Stochastic Processes in Discrete Time - Stochastik II

Lecture notes

Prof. Dirk Becherer

Institut für Mathematik
Humboldt Universität zu Berlin

February 12, 2018

Contents

Chapter 1

Conditional expectations

To define conditional expectation, we will need the following theory.

1.1 The Hilbert space L^2 and orthogonal projections

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ is the space of (equivalence classes of) *square-integrable real-valued* random variables with the scalar product

$$\langle X, Y \rangle := \mathbb{E}[XY]$$

that induces the norm $\|X\| := \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]}$. From measure and integration theory we know that L^2 is a complete normed space (hence a Banach space), with the scalar product $\langle \cdot, \cdot \rangle$, i.e. it is even a Hilbert space.

Theorem 1.1. *Let K be a closed subspace in L^2. Then for any $Y \in L^2$ there exists an element $\hat{Y} \in K$ such that*

(i) $\|Y - \hat{Y}\| = d(Y, K) := \inf \left\{ \|Y - X\| \,\middle|\, X \in K \right\}$,

(ii) $Y - \hat{Y} \perp X$ *for all* $X \in K$, *i.e.* $Y - \hat{Y} \in K^{\perp}$.

Properties (i) and (ii) are equivalent and $\hat{Y}$ is uniquely described through (i) or (ii).

Remark 1.1. The result holds as well for general Hilbert spaces: Note that in the proof we just used the inner-product-structure but not the specific form of this inner product.

Definition 1.2 ("best forecast in L^2 sense")**.** $\hat{Y}$ from Theorem 1.1 is called *the orthogonal projection* of Y into K (in the Hilbert space L^2).

1.2 Conditional expectations: Definition and construction

Let us fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Conditional expectations will be essential for the definition of martingales. A conditional expectation could be seen as the best approximation in a mean-squared-error sense for a real valued random variable given some partial information (modelled as a sub-σ-field $\mathcal{G}$ of $\mathcal{F}$).

More precisely and more generally, we define

Definition 1.3 (Conditional expectation)**.** Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Let also X be a random variable with $X \geqslant 0$ or $X \in L^1$. Then a random variable Y with the properties

(i) Y is $\mathcal{G}$-measurable,

(ii) $\mathbb{E}[Y\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$ for each $G \in \mathcal{G}$,

is called (a version of) the conditional expectation of X with respect to $\mathcal{G}$. We write $Y = \mathbb{E}[X|\mathcal{G}]$.

Remark 1.2.

- Intuitively, (i) means "based on information $\mathcal{G}$", while (ii) means "best forecast".
- Using (i) and (ii) only, it is straightforward to check that

$$X \geqslant 0 \Longrightarrow Y \geqslant 0,$$

$$X - \text{integrable} \Longrightarrow Y - \text{integrable}.$$

(By (i), note that $\{\hat{Y} \geqslant 0\}$, $\{\hat{Y} < 0\} \in \mathcal{G}$; use (ii) in Definition 1.3)

Exercise 1.1. Let Y be a $\mathcal{G}$-measurable integrable random variable for a σ-field $\mathcal{G}$. Then the following are equivalent

(a) Y satisfies (ii) in Definition 1.3 for every $G \in \mathcal{G}$, i.e.

$$\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[Y\mathbb{1}_G] \quad \forall G \in \mathcal{G};$$

(b) Y satisfies

$$\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[Y\mathbb{1}_G] \quad \forall G \in \mathcal{E}$$

for a class $\mathcal{E} \subset 2^\Omega$ with $\Omega \in \mathcal{E}$, $\mathcal{G} = \sigma(\mathcal{E})$ and $\mathcal{E}$ - $\cap$-closed;

(c) Y satisfies

$$\mathbb{E}[XZ] = \mathbb{E}[YZ]$$

for all $\mathcal{G}$-measurable, bounded (and[1] non-negative) random variables Z.

Sidenote: An analogous statement holds with Y being nonnegative instead of integrable, if one requires $Z \geqslant 0$ (instead of boundedness) in part (c).

Theorem 1.4. *Let X be a random variable with $X \geqslant 0$ (or $X \in L^1$) and let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Then:*

(1) Conditional expectation $\mathbb{E}[X|\mathcal{G}]$ exists and satisfies $\mathbb{E}[X|\mathcal{G}] \geqslant 0$ (or $\mathbb{E}[X|\mathcal{G}] \in L^1$ respectively).

(2) The conditional expectation is $\mathbb{P}$-a.s. unique, i.e. if Y and Y' are both versions of the conditional expectation, then $Y = Y'$ $\mathbb{P}$-a.s.

[1] optional

Remark 1.3. In part (b) of the proof above we constructed $\mathbb{E}[X|\mathcal{G}]$ for $X \in L^2$ as the orthogonal projection onto the subspace $L^2_{\mathcal{G}}$, i.e. the conditional expectation minimizes $\mathbb{E}[(X-Y)^2] = \|X-Y\|^2_{L^2}$ over all $Y \in L^2_{\mathcal{G}}$, which justifies the description of the conditional expectation as "the best" projection (for $X \in L^2$).

In simple cases conditional expectations can be constructed explicitly.

Example 1.1. Let $\mathcal{G}$ be a σ-field that is generated by a finite number of sets, i.e. $\mathcal{G} = \sigma(A_1, \dots, A_N)$ for a finite number of sets $A_i \in \mathcal{F}$. For such partitions, we can find a disjoint partition $\Omega = \bigsqcup_{i=1}^n B_i$ with $B_i \cap B_j = \varnothing$ for $i \neq j$ with $\mathcal{G} = \sigma(B_1, \dots, B_n)$ ($\bigsqcup \equiv$ disjoint union). Then every $\mathcal{G}$-measurable real-valued random variable Y has the form

$$Y = \sum_{i=1}^n y_i \mathbb{1}_{B_i}, \quad y_i \in \mathbb{R}$$

($\rightsquigarrow$ Exercise: $\Leftarrow$ is clear, $\Rightarrow$ follows by contradiction).

Let $X \in L^1$ or $\geqslant 0$. Then we have the following explicit representation for the conditional expectation of X given $\mathcal{G}$:

$$\mathbb{E}[X|\mathcal{G}] = \sum_{i=1}^n \frac{\mathbb{E}[X\mathbb{1}_{B_i}]}{\mathbb{P}[B_k]} \mathbb{1}_{B_i} \qquad \text{(with the convention } 0/0 \equiv 0\text{).}$$

There is an analogous formula if $\mathcal{G}$ is generated by a countable (possibly infinite) set.

For example, if $X = f(U)$ with U uniformly distributed on $[0,1]$ and (B_k) some partition of [0,1], then the formula above shows that $\mathbb{E}[X|\mathcal{G}]$ is a piecewise-constant function on each interval B_k with constant value equal to the mean of $f(U)$ on this interval.

Intuitive interpretation of conditional expectations with conditional probabilities: for $\mathbb{P}[B_i] > 0$ the elementary conditional probability is defined by $\mathbb{P}[\cdot|B_i] := \mathbb{P}[\,\cdot\, \cap B_i]/\mathbb{P}[B_i]$, and the conditional expectation of X given B_i could be defined by

$$\mathbb{E}[X|B_i] = \int X d\mathbb{P}[\cdot|B_i] = \frac{\mathbb{E}[X\mathbb{1}_{B_i}]}{\mathbb{P}[B_i]},$$

such that $\mathbb{E}[X|\mathcal{G}] = \sum_{i=1}^n \mathbb{E}[X|B_i] \cdot \mathbb{1}_{B_i}$.

Remark 1.4. A similar formula holds when $\mathcal{G}$ is generated by a countably infinite partition. Check that (i) and (ii) in Definition 1.3 are fulfilled (exercise).

Remark 1.5. If $\mathcal{G} = \sigma(Z)$ for a random variable Z, we define $\mathbb{E}[X|Z] := \mathbb{E}[X|\sigma(Z)]$. Then $\mathbb{E}[X|Z]$ is of the form $h(Z)$ for a measurable function h, since σ-measurable. We often write $h(z) = \mathbb{E}[X|Z = z]$, or more precise would be $h(z) = \mathbb{E}[X|Z]\big|_{Z=z}$.

1.3 Properties of conditional expectations

Throughout this section, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$.

Theorem 1.5. *Let $X \geqslant 0$ or in L^1. Then*

(1) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$*;*

(2) *X: $\mathcal{G}$-measurable* $\Longrightarrow \mathbb{E}[X|\mathcal{G}] = X$ *a.s.;*

(3) Linearity: $X_{1,2} \in L^1$, $a, b \in \mathbb{R} \Longrightarrow \mathbb{E}[aX_1 + bX_2|\mathcal{G}] = a\mathbb{E}[X_1|\mathcal{G}] + b\mathbb{E}[X_2|\mathcal{G}]$*;*

(4) Monotonicity: $X_{1,2} \geqslant 0$ *or* $\in L^1$ *with* $X_1 \leqslant X_2$ *a.s.* $\Longrightarrow \mathbb{E}[X_1|\mathcal{G}] \leqslant \mathbb{E}[X_2|\mathcal{G}]$ *a.s.;*

(5) Projectivity, "tower property": $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ *- σ-fields, then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] \quad \textit{a.s.;}$$

(6) "Taking out what is known": *Z: $\mathcal{G}$-measurable, $Z \cdot X$ and X are $\geqslant 0$ or in L^1, then*

$$\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}] \quad a.s.;$$

(7) If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ *a.s.*

Example 1.2 (Application of conditional expectations: martingales)**.** Let $(\mathcal{F}_k)_{k\in\mathbb{N}_0}$ with $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$, be an increasing family of σ-fields. Such a sequence $(\mathcal{F}_k)$ is called a *filtration*. A process $(M_k)_{k\in\mathbb{N}_0}$ with the properties

(1) $M_k \in L^1$ for every $k \in \mathbb{N}_0$,

(2) Any M_k, $k \in \mathbb{N}_0$, is $\mathcal{F}_k$-measurable, (We say: (M_k) is *adapted to* $(\mathcal{F}_k)$.)

(3) $\mathbb{E}[M_k|\mathcal{F}_l] = M_l$ a.s. $\forall l \leqslant k$,

is called a *martingale*.

(a) *Random walk:* Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with $Y_k \in L^1$, $\mathbb{E}[Y_i] = 0$. Then $M_k := \sum_{i=1}^k Y_i$ is a martingale with respect to the filtration $(\mathcal{F}_k) = (\sigma(Y_i : \ i \leqslant k))$ ($\rightsquigarrow$ Exercise).

(b) Let $X \in L^1$ and $(\mathcal{F}_k)_{k\in\mathbb{N}_0}$ be a filtration. Then $M_k := \mathbb{E}[X|\mathcal{F}_k]$, $k \in \mathbb{N}_0$, is a martingale.

(c) *"Multiplicative random walk":* $Y_1, Y_2, \cdots$, are i.i.d. random variables such that $Y_i \geqslant 0$ $\forall i$ and bounded. Let $\mathcal{F}_k = \sigma(Y_i : \ i \leqslant k)$. Then $M_k := \prod_{i=1}^k Y_i$, $k = 1, 2, \ldots$, is a martingale if $\mathbb{E}[Y_i] = 1$ (noting $M_0 = 1$).

Theorem 1.6 (Limiting theorems and inequalities for conditional expectations)**.**

(1) **Monotone convergence:** *Let* $0 \leqslant X_n \nearrow X$ *a.s. Then*

$$\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}] \quad a.s.$$

(2) **Fatou's lemma:** *Let $X_n \geqslant 0$ for all n. Then*

$$\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \quad a.s.$$

(3) **Dominated convergence theorem (theorem of Lebesgue):** *If* $X_n \overset{a.s.}{\to} X$ *and there exists* $Z \in L^1$ *such that* $|X_n| \leqslant Z$*, then*

$$\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}] \quad a.s.$$

(4) **Jensen's inequality:** *Let* $X \in L^1$ *and* $u : \mathbb{R} \to \mathbb{R}$ *be convex with* $u(X) \in L^1$*. Then*

$$\mathbb{E}[u(X)|\mathcal{G}] \geqslant u(\mathbb{E}[X|\mathcal{G}]) \quad a.s.$$

A simple corollary of Jensen's inequality gives us

Corollary 1.7. *For* $p \geqslant 1$*, the conditional expectation maps* L^p *into itself, i.e. for* $X \in L^p$ *we have* $\mathbb{E}[X|\mathcal{G}] \in L^p$*, and*

$$\|\mathbb{E}[X|\mathcal{G}]\|_p \leqslant \|X\|_p.$$

Example 1.3. (1) ***Conditional expectation if joint density exists:***

Let X and Y be real-valued random variables with joint density $f(x, y)$. Let g be a bounded (or non-negative) measurable function. Then

$$\mathbb{E}[g(X)|Y] \equiv \mathbb{E}[g(X)|\sigma(Y)] = h(Y)$$

for

$$h(y) := \frac{\int g(x) f(x,y)\,\mathrm{d}x}{\int f(x,y)\,\mathrm{d}x} = \int g(x) \frac{f(x,y)}{\int f(u,y)\,\mathrm{d}u}\,\mathrm{d}x \quad \Big(\equiv \int g(x) \underbrace{f_{X|Y}(x|y)}_{\text{cond. density}} \,\mathrm{d}x \Big),$$

where the right-hand side is defined finitely, i.e. $h = 0$ when the denominator in the formula is 0. (Exercise)

(2) ***Bayes' formula:***

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and consider the probability measure $\mathbb{Q}$ with density Z with respect to $\mathbb{P}$, i.e. $\mathbb{Q}[A] = \mathbb{E}_{\mathbb{P}}[Z \mathbb{1}_A]$ for every $A \in \mathcal{F}$ (in particular $\mathbb{E}[Z] = 1$, $Z \geqslant 0$ P-a.s.). Then for any random variable $X \geqslant 0$ (or in L^1) and sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$ we have

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}] = \frac{\mathbb{E}_{\mathbb{P}}[X \cdot Z|\mathcal{G}]}{Z_{\mathcal{G}}},$$

where $Z_{\mathcal{G}} := \mathbb{E}_{\mathbb{P}}[Z|\mathcal{G}]$. Note that the right-hand side is defined $\mathbb{Q}$-a.s.; on the $\mathbb{Q}$-nullset where the (version of the) denominator is 0, one could set the conditional expectation to 0.

Remark: $Z_{\mathcal{G}}$ is the density of $\mathbb{Q}$ with respect to $\mathbb{P}$ on $\mathcal{G}$ in the sense that $\mathbb{Q}[G] = \mathbb{E}_{\mathbb{P}}[Z_{\mathcal{G}} \mathbb{1}_G]$ for every $G \in \mathcal{G}$.

(3) ***Wald's identities:***

Let $X_1, X_2, \ldots,$ be i.i.d. square integrable random variables and let N be an $\mathbb{N}$-valued random variable that is independent of $(X_i)_{i \in \mathbb{N}}$. Consider the random variable

$S := \sum_{i=1}^{N} X_i$ (interpretation: "total claim size of an insurance company"). We are interested in computing some statistics of S, e.g. moments: mean and variance. We do so by using the Theorem 1.8 with $\mathcal{G} := \sigma(N)$:

$$\begin{aligned}\mathbb{E}[S] = \mathbb{E}\Big[\mathbb{E}\Big[\sum_{i=1}^{N} X_i \;\Big|\; \mathcal{G}\Big]\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big]\Big|_{n=N}\Big] &= \mathbb{E}\Big[n\mathbb{E}[X_1]\big|_{n=N}\Big] \\ &= \mathbb{E}[N]\cdot\mathbb{E}[X_1] \quad \text{(Wald's first identity).}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[S^2] = \mathbb{E}\Big[\mathbb{E}\Big[\Big(\sum_{i=1}^{n} X_i\Big)^2\Big]\Big|_{n=N}\Big] &= \mathbb{E}\Big[\Big(\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) + \mathbb{E}\Big[\Big(\sum_{i=1}^{n} X_i\Big)\Big]^2\Big)\Big|_{n=N}\Big] \\ &= \mathbb{E}\left[\left(n\mathrm{Var}\left(X_1\right) + n^2\mathbb{E}\left[(X_1)\right]^2\right)\Big|_{n=N}\right] \\ &= \mathbb{E}[N]\cdot\mathrm{Var}[X_1] + \mathbb{E}[N^2]\cdot\mathbb{E}[X_1]^2,\end{aligned}$$

and thus

$$\begin{aligned}\mathrm{Var}[S] &= \mathbb{E}[N]\cdot\mathrm{Var}[X_1] + \mathbb{E}[X_1]^2\cdot\big(\mathbb{E}[N^2] - \mathbb{E}[N]^2\big) \\ &= \mathbb{E}[N]\cdot\mathrm{Var}[X_1] + \mathbb{E}[X_1]^2\cdot\mathrm{Var}[N] \qquad \text{(Wald's second identity).}\end{aligned}$$

(4) Let $X_1, \ldots, X_n$ be i.i.d., $X_1 \in L^1$. Set $S_n := \sum_{i=1}^{n} X_i$. Then we have

$$\mathbb{E}[X_i|S_n] = \frac{S_n}{n}.$$

Now, consider random variables X and Y defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in the measurable spaces $(S_X, \mathcal{S}_X)$ and $(S_Y, \mathcal{S}_Y)$ respectively.

Theorem 1.8. *Let X and Y be independent, $g(x, y)$ be a measurable real-valued function taking non-negative values (or $g(X, Y) \in L^1$). Then*

$$\mathbb{E}[g(X,Y)|Y] = \mathbb{E}[g(X,y)]\big|_{y=Y}.$$

More precisely, $h(y) := \mathbb{E}[g(X, y)]$ is a measurable function and the random variable $h(Y)$ is a version of $\mathbb{E}[g(X, Y)|Y]$.

1.4 Regular conditional probability distributions

For this section, fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sub-$\sigma$-field $\mathcal{G} \subseteq \mathcal{F}$, a measurable space $(S, \mathcal{S})$, and a random variable $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$.

Motivation: for any $B \in \mathcal{S}$, $\mathbb{1}_B(X) = \mathbb{1}_{\{X \in B\}}$ is a bounded random variable, hence (a version of) the conditional expectation $\mathbb{E}[\mathbb{1}_B(X)|\mathcal{G}] \equiv \mathbb{P}[X \in B|\mathcal{G}]$ exits, it is $\mathcal{G}$-measurable, valued in $[0, 1]$ a.s. This leads to a mapping

$$\begin{aligned}\Omega \times \mathcal{S} &\mapsto [0,1] \\ (\omega, B) &\mapsto \mathbb{P}[X \in B|\mathcal{G}][\omega]\end{aligned}$$

for any chosen versions (for each B) of the conditional expectations. However, the definition of such a mapping would depend on the choice of versions.

Question: Can we choose versions of the conditional expectations such that $B \mapsto \mathbb{P}[X \in B|\mathcal{G}]$ is a probability measure for any $\omega \in \Omega$ (or a.e.)? Problem: In general there a uncountably many B's.

Definition 1.9. (a) *A stochastic kernel* (or Markov kernel) from a measurable space $(S_1, \mathcal{S}_1)$ into another one $(S_2, \mathcal{S}_2)$ is a mapping $K : S_1 \times \mathcal{S}_2 \to [0, 1]$ such that

(i) for all $x_1 \in S_1$, $B \mapsto K(x_1, B)$ is a probability measure on $\mathcal{S}_2$,

(ii) for all $A_2 \in \mathcal{S}_2$, $x_1 \mapsto K(x_1, A_2)$ is a $\mathcal{S}_1 - \mathcal{B}([0, 1])$-measurable mapping.

