}, \dots, Z_{t_n+s}) = Ae^{i(\omega s+U)}(e^{i\omega t_1}, \dots, e^{i\omega t_n}).$$

Since $e^{i(\omega s+U)} \overset{d}{=} e^{iU}$ is uniformly distributed on S^1, we conclude $(Z_{t_1+s}, \dots, Z_{t_n+s}) \overset{d}{=} (Z_{t_1}, \dots, Z_{t_n})$ and (Z_t) is stationary. This implies that also (X_t) is stationary.

(b) Let $(X_t, t \in T)$ be a Gaussian process with expectation function $\mu(t)$ and covariance function $c(t, s)$. If (X_t) is stationary, then $X_t \overset{d}{=} X_s$ implies $\mu(t) = \mu(s)$ and μ must be constant. Furthermore, $(X_t, X_{t+u}) \overset{d}{=} (X_s, X_{s+u})$ implies $c(t, t+u) = c(s, s+u)$ and the covariance function satisfies $c(t, s) = c(0, |t-s|)$. It is easy to see that these properties of μ and c conversely imply that (X_t) is stationary.

More generally, any process (X_t) with $X_t \in L^2$, constant expectation function $\mu(t)$ and covariance function $c(t, s) = c(0, |t-s|)$, only depending on the distance of time points, is called weakly stationary.

5.4 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 $i \in S$.

5.5 Lemma. *A time-homogeneous Markov chain with invariant initial distribution is stationary.* ▶ Exercise

5.6 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 $A \in \mathscr{F}$.

5.7 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

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

(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.

Proof.

(a) Let $\pi_{\{0,\dots,n\}}((\omega_k)_{k\geqslant 0}) = (\omega_0, \dots, \omega_n)$ denote the projection onto the first $(n+1)$ coordinates. Consider a cylinder set $A = \pi^{-1}_{\{0,\dots,n\}}(B_n)$ with $B_n \in \mathcal{S}^{\otimes(n+1)}$. Then:

$$\mathbb{P}^X(A) = \mathbb{P}((X_0, \dots, X_n) \in B_n) \stackrel{!}{=} \mathbb{P}((X_1, \dots, X_{n+1}) \in B_n) = \mathbb{P}(T \circ X \in A).$$

Since the cylinder sets form an $\cap$-stable generator of $\mathcal{S}^{\otimes \mathbb{N}_0}$ the probability measures $\mathbb{P}^X$ and $\mathbb{P}^{T\circ X}$ (note $\mathbb{P}^{T\circ X}(A) = \mathbb{P}(T \circ X \in A)$) coincide on $\mathcal{S}^{\otimes \mathbb{N}_0}$ and T is measure-preserving.

(b) We calculate, using measure preservation $\mathbb{P}(T^{-m}(B)) = \mathbb{P}(B)$:

$$\begin{aligned}\mathbb{P}((X_m, \dots, X_{m+n}) \in A_n) &= \mathbb{P}((Y \circ T^m, \dots, Y \circ T^{m+n}) \in A_n) \\ &= \mathbb{P}(T^{-m}(Y \circ T^0, \dots, Y \circ T^n)^{-1}(A_n)) \\ &\stackrel{!}{=} \mathbb{P}((Y \circ T^0, \dots, Y \circ T^n)^{-1}(A_n)) \\ &= \mathbb{P}((X_0, \dots, X_n) \in A_n).\end{aligned}$$

□

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

5.9 Remark. Null and one sets are always invariant events. An ergodic transformation leaves no other events fixed than those.

5.10 Example. Consider $\Omega = \{1, 2, 3, 4, 5\}$, $\mathscr{F} = \mathcal{P}(\Omega)$ and the permutation $T \in S_5$ with cycles $T = (1, 2, 3)(4, 5)$. Then T preserves the measure $\mathbb{P}$ whenever $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\})$ and $\mathbb{P}(\{4\}) = \mathbb{P}(\{5\})$. If all probabilities are non-zero, then $\mathscr{I}_T = \{\varnothing, \Omega, \{1, 2, 3\}, \{4, 5\}\}$ holds. For $X_n(\omega) = \mathbf{1}_{\{1,4\}}(T^n(\omega))$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X_i(\omega) = \begin{cases} 1/3, & \text{if } \omega \in \{1, 2, 3\} \\ 1/2, & \text{if } \omega \in \{4, 5\}. \end{cases}$$

Note that the limit can be written as $\mathbb{E}[X_0 \mid \mathscr{I}_T]$.

5.11 Lemma. ► Exercise *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 $\mathbb{P}(A \Delta B) = 0$.*

5.12 Example.

(a) Suppose $(X_n)_{n \geqslant 0}$ are i.i.d. $(S, \mathscr{S})$-valued random variables. Then they form a stationary process X. Moreover, below we shall see that $T^{-1}(A) = A$ for the left shift T on $(S^{\mathbb{N}_0}, \mathcal{S}^{\otimes \mathbb{N}_0}, \mathbb{P}^X)$ implies $A \in \bigcap_{n \geqslant 1} \sigma(\pi_k,\, k \geqslant n)$ with the projection π_k on the k-th coordinate. This means that A lies in the terminal (asymptotic) σ-algebra of (π_n). Since (π_n) under $\mathbb{P}_X$ are as (X_n) under $\mathbb{P}$ distributed and therefore independent, Kolmogorov's 0-1 law implies $\mathbb{P}(A) \in \{0, 1\}$. Using part (b) of the preceding lemma, we infer that T is ergodic. We also say that (X_n) forms an ergodic process.

It remains to show that $T^{-1}(A) = A$ implies $A \in \sigma(\pi_k, k \geqslant n)$ for all $n \geqslant 1$. We have $\omega \in T^{-1}(A) \iff (\omega_1, \omega_2, \ldots) \in A$ and thus $A = T^{-1}(A) \in \sigma(\pi_k,\, k \geqslant 1)$. Inductively, we conclude $A = T^{-n}(A) \in \sigma(\pi_k,\, k \geqslant n)$.

(b) Let $\Omega = [0, 1)$, $\mathscr{F} = \mathfrak{B}_\Omega$, $\mathbb{P}$ Lebesgue measure on $[0, 1)$ and $T(\omega) = (\omega + \vartheta)$ mod 1 for some fixed $\vartheta \in \mathbb{R}$ (model for rotation by an angle $2\pi\vartheta$). Properties of Lebesgue measure ensure that T is measure-preserving. Cases:

(i) $\vartheta = p/q \in \mathbb{Q}$: Consider $A = \bigcup_{k=0}^{q-1} [k/q, (k + 1/2)/q)$. Then $A = (A + p/q) \mod 1$ implies that A is T-invariant, but $\mathbb{P}(A) = 1/2$ holds. T is not ergodic (it is *periodic*).

(ii) $\vartheta \in \mathbb{R} \setminus \mathbb{Q}$: Let $f : [0, 1] \to \mathbb{R}$ be bounded, measurable and $\mathbb{P}$-a.s. invariant, i.e. $f \circ T = f$ Lebesugue-almost everywhere. Since $f \in L^2([0, 1))$ holds, there is a Fourier series expansion $f(x) = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k x}$ in $L^2([0, 1))$ with coefficients $(c_k) \in \ell^2$. This gives

$$f \circ T(x) = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k (x + \vartheta)} = \sum_{k \in \mathbb{Z}} (c_k e^{2\pi i k \vartheta}) e^{2\pi i k x}.$$

The invariance $f = f \circ T$ and the uniqueness of the Fourier coefficients (c_k) implies $c_k = c_k e^{2\pi i k \vartheta}$ for all $k \in \mathbb{Z}$. Due to $e^{2\pi i k \vartheta} \neq 1$ for $k \in \mathbb{Z} \setminus \{0\}$ (ϑ is irrational!), we infer $c_k = 0$ for $k \neq 0$ and f must be $\mathbb{P}$-a.s. constant. By criterion (a) in the lemma above, T is ergodic. This is Weyl's Equidistribution Theorem (1909), which has many connections to number theory, harmonic analysis and pseudo-random number generation.

5.2 Ergodic theorems

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

Proof. For $j = 1, \ldots, n$ we have

$$Y + M_n \circ T \geqslant Y + M_{n-1} \circ T \geqslant Y + S_{j-1} \circ T = S_j.$$

Hence, $M_n(\omega) > 0$ implies $M_n(\omega) = \max_{j=1,\dots,n} S_j(\omega)$ and thus $Y(\omega) + M_n(T(\omega)) \geqslant M_n(\omega)$. We conclude, using $M_n \geqslant 0$ and measure preservation of T, that

$$\mathbb{E}[(Y + M_n \circ T)\mathbf{1}(M_n > 0)] \geqslant \mathbb{E}[M_n \mathbf{1}(M_n > 0)] = \mathbb{E}[M_n] = \mathbb{E}[M_n \circ T].$$

From this $\mathbb{E}[Y\mathbf{1}(M_n > 0)] \geqslant 0$ follows. □

5.14 Theorem (Birkhoff's ergodic theorem, 1931)**.** *Let $X \in L^1$ and T be measure-preserving on $(\Omega, \mathscr{F}, \mathbb{P})$. Then:*

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

If T is even ergodic, then

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

Proof. We put $A_n := \frac{1}{n}\sum_{k=0}^{n-1} X \circ T^k$, $\overline{X} = \limsup_{n\to\infty} A_n$, $\underline{X} = \liminf_{n\to\infty} A_n$ and split the proof in several steps.

