Probability Theory

Wilhelm Stannat

Technische Universität Darmstadt

Winter Term 2007/08

Corrected version

This text is a summary of the lecture on Probability Theory held at the TU Darmstadt in Winter Term 2007/08.

Please email all misprints and mistakes to

`stannat@mathematik.tu-darmstadt.de`

Bibliography

1. Bauer, H., *Probability theory*, de Gruyter, 1996.
2. Bauer, H., *Maß- und Integrationstheorie*, de Gruyter, 1996.
3. Billingsley, P., *Probability and Measure*, Wiley, 1995.
4. Billingsley, P., *Convergence of probability measures*, Wiley, 1999.
5. Dudley, R.M., *Real analysis and probability*, Cambridge University Press, 2002.
6. Elstrodt, J., *Maß- und Integrationstheorie*, Springer, 2005.
7. Feller, W., *An introduction to probability theory and its applications*, Vol. 1 & 2, Wiley, 1950.
8. Halmos, P.R., *Measure Theory*, Springer, 1974.
9. Klenke, A., *Wahrscheinlichkeitstheorie*, Springer, 2006.
10. Shiryaev, A.N., *Probability*, Springer, 1996.

1 Basic Notions

1 Probability spaces

Probability theory is the mathematical theory of randomness. The basic notion is that of a random experiment, which is an event whose outcome is not predictable and can only be determined after performing them and then observing the outcome.

Probability theory tries to quantify the possible outcomes by attaching a probability to every event. This is of importance for example for an insurance company when asking the question what is a fair price of an insurance against events like fire or death that are events that can happen but need not happen.

The set of all possible outcomes of a random experiment is denoted by Ω. The set Ω may be finite, infintely countable or even uncountable.

Example 1.1. Examples of random experiments and corresponding Ω:

(i) **Coin tossing** The possible outcomes of tossing a coin are either "head" or "tail". Denoting one outcome by "0" and the other one by "1", the set of all possible outcomes is given by $\Omega = \{0, 1\}$.

(ii) **Tossing a coin n times** In this case any sequence of zeros and ones (alias heads or tails) of length n are considered as one possible outcome; hence

$$\Omega = \big\{(x_1, x_2, \dots, x_n) \,\big|\, x_i \in \{0,1\}\big\} =: \{0,1\}^n$$

is the space of all possible outcomes.

(iii) **Tossing a coin infintely many times** In this case

$$\Omega = \big\{(x_i)_{i\in\mathbb{N}} \,\big|\, x_i \in \{0,1\}\big\} =: \{0,1\}^{\mathbb{N}}.$$

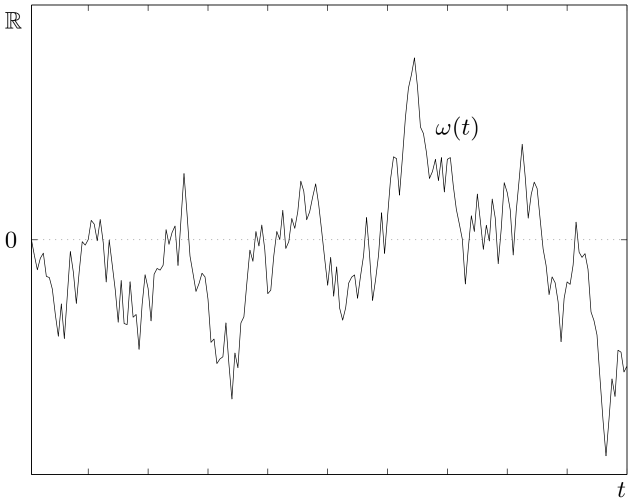
In this case Ω is uncountable in contrast to the previous examples. We can identify Ω with the set $[0,1] \subset \mathbb{R}$ using the binary expansion

$$x = \sum_{i=1}^{\infty} x_i 2^{-i}\,.$$

(iv) **A random number between 0 and 1** $\Omega = [0,1]$.

(v) **Continuous stochastic processes, e.g. Brownian motion on $\mathbb{R}$** Any continuous real-valued function defined on $[0,1] \subset \mathbb{R}$ is a possible outcome. In this case $\Omega = \mathcal{C}\big([0,1]\big)$.

Example:



Events
Subsets $A \subset \Omega$ are called *events*. If $\omega \in A$ we say that A *has occured*.

- *elementary events* $A = \{\omega\}$ for some $\omega \in \Omega$
- the *impossible event* $A = \emptyset$ and the *certain event* $A = \Omega$
- *"A does not occur"* $A^c = \Omega \setminus A$

Combination of events

$A_1 \cup A_2, \quad \bigcup_i A_i$	"at least one of the events A_i occur"
$A_1 \cap A_2, \quad \bigcap_i A_i$	"all of the events A_i occur"
$\limsup_{n\to\infty} A_n := \bigcap_n \bigcup_{m \geqslant n} A_m$	"infinitely many of the A_m occur"
$\liminf_{n\to\infty} A_n := \bigcup_n \bigcap_{m \geqslant n} A_m$	"all but finitely many of the A_m occur"

Example 1.2. (i) **Coin tossing** "1 occurs": $A = \{1\}$

(ii) **Tossing a coin n times** "tossing k ones":

$$A = \Big\{(x_1, \ldots, x_n) \in \{0,1\}^n \;\Big|\; \sum_{i=1}^n x_i = k\Big\}.$$

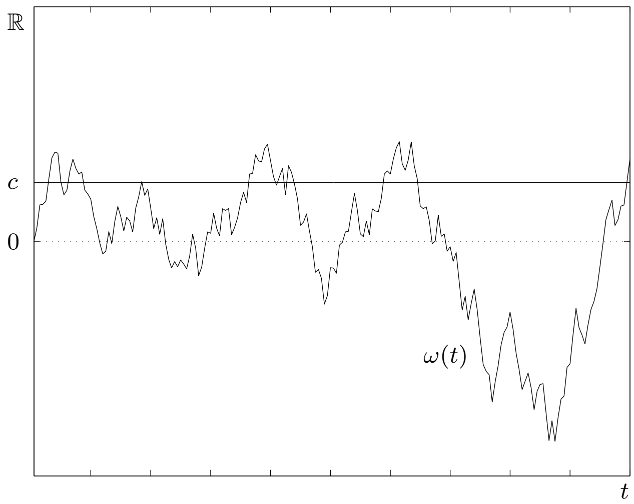
(iii) **Tossing a coin infinitely many times** "relative frequency of 1 equals p":

$$A = \Big\{(x_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} \;\Big|\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n x_i = p\Big\}.$$

(iv) **random number** 0 **and** 1. "number $\in [a,b]$": $A = [a,b] \subset \Omega = [0,1]$.

(v) **Continuous stochastic processes** "exceeding level c":

$$A = \left\{ \omega \in \mathcal{C}\big([0,1]\big) \;\middle|\; \max_{0 \leqslant t \leqslant 1} \omega(t) > c \right\}.$$



Let Ω be countable. A *probability function* p on Ω is a function

$$p : \Omega \to [0,1] \text{ with } \sum_{\omega \in \Omega} p(\omega) = 1\,.$$

Given any subset $A \subset \Omega$, its probability $P(A)$ can then be defined by simply adding up

$$P(A) = \sum_{\omega \in A} p(\omega)\,.$$

In the uncountable case, however, there is no reasonable way of adding up an uncountable set of numbers. There is no way to build a reasonable theory by starting with probability functions specifying the probability of individual outcomes. The best way out is to specify directly the probability of events. In the uncountable case it is not possible in general to consider the power set $\mathcal{P}(\Omega)$, i.e. the collection of *all* subsets of Ω (including the empty set $\emptyset$ and the whole set Ω) but only a certain subclass. On the other hand $\mathcal{A}$ should satisfy some minimal requirements specified in the following definition:

Definition 1.3. $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called a σ-*algebra* if

(i) $\Omega \in \mathcal{A}$.

(ii) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$.

(iii) $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, implies $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

Remark 1.4. *(i) Let $\mathcal{A}$ be a σ-algebra. Then:*

- $\emptyset = \Omega^c \in \mathcal{A}$.
- $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *implies*

$$\bigcap_{i\in\mathbb{N}} A_i = \Big(\bigcup_{i\in\mathbb{N}} A_i^c\Big)^c \in \mathcal{A}.$$

- $A_1, \dots, A_n \in \mathcal{A}$ *implies*

$$\bigcup_{i=1}^n A_i \in \mathcal{A} \quad \textit{and} \quad \bigcap_{i=1}^n A_i \in \mathcal{A}$$

(consider $A_m = \emptyset$ for all $m > n$, resp. $A_m = \Omega$ for all $m > n$).

- $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *implies*

$$\bigcap_n \bigcup_{m\geqslant n} A_m \in \mathcal{A} \quad \textit{and} \quad \bigcup_n \bigcap_{m\geqslant n} A_m \in \mathcal{A}.$$

(ii) the power set $\mathcal{P}(\Omega)$ is a σ-algebra.

(iii) Let I be an index set (not necessarily countable) and for any $i \in I$, let $\mathcal{A}_i$ be a σ-algebra. Then $\bigcap_{i\in I} \mathcal{A}_i$ is again a σ-algebra.

(iv) **Typical construction of a σ-algebra** *Let $\mathcal{A}_0 \neq \emptyset$ be a class of events. Then*

$$\sigma(\mathcal{A}_0) := \bigcap_{\substack{\mathcal{B} \text{ is } \sigma\text{-algebra}, \\ \mathcal{A}_0 \subset \mathcal{B}}} \mathcal{B}.$$

$\sigma(\mathcal{A}_0)$ is the smallest σ-algebra containing $\mathcal{A}_0$. $\sigma(\mathcal{A}_0)$ is called the σ-algebra generated by $\mathcal{A}_0$.

Example 1.5. Let Ω be a topological space, and $\mathcal{A}_o$ be the collection of open subsets of Ω. Then $\mathcal{B}(\Omega) := \sigma(\mathcal{A}_o)$ is called the *Borel-σ-algebra* of Ω, or σ-algebra of *Borel-subsets*.

Example of Borel-subsets: closed sets, countable unions of closed sets, etc...

Note: not every subset of a topological space is a Borel-subset, e.g. $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.

Definition 1.6. Let $\Omega \neq \emptyset$ and $\mathcal{A} \subset \mathcal{P}(\Omega)$ a σ-algebra. A mapping $P : \mathcal{A} \to [0, \infty]$ is called a *measure* (on $(\Omega, \mathcal{A})$) if:

- $P(\emptyset) = 0$
-

$$P\Big(\dot{\bigcup_{i\in\mathbb{N}}} A_i\Big) = \sum_{i=1}^{\infty} P(A_i) \qquad \text{("}\sigma\text{-additivity")}$$

for all pairwise disjoint $A_i \in \mathcal{A}$, $i \in \mathbb{N}$.

P is called a *probability measure* if in addition

- $P(\Omega) = 1$.

In this case $(\Omega, \mathcal{A}, P)$ is called a *probability space*. The pair $(\Omega, \mathcal{A})$ of a set Ω together with a σ-algebra is called a *measurable space*.