(b) *A regular conditional probability* distribution of a random variable X taking values in $(S, \mathcal{S})$ given the σ-field $\mathcal{G}$ is a stochastic kernel K from $(\Omega, \mathcal{G})$ to $(S, \mathcal{S})$ such that for every $B \in \mathcal{S}$, $K(\cdot, B)$ is a version of $\mathbb{E}[\mathbb{1}_B(X)|\mathcal{G}] \equiv \mathbb{P}[X \in B|\mathcal{G}]$.

The existence results below will make use of the so-called "Monotone Class Theorem" (MCT). That is why we explain it first.

Definition 1.10. Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mathcal{D} \subseteq 2^\Omega$ be a class (familiy) of sets which is closed with respect to differences and increasing limits, that means

- $A, B \in \mathcal{D}$ with $B \subseteq A$ implies $A \backslash B \in \mathcal{D}$,
- $A_k \in \mathcal{D}$, $k \in \mathbb{N}$, with $A_1 \subseteq A_2 \subseteq \cdots$, implies $A := \bigcup_{k=1}^\infty A_k \in \mathcal{D}$.

We may call such $\mathcal{D}$ an "*m-class*"[2]. If additionally $\Omega \in \mathcal{D}$, then one calls $\mathcal{D}$ a *Dynkin system*.

Note that a Dynkin system contains any monotone limit of its elements, being sets.

Theorem 1.11 (Monotone class theorem for classes of sets, cf. [Klenke06] (Thm. 1.19) or [JacPro03] (ch.6))**.** *Let $\mathcal{E} \subseteq 2^\Omega$ be closed under finite intersections (and $\Omega \in \mathcal{E}$). Then the smallest Dynkin system $\delta(\mathcal{E})$ (or m-class when $\Omega \in \mathcal{E}$) containing $\mathcal{E}$ is equal to the σ-field $\sigma(\mathcal{E})$ generated by $\mathcal{E}$, i.e.*

$$\delta(\mathcal{E}) = \sigma(\mathcal{E}).$$

Note that "$\subseteq$" is clear, "$\supseteq$" is the main statement.

Corollary 1.12 (a simple MCT for functions)**.** *Let $\mathcal{E}$ be as in Theorem 1.11 with $\Omega \in \mathcal{E}$. Let $\mathcal{H}$ be a vector space of functions $\Omega \mapsto \mathbb{R}$ such that*

- *$\mathcal{H}$ contains all functions $\mathbb{1}_E$, $E \in \mathcal{E}$,*
- *if $f_n \in \mathcal{H}$, $n \in \mathbb{N}$, $0 \leqslant f_1 \leqslant f_2 \leqslant \cdots$, and $f :=\uparrow - \lim_{n\to\infty} f_n$ is bounded, then $f \in \mathcal{H}$.*

Then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{E})$-measurable functions.

Theorem 1.13. *Let $S = \mathbb{R}$ with the σ-field $\mathcal{S} = \mathbb{B}(\mathbb{R})$, and X be a $(S, \mathcal{S})$-valued random variable. For every sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$ there exists a regular conditional probability of X given $\mathcal{G}$.*

[2]this is not standard in the literature

Remark 1.6. The important property about $\mathbb{R}$ which we will use is that there exists a countable dense subset (e.g. $\mathbb{Q}$), i.e. that $\mathbb{R}$ is separable.

Remark 1.7. If $(S, \mathcal{S})$ as a measurable space is "similar" (isomorphic) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then one can construct a regular conditional probability on it.

Definition 1.14. A measurable space $(S, \mathcal{S})$ is called *Borel space* if there exists $A \in \mathcal{B}(\mathbb{R})$ and a bijection $\Psi : S \to A$ such that Ψ and Ψ^{-1} are measurable with respect to $(S, \mathcal{S})$ and $(A, \mathcal{B}(A))$.

Remark 1.8. Such Ψ induces a bijection between $\mathcal{S}$ and $\mathcal{B}(A)$ by[3] $B \mapsto \Psi(B) = (\Psi^{-1})^{-1}(B) \mapsto \Psi^{-1}(\Psi(B)) = B$. That means, the σ-fields are isomorphic. One could write briefly $\Psi(\mathcal{S}) = (\Psi^{-1})^{-1}(\mathcal{S}) = \mathcal{B}(A)$ and $\Psi^{-1}(\mathcal{B}(A)) = \mathcal{S}$.

Theorem 1.15. *Let $(S, \mathcal{S})$ be a Borel space, X be a random variable valued in $(S, \mathcal{S})$ and defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let also $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field. Then there exists a regular conditional probability for X given $\mathcal{G}$.*

Definition 1.16. *A Polish space* is a complete, separable and metrizable topological space.

Theorem 1.17. *Let S be a Polish space with $\mathcal{S} = \mathcal{B}(S)$ (the σ-algebra generated by the topology). Then $(S, \mathcal{S})$ is a Borel space.*

Example 1.4 (Examples of Polish spaces)**.**

- $(\mathbb{R}^d, \mathcal{B}^d)$;
- $(\mathbb{Z}^d, 2^{\mathbb{Z}^d})$;
- closed subsets A of $\mathbb{R}^d$ with $\mathcal{B}(A)$;
- $C([0,T], \mathbb{R})$ with $\|\cdot\|_\infty$.

Lemma 1.18 ("Lifting" of a kernel)**.** *Let X be a random variable taking values in $(S_1, \mathcal{S}_1)$, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$; let $\sigma(X) =: \mathcal{G} \subset \mathcal{F}$, and K be a stochastic kernel from $(\Omega, \sigma(X))$ to $(S_2, \mathcal{S}_2)$. Then there exists a unique stochastic kernel $\widetilde{K}$ from $(X(\Omega), X(\Omega) \cap \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$ such that $\widetilde{K}(X(\omega), B) = K(\omega, B)$ for all $\omega \in \Omega$ and $B \in \mathcal{S}_2$.*

Terminology: If K is the regular conditional probability of a random variable Y given $\sigma(X)$, then one calls $\widetilde{K}$ the regular conditional probability of Y given X.

$$\begin{tikzcd}
(S_1, \mathcal{S}_1) \arrow[drr, dashed, "\widetilde{K}"] & & \\
(\Omega, \mathcal{G} \equiv \sigma(X)) \arrow[u, "X"] \arrow[rr, "K"] & & (S_2, \mathcal{S}_2)
\end{tikzcd}$$

Figure 1.4.1: "Lifting" of a kernel K to a kernel $\widetilde{K}$ via the random variable X

[3] Note that each preimage here can be read as an inverse image.

Example 1.5 (Application example)**.** Let K be a regular conditional probability of a random variable Y (taking values in $S_Y, \mathcal{S}_Y)$) given $\mathcal{G} = \sigma(X)$ for some random variable X. Then for every measurable function $g \geqslant 0$ (or g bounded, or $g(Y) \in L^1$) we have

$$\mathbb{E}[g(Y)|X] = \int g(y)K(x, \mathrm{d}y)\Big|_{x=X} \quad \text{a.s.} \tag{1.4.1}$$

Remark 1.9. What if $g(X) \in L^1$ but not bounded? Then (1.4.1) would hold for g^+ and g^-. The decomposition $g = g^+ - g^-$ yields a.s. finiteness and equality (1.4.1), thus $\mathbb{P}[\text{“}\infty - \infty\text{''}] = \mathbb{P}\left[\int g^+ \,\mathrm{d}K = \int g^- \,\mathrm{d}K = \infty\right] = 0$, i.e. the right-hand side above is well-defined for $\mathbb{P} \circ X^{-1}$-almost all values of $x = X(\omega)$. Hence, the claim will hold true; one could define the integral at the RHS (for all values x) to be $-\infty$ for those x where it is (as a measure-integral) not well-defined, to have RHS defined for all x with a.s. equality holding true.

1.5 Disintegration and semi-direct products

Setting to consider: Given a stochastic kernel K from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$, a probability measure $\mathbb{P}_1$ on $(S_1, \mathcal{S}_1)$, $\Omega = S_1 \times S_2$, $\mathcal{F} = \mathcal{S}_1 \otimes \mathcal{S}_2$, and projections $X_i : \Omega \to S_i$, $\omega = (x_1, x_2) \mapsto x_i$. The aim is to construct a probability measure $\mathbb{P}$ on $\mathcal{F}$ such that $\mathbb{P} \circ X_1^{-1} = \mathbb{P}_1$ and K is the regular conditional probability distribution of X_2 given X_1.

Example 1.6. Let S_i be countable sets, $\mathcal{S}_i = 2^{S_i}$ $(i = 1, 2)$, $\mathcal{F} = 2^{\Omega}$. Any probability measure $\mathbb{P}$ on $\mathcal{F}$ is determined by the probability weights

$$\begin{aligned}\mathbb{P}[\{(x_1, x_2)\}] &= \mathbb{P}[X_1 = x_1, X_2 = x_2]\\ &= \mathbb{P}[X_1 = x_1] \cdot \mathbb{P}[X_2 = x_2 | X_1 = x_1]\\ &= \mathbb{P}_1[x_1] \cdot K(x_1, x_2).\end{aligned}$$

Hence, $\mathbb{P}_1$ and K determine $\mathbb{P}$ on $\mathcal{F}$. Furthermore, for any $f : (\Omega, \mathcal{F}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ the following formula holds:

$$\begin{aligned}\mathbb{E}[f(X_1, X_2)] &= \int_\Omega f \,\mathrm{d}\mathbb{P}\\ &= \sum_{x_1 \in S_1} \sum_{x_2 \in S_2} f(x_1, x_2) K(x_1, x_2) \mathbb{P}_1(x_1)\\ &= \int_{S_1} \left(\int_{S_2} f(x_1, x_2) K(x_1, \mathrm{d}x_2) \right) \mathbb{P}_1(\mathrm{d}x_1).\end{aligned}$$

In other words, the probability measure $\mathbb{P}$ and expectations with respect to it are being determined by $\mathbb{P}_1$ and K via a representation similar to the one from Fubini's theorem.

Theorem 1.19 (Disintegration)**.** *In the setting under consideration, there exists a probability measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$ *such that*

(a)

$$\int_\Omega f\,\mathrm{d}\mathbb{P} = \int_{S_1}\left(\int_{S_2} f(x_1,x_2)K(x_1,\mathrm{d}x_2)\right)\mathbb{P}_1(\mathrm{d}x_1) \quad \textit{for all measurable } f \geqslant 0.$$

In particular, for all $A \in \mathcal{F}$ *the sections* $A_{x_1} := \{x_2 \in S_2 |\ (x_1,x_2) \in A\}$ *are in* $\mathcal{S}_2$*, and*

(b) $\mathbb{P}[A] = \int_{S_1} K(x_1, A_{x_1})\mathbb{P}_1(\mathrm{d}x_1)$,

(c) $\mathbb{P}[A_1 \times A_2] = \int_{A_1} K(x_1, A_2)\mathbb{P}_1(\mathrm{d}x_1)$ *for all* $A_i \in \mathcal{S}_i$, $i = 1, 2$.

The probability measure $\mathbb{P}$ *is uniquely characterized by (c).*

Notation: For this unique measure $\mathbb{P}$ we write

$$\mathbb{P} = \mathbb{P}_1 \otimes K\,.$$

One calls P the (semi-direct) product of $\mathbb{P}_1$ with K.

Remark 1.10. Theorem 1.19 implies the standard statement of the Fubini's theorem for product measures $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ (take $K(x_1, \cdot) \equiv P_2(\cdot)$) (cf. [Klenke06, Theorem 14.14 and 14.16]).

Terminology: *The marginal distributions* of $\mathbb{P}$ on $\mathcal{S}_1 \otimes \mathcal{S}_2$ are $\mathbb{P}_{X_1} := \mathbb{P} \circ X_1^{-1}$ and $\mathbb{P}_{X_2} := \mathbb{P} \circ X_2^{-1}$. Then $K(x_1, \cdot)$ is the (regular) conditional probability distribution of X_2 given X_1.

Example 1.7. If X_1, X_2 are discrete random variables, we get using elementary conditional probabilities that

$$\mathbb{P}[X_2 \in A_2 | X_1 = x_1] = \frac{\mathbb{P}[\{X_2 \in A_2\} \cap \{X_1 = x_1\}]}{\mathbb{P}_1[\{X_1 = x_1\}]},$$

and hence $K(x_1, A_2) = \mathbb{P}[X_2 \in A_2 | X_1 = x_1]$ because $\mathbb{P}[\{X_2 \in A_2\} \cap \{X_1 = x_1\}] = \mathbb{P}_1[\{X_1 = x_1\}] \cdot K(x_1, A_2)$ by (c) in Theorem 1.19.

Remark 1.11. We can obtain a canonical generalization from 2 to n factors (finite number of factors) in Theorem 1.19 by induction. For a countably infinite number of factors the existence of a measure $\mathbb{P}$ will be proven later in the lecture.

Now, let $(S_i, \mathcal{S}_i)$ be measurable spaces for $1 \leqslant i \leqslant n$, $\mathbb{P}_1$ be a probability measure on $(S_1, \mathcal{S}_1)$ and K_i be stochastic kernels from $(\times_{k=1}^{i-1} S_k, \bigotimes_{k=1}^{i-1} \mathcal{S}_k)$ to $(S_i, \mathcal{S}_i)$ for $2 \leqslant i \leqslant n$. Set $\Omega := \times_1^n S_i$ and $\mathcal{F} := \bigotimes_1^n \mathcal{S}_i$.

Corollary 1.20. *In the aforementioned setting, there exists a probability measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$ *such that for every measurable* $f \geqslant 0$

(a)

$$\int_\Omega f\mathrm{d}\mathbb{P} = \int_{S_1}\int_{S_2}\cdots\int_{S_n} f(x_1,\cdots,x_n)K_n((x_1,\cdots,x_{n-1}),\mathrm{d}x_n)\cdots K_2(x_1,\mathrm{d}x_2)\mathbb{P}_1(\mathrm{d}x_1),$$

(b) and, in particular, for any $A_i \in \mathcal{S}_i$

$$\mathbb{P}\left[\bigtimes_1^n A_i\right] = \int_{A_1}\int_{A_2}\cdots\int_{A_{n-1}} K_n((x_1,\dots,x_{n-1}),A_n)K_{n-1}((x_1,\cdots,x_{n-2}),\mathrm{d}x_{n-1})$$
$$\cdots K_2(x_1,\mathrm{d}x_2)\mathbb{P}_1(\mathrm{d}x_1);$$

with probability measure $\mathbb{P}$ being uniquely defined in this way.

Remark 1.12. Remark for product measures $\mathbb{P} = \bigotimes_1^n \mathbb{P}_i$:

1. The construction above works for general measures (instead of just probability measures); the uniqueness argument in the Fubini's thm. needed σ-finiteness of $\mathbb{P}$.
2. One can consider $f \geqslant 0$, or bounded or $f \in L^1(\mathbb{P})$ in the theorem of Fubini for product measures; in the L^1-case, the inner integrals are finitely defined a.e. (wrt. the respective finite dim. distribution given by the outer intergrals); integral wrt to kernels could be defined (in generalization of the ususal measure integral) pointwise everywhere, defining their value as the usual difference of the integral of the positiv and negative integrand where that is defined, and letting it be $-\infty$ elsewhere.

Theorem 1.21. *Let X_1, X_2 be random variables defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and taking values in the measurable spaces $(S_1,\mathcal{S}_1)$ and $(S_2,\mathcal{S}_2)$ respectively. Let $\mathbb{P}_1$ be a measure on $\mathcal{S}_1$ and K be a stochastic kernel from S_1 to S_2. Suppose that the joint distribution (X_1,X_2) on $(S_1\times S_2,\mathcal{S}_1\otimes\mathcal{S}_2)$ is $\mathbb{P}_1\otimes K$. Then for any $\mathcal{S}_1\otimes\mathcal{S}_2$-measurable real-valued function $F\geqslant 0$, the following holds*

$$\begin{aligned}\mathbb{E}[F(X_1,X_2)|X_1](\omega) &= \int_{S_2} F(X_1(\omega),x_2)K(X_1(\omega),\mathrm{d}x_2)\\ &= h(X_1(\omega)),\end{aligned}$$

for $h(x_1) := \int_{S_2} F(x_1,x_2)K(x_1,\mathrm{d}x_2)$.

Remark 1.13. This statement generalizes a previous result: see Example 1.5.

Remark 1.14. A simple corollary of Theorem 1.21 is Theorem 1.8. Indeed, this follows when $\mathbb{P}_1\otimes K = \mathbb{P}_1\otimes\mathbb{P}_2$ (i.e. independent case, without dependence of K from its first argument).

Theorem 1.22. *Let $(S_i,\mathcal{S}_i)$ be measurable spaces for $i=1,2$, and $(S_2,\mathcal{S}_2)$ be a Polish space with $\mathcal{S}_2 = \mathcal{B}(S_2)$. Then every probability distribution $\mathbb{Q}$ on $(S_1\times S_2,\mathcal{S}_1\otimes\mathcal{S}_2) =: (\Omega,\mathcal{F})$ is of the form $\mathbb{Q} = \mathbb{Q}_1\otimes\widetilde{K}$ for a probability measure $\mathbb{Q}_1$ on $\mathcal{S}_1$ and a stochastic kernel $\widetilde{K}$ from $(S_1,\mathcal{S}_1)$ to $(S_2,\mathcal{S}_2)$.*

An analogous statement holds for measures on product spaces with finitely many factors $(S_i,\mathcal{S}_i), i=1,\cdots,n,$. Proof by induction.

Example 1.8 (Markov process in discrete finite time with a general state space)**.** Let $(S,\mathcal{S})$ be a measurable space, $\mathbb{P}_0$ a probability measure on $\mathcal{S}$ (starting initial distribution for the process), and K be a stochastic kernel from $(S,\mathcal{S})$ to $(S,\mathcal{S})$. An S-valued process $(X_n)_{n\geqslant 0}$ on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ is called *(time-homogeneous) Markov process* with initial distribution $\mathbb{P}_0$ and transition kernel K if

(i) $\mathbb{P}_{X_0} \equiv \mathbb{P} \circ X_0^{-1} = \mathbb{P}_0$,

(ii) for every $n \geqslant 0$ and $B \in \mathcal{S}$ holds

$$\mathbb{P}[X_i \in B | X_{i-1}, X_{i-2}, \cdots, X_0] = \mathbb{P}[X_i \in B | X_{i-1}] = K(X_{i-1}, B) \quad \text{a.s..}$$

For a given $\mathbb{P}_0$ and K, there exists a Markov process $(X_k)_{k=0,1,\ldots,N}$ for any given finite time horizon $N \in \mathbb{N}$, and the joint law of $(X_0, X_1, \cdots, X_N)$ is given by $\mathbb{P}_0 \otimes \underbrace{K \otimes \cdots \otimes K}_{N \text{ times}}$. (For infinite time horizon, we will construct such a process only later in the lecture).

Existence: Take $\Omega = \times_0^N S_i$, $\mathcal{F} = \bigotimes_0^N \mathcal{S}_i$, $\mathbb{P} := \mathbb{P}_0 \otimes \underbrace{K \otimes \cdots \otimes K}_{N \text{ times}}$ with kernels $K_i((x_0, \ldots, x_{i-1}), dx_i) = K(x_{i-1}, dx_i)$ as for Cor. 1.20 but with the conditional distribution of the next future state x_i depending on the 'past' $(x_0, \ldots, x_{i-1})$ only through the 'present state' x_{i-1} (Markov property). Provided that the Kernels K do not depend on i one speaks of a time-homogeneous Markov process. (If one considers more generally kernels $K_i(x_{i-1}, dx_i)$ with K_i not being the same for all i one speaks of a time-inhomogeneous Markov process and can do a likewise construction.)