$\overline{X}, \underline{X}$ **are (strictly) T-invariant:** We have $\frac{n+1}{n} A_{n+1} = A_n \circ T + \frac{1}{n} X$. From $\frac{1}{n} X \to 0$ we deduce

$$\overline{X} = \limsup_{n\to\infty} A_{n+1} = \limsup_{n\to\infty} \tfrac{n+1}{n} A_{n+1} = \limsup_{n\to\infty} A_n \circ T = \overline{X} \circ T.$$

Analogously, $\underline{X} = \underline{X} \circ T$ follows.

$\overline{X} = \underline{X}$ **$\mathbb{P}$-a.s.:** Apply the maximal lemma to $Y = (X - b)\mathbf{1}(\underline{X} < a, \overline{X} > b)$ for some $a < b$. In the notation of the maximal lemma we have as $n \uparrow \infty$

$$\{M_n > 0\} \uparrow \Big\{\sup_{n\geqslant 1} S_n > 0\Big\} = \Big\{\sup_{n\geqslant 1} \frac{1}{n} S_n > 0\Big\}$$
$$= \Big\{\sup_{n\geqslant 1} (A_n - b)\mathbf{1}(\underline{X} < a, \overline{X} > b) > 0\Big\} = \{\underline{X} < a, \overline{X} > b\}.$$

By the maximal lemma and dominated convergence we find

$$0 \leqslant \mathbb{E}[(X-b)\mathbf{1}(\underline{X} < a, \overline{X} > b)\mathbf{1}(M_n > 0)] \to \mathbb{E}[(X-b)\mathbf{1}(\underline{X} < a, \overline{X} > b)].$$

Consequently, the limit is non-negative and

$$\mathbb{E}[X\mathbf{1}(\underline{X} < a, \overline{X} > b)] \geqslant b\,\mathbb{P}(\underline{X} < a, \overline{X} > b).$$

The analogous argument for $Y = (a - X)\mathbf{1}(\underline{X} < a, \overline{X} > b)$ yields

$$\mathbb{E}[X\mathbf{1}(\underline{X} < a, \overline{X} > b)] \leqslant a\,\mathbb{P}(\underline{X} < a, \overline{X} > b), \text{ implying } \mathbb{P}(\underline{X} < a, \overline{X} > b) = 0.$$

Taking the union of all these null events with $a, b \in \mathbb{Q}$ and $a < b$, we conclude $\mathbb{P}(\underline{X} < \overline{X}) = 0$ and thus $\underline{X} = \overline{X}$ $\mathbb{P}$-a.s.

(A_n) **is uniformly integrable:** Recall that (Z_n) are uniformly integrable if and only if $\sup_n \mathbb{E}[|Z_n|] < \infty$ and

$$\forall \varepsilon > 0 \, \exists \delta > 0 : \mathbb{P}(B) < \delta \Rightarrow \sup_n \mathbb{E}[|Z_n| \mathbf{1}_B] < \varepsilon.$$

Since T is measure-preserving, $X \circ T^k$ has the same distribution as X such that $(X \circ T^k)_{k \geqslant 0}$ is uniformly integrable. For $\varepsilon > 0$ choose $\delta > 0$ such that $\mathbb{P}(B) < \delta$ implies $\mathbb{E}[|X \circ T^k| \mathbf{1}_B] < \varepsilon$. Then for events B with $\mathbb{P}(B) < \delta$

$$\sup_n \mathbb{E}[|A_n| \mathbf{1}_B] \leqslant \sup_n \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}[|X \circ T^k| \mathbf{1}_B] < \varepsilon$$

follows. Noting $\mathbb{E}[|A_n|] \leqslant \mathbb{E}[|X|]$, we conclude that (A_n) is uniformly integrable.

$\underline{X} = \overline{X} = \mathbb{E}[X \mid \mathscr{I}_T]$ **and L^1-convergence:** Since $A_n \to \underline{X} = \overline{X}$ $\mathbb{P}$-a.s. and (A_n) is uniformly integrable, the convergence also holds in L^1. Moreover, L^1-convergence implies L^1-convergence of the conditional expectations:

$$\mathbb{E}\Big[|\mathbb{E}[A_n \mid \mathscr{I}_T] - \mathbb{E}[\overline{X} \mid \mathscr{I}_T]|\Big] \leqslant \mathbb{E}[\mathbb{E}[|A_n - \overline{X}| \mid \mathscr{I}_T]] = \|A_n - \overline{X}\|_{L^1} \to 0.$$

Now, $\mathbb{E}[A_n \mid \mathscr{I}_T] = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}[X \circ T^k \mid \mathscr{I}_T] = \mathbb{E}[X \mid \mathscr{I}_T]$ holds because of $\mathbb{E}[X \circ T^k \mid \mathscr{I}_T] = \mathbb{E}[X \mid \mathscr{I}_T]$: for $A \in \mathscr{I}_T$ we check

$$\mathbb{E}[\mathbb{E}[X \circ T^k \mid \mathscr{I}_T] \mathbf{1}_A] = \mathbb{E}[(X \circ T^k) \mathbf{1}_A] = \mathbb{E}[(X \circ T^k)(\mathbf{1}_A \circ T^k)] = \mathbb{E}[X \mathbf{1}_A].$$

From $\mathbb{E}[X \mid \mathscr{I}_T] = \mathbb{E}[A_n \mid \mathscr{I}_T] \to \mathbb{E}[\overline{X} \mid \mathscr{I}_T] = \overline{X}$ in L^1 we infer $\underline{X} = \overline{X} = \mathbb{E}[X \mid \mathscr{I}_T]$ $\mathbb{P}$-a.s.

T **ergodic implies $\mathbb{E}[X \mid \mathscr{I}_T] = \mathbb{E}[X]$ $\mathbb{P}$-a.s.:** The above lemma asserts that the left-hand side is $\mathbb{P}$-a.s. constant. Taking expectations, the constant must be $\mathbb{E}[X]$.

□

5.15 Theorem (L^p-ergodic theorem, L^2-version by von Neumann 1932)**.** *For $X \in L^p$, $p \geqslant 1$, and measure-preserving T on $(\Omega, \mathscr{F}, \mathbb{P})$ we have*

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

Proof. ► Exercise □

5.16 Corollary. *Let $(X_k, k \geqslant 0)$ be an ergodic process in L^1 (i.e. $X_k \in L^1$ and the left shift on $(\mathbb{R}^{\mathbb{N}_0}, \mathfrak{B}_{\mathbb{R}}^{\otimes \mathbb{N}_0}, \mathbb{P}^X)$ is ergodic). Then*

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

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

5.17 Corollary. *A measure-preserving transformation T on $(\Omega, \mathscr{F}, \mathbb{P})$ is ergodic if and only if*

$$\forall A, B \in \mathscr{F}: \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}(A \cap T^{-k}(B)) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Proof. ▶ Exercise □

5.18 Remark. If even $\lim_{k\to\infty} \mathbb{P}(A \cap T^{-k}(B)) = \mathbb{P}(A)\,\mathbb{P}(B)$ for all $A, B \in \mathscr{F}$ holds (*A and $T^{-k}(B)$ are asymptotically independent*), then T is called mixing.

5.3 The structure of the invariant measures

5.19 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 T-invariant probability measures is denoted by $\mathscr{M}_T$.

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

Proof. For $\mu_1, \mu_2 \in \mathscr{M}_T$, $\alpha \in (0,1)$ consider $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. Then μ is again a probability measure and satisfies for $A \in \mathscr{F}$

$$\mu(T^{-1}(A)) = \alpha\mu_1(T^{-1}(A)) + (1-\alpha)\mu_2(T^{-1}(A)) = \alpha\mu_1(A) + (1-\alpha)\mu_2(A) = \mu(A).$$

Hence, $\mu \in \mathscr{M}_T$ and $\mathscr{M}_T$ is convex. □

5.21 Proposition. *If μ and ν are distinct ergodic measures, then $\mu \perp \nu$.*

Proof. Choose $A \in \mathscr{F}$ with $\mu(A) \neq \nu(A)$. The ergodic theorem implies

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A \circ T^k \to \begin{cases} \mu(A), & \mu\text{-a.s.}, \\ \nu(A), & \nu\text{-a.s.} \end{cases}$$

Hence, for $\Omega_\mu = \{\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A \circ T^k = \mu(A)\}$ we have $\mu(\Omega_\mu) = 1$, $\nu(\Omega_\mu) = 0$. □

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

Proof. Suppose first that μ is not ergodic. Then there is some $A \in \mathscr{F}$ with $T^{-1}(A) = A$ and $\mu(A) \in (0,1)$. Introduce the probability measures $\mu_1 = \mu(\bullet \mid A)$, $\mu_2 = \mu(\bullet \mid A^C)$. Then $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ holds with $\alpha = \mu(A) \in (0,1)$. Moreover,

$$\mu_1(T^{-1}(B)) = \frac{\mu(T^{-1}(B) \cap A)}{\mu(A)} = \frac{\mu(T^{-1}(B \cap A))}{\mu(A)} = \frac{\mu(B \cap A)}{\mu(A)} = \mu_1(B)$$

for $B \in \mathscr{F}$ shows that μ_1 is T-invariant. Similarly, μ_2 is T-invariant. Therefore μ is a strict convex combination of $\mu_1, \mu_2 \in \mathscr{M}_T$ and thus not extremal.