Example 1.7. (i) **Coin tossing** Let $\mathcal{A} := \mathcal{P}(\Omega) = \big\{\emptyset, \{0\}, \{1\}, \{0,1\}\big\}$.
Tossing a *fair coin* means "head" and "tail" have equal probability 0.5, hence:

$$P(\{0\}) := P(\{1\}) := \frac{1}{2}, \quad P(\emptyset) := 0, \quad P(\underbrace{\{0,1\}}_{=\Omega}) := 1.$$

(iii) **Tossing a coin infinitely many times** $\Omega = \{0,1\}^{\mathbb{N}}$.
Let $\mathcal{A} := \sigma(\mathcal{A}_0)$ where

$$\mathcal{A}_0 := \big\{B \subset \Omega \bigm| \exists\, n \in \mathbb{N} \text{ and } B_0 \in \mathcal{P}\big(\{0,1\}^n\big), \\ \text{such that } B = B_0 \times \{0,1\} \times \{0,1\} \times \dots\big\}.$$

An event B is contained in $\mathcal{A}_0$ if it depends on *finitely* many tosses.
Fix $\bar{x}_1, \dots, \bar{x}_n \in \{0,1\}$ and define

$$P\big(\underbrace{\big\{(x_1, x_2, \dots) \in \{0,1\}^{\mathbb{N}} \bigm| x_1 = \bar{x}_1, \dots, x_n = \bar{x}_n\big\}}_{\in \mathcal{A}_0}\big) := 2^{-n}.$$

P can be extended to a probability measure on $\mathcal{A} = \sigma(\mathcal{A}_0)$. For this probability measure we have that

$$P\left(\left\{(x_1, x_2, \dots) \in \{0,1\}^{\mathbb{N}} \;\middle|\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{2}\right\}\right) = 1.$$

(Proof: Later!)

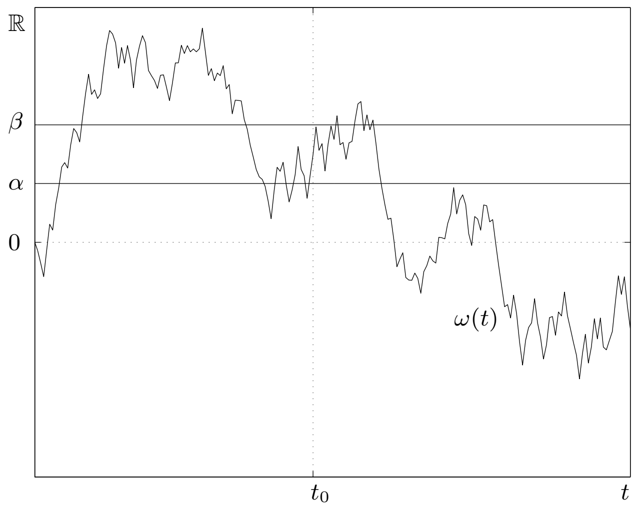
(v) **Continuous stochastic processes** $\Omega = \mathcal{C}([0,1])$, $P =$ Wiener measure ("Brownian motion") For fixed $t_0 \in \mathbb{R}_+$ and $\alpha, \beta \in \mathbb{R}$ we have that

$$P\big(\big\{\omega \bigm| \omega(t_0) \in [\alpha, \beta]\big\}\big) := \frac{1}{\sqrt{2\pi t_0}} \int_\alpha^\beta e^{-\frac{x^2}{2t_0}} \,\mathrm{d}x$$

(Gaussian or normal distribution).

What is now the probability $P\big(\big\{\omega \bigm| \max\limits_{0 \leqslant t \leqslant 1} \omega(t) > c\big\}\big)$?

Answer to this question: Later!



Remark 1.8. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $A_1, \ldots, A_n \in \mathcal{A}$ be pairwise disjoint. Then*

$$P\Big(\dot{\bigcup_{i\leqslant n}} A_i\Big) = \sum_{i=1}^{n} P(A_i)$$

(simply let $A_m = \emptyset$ for all $m > n$). In particular:

$$\begin{aligned} A, B \in \mathcal{A},\ A \subset B &\Rightarrow P(B) = P(A) + P(B \setminus A) \\ &\Rightarrow P(B \setminus A) = P(B) - P(A), \end{aligned}$$

and $P(A^c) = P(\Omega \setminus A) = P(\Omega) - P(A) = 1 - P(A)$.

P *is* subadditive, *that is, for $A, B \in \mathcal{A}$*

$$\begin{aligned} P(A \cup B) &= P\big(A \,\dot\cup\, [B \setminus (A \cap B)]\big) \\ &= P(A) + P(B) - P(A \cap B) \\ &\leqslant P(A) + P(B), \end{aligned} \tag{1.1}$$

and by induction one obtains Sylvester's formula*:*
Let I be a finite index set, A_i, $i \in I$, be a collection of subsets in $\mathcal{A}$ (not necessarily disjoint). Then

$$\begin{aligned} P\Big(\bigcup_{i\in I} A_i\Big) \quad &= \quad \sum_{\substack{J \subset I, \\ J \neq \emptyset}} (-1)^{|J|-1} \cdot P\Big(\bigcap_{j\in J} A_j\Big) \\ &\overset{I=\{1,\ldots,n\}}{=} \sum_{k=1}^{n} (-1)^{k-1} \cdot \sum_{1\leqslant i_1<\cdots<i_k\leqslant n} P(A_{i_1} \cap \cdots \cap A_{i_k}). \end{aligned} \tag{1.2}$$

Proposition 1.9. *Let $\mathcal{A}$ be a σ-algebra, and $P : \mathcal{A} \to \mathbb{R}_+$ be a mapping with $P(\Omega) = 1$. Then the following are equivalent:*

(i) P *is a probability measure.*

(ii) P *is* additive, *that is,* $A, B \in \mathcal{A}$, $A \cap B = \emptyset$ *implies* $P(A \mathbin{\dot\cup} B) = P(A) + P(B)$, *and*

continuous from below, *that is,* $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *with* $A_i \subset A_{i+1}$ *for all* $i \in \mathbb{N}$, *implies*

$$P\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \lim_{i\to\infty} P(A_i).$$

(iii) P *is additive and* continuous from above, *that is* $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, *with* $A_i \supset A_{i+1}$ *for all* $i \in \mathbb{N}$, *implies*

$$P\Big(\bigcap_{i\in\mathbb{N}} A_i\Big) = \lim_{i\to\infty} P(A_i).$$

Corollary 1.10 (σ-subadditivity)**.** *Let* A_i, $i \in \mathbb{N}$, *be a sequence of subsets in* $\mathcal{A}$ *(not necessarily pairwise disjoint). Then:*

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \leqslant \sum_{i=1}^{\infty} P(A_i).$$

Proof.

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \overset{1.9}{=} \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} A_i\Big) \overset{(1.1)}{\leqslant} \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$ □

Lemma 1.11 (Borel-Cantelli)**.** *Let* $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. *Then*

$$\sum_{i=1}^{\infty} P(A_i) < \infty \quad \Rightarrow \quad P\Big(\underbrace{\bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m}_{=:\limsup\limits_{n\to\infty} A_n}\Big) = 0.$$

Proof. Since

$$\bigcup_{m\geqslant n} A_m \overset{n\to\infty}{\searrow} \bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m$$

the continuity from above of P implies that

$$P\big(\limsup_{n\to\infty} A_n\big) \overset{1.9}{=} \lim_{n\to\infty} P\Big(\bigcup_{m\geqslant n} A_m\Big) \overset{1.10}{\leqslant} \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) = 0,$$

since $\sum_{m=1}^{\infty} P(A_m) < \infty$. □

Example 1.12. (i) **Uniform distribution on** $[0,1]$ Let $\Omega = [0,1]$ and $\mathcal{A}$ be the Borel-σ-algebra on Ω $(= \sigma\big(\{[a,b] \mid 0 \leqslant a \leqslant b \leqslant 1\}\big))$. Let P be the restriction of the Lebesgue measure on the Borel subset of $\mathbb{R}$ to $[0,1]$. Then $(\Omega, \mathcal{A}, P)$ is a probability space. The probability measure P is called the uniform distribution on $[0,1]$, since $P([a,b]) = b - a$ for any $0 \leq a \leq b \leq 1$ (*translation invariance*).

(ii) **Dirac-measure** Let $\Omega \neq \emptyset$ and $\omega_0 \in \Omega$. Let $\mathcal{A}$ be an arbitrary σ-algebra on Ω (e.g. $\mathcal{A} = \mathcal{P}(\Omega)$). Then

$$P(A) := 1_A(\omega_0) := \begin{cases} 1 & \text{if } \omega_0 \in A \\ 0 & \text{if } \omega_0 \notin A. \end{cases}$$

defines a probability measure on $\mathcal{A}$. P is called the *Dirac-measure in* ω_0, denoted by $P = \delta_{\omega_0}$ or $P = \varepsilon_{\omega_0}$.

(iii) **Convex combinations of probability measures** Let $\Omega \neq \emptyset$ and $\mathcal{A}$ be a σ-algebra of subsets of Ω. Let I be a countable index set. Let P_i, $i \in I$, be a family of probability measures on $(\Omega, \mathcal{A})$, and $\alpha_i \in [0,1]$, $i \in I$, be such that $\sum_{i \in I} \alpha_i = 1$. Then $P := \sum_{i \in I} \alpha_i \cdot P_i$ is again a probability measure on $(\Omega, \mathcal{A})$.

This holds in particular for

$$P := \sum_{i \in I} \alpha_i \cdot \delta_{\omega_i}$$

if $\omega_i \in \Omega$, $i \in I$.

2 Discrete models

Throughout the whole section

- $\Omega \neq \emptyset$ countable
- $\mathcal{A} = \mathcal{P}(\Omega)$ and
- $\omega \in \Omega$ an elementary event.

Proposition 2.1. *(i) Let $p : \Omega \to [0,1]$ be a function with $0 \leqslant p(\omega) \leqslant 1$ for all $\omega \in \Omega$ and $\sum_{\omega \in \Omega} p(\omega) = 1$ (p is called a probability distribution). Then*

$$P(A) := \sum_{\omega \in A} p(\omega) \qquad \forall A \subset \Omega$$

defines a probability measure on $(\Omega, \mathcal{A})$.

(ii) Every probability measure P on $(\Omega, \mathcal{A})$ is of this form, with $p(\omega) := P(\{\omega\})$ for all $\omega \in \Omega$.

Proof. (i)

$$P = \sum_{\omega \in \Omega} p(\omega) \cdot \delta_\omega \,.$$

(ii) Exercise. $\square$

Example 2.2 (Laplace probability space)**.** Fundamental example in the discrete case that forms the basis of many other discrete models.

Let Ω be a nonempty finite set (that is $0 < |\Omega| < \infty$). Define

$$p(\omega) = \frac{1}{|\Omega|} \qquad \forall \omega \in \Omega\,.$$

Then

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{number of convenient outcomes}}{\text{number of possible outcomes}}.$$

Hence measure theoretic problems reduce to combinatorial problems in the discrete case.

P is called *uniform distribution* on Ω, because every elementary event $\omega \in \Omega$ has the same probability $\frac{1}{|\Omega|}$.

Example 2.3. (i) **random permutations** Let $M := \{1, \dots, n\}$ and $\Omega :=$ all permutations of M. Then $|\Omega| = n!$. Let P be the uniform distribution on Ω.

Problem: What is the probability P("at least one fixed point")?