Then it is easy to check that (i) and (ii) hold for the canonical projections $X_0, \cdots, X_N$ on the product space:

- Clearly $\mathbb{P}_{X_0} = \mathbb{P}_0$ by construction.
- By the generalized Fubini's theorem (Theorem 1.19, extension for more kernels) we have for all $B_k \in \mathcal{S}$

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{B_n}(X_n) \cdot \prod_0^{n-1} \mathbb{1}_{B_k}(X_k)] &= \int_{B_0} \cdots \int_{B_{n-1}} K(x_{n-1}, B_n) K(x_{n-2}, \mathrm{d}x_{n-1}) \cdots \mathbb{P}_0(\mathrm{d}x_0) \\ &= \mathbb{E}[K(X_{n-1}, B_n) \cdot \prod_0^{n-1} \mathbb{1}_{B_k}(X_k)]. \end{aligned}$$

  Hence, $\mathbb{P}[X_n \in B | X_{n-1}, \cdots, X_0] = K(X_{n-1}, B)$ since $\times_0^{n-1}\{X_k \in B_k\}$ forms an intersection stable generator of $\sigma(X_0, \cdots, X_{n-1})$ (cf. Exercise 1.1). Analogously, taking $B_k = S$ for $k = 0, \cdots, n-2$ we get $\mathbb{P}[X_n \in B | X_{n-1}] = K(X_{n-1}, B)$.

Uniqueness of the finite dimensional distribution: follows by uniqueness of the generalized Fubini theorem (Cor.1.20), noting that (ii) implies (with notations as from the proof there) that $\mathbb{P}_{k+1} = \mathbb{P}_k \otimes K$ for $k = 1, \ldots n$ is the joint distribution of $X_0, \ldots, X_k$.

Chapter 2

Martingales

2.1 Definition and interpretation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Consider discrete time $n = 0, 1, \ldots,$ with $n \in \{0, 1, \ldots, T\}$ (finite horizon) or $n \in \mathbb{N}_0$ (infinite horizon).

Definition 2.1. (a) *A filtration* $(\mathcal{F}_n)_{n \geqslant 0}$ on $\mathcal{F}$ is an increasing family of σ-fields $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$. Set $\mathcal{F}_\infty := \sigma\left(\bigcup_{n \geqslant 0} \mathcal{F}_n\right) \subset \mathcal{F}$.

(b) A stochastic process $(X_n)_{n \geqslant 0}$ is *adapted* to the filtration $(\mathcal{F}_n)_{n \geqslant 0}$ if X_n is $\mathcal{F}_n$-measurable for each $n \geqslant 0$.

(c) *The natural filtration* of a process $(X_n)_{n \geqslant 0}$ is the smallest filtration with respect to which $(X_n)_{n \geqslant 0}$ is an adapted process, i.e. $(\mathcal{G}_n)_{n \geqslant 0}$ for $\mathcal{G}_n = \sigma(Y_k : \; k \leqslant n)$.

Remark 2.1. Filtrations are models for information flow, $\mathcal{F}_n$ represents information available at time n.

Definition 2.2. A stochastic process $(X_n)_{n \geqslant 0}$ is called *martingale* with respect to the filtration $(\mathcal{F}_n)_{n \geqslant 0}$ and a probability measure $\mathbb{P}$ if

(i) $(X_n)_{n \geqslant 0}$ is adapted,

(ii) $(X_n)_{n \geqslant 0}$ is integrable, i.e. $X_n \in L^1(\mathbb{P})$ for all $n \geqslant 0$,

(iii) $\mathbb{E}[X_n | \mathcal{F}_{n-1}] = X_{n-1}$ $\mathbb{P}$-a.s. for all $n \geqslant 1$.

The process $(X_n)_{n \geqslant 0}$ is a super-/sub-martingale if (iii) only holds with "$\leqslant$" / "$\geqslant$".

Remark 2.2.

- Note that (iii) is equivalent to $\mathbb{E}[X_m | \mathcal{F}_n] = X_n$ $\mathbb{P}$-a.s. for all $m \geqslant n$.
- A super(sub)-martingale is a an adapted process that decreases (increases) in conditional mean.

Example 2.1. (1) **Additive martingales:** let $Y_k \in L^1$ be i.i.d. with $\mathbb{E}[Y_1] = 0$, and consider $X_n := \sum_{k=1}^n Y_k$. Then $(X_n)_{n \geqslant 1}$ is a martingale with respect to $\mathcal{F}_n = \sigma(Y_k : k \leqslant n)$.

(2) **Multiplicative martingales:** let Y_k be i.i.d. with $Y_k \in L^1$ and $\mathbb{E}[Y_k] = 1$. Consider $X_n := \prod_{k=1}^n Y_k$. Then $(X_n)_{n\geqslant 1}$ is a martingale with respect to $\mathcal{F}_n = \sigma(Y_k : \ k \leqslant n)$.

(3) For $X \in L^1(\mathbb{P})$, let $X_n := \mathbb{E}[X|\mathcal{F}_n]$, $n \geqslant 0$, for some filtration $(\mathcal{F}_n)_{n\geqslant 0}$. Then $(X_n)_{n\geqslant 0}$ is a martingale w.r.t $(\mathcal{F}_n)_{n\geqslant 0}$.

Definition 2.3. A process $(H_n)_{n\geqslant 1}$ is called *predictable* if H_n is $\mathcal{F}_{n-1}$-measurable for all $n \geqslant 1$ (sometimes also called "previsible").

For a stochastic process (X_n) and a predictable process $(H_n)_{n\geqslant 1}$, we define by $H \cdot X$ the process

$$Y_n \equiv (H \cdot X)_n = \sum_{k=1}^n H_k \Delta X_k, \quad n \geqslant 0 \quad (Y_0 \equiv 0).$$

Interpretation: "cumulative gain/loss from betting according to the strategy H". Y "$= \int H \mathrm{d}X$" is a discrete-time version of the stochastic integral of H with respect to X. If X is a martingale, then $H \cdot X$ is also called *a martingale transform*.

Theorem 2.4. *(a) Let H be predictable, bounded (i.e. there exists $K < \infty$ such that $|H_n| \leqslant K$ for all n), and X be a martingale. Then $Y = H \cdot X$ is a martingale.*

(b) Let H be predictable, bounded and non-negative, X be a supermartingale. Then $Y = H \cdot X$ is a supermartingale.

(c) Instead of boundedness of H in (a) or (b), it suffices that $H \in L^2$ when $X \in L^2$.

Definition 2.5. A map $\tau : \Omega \to \overline{\mathbb{N}_0} = \{0, 1, \ldots, \infty\}$ is called *stopping time* with respect to a filtration $(\mathcal{F}_n)_{n\geqslant 0}$ if $\{\tau \leqslant n\} \in \mathcal{F}_n$ for all $n \geqslant 0$.

Remark 2.3. $\tau : \Omega \to \overline{\mathbb{N}_0}$ is a stopping time iff $\{\tau = n\} \in \mathcal{F}_n$ for all n since (note: notation being for disjoint unions) $\{\tau \leqslant n\} = \bigsqcup_{k=0}^n \{\tau = k\}$ and $\mathcal{F}_k \subset \mathcal{F}_n$ for $k \leqslant n$.

Example 2.2.

- Let $(X_n)_{n\in\mathbb{N}}$ be an adapted process taking values in the measurable space $(S, \mathcal{S})$ and let $B \in \mathcal{S}$. Then, "the first entry time of X in B":

  $$\tau := \inf\{n \in \mathbb{N}_0 |\ X_n \in B\} \qquad (\inf \varnothing := \infty),$$

  is a stopping time. Indeed, $\{\tau \leqslant n\} = \bigcup_{k=0}^n \underbrace{\{X_k \in B\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n} \in \mathcal{F}_n$.

- Typically, a "last entry time": $\tau := \sup\{n \in \mathbb{N}_0 |\ X_n \in B\}$ would not be a stopping time as $\{\tau \leqslant n\}$ would depend on foresight (future) knowledge,

- Let τ, τ' be stopping times. Then $\tau \wedge \tau'$, $\tau \vee \tau'$, $\tau + \tau'$ are also stopping times.

- Every deterministic time n is a stopping time. In particular, $\tau \wedge n$ is a stopping time for each stopping time τ.

Remark 2.4. Intuitively, Definition 2.5 says that it is known at any time n whether stopping has occured up to that time.

Note that a finite horizon process $(X_k)_{k=0,\ldots,N}$ can be naturally embedded into the infinite horizon setting by letting $X_k := X_N$ for $k > N$ (and $\mathcal{F}_k = a\mathcal{F}_N$) for $k > N$.

Now, consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$.

Definition 2.6. Let $X = (X_n)$ be adapted process and τ be a stopping time. Then $X^\tau \equiv (X_n^\tau)$ with $X_n^\tau := X_{n\wedge\tau}$ (i.e. $X_n^\tau(\omega) = X_{n\wedge\tau(\omega)}(\omega)$ for $\omega \in \Omega$) is called the process X stopped at time τ.

Remark 2.5. Note the difference

- X^τ is a process, while
- X_τ is a random variable: $X_\tau : \omega \mapsto X_{\tau(\omega)}$ (which requires τ to be finite or, otherwise, X_∞ to be defined).

Exercise 2.1. (a) If X is an adapted process and τ is a finite stopping time, then X_τ is a random variable.

(b) If X is an adapted process and τ is a stopping time, then X^τ is an adapted process.

(c) If τ and τ' are stopping times, then $\tau \wedge \tau'$, $\tau \vee \tau'$ and $\tau + \tau'$ are stopping times.

Now, let τ be a stopping time and let $H_n^{(\tau)} := \mathbb{1}_{\{n \leqslant \tau\}}$. Note that $\{n \leqslant \tau\} = \Omega \backslash \{\tau \leqslant n-1\} \in \mathcal{F}_{n-1}$, and thus $(H_n^{(\tau)})$ is a previsible process. For a process (X_n) we have

$$(H^{(\tau)} \cdot X)_n = \sum_{k=1}^{n} \mathbb{1}_{\{k \leqslant \tau\}} \cdot (X_k - X_{k-1}) = X_{n\wedge\tau} - X_0 = X_n^\tau - X_0.$$

By Theorem 2.4 we thus obtain

Theorem 2.7 (Stopping theorem, Doob, version I)**.**

(a) Let (X_n) be a supermartingale τ be a stopping time. Then X^τ is a supermartingale. In particular, $\mathbb{E}[X_n^\tau | \mathcal{F}_k] \leqslant X_k^\tau$ and $\mathbb{E}[X_n^\tau] \leqslant \mathbb{E}[X_0]$ for all $k \leqslant n$.

(b) Let (X_n) be a martingale τ be a stopping time. Then X^τ is a martingale. In particular, $\mathbb{E}[X_n^\tau | \mathcal{F}_k] = X_k$ and $\mathbb{E}[X_n^\tau] = \mathbb{E}[X_0]$ for all $k \leqslant n$.

Remark 2.6.
- If τ is a bounded stopping time, i.e. if there exists a deterministic constant $N < \infty$ such that $\tau \leqslant N$, then Theorem 2.7 gives that $\mathbb{E}[X_\tau] \leqslant \mathbb{E}[X_0]$ (and "=" in (b)). Indeed, simply choose $n \geqslant N$.
- Stopped supermartingale remains supermartingale.

Next aim will be obtain statements as in Theorem 2.7 for σ-fields from stopping times. For that purpose, we need to define the stopped-σ-field.

Definition 2.8. Let τ be a stopping time with respect to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ in $\mathcal{F}$. Then

$$\mathcal{F}_\tau := \{A \in \mathcal{F} |\ A \cap \{\tau \leqslant n\} \in \mathcal{F}_n\ \forall n \geqslant 0\}$$

is called the σ-field of the events observable until time τ.

Exercise 2.2.
1. for deterministic stopping times $\tau = k$ we have $\mathcal{F}_\tau = \mathcal{F}_k$, hence notation in non-ambiguous.

2. Prove that $\mathcal{F}_\tau$ is a σ-field.

3. If X is an adapted process and τ is a finite stopping time, then X_τ is $\mathcal{F}_\tau$-measurable.

Lemma 2.9. *Let τ and τ' be stopping times with $\tau \leqslant \tau'$. Then $\mathcal{F}_\tau \subset \mathcal{F}_{\tau'}$.*

Corollary 2.10 (Corollary of Theorem 2.7)**.** *Let $X = (X_n)_{n\geqslant 0}$ be a martingale (or supermartinga) and σ, τ be stopping times with $\sigma \leqslant \tau$ a.s.*

(a) Then

$$\mathbb{E}[X_{\tau\wedge n}|\mathcal{F}_{\sigma\wedge n}] \overset{(\leqslant)}{=} X_{\sigma\wedge n} \quad a.s.$$

(b) If, moreover, τ and σ are bounded, then

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \overset{(\leqslant)}{=} X_\sigma \quad a.s.$$

Corollary 2.11. *1. Let (X_n) be a martingale. Then for any bounded stopping time τ holds $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.*

2. Let (X_n) be an adapted and integrable process (i.e. $X_n \in L^1$ for all n) such that for each bounded stopping time τ it holds that $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$. Then (X_n) is a martingale.

2.2 Martingale convergence theorems for a.s. convergence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $(\mathcal{F}_n)_{n\geqslant 0}$ a filtration. Let also $(X_n)_{n\geqslant 0}$ be a real-valued stochastic process. For $a, b \in \mathbb{R}$ with $a < b$ define

$$U_{a,b}^N := \#\text{ of times the process } X \text{ crosses } [a,b] \text{ for time [0,N]}, \qquad N \in \mathbb{N}.$$

In other words,

$$U_{a,b}^N = \sup\{k|\ T_k \leqslant N\},$$

where T_k is defined as follows:

$$S_0 = T_0 = 0, \quad S_k := \inf\{n|\ n \geqslant T_{k-1}, X_n \leqslant a\}, \quad T_k := \inf\{n|\ n \geqslant S_k, X_n \geqslant b\}, \quad k \geqslant 1.$$

For X being adapted, S_k and T_k are stopping times.

Lemma 2.12 (Upcrossing lemma)**.** *Let $(X_n)_{n\geqslant 0}$ be a supermartingale. Then*

$$\mathbb{E}[U_{a,b}^N] \leqslant \frac{1}{b-a}\mathbb{E}[(X_N - a)^-]. \tag{2.2.1}$$

Remark 2.7. • With $U_{a,b} = \uparrow\text{-}\lim_{N\to\infty} U_{a,b}^N = \#$ of all upcrossings, the monotone convergence theorem gives

$$\mathbb{E}[U_{a,b}] \leqslant \frac{1}{b-a}\sup_{N\in\mathbb{N}}\mathbb{E}[(X_N - a)^-]. \tag{2.2.2}$$

- The right-hand side in (2.2.2) is finite if and only if $\sup_{N \in \mathbb{N}} \mathbb{E}[X_N^-]$ is finite.
- This condition is satisfied if $X_N^- = 0$ for all N (i.e. $X \geqslant 0$) or if $\sup_{N \in \mathbb{N}} \mathbb{E}[|X_N|] < \infty$ (i.e. X is L^1-bounded). Indeed, the latter condition suffices since $\mathbb{E}[X_N^-] \leqslant \mathbb{E}[|X_N|]$.

Theorem 2.13 (Doob's martingale convergence theorem I)**.** *Let $(X_n)_{n \geqslant 0}$ be a supermartingale with $\sup_{N \geqslant 0} \mathbb{E}[X_N^-] < \infty$. Then the a.s. limit $X_\infty = \lim_{n \to \infty} X_n$ exists and is in L^1. In particular, X_∞ is finite a.s. and every non-negative supermartingale converges almost surely against a finite and integrable limit.*

Example 2.3. (1) Let Z_i be i.i.d. with $Z_i \sim \mathcal{N}(0,1)$ and let $a \geqslant \sigma^2$ for $\sigma > 0$ be fixed. Consider the random variables

$$Y_i = \exp\left(\sigma Z_i - \frac{1}{2}a\right).$$

Then $\mathbb{E}[Y_i] \leqslant 1$ because $a \geqslant \sigma^2$. Consider the non-negative process $X_n := \prod_1^n Y_i$. It is a non-negative supermartingale (cf. Example 2.1, 2). Hence, Theorem 2.13 implies that X_n converges a.s. Question: What is the limit? (Exercise)

(2) Consider a random walk $S_n = \sum_1^n Y_n$, where Y_i are i.i.d. and $p = \mathbb{P}[Y_i = +1] = 1 - \mathbb{P}[Y_i = -1]$.

Symmetric case: $p = 1/2$. In this case, S_n is a martingale (cf. Exercise 2.1, 1). Consider the stopping time $T_c := \inf\{n |\ S_n = c\}$ for some $c \in \mathbb{Z}$. Then Theorem 2.7 gives that $(S_n^{T_c})$ is a martingale. It is bounded from above if $c > 0$ and from below if $c < 0$. By the martingale convergence theorem, $\lim_{n \to \infty} S_n^{T_c}$ exists a.s. and therefore $\mathbb{P}[T_c < \infty] = 1$ for every $c \in \mathbb{Z}$. Thus,

$$\bigcap_{c \in \mathbb{Z}} \{T_c < \infty\} = \mathbb{P}[\limsup S_n = +\infty,\ \liminf S_n = -\infty] = 1.$$

Asymmetric case: $p \neq 1/2$. Then one can check that $\left(\frac{1-p}{p}\right)^{S_n}$ is a (non-trivial) non-negative martingale ($\rightsquigarrow$ exercise). Hence, Theorem 2.13 implies that it converges a.s. The limit is 0 because the Law of large numbers gives

$$\frac{1}{n} S_n \to 2p - 1 = \mathbb{E}[X_1],$$

and hence $S_n \to +\infty$ if $p > 1/2$ and $S_n \to -\infty$ if $p < 1/2$. In both cases, $\left(\frac{1-p}{p}\right)^{S_n} \to 0$. Note that $\left(\frac{1-p}{p}\right)^{S_n}$ DOES NOT converge to 0 in L^1 (expectations being 1 $\forall n$).

2.3 Uniform integrability and martingale convergence in L^1

Definition 2.14. A family $\mathcal{X}$ of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *uniformly integrable* if

$$\lim_{c \to \infty} \sup_{X \in \mathcal{X}} \mathbb{E}[|X| \mathbb{1}_{\{|X| \geqslant c\}}] = 0.$$

Notation: $\mathbb{E}[|X|; |X| \geqslant c] := \mathbb{E}[|X| \mathbb{1}_{\{|X| \geqslant c\}}]$ or $\mathbb{E}[Y;\ B] = \mathbb{E}[Y \mathbb{1}_B]$ (common notation, as e.g. in [Will91]); Writing $\mathcal{X}$ "is UI" is an abbrevation for "is uniformly integrable".

Example 2.4. Let $Y \in L^1$. Then the family $\mathcal{X} = \{\mathbb{E}[Y|\mathcal{G}]|\ \mathcal{G}$ is a sub-σ-field of $\mathcal{F}\}$ is uniformly integrable (exercise, or [Will91, Chapter 13.1], or example 2.5)

Lemma 2.15. *Each of the following is a sufficient condition for a family $\mathcal{X}$ to be uniformly integrable.*

(a) $\sup_{X \in \mathcal{X}} \mathbb{E}[|X|^p] < \infty$ *for some* $p > 1$.

(b) There exists $Y \in L^1$, such that $|X| \leqslant Y$ a.s. for all $X \in \mathcal{X}$.

Theorem 2.16. *Let $\mathcal{X}$ be a family of random variables. The following are equivalent*

(1) $\mathcal{X}$ is uniformly integrable;

(2) $\mathcal{X}$ is L^1-bounded (i.e. $\sup_{X \in \mathcal{X}} \mathbb{E}[|X|] < \infty$*) and $\forall \varepsilon > 0\ \exists \delta > 0$ such that*

$$\sup_{X \in \mathcal{X}} \mathbb{E}[|X| \mathbb{1}_A] < \varepsilon \quad \forall A \in \mathcal{F} \text{ with } \mathbb{P}[A] < \delta;$$

(3) (De La Vallée Poussin criterion:) There exists a function $G : [0, \infty) \to [0, \infty)$ with $\lim_{x \to \infty} \frac{G(x)}{x} = +\infty$ *and* $\sup_{X \in \mathcal{X}} \mathbb{E}[G(|X|)] < \infty$*; moreover, G can be chosen as increasing and convex.*

Example 2.5 (for application of the de La Vallée Poussin criterion)**.** (a) If $\mathcal{X}$ is L^p-bounded for some $p > 1$, then $\mathcal{X}$ is UI (choose $G(x) = |x|^p$ for dLVP in part (3)).