Next, let us show that $\mu, \nu \in \mathscr{M}_T$, $\nu \ll \mu$ and μ ergodic implies $\mu = \nu$. Indeed, the ergodic theorem yields μ-a.s. and because of $\nu \ll \mu$ also ν-a.s.

$$\forall A \in \mathscr{F}: \; \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A \circ T^k \to \mu(A).$$

Dominated convergence yields $\mu = \nu$:

$$\forall A \in \mathscr{F}: \; \nu(A) = \int \Big(\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A \circ T^k \Big) \, d\nu \to \int \mu(A) \, d\nu = \mu(A).$$

Now, if μ is ergodic and $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ holds with $\mu_1, \mu_2 \in \mathscr{M}_T$, $\alpha \in (0,1)$, then we have $\mu_1, \mu_2 \ll \mu$, thus implying $\mu_1 = \mu_2 = \mu$. This shows that μ is an extremal point of $\mathscr{M}_T$. □

5.23 Corollary. *If T possesses exactly one invariant probability measure μ, then μ is ergodic.*

Proof. μ is an extremal point of $\mathscr{M}_T = \{\mu\}$. □

5.4 Application to Markov chains

In this section $(X_n, \, n \geqslant 0)$ always denotes a time-homogeneous Markov chain with discrete state space $(S, \mathcal{P}(S))$, realized as coordinate process on $\Omega = S^{\mathbb{N}_0}$ with σ-algebra $\mathscr{F} = \mathcal{P}(S)^{\otimes \mathbb{N}_0}$, filtration $\mathscr{F}_n = \mathscr{F}_n^X$ and measure $\mathbb{P}_\mu$, where μ denotes the initial distribution of X_0. We write short $\mathbb{P}_x := \mathbb{P}_{\delta_x}$.

5.24 Definition. For $x \in S$ let $T_x := \inf\{n \geqslant 1 \,|\, X_n = x\}$ be the first time of visiting x after starting. A state $x \in S$ is called recurrent (rekurrent) if $\mathbb{P}_x(T_x < \infty) = 1$ and transient (transient) otherwise. A recurrent state $x \in S$ is called positive-recurrent if $\mathbb{E}_x[T_x] < \infty$, otherwise it is called null-recurrent. The Markov chain (X_n) is irreducible if $\mathbb{P}_x(T_y < \infty) > 0$ holds for any $x, y \in S$.

5.25 Example.

(a) Consider the simple random walk $(S_n)_{n \geqslant 0}$ on $\mathbb{Z}$ with $S_n = S_{n-1} + X_n$ and (X_n) independent with $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = -1) = 1 - p$ for some $p \in (0,1)$. This forms an irreducible Markov chain with $S_0 = x$ under $\mathbb{P}_x$: $\mathbb{P}_x(T_y = y - x) = p^{y-x} > 0$ for $y > x$, $\mathbb{P}_x(T_y = x - y) = (1-p)^{x-y} > 0$ for $y < x$ and $\mathbb{P}_x(T_x = 2) = p(1-p) > 0$ for any x. By translation invariance $\mathbb{P}_x(S_n = m + x) = \mathbb{P}_0(S_n = m)$, $m \in \mathbb{Z}$, all states are either recurrent or transient. From the above analysis of the simple random walk we can conclude that $\mathbb{P}_0(T_0 < \infty) = 1$ holds if and only if the walk is symmetric ($p = 1/2$). In that case it is null recurrent because $\mathbb{E}_0[T_0] = \infty$ and thus also $\mathbb{E}_x[T_x] = \infty$ for all $x \in \mathbb{Z}$. For $p \neq 1/2$ the simple random walk is transient.

(b) The Markov chain on $S=\{1,2\}$ with transition probabilities $p_{11}=p_{22}=p$, $p_{12}=p_{21}=1-p$ is irreducible for $p\in[0,1)$. Both states are positive-recurrent for any $p\in[0,1]$, e.g.

$$\mathbb{E}_1[T_1]=p+\sum_{k=2}^{\infty}(1-p)^2p^{k-2}k=p+(2-p)\mathbf{1}(p<1)=2\mathbf{1}(p<1)+\mathbf{1}(p=1).$$

The case $p=1$ here furnishes a trivial example of a reducible Markov chain that is positive-recurrent. For $p\in[0,1)$ there is one invariant initial distribution, namely $\mu(\{1\})=\mu(\{2\})=1/2$. For $p=1$ any initial distribution is invariant.

5.26 Proposition. *An irreducible Markov chain on a finite state space S has only positive-recurrent states: $\forall\, x\in S:\ \mathbb{E}_x[T_x]<\infty$.*

Proof (presented in class). Since the chain is irreducible and S is finite, there are $r\in\mathbb{N}$ and $\varepsilon>0$ such that $\mathbb{P}_y(T_x\leqslant r)\geqslant\varepsilon$ for all $x,y\in S$ (pick suitable $r_{xy}\in\mathbb{N},\varepsilon_{xy}>0$ for each pair (x,y) and take $r=\max_{x,y}r_{xy}$, $\varepsilon=\min_{x,y}\varepsilon_{xy}$). The Markov property gives for $k\geqslant 2$

$$\begin{aligned}\mathbb{P}_x(T_x>kr)&=\sum_{y\in S\setminus\{x\}}\mathbb{P}_x(\forall\, 1\leqslant n\leqslant kr:\ X_n\neq x,\ X_{(k-1)r}=y)\\&=\sum_{y\in S\setminus\{x\}}\mathbb{P}_x(\forall\, 1\leqslant n\leqslant (k-1)r:\ X_n\neq x,\ X_{(k-1)r}=y)\,\mathbb{P}_y(T_x>r)\\&\leqslant\mathbb{P}_x(T_x>(k-1)r)(1-\varepsilon).\end{aligned}$$

Recursively, we obtain $\mathbb{P}_x(T_x>kr)\leqslant(1-\varepsilon)^k$ for all $k\in\mathbb{N}$ and thus $\mathbb{E}_x[T_x]=\sum_{n=0}^{\infty}\mathbb{P}_x(T_x>n)\leqslant\sum_{r=0}^{\infty}r\,\mathbb{P}_x(T_x>kr)<\infty$. □

5.27 Proposition. *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 defines an invariant initial distribution.

Proof. From $\mathbb{E}_x[T_x]<\infty$ it follows by the Fubini-Tonelli Theorem that

$$\sum_{y\in S}\mu(\{y\})=\frac{\mathbb{E}_x[\sum_{n=0}^{T_x-1}\sum_{y\in S}\mathbf{1}(X_n=y)]}{\mathbb{E}_x[T_x]}=1$$

and μ defines a probability measure. We have to prove $\mathbb{P}_\mu(X_1=z)=\mu(\{z\})$ for all $z\in S$. Putting $q_n(x,y):=\mathbb{P}_x(X_n=y,T_x>n)$ and denoting the transition probability $\mathbb{P}_y(X_1=z)$ by p_{yz}, it suffices to show

$$\forall z\in S:\ \sum_{y\in S}\sum_{n=0}^{\infty}q_n(x,y)p_{yz}=\sum_{n=0}^{\infty}q_n(x,z).$$

The Markov property yields $q_n(x,y)p_{yz} = \mathbb{P}_x(X_n = y, T_x > n, X_{n+1} = z)$ because $\{T_x > n\} = \{X_k \neq x, k = 1,\dots,n\}$ is $\mathscr{F}_n^X$-measurable. For $z \neq x$ this gives

$$\sum_{y\in S} q_n(x,y)p_{yz} = \mathbb{P}_x(T_x > n, X_{n+1} = z) = \mathbb{P}_x(T_x > n+1, X_{n+1} = z) = q_{n+1}(x,z).$$

Since $q_0(x,z) = 0$ for $x \neq z$ this shows $\sum_{y\in S}\sum_{n=0}^\infty q_n(x,y)p_{yz} = \sum_{n=0}^\infty q_n(x,z)$. In the case $z = x$ we have $\{T_x > n, X_{n+1} = x\} = \{T_x = n+1\}$ and $q_0(x,x) = 1$, $q_n(x,x) = 0$, $n \geqslant 1$. The above argument gives

$$\sum_{y\in S}\sum_{n=0}^\infty q_n(x,y)p_{yx} = \sum_{n=0}^\infty \mathbb{P}_x(T_x > n, X_{n+1} = x) = 1 = \sum_{n=0}^\infty q_n(x,x).$$

Alternatively, for $z = x$ we could argue via the first part which by summing over all $z \neq x$ gives $\mathbb{P}_\mu(X_1 \neq x) = \mu(S \setminus \{x\})$. □

5.28 Lemma. *An irreducible Markov chain has at most one invariant initial distribution μ. If it exists, it satisfies $\mu(\{x\}) > 0$ for all $x \in S$.*

Proof. Suppose μ and ν are distinct invariant initial distributions. Choose weights (w_n) with $w_n > 0$ and $\sum_n w_n = 1$. Define $\tilde{p}_{xy} = \sum_{n=1}^\infty w_n p_{xy}^{(n)}$ with the n-step transition probabilities $p_{xy}^{(n)} = \mathbb{P}_x(X_n = y)$. By irreducibility, $\tilde{p}_{xy} > 0$ holds for all $x,y \in S$. Introduce $\pi_x = \mu_x - \nu_x$ with $\mu_x = \mu(\{x\})$, $\nu_x = \nu(\{x\})$. Then for all $y \in S$

$$\sum_{x\in S}\pi_x\tilde{p}_{xy} = \sum_{n=1}^\infty w_n \sum_{x\in S}(\mu_x - \nu_x)p_{xy}^{(n)} = \sum_{n=1}^\infty w_n(\mu_y - \nu_y) = \pi_y.$$

Since $\mu \neq \nu$ implies that there are $x_1, x_2 \in S$ with $\pi_{x_1} < 0$, $\pi_{x_2} > 0$, this yields the strict inequality

$$\sum_{y\in S}|\pi_y| = \sum_{y\in S}\Big|\sum_{x\in S}\pi_x\tilde{p}_{xy}\Big| < \sum_{x,y\in S}|\pi_x|\tilde{p}_{xy} = \sum_{x\in S}|\pi_x|.$$

This contradiction proves that an invariant initial distribution is unique.