Consider the event $A_i := \{\omega \,|\, \omega(i) = i\}$ (fixed point at position i). Then Sylvester's formula (cf. (1.2)) implies that

$$\begin{aligned}
P(\text{"at least one fixed point"}) &= P\Big(\bigcup_{i=1}^n A_i\Big) \\
&\overset{(1.2)}{=} \sum_{k=1}^n (-1)^{k+1} \cdot \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \underbrace{P(A_{i_1} \cap \cdots \cap A_{i_k})}_{= \frac{(n-k)!}{n!} (k \text{ pos. fixed})} \\
&= \sum_{k=1}^n (-1)^{k+1} \cdot \binom{n}{k} \frac{(n-k)!}{n!} = -\sum_{k=1}^n \frac{(-1)^k}{k!}.
\end{aligned}$$

Consequently,

$$P(\text{"no fixed point"}) = 1 + \sum_{k=1}^n \frac{(-1)^k}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!} \xrightarrow{n \to \infty} e^{-1}$$

and thus for all $k \in \{0, \dots, n\}$:

$$\begin{aligned}
&P(\text{"exactly } k \text{ fixed points"}) \\
&= \underbrace{\frac{1}{n!}}_{\substack{\text{all poss.} \\ \text{outcomes}}} \cdot \underbrace{\binom{n}{k}}_{\substack{\text{all } \omega \\ \text{with } k \text{ positions} \\ \text{fixed}}} \cdot \underbrace{(n-k)! \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}}_{\substack{n-k \text{ positions} \\ \text{without fixed points}}} = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}.
\end{aligned}$$

Asymptotics as $n \to \infty$:

$$P(\text{“exactly } k \text{ fixed points”}) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \xrightarrow{n\to\infty} \frac{1}{k!} \cdot e^{-1}$$

(Poisson distribution with parameter $\lambda = 1$).

Recall: *Poisson distribution with parameter* λ *(*> 0*)*

$$\pi_\lambda := e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot \delta_k.$$

(ii) n **experiments with state space** S**,** $|S| < \infty$

$$\Omega := \{\omega = (x_1, \ldots, x_n) \mid x_i \in S\}, \quad |\Omega| = |S|^n.$$

Let P be the uniform distribution on Ω.

Fix a subset $S_0 \subset S$, such that $x_i \in S_0$ is called a “success”, hence $p := \frac{|S_0|}{|S|}$ is the probability of success.

What is the probability of the event $A_k =$ “(exactly) k successes”, $k = 0, \ldots, n$?

$$P(A_k) = \frac{|A_k|}{|\Omega|} = \binom{n}{k} \cdot \frac{|S_0|^k \cdot |S \setminus S_0|^{n-k}}{|S|^n} = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

(Binomial distribution with parameters n, p).

Recall: *Binomial distribution with parameters* n, p

$$B(n,p) := \beta_n^p := \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot \delta_k.$$

Let $p := P(\text{“Success in the } i\text{-th experiment”})$, $i = 1, \ldots, n$, and consider the asymptotics of the binomial distribution as $n \to \infty$ for $p_n := \frac{\lambda}{n}$. Then

$$\begin{aligned}
&\binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&\quad = \quad \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}}_{\xrightarrow{n\to\infty} 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n-k}}_{\xrightarrow{n\to\infty} e^{-\lambda}} \\
&\quad \xrightarrow{n\to\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} \qquad (k = 0, 1, 2, \ldots)
\end{aligned}$$

(Poisson distribution with parameter λ).

(iii) **Urn model (for example: opinion polls, samples, ...)** We consider an urn containing N balls, K red and $N-K$ black. Suppose that $n \leqslant N$ balls are sampled without replacement. What is the probability that exactly k balls in the sample are red?

typical application: suppose that a small lake contains an (unknown) number N of fish. To estimate N one can do the following: K fish will be marked by red and after that n ($n \leq N$) fish are "sampled" from the lake. If k is the number of marked fish in the sample, $\hat{N} := K \cdot \frac{n}{k}$ is an estimation of the unknown number N.

Model:
Let Ω be all subsets of $\{1, \dots, N\}$ having cardinality n, hence

$$\Omega := \{\omega \in \mathcal{P}(\{1, \dots, N\}) \mid |\omega| = n\}, \quad |\Omega| = \binom{N}{n}$$

and let P be the uniform distribution on Ω. Consider the event $A_k :=$ "exactly k red". Then

$$|A_k| = \binom{K}{k}\binom{N-K}{n-k}$$

so that

$$P(A_k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \; (k = 0, \dots, n) \quad \text{hypergeometric distr.}$$

Asymptotics for $N \to \infty$, $K \to \infty$ with $p := \frac{K}{N}$ and n fixed:

$$P(A_k) \longrightarrow \binom{n}{k} \cdot p^k (1-p)^{n-k} \qquad (k = 0, \dots, n)$$

3 Transformations of probability spaces

Throughout this section let $(\Omega, \mathcal{A})$ and $(\tilde{\Omega}, \tilde{\mathcal{A}})$ be measurable spaces.

Definition 3.1. A mapping $T : \Omega \to \tilde{\Omega}$ is called $\mathcal{A}/\tilde{\mathcal{A}}$-*measurable* (or simply *measurable*), if $T^{-1}(\tilde{A}) \in \mathcal{A}$ for all $\tilde{A} \in \tilde{\mathcal{A}}$.
Notation:

$$\{T \in \tilde{A}\} := T^{-1}(\tilde{A}) = \{\omega \in \Omega \mid T(\omega) \in \tilde{A}\}.$$

Remark 3.2. *(i) Clearly, if $\mathcal{A} := \mathcal{P}(\Omega)$ then every mapping $T : \Omega \to \tilde{\Omega}$ is measurable.*

(ii) Sufficient criterion for measurability *Suppose that $\tilde{\mathcal{A}} := \sigma(\tilde{\mathcal{A}}_0)$ for some collection of subsets $\tilde{\mathcal{A}}_0 \subset \mathcal{P}(\Omega)$. Then T is $\mathcal{A}/\tilde{\mathcal{A}}$-measurable, if $T^{-1}(\tilde{A}) \in \mathcal{A}$ for all $\tilde{A} \in \tilde{\mathcal{A}}_0$.*

(iii) Let $\Omega, \tilde{\Omega}$ be topological spaces, and $\mathcal{A}, \tilde{\mathcal{A}}$ be the associated Borel σ-algebras. Then:

$$T : \Omega \to \tilde{\Omega} \text{ is continuous} \quad \Rightarrow \quad T \text{ is } \mathcal{A}/\tilde{\mathcal{A}}\text{-measurable.}$$

(iv) Let $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2, 3$, be measurable spaces, and $T_i : \mathcal{A}_i \to \mathcal{A}_{i+1}$, $i = 1, 2$, measurable mappings. Then:

$$T_2 \circ T_1 \text{ is } \mathcal{A}_1/\mathcal{A}_3\text{-measurable.}$$

Proof. (ii) $\{\tilde{A} \in \mathcal{P}(\tilde{\Omega}) \,|\, T^{-1}(\tilde{A}) \in \mathcal{A}\}$ is a σ-algebra containing $\mathcal{A}_0$. Consequently,

$$\sigma(\tilde{\mathcal{A}}_0) \subset \big\{\tilde{A} \in \mathcal{P}(\tilde{\Omega}) \;\big|\; T^{-1}(\tilde{A}) \in \mathcal{A}\big\}.$$

(iii) Easy consequence of (ii).

(iv) Exercise. □

Definition 3.3. Let $T : \bar{\Omega} \to \Omega$ be a mapping and let $\mathcal{A}$ be a σ-Algebra of subsets of Ω. The system

$$\sigma(T) := \big\{T^{-1}(A) \;\big|\; A \in \mathcal{A}\big\}$$

is a σ-algebra of subsets of $\bar{\Omega}$; $\sigma(T)$ is called the *σ-algebra generated by T*. More precisely: $\sigma(T)$ is the smallest σ-algebra $\bar{\mathcal{A}}$, such that T is $\bar{\mathcal{A}}/\mathcal{A}$-measurable.

Proposition 3.4. *Let $T : \Omega \to \tilde{\Omega}$ be $\mathcal{A}/\tilde{\mathcal{A}}$-measurable and P be a probability measure on $(\Omega, \mathcal{A})$. Then*

$$\tilde{P}(\tilde{A}) := T(P)(\tilde{A}) := P\big(T^{-1}(\tilde{A})\big) =: P[T \in \tilde{A}], \quad \tilde{A} \in \tilde{\mathcal{A}},$$

defines a probability measure on $(\tilde{\Omega}, \tilde{\mathcal{A}})$. $\tilde{P}$ is called the induced measure on $(\tilde{\Omega}, \tilde{\mathcal{A}})$ *or the* distribution of T under P*.*

Notation: $\tilde{P} = P \circ T^{-1}$ or $\tilde{P} = T(P)$.

Proof. Clearly, $\tilde{P}(\tilde{A}) \geqslant 0$ for all $\tilde{A} \in \tilde{\mathcal{A}}$, $\tilde{P}(\emptyset) = 0$ and $\tilde{P}(\tilde{\Omega}) = 1$. For pairwise disjoint $\tilde{A}_i \in \tilde{\mathcal{A}}$, $i \in \mathbb{N}$, $T^{-1}(\tilde{A}_i)$ are pairwise disjoint too, hence

$$\tilde{P}\Big(\dot{\bigcup_{i\in\mathbb{N}}} \tilde{A}_i\Big) = P\Big(\underbrace{T^{-1}\Big(\dot{\bigcup_{i\in\mathbb{N}}} \tilde{A}_i\Big)}_{\dot{\bigcup}_{i\in\mathbb{N}} T^{-1}(\tilde{A}_i)}\Big) \overset{\substack{P \text{ is} \\ \sigma\text{-additive}}}{=} \sum_{i=1}^{\infty} P\big(T^{-1}(\tilde{A}_i)\big) = \sum_{i=1}^{\infty} \tilde{P}(\tilde{A}_i). \qquad \square$$

Remark 3.5. *Let $T(\Omega)$ be countable, so that $T(\Omega) = \{\tilde{\omega}_i \,|\, i \in \mathbb{N}\}$, then*

$$\tilde{P} = \sum_{i=1}^{\infty} P[T = \tilde{\omega}_i] \cdot \delta_{\tilde{\omega}_i}.$$

Proof. For any $\tilde{A} \in \tilde{\mathcal{A}}$ can be written as

$$\left\{T \in \tilde{A}\right\} = \dot{\bigcup_{\{i \in \mathbb{N} \,|\, T = \tilde{\omega}_i \in \tilde{A}\}}} \{\tilde{\omega}_i\}$$

so that

$$\tilde{P}(\tilde{A}) = P[T \in \tilde{A}] = \sum_{i=1}^{\infty} P[T = \tilde{\omega}_i] \cdot \underbrace{1_{\tilde{A}}(\tilde{\omega}_i)}_{=\delta_{\tilde{\omega}_i}(\tilde{A})} = \Big(\sum_{i=1}^{\infty} P[T = \tilde{\omega}_i] \cdot \delta_{\tilde{\omega}_i}\Big)(\tilde{A}). \qquad \square$$

Example 3.6. Infinitely many coin tosses: existence of a probability measure. Let $\Omega := [0,1]$ and $\mathcal{A}$ be the Borel σ-algebra on $[0,1]$. Let P be the restriction of the Lebesgue measure on $[0,1]$. Let

$$\tilde{\Omega} := \left\{\tilde{\omega} = (x_n)_{n \in \mathbb{N}} \;\middle|\; x_i \in \{0,1\} \;\forall\, i \in \mathbb{N}\right\} = \{0,1\}^{\mathbb{N}}.$$

Define $X_i : \tilde{\Omega} \to \{0,1\}$ by

$$X_i\big((x_n)_{n \in \mathbb{N}}\big) := x_i, \quad i \in \mathbb{N},$$

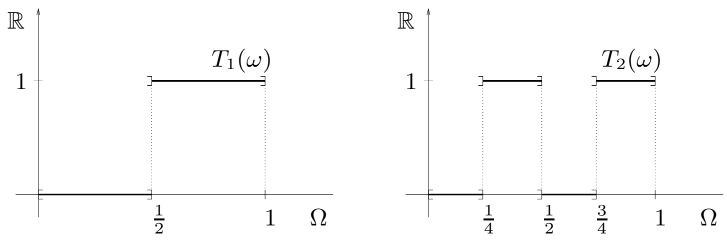
and let

$$\tilde{\mathcal{A}} := \sigma\big(\big\{\{X_i = 1\} \;\big|\; i \in \mathbb{N}\big\}\big).$$