(b) Let $Z \in L^1$. Then $\mathcal{X} := \{\mathbb{E}[Z|\mathcal{G}] \mid \mathcal{G}$ is a sub-σ-field of $\mathcal{F}\}$ is uniformly integrable.

Theorem 2.17. *For a random variable X_∞ and a sequence $(X_n)_{n \in \mathbb{N}}$ in L^1, the following are equivalent:*

(1) (X_n) converges to X_∞ in L^1 and $X_\infty \in L^1$;

(2) (X_n) is UI and converges to X_∞ in probability.

Remark 2.8.
- This generalizes the dominated convergence theorem.

Remark 2.9. On the connection of uniform integrability to compactness in the weak topology $\sigma(L^1, L^\infty)$ on L^1:

- *The weak topology* $\sigma(L^1, L^\infty)$ is the coarsest topology on L^1 with respect to which all mappings $f_Z(X) := \mathbb{E}[ZX]$, $Z \in L^\infty$, are continuous. A sequence (X_n) in L^1 converges to $X_w \in L^1$ weakly if

$$\lim_n \mathbb{E}[Z X_n] = \mathbb{E}[Z X_w] \quad \forall Z \in L^\infty.$$

- A set $K \subset L^1$ is weakly (relatively) compact if K (resp. its closure) is compact with respect to the topology $\sigma(L^1, L^\infty)$.

Theorem 2.18 (Dunford, Pettis)**.** *For a set $K \subset L^1$ the following are equivalent*

(1) K is UI,

(2) K is weakly relatively compact,

(3) K is weakly relatively sequentially compact, i.e. any sequence in K has a subsequence that converges weakly (with limit in the closure of K).

Remark 2.10.

- "(2) $\Leftrightarrow$ (3)" holds on every Banach space (result by Eberlein and Smulian).
- Weak convergence of probability measures (will come later in the lecture) is just the weak-$*$-convergence of probability measures as subset of the dual space of continuous and bounded functionals.

Theorem 2.19 (Martingale convergence theorem II, L^1-convergence)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a martingale. Then the following are equivalent.*

(1) There exists a random variable $Y \in L^1$ such that $X_n = \mathbb{E}[Y|\mathcal{F}_n]$ a.s.

(2) $(X_n)_{n\in\mathbb{N}_0}$ converges in L^1 against an $\mathcal{F}_\infty$-measurable random variable X_∞.

(3) There exists a random variable $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $(X_n)_{n\in\overline{\mathbb{N}_0}}$ is a martingale (on $n \in \overline{\mathbb{N}_0}$).

(4) $(X_n)_{n\in\mathbb{N}_0}$ is UI.

Exercise 2.3. (L^p-convergence of martingales): Let $p \in (1, \infty)$. Assume that $(X_n)_{n\in\mathbb{N}_0}$ is martingale that is bounded in L^p. Then the convergence statement (2) of Theorem 2.19 holds even with convergence in L^p; i.p. X_∞ is in L^p and the other statements of that theorem hold true too (with $Y = X_\infty \in L^p$).

Remark 2.11 (On the relation of Y and X_∞)**.** In general,

$$\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A] \quad \forall A \in \mathcal{F}_n,\ \forall n \in \mathbb{N}_0.$$

Therefore,

$$\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A] \quad \forall A \in \bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n.$$

Hence, by a monotone class / Dynkin-type of argument it follows that

$$\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A] \quad \forall A \in \sigma\left(\bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n\right) \equiv \mathcal{F}_\infty,$$

i.e. $X_\infty = \mathbb{E}[Y|\mathcal{F}_\infty]$.

Corollary 2.20. *Let $Y \in L^1$ and $X_n := \mathbb{E}[Y|\mathcal{F}_n]$. Then $(X_n)_{n\in\mathbb{N}_0}$ is a uniformly integrable martingale and $X_n \to X_\infty := \mathbb{E}[Y|\mathcal{F}_\infty]$, where the convergence is a.s. and in L^1.*

Corollary 2.21 (Kolmogorov's 0-1-law)**.** *Let $Y_0, Y_1, \ldots,$ be independent random variables. Consider the σ-fields $\mathcal{A}_n := \sigma(Y_{n+1}, Y_{n+2}, \ldots)$ and $\mathcal{A} = \bigcap_{n\in\mathbb{N}} \mathcal{A}_n$. Then for every $A \in \mathcal{A}$ we have either $\mathbb{P}[A] = 0$ or $\mathbb{P}[A] = 1$.*

Example 2.6 (Approximation of functions)**.** Consider $\Omega = [0,1]$, $\mathcal{F} = \mathbb{B}(\Omega)$, $\mathbb{P}$ - the Lebesgue measure on Ω. Consider the filtration

$$\mathcal{F}_n := \sigma\left(\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right) \,\middle|\, j \leqslant 2^n - 1\right), \qquad n \in \mathbb{N}_0.$$

For a function $f \in L^1$, let $f_n := \mathbb{E}[f|\mathcal{F}_n]$. Then $f_n \to f$ in L^1 and almost everywhere.

To be able to apply the results, one needs to check that f is $\mathcal{F}_\infty$-measurable, i.e. that $\mathcal{B}([0,1]) = \mathcal{F}_\infty$. But this is true since for every $[a,b]$ with $a, b \in [0,1]$ we can take dyadic approximations $a_n \uparrow a$ and $b_n \downarrow b$ such that $[a_n, b_n) \in \mathcal{F}_n$; then $[a,b] = \bigcap_n [a_n, b_n) \in \mathcal{F}_\infty$.

2.4 Martingale inequalities

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ be a filtration.

Lemma 2.22. *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a supermartingale with $X \geqslant 0$. Then for any $\overline{\mathbb{N}_0}$-valued stopping time τ holds*

$$\mathbb{E}[X_0] \geqslant \mathbb{E}[X_\tau] \geqslant \mathbb{E}[X_\tau \mathbb{1}_{\{\tau<\infty\}}].$$

(in particular, X_τ is also defined on the set $\{\tau = \infty\}$ a.s.)

Theorem 2.23 (Doob's maximal inequality for probabilities)**.**

(1) Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a supermartingale with $X \geqslant 0$. Then for every $c > 0$

$$\mathbb{P}\left[\sup_{n\in\mathbb{N}_0} X_n \geqslant c\right] \leqslant \frac{\mathbb{E}[X_0]}{c}.$$

(2) Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a submartingale with $X \geqslant 0$. Then for every $c > 0$ and $N \in \mathbb{N}_0$

$$\begin{aligned}\mathbb{P}\left[\max_{n=0,1,\ldots,N} X_n \geqslant c\right] &\leqslant \frac{1}{c}\mathbb{E}[X_N \mathbb{1}_{\{\max_{n=0,1,\ldots,N} X_n \geqslant c\}}] \\ &\leqslant \frac{1}{c}\mathbb{E}[X_N].\end{aligned}$$

Corollary 2.24. *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a martingale. Then*

$$\mathbb{P}\left[\max_{n=0,1,\ldots,N} |X_n| \geqslant c\right] \leqslant \frac{1}{c}\mathbb{E}[|X_N|].$$

Example 2.7 (Ruin probabilities in insurance mathematics)**.** Consider the following insurance balance process

$$X_n = X_{n-1} + c_n - Y_n \quad n \in \mathbb{N},$$

where $X_0 = x_0 \in \mathbb{R}_+$ denotes the starting capital, c_n is deterministic and denotes the premiums at time n, and Y_n is stochastic and denotes the claims at time n. Assume that (X_n) is adapted to $(\mathcal{F}_n)$ (e.g. $\mathcal{F}_n = \sigma(Y_k : \ l \leqslant n)$). Let $\tau := \inf\{n \in \mathbb{N} | \ X_n \leqslant 0\}$; this is the time of technical ruin of the insurance company.

A main question in Ruin theory is to estimate $\mathbb{P}[\tau < \infty]$ (ruin probability). Note that $\{\tau < \infty\} \subset \{\inf_n X_n \leqslant 0\} = \{\sup_n(-X_n) \geqslant 0\}$. The idea is to look for a decreasing function $u : \mathbb{R} \to \mathbb{R}_+$ such that $u(X_n)$ is a supermartingale. This way we can apply the Doob's inequality (Theorem 2.23, (1)) to estimate

$$\mathbb{P}[\tau < \infty] \leqslant \mathbb{P}[\sup_n (u(X_n)) \geqslant u(0)] \leqslant \frac{u(X_0)}{u(0)} = \frac{u(x_0)}{u(0)}.$$

A specific example is when $\mathbb{E}[\exp(-\lambda(c_n - Y_n))|\mathcal{F}_{n-1}] \leqslant 1$ for all $n \in \mathbb{N}$, where $\lambda > 0$ is fixed. In this case $u(X_n) := \exp(-\lambda X_n)$ is a supermartingale, and thus

$$\mathbb{P}[\tau < \infty] \leqslant e^{-\lambda x_0}.$$

For instance, if Y_n are i.i.d. and $c_n \equiv c$, then the moment estimate $\mathbb{E}[e^{\lambda Y_1}] \leqslant e^{\lambda c}$ for some $\lambda > 0$ sufficies.

Remark:

- exponential decay of ruin probability $e^{-\lambda x_0}$ in initial capital $x_0 > 0$;
- but assuming exponential moments for claim sizes Y_n is restrictive.

The next statement is a preparation for the Doob's martingale inequalities for L^p-moments.

Lemma 2.25. *Let $U, V \geqslant 0$ be random variables with $c \cdot \mathbb{P}[V \geqslant c] \leqslant \mathbb{E}[U \mathbb{1}_{\{V \geqslant c\}}]$ for every $c > 0$. Then for any measurable function $\psi : [0, \infty) \to [0, \infty)$ with $\Psi(v) := \int_0^v \psi(z)\, \mathrm{d}z$ we have*

$$\mathbb{E}[\Psi(V)] \leqslant \mathbb{E}\left[U \int_0^V \frac{1}{c}\psi(c)\, \mathrm{d}c\right].$$

Theorem 2.26 (Doob's inequality for L^p-moments)**.** *Let $(X_n)_{n\in\mathbb{N}_0}$ be a non-negative submartingale (or a general martingale). Set $X_\infty^* := \sup_{n\in\mathbb{N}_0} |X_n|$. Then for every $p > 1$ there exist constants $c_p, C_p > 0$, depending only on p but not on X, such that*

$$c_p \cdot \sup_{n\in\mathbb{N}_0} \|X_n\|_{L^p} \overset{(a)}{\leqslant} \|X_\infty^*\|_{L^p} \overset{(b)}{\leqslant} C_p \cdot \sup_{n\in\mathbb{N}_0} \|X_n\|_{L^p}.$$

Remark 2.12.

- For $p > 1$ and a non-negative submartingale X we have: X is bounded in $L^p \iff X_\infty^*$ is in L^p.
- As we saw in the proof of Theorem 2.26, the inequality (a) holds even for $p = 1$, but in general (b) does not necessarily hold for $p = 1$.
- Note that, by stopping, the inequality holds up to any (finite) time horizon.

2.5 The Radon-Nikodym theorem and differentiation of measures

Let $\mathbb{P}$ and $\mathbb{Q}$ be finite measures on $(\Omega, \mathcal{F})$.

Definition 2.27. (a) $\mathbb{Q}$ is *absolutely continuous* w.r.t $\mathbb{P}$ on $\mathcal{F}$, denoted by $\mathbb{Q} \ll \mathbb{P}$, if the following holds

$$\forall A \in \mathcal{F}: \qquad \mathbb{P}[A] = 0 \Rightarrow \mathbb{Q}[A] = 0.$$

(b) $\mathbb{Q}$ is *equivalent* to $\mathbb{P}$ on $\mathcal{F}$, denoted by $\mathbb{Q} \approx \mathbb{Q}$, if $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$ on $\mathcal{F}$, i.e. the null sets of $\mathbb{P}$ and $\mathbb{Q}$ are the same.

(c) $\mathbb{Q}$ is *singular* to $\mathbb{P}$ on $\mathcal{F}$, denoted by $\mathbb{Q} \perp \mathbb{P}$, if there exists $A \in \mathcal{F}$ with $\mathbb{P}[A] = 0$ and $\mathbb{Q}[A^c] = 0$, i.e. $\mathbb{Q}$ is supported on a $\mathbb{P}$-null set and vice versa.

Example 2.8. (1) Let $\mathbb{P}$ be a continuous distribution on $\mathbb{R}$ and $\mathbb{Q}$ be a discrete distribution on $\mathbb{R}$. Then $\mathbb{P} \perp \mathbb{Q}$.

(2) Let $\mathbb{P}$ and $\mathbb{Q}$ be continuous distributions on $\mathbb{R}^d$ with density functions f and g (w.r.t Lebesgue measure λ). Then

$$\begin{aligned} \mathbb{Q} \ll \mathbb{P} \quad &\Longleftrightarrow \quad \{g = 0\} \supseteq \{f = 0\} \quad \lambda\text{-a.e.}, \\ \mathbb{Q} \approx \mathbb{P} \quad &\Longleftrightarrow \quad \{g = 0\} = \{f = 0\} \quad \lambda\text{-a.e.}. \end{aligned}$$

(3) Let $X_1, X_2, \ldots,$ be i.i.d Bernouli random variables, $\{0,1\}$-valued, with success probability p and q under $\mathbb{P}$ and $\mathbb{Q}$ respectively. $\mathbb{P}$ and $\mathbb{Q}$ are measures on $\Omega = \{0,1\}^{\mathbb{N}}$, $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$, $\mathcal{F} := \bigvee_{n \in \mathbb{N}} \mathcal{F}_n = \sigma(X_n | \, n \in \mathbb{N})$. Then $\mathbb{P} \approx \mathbb{Q}$ on $\mathcal{F}_n$ for every $n \in \mathbb{N}$ if $p, q \in (0,1)$: for every non-empty $A \in \mathcal{F}_n$ we have that $\mathbb{P}[A] > 0$ and $\mathbb{Q}[A] > 0$. However, for $p \neq q$ we have that $\mathbb{Q} \perp \mathbb{P}$ on $\mathcal{F}$. Indeed, by the Law of large numbers we have that $\frac{1}{n}\sum_{k=1}^n X_k \to p$ $\mathbb{P}$-a.s., and $\frac{1}{n}\sum_{k=1}^n X_k \to q$ $\mathbb{Q}$-a.s., i.e. for $A := \left\{\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k = q\right\} \in \mathcal{F}$ we have that $\mathbb{P}[A] = 0$ and $\mathbb{Q}[A] = 1$.

Lemma 2.28. *Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$, $X \geqslant 0$ be a random variable in $L^1(\mathbb{P})$, i.e. $\mathbb{E}_{\mathbb{P}}[X] < \infty$. Then*

$$\mathbb{Q}[A] := \mathbb{E}_{\mathbb{P}}[X \mathbb{1}_A] = \int_A X \, \mathrm{d}\mathbb{P}, \qquad A \in \mathcal{F},$$

defines a finite measure on $(\Omega, \mathcal{F})$ with $\mathbb{Q} \ll \mathbb{P}$.

Remark 2.13.
- For $Y \geqslant 0$, $\mathcal{F}$-measurable, we have $\mathbb{E}_{\mathbb{Q}}[Y] = \mathbb{E}_{\mathbb{P}}[XY]$ (by standard approximation).
- The Radon-Nikodym theorem will show that $\mathbb{Q} \ll \mathbb{P}$ is sufficient for having a representation as in Lemma 2.28.

More generally, we will next study the Lebesgue decomposition.

Definition 2.29. *Lebesgue decomposition* of a probability measure $\mathbb{Q}$ with respect to a probability measure $\mathbb{P}$ on $\mathcal{F}$ is a decomposition of the form

$$\boxed{\mathbb{Q}[A] = \mathbb{E}_{\mathbb{P}}[X\mathbb{1}_A] + \mathbb{Q}[A \cap \{X = \infty\}] \qquad \forall A \in \mathcal{F}}, \tag{2.5.1}$$

where $X \in L^1(\mathbb{P})$ is $\mathcal{F}$-measurable and non-negative. Such a random variable is called *generalized Radon-Nikodym derivative* of $\mathbb{Q}$ w.r.t $\mathbb{P}$ on $\mathcal{F}$, denoted by

$$X = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}}.$$

Remark 2.14. A Lebesgue decomposition of a measure $\mathbb{Q}$ w.r.t $\mathbb{P}$ is a decomposition of $\mathbb{Q}$ into an absolutely continuous part and a singular part. More precisely, let $\mathbb{Q}_a(A) := \mathbb{E}_{\mathbb{P}}[X\mathbb{1}_A]$ and $\mathbb{Q}_s[A] := \mathbb{Q}[A \cap \{X = \infty\}]$ for $A \in \mathcal{F}$. Then $\mathbb{Q} = \mathbb{Q}_a + \mathbb{Q}_s$; also, $\mathbb{Q}_a \ll \mathbb{P}$ on $\mathcal{F}$ by Lemma 2.28 and $\mathbb{Q}_s \perp \mathbb{P}$ because $\mathbb{P}[X = \infty] = 0$ since $X \in L^1(\mathbb{P})$.

Lemma 2.30. *Let $\mathbb{Q}$ have a Lebesgue decomposition w.r.t $\mathbb{P}$ on $\mathcal{F}$ with a generalized Radon-Nikodym derivative X. Then*

(1) $X > 0$ $\mathbb{Q}$-a.s.

(2) $\mathbb{P}$ has a Lebesgue decomposition w.r.t $\mathbb{Q}$ on $\mathcal{F}$ with a generalized Radon-Nikodym derivative given by

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\Big|_{\mathcal{F}} = \frac{1}{X}.$$

Example 2.9. Let $\mathcal{F}$ be generated by a countable partition of Ω: $\Omega = \bigsqcup_{i=1}^{\infty} B_i$ for (disjoint) $B_i \in \mathcal{F}$ and $\mathcal{F} = \sigma(B_i : i \in \mathbb{N})$. Set

$$\begin{aligned} X &= \sum_{i=1}^{\infty} \mathbb{1}_{B_i} \cdot \left(\frac{\mathbb{Q}[B_i]}{\mathbb{P}[B_i]} \mathbb{1}_{\{\mathbb{P}[B_i] \neq 0\}} + \infty \cdot \mathbb{1}_{\{\mathbb{P}[B_i]=0\}} \right) \\ &= \begin{cases} \mathbb{Q}[B_i]/\mathbb{P}[B_i] & \text{if } \omega \in B_i \text{ with } \mathbb{P}[B_i] \neq 0 \\ \infty & \text{if } \omega \in B_i \text{ with } \mathbb{P}[B_i] = 0 \end{cases}. \end{aligned}$$

Then $\mathbb{P}[X = \infty] = 0$ and $\mathbb{E}_{\mathbb{P}}[X] = \sum_{i:\mathbb{P}[B_i]\neq 0} \mathbb{Q}[B_i]$ (being $< \infty$ because $\mathbb{Q}$ is a probability measure). Each set $A \in \mathcal{F}$ is of the form $A = \bigsqcup_{i \in J} B_i$ with $J \subseteq \mathbb{N}$. Hence

$$\begin{aligned} \mathbb{Q}[A] = \sum_{j \in J} \mathbb{Q}[B_j] &= \sum_{\substack{j \in J \\ \mathbb{P}[B_j] \neq 0}} \frac{\mathbb{Q}[B_i]}{\mathbb{P}[B_i]} \mathbb{P}[B_i] + \sum_{\substack{j \in J \\ \mathbb{P}[B_j] = 0}} \mathbb{Q}[B_i] \\ &= \mathbb{E}_{\mathbb{P}}[X\mathbb{1}_A] + \mathbb{Q}(A \cap \{X = \infty\}), \end{aligned}$$

i.e. $X = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}}$.