Choose some $x \in S$ with $\mu(\{x\}) > 0$. By irreducibility, for any $y \in S$ there is $n \in \mathbb{N}$ with $\mathbb{P}_x(X_n = y) > 0$. This implies $\mu(\{y\}) = \mathbb{P}_\mu(X_n = y) \geqslant \mu(\{x\})\,\mathbb{P}_x(X_n = y) > 0$, as asserted. □

5.29 Lemma. *If $x \in S$ is a recurrent state of an irreducible Markov chain, then $\mathbb{P}_y(T_x < \infty) = 1$ holds for all $y \in S$.*

Proof. By irreducibility, there is a minimal $n \in \mathbb{N}$ with $\mathbb{P}_x(X_n = y) > 0$ for $y \in S$. Then there exist $x_1,\dots,x_n \in S$ with $x_n = y$ and $x_k \neq x$, $k = 1,\dots,n-1$, such that $\mathbb{P}_x(X_1 = x_1,\dots,X_n = x_n) > 0$. Applying the Markov property in the last step, we obtain for $y \neq x$

$$\begin{aligned}
0 = \mathbb{P}_x(T_x = \infty) &\geqslant \mathbb{P}_x(X_1 = x_1,\dots,X_n = x_n, T_x = \infty)\\
&= \mathbb{P}_x(X_1 = x_1,\dots,X_n = x_n, X_{n+k} \neq x, k \geqslant 1)\\
&= P_x(X_1 = x_1,\dots,X_n = x_n)\,\mathbb{P}_{x_n}(X_k \neq x, k \geqslant 1).
\end{aligned}$$

This shows $\mathbb{P}_y(T_x = \infty) = 0$. □

5.30 Remark. One can show further that then also $\mathbb{P}_y(T_y < \infty) = 1$ and in an irreducible Markov chain either all states are recurrent or all are transient, see Klenke.

5.31 Theorem. *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 Proposition 5.27.*

Proof. We know already that μ is the unique invariant initial distribution of (X_n). It remains to show that $A \in \mathcal{P}(A)^{\otimes \mathbb{N}_0}$, $T^{-1}(A) = A$ for the left shift T on $S^{\mathbb{N}_0}$ implies $\mathbb{P}_\mu(A) \in \{0,1\}$.

Suppose $\mathbb{P}_\mu(A) > 0$. Then for any $x_0, \dots, x_n \in S$ by $A = T^{-n}A$ and the Markov property

$$\begin{aligned}\mathbb{P}_\mu(X_0 = x_0, \dots, X_n = x_n \,|\, A) &= \frac{\mathbb{P}_\mu(X_0 = x_0, \dots, X_n = x_n, (X_{n+k})_{k\geqslant 0} \in A)}{\mathbb{P}_\mu(A)} \\ &= \mu(\{x_0\}) \prod_{k=1}^n p_{x_{k-1},x_k} \frac{\mathbb{P}_{x_n}(A)}{\mathbb{P}_\mu(A)}.\end{aligned}$$

Below, we shall see that $\mathbb{P}_{x_n}(A) = \mathbb{P}_\mu(A)$ holds such that

$$\mathbb{P}_\mu(X_0 = x_0, \dots, X_n = x_n \,|\, A) = \mathbb{P}_\mu(X_0 = x_0, \dots, X_n = x_n)$$

follows. This implies, however, that $\mathbb{P}_\mu(\bullet \,|\, A)$ and $\mathbb{P}_\mu$ coincide on all cylinder sets of $\mathcal{P}(S)^{\otimes \mathbb{N}_0}$. By the uniqueness theorem, they coincide everywhere and in particular we have $\mathbb{P}_\mu(A) = \mathbb{P}_\mu(A \,|\, A) = 1$. This proves $\mathbb{P}_\mu(A) \in \{0,1\}$.

Finally, we set $f(y) := \mathbb{P}_y(A)$, $y \in S$, and show $f(y) = f(x)$ constant. First note that by the invariance of A and the Markov property

$$f(y) = \mathbb{P}_y(T^{-1}(A)) = \mathbb{P}_y((X_{k+1})_{k\geqslant 0} \in A) = \sum_{z\in S} p_{y,z}\, \mathbb{P}_z(A) = \mathbb{E}_y[f(X_1)],\ y \in S.$$

(A function f with this property is called harmonic). This implies that $M_n := f(X_n)$ is a martingale with respect to $(\mathscr{F}_n^X)$ under $\mathbb{P}_y$:

$$\mathbb{E}_y[M_n \,|\, \mathscr{F}_{n-1}^X] = \mathbb{E}_y[f(X_n) \,|\, X_0, \dots, X_{n-1}] = \mathbb{E}_{X_{n-1}}[f(X_1)] = f(X_{n-1}) = M_{n-1}.$$

Since T_x is a $\mathbb{P}_y$-a.s. finite stopping time by the above lemma and (M_n) is a bounded martingale ($M_n \in [0,1]$), optional stopping yields

$$f(x) = \mathbb{E}_y[M_{T_x}] = \mathbb{E}_y[M_0] = f(y).$$

Consequently, $\mathbb{P}_\mu(A) = \sum_{z\in S} \mu(\{z\})\, \mathbb{P}_z(A) = \mathbb{P}_y(A)$ holds for any $y \in S$. $\square$

5.32 Theorem. *If an irreducible Markov chain $(X_n,\, n \geqslant 0)$ is ergodic with invariant initial distribution μ, then all its states are positive-recurrent and $\mu(\{x\}) = 1/\,\mathbb{E}_x[T_x]$, $x \in S$, holds.*

Proof (not presented). From the lemma above we know $\mu(\{x\}) > 0$ for all $x \in S$. The ergodic theorem yields

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}(X_k = x) \to \mu(\{x\}) > 0 \ \mathbb{P}_\mu\text{-a.s.}$$

This implies $T_x < \infty$ $\mathbb{P}_\mu$-a.s. and, using $\mathbb{P}_\mu(T_x = \infty) = 0$ as well as $\mu(\{x\}) > 0$, we have $\mathbb{P}_x(T_x = \infty) = 0$ for all $x \in S$. Hence, all states $x \in S$ are recurrent.

Define $\sigma_x^n := \sup\{m \leqslant n \,|\, X_m = y\} \in \{0, \dots, n\} \cup \{-\infty\}$ (last visit to x before n, no stopping time). Then for $k = 0, \dots, n$

$$\begin{aligned}
\mathbb{P}_\mu(\sigma_x^n = k) &= \mathbb{P}_\mu(X_k = x, X_{k+1} \neq x, \dots, X_n \neq x) \\
&= \mu(\{x\})\, \mathbb{P}_x(X_1 \neq x, \dots, X_{n-k} \neq x) \\
&= \mu(\{x\})\, \mathbb{P}_x(T_x \geqslant n - k + 1).
\end{aligned}$$

By the above lemma $\mathbb{P}_x(T_x < \infty) = 1$ implies $\mathbb{P}_y(T_x < \infty) = 1$ for any $y \in S$ and therefore $\mathbb{P}_\mu(T_x < \infty) = 1$. This yields for $n \to \infty$

$$\begin{aligned}
1 &= \sum_{k=0}^{n} \mathbb{P}_\mu(\sigma_x^n = k) + \mathbb{P}_\mu(\sigma_x^n = -\infty) \\
&= \mu(\{x\}) \sum_{k=0}^{n} \mathbb{P}_x(T_x \geqslant n - k + 1) + \mathbb{P}_\mu(T_x \geqslant n + 1) \\
&\to \mu(\{x\}) \sum_{j=1}^{\infty} \mathbb{P}_x(T_x \geqslant j) = \mu(\{x\})\, \mathbb{E}_x[T_x].
\end{aligned}$$

This proves $\mathbb{E}_x[T_x] = \frac{1}{\mu(\{x\})} < \infty$ for all $x \in S$ and all states are positive-recurrent. □

5.33 Remark.

(a) Slightly more generally, for an irreducible Markov chain we have equivalence between:

(i) all states are positive-recurrent;

(ii) there is a positive-recurrent state;

(iii) there is an invariant initial distribution;

(iv) there is exactly one invariant initial distribution;

(v) the Markov chain is ergodic under the invariant initial distribution.