Note that $\tilde{\mathcal{A}} = \sigma(\mathcal{A}_0)$, where $\mathcal{A}_0$ is the algebra of cylindrical subsets of Example 1.7 (iii). The binary expansion of some $\omega \in [0,1]$ defines a mapping

$$\begin{aligned} T : \Omega &\to \tilde{\Omega} \\ \omega &\mapsto T(\omega) = (T_1\omega, T_2\omega, \dots), \end{aligned}$$

with



(and similar for T_3, T_4, ...). Note that $T_i = X_i \circ T$ for all $i \in \mathbb{N}$. T is $\mathcal{A}/\tilde{\mathcal{A}}$-measurable, since

$$T^{-1}\big(\{X_i = 1\}\big) = \{T_i = 1\} = \text{finite union of intervals} \in \mathcal{A}\,.$$

Define $\tilde{P} := P \circ T^{-1}$. For fixed $(x_1, \dots, x_n) \in \{0,1\}^n$ we now obtain

$$\begin{aligned} \tilde{P}[X_1 = x_1, \dots, X_n = x_n] &= \tilde{P}\Big(\bigcap_{i=1}^{n} \{X_i = x_i\}\Big) \\ &= P[\text{interval of length } 2^{-n}] = 2^{-n}\,. \end{aligned}$$

Hence, for any fixed n, $\tilde{P}$ coincides with the probability measure for n coin tosses (= uniform distribution on binary sequences of length n). We have thus shown the existence of a probability measure $\tilde{P}$ on $(\tilde{\Omega}, \tilde{\mathcal{A}})$ and solved part of the problem of 1.7.

Uniqueness of $\tilde{P}$ later!

4 Random variables

Let $(\Omega, \mathcal{A})$ be a measurable space.

Definition 4.1. A *random variable*(on $(\Omega, \mathcal{A})$) is a ($\mathcal{A}$-) measurable map $X : \Omega \to \mathbb{R}$ resp. $X : \Omega \to \bar{\mathbb{R}}$ (with $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$), where $\mathbb{R}$, resp. $\bar{\mathbb{R}}$, is endowed with the Borel σ-algebra.

(Note: $\mathcal{B}(\bar{\mathbb{R}}) = \big\{ B \subset \bar{\mathbb{R}} \bigm| B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R}) \big\}$)

Remark 4.2. *(i) $X : \Omega \to \bar{\mathbb{R}}$ is a random variable if for all $c \in \mathbb{R}$ $\{X \leqslant c\} \in \mathcal{A}$.*

(ii) If $\mathcal{A} = \mathcal{P}(\Omega)$, then every function from Ω to $\bar{\mathbb{R}}$ is a random variable on $(\Omega, \mathcal{A})$.

(iii) Let X be a random variable on $(\Omega, \mathcal{A})$ and $h : \mathbb{R} \to \mathbb{R}$ (resp. $\bar{\mathbb{R}} \to \bar{\mathbb{R}}$) be a measurable mapping. Then $h(X)$ is a random variable too.

Examples: $|X|$, X^2, $|X|^p$, e^X, ...

(iv) The family of random variables on $(\Omega, \mathcal{A})$ is closed under countable operations.

In particular, if $X_1, X_2, \dots$ are random variables, then

- $\sum_i \alpha_i \cdot X_i$
- $\sup_i X_i$, $\inf_i X_i$
- $\limsup_{i\to\infty} X_i$, $\liminf_{i\to\infty} X_i$

are random variables too.

Proof. (i) Obvious, since $\sigma(\{]-\infty, c] : c \in \mathbb{R}\}) = \mathcal{B}(\bar{\mathbb{R}})$. It suffices to assume $\{X \leq c\} \in \mathcal{A}$ for all $c \in \mathbb{Q}$ or any other dense subset of $\mathbb{R}$ (Exercise!).

(ii) and (iii) Obvious.

(iv) for example:

- supremum

$$\big\{(\sup_i X_i) \leqslant c\big\} = \bigcap_{i\in\mathbb{N}} \underbrace{\{X_i \leqslant c\}}_{\in\mathcal{A}} \in \mathcal{A}.$$

- sum

$$\big\{X + Y < c\big\} = \bigcup_{\substack{r,s\in\mathbb{Q} \\ r+s<c}} \underbrace{\{X < r\} \cap \{Y < s\}}_{\in\mathcal{A}} \in \mathcal{A}$$

□

Important examples

Example 4.3. (i) **Indicator functions** of an event $A \in \mathcal{A}$:

$$\omega \mapsto 1_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases} \qquad \text{alternative notation } \mathbb{I}_A$$

is a random variable, because

$$\{1_A \leqslant c\} = \begin{cases} \emptyset & \text{if } c < 0 \\ A^c & \text{if } 0 \leqslant c < 1 \\ \Omega & \text{if } c \geqslant 1. \end{cases}$$

(ii) **simple random variables**

$$X = \sum_{i=1}^{n} c_i \cdot 1_{A_i}, \quad c_i \in \mathbb{R},\ A_i \in \mathcal{A},$$

Note: any finite-valued random variable is simple, because $X(\Omega) = \{c_1, \ldots, c_n\}$ implies

$$X = \sum_i c_i 1_{A_i} \quad \text{for} \quad X^{-1}\big(\{c_i\}\big) =: A_i.$$

Proposition 4.4 (Structure of random variables on $(\Omega, \mathcal{A})$)**.** *Let X be a random variable on $(\Omega, \mathcal{A})$. Then:*

(i) $X = X^+ - X^-$, with

$$X^+ := \max(X, 0), \quad X^- := -\min(X, 0) \quad \textit{(random variables!)}.$$

(ii) Let $X \geqslant 0$. Then there exists a sequence of simple random variables X_n, $n \in \mathbb{N}$, with $X_n \leqslant X_{n+1}$ and $X = \lim_{n\to\infty} X_n$.

Proof. (of (ii))

$$X_n := \sum_{i=0}^{n2^n-1} \frac{i}{2^n} 1_{\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\}} + n 1_{\{X \geq n\}}$$

□

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 4.5. Let X be a random variable on $(\Omega, \mathcal{A})$ with

$$\min\left(\int X^+ \, \mathrm{d}P, \int X^- \, \mathrm{d}P\right) < \infty. \tag{1.3}$$

Then

$$\mathbb{E}[X] := \int X \, \mathrm{d}P \quad \left(= \int_\Omega X \, \mathrm{d}P\right)$$

is called the *expectation* of X (w.r.t. P).

Definition/Construction

of the integral w.r.t. P

Let X be a random variable.

1. If $X = 1_A$, $A \in \mathcal{A}$, define
$$\int X \,\mathrm{d}P := P(A)\,.$$

2. If $X = \sum_{i=1}^n c_i \cdot 1_{A_i}$, $c_i \in \mathbb{R}$, $A_i \in \mathcal{A}$, define
$$\int X \,\mathrm{d}P := \sum_{i=1}^n c_i \cdot P(A_i)$$
(independent of the particular representation of X!)

3. $X \geq 0$, then there exist X_n simple, $X_n \geq 0$ (see 4.4) with $X_n \nearrow X$. Define
$$\int X \,\mathrm{d}P := \lim_{n\to\infty} \int X_n \,\mathrm{d}P \quad \big(\in [0,\infty]\big).$$
(independent of the particular choice for X_n!)

4. for general X, decompose $X = X^+ - X^-$ and define
$$\big(\mathbb{E}[X] =\big) \quad \int X \,\mathrm{d}P := \int X^+ \,\mathrm{d}P - \int X^- \,\mathrm{d}P.$$
(well-defined, if (1.3) satisfied.)

Definition 4.6. The *set of all P-integrable random variables* is defined by
$$\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{A}, P) := \big\{X \text{ r.v.} \bigm| \mathbb{E}\big[|X|\big] < \infty\big\}.$$

In the following let us introduce the following notion: A property E of points $\omega \in \Omega$ holds P-almost surely (P-a.s.), if there exists a measurable null-set N, i.e. a set $N \in \mathcal{A}$ with $P[N] = 0$, such that every $\omega \in \Omega \setminus N$ has property E.

If
$$\mathcal{N} := \{X \text{ r.v.} \mid X = 0 \ P\text{-a.s.}\}$$
then the quotient space
$$L^1 := {}^{\mathcal{L}^1}\!\big/_{\mathcal{N}}$$
is a Banach space w.r.t. the norm $\mathbb{E}\big[|X|\big]$.

Remark 4.7. *Special case: X random variable, $X \geq 0$, $X(\Omega)$ countable. Then*

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{x\in X(\Omega)} x \cdot 1_{\{X=x\}}\Big] \overset{\substack{\text{“3.” and}\\ \text{“2.” above}}}{=} \sum_{x\in X(\Omega)} x \cdot P[X=x]. \tag{1.4}$$

Similarly for X not necessarily finite, but $\mathbb{E}[X]$ well-defined:

$$\mathbb{E}[X] = \sum_{\substack{x\in X(\Omega),\\ x\geqslant 0}} x \cdot P[X=x] - \sum_{\substack{x\in X(\Omega),\\ x<0}} (-x)\cdot P[X=x].$$

If, in addition, Ω is countable, and $X \geqslant 0$, then

$$X = \sum_{\omega\in\Omega} X(\omega)\cdot 1_{\{\omega\}}\,, \quad \textit{and}$$

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\cdot \mathbb{E}[1_{\{\omega\}}] = \sum_{\omega\in\Omega} X(\omega)\cdot \underbrace{P(\{\omega\})}_{=:p(\omega)} = \sum_{\omega\in\Omega} p(\omega)\cdot X(\omega).$$

Example 4.8. Infinitely many coin tosses with a fair coin Let $\Omega = \{0,1\}^{\mathbb{N}}$. $\mathcal{A}$ and P as in 3.6

(i) **Expectation of the i^{th} coin toss** $X_i\big((x_n)_{n\in\mathbb{N}}\big) := x_i$

$$\mathbb{E}[X_i] \overset{(1.4)}{=} 1\cdot P[X_i = 1] + 0\cdot P[X_i = 0] = \frac{1}{2}.$$

(ii) **Expectation of number of “successes”**
$S_n := X_1 + \cdots + X_n$ = number of “successes” (= Ones) in n tosses
Then

$$P[S_n = k] = \sum_{\substack{(x_1,\dots,x_n)\in\{0,1\}^n\\ \text{mit}\\ x_1+\cdots+x_n=k}} P[X_1 = x_1,\dots,X_n = x_n] = \binom{n}{k}\cdot 2^{-n}\,, k = 0,1,\dots,n$$

Hence

$$\mathbb{E}[S_n] \overset{(1.4)}{=} \sum_{k=0}^{n} k\cdot P[S_n = k] = \sum_{k=1}^{n} k\cdot \binom{n}{k}\cdot 2^{-n} = \frac{n}{2}.$$

Easier: Once we have noticed that $\mathbb{E}[\,\cdot\,]$ is linear (see next proposition):

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] \overset{\text{(i)}}{=} \frac{n}{2}.$$

(iii) **Waiting time until first success** Let

$$T(\omega) := \min\{n \in \mathbb{N} \mid X_n(\omega) = 1\}$$
$$= \text{ waiting time until first success.}$$

Then

$$P[T = k] = P[X_1 = \cdots = X_{k=1} = 0,\ X_k = 1] = 2^{-k},$$

so that

$$\mathbb{E}[T] \overset{(1.4)}{=} \sum_{k=1}^{\infty} k \cdot P[T = k] = \sum_{k=1}^{\infty} k \cdot 2^{-k} = \frac{1}{2} \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^{k-1} = 2.$$

(Recall: $\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right) = \frac{\mathrm{d}}{\mathrm{d}q} \sum_{k=1}^{\infty} q^k = \sum_{k=1}^{\infty} kq^{k-1}$)

Remark 4.9. $X = Y$ P-a.s., *i.e.* $P[X = Y] = 1$, *implies* $\mathbb{E}[X] = \mathbb{E}[Y]$.