Remark 2.15. The example showed that the Lebesgue decomposition for any two measures $\mathbb{Q}$ and $\mathbb{P}$ on a σ-field generated by a countable partition always exists!

Let us fix the set-up for the statements that follow. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$ with a filtration $(\mathcal{F}_n)$. Set $\mathcal{F}_\infty := \bigvee_n \mathcal{F}_n$. Assume that Lebesgue decompositions of $\mathbb{Q}$ with respect to $\mathbb{P}$ on any $\mathcal{F}_n$ exist, providing random variables $X_n \geqslant 0$, $X_n \in L^1(\mathbb{P}, \mathcal{F}_n)$ such that

$$\mathbb{Q}[A] = \underbrace{\mathbb{E}_{\mathbb{P}}[X_n \mathbb{1}_A]}_{=:\mathbb{Q}_a[A]} + \underbrace{\mathbb{Q}[A \cap \{X_n = \infty\}]}_{=:\mathbb{Q}_s[A]} \quad \forall A \in F_n$$

Remark 2.16.
- If $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}_n$, then $\mathbb{Q}_s[A] = 0$ and $\mathbb{Q} = \mathbb{Q}_a$ on $\mathcal{F}_n$.
- If $\mathbb{Q} \perp \mathbb{P}$ (on $\mathcal{F}_\infty$) with gen.RN-derivative X_∞, then $\mathbb{Q}_s = \mathbb{Q}$, $\mathbb{Q}_a = 0$, and $X_\infty = 0$ $\mathbb{P}$-a.s.

Theorem 2.31. *In the set-up defined above, the following holds true*

(1) (X_n) is a $\mathbb{P}$-supermartingale and $(\frac{1}{X_n})$ is a $\mathbb{Q}$-supermartingale.

(2) $X_\infty := a.s\text{-}\lim_{n\to\infty} X_n$ exists and

$$X_\infty = \begin{cases} \in [0,\infty) & \mathbb{P}\text{-}a.s. \\ \in (0,\infty] & \mathbb{Q}\text{-}a.s. \end{cases}.$$

(3) The Lebesgue decomposition of $\mathbb{Q}$ w.r.t $\mathbb{P}$ on $\mathcal{F}_\infty$ is given by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_\infty} = X_\infty.$$

(4) $\mathbb{Q}$ is locally absolutely continuous w.r.t $\mathbb{P}$ ("$\mathbb{Q} \overset{loc}{\ll} \mathbb{P}$"), i.e.

$$\mathbb{Q} \ll \mathbb{P} \quad \text{on } \mathcal{F}_n \ \forall n,$$

if and only if (X_n) is a $\mathbb{P}$-martingale with $\mathbb{E}_{\mathbb{P}}[X_n] = 1$.

(5) $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}_\infty$ if and only if $\mathbb{Q} \overset{loc}{\ll} \mathbb{P}$ and (X_n) is $\mathbb{P}$-UI, or (equivalently) if $\mathbb{Q}[X_\infty = \infty] = 0$.

Definition 2.32. A σ-field is called *separable* if it is generated by a countable family $(A_i)_{i\in\mathbb{N}}$ of sets.

Example 2.10. (1) $\mathcal{B}(\mathbb{R}^d)$ is separable (generated by rectangles with rational vertices).

(2) Borel σ-field of a Polish space is separable.

Remark:

- The sets A_i, $i \in \mathbb{N}$, do not need to be disjoint. But for any n one can find finitely many disjoint sets $A'_1, \ldots, A'_m$ such that $\sigma(A_1, \ldots, A_n) = \sigma(A'_1, \ldots, A'_m)$. For instance, consider the finite partition generated by all intersections of sets A_i or A_i^c, for $i \leqslant n$. In this sense, each $\mathcal{F}_n = \sigma(A_1, \ldots, A_n)$ may be assumed to be generated by a finite number of disjoint sets, or by a finite partition of Ω.

- Thus, every separable σ-field can be represented as $\bigvee_{n\in\mathbb{N}} \mathcal{F}_n$ as in Thm.2.31 with each $\mathcal{F}_n$ being generated by a finite partition.

- Recall that for every σ-field $\mathcal{F}_n$ generated by a countable partition the Lebesgue decomposition of $\mathbb{Q}$ with respect to $\mathbb{P}$ exists for any probability measures $\mathbb{P}$ and $\mathbb{Q}$ (see Example).

Theorem 2.33 (Radon-Nikodym)**.** *Let $\mathbb{Q}$ and $\mathbb{P}$ be finite measures on $(\Omega, \mathcal{F})$ with $\mathbb{Q} \ll \mathbb{P}$. Then there exists $X \in L^1(\mathbb{P}, \mathcal{F})$, $X \geqslant 0$, such that*

$$\mathbb{Q}[A] = \mathbb{E}_{\mathbb{P}}[X \mathbb{1}_A] = \int_A X \, \mathrm{d}\mathbb{P} \qquad \forall A \in \mathcal{F}.$$

We write $X = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}}$ and call it the Radon-Nikodym derivative (or density) of $\mathbb{Q}$ with respect to $\mathbb{P}$ on $\mathcal{F}$.

The same statement holds more generally for P being only σ-finite. For both $\mathbb{P}, \mathbb{Q}$ being only σ-finite, the statement holds too except that the density $X \geqslant 0$ does not need to be in $L^1(\mathbb{P})$.

Note that the density is unique up to equality P-almost everywhere..

Example 2.11 (Likelihood-ratio test)**.** Let $(Y_i)_{i\in\mathbb{N}}$ be i.i.d. under $\mathbb{P}$ and $\mathbb{Q}$ real-valued random variables with Y_i having distribution μ under $\mathbb{P}$ and ν under $\mathbb{Q}$.

Question: are successive observations of (Y_i) being drawn from $\mathbb{P}$ or $\mathbb{Q}$?

Set-up: $\Omega = \mathbb{R}^{\mathbb{N}}$ with $Y_i(\omega) = \omega_i$ (coordinate mappings), $\mathcal{F}_n = \sigma(Y_k : \ k \leqslant n)$, $\mathcal{F} = \mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$, $\mathbb{P} = \mu^{\otimes\mathbb{N}}$ and $\mathbb{Q} = \nu^{\otimes\mathbb{N}}$.

Note that knowing the whole sequence, the decision between $\mathbb{Q}$ and $\mathbb{P}$ is easy: for every $B \in \mathcal{B}(\mathbb{R})$ we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{1}_B(Y_i) = \begin{cases} \mu(B) & \mathbb{P}\text{-a.s.,} \\ \nu(B) & \mathbb{Q}\text{-a.s.} \end{cases}$$

So take $B \in \mathcal{B}(\mathbb{R})$ such that $\mu(B) \neq \nu(B)$ and observe the limit, or rather the sequence, above. What to conclude based on only finitely many observations?

Suppose that $\nu \ll \mu$ with $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = h$ on $\mathbb{R}$. For instance, μ and ν have densities f and g w.r.t the Lebesgue measure λ on $\mathbb{R}$, in which case one needs $\{g = 0\} \supseteq \{f = 0\}$ λ a.s., and $h = g/f$:

$$\nu(B) = \int_B g \, \mathrm{d}\lambda = \int_B \frac{g}{f} f \, \mathrm{d}\lambda = \mathbb{E}_\mu[h \mathbb{1}_B] \qquad \forall B \in \mathcal{B}(\mathbb{R}).$$

Then $\mathbb{Q} \overset{\text{loc}}{\ll} \mathbb{P}$ and $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}_n} = X_n := \prod_1^n h(Y_i)$: Indeed, let $A = \times_1^n Y_i^{-1}(B_i)$ for $B_i \in \mathcal{B}(\mathbb{R})$, then

$$
\begin{aligned}
\mathbb{Q}[A] &= \prod_1^n \mathbb{Q}[Y_i \in B_i] && (Y_i \text{ are } \mathbb{Q}\text{-i.i.d.}) \\
&= \prod_1^n \int_{B_i} g \,\mathrm{d}\lambda = \prod_1^n \int_{B_i} \underbrace{\frac{g}{f}}_{=h} f \,\mathrm{d}\lambda \\
&= \prod_1^n \mathbb{E}_{\mathbb{P}}[h(Y_i)\mathbb{1}_{B_i}(Y_i)] \\
&= \mathbb{E}_{\mathbb{P}}\left[\left(\prod_1^n h(Y_i)\right) \mathbb{1}_{\times_1^n B_i}(Y_1, \ldots, Y_n)\right] && (Y_i \text{ are } \mathbb{P}\text{-i.i.d.}) \\
&= \mathbb{E}_{\mathbb{P}}[X_n \mathbb{1}_A],
\end{aligned}
$$

and since such sets form $\cap$-stable generator of $\mathcal{B}(\mathbb{R}^n)$, we conclude that $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}_n} = X_n$ (using standard arguments/monotone class theorem).

However, if $\mu \neq \nu$, then $\mathbb{Q} \perp \mathbb{P}$ on $\mathcal{F}_\infty$ (take B such that $\mu(B) \neq \nu(B)$, then use the Law of large numbers to construct a suitable set). Also, by Theorem 2.31 and Remark 2.16 we have that $\lim_{n\to\infty} X_n = X_\infty = 0$ $\mathbb{P}$-a.s. and $\lim_{n\to\infty} X_n = \infty$ $\mathbb{Q}$-a.s.

Decision between $\mathbb{P}$ and $\mathbb{Q}$ can be made by observing the sequence of the likelihood ratios X_n. More precisely, the law of large numbers gives

$$
\frac{1}{n} \log X_n = \frac{1}{n} \sum_1^n \log h(Y_i) \to \begin{cases} \mathbb{E}_{\mathbb{Q}}[\log h(Y_1)] & \mathbb{Q}\text{-a.s.,} \\ \mathbb{E}_{\mathbb{P}}[\log h(Y_1)] & \mathbb{P}\text{-a.s.,} \end{cases}
$$

assuming that $\log h(Y_1) \in L^1(\mathbb{Q})$ or $\in L^1(\mathbb{P})$ respectively.

Set

$$
\mathbb{E}_{\mathbb{Q}}[\log h(Y_1)] = \int \log h \,\mathrm{d}\nu = \int \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\nu =: H(\nu|\mu),
$$

the relative entropy of ν with respect to μ, and we have respectively when also $\mu \ll \nu$, i.e. $\mu \approx \nu$,

$$
\mathbb{E}_{\mathbb{P}}[\log h(Y_1)] = \int \log h \,\mathrm{d}\mu = \int -\log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \,\mathrm{d}\mu = -H(\mu|\nu).
$$

Exercise: $H(\nu|\mu) \geqslant 0$ with equality iff $\mu = \nu$.

If both entropies are finite, we have asymptotically

$$
X_n \sim \begin{cases} \exp(n \cdot H(\nu|\mu)) & \mathbb{Q}\text{-a.s.,} \\ \exp(-n \cdot H(\mu|\nu)) & \mathbb{P}\text{-a.s.} \end{cases}
$$

The relative entropy intuitively gives the rate of convergence of (X_n) towards $+\infty$ or 0.

Example 2.12 ("Exponential tilting")**.** (a) Let $Y_i \sim \mathrm{Exp}(1)$ i.i.d. under $\mathbb{P}$ (with $\mu := \mathrm{Exp}(1)$). Can we construct a change to an absolutely continuous probability measure $\mathbb{Q}$ such that $Y_1, \ldots, Y_n$ are i.i.d. with distribution $\mathrm{Exp}(\lambda) =: \nu$ under $\mathbb{Q}$ for a given parameter $\lambda > 0$?

Ansatz: For $h(x) = \frac{d\nu}{dx} \cdot \frac{dx}{d\mu}(x) = \frac{\lambda e^{-\lambda x}}{e^{-x}}$, note that $h \geqslant 0$ and $\mathbb{E}_\mu[h] = 1$. So if we define

$$\nu(B) := \mathbb{E}_\mu[\mathbb{1}_B h] \qquad \forall B \in \mathcal{B}(\mathbb{R}),$$

we have that $\nu = \mathrm{Exp}(\lambda)$ and with

$$X_n := \prod_1^n h(Y_i) = \frac{d\nu^{\otimes n}}{d\mu^{\otimes n}}$$

one can describe a probability distribution $d\mathbb{Q} = X_n d\mathbb{P}$ (on any $\mathcal{F}_n$): Indeed, for all $B_i \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned}
\mathbb{Q}[Y \in B] &= \mathbb{E}_\mathbb{P}[\prod_1^n h(Y_i)\mathbb{1}_B(Y_1, \dots, Y_n)] \\
&= \prod_{i=1}^n \mathbb{E}_\mathbb{P}[h(Y_i)\mathbb{1}_{B_i}(Y_i)] && (Y_i \text{ are i.i.d.}) \\
&= \prod_{i=1}^n \int_{B_i} h(x)e^{-x}\,dx = \prod_{i=1}^n \int_{B_i} \lambda \exp(-\lambda x)\,dx.
\end{aligned}$$

Thus, $Y_1, \dots, Y_n$ are i.i.d. and $\mathrm{Exp}(\lambda)$-distributed under $\mathbb{Q}$.

However, the measure $\mathbb{Q} = \nu^{\otimes\mathbb{N}}$, under which all Y_i, $i \in \mathbb{N}$, are i.i.d. $\mathrm{Exp}(\lambda)$-distributed, is singular to $\mathbb{P} = \mu^{\otimes\mathbb{N}}$ on $\mathcal{F}_\infty = \sigma(Y_i : \ i \in \mathbb{N})$, although $\nu^{\otimes n} \approx \mu^{\otimes n}$, i.e. $\mathbb{Q} \approx \mathbb{P}$ on $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$.

Note that (X_n) is a $\mathbb{P}$-martingale but not $\mathbb{P}$-UI.

(b) Let $Y_i \sim \mathcal{N}(0,1)$ be i.i.d. under $\mathbb{P}$ ($\mu := \mathcal{N}(0,1)$). For $m \in \mathbb{R}$, define ν on $\mathcal{B}(\mathbb{R})$ through $h(x) = \exp\left(mx - \frac{1}{2}m^2\right) = \exp\left(-\frac{(x-m)^2}{2}\right) / \exp\left(-x^2/2\right)$ by

$$\begin{aligned}
\nu(B) = \mathbb{E}_\mu[h\mathbb{1}_B] &= \int_B h\,d\mu \\
&= \int_B \frac{\varphi_{m,1}}{\varphi_{0,1}}\varphi_{0,1}\,dx \\
&= \int_B \varphi_{m,1}\,dx,
\end{aligned}$$

where $\varphi_{a,b}$ is the density of $\mathcal{N}(a,b)$ w.r.t dx. Thus, $\nu = \mathcal{N}(m,1)$ distribution. Now set

$$\frac{d\nu^{\otimes n}}{d\mu^{\otimes n}} = \prod_1^n h(Y_i) =: X_n.$$

(X_n) is a $\mathbb{P}$-martingale but not $\mathbb{P}$-UI because $\mu^{\otimes\mathbb{N}} \perp \nu^{\otimes\mathbb{N}}$ on the canonical probability space $(\mathbb{R}^\mathbb{N}, \mathcal{B}(\mathbb{R}^\mathbb{N}))$ since $\mu \neq \nu$ although $\mu^{\otimes n} \approx \nu^{\otimes n}$ for every $n \in \mathbb{N}$ (because $X_n > 0$).

2.6 A brief excursion on Markov chains

We do a brief excusion on Markov chains, postponing question about construction of the respective stochastic model to the next chapter. Indeed, the construction is a direct application (using Markov kernels) of the general contruction for stochastic processes in discrete time provided there. The idea of Markov transition kernels also provides an intuitive example that motivates the more general framework (more general kernels) in the next chapter.

- Markov chain: Markov process in discrete time with countable (discrete) state space.
- Setting: E is a countable state space, $i \in E$ are the states, $\mathbb{N} = \{0, 1, \ldots\}$ is the time index set.

Definition 2.34 (Markov property)**.** A process $(X_n)_{n\in\mathbb{N}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has the *Markov property* (w.r.t. $\mathbb{P}$) if $\forall n\ \forall i_0, i_1, \ldots, i_{n+1} \in E$

$$\mathbb{P}\left[X_{n+1} = i_{n+1} \middle| X_{[0,n]} = i_{[0,n]}\right] = \mathbb{P}\left[X_{n+1} = i_{n+1} \middle| X_n = i_n\right]$$

(whereever defined), i.e. if $P\left[X_{n+1} = i_{n+1} \middle| X_{[0,n]}\right] = \mathbb{P}\left[X_{n+1} = i_{n+1} \middle| X_n\right]$ a.e..

- Notational abbreviation used: $i_{[s,t]} \equiv (i_s, \ldots, i_t) \in E^{t-s+1}$, $X_{[s,t]} \equiv (X_s, \ldots, X_t)$.
- On E we consider the *discrete topology*, $\mathcal{B}(E) = 2^E$.
- Stochastic kernel $K : E \times \mathcal{B}(E) \to [0, 1]$ from E to E is determined by a *stochastic matrix* $\mathcal{P} = (p_{ij})_{i,j\in E}$ with $p_{ij} = K(i, \{j\})$.
- A stochastic kernel K is also called *Markov kernel*, and $\mathcal{P}$ is called *transition probability matrix*, as K (resp. $\mathcal{P}$) describes the dynamics of transitions for a *Markov chain.*

Definition 2.35 (Markov chain)**.** $(X_n)_{n\in\mathbb{N}}$ is called (time-homogeneous) *Markov chain* (under $\mathbb{P}$) with initial (starting) distribution ν on E and stochastic kernel K represented by a stochastic matrix $\mathcal{P}$, if $\forall n \in \mathbb{N}\ \forall i_0, \ldots, i_{n+1} \in E$

$$\mathbb{P}\left[X_{n+1} = i_{n+1} \middle| X_{[0,n]} = i_{[0,n]}\right] = p_{i_n i_{n+1}} = K(i_n, \{i_{n+1}\}) \tag{2.6.1}$$

(whereever conditional probability is defined), i.e. if

$$\mathbb{P}\left[X_{n+1} = i_{n+1} \middle| X_{[0,n]}\right] = p_{i_n i_{n+1}}\big|_{i_n = X_n} = K(X_n, \{i_{n+1}\}) \quad \text{a.e..}$$

Notation: given a transition matrix/kernel, the distribution of the process $(X_n)_{n\in\mathbb{N}}$ is determined by ν and one often writes $\mathbb{P}_\nu$ for $\mathbb{P}$ to indicate the initial distribution, or $\mathbb{P}_i$ for $\mathbb{P}_\nu$ when $\nu = \delta_i$ is the point mass at i.

In the sequel we will work on the ***canonical probability space*** of paths for a Markov chain: $\Omega = E^{\mathbb{N}}$, $\mathcal{F} = \sigma(\pi_n :\ n \in \mathbb{N}) = \sigma(X_n :\ n \in \mathbb{N}) = \mathcal{B}(E)^{\mathbb{N}}$, $X_n = \pi_n$ for the canonical projections $\pi_n(\omega) = \omega_n$. This permits to define a family $\mathbb{P}_i$, $i \in E$, of probability measures *simultaneously* on the same space $(\Omega, \mathcal{F})$ with the canonical mapping $X : \mathbb{N} \times \Omega \to E$,

$X_n(\omega) = \omega_n$, such that $X = (X_n)_{n\in\mathbb{N}}$ is a Markov chain with initial distribution δ_i (starting at $i \in E$) under $\mathbb{P}_i$ for any i. By Th. 2.36(a) one can also define $\mathbb{P}_\nu$ on $(\Omega, \mathcal{F})$ for any initial probability distribution ν on E.