For a finite state space S all five properties are always satisfied. See Klenke for further details.

(b) If we ask for more than an analogue of the law of large numbers, e.g. convergence in distribution or a central limit theorem, then ergodicity is clearly not sufficient. A toy example is given by the state space $S = \{1, \dots, m\}$ and a cyclic permutation $T = (1\,2 \cdots m)$. Under the uniform distribution $\mu(\{i\}) = 1/m$, $i = 1, \dots, m$, T is ergodic, and the fluctuations of the relative frequencies satisfy $|\frac{1}{n}\sum_{k=0}^{n-1}(\mathbf{1}(T^k(x) = y) - 1/m)| \leqslant 1/n$ for $x, y \in S$. This is much smaller than the order $1/\sqrt{n}$ from the central limit theorem. For Markov chains it leads to the concept of aperiodic chains and more generally to mixing or weak dependence conditions for stationary processes.

6 Weak convergence

6.1 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 $C_b(S)$.

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

6.2 Example. For $x_n \to x$ in S the point measures δ_{x_n} converge weakly to δ_x. Note that for $x_n \neq x$, $n \geqslant 1$, we have $0 = \delta_{x_n}(\{x\} \not\xrightarrow{w} \delta_x(\{x\}) = 1$. In general, we cannot expect that $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ implies $\mathbb{P}_n(A) \to \mathbb{P}(A)$ for an event A.

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

Proof. For $f \in C_b(T)$ we have $f \circ g \in C_b(S)$. Hence, $X_n \xrightarrow{d} X$ implies $\mathbb{E}[f(g(X_n))] \to \mathbb{E}[f(g(X))]$ and therefore $g(X_n) \xrightarrow{d} g(X)$. □

6.4 Example. If real-valued random variables X_n satisfy $X_n \xrightarrow{d} N(0,1)$, then $aX_n + b \xrightarrow{d} N(b, a^2)$ and $X_n^2 \xrightarrow{d} \chi_1^2$ follow.

6.5 Definition. Let $(X_n), X$ be random variables with values in a Polish space $(S, \mathfrak{B}_S)$. Then (X_n) converges in probability or stochastically to X, notation $X_n \xrightarrow{\mathbb{P}} X$ if

$$\forall \varepsilon > 0: \lim_{n\to\infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0.$$

6.6 Remark. We need that S is Polish to justify that $d(X_n, X)$ is a real-valued random variable ▶ Exercise . Note that $X_n \xrightarrow{\mathbb{P}} X$ is therefore equivalent to the stochastic convergence $d(X_n, X) \xrightarrow{\mathbb{P}} 0$ for real-valued random variables. This allows to transfer many results for stochastic convergence from $\mathbb{R}$ to general Polish S. In particular, $X_n \xrightarrow{\mathbb{P}} X$ implies $X_n \xrightarrow{d} X$ ▶ Exercise .

6.7 Theorem (Portmanteau Lemma, Alexandrov 1940)**.** *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* $\mathbb{P}(\partial A) = 0 :\ \lim_{n\to\infty} \mathbb{P}_n(A) = \mathbb{P}(A)$*.*

6.8 Remark. The topological boundary of a set A is defined as $\partial A = \bar{A} \setminus A^\circ$ where $\bar{A}$ is the closure and A° is the interior of A.

6.9 Example. For probability measures on $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ this shows that $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ implies $F_n(x) := \mathbb{P}_n((-\infty, x]) \to \mathbb{P}((-\infty, x]) = F(x)$ for all $x \in \mathbb{R}$ with $\mathbb{P}(\partial(-\infty, x]) = \mathbb{P}(\{x\}) = 0$. These are exactly the continuity points of the distribution function $\mathbb{P}$ and we find back the result from Stochastics I on the pointwise convergence of the distribution functions at those continuity points.

Proof. (a)$\Rightarrow$(b): Let $U \subseteq S$ be open, $F = S \setminus U$. Then $x \mapsto \operatorname{dist}_F(x) = \inf_{y\in F} d(x, y)$ is continuous, which follows from triangle inequality. Therefore $f_m(x) := (m \operatorname{dist}_F(x)) \wedge 1$, $m \in \mathbb{N}$, lies in $C_b(S)$ and satisfies $f_m \uparrow \mathbf{1}_U$ as $m \uparrow \infty$. From (a) we deduce for any $m \in \mathbb{N}$

$$\liminf_{n\to\infty} \mathbb{P}_n(U) \geqslant \liminf_{n\to\infty} \int f_m \, d\,\mathbb{P}_n = \int f_m \, d\,\mathbb{P}\,.$$

Monotone convergence gives $\lim_{m\to\infty} \int f_m d\,\mathbb{P} = \int \mathbf{1}_U d\,\mathbb{P} = \mathbb{P}(U)$ and thus (b).

(b) $\iff$ (c) follows directly by taking complements.

(b,c)$\Rightarrow$(d): For all $A \in \mathfrak{B}_S$ we have by (b) and (c)

$$\begin{aligned}\mathbb{P}(A^\circ) &\leqslant \liminf_{n\to\infty} \mathbb{P}_n(A^\circ) \leqslant \liminf_{n\to\infty} \mathbb{P}_n(A) \\ &\leqslant \limsup_{n\to\infty} \mathbb{P}_n(A) \leqslant \limsup_{n\to\infty} \mathbb{P}_n(\bar{A}) \leqslant \mathbb{P}(\bar{A}).\end{aligned}$$

If $\mathbb{P}(\partial A) = \mathbb{P}(\bar{A}) - \mathbb{P}(A^\circ) = 0$ holds, then we have equality everywhere and (d) follows.

(d)$\Rightarrow$ (a): Let $f \in C_b(S)$. Since the preimages $(f^{-1}(\{y\}))_{y\in\mathbb{R}}$ are pairwise disjoint, there are at most countably many y with $\mathbb{P}(f^{-1}(\{y\})) > 0$. Without loss of generality assume $\mathbb{P}(f^{-1}(\{0\})) = 0$ and consider

$$B_{k,\varepsilon} = f^{-1}([k\varepsilon, (k+1)\varepsilon)), \quad k \in \mathbb{Z},\, \varepsilon > 0.$$

By continuity of f, we have $\partial B_{k,\varepsilon} \subseteq f^{-1}(\{k\varepsilon\}) \cup f^{-1}(\{(k+1)\varepsilon\})$ and there is a sequence $\varepsilon_m \downarrow 0$ with $\mathbb{P}(\partial B_{k,\varepsilon_m}) = 0$ for all m, k. Using

$$\sum_k k\varepsilon_m \mathbf{1}_{B_{k,\varepsilon_m}} \leqslant f \leqslant \sum_k (k+1)\varepsilon_m \mathbf{1}_{B_{k,\varepsilon_m}},$$

where k runs through the finite set $\{k \in \mathbb{Z} \mid |k| \leqslant \|f\|_\infty/\varepsilon_m + 2\}$, we obtain

$$\begin{aligned}\int f\,d\mathbb{P} - \varepsilon_m &\leqslant \sum_k k\varepsilon_m\,\mathbb{P}(B_{k,\varepsilon_m}) = \lim_{n\to\infty}\sum_k k\varepsilon_m\,\mathbb{P}_n(B_{k,\varepsilon_m})\\ &\leqslant \liminf_{n\to\infty}\int f\,d\mathbb{P}_n \leqslant \limsup_{n\to\infty}\int f\,d\mathbb{P}_n \leqslant \lim_{n\to\infty}\sum_k (k+1)\varepsilon_m\,\mathbb{P}_n(B_{k,\varepsilon_m})\\ &= \sum_k (k+1)\varepsilon_m\,\mathbb{P}(B_{k,\varepsilon_m}) \leqslant \int f\,d\mathbb{P} + \varepsilon_m.\end{aligned}$$

With $\varepsilon_m \downarrow 0$ we obtain equality everywhere in the limit and $\lim_{n\to\infty}\int f\,d\mathbb{P}_n = \int f\,d\mathbb{P}$ follows. □

6.10 Remark. The proof of (a)⇒(b) only uses weak convergence to ensure that integrals over the functions f_m converge, which are truncations of dist_F-functions for closed sets F. By the equivalence statement, this is already sufficient for weak convergence, that is convergence of integrals over all $C_b(S)$-functions. In particular, it suffices to check convergence over bounded Lipschitz-functions. Next we formulate and prove this separately because the approximation argument used is also valuable for other properties of weak convergence.

6.11 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.*

Proof. Below we shall prove that for any $f \in C_b(S)$ there is a sequence (g_k) of Lipschitz-continuous functions with $g_k \uparrow f$ and $g_k \geqslant -\|f\|_\infty$. By assumption, this gives for any $k \in \mathbb{N}$:

$$\liminf_{n\to\infty}\int f\,d\mathbb{P}_n \geqslant \liminf_{n\to\infty}\int g_k\,d\mathbb{P}_n = \int g_k\,d\mathbb{P}\,.$$

By monotone convergence, $\int g_k\,d\mathbb{P} \uparrow \int f\,d\mathbb{P}$ holds such that $\liminf_{n\to\infty}\int f\,d\mathbb{P}_n \geqslant \int f\,d\mathbb{P}$. The same argument applied to $-f$ yields $\liminf_{n\to\infty}\int(-f)\,d\mathbb{P}_n \geqslant \int(-f)\,d\mathbb{P}$ and thus $\lim_{n\to\infty}\int f\,d\mathbb{P}_n = \int f\,d\mathbb{P}$.