Proposition 4.10. $X \mapsto \mathbb{E}[X]$ *is a* positive linear functional *on* $\mathcal{L}^1$, *i.e.*

(i) $X \geqslant 0$ *P-a.s. implies* $\mathbb{E}[X] \geqslant 0$.

(ii) $\mathbb{E}\Big[\sum_{i=1}^{n} c_i \cdot X_i\Big] = \sum_{i=1}^{n} c_i \cdot \mathbb{E}[X_i]$.

((i) and (ii) imply monotonicity of $\mathbb{E}[\,\cdot\,]$: $X \leqslant Y \ \Rightarrow\ \mathbb{E}[X] \leqslant \mathbb{E}[Y]$.*)*

Proof. See text books on measure theory. □

In addition $X \mapsto \mathbb{E}[X]$ is continuous w.r.t. monotone increasing sequences, i.e. the following proposition holds:

Proposition 4.11 (monotone integration, B. Levi)**.** *Let* X_n *random variables with* $0 \leqslant X_1 \leqslant X_2 \leqslant \ldots$ *. Then:*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\big[\lim_{n\to\infty} X_n\big].$$

Proof. See text books on measure theory. □

Corollary 4.12. *Let* $X_n \geqslant 0$, $n \in \mathbb{N}$, *be random variables. Then*

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

Lemma 4.13 (Fatou's lemma)**.** *Let* $X_n \geq 0$, $n \in \mathbb{N}$, *be random variables (or more general* $X_n \geqslant Y \in \mathcal{L}^1$*). Then*

$$\mathbb{E}\big[\liminf_{n\to\infty} X_n\big] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Proof.

$$\begin{aligned}\mathbb{E}\big[\liminf_{n\to\infty} X_n\big] &= \mathbb{E}\Big[\lim_{n\to\infty}\big(\inf_{k\geqslant n} X_k\big)\Big] \overset{\text{Levi}}{=} \lim_{n\to\infty} \mathbb{E}\big[\inf_{k\geqslant n} X_k\big] \\ &\leqslant \lim_{n\to\infty}\inf_{k\geqslant n}\mathbb{E}[X_k] = \liminf_{n\to\infty}\mathbb{E}[X_n].\end{aligned}$$

□

Proposition 4.14 (Lebesgue's dominated convergence)**.** *Let X_n, $n \in \mathbb{N}$ be random variables and $Y \in \mathcal{L}^1$ with $|X_n| \leqslant Y$ P-a.s. Suppose that the pointwise limit $\lim_{n\to\infty} X_n$ exists P-a.s., then*

$$\mathbb{E}\big[\lim_{n\to\infty} X_n\big] = \lim_{n\to\infty}\mathbb{E}[X_n].$$

Proof.

$$\begin{aligned}\underbrace{\mathbb{E}\big[Y - \limsup_{n\to\infty} X_n\big]}_{\geqslant 0} &= \mathbb{E}\big[\liminf_{n\to\infty}\underbrace{(Y - X_n)}_{\geqslant 0}\big] \overset{\text{Fatou}}{\leqslant} \liminf_{n\to\infty}\mathbb{E}[Y - X_n] \\ &= \mathbb{E}[Y] - \limsup_{n\to\infty}\mathbb{E}[X_n].\end{aligned}$$

Next, $\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n = \lim_{n\to\infty} X_n$ P-a.s. implies

$$\begin{aligned}\mathbb{E}\big[\lim_{n\to\infty} X_n\big] &= \mathbb{E}\big[\liminf_{n\to\infty} X_n\big] \overset{\text{Fatou}}{\leqslant} \liminf_{n\to\infty}\mathbb{E}[X_n] \leqslant \limsup_{n\to\infty}\mathbb{E}[X_n] \\ &\leqslant \mathbb{E}\big[\limsup_{n\to\infty} X_n\big] = \mathbb{E}\big[\lim_{n\to\infty} X_n\big]\end{aligned}$$

□

Example 4.15. Tossing a fair coin Consider the following simple game: A fair coin is thrown and the player can invest an arbitrary amount of Euros on either "head" or "tail". If the right side shows up, the player gets twice his investment back, otherwise nothing.

Suppose now a player plays the following bold strategy: he doubles his investment until his first success. Assuming the initial investment was 1 Euro, the investment in the n^{th} round is given by

$$X_n = 2^{n-1}\cdot 1_{\{T>n-1\}},$$

where T = waiting time until the first "'1"'. Then

$$\mathbb{E}[X_n] = 2^{n-1}\cdot\underbrace{P[T > n-1]}_{=(\frac{1}{2})^{n-1}} = 1.$$

whereas on the other hand $\lim_{n\to\infty} X_n = 0$ P-a.s. (more precisely: for all $\omega \neq (0,0,0,\dots)$).

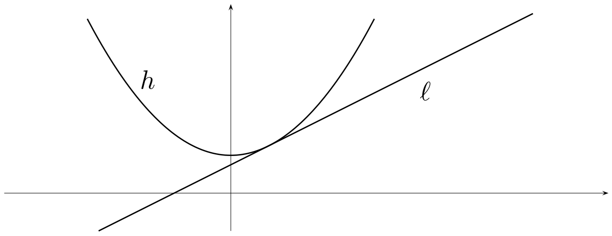
5 Inequalities

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Proposition 5.1 (Jensens' inequality)**.** *Let h be a convex function defined on some interval $I \subseteq \mathbb{R}$, X in $\mathcal{L}^1$ with $X(\Omega) \subset I$. Then $\mathbb{E}(X) \in I$ and*

$$h\big(\mathbb{E}[X]\big) \leqslant \mathbb{E}\big[h(X)\big] .$$

Proof. W.l.o.g. we may asssume that X not P-a.s. equal to a constant function. Then $x_0 := \mathbb{E}[X] \in \mathring{I}$. Since h is convex, there exists an affine linear function ℓ with $\ell(x_0) = h(x_0)$ and $\ell \leqslant h$ ("support tangent line").



Consequently,

$$h\big(\mathbb{E}[X]\big) = \ell\big(\mathbb{E}[X]\big) \overset{\text{linearity}}{=} \mathbb{E}\big[\ell(X)\big] \overset{\text{monotonicity}}{\leqslant} \mathbb{E}\big[h(X)\big] . \qquad \square$$

Example 5.2.

$$\mathbb{E}[X]^2 \leqslant \mathbb{E}[X^2].$$

Moreover, for $0 < p \leqslant q$:

$$\underbrace{\mathbb{E}\big[|X|^p\big]^{\frac{1}{p}}}_{=:\|X\|_p} \leqslant \underbrace{\mathbb{E}\big[|X|^q\big]^{\frac{1}{q}}}_{=:\|X\|_q} .$$

Proof. $h(x) := |x|^{\frac{q}{p}}$ is convex. Since $\big(|X| \wedge n\big)^p \in \mathcal{L}^1$ for $n \in \mathbb{N}$, we obtain that

$$\Big(\mathbb{E}\Big[\big(|X| \wedge n\big)^p\Big]\Big)^{\frac{q}{p}} \leqslant \mathbb{E}\Big[\big(|X| \wedge n\big)^q\Big] ,$$

which implies the assertion taking the limit $n \to \infty$. $\square$

Definition 5.3. For $1 \leqslant p < \infty$ let

$$\mathcal{L}^p := \big\{X \bigm| X \text{ r.v. and } \mathbb{E}\big[|X|^p\big] < \infty\big\}.$$

$\mathcal{L}^p$ is called the *set of p-integrable random variables.*

Remark 5.4. *(i) If $1 \leqslant p \leqslant q$ then $\mathcal{L}^q \subset \mathcal{L}^p$.*

(ii) *Let $\mathcal{N} = \{X \mid X$ r.v. and $X = 0$ P-a.s.$\}$, and $p \geqslant 1$. Then $\mathcal{N} \subset \mathcal{L}^p$ is a linear subspace and the quotient space*

$$L^p := {}^{\mathcal{L}^p}\!\big/_{\mathcal{N}}$$

is a Banach space w.r.t. $\|\cdot\|_p$ (i.e. a complete normed vector space)

Proposition 5.5. *Let X be a random variable, $h \geqslant 0$ monotone increasing. Then*

$$h(c)\, P[X \geqslant c] \leqslant \mathbb{E}\big[h(X)\big] \qquad \forall\, c > 0.$$

Proof.

$$\begin{aligned} h(c)\, P[X \geqslant c] &\leqslant h(c)\, P\big[h(X) \geqslant h(c)\big] = \mathbb{E}\big[h(c)\, 1_{\{h(X) \geqslant h(c)\}}\big] \\ &\leqslant \mathbb{E}\big[h(X)\big]. \end{aligned}$$

□

Corollary 5.6. (i) **Markov inequality** *$h(x) = |x|$ (increasing on $\bar{\mathbb{R}}_+$). Then*

$$P\big[|X| \geqslant c\big] \leqslant \frac{1}{c}\, \mathbb{E}\big[|X|\big] \qquad \forall\, c > 0\,.$$

In particular,

$$\begin{aligned} \mathbb{E}\big[|X|\big] = 0 \quad &\Rightarrow \quad |X| = 0 \;\; P\text{-a.s.} \\ \mathbb{E}\big[|X|\big] < \infty \quad &\Rightarrow \quad |X| < \infty \;\; P\text{-a.s.} \end{aligned}$$

(ii) **Chebychev's inequality** *$h(x) = x^2$ and $X \in \mathcal{L}^2$ implies*

$$P\Big[\big|X - \mathbb{E}[X]\big| \geqslant c\Big] \leqslant \frac{1}{c^2}\, \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] \quad \Big(= \frac{\mathrm{var}(X)}{c^2}\Big).$$

6 Variance and Covariance

Let $(\Omega, \mathcal{A}, P)$ be a probability space.
$\mathbb{E}[X]$ = "average value" of $X(\omega)$, $\omega \in \Omega$ (forecast value)

Remark 6.1. *Let P be the uniform distribution on $\Omega = \{\omega_1, \dots, \omega_n\}$, then*

$$\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^{n} X(\omega_i) = \text{ arithmetic mean of} X(\omega_1), \dots, X(\omega_n)\,.$$

Definition 6.2. Let $X \in \mathcal{L}^1$. Then

$$\mathrm{var}(X) := \sigma^2(X) := \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] \quad \big(\in [0, \infty]\big)$$

is called the *variance* of X (mean square forecast error)

The variance is a measure for fluctuations of X around $\mathbb{E}[X]$, resp. a measure for "dispersion" or for "risk".

$\sigma(X) := \sqrt{\mathrm{var}(X)}$ is called *standard deviation*.