Any question that depends only the distributional properties of the Markov process (e.g. about probabilities or expectations depending on the distribution of the process (X_n)) can be addressed on the canonical space, since the distribution on $(E^{\mathbb{N}}, \mathcal{B}(E)^{\otimes\mathbb{N}})$ of a Markov chain X (defined on some probability space) with initial distribution $\nu := P \circ X_0^{-1}$ and given transition matrix is identical to the distribution of the canonical process π on the canonical space under $\mathbb{P}_\nu$.

Indeed part (a) of Theorem 2.36 shows that the distribution of $X_{[0,n]}$ under $\mathbb{P}$ is uniquely determined on $\mathcal{F}_n$ for any $n \in \mathbb{N}$. As $\cup_n \mathcal{F}_n$ is an intersection-stable generator of $\mathcal{F}$, the distributions of the processes X and π on $\mathcal{B}(E)^{\otimes\mathbb{N}}$ are the same (uniqueness) - if they exists. Existence we postpone to the next chapter - it will be obtained by Kolmogorov's consistency theorem.

Theorem 2.36. *Let $(X_n)_{n\in\mathbb{N}}$ be E-valued stochastic process.*

(a) $(X_n)_{n\in\mathbb{N}}$ is a Markov chain (under P) with transition kernel (matrix) K ($\mathcal{P}$) and initial distribution $\mathbb{P} \circ X_0^{-1} = \nu$ if and only if $\forall n \in \mathbb{N}\ \forall i_0, \ldots, i_n \in E$

$$\mathbb{P}\left[X_{[0,n]} = i_{[0,n]}\right] = \nu(i_0) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.$$

(b) If $(X_n)_{n\in\mathbb{N}}$ is a Markov chain, then $\forall n < m$, $i_n \in E$, $A \subset E^n$, $B \subset E^{m-n}$

$$\begin{aligned}\mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_{[0,n-1]} \in A, X_n = i_n\right] &= \mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_n = i_n\right] \\ &= \mathbb{P}_{i_n}\left[X_{[1,m-n]} \in B\right] \end{aligned} \tag{2.6.2}$$

(whereever defined), i.e.

$$\mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_0, \ldots, X_n\right] = \mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_n\right],$$

meaning that future evolution of Markov chains only depends on the history up to today through the present state.

Theorem 2.37 (Conditional independence of past and future given the present)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain. Then $\forall n < m$, $i_n \in E$, $A \subset E^n$, $B \subset E^{m-n}$ we have*

$$\mathbb{P}\left[X_{[0,n-1]} \in A, X_{[n+1,m]} \in B \middle| X_n = i_n\right] = \mathbb{P}\left[X_{[0,n-1]} \in A \middle| X_n = i_n\right] \cdot \mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_n = i_n\right]$$

(whereever defined), i.e.

$$\mathbb{P}\left[X_{[0,n-1]} \in A, X_{[n+1,m]} \in B \middle| X_n\right] = \mathbb{P}\left[X_{[0,n-1]} \in A \middle| X_n\right] \cdot \mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_n\right] \quad a.e.$$

2.6.1 First-entry times of Markov chains

We consider the natural filtration $\mathcal{F}_n = \sigma(X_k :\ k \leqslant n)$ generated by the process (X_n). With respect to $(\mathcal{F}_n)$, the first-entry times of a set $A \subset E$

$$\begin{aligned} T_A &:= \inf\{n \geqslant 1 :\ X_n \in A\}, \\ H_A &:= \inf\{n \geqslant 0 :\ X_n \in A\}, \end{aligned}$$

are stopping times, possibly taking value $+\infty$. Note that $\mathbb{P}_i[H_A = 0] = 1$ for $i \in A$ and $\mathbb{P}_i[H_A = T_A] = 1$ for $i \notin A$. Define $h_A(i) := \mathbb{P}_i[H_A < \infty]$ as the probability of reaching A eventually when starting at i, and define $k_A(i) := \mathbb{E}_i[H_A]$ as the expected time for reaching A from i (expectation is taken under $\mathbb{P}_i$).

Theorem 2.38. *Let $A \subset E$ be non-empty. Then $h_A : E \to [0,1]$ is the smallest function $E \mapsto [0,\infty)$ solving*

$$h_A(i) = \sum_{j \in E} p_{ij} h_A(j) \quad \forall i \in E \backslash A$$

with boundary conditions $h_A(i) = 1$ for $i \in A$.

For the proof of the theorem we will need the following lemma.

Lemma 2.39. *For every $i \in E \backslash A$, $n \in \mathbb{N}$, $A \neq \varnothing$ we have*

$$\mathbb{P}_i[T_A \leqslant n+1] = \sum_{j \in E} p_{ij} \mathbb{P}_j[H_A \leqslant n].$$

Theorem 2.40. *k_A is the minimal function $E \mapsto [0,\infty)$ solving*

$$k_A(i) = 1 + \sum_{j \notin A} p_{ij} k_A(j) \quad \forall i \notin A$$

with boundary condition $k_A(i) = 0$ for $i \in A$.

2.6.2 Further properties of Markov chains

For $n \in \mathbb{N}$ define the shift operator $\theta_n : \Omega \to \Omega$, $\omega = (\omega_j)_{j \in \mathbb{N}} \mapsto (\omega_{n+j})_{j \in \mathbb{N}}$.

Theorem 2.41. *Let $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable and nonnegative (or bounded). Then for any $n \geqslant 0$ we have*

$$\mathbb{E}_\nu[g \circ \theta_n \,|\, \mathcal{F}_n] = \mathbb{E}_i[g]\big|_{i = X_n} =: \mathbb{E}_{X_n}[g]$$

Instead of just 1-step transition probabilities, one can also consider multi-step transition probabilities

$$p_{ij}^{(n)} := \mathbb{P}[X_{k+n} = j | X_k = i] = P_i[X_n = j]$$

and those are being related according to

Theorem 2.42 (Chapman-Kolmogorov)**.** *For $m, n \in \mathbb{N}$, $i, j \in E$ we have*

$$\mathbb{P}_i[X_{n+m} = j] = \sum_{y \in E} \mathbb{P}_i[X_n = y] \mathbb{P}_y[X_m = j]\,.$$

For Markov processes in general the analogue of the next property would be stronger than the Markov property itself, but for Markov chains both are equivalent:

Theorem 2.43. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain and $T < \infty$ be a finite stopping time, with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then*

$$\mathbb{P}\left[X_{[T+1,T+k]} \in B \middle| X_T = i, A\right] = \mathbb{P}_i\left[X_{[1,k]} \in B\right], \tag{2.6.3}$$

for $k \in \mathbb{N}$, $A \in \mathcal{F}_T$, $B \subset E^k$, $i \in E$ (wherever defined), such that

$$\mathbb{P}\left[X_{[T+1,T+k]} \in B \middle| \mathcal{F}_T\right] = \mathbb{P}_i\left[X_{[1,k]} \in B\right]\Big|_{i=X_T} =: \mathbb{P}_{X_T}\left[X_{[1,k]} \in B\right]. \tag{2.6.4}$$

Remark 2.17. For $T \equiv n$, $m = n + k$, we have for $A = \{X_{[0,n]} \in \tilde{A}\}$ for some $\tilde{A} \in E^{n+1}$ in (2.6.3):

$$\mathbb{P}\left[X_{[n+1,m]} \in B \middle| X_n = i, X_{[0,T]} \in \tilde{A}\right] = \mathbb{P}_i\left[X_{[1,m-n]} \in B\right],$$

i.e. the ordinary Markov property.

2.6.3 Branching Process

We consider the branching process (Galton-Watson process) as a simple Markovian model for a growing population.
Let $(\xi_i^n)_{n,i\in\mathbb{N}_1}$ be an array of iid., $\mathbb{N} = \mathbb{N}_0$-valued random variables.
Define the process $(Z_n)_{n\in\mathbb{N}}$ by $Z_0 := 1$ and

$$Z_{n+1} := \sum_{i=1}^{Z_n} \xi_i^{n+1},$$

Then (Z_n) is a Markov chain, taking values in $E = \mathbb{N}$, with respect to filtration

$$\mathcal{F}_n := \sigma(\xi_i^k : k \leqslant n, i \in \mathbb{N}),$$

with transition-matrix (p_{ij}) and -kernel K given by

$$\mathbb{P}[Z_{n+1} = j | Z_n = i]\& = \mathbb{P}[\xi_1^{n+1} + \ldots + \xi_i^{n+1} \ = \ j] =: p_{ij} =: K(i, \{j\});$$

indeed we have

$$\mathbb{P}[Z_{n+1} = j | \mathcal{F}_n] = \mathbb{P}\left[\sum_{i=1}^{Z_n} \xi_i^{n+1} \middle| \mathcal{F}_n\right] = \mathbb{P}[Z_{n+1} = j | Z_n], \quad j \in E,$$

implying (by monotone convergence)

$$P[Z_{n+1} \in B | \mathcal{F}_n] = \mathbb{P}[Z_{n+1} \in B | Z_n], \qquad B \subset E.$$

Let $m := E[\xi_i^n]$ and $\sigma^2 := \mathrm{Var}(\xi_i^n)$. Assume $m \in (0, \infty)$. (Case $m = 0$ is trivial .)

Theorem 2.44. *Assume $\sigma^2 \in (0, \infty)$ and let $W_n := Z_n/m^n$.*
Then (W_n) is a martingale,the almost-sure limit $\lim_n W_n =: W_\infty$ exists, and

$$m > 1 \Leftrightarrow \mathbb{E}[W_\infty] = 1 \Leftrightarrow \mathbb{E}[W_\infty] > 0\,.$$

Chapter 3

Construction of stochastic processes

3.1 General construction of stochastic processes

The aim of this section is to explain general concepts and construction of mathematical model (probability space) carrying general stochastic processes.

Let $(S, \mathcal{S})$ be a measurable space, S is the space of observations, $(X_i)_{i\in I}$ is some family of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $(S, \mathcal{S})$, with some general index set I.

Definition 3.1. (and Lemma) Let the index set I be non-empty, $(S, \mathcal{S})$ be a measurable space, $(X_i)_{i\in I}$ be a stochastic process (family of random variables valued in $(S, \mathcal{S})$). For $J \subset I$, J finite, the mapping $(X_i)_{i\in J} : \Omega \to S^J$ induces a measure $\mathbb{Q}^{(J)}$ on $\mathcal{S}^J := \bigotimes_{i\in J} \mathcal{S}$ by

$$\mathbb{Q}^{(J)} = \mathbb{P} \circ ((X_i)_{i\in J})^{-1},$$

being the "image measure" (distribution) of $(X_i)_{i\in J}$. The family $\{\mathbb{Q}^{(J)} |\ J \subset I,\ J \text{ finite}\}$ is called *the family of finite-dimensional marginal distributions* of $(X_i)_{i\in I}$ under $\mathbb{P}$. This family is *consistent* in the sense that for any finite $J' \subset J$ we have

$$\mathbb{Q}^{(J')} = \mathbb{Q}^{(J)} \circ (\pi^J_{J'})^{-1},$$

where $\pi^J_{J'} : S^J \to S^{J'}$ are the *canonical projections*: $\pi^J_{J'}((s_i)_{i\in J}) = (s_i)_{i\in J'}$.

On $S^I = \times_{i\in I} S$ we define the coordinate mappings $\pi_j = \pi^I_{\{j\}} : \omega = (s_i)_{i\in I} \mapsto s_j$. $\mathcal{S}^I = \bigotimes_{i\in I} \mathcal{S}$ is the *product-σ-field* given by $\sigma(\pi_i :\ i \in I)$, i.e. the smallest σ-field on S^I so that all π_i are measurable. An intersection-stable generator of $\mathcal{S}^I$ is given by the system $\mathcal{Z}^R$ of all *cylinder sets* of the form $Z^R = \{\omega \in S^I |\ \omega_i \in A_i \text{ for } i \in J\}$ with $A_i \in \mathcal{S}$ and $J \subset I$, J – finite. A process $(X_i)_{i\in I}$ induces also a distribution on $\mathcal{S}^I$ as the image measure of $\mathbb{P}$ under the mapping $(X_i)_{i\in I} : \Omega \to S^I$. Studying the probabilistic properties of $(X_i)_{i\in I}$ requires to know this *distribution* $\mathbb{Q} = \mathbb{P} \circ ((X_i)_{i\in I})^{-1}$ *of the process* X.

Remark 3.1.

- More generally, we can also consider a family of different measurable spaces $(S_i, \mathcal{S}_i)$, $i \in I$; and in this case $(S^I, \mathcal{S}^I) := (\times_{i\in I} S_i, \bigotimes_{i\in I} \mathcal{S}_i)$ where $\bigotimes_{i\in I} \mathcal{S}_i = \sigma(\pi_i :\ i \in I)$.
- $\mathcal{S}^I = \bigotimes_{i\in I} \mathcal{S}_i$ in general is not the same as $\sigma(\{\times_{i\in I} A_i |\ A_i \in \mathcal{S}_i\})$ if I is uncountable. If I is countable, then the two σ-fields coincide (exercise).

Question: How to construct a model $(\Omega, \mathcal{F}, \mathbb{P})$ and $(X_i)_{i\in I}$ such that X has a given "stochastic structure"?

Canonical model: Choose $\Omega = S^I$, $\mathcal{F} = \mathcal{S}^I$, and $X_i = \pi_i$ as canonical coordinate projections. Hence, the question reduces to the construction of a measure $\mathbb{P}$ on $(S^I, \mathcal{S}^I)$ with some "structure" (on canonical probability space, $\mathbb{Q} = \mathbb{P}$ in the above notation).

There are alternative conceptual ways to impose "stochastic structure" in general:

(1) specify a consistent family of finite-dimensional distributions;

(2) specify some initial distribution for one (first) coordinate and a "transition rule" on how to sample further coordinates from given ones.

Lemma 3.2. *A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}) = (S^I, \mathcal{S}^I)$ is uniquely determined by its finite-dimensional marginal distributions, i.e. for any consistent family of finite-dimensional distributions $\{\mathbb{Q}^{(J)} |\ J \subset I,\ J\text{- finite}\}$, there exists at most one $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that $\mathbb{Q}^{(J)} = \mathbb{P} \circ (\pi^I_J)^{-1}$ for every finite $J \subset I$.*

Let $(S, \mathcal{S})$ be a measurable space, $\mathbb{P}_0$ be a probability measure on $(S, \mathcal{S})$, $I = \mathbb{N} := \{0, 1, \ldots\}$. For any $n \geqslant 1$, let K_n be a stochastic kernel from $(S^n, \mathcal{S}^n)$ to $(S, \mathcal{S})$. Consider the following iterative construction of measures (as semidirect products of measures and kernels):

$$\mathbb{P}^{(0)} = \mathbb{P}_0, \quad \mathbb{P}^{(1)} = \mathbb{P}_0 \otimes K_1, \quad \ldots \quad \mathbb{P}^{(n)} = \mathbb{P}^{(n-1)} \otimes K_n, \quad \ldots$$

For every $f \geqslant 0$, $\mathcal{S}^{n+1}$-measurable function, we have

$$\int_{S^{n+1}} f \,\mathrm{d}\mathbb{P}^{(n)} = \int_S \mathbb{P}_0(\mathrm{d}x_0) \int_S K_1(x_0, \mathrm{d}x_1) \cdots \int_S K_n((x_0, \ldots, x_{n-1}), \mathrm{d}x_n) \cdot f(x_0, \ldots, x_n).$$

Let $\Omega = S^{\mathbb{N}} = \{\omega = (x_0, x_1, \ldots) |\ x_i \in S,\ i \in \mathbb{N}\}$; this is the space of all trajectories. Let $X_n(\omega) = \pi_n(\omega) = x_n$ for $n \in \mathbb{N}$; this process describes the state at time n. Set $\mathcal{F}_n := \sigma(X_0, \ldots, X_n)$ for all $n \in \mathbb{N}$, describing the information up to time n, and let $\mathcal{F} := \sigma(X_0, X_1, \ldots) = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right) = \mathcal{S}^{\mathbb{N}}$. Note that $\mathcal{F}_n$ is a σ-field on $S^{\mathbb{N}}$ and each $A \in \mathcal{F}_n$ has the form $A_n \times S \times S \times \cdots$ for some $A_n \in \mathcal{S}^{n+1}$.

Remark 3.2. Consistency of $P^{(0)}$, $P^{(1)}$, … : For $A_n \in \mathcal{F}_n$, we have $\underbrace{(A_n \times S)}_{=: A_{n+1} \in \mathcal{S}^{n+2}} \times S \times \cdots = A_{n+1} \times S \times \cdots$, and $\mathbb{P}^{(n+1)}[A_n \times S] = \mathbb{P}^{(n)} \otimes K_{n+1}[A_n \times S] = \mathbb{P}^{(n)}[A_n]$. Similarly, $\mathbb{P}^{(m)}[A_n \times \underbrace{S \times \cdots \times S}_{m-n}] = \mathbb{P}^{(n)}[A_n]$ for every $m > n$ and $A_n \in \mathcal{F}_n$.

Theorem 3.3 (Ionescu Tulcea). *In the discrete time situation $I = \mathbb{N}$, there exists a unique probability measure $\mathbb{P}$ on $\mathcal{S}^{\mathbb{N}}$ satisfying*

$$\mathbb{P}[A_n \times S \times \cdots] = \mathbb{P}^{(n)}[A_n] \quad \forall A_n \in \mathcal{S}^{n+1}, \tag{3.1.1}$$

or equivalently $\forall n$ and $f \geqslant 0$, $\mathcal{S}^{n+1}$-measurable,

$$\int_\Omega f(X_0, \ldots, X_n) \,\mathrm{d}\mathbb{P} = \int_S \mathbb{P}_0(\mathrm{d}x_0) \int_S K_1(x_0, \mathrm{d}x_1) \cdots \int_S K_n((x_0, \ldots, x_{n-1}), \mathrm{d}x_n) \cdot f(x_0, \ldots, x_n). \tag{3.1.2}$$

Notation and example: If in Theorem 3.3 we have that $K_n((x_0, \ldots, x_{n-1}), \mathrm{d}x_n) = \mathbb{P}_n(\mathrm{d}x_n)$ for some probability measure $\mathbb{P}_n$ on $\mathcal{S}$ for all n, i.e. does not depend on $x_0, \ldots, x_{n-1}$, then the unique measure $\mathbb{P}$ (from Theorem 3.3) is called the product measure of the $\mathbb{P}_n$'s, and we write

$$\mathbb{P} = \bigotimes_{n=0}^{\infty} \mathbb{P}_n.$$

In the general case we write

$$\mathbb{P} = \mathbb{P}_0 \otimes \left(\bigotimes_{n=1}^{\infty} K_n \right)$$

using notation of semidirect products.

For $\mathbb{P} = \bigotimes_{n=0}^{\infty} \mathbb{P}_n$ we have that $\mathbb{P}_n$ is the marginal distribution of the n-th coordinate and $\mathbb{P}^{(n)} = \bigotimes_{k=0}^{n} \mathbb{P}_k$ for all $n \geqslant 0$, i.e. in particular the coordinate mappings are independent.

Corollary 3.4. *The probability measure $\mathbb{P}$ from Theorem 3.3 is the product of its one-dimensional marginal distributions $\mathbb{P}_n := \mathbb{P} \circ \pi_n^{-1} \equiv \mathbb{P} \circ X_n^{-1}$ if and only if the random variables $(X_n)_{n \geqslant 0}$ are independent under $\mathbb{P}$.*

Example 3.1. If $K_n((x_0, \ldots, x_{n-1}), \cdot) = K_n(x_{n-1}, \cdot)$ does not depend on $x_0, x_1, \ldots, x_{n-2}$ but only on x_{n-1} for all n, then the coordinate process $(X_n)_{n \geqslant 0}$ is a (time-inhomogeneous) discrete-time Markov process. If $K_n = K$ (does not depend on n), then $(X_n)_{n \geqslant 0}$ is a time-homogeneous Markov process.