To construct (g_k) assume without loss of generality $f \geqslant 0$ and set

$$h_{m,r}(x) = (m\,\mathrm{dist}_{\{y\in S\,|\,f(y)\leqslant r\}}(x)) \wedge r, \quad r > 0, m \in \mathbb{N}\,.$$

By triangle inequality, we have $|h_{m,r}(x) - h_{m,r}(y)| \leqslant md(x,y)$ for $x, y \in S$ and $h_{m,r}$ is a bounded Lipschitz-continuous function. By construction, we have $f(x) \leqslant r \Rightarrow h_{m,r}(x) = 0$ and $h_{m,r}(x) \in [0,r]$. This shows $0 \leqslant h_{m,r} \leqslant f$. For $f(x) - \varepsilon < r < f(x)$ we have $\lim_{m\to\infty} h_{m,r}(x) = r > f(x) - \varepsilon$, implying $f(x) = \sup_{m\in\mathbb{N}, r>0, r\in\mathbb{Q}} h_{m,r}(x)$. Let $(m_k, r_k)_{k\geqslant 1}$ be an enumeration of $\{(m,r) \mid m \in \mathbb{N}, r > 0, r \in \mathbb{Q}\}$, then $g_k := \max(h_{m_1,r_1}, \ldots, h_{m_k,r_k})$ is Lipschitz-continuous with $g_k \geqslant 0$ and $g_k \uparrow f$. □

6.12 Lemma (Slutsky, 1925)**.** *Let (S,d) be Polish. We have for $(S, \mathfrak{B}_s)$-valued random variables (X_n), (Y_n)*

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

Proof. Let $f \in C_b(S)$ be Lipschitz-continuous with constant L. Then for any $\varepsilon > 0$

$$\begin{aligned}\limsup_{n\to\infty}|\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)]| &\leqslant \limsup_{n\to\infty}\mathbb{E}[|f(X_n) - f(Y_n)|] \\ &\leqslant \limsup_{n\to\infty}\mathbb{E}[|f(X_n) - f(Y_n)|\mathbf{1}(d(X_n, Y_n) > \varepsilon)] + L\varepsilon \\ &\leqslant 2\|f\|_\infty \limsup_{n\to\infty}\mathbb{P}((d(X_n, Y_n) > \varepsilon) + L\varepsilon = L\varepsilon.\end{aligned}$$

With $\varepsilon \downarrow 0$ we conclude $\lim_{n\to\infty}|\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)]| = 0$. This yields

$$\lim_{n\to\infty}\mathbb{E}[f(Y_n)] = \lim_{n\to\infty}\mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

By the preceding proposition, we obtain $Y_n \xrightarrow{d} X$. □

6.13 Corollary (Slutsky). *Let (S, d) be Polish. If $(S, \mathfrak{B}_s)$-valued random variables satisfy $Y_n \xrightarrow{\mathbb{P}} a$, $a \in S$ deterministic, and $X_n \xrightarrow{d} X$, then $(X_n, Y_n) \xrightarrow{d} (X, a)$ holds. In particular, for $S = \mathbb{R}$ we have $X_n Y_n \xrightarrow{d} aX$ and $X_n + Y_n \xrightarrow{d} X + a$.*

Proof. Note that the space S^2, equipped with the product metric $d_2((x_1, x_2), (y_1, y_2)) = d(x_1, y_1) + d(x_2, y_2)$, is again Polish, cf. Section 2.2. On one hand, we have $d_2((X_n, Y_n), (X_n, a)) = d(Y_n, a) \xrightarrow{\mathbb{P}} 0$ due to $Y_n \xrightarrow{\mathbb{P}} a$. On the other hand, $(X_n, a) \xrightarrow{d} (X, a)$ follows from

$$f \in C_b(S^2) \Rightarrow f(\bullet, a) \in C_b(S) \Rightarrow \mathbb{E}[f(X_n, a)] \to \mathbb{E}[f(X, a)].$$

Applying the Slutsky Lemma in S^2, we conclude $(X_n, Y_n) \xrightarrow{d} (X, a)$. Noting that $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$ are both continuous mappings from $\mathbb{R}^2$ to $\mathbb{R}$, the continuous mapping theorem shows that $(X_n, Y_n) \xrightarrow{d} (X, a)$ implies $X_n + Y_n \xrightarrow{d} X + a$, $X_n Y_n \xrightarrow{d} Xa$. □

6.2 Tightness

6.14 Definition. A family $(\mathbb{P}_i)_{i\in I}$ of probability measures on $(S, \mathfrak{B}_S)$ is called <u>(weakly) relatively compact</u> if each sequence $(\mathbb{P}_{i_k})_{k\geqslant 1}$ has a weakly convergent subsequence. This means that there is a probability measure $\mathbb{P}$ and a subsequence (i_{k_l}) such that $\mathbb{P}_{i_{k_l}} \xrightarrow{w} \mathbb{P}$ as $l \to \infty$. The family $(\mathbb{P}_i)_{i\in I}$ is <u>(uniformly) tight</u> (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 $i \in I$.

6.15 Remark. One can show that the set $M(S)$ of all probability measures on a Polish space $(S, \mathfrak{B}_s)$ under weak convergence is again Polish, see ► Exercise and Dudley. In particular, sequential compactness and compactness are identical and 'relatively compact' just means that the closure is compact.

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

Proof. Similar to Ulam's Theorem ► Exercise . □

6.17 Theorem (Prohorov, 1956)**.** *Any tight family of probability measures on a Polish space is weakly relatively compact.*

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

6.3 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. Then C forms a Polish space. The Borel σ-algebra $\mathfrak{B}_C$ is generated by the coordinate projections $\pi_t : C \to \mathbb{R}$, $\pi_t(f) = f(t)$ for all $t \in [0,T]$ and $t \geqslant 0$, respectively. Hence, a probability measure $\mathbb{P}$ is uniquely determined by its finite-dimensional distributions $(\mathbb{P}^{\pi_{t_1,\dots,t_m}})_{m,t_1,\dots,t_m}$ with $\pi_{t_1,\dots,t_m}(f) = (f(t_1),\dots,f(t_m))$.

6.19 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 $\mathbb{P}_n^{\pi_{t_1,\dots,t_m}}$ converge weakly to $\mathbb{P}^{\pi_{t_1,\dots,t_m}}$ and $(\mathbb{P}_n)$ is tight.*

Proof. Since C is Polish, $(\mathbb{P}_n)$ is weakly relatively compact if and only if it is tight by Prohorov's Theorem. Therefore, $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ implies that $(\mathbb{P}_n)$ is tight. Since all $\pi_{t_1,\dots,t_m} : S \to \mathbb{R}^m$ are continuous, the continuous mapping theorem, applied to probability measures, gives $\mathbb{P}_n^{\pi_{t_1},\dots\pi_{t_m}} \xrightarrow{w} \mathbb{P}^{\pi_{t_1,\dots,t_m}}$.

Conversely, if $(\mathbb{P}_n)$ is tight, then any subsequence (n_k) has a subsubsequence (n_{k_l}) with $\mathbb{P}_{n_{k_l}} \xrightarrow{d} \mathbb{Q}$ for some probability measure $\mathbb{Q}$. By continuous mapping, also the finite-dimensional distributions converge to those of $\mathbb{Q}$. By assumption, they also converge to the finite-dimensional distributions of $\mathbb{P}$. By the above uniqueness result, we infer $\mathbb{Q} = \mathbb{P}$. If $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ were not true, then there would be a subsequence (n_k) and some $\varepsilon > 0$, $f \in C_b(S)$ such that $|\int f\,d\mathbb{P}_{n_k} - \int f\,d\mathbb{P}| \geqslant \varepsilon$ for all k. From (n_k), however, we could extract a subsubsequence (n_{k_l}) with $\mathbb{P}_{n_{k_l}} \xrightarrow{w} \mathbb{P}$ and thus $|\int f\,d\mathbb{P}_{n_{k_l}} - \int f\,d\mathbb{P}| \to 0$. This contradiction proves $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$. □

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

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

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

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

(b) $\lim_{\delta\to 0}\sup_{f\in A}\omega_\delta(f) = 0$ *(equi-continuity, gleichgradige Stetigkeit).*

Proof. e.g. Dirk Werner, *Einführung in die Höhere Analysis*, Springer. □

6.22 Remark. Relative compactness in $C(\mathbb{R}^+)$ holds if (a) holds and (b) is satisfied for all $T > 0$, i.e. $\lim_{\delta\to 0}\sup_{f\in A}\max\{|f(s)-f(t)| \mid s,t \in [0,T],\ |s-t| \leqslant \delta\} = 0$ holds for all $T > 0$. ► Exercise

6.23 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* $\varepsilon>0$.