Remark 6.3. *(i)*

$$\mathrm{var}(X) = \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2.$$

(ii) $\mathrm{var}(X) = 0 \quad \Leftrightarrow \quad P\big[X = \mathbb{E}[X]\big] = 1$
i.e. X *behaves* deterministically

(iii) $\mathrm{var}(X) < \infty \quad \Leftrightarrow \quad X \in \mathcal{L}^2.$

Definition 6.4. Let X, $Y \in \mathcal{L}^2$. Then

$$\mathrm{cov}(X,Y) := \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\Big] = \mathbb{E}[XY] - \mathbb{E}[X]\cdot\mathbb{E}[Y] \tag{1.5}$$

is called the *covariance* of X and Y.

$$\varrho(X,Y) := \frac{\mathrm{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

is caled the *correlation* of X and Y.

Remark 6.5 (properties of the covariance)**.** *(i) Let $X \in \mathcal{L}^1$. Then*

$$\mathrm{var}(aX + b) = a^2 \cdot \mathrm{var}(X).$$

(ii) Let $X, Y \in \mathcal{L}^2$. Then

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y).$$

Definition 6.6. Two random variables $X, Y \in \mathcal{L}^2$ are called *uncorrelated*, if

$$\mathrm{cov}(X,Y) = 0 \quad \big(\Leftrightarrow \quad \mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)\big). \tag{1.6}$$

Proposition 6.7 (Cauchy-Schwarz)**.** *Let X and $Y \in \mathcal{L}^2$. Then*

$$X \cdot Y \in \mathcal{L}^1 \quad \textit{and} \quad \big|\mathrm{cov}(X,Y)\big| \leqslant \sigma(X)\cdot\sigma(Y)$$

(resp. $\varrho(X,Y) \in [-1,1]$).

Proof. Let $X, Y \in \mathcal{L}^2$. Then $X + Y \in \mathcal{L}^2$, hence

$$2 \cdot XY = (X+Y)^2 - X^2 - Y^2 \quad \in \mathcal{L}^1.$$

Proof of the inequality: W.l.o.g. let $\mathrm{var}(X) > 0$ and $\mathrm{var}(Y) > 0$
(otherwise $X = \mathbb{E}[X]$ P-a.s. or $Y = \mathbb{E}[Y]$ P-a.s., so that
$\mathrm{cov}(X,Y) = 0$ and die inequality is trivial)

Let $a := \frac{\sigma(X)}{\sigma(Y)}$. Then

$$\begin{aligned}
0 &\leqslant \mathbb{E}\left[\Big(X - \mathbb{E}[X] - a\big(Y - \mathbb{E}[Y]\big)\Big)^2\right] \\
&= \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] - 2a \cdot \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\Big] \\
&\quad + a^2 \cdot \mathbb{E}\Big[\big(Y - \mathbb{E}[Y]\big)^2\Big] \\
&= \sigma^2(X) - 2a \cdot \operatorname{cov}(X,Y) + a^2 \cdot \sigma^2(Y) \\
&= 2\sigma^2(X) - 2 \cdot \frac{\sigma(X)}{\sigma(Y)} \cdot \operatorname{cov}(X,Y).
\end{aligned}$$

□

Example 6.8. Tossing a coin with probability $p \in [0,1]$ for success

$$\begin{aligned}
&\Omega = \big\{\omega = (x_1, x_2, \dots) \;\big|\; x_i \in \{0,1\}\big\} = \{0,1\}^{\mathbb{N}}, \\
&X_i : \Omega \to \{0,1\} \quad \text{mit} \quad X_i\big((x_n)_{n\in\mathbb{N}}\big) = x_i, \quad i \in \mathbb{N}, \\
&\mathcal{A} = \sigma\big(\big\{\{X_i = 1\} \;\big|\; i \in \mathbb{N}\big\}\big).
\end{aligned}$$

Then there exists a unique probability measure $P = P_p$ on $(\Omega, \mathcal{A})$ with

$$P_p[X_{i_1} = x_1, \dots, X_{i_n} = x_n] = p^{\sum\limits_{i=1}^n x_i} \cdot (1-p)^{n - \sum\limits_{i=1}^n x_i}.$$

(existence for $p = \frac{1}{2}$ in example 3.6, existence for general $p \neq \frac{1}{2}$ later, uniqueness later).

Then $P[X_i = 1] = p$ and $P[X_i = 1,\, X_j = 1] = p^2$ for all $i \neq j$. Consequently,

$$\mathbb{E}[X_i] = p \quad \text{and} \quad \operatorname{var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = p - p^2 = p(1-p)$$

and for $i \neq j$

$$\operatorname{cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - p^2 = 0$$

so that $X_1, X_2, \dots$ are pairwise uncorrelated (in fact even independent, see below)

Let $S_n := X_1 + \dots + X_n$ be the number of successes. Then

$$\mathbb{E}[S_n] = np \quad \text{and} \quad \operatorname{var}(S_n) = np(1-p).$$

If $X := \sum_{n=1}^\infty 2^{-n} X_n$ then

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{n=1}^\infty 2^{-n} X_n\Big] \overset{\text{Levi}}{=} \sum_{n=1}^\infty \mathbb{E}[2^{-n} X_n] = \sum_{n=1}^\infty 2^{-n} p = p$$

and using Levi and the fact that $X_1, X_2, \dots$ are pairwise uncorrelated, we conclude that

$$\operatorname{var}(X) = \sum_{n=1}^\infty 2^{-2n} \cdot p(1-p) = \frac{1}{3} \cdot p(1-p).$$

Finally, let T be the waiting time until the first "'success"'. Then

$$\begin{aligned} P_p[T=n] &= P_p\big[X_1=\cdots=X_{n-1}=0,\ X_n=1\big] \\ &= (1-p)^{n-1}p \qquad \text{(geometric distribution)} \end{aligned}$$

then

$$\begin{aligned} \mathbb{E}[T] &= \mathbb{E}\Big[\sum_{n=1}^{\infty} n\cdot 1_{\{T=n\}}\Big] = \sum_{n=1}^{\infty} n\cdot P_p[T=n] \\ &= \underbrace{\sum_{n=1}^{\infty} n\cdot(1-p)^{n-1}p}_{\substack{\text{"derivative of the}\\ \text{geometr. series"}}} = \frac{1}{p}, \end{aligned}$$

and analogously

$$\operatorname{var}(T) = \cdots = \frac{1-p}{p^2}.$$

7 The (strong and the weak) law of large numbers

Let

- $(\Omega, \mathcal{A}, P)$ be a probability space
- $X_1, X_2, \ldots \in \mathcal{L}^2$ r.v. with
  - X_i uncorrelated, i.e. $\operatorname{cov}(X_i, X_j) = 0$ for $i \neq j$
  - bounded variances, i.e. $\sup_{i\in\mathbb{N}} \underbrace{\operatorname{var}(X_i)}_{\substack{=\sigma^2(X_i)\\ =:\sigma_i^2}} < \infty$.

Let

$$S_n := X_1 + \cdots + X_n$$

so that $\frac{S_n(\omega)}{n}$ is the arithmetic mean of the first n observations $X_1(\omega), \ldots, X_n(\omega)$ ("empirical mean")

Our aim in this section is to show that randomness in the empirical mean vanishes for increasing n, i.e.

$$\frac{S_n(\omega)}{n} \overset{n \text{ large}}{\sim} \frac{\mathbb{E}[S_n]}{n},$$

resp.

$$\frac{S_n(\omega)}{n} \overset{n \text{ large}}{\sim} m \qquad \text{if } \mathbb{E}[X_i] = m\,.$$

Remark 7.1. *W.l.o.g. we may assume that* $\mathbb{E}[X_i] = 0$ *for all* i*, because:*

- $\tilde{X}_i := X_i - \mathbb{E}[X_i]$ *"centered"*
- $\operatorname{cov}(\tilde{X}_i, \tilde{X}_j) = \operatorname{cov}(X_i, X_j) = 0$ *for* $i \neq j$
- $\operatorname{var}(\tilde{X}_i) = \operatorname{var}(X_j)$.

Proposition 7.2.

$$\lim_{n\to\infty} \mathbb{E}\Big[\Big(\frac{S_n}{n} - \frac{\mathbb{E}[S_n]}{n}\Big)^2\Big] = 0\,.$$

$$(\textit{resp.} \qquad \lim_{n\to\infty} \mathbb{E}\Big[\Big(\frac{S_n}{n} - m\Big)^2\Big] = 0 \qquad \textit{if } E[X_i] = m)$$

Proof.

$$\begin{aligned}\mathbb{E}\Big[\Big(\frac{S_n}{n} - \frac{\mathbb{E}[S_n]}{n}\Big)^2\Big] &= \operatorname{var}\Big(\frac{S_n}{n}\Big) = \frac{1}{n^2} \cdot \operatorname{var}(S_n) \\ &\overset{\text{Bienaymé}}{=} \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \leqslant \frac{1}{n} \cdot \text{const.} \xrightarrow{n\to\infty} 0.\end{aligned}$$

□

Remark 7.3. *mere functional analytic fact: in the Hilbert space* $L^2 = {}^{\mathcal{L}^2}\!/_{\sim}$ *(with the scalar product* $(X,Y) = \mathbb{E}[X \cdot Y]$ *and norm* $\|\cdot\| = (\,\cdot\,,\,\cdot\,)^{\frac{1}{2}}$*), the arithmetic mean of bounded, orthogonal vectors converges to zero:*

$$\Big\|\frac{S_n}{n}\Big\|^2 = \frac{1}{n^2} \cdot \|S_n\|^2 = \frac{1}{n^2}\sum_{i=1}^n \|X_i\|^2 \xrightarrow{n\to\infty} 0.$$

Chebychev's inequality immediately implies the following

Proposition 7.4 ("Weak law of large numbers")**.** *Let* $X_1, X_2, \ldots \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ *uncorrelated r.v. with bounded variances and* $\mathbb{E}[X_i] = m \;\forall\, i$*. Then for all* $\varepsilon > 0$*:*

$$\lim_{n\to\infty} P\Big[\Big|\frac{S_n}{n} - m\Big| \geqslant \varepsilon\Big] = 0\,.$$

"convergence in probability of $\frac{S_n}{n}$ *towards* m*"*

Proof. Chebychev's inequality implies

$$P\Big[\Big|\frac{S_n}{n} - m\Big| \geqslant \varepsilon\Big] \leqslant \frac{1}{\varepsilon^2} \cdot \operatorname{var}\Big(\frac{S_n}{n}\Big) \to 0 \text{ if } n \to \infty.$$

□

Example 7.5. Bernoulli experiments with parameter $p \in [0,1]$ Let $X_i(\omega) = x_i$ and $P_p[X_i = 1] = p$, hence $\mathbb{E}_p[X_i] = p$ and $\operatorname{var}(X_i) = p(1-p)$ $(\leqslant \frac{1}{4})$

Then

$$P_p\Big[\Big|\underbrace{\frac{S_n}{n}}_{\substack{\text{rel. freq.}\\ \text{of "1"}}} - p\Big| \geqslant \varepsilon\Big] \xrightarrow{n\to\infty} 0; \tag{1.7}$$

(J. Bernoulli: Ars Conjectandi)

Interpretation of the success probability p as relative frequency.

Problem: to infer p from (1.7) one uses a probabilistic statement w.r.t. a probability measure P_p that is defined with the help of p.