Next goal is to construct stochastic processes on arbitrary (possibly uncountable) index set I with state space $(S, \mathcal{S})$. Let $\Omega = S^I$, $\mathcal{F} = \mathcal{S}^I$, and $X_i : \Omega \to S$ for $i \in I$ - the coordinate mappings.

Theorem 3.5 (Kolmogorov's consistency theorem)**.** *Let $\{\mathbb{Q}^{(J)} |\ J \subset I,\ J \text{ - finite}\}$ be a consistent family of finite-dimensional distributions $\mathbb{Q}^{(J)}$ on $(S^J, \mathcal{S}^J)$ and let S be a Polish space with Borel σ-field $\mathcal{S} = \mathcal{B}(S)$. Then there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}) = (S^I, \mathcal{S}^I)$ which has the $Q^{(J)}$'s as its finite-dimensional marginal distributions.*

Lemma 3.6. *Let I be an arbitrary index set and $\mathcal{S}^I \equiv \bigotimes_{i \in I} \mathcal{S} \equiv \sigma(X_i :\ i \in I)$ be the product σ-field. Then we have*

(a)

$$\mathcal{S}^I = \bigcup_{\substack{J \subset I \\ J \text{ - countable}}} \sigma(X_j :\ j \in J).$$

(b) $\mathcal{S}^I = \sigma\left(\left\{ (\times_{j \in J} A_j) \times (\times_{i \in I \setminus J} S) \;\middle|\; A_j \in \mathcal{S}\ \forall j \in J,\ J \subset I,\ J \text{ - countable} \right\} \right).$

3.2 Gaussian Processes

Let I be an arbitrary index set, $(X_i)_{i \in I}$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 3.7. (a) $X = (X_i)_{i \in I}$ is called *Gaussian process* if all finite dimensional marginal distributions are multivariate normal ("Gaussian"), i.e. $\forall n \in N$, $i_1, \ldots, i_n \in I$, $a \in \mathbb{R}^n$ holds that $\sum_1^n a_k X_{i_k}$ is a univariate normal random variable (possibly degenerate, i.e. with zero variance).

(b) X is called *centered Gaussian process* if it is a Gaussian process with $\mathbb{E}[X_i] = 0$ for every $i \in I$.

(c) A family $(X_i^k)_{i \in I_k}$, $k \in K$, of stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a Gaussian process if $\{X_i^k |\ k \in K,\ i \in I_k\} = (X_i^k)_{(k,i) \in \tilde{I}}$ with $\tilde{I} := \{(k,i) |\ k \in K,\ i \in I_k\}$ is a Gaussian process.

Theorem 3.8. *(a) Let $X = (X_i)_{i \in I}$ be a Gaussian process. Then its distribution is uniquely determined by $\mathbb{E}[X_i]$, $i \in I$, and $Cov(X_i, X_j)$, $i, j \in I$.*

(b) The process $X = (X_i)_{i \in I}$ is a Gaussian process with mean values $\mathbb{E}[X_i]$, $i \in I$, if and only if the process $\widetilde{X} = (X_i - \mathbb{E}[X_i])_{i \in I}$ is a centered Gaussian process (with the covariance function for both processes being the same).

(c) ***Existence:*** *Let $\Gamma : I \times I \to \mathbb{R}$ be such that for every finite J, $J \subset I$, the matrix $(\Gamma(i,j))_{i,j \in J}$ is symmetric and positive semi-definite. Then there exists a centered Gaussian process $(X_i)_{i \in I}$ on some probability space with $Cov(X_i, X_j) = \Gamma(i,j)$ for all $i, j \in I$.*

Example 3.2. Consider the space $L^2(\mathbb{R}_+, \mathrm{d}x)$ with the scalar product $\langle f, g \rangle = \int_{\mathbb{R}_+} fg \, \mathrm{d}x$. This space is a separable Hilbert space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space carrying a sequence $(\varepsilon_i)_{i \in \mathbb{N}}$ of i.i.d. $\mathcal{N}(0,1)$ random variables. The subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ which is spanned by $(\varepsilon_i)_{i \in \mathbb{N}}$ is also a separable Hilbert space with the scalar product $\langle X, Y \rangle = \mathbb{E}[XY] = \int_\Omega XY \, \mathrm{d}\mathbb{P}$.

Now, an isometry between these two Hilbert spaces is given, after choosing an ONB $(b_i)_{i \in \mathbb{N}}$ of $L^2(\mathbb{R}_+, \mathrm{d}x)$, by $b_i \leftrightarrow \varepsilon_i$, $i \in \mathbb{N}$, inducing

$$L^2(\mathbb{R}_+, \mathrm{d}x) \ni h = \sum \alpha_i b_i \quad \longleftrightarrow \quad \sum \alpha_i \varepsilon_i =: X_h \in L^2(\Omega, \mathcal{F}, \mathbb{P}),$$

where $\alpha_i = \langle h, b_i \rangle$ for $i \in \mathbb{N}$.

Since ε_i are Gaussian, then one can prove that $(X_h)_{h \in L^2(\mathbb{R}_+, \mathrm{d}x)}$ is a centered Gaussian process with covariance $\mathrm{Cov}(X_h, X_g) = \langle h, g \rangle$

Take $h = \mathbb{1}_{[0,t]}$ and set $W_t := X_{\mathbb{1}_{[0,t]}}$ for every $t \in \mathbb{R}_+$. The process $(W_t)_{t \in \mathbb{R}_+}$ is a centered Gaussian process, since $\{\mathbb{1}_{[0,t]} |\ t \in \mathbb{R}_+\} \subset L^2(\mathbb{R}_+, \mathrm{d}x)$, and its covariance function is given by

$$\begin{aligned} \mathrm{Cov}(W_s, W_t) &= \mathrm{Cov}(X_{\mathbb{1}_{[0,s]}}, X_{\mathbb{1}_{[0,t]}}) \\ &= \langle \mathbb{1}_{[0,s]}, \mathbb{1}_{[0,t]} \rangle && \text{(isometry)} \\ &= \int_0^\infty \mathbb{1}_{[0,s]} \cdot \mathbb{1}_{[0,t]} \, \mathrm{d}x \\ &= s \wedge t. \end{aligned}$$

This will be the covariance structure of Brownian motion, also known as Wiener process (after Norbert Wiener, 1894-1964).

Example 3.3. In this example, we will construct a Gaussian process with the same covariance structure as the one from the previous example using a different approach.

Let $\mathbb{P}_0 = \delta_0$ and

$$K_{s,t}(x, \mathrm{d}y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{|y-x|^2}{2|t-s|}\right) \mathrm{d}y, \qquad t > s \geqslant 0.$$

For every finite index set $J = \{t_1, \ldots, t_n\}$, $n \in \mathbb{N}$, define the finite-dimensional distributions

$$\mathbb{Q}^{(J)}(\bigtimes_1^n A_i) = \int_{\mathbb{R}} \mathbb{P}_0(\mathrm{d}x_0) \int_{A_{t_1}} K_{0,t_1}(x_0, \mathrm{d}x_{t_1}) \cdots \int_{A_{t_n}} K_{t_{n-1},t_n}(x_{t_{n-1}}, \mathrm{d}x_{t_n}) \quad \forall A_{t_i} \in \mathcal{B}(\mathbb{R}).$$

This family of measures is consistent. Indeed, $\delta_x \otimes K_{s,t}(x, \mathrm{d}y) \sim x + \mathcal{N}(0, s-t)$ and $\mathcal{N}(0, t-s) + \mathcal{N}(0, u-t) = \mathcal{N}(0, u-s)$ for all $u > t > s \geqslant 0$. Hence, by the Kolmogorov's consistency theorem (Theorem 3.5) there exists a measure $\mathbb{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R})^{[0,\infty)})$ such that the canonical process $X = (X_t)_{t \geqslant 0}$ has $\{Q^{(J)} | \ J \text{ - finite}\}$ as finite-dimensional marginal distributions that are multivariate normal distributions. Therefore, X is a Gaussian process under $\mathbb{P}$, centered and with covariances: for $s < t$

$$\begin{aligned}
\operatorname{Cov}(X_s, X_t) &= \mathbb{E}[X_s X_t] \\
&= \mathbb{E}[X_s(X_s + \underbrace{X_t - X_s}_{\perp X_s})] \\
&= \mathbb{E}[X_s^2] = s \\
&= s \wedge t.
\end{aligned}$$

Remark 3.3. Note that (X_t) and (W_t) (from the previous example) have the same distribution on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R})^{[0,\infty)})$. The process X starts at 0, $X_t - X_s$ is $\mathcal{N}(0, t-s)$-distributed, and $X_t - X_s$ is independent of the past of the process until time s for $t > s \geqslant 0$. This Gaussian process has the properties which define Brownian motion except the continuity of paths, it is sometimes called pre-Brownian motion.

3.3 Brownian motion

Definition 3.9. Processes (X_t), (Y_t), $t \in I$, given on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, are called

(a) *versions* or *modifications* of each other if

$$\forall t \in I \ \exists N_t \in \mathcal{F}: \quad \mathbb{P}[N_t] = 0 \quad \& \quad X_t(\omega) = Y_t(\omega) \quad \forall \omega \notin N_t,$$

i.e.

$$\mathbb{P}[X_t = Y_t] = 1 \qquad \forall t \in I;$$

(b) *indistinguishable* if there exists a null set N such that

$$X_t(\omega) = Y_t(\omega) \qquad \forall t \in I, \ \forall \omega \notin N,$$

i.e.

$$\mathbb{P}[X_t = Y_t \quad \text{for all } t \in I] = 1.$$

Theorem 3.10 (Continuity criterion from Kolmogorov and Chentsov)**.** *Let* $(X_t)_{t\in[0,1]}$ *be an* $\mathbb{R}^d$*-valued process, with constants* $\alpha, \beta,\ C > 0$ *such that*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leqslant C \cdot |t-s|^{1+\beta} \qquad \forall s,t \in [0,1].$$

Then there exists a version $(Y_t)_{t\in[0,1]}$ *of* $(X_t)_{t\in[0,1]}$ *with (globally) Hölder-continuous paths of order* $\lambda \in \left[0, \frac{\beta}{\alpha}\right)$*.*

Remark 3.4.

- A function $g : [0,1] \to \mathbb{R}^d$ is (globally) Hölder-continuous of order λ if

$$\sup_{\substack{s,t\in[0,1]\\ s\neq t}} \frac{\|g(s)-g(t)\|}{|s-t|^\lambda} < \infty.$$

- If $X = (X_t)_{t\in[0,1]}$ is a process such that $X_t - X_s \sim \mathcal{N}(0, t-s)$ for every $t, s \in [0,1]$ with $t > s$, then $\mathbb{E}[|X_t - X_s|^4] = 3|t-s|^2$ ($\mathbb{E}[Z^4] = 3\sigma^4$ for $Z \sim \mathcal{N}(0,\sigma^2)$). Hence, (X_t) has a Hölder-continuous version with $\alpha = 4$, $\beta = 1$ and $c = 3$ in the previous theorem.

Definition 3.11 (Brownian motion)**.** An $\mathbb{R}^d$-valued stochastic process $(B_t)_{t\in\mathbb{R}_+}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called

(1) $[d = 1]$ a (1-dimensional) Brownian motion if

   (a) $B_0 = 0$,

   (b) $\forall n \in \mathbb{N}$, $0 \leqslant t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, we have that $\Delta B_{t_n} := B_{t_n} - B_{t_{n-1}}$ are $\mathcal{N}(0, t_n - t_{n-1})$–distributed and independent (increments over disjoint intervals are independent),

   (c) $\mathbb{P}$-a.a. paths are continuous;

(2) $[d \geqslant 1]$ a (d-dimensional) Brownian motion if all coordinate processes B^j, $j = 1, \ldots, d$, of $B = (B^j)_{j=1,\ldots,d}$ are independent 1-dimensional Brownian motions;

(3) a Brownian motion with respect to a given filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ on $\mathcal{F}$ if it is a d-dimensional Brownian motion with increments $B_t - B_s$ independent of $\mathcal{F}_s$ for all $t \geqslant s \geqslant 0$.

Theorem 3.12 (Existence and Hölder-continuity of Brownian motion)**.** *Let* $X = (X_t)_{t\in\mathbb{R}_+}$ *be a centered real-valued Gaussian process with covariance function* $Cov(X_s, X_t) = s \wedge t$ *(which exists, by previous results). Then there exists a version* B *of* X *such that* B *has continuous paths that are Hölder-continuous of any order* $\lambda < 1/2$ *on any compact interval* $[0,n] \subset \mathbb{R}_+$*. In particular, a Brownian motion (d-dimensional) does exist.*

Remark 3.5. Let $(X_t)_{t\in\mathbb{R}_+}$ and $(Y_t)_{t\in\mathbb{R}_+}$ be right-continuous processes. If X is a version of Y, then X and Y are indistinguishable. The reason is that a right-continuous function $\mathbb{R}_+ \mapsto \mathbb{R}^d$ is determined by its values on a dense countable subset on $\mathbb{R}_+$.

Definition 3.13. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) *The natural filtration* $(\mathcal{F}_t^\circ)_{t\geqslant 0}$ generated by X is the filtration $\mathcal{F}_t^\circ = \sigma(X_s : \ s \leqslant t)$, i.e. the smallest filtration with respect to which X is adapted.

(b)
- A filtration $(\mathcal{F}_t)_{t\geqslant 0}$ is *complete* (w.r.t the measure $\mathbb{P}$) if each $\mathcal{F}_t$ is complete w.r.t $\mathbb{P}$ (i.e. contains all subsets of $\mathbb{P}$-null sets).
- A filtration $(\mathcal{F}_t)_{t\geqslant 0}$ is called *right-continuous* if

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} \qquad \forall t \in \mathbb{R}_+.$$

- A filtration $(\mathcal{F}_t)_{t\geqslant 0}$ *satisfies the usual conditions* if it is complete (w.r.t $\mathbb{P}$) and right-continuous.

(c) The *usual filtration* $(\mathcal{F}_t^X)$ generated by a process X is

$$\mathcal{F}_t^X := \bigcap_{\varepsilon>0} \left(\mathcal{F}_{t+\varepsilon}^\circ\right)^{\mathbb{P}} \qquad \forall t \in \mathbb{R}_+,$$

where $(\cdot)^{\mathbb{P}}$ means completion with respect to $\mathbb{P}$.

Lemma 3.14. *If a filtration $(\mathcal{F}_t)$ is right-continuous, then its completion $(\mathcal{F}_t^{\mathbb{P}})$ is also right-continuous, hence it satisfies the usual conditions.*

Theorem 3.15. *Let $(\mathcal{F}_t)_{t\geqslant 0}$ denote the completion of the natural filtration of the Brownian motion $(B_t)_{t\geqslant 0}$. Then $(\mathcal{F}_t)_{t\geqslant 0}$ is right-continuous, i.e. it satisfies the usual conditions.*

Remark 3.6. A Brownian motion is an $(\mathcal{F}_t)$-Brownian motion with respect to its usual filtration $=: (\mathcal{F}_t)$.

Corollary 3.16. *For $(\mathcal{F}_t)_{t\geqslant 0}$ from Theorem 3.15, $\mathcal{F}_0 = (\mathcal{F}_0^\circ)^{\mathbb{P}}$ is trivial, i.e. for every $A \in \mathcal{F}_0$ we have $\mathbb{P}[A] \in \{0, 1\}$.*

Example 3.4. If B is a Brownian motion, then

$$A = \{\exists \varepsilon > 0 \ B_t \neq 0 \text{ in } (0, \varepsilon]\} \in \mathcal{F}_{0+}^B$$

and hence $\mathbb{P}[A] \in \{0, 1\}$. However, $\mathbb{P}[A] \neq 1$ (e.g. $\mathbb{P}[B_{\varepsilon/3} > 0, B_{2\varepsilon/3} < 0] > 0$).

Similarly, $\mathbb{P}[C] = 0$ for $C := \{\forall \varepsilon > 0 \ \exists t^+, t^- \in (0, \varepsilon) \ : B_{t^+} > 0 \text{ and } B_{t^-} > 0\}$

Chapter 4

Weak convergence and tightness of probability measures

4.1 Weak convergence

Setting: (S,d) is a metric space with $\mathcal{S} = \mathcal{B}(S)$ (the Borel σ-field), μ and μ_n, $n \in \mathbb{N}$, are probability measures on $(S,\mathcal{S})$.

Remark 4.1. We know from the course Stochastics I that convergence notions like "$\mu_n(A) \to \mu(A)\ \forall A \in \mathcal{S}$" or "$|\mu_n - \mu(A)| \equiv \sup_{A\in\mathcal{S}} |\mu_n(A) - \mu| \to 0$" would be too restrictive, e.g. for the central limit theorem.

Definition 4.1. (a) $(\mu_n)_{n\in\mathbb{N}}$ *converges weakly* to μ, write $\mu_n \Rightarrow \mu$, if

$$\int_S h \,\mathrm{d}\mu_n \to \int_S h \,\mathrm{d}\mu \qquad \forall h \in C_b(S),$$

where $C_b(S)$ denotes the space of all bounded continuous functions on S.

(b) A sequence of $(S,\mathcal{S})$-valued random variables X_n defined on some probability spaces $(\Omega_n,\mathcal{F}_n,\mathbb{P}_n)$ *converges in distribution* to an $(S,\mathcal{S})$-valued random variable X defined on $(\Omega,\mathcal{F},\mathbb{P})$, and write $X_n \xrightarrow{\mathcal{D}} X$, if $\mu_n := \mathbb{P}_n \circ X_n^{-1} \Rightarrow \mu := \mathbb{P} \circ X^{-1}$.

Remark 4.2.
- It is essential in (b) that X_n, X, are $(S,\mathcal{S})$-valued (the same space).
- Notations in the literature for this type of convergence: $X_n \xrightarrow{\mathcal{D}} X$, $X_n \xrightarrow{\mathcal{L}} X$ (convergence in law), $\mathcal{L}(X_n|\mathbb{P}_n) \to \mathcal{L}(X|\mathbb{P})$, $\mu_n \xrightarrow{\mathrm{w}} \mu$.

Theorem 4.2 (Portemanteau theorem)**.** *The following are equivalent*

(1) $\mu_n \Rightarrow \mu$ *as* $n \to \infty$.

(2) $\int h \,\mathrm{d}\mu_n \to \int h \,\mathrm{d}\mu$ *for every* $h \in C_b(S)$ *that is uniformly continuous.*

(3) $\limsup_n \mu_n(F) \leqslant \mu(F)$ *for every closed set* $F \subset S$.

(4) $\liminf_n \mu_n(G) \geqslant \mu(G)$ *for every open set* $G \subset S$.

(5) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for every* μ*-boundary-less* $A \in \mathcal{S}$*, i.e.* $A \in \mathcal{S}$ *with* $\mu(\overline{A} \backslash A^\circ) = 0$.

Sometimes it is very useful that it suffices to have $\mu_n(A) \to \mu(A)$ for only a subfamily of sets A.

Corollary 4.3. *Let* $\mathcal{E} \subset \mathcal{S}$ *be* $\cap$*-stable such that*

(1) any open $G \subset S$ *can be represented as a countable union of sets of* $\mathcal{E}$*,*

(2) $\mu_n(A) \to \mu(A)$ *as* $n \to \infty$ *for every* $A \in \mathcal{E}$*.*

Then $\mu_n \Rightarrow \mu$*.*

Remark 4.3. For separable (S, d) one can construct a countable set $\mathcal{E}$ as in Corollary 4.3: take the balls of rational radii centered at a countable dense subset of S, and consider all finite intersection of such balls.