Proof. If $(\mathbb{P}_n)$ is tight, then for $\eta>0$ there is a compact set K_η with $\mathbb{P}_n(K_\eta)\geqslant 1-\eta$, $n\geqslant 1$. By Arzelà-Ascoli, for any $\varepsilon>0$ we have

$$K_\eta\subseteq\{|f(0)|\leqslant R_\eta,\omega_{\delta_{\eta,\varepsilon}}(f)<\varepsilon\}$$

for sufficiently large $R_\eta>0$ and small $\delta_{\eta,\varepsilon}$. With $\eta\downarrow 0$ we deduce

$$\lim_{R\to\infty}\sup_n \mathbb{P}_n(\{|f(0)|\geqslant R\})=0,\quad \lim_{\delta\to 0}\sup_{n\to\infty}\mathbb{P}_n(\{\omega_\delta(f)\geqslant\varepsilon\})=0,$$

which implies (a), (b) due to $\limsup\leqslant\sup$.

Conversely, given (a), (b) and $\eta>0$ choose $R>0$ with

$$\mathbb{P}_n(\{|f(0)|\leqslant R\})\geqslant 1-\eta/2$$

and for each $k\in\mathbb{N}$ some $\delta_k>0$ with

$$\liminf_{n\to\infty}\mathbb{P}_n(\{\omega_{\delta_k}(f)\leqslant 1/k\})\geqslant 1-\eta 2^{-k-1}.$$

By Arzelà-Ascoli, the set $K_\eta=\{|f(0)|\leqslant R\}\cap\bigcap_{k\geqslant 1}\{\omega_{\delta_k}(f)\leqslant 1/k\}$ is relatively compact and satisfies $\limsup_{n\to\infty}\mathbb{P}_n(K_\eta^C)\leqslant\eta/2+\sum_{k\geqslant 1}\eta 2^{-k-1}=\eta$. This shows for the compact set $\overline{K}_\eta$: $\liminf_{n\to\infty}\mathbb{P}_n(\overline{K}_\eta)\geqslant 1-\eta$. Pick $n_0\in\mathbb{N}$ such that for $n\geqslant n_0$ we have $\mathbb{P}_n(\overline{K}_\eta)\geqslant 1-2\eta$ and then some compact set $\tilde{K}_\eta$ with $\mathbb{P}_n(\tilde{K}_\eta)\geqslant 1-2\eta$ for the finitely many indices $n=1,\dots,n_0-1$. Then $\inf_{n\geqslant 1}\mathbb{P}_n(\overline{K}_\eta\cup\tilde{K}_\eta)\geqslant 1-2\eta$ holds with the compact set $\overline{K}_\eta\cup\tilde{K}_\eta$. This gives tightness. □

6.24 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* $\varepsilon>0$.

6.25 Remark. Tightness on $\mathfrak{B}_{C(\mathbb{R}^+)}$ follows if conditions (a), (b') are satisfied for all $T>0$. The advantage of (b') compared to (b) is that the maximum over t is pulled out of the probability.

Proof. To prove (b')$\Rightarrow$ (b) in the corollary, let $\varepsilon,\eta>0$ and choose $m,n_0\in\mathbb{N}$ with

$$\forall\, n\geqslant n_0,\ t\in[0,1-m^{-1}]:\ \frac{1}{m^{-1}}\,\mathbb{P}_n\Big(\{\max_{s\in[t,t+m^{-1}]}|f(s)-f(t)|\geqslant\varepsilon/2\}\Big)\leqslant\frac{\eta}{2T}.$$

Suppose $\omega_{(2m)^{-1}}(f) \geqslant \varepsilon$ for some $f \in C([0,T])$. Then there are $t < s \leqslant t + (2m)^{-1}$ with $|f(t) - f(s)| \geqslant \varepsilon$. For $k \in \{0, \dots, 2m\lceil T\rceil - 2\}$ with $\frac{k}{2m} \leqslant t < s \leqslant \frac{k}{2m} + \frac{1}{m}$ this implies $|f(t) - f(\frac{k}{2m})| \geqslant \varepsilon/2$ or $|f(s) - f(\frac{k}{2m})| \geqslant \varepsilon/2$. Consequently, we obtain for $n \geqslant n_0$

$$\begin{aligned}\mathbb{P}_n(\{\omega_{(2m)^{-1}}(f) \geqslant \varepsilon\}) &\leqslant \sum_{k=0}^{2m\lceil T\rceil-2} \mathbb{P}_n\Big(\Big\{\max_{s\in[\frac{k}{2m},\frac{k}{2m}+\frac{1}{m}]} |f(s) - f(\tfrac{k}{2m})| \geqslant \varepsilon/2\Big\}\Big)\\ &\leqslant 2mTm^{-1}\frac{\eta}{2T} = \eta.\end{aligned}$$

For $\eta \downarrow 0$, noting the monotonicity of $\delta \mapsto \mathbb{P}_n(\{\omega_\delta(f) \geqslant \varepsilon\})$, this gives condition (b). □

6.26 Theorem (Kolmogorov, Centsov 1956). *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") $\exists \alpha, \beta > 0, K > 0\, \forall n \geqslant 1, s,t \in [0,T] : \mathbb{E}[|X_n(s) - X_n(t)|^\alpha] \leqslant K|s-t|^{1+\beta}$.

Proof. Putting $\mathbb{P}_n = \mathbb{P}^{X_n}$, condition (a) in the lemma is verified directly. For condition (b') we have to prove for some sequence $\delta_r \to 0$

$$\forall \varepsilon > 0 : \lim_{r\to\infty} \limsup_{n\to\infty} \sup_{t\in[0,T-\delta_r]} \delta_r^{-1}\, \mathbb{P}_n(\{\max_{s\in[t,t+\delta_r]} |f(s) - f(t)| \geqslant \varepsilon\}) = 0.$$

For simplicity let us consider $\mathbb{P}_n(\{\max_{s\in[t,t+\delta_r]}|f(s) - f(t)| \geqslant \varepsilon\})$ for $t = 0$. All arguments will remain valid for general t if s is replaced by $s - t$ in the sequel.

We apply a so-called chaining technique. Consider $\delta_r = 2^{-r}$, $r \geqslant 1$, $D_m = \{k2^{-m} \mid k \in \mathbb{N}_0\}$, $D = \bigcup_{m\in\mathbb{N}} D_m$ (dyadic numbers) and $\gamma \in (0, \beta/\alpha)$. By Markov's inequality and the assumption, we obtain

$$\forall c > 0 : \mathbb{P}(|X_n(k2^{-j}) - X_n((k-1)2^{-j})| \geqslant c2^{-\gamma j}) \leqslant c^{-\alpha} K 2^{-j(1+\beta-\alpha\gamma)}.$$

For $s \in [0, \delta_r) \cap D$ we have $s \in D_m$ for some $m \in \mathbb{N}$ and $s < 2^{-r}$. Writing $s = \sum_{l=r+1}^m b_l 2^{-l}$ with $b_l \in \{0,1\}$ and introducing $s_j = \sum_{l=r+1}^j b_l 2^{-l}$, we note $s_j \in D_j$, $|s_{j+1} - s_j| \leqslant 2^{-(j+1)}$ and $s_m = s$. With $s_r := 0$ we have the telescoping sum

$$X_n(s) - X_n(0) = \sum_{j=r+1}^m (X_n(s_j) - X_n(s_{j-1})).$$

If $|X_n(s_j) - X_n(s_{j-1})| < c2^{-\gamma j}$ with $c = 2 - 2^{1-\gamma}$ holds for all $j \geqslant r+1$, then $|X_n(s) - X_n(0)| < 2^{-\gamma r}$ follows by evaluating the geometric series. This shows

$$\bigcup_{s\in[0,\delta_r)\cap D} \big\{|X_n(s) - X_n(0)| \geqslant 2^{-\gamma r}\big\} \subseteq \bigcup_{\substack{j\geqslant r+1\\ 1\leqslant k\leqslant 2^{j-r}}} \big\{|X_n(k2^{-j}) - X_n((k-1)2^{-j})| \geqslant c2^{-\gamma j}\big\}.$$

Using that D is dense and X_n is continuous, we conclude

$$\mathbb{P}\Big(\max_{s\in[0,\delta_r]}|X_n(s)-X_n(0)|>2^{-\gamma r}\Big)\leqslant \mathbb{P}\Big(\bigcup_{s\in[0,\delta_r)\cap D}\{|X_n(s)-X_n(0)|\geqslant 2^{-\gamma r}\}\Big)$$

$$\leqslant \sum_{j\geqslant r+1}\sum_{k=1}^{2^{j-r}}\mathbb{P}(|X_n(k2^{-j})-X_n((k-1)2^{-j})|\geqslant c2^{-\gamma j})$$

$$\leqslant \sum_{j\geqslant r+1}\sum_{k=1}^{2^{j-r}} c^{-\alpha}K2^{-j(1+\beta-\alpha\gamma)}$$

$$= c^{-\alpha}K\delta_r\sum_{j\geqslant r+1}2^{-j(\beta-\alpha\gamma)} = c^{-\alpha}K\delta_r(2-2^{1-\beta+\alpha\gamma})^{-1}2^{-jr}.$$

Thus, for general $t\in[0,T-\delta_r]$ we obtain

$$\delta_r^{-1}\,\mathbb{P}_n(\{\max_{s\in[t,t+\delta_r]}|f(s)-f(t)|>2^{-\gamma r}\})\leqslant c^{-\alpha}K(2-2^{1-\beta+\alpha\gamma})^{-1}2^{-jr}\to 0$$

uniformly in n and t as $\delta_r=2^{-r}\to 0$. By monotonicity, this convergence continues to hold if '$>2^{-\gamma r}$' inside the probability is replaced by '$\geqslant\varepsilon$' for some fixed $\varepsilon>0$, which was to be proved. □

6.27 Remark. Using the previous remark, the Kolmogorov-Centsov Theorem extends to the case $C(\mathbb{R}^+)$, when the conditions hold for all $T>0$, i.e. the moment condition is satisfied for all $s,t\geqslant 0$.