Example 7.6. Application: uniform approximation of $f \in C\big([0,1]\big)$ **with Bernstein polynomials**

Let

$$\begin{aligned} B_n(p) &:= \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big) \cdot \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big) \cdot P_p[S_n = k] \\ &= E_p\left[f\left(\frac{S_n}{n}\right)\right]. \end{aligned}$$

$\varepsilon > 0$: f uniformly continuous $\Rightarrow \exists \delta = \delta(\varepsilon) > 0$ such that

$$\sup_{\substack{x,y\in[0,1],\\ |x-y|\leqslant\delta}} \big|f(x) - f(y)\big| \leqslant \varepsilon.$$

Consequently,

$$\begin{aligned} \big|B_n(p) - f(p)\big| &= \left|\mathbb{E}_p\Big[f\Big(\frac{S_n}{n}\Big) - f(p)\Big]\right| \leqslant \mathbb{E}_p\Big[\Big|f\Big(\frac{S_n}{n}\Big) - f(p)\Big|\Big] \\ &= \mathbb{E}_p\Big[\Big|f\Big(\frac{S_n}{n}\Big) - f(p)\Big| \cdot 1_{\{|\frac{S_n}{n}-p|\leqslant\delta\}}\Big] + \mathbb{E}_p\Big[\Big|f\Big(\frac{S_n}{n}\Big) - f(p)\Big| \cdot 1_{\{|\frac{S_n}{n}-p|>\delta\}}\Big] \\ &\leqslant \varepsilon \cdot P_p\Big[\Big|\frac{S_n}{n} - p\Big| \leqslant \delta\Big] + 2\|f\|_\infty \underbrace{P_p\Big[\Big|\frac{S_n}{n} - p\Big| > \delta\Big]}_{\leq \frac{1}{\delta^2 n} p(1-p) \leq \frac{1}{4\delta^2 n}}. \end{aligned}$$

Consequently,

$$\limsup_{n\to\infty} \|B_n - f\|_\infty \leqslant \varepsilon \ \ \forall\, \varepsilon > 0, \quad \text{hence} \quad \lim_{n\to\infty} \|B_n - f\|_\infty = 0.$$

From convergence in probability to a.s.-convergence:

Lemma 7.7. *Let* $Z_1, Z_2, \dots$ *be r.v. on* $(\Omega, \mathcal{A}, P)$*. If for all* $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P\big[|Z_n| \geqslant \varepsilon\big] < \infty$$

(= "fast convergence in probability towards 0"), then

$$P\big(\{\omega \mid \lim_{n\to\infty} Z_n(\omega) = 0\}\big) = 1.$$

(= "almost sure convergence towards 0").

Proof. The lemma of Borel-Cantelli (Lemma 1.11) implies that for all $\varepsilon > 0$

$$P\Big(\limsup_{n\to\infty}\big\{|Z_n| > \varepsilon\big\}\Big) = 0.$$

It follows that for all $k \in \mathbb{N}$ there exists $N_k \in \mathcal{A}$ with $P\,[N_k] = 0$ such that

$$\limsup_{n\to\infty}\big|Z_n(\omega)\big| \leqslant \frac{1}{k} \qquad \forall\,\omega \in \Omega \setminus N_k\,.$$

Hence for all $\omega \notin N := \bigcup_{k=1}^{\infty} N_k$ (note that $P[N] = 0$!)

$$\lim_{n\to\infty}\big|Z_n(\omega)\big| = 0.$$

□

Proposition 7.8 ("Strong law of large numbers"). *Let $X_1, X_2, \ldots \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ be uncorrelated with $\sup_{i\in\mathbb{N}} \sigma^2(X_i) = c < \infty$. Then:*

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} - \frac{\mathbb{E}[S_n]}{n} = 0 \quad P\text{-a.s.}$$

(resp., if $\mathbb{E}[X_i] = m$: $\lim_{n\to\infty} \frac{S_n(\omega)}{n} = m$ P-a.s.).

Proof. Again w.l.o.g. we may assume that $\mathbb{E}[X_i] = 0$.

1. Step Fast convergence in probability towards 0 along the subsequence $n_k = k^2$
   For all $\varepsilon > 0$

$$P\Big[\Big|\frac{S_{k^2}}{k^2}\Big| > \varepsilon\Big] \overset{\text{Chebychev}}{\leqslant} \frac{1}{\varepsilon^2 k^4}\operatorname{var}(S_{k^2}) \leq \frac{c}{\varepsilon^2 k^2}\,.$$

   Consequently, Lemma 7.7 implies that

$$\lim_{k\to\infty} \frac{S_{k^2}(\omega)}{k^2} = 0 \qquad \forall\,\omega \notin N_1 \text{ with } P[N_1] = 0\,.$$

2. Step Let $D_k := \max_{k^2 \leqslant l < (k+1)^2} |S_l - S_{k^2}|$. We will show in the following fast convergence in probability of $\frac{D_k}{k^2}$ towards 0:

For all $\varepsilon > 0$:

$$P\left[\frac{D_k}{k^2} \geqslant \varepsilon\right] = P\left(\bigcup_{l=k^2}^{k^2+2k} \{|S_l - S_{k^2}| \geqslant \varepsilon k^2\}\right)$$

$$\leqslant \sum_{l=k^2+1}^{k^2+2k+1} \underbrace{P\left(\{|S_l - S_{k^2}| \geqslant \varepsilon k^2\}\right)}_{\overset{\text{Chebychev}}{\leqslant} \frac{1}{\varepsilon^2 k^4} \underbrace{(l-k^2)}_{\leqslant 2k+1} \cdot c} \leqslant \frac{(2k+1)(2k+1)\cdot c}{\varepsilon^2 k^4}$$

$$= \frac{9c}{\varepsilon^2 k^2}$$

Lemma 7.7 now implies that

$$\lim_{k\to\infty} \frac{D_k(\omega)}{k^2} = 0 \qquad \forall \omega \notin N_2 \text{ with } P[N_2] = 0.$$

3. Step For $n \in \mathbb{N}$ and $k = k(n) \in \mathbb{N}$ with $k^2 \leqslant n < (k+1)^2$ we obtain that

$$\frac{|S_n(\omega)|}{n} \leqslant \frac{|S_{k^2}(\omega)| + D_k(\omega)}{k^2} \xrightarrow{n\to\infty} 0 \qquad \forall\, \omega \notin N_1 \cup N_2 \qquad \square$$

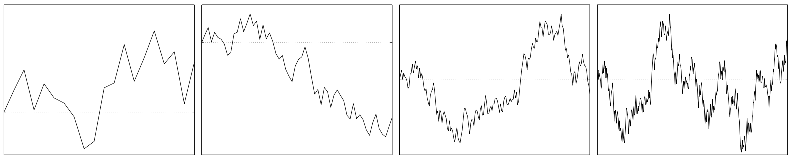
Example 7.9. Bernoulli experiments with $p \in [0,1]$

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \longrightarrow p \quad P_p\text{-a.s. (E. Borel 1909)}$$

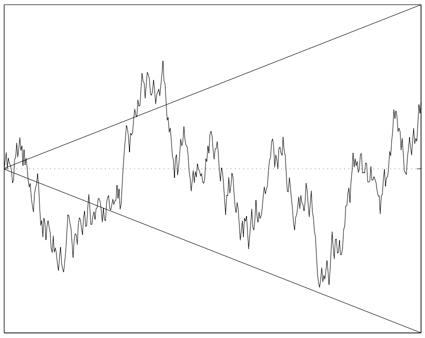
Consider the experiment of tossing a fair coin ($p = \frac{1}{2}$), $Y_i := 2X_i - 1$

$$S_n = Y_1 + \cdots + Y_n$$
$$= \text{position of a particle undergoing a "random walk" on } \mathbb{Z}$$

Increasing refinement of the random walk yields the *Brownian motion*:



The strong law of large numbers implies that $\frac{S_n(\omega)}{n} \to 0$ P-a.s.
In particular, fluctuations are growing smaller than linear.



A precise description of the order of fluctuations is provided by the law of the iterated logarithm:

$$\limsup_{n\to\infty} \frac{S_n(\omega)}{\sqrt{2n\log\log n}} = +1 \qquad P-a.s.$$
$$\liminf_{n\to\infty} \frac{S_n(\omega)}{\sqrt{2n\log\log n}} = -1 \qquad P-a.s.$$

8 Convergence and uniform integrability

Definition 8.1. Let $X, X_1, X_2, \ldots$ be r.v. on $(\Omega, \mathcal{A}, P)$.

(i) $\mathcal{L}^p$**-convergence** $(p \geqslant 1)$

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|^p\big] = 0$$

(alternative notation $\lim_{n\to\infty}\|X_n - X\|_p = 0$)

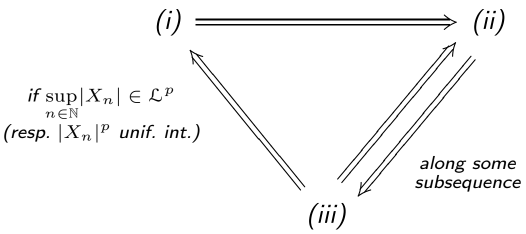
(ii) **Convergence in probability**

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} P\big[|X_n - X| \geqslant \varepsilon\big] = 0.$$

(iii) P**-a.s. convergence**

$$P\big[\lim_{n\to\infty} X_n = X\big] = 1.$$

Proposition 8.2 (Comparison of the three types of convergence)**.**



Proof. **(i)⇒(ii):** Chebychev's inequality implies:

$$P\big[|X_n - X| \geqslant \varepsilon\big] \leqslant \frac{\mathbb{E}\big[|X_n - X|^p\big]}{\varepsilon^p}.$$

(iii)⇒(ii):

$$\big\{\lim_{n\to\infty} X_n = X\big\} = \bigcap_{k=1}^{\infty} \underbrace{\bigcup_{m=1}^{\infty} \bigcap_{n\geqslant m} \Big\{|X_n - X| \leqslant \frac{1}{k}\Big\}}_{=:A_k}.$$

Then $P\big[\lim_{n\to\infty} X_n = X\big] = 1$ implies $P(A_k) = 1$ for all $k \in \mathbb{N}$. Continuity of P from above (cf. Proposition 1.9) implies that

$$\begin{aligned} 1 = P[A_k] &\overset{1.9}{=} \lim_{m\to\infty} P\left(\bigcap_{n\geqslant m} \Big\{|X_n - X| \leqslant \frac{1}{k}\Big\}\right) \\ &\leqslant \limsup_{m\to\infty} P\left[|X_m - X| \leqslant \frac{1}{k}\right] \leqslant 1 \end{aligned}$$

Consequently,

$$\lim_{m\to\infty} P\left[|X_m - X| > \frac{1}{k}\right] = 0.$$

(iii)⇒(i): $Y := \sup_{n\in\mathbb{N}}|X_n| \in \mathcal{L}^p$, $\lim_{n\to\infty} X_n = X$ P-a.s. implies $|X| \leqslant Y$
In particular, $|X_n - X|^p \leqslant 2^p Y^p \in \mathcal{L}^1$.
$\lim_{n\to\infty}|X_n - X|^p = 0$ P-a.s. now implies with Lebesgue's dominated convergence

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|^p\big] = 0.$$

(ii)⇒(iii): $\forall l \in \mathbb{N}$ we can find by iteration a subsequence $\exists\big(n_k(l)\big)_k$ of $\big(n_k(l-1)\big)_k$ such that

$$\sum_{k=1}^{\infty} P\left[\big|X_{n_k(l)} - X\big| \geqslant \frac{1}{l}\right] < \left(\frac{1}{2}\right)^l$$

Here, we let $n_k(0) = k$. Extracting the diagonal sequence $\big(n_k(k)\big)_k$ we conclude that for all $l \in \mathbb{N}$

$$\sum_{k=1}^{\infty} P\left[\big|X_{n_k(k)} - X\big| \geqslant \frac{1}{l}\right] \leq \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k < \infty\,.$$

Lemma 7.7 now implies

$$\lim_{k\to\infty} X_{n_k}(k) = X \quad P\text{-a.s.} \qquad \square$$

Remark 8.3. *The diagram can be complemented as follows:*

- *(ii)⇒(i) holds, if* $\sup_{n\in\mathbb{N}}|X_n| \in \mathcal{L}^p$ *(resp.* $|X_n|^p$ *uniformly integrable)(see Proposition 8.4 and Remark 8.8 below)*
- *in general (i)⇏(iii) and (iii)⇏(i) (hence (ii)⇏(i) too). For examples see Exercises.*

The next Proposition is the definitive version of Lebesgue's theorem on dominated convergence.