4.2 Tightness and Prohorov's theorem

Let (S, d) be a metric space and $\mathcal{S} = \mathcal{B}(S)$ be the Borel σ-field. Consider be the set of all probability distributions on $(S, \mathcal{S})$ denoted by $\mathcal{M}_1(S)$. By "$\mu_n \Rightarrow \mu$" we will denote weak convergence.

Remark 4.4.

- The topology of weak convergence is described by the neighborhood-basis (which is also a basis of the topology)

$$U_{\varepsilon,h_1,\dots,h_n}(\mu) := \left\{ \nu \in \mathcal{M}_1(S) \;\middle|\; \left|\int h_i \,\mathrm{d}\nu - \int h_i \,\mathrm{d}\mu\right| < \varepsilon, i = 1, \dots, n \right\}, \quad n \in \mathbb{N},\ \varepsilon > 0,\ h_i \in C_b(S$$

  Open sets of $\mathcal{M}_1(S)$ are sets which contain for any element a neighborhood of this form. Note that basis sets above are itself open, and any open set is an (arbitrary) union of such basis sets.

- Recall (cf. [Alt02, Chapter 2.5]): A set M in a topological space is
  - M is compact :$\Leftrightarrow$ any open cover has a finite subcover.
  - M is sequentially (relatively) compact :$\Leftrightarrow$ any sequence in M ($\overline{M}$) has a subsequence converging in M ($\overline{M}$). Note that compact is not necessarily cvalent to sequentially compact.
  - M is precompact (totally bounded) in a metric space :$\Leftrightarrow$ It has, for any $\varepsilon > 0$, a finite covering with ε-balls.
  - for M set in a metric space,

$$\begin{aligned} \text{compact} &\Leftrightarrow \text{sequentially compact} \\ &\Leftrightarrow \text{precompact and complete} \\ &\Rightarrow M \text{ is separable} \end{aligned}$$

  - A subset M of a complete metric space is precompact iff it is relatively compact.

- A natural question: is the topology of weak convergence in $\mathcal{M}_1(S)$ metrizable? Answer: yes, if S is separable ("Prohorov's metric").

Example 4.1. Consider $S = \mathbb{R}$ with the usual metric, (x_n) a sequence in $\mathbb{R}$ and $\mu_n := \delta_{x_n}$. If $\lim_{n\to\infty} |x_n| = \infty$, there exists no convergence subsequence in general. Consider for example $x_n = n$. Then $\int h \, d\mu_n = h(n)$ and this does not necessarily converge, i.e. μ_n does not converge weakly to some measure. However, if (x_n) is bounded, then there exists a subsequence (x_{n_k}) converging to some $x \in \mathbb{R}$. In this case, it is easy to see that $\mu_{n_k} \Rightarrow \delta_x$.

Definition 4.4. A set $M \subset \mathcal{M}_1(S)$ is called *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset S$ such that $\mu(K) \geqslant 1 - \varepsilon$ for every $\mu \in M$.

Example 4.2. The following is a sufficient condition for tightness: there exists a function $h : S \to \mathbb{R}$ such that $\{|h| \leqslant c\}$ is compact for every $c \geqslant 0$ and

$$\sup_{\mu \in M} \int h \, d\mu < \infty.$$

For instance, $S = \mathbb{R}$, $h(x) = |x|^p$ for some $p > 0$. Then the set of measures with bounded p-moments will form a tight set.

Theorem 4.5 (Prohorov)**.** *Let $\mathcal{M} \subset \mathcal{M}_1(S)$.*

(1) If $\mathcal{M}$ is tight, then $\mathcal{M}$ is relatively sequentially compact.

(2) Let S be complete and separable. If $\mathcal{M}$ is relatively sequentially compact, then $\mathcal{M}$ is tight.

Example 4.3. Let (μ_n) be a tight sequence of probability measures. Assume that any weakly convergent subsequence has weak limit μ. Then the sequence converges: $\mu_n \Rightarrow \mu$.

Proposition 4.6. *If S is compact, then $\mathcal{M}_1(S)$ is compact and sequentially compact.*

For probability measures μ and μ_n, $n \in \mathbb{N}$, on $(\mathbb{R}^d, \mathcal{B}^d)$, let φ_{μ_n}, φ_μ denote the characteristic functions of μ_n, μ; and let X_n, X denote random variables with distributions $\mu_n = \mathbb{P}_n \circ X_n^{-1}$, $\mu = \mathbb{P} \circ X^{-1}$, respectively. We recall Levy's continuity theorem for characteristic functions. It states that the following are equivalent

(1) There exists a probability measure μ on $\mathcal{B}^d$ such that $\mu_n \Rightarrow \mu$.

(2) There exists a function $\varphi : \mathbb{R}^d \to \mathbb{C}$ such that $\varphi_{\mu_n}(u) \to \varphi(u)$ for all $u \in \mathbb{R}^d$ and φ is (partially) continuous at 0 (in which case $\varphi = \varphi_\mu$).

The proof for $d = 1$ is usually shown in the lecture Stochastics I. The proofs for the case $d \geqslant 1$ for Levy's continuity theorem, and likewise the proof for the Cramer-Wold theorem (cf.below), make use of Prohorov's Theorem 4.5 (cf. [Klenke06, Thms 15.23 and 15.55], elaborate the details) (for part (b), by using part (a) and the already known result for dimension $d = 1$). :

Proposition 4.7. *With the previous notations, we have* *(a) (Cramér-Wold theorem): Then*

$$\mu_n \Rightarrow \mu \quad \iff \quad \langle y, X_n \rangle \xrightarrow{\mathcal{D}} \langle y, X \rangle \quad \forall y \in \mathbb{R}^d.$$

(b) Levy's continuity theorem (cf.above) on characteristic functions holds in multiple dimensions on $\mathbb{R}^d$, $d \in \mathbb{N}$.

Remark 4.5 (On $\mathcal{M}_1(S)$ being metrizable and the Prohorov metric)**.** (1)

$$\rho(\mu, \nu) := \inf\{\varepsilon > 0 | \ \mu(A) \leqslant \varepsilon + \nu(U_\varepsilon(A)), \ \nu(A) \leqslant \varepsilon + \mu(U_\varepsilon(A)) \quad \forall A \in \mathcal{S}\}$$

defines a metric on $\mathcal{M}_1(S)$ called *the Prohorov metric.*

(2) $\rho(\mu_n, \mu) \to 0$ implies $\mu_n \Rightarrow \mu$.

(3) If S is separable, then the converse is true: $\mu_n \Rightarrow \mu$ implies that $\rho(\mu_n, \mu) \to 0$. Hence, $\mathcal{M}_1(S)$ with the topology of weak convergence is metrizable by the Prohorov metric.

For reference, cf. [Klenke06].

4.3 Weak convergence on $S = C[0, 1]$

We consider the space $S = C[0, 1]$ of continuous real-valued functions on $[0, 1]$ with the sup-norm $\|x\| = \sup_{t \in [0,1]} |x(t)|$ (inducing the metric $d(x, y) = \|x - y\|$). With this topology, S is a separable Banach space .

Lemma 4.8. *(a) The Borel σ-field $\mathcal{S} = \mathcal{B}(S)$ is generated by the cylinder sets of the form*

$$Z = \{x \in S | \ x(t_i) \in A_i, \ i = 1, \ldots, n\}$$

with $0 \leqslant t_1 \leqslant \ldots \leqslant t_n \leqslant 1$, $A_i \in \mathcal{B}(R)$ for $i = 1, \ldots, n$, $n \in \mathbb{N}$, i.e. $\sigma(\{Z \text{ sets}\}) = \mathcal{B}(S) = \mathcal{S}$.

(b) $\mathcal{B}(S) = \sigma(\pi_t | \ t \in [0, 1])$ with $\pi_t : x \mapsto x(t)$ being the coordinate projections.

(c) The set $\mathcal{Z}$ of all cylinder sets is $\cap$-stable.

Any measure μ on $C[0, 1]$ is determined by its finite-dimensional marginal distributions $\mu_J = \mu \circ \pi_J^{-1}$ with $J \subset [0, 1]$ - finite, where $\pi_J = \pi_J^{[0,1]}$, $x \mapsto (x(t_j))_{t_j \in J}$, are the canonical projections, since cylinder sets are $\cap$-stable generator of $\mathcal{B}(C[0, 1])$ by Lemma 4.8.

Lemma 4.9. *(a) If $\mu_n \Rightarrow \mu$ for some $\mu, \mu_n \in \mathcal{M}_1(C[0, 1])$, then all finite-dimensional marginal distributions of μ_n converge to the finite-dimensional marginal distributions of μ, i.e.*

$$\mu_{n,J} \Rightarrow \mu_J \qquad \forall J \subset [0, 1] \text{ - finite.}$$

(b) The converse implication is not true.

Theorem 4.10. *For probability measures μ_n, μ, on $(C[0, 1], \mathcal{B}(C[0, 1]))$, the following are equivalent* *(a)* $\mu_n \Rightarrow \mu$,

(b) *all finite-dimensional marginal distributions of* μ_n *converge against these of* μ, *i.e.* $\mu_{n,J} \equiv \mu_n \circ \pi_J^{-1} \Rightarrow \mu \circ \pi_J^{-1} \equiv \mu_J$ *for all finite* $J \subset [0,1]$, *and the sequence* $(\mu_n)_{n\in\mathbb{N}}$ *is tight.*

For $x \in C[0,1]$ we define

$$W_\delta(x) := \sup\{|x(t) - x(s)| \mid s,t \in [0,1] \text{ with } |t-s| \leqslant \delta\}, \qquad \delta > 0.$$

Note that for any $x \in C[0,1]$ holds $\lim_{\delta\downarrow 0} W_\delta(x) = 0$ since continuity on $[0,1]$ implies uniform continuity.

Proposition 4.11 (Arzelà-Ascoli). *A set* $A \subset C[0,1]$ *is relatively compact in* $C[0,1]$ *if and only if the following two conditions are satisfied:*

(1) *A is uniformly bounded, i.e.* $\sup_{x\in A} \|x\| < \infty$,

(2) *A is equi-continuous , i.e.* $\lim_{\delta\downarrow 0} \sup_{x\in A} W_\delta(x) = 0$.

Given (2) in Proposition 4.11, one can replace (1) by

$$\text{(1') } A \text{ is uniformly bounded at 0, i.e. } \sup_{x\in A} |x(0)| < \infty.$$

Theorem 4.12. *Let* $\mathcal{M} \subset \mathcal{M}_1(C[0,1])$ *and define* $\mu^0 := \mu \circ \pi_{\{0\}}^{-1} = \mu(\{x(0) \in \cdot\})$ *for* $\mu \in \mathcal{M}$ *. Then* $\mathcal{M}$ *is tight if and only if* $\{\mu^0 | \ \mu \in \mathcal{M}\}$ *is tight on* $\mathbb{R}$ *and*

$$\lim_{\delta\downarrow 0} \sup_{\mu\in\mathcal{M}} \mu(\{W_\delta(x) \geqslant \varepsilon\}) = 0 \qquad \forall \varepsilon > 0\,. \tag{4.3.1}$$

Remark 4.6. (a) In Theorem 4.12 the set $\{\mu^0 | \ \mu \in \mathcal{M}\}$ would be trivially tight if for example $\mu(\{x| \ x(0) = 0\}) = 1$ for all $\mu \in \mathcal{M}$.

(b) If $\mathcal{M} = \{\mu_n | \ n \in \mathbb{N}\}$, then (4.3.1) is equivalent to

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mu_n(W_\delta(x) \geqslant \varepsilon) = 0. \tag{4.3.2}$$

Chapter 5

Donsker's invariance principle

If $(B_t)_{t\in[0,1]}$ is a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the mapping $B : \Omega \to C[0,1], \omega \mapsto (t \mapsto B_t(\omega))$, induces a probability measure on $(C[0,1], \mathcal{B}(C[0,1]))$, namely the distribution $\mu = \mathbb{P} \circ (B_{[0,1]})^{-1}$ of $B_{[0,1]}$ under $\mathbb{P}$.

Definition 5.1. The measure μ on $(C[0,1], \mathcal{B}(C[0,1]))$ whose finite dimensional distributions are centered multivariate Gaussian with Covariance function $s \wedge t$ $(s,t \in [0,1])$ (i.e. the distribution of Brownian motion) is called *the Wiener measure* (for time interval $[0,1]$).

Remark 5.1. (a) Existence of the Wiener measure is equivalent to existence of Brownian motion (though its definition does not require existence of the latter). Indeed, taking $\Omega = C[0,1]$, $\mathcal{F} = \mathcal{B}(C[0,1])$, and $X_t(\omega) = \omega(t)$ as the coordinate process, $\mathbb{P} = \mu$ the Wiener measure, then $(X_t)_{t\in[0,1]}$ is a BM (on $[0,1]$) with respect to the natural filtration $\mathcal{F}_t = \sigma(X_s : \ s \leqslant t)$.

(b) We already know existence of Brownian motion as a continuous centered Gaussian process X with covariance function $\mathrm{Cov}(X_t, X_s) = t \wedge s$ (Kolmogorov's consistency theorem together with Kolmogorov-Chentsov continuity theorem for a continuous modification). Hence, the Wiener measure exists.

(c) The Wiener measure is defined analogously for any time interval, e.g. for $[0,\infty)$.

Setting now: $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space with $(Y_k)_{k\in\mathbb{N}}$ a sequence of real-valued i.i.d.'s with $\mathbb{E}[Y_1] = 0$, $\mathrm{Var}(Y_1) = 1$. Define $S_n := \sum_{k=1}^n Y_k$, $S_0 := 0$ - random walk with i.i.d. increments. Consider the rescaled random variables $\frac{1}{\sqrt{n}} S_l$, $l = 0, \ldots, n$, on equidistant time grid on $[0,1]$ with linear interpolation:

$$X_t^n = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} + \frac{1}{\sqrt{n}} Y_{\lfloor nt \rfloor + 1}(nt - \lfloor nt \rfloor), \qquad t \in [0,1].$$

X^n is a continuous stochastic process, $X^n_\cdot : \Omega \to C[0,1]$ is $\mathcal{F} - \mathcal{B}(C[0,1])$-measurable: all $X_t^n = \pi_t \circ X^n_\cdot$ are $\mathcal{F} - \mathcal{B}(\mathbb{R})$-measurable and $\sigma(\pi_t : \ t \in [0,1]) = \mathcal{B}(C[0,1])$ by Lemma 4.8. Denote by $\mu_n = \mathbb{P} \circ (X^n)^{-1} \in \mathcal{M}_1(C[0,1])$ the distribution of X^n under $\mathbb{P}$.

Theorem 5.2 (Donsker)**.** *Under the above considerations, it holds that the sequence* (μ_n) *converges weakly and the limit is the Wiener measure* μ*, i.e.*

$$\mu_n \Rightarrow \mu \qquad (\equiv \textit{Wiener measure on } C[0,1]);$$

that means that $X^n_{\cdot} \xrightarrow{\mathcal{D}} B_{\cdot}$ *for Brownian motion* B *(one-dimensional, on* $[0,1]$*).*

Remark 5.2. (1) Existence and construction of μ.

(2) *Invariance principle:* only requirement for $(Y_k)_{k\in\mathbb{N}}$ is that they are i.i.d. with second moments.

(3) Functional version of the CLT: not only the convergence $\frac{1}{\sqrt{n}}S_n = X^n_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, but convergence of processes.

(4) Instead of sequences $(Y_k)_{k\in\mathbb{N}}$, the result also holds for "triangle-schemes" where each X^n relies on other increments $(Y^n_k)_{k\in\mathbb{N}}$ ...

(5) For continuous functions $F : C[0,1] \to S$, we get that $F(X^n) \xrightarrow{\mathcal{D}} F(B)$, where B is a Brownian motion. This is used, for instance, in financial mathematics to show that the CRR binomial model converges to the Black-Scholes model.

Theorem 5.3 (Multivariate CLT)**.** *Let* $(X_k)_{k\in\mathbb{N}} = ((X^1_k, \ldots, X^d_k))_{k\in\mathbb{N}}$*,* $k \in \mathbb{N}$*, be* $\mathbb{R}^d$*-valued i.i.d. random variables with* $\mathbb{E}[X_1] = b \in \mathbb{R}^d$*,* $Cov(X^j_1, X^l_1) = \Sigma_{jl}$*,* $j, l \in \{1, \ldots, d\}$*. Let also* $S^*_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n (X_k - b)$*. Then* $S^*_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma)$ *as* $n \to \infty$*.*

Lemma 5.4. *Let* $(U_n)_{n\in\mathbb{N}}$*,* U*, be random variables on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *taking values in a normed space* $(S, \|\cdot\|)$*.*

(1) Let (V_n) *be other* S*-valued random variables with* $V_n \xrightarrow{\mathbb{P}} 0$*. If* S *is separable and* $U_n \xrightarrow{\mathcal{D}} U$*, then* $U_n + V_n \xrightarrow{\mathcal{D}} U$*.*

(2) Let (c_n) *be a sequence in* $\mathbb{R}$ *such that* $c_n \to c \in \mathbb{R}$ *as* $n \to \infty$*. If* $U_n \xrightarrow{\mathcal{D}} U$*, then* $c_nU_n \xrightarrow{\mathcal{D}} cU$*.*

Proposition 5.5. *With the set-up from above, the finite-dimensional marginal distributions of* $(\mu_n)_{n\in\mathbb{N}}$ *converge weakly to the finite-dimensional marginal distributions of* μ*, i.e.*

$$\underbrace{\mu_n \circ (\pi^{[0,1]}_J)^{-1}}_{=:\mu^J_n} \quad\Longrightarrow\quad \underbrace{\mu \circ (\pi^{[0,1]}_J)^{-1}}_{=:\mu^J} \qquad \forall \textit{ finite } J \subset [0,1],$$

where $\pi^{[0,1]}_J$ *denotes the projection* $x \mapsto (x(j))_{j\in J}$*.*

Lemma 5.6 (maximum inequality, Ottaviani)**.** *Let* $U_1, \ldots, U_n$ *be independent random variables with* $\mathbb{E}[U_i] = 0$*,* $Var(U_i) = E[U_i^2] = 1$*. Set* $Z_k := \frac{1}{\sqrt{n}}\sum_{i=1}^k U_i$*. Then we have*

$$\mathbb{P}\left[\max_{k=1,\ldots,n} |Z_k| > 2\alpha\right] \leqslant \frac{1}{1-\frac{1}{\alpha^2}}\mathbb{P}[|Z_n| > \alpha] \qquad \forall \alpha > 1.$$

Proposition 5.7. *In the set-up from above, the sequence* $(\mu_n)_{n\in\mathbb{N}}$ *is tight.*

Bibliography

[Aleks01] Aleksandrov P.S., *Lehrbuch der Mengenlehre*, Harri Deutsch (Verlag), 2001.

[Alt02] Alt H. W., *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*, Springer, 2002.

[Bauer01] Bauer H., *Measure and Integration Theory*, Walter de Gruyter, 2001.

[Dieu60] Dieudonné J., *Foundations of Modern Analysis*, New York, Academic Press, 1960.

[DellMey79] Dellacherie C., Meyer P.-A., *A Probabilities and Potential*, North-Holland Mathematics Studies, 1979.

[JacPro03] Jacod J., Protter P., *Probability Essentials*, Universitext, Springer, 2003.

[Klenke06] Klenke A., *Probability Theory: A Comprehensive Course.*, Universitext, Springer, 2006.

[Kall02] Kallenberg O., *Foundations of Modern Probability*, Springer, 2002.

[Will91] Williams D., *Probability with Martingales*, Cambridge University Press, 1991.

[a.s. wordpress] http://almostsure.wordpress.com/2009/12/06/upcrossings-downcrossings-and-martingale-convergence/