7 Invariance principle and the empirical process

7.1 Invariance principle and Brownian motion

7.1 Definition. A process $(B_t,\, t\geqslant 0)$ is called Brownian motion (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 $0\leqslant t_0<t_1<\cdots<t_m$ 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.

7.2 Remark. The existence of Brownian motion is non-trivial. Without the continuity assumption this follows from the construction of Gaussian processes by Kolmogorov's consistency theorem. Yet, the set of continuous paths is not even measurable with respect to the product σ-algebra ▶ Exercise. Here, existence of Brownian motion will be a consequence of laws of rescaled random walks converging in $C([0,1])$ or $C(\mathbb{R}^+)$ towards a limit law under which the coordinate projections form a Brownian motion.

7.3 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.*

Proof. Because of $Y_n(t) = B_t = 0$ we just consider increments along $0 = t_0 < t_1 < \ldots < t_m \leqslant 1$. We write

$$Y_n(t_j) = \sum_{i=1}^{j} Z_i^{(n)} + \frac{nt_j - \lfloor nt_j \rfloor}{\sqrt{n}} X_{\lfloor nt_j \rfloor + 1} \text{ with } Z_i^{(n)} = \frac{1}{\sqrt{n}} \sum_{k=\lfloor nt_{i-1} \rfloor + 1}^{\lfloor nt_i \rfloor} X_k.$$

The central limit theorem yields for $n \to \infty$

$$\frac{\sqrt{n}}{\sqrt{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor}} Z_i^{(n)} \xrightarrow{d} N(0,1).$$

Since $\frac{\sqrt{n}}{\sqrt{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor}} \to \frac{1}{\sqrt{t_i - t_{i-1}}}$ holds, Slutsky's Lemma and scaling of the normal distribution imply $Z_i^{(n)} \xrightarrow{d} \bar{Z}_i \sim N(0, t_i - t_{i-1})$. Now observe that $Z_1^{(n)}, \ldots, Z_m^{(n)}$ are independent for each n. A simple consequence of the definition is that then $(Z_1^{(n)}, \ldots, Z_m^{(n)}) \xrightarrow{d} (\bar{Z}_1, \ldots, \bar{Z}_m)$ holds with independent $\bar{Z}_1, \ldots, \bar{Z}_m$. By continuous mapping, we conclude

$$\begin{aligned}(Z_1^{(n)}, Z_1^{(n)} + Z_2^{(n)}, \ldots, Z_1^{(n)} + \cdots + Z_m^{(n)}) &\xrightarrow{d} (\bar{Z}_1, \bar{Z}_1 + \bar{Z}_2, \ldots, \bar{Z}_1 + \cdots + \bar{Z}_m) \\ &\stackrel{d}{=} (B_{t_1}, B_{t_2}, \ldots, B_{t_m}).\end{aligned}$$

Let us remark that this step follows more easily by the multivariate central limit theorem.

Finally, observe $\mathbb{E}[(\frac{nt_j - \lfloor nt_j \rfloor}{\sqrt{n}} X_{\lfloor nt_j \rfloor + 1})^2] \leqslant \frac{1}{n} \mathbb{E}[X_1^2] \to 0$ such that another application of Slutsky's Lemma gives $(Y_n(t_1), Y_n(t_2), \ldots, Y_n(t_m)) \xrightarrow{d} (B_{t_1}, B_{t_2}, \ldots, B_{t_m})$. □

7.4 Theorem (Invariance principle, functional CLT, Donsker 1951)**.** *Suppose* $(X_k)_{k\geqslant 1}$ *are i.i.d.,* $X_k \in L^2$, $\mathbb{E}[X_k] = 0$, $\operatorname{Var}(X_k) = 1$ *and set* $S_n := \sum_{k=1}^n X_k$, $S_0 = 0$. *The rescaled, linearly interpolated random walk*

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

satisfies $Y^{(n)} \xrightarrow{d} B$ *with a Brownian motion* $(B_t,\, 0 \leqslant t \leqslant 1)$ *and convergence in distribution on* $(C([0,1]), \mathfrak{B}_{C([0,1])})$. *In particular, Brownian motion exists for* $t \in [0,1]$.

Proof. Here we give the proof under the additional assumption $X_k \in L^4$ which permits an application of the Kolmogorov-Centsov criterion. Due to $Y_n(0) = B_0 = 0$ and the lemma on the convergence of the finite-dimensional distributions it suffices to check tightness via

$$\exists K > 0 \,\forall s, t \in [0,1]: \; \mathbb{E}[(Y_n(t) - Y_n(s))^4] \leqslant K(t-s)^2.$$

From tightness it follows in particular that the limit law (so-called Wiener measure) exists on $(C([0,1]), \mathfrak{B}_{C([0,1])})$. Under this law $B_t(\omega) = \omega(t)$ for $\omega \in \Omega = C([0,1])$ is clearly continuous in t and has the correct finite-dimensional distributions such that $(B_t, 0 \leqslant t \leqslant 1)$ forms a Brownian motion. Let us write for $t > s$

$$Y_n(t) - Y_n(s) = \frac{1}{\sqrt{n}}(S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor}) + \frac{nt - \lfloor nt \rfloor}{\sqrt{n}} X_{\lfloor nt \rfloor + 1} - \frac{ns - \lfloor ns \rfloor}{\sqrt{n}} X_{\lfloor ns \rfloor + 1}.$$

We shall use $(A+B)^4 \leqslant 2^3(A^4 + B^4)$ several times, but in general just write C_i, $i = 1, \dots$ for some numerical constants.

In the case $t - s \geqslant \frac{1}{n}$ we have $\mathbb{E}[(\frac{nt - \lfloor nt \rfloor}{\sqrt{n}} X_{\lfloor nt \rfloor + 1})^4] \leqslant n^{-2}\,\mathbb{E}[X_1^4] \leqslant C_1(t-s)^2$ and similarly for the term in s instead of t. By the independence of (X_k) and $\mathbb{E}[X_k] = 0$ we have for $L > l$

$$\begin{aligned}
\mathbb{E}[(S_L - S_l)^4] &= \sum_{k=l+1}^{L} \mathbb{E}[X_k^4] + 2 \sum_{l+1 \leqslant k_1 < k_2 \leqslant L} \mathbb{E}[X_{k_1}^2]\,\mathbb{E}[X_{k_2}^2] \\
&= (L-l)\,\mathbb{E}[X_1^4] + (L-l)(L-l+1)\,\mathbb{E}[X_1^2]^2 \leqslant C_2(L-l)^2.
\end{aligned}$$

For $n(t-s) \geqslant 1$ this shows

$$\mathbb{E}[(\tfrac{1}{\sqrt{n}}(S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor}))^4] \leqslant C_2 n^{-2}(\lfloor nt \rfloor - \lfloor ns \rfloor)^2 \leqslant C_3(t-s)^2.$$

We conclude $\mathbb{E}[(Y_n(t) - Y_n(s))^4] \leqslant C_4(t-s)^2$ provided $t - s \geqslant \frac{1}{n}$.

If $t - s < \frac{1}{n}$ and $\lfloor nt \rfloor = \lfloor ns \rfloor$ holds, then $Y_n(t) - Y_n(s) = \sqrt{n}(t-s)X_{\lfloor nt \rfloor + 1}$. This gives $\mathbb{E}[(Y_n(t) - Y_n(s))^4] \leqslant C_5 n^2 (t-s)^4 \leqslant C_5(t-s)^2$. If $t - s < \frac{1}{n}$ and $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$ holds, then we have $Y_n(t) - Y_n(s) = \frac{nt - \lfloor nt \rfloor}{\sqrt{n}} X_{\lfloor nt \rfloor + 1} + \frac{\lfloor nt \rfloor - ns}{\sqrt{n}} X_{\lfloor nt \rfloor}$. This implies

$$\begin{aligned}
\mathbb{E}[(Y_n(t) - Y_n(s))^4] &\leqslant C_6\Big(\frac{(nt - \lfloor nt \rfloor)^4}{n^2}\,\mathbb{E}[X_1^4] + \frac{(\lfloor nt \rfloor - ns)^4}{n^2}\,\mathbb{E}[X_1^4]\Big) \\
&\leqslant C_7 n^{-2}(nt - ns)^4 \leqslant C_7(t-s)^2.
\end{aligned}$$

With $K = C_4 \vee C_5 \vee C_7$ the Kolmogorov-Centsov criterion is satisfied. □

7.5 Corollary. *Brownian motion exists on the positive real line.*

7.6 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:* $\max_{0 \leqslant t \leqslant 1} B_t \stackrel{d}{=} |B_1|$.