Proposition 8.4. *Let* $X_n \in \mathcal{L}^1$ *and* X *be r.v. Then the following statements are equivalent:*

(i) $\lim_{n\to\infty} X_n = X$ *in* $\mathcal{L}^1$*.*

(ii) $\lim_{n\to\infty} X_n = X$ *in probability and* $(X_n)_{n\in\mathbb{N}}$ *uniformly integrable.*

Corollary 8.5. $\lim_{n\to\infty} X_n = X$ P*-a.s. and* $(X_n)_{n\in\mathbb{N}}$ *uniformly integrable implies*

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

Definition 8.6. A family $(X_i)_{i\in I} \subset \mathcal{L}^1$ of r.v. is called *uniformly integrable* if

$$\lim_{c\to\infty} \sup_{i\in I} \underbrace{\int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P}_{=:\mathbb{E}[1_{\{|X_i|\geqslant c\}}\cdot|X_i|]} = 0.$$

Lemma 8.7 (ε-δ criterion)**.** *Let* $(X_i)_{i\in I} \subset \mathcal{L}^1$*. Then the following statements are equivalent:*

(i) $(X_i)_{i\in I}$ *is uniformly integrable.*

(ii) $\sup_{i\in I} \mathbb{E}\big[|X_i|\big] < \infty$ *and* $\forall\, \varepsilon > 0$ $\exists \delta > 0$ *such that*

$$P(A) < \delta \quad \Rightarrow \quad \int_A |X_i| \,\mathrm{d}P < \varepsilon \quad \forall\, i \in I.$$

Proof. **(i)⇒(ii):** $\exists c$ such that $\sup_{i\in I} \int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P \leqslant 1$ Consequently,

$$\begin{aligned}\sup_{i\in I} \int |X_i| \,\mathrm{d}P &= \sup_{i\in I} \left\{ \int_{\{|X_i|<c\}} |X_i| \,\mathrm{d}P + \int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P \right\} \\ &\leqslant c + 1 < \infty.\end{aligned}$$

Let $\varepsilon > 0$. Then there exists $c \geqslant 0$ such that

$$\sup_{i\in I} \int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P < \frac{\varepsilon}{2}.$$

For $\delta := \frac{\varepsilon}{2c}$ and $A \in \mathcal{A}$ mit $P[A] < \delta$ we now conclude

$$\int_A |X_i| \,\mathrm{d}P = \int_{A\cap\{|X_i|<c\}} |X_i| \,\mathrm{d}P + \int_{A\cap\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P$$
$$\leqslant c \int_A \,\mathrm{d}P + \int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P < c \cdot P[A] + \frac{\varepsilon}{2} < \varepsilon.$$

(ii)⇒(i): Let $\varepsilon > 0$ and δ be as in (ii). Let c be so large such that Markov's inequality implies

$$\frac{1}{c} \cdot \sup_{i\in I} \mathbb{E}\big[|X_i|\big] < \delta \quad \Rightarrow \quad P\big[|X_i| \geqslant c\big] \leqslant \frac{1}{c} \cdot \mathbb{E}\big[|X_i|\big] < \delta\,.$$

It follows in particular that

$$\int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P < \varepsilon \quad \forall\, i \in I \quad \Rightarrow \quad \sup_{i\in I} \int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P \leq \varepsilon\,. \qquad \square$$

Remark 8.8. *(i) Existence of dominating integrable r.v. implies uniform integrability:* $|X_i| \leq Y \in \mathcal{L}^1\ \forall i \in I$

$$\Rightarrow \qquad \int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P \leqslant \int_{\{Y\geqslant c\}} Y \,\mathrm{d}P = \mathbb{E}\big[1_{\{Y\geqslant c\}} \cdot Y\big] \xrightarrow{\textit{Lebesgue} \quad c\nearrow\infty} 0,$$

since $1_{\{Y\geqslant c\}} \cdot Y \xrightarrow{c\to\infty} 0$ *P-a.s. (Markov's inequality)*
In particular, I *finite* $\Rightarrow (X_i)_{i\in I} \subset \mathcal{L}^1$ *uniformly integrable.*

(ii) Let $(X_i)_{i\in I}$*,* $(Y_i)_{i\in I}$ *be uniformly integrable,* $\alpha, \beta \in \mathbb{R}$

$$\Rightarrow \qquad (\alpha X_i + \beta Y_i)_{i\in I} \qquad \text{uniformly integrable}$$

(see Exercises)

Proof of Proposition 8.4. **(i)⇒(ii):** see Exercises. (Hint: Use Lemma 8.7).

(ii)⇒(i): a) $X \in \mathcal{L}^1$, because there exists a subsequence (n_k) such that $\lim_{k\to\infty} X_{n_k} = X$ P-a.s., so that

$$\mathbb{E}\big[|X|\big] = \mathbb{E}\big[\liminf_{k\to\infty} |X_{n_k}|\big] \overset{\text{Fatou}}{\leqslant} \liminf_{k\to\infty} \mathbb{E}\big[|X_{n_k}|\big] \leqslant \sup_{n\in\mathbb{N}} \mathbb{E}\big[|X_n|\big] < \infty.$$

b) W.l.o.g. $X = 0$ (because $(X_n)_{n\in\mathbb{N}}$ uniformly integrable implies $(X_n - X)_{n\in\mathbb{N}}$ uniformly integrable too by Remark 8.8)
Let $\varepsilon > 0$. Then there exists $\delta > 0$ such that for all $A \in \mathcal{A}$ with $P(A) < \delta$ it follows that $\int_A |X_n| \,\mathrm{d}P < \frac{\varepsilon}{2}$.

Since $X_n \to 0$ in probability, there exists $n_0 \in \mathbb{N}$, such that $P\big[|X_n| \geqslant \frac{\varepsilon}{2}\big] < \delta\ \forall\, n \geq n_0$. Hence, for $n \geq n_0$

$$\mathbb{E}\big[|X_n|\big] = \underbrace{\int_{\{|X_n|<\frac{\varepsilon}{2}\}} |X_n| \,\mathrm{d}P}_{<\frac{\varepsilon}{2}} + \underbrace{\int_{\{|X_n|\geqslant\frac{\varepsilon}{2}\}} |X_n| \,\mathrm{d}P}_{<\frac{\varepsilon}{2}} < \varepsilon,$$

and thus $\lim_{n\to\infty} \mathbb{E}[|X_n|] = 0$. $\square$

Corollary 8.9. $\lim_{n\to\infty} X_n = X$ *in probability and* $\big(|X_n|^p\big)_{n\in\mathbb{N}}$ *uniformly integrable,* $p > 0$

$$\Rightarrow \qquad \lim_{n\to\infty} X_n = X \quad \textit{in } \mathcal{L}^p.$$

Proof. $\lim_{n\to\infty}|X_n - X|^p \to 0$ in probability and since

$$|X_n - X|^p \leqslant 2^p \cdot \underbrace{\big(|X_n|^p + |X|^p\big)}_{\text{unif. integrable}},$$

$(|X_n - X|^p)_{n\in\mathbb{N}}$ is uniformly integrable too. Proposition 8.4 implies

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|^p\big] = 0.$$

□

Proposition 8.10. *Let* $g : [0,\infty[\to [0,\infty[$ *be measurable with* $\lim_{x\to\infty} \frac{g(x)}{x} = \infty$. *Then*

$$\sup_{i\in I} \mathbb{E}\left[g\big(|X_i|\big)\right] < \infty \qquad \Rightarrow \qquad (X_i)_{i\in I} \textit{ uniformly integrable}$$

Proof. Let $\varepsilon > 0$. Choose $c > 0$, such that $\frac{g(x)}{x} \geqslant \frac{1}{\varepsilon}\sup_{i\in I}\mathbb{E}\big[g\big(|X_i|\big)\big] + 1$ for $x \geqslant c$. Then

$$\begin{aligned}\int_{\{|X_i|\geqslant c\}} |X_i| \,\mathrm{d}P &= \int_{\{|X_i|\geqslant c\}} g\big(|X_i|\big) \cdot \frac{|X_i|}{g\big(|X_i|\big)} \,\mathrm{d}P \\ &\leqslant \frac{\varepsilon}{\sup\limits_{j\in I}\mathbb{E}\big[g\big(|X_j|\big)\big] + 1} \cdot \int_{\{|X_i|\geqslant c\}} g\big(|X_i|\big) \,\mathrm{d}P \leqslant \varepsilon \qquad \forall i \in I\end{aligned}$$

□

Example 8.11. (i) $p > 1$, $\sup_i \mathbb{E}[|X_i|^p] < \infty \Rightarrow (X_i)_{i\in I}$ uniformly integrable

(ii) ("finite entropy condition")

$$\sup_{i\in I}\mathbb{E}\Big[|X_i|\cdot\log^+\big(|X_i|\big)\Big] < \infty \qquad \Rightarrow \qquad (X_i)_{i\in I} \text{ uniformly integrable} \tag{1.8}$$

Example 8.12. Application to the strong law of large numbers Let $X_1, X_2, \dots$ be r.v. in $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ with $\mathbb{E}[X_i] = m$ for all $i \in \mathbb{N}$. Suppose that

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{n\to\infty} m \quad P-a.s. \tag{1.9}$$

Problem: Under what conditions do we have $\mathcal{L}^1$-convergence?

We have seen in Proposition **??** that (1.9) holds if $\sup_{i\in\mathbb{N}}\mathbb{E}[X_i^2] < \infty$ and $(X_i)_{i\in\mathbb{N}}$ uncorrelated ("$\mathcal{L}^2$-case"). We will see below that (1.9) always holds if $X_i \in \mathcal{L}^1$ are pairwise independent, identically distributed.

Solution:

$$\sup_{i\in\mathbb{N}}\mathbb{E}\Big[|X_i|\cdot\log^+\big(|X_i|\big)\Big] < \infty \quad \text{implies} \quad \lim_{n\to\infty}\frac{S_n}{n} = m \quad \text{in } \mathcal{L}^1.$$

In particular, in the situation of Proposition 7.7 (check) it follows that $\lim_{n\to\infty} \frac{S_n}{n} = m$ in $\mathcal{L}^1$.
Proof: $g(x) := x \cdot \log^+(x)$ ist monotone increasing and convex. Consequently,

$$\mathbb{E}\left[g\left(\frac{|S_n|}{n}\right)\right] \overset{\text{monotonicity}}{\leqslant} \mathbb{E}\left[g\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\right)\right] \overset{\text{convexity}}{\leqslant} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} g\big(|X_i|\big)\right] \leqslant \text{const.} \qquad \forall n\,.$$

Consequently, $\left(\frac{S_n}{n}\right)_{n\in\mathbb{N}}$ is uniformly integrable and thus

$$\lim_{n\to\infty} \frac{S_n}{n} = m \quad \text{in } \mathcal{L}^1.$$

One complementary remark concerning Lebesguet's dominated convergence theorem.

Proposition 8.13. *Let $X_n \geq 0$, $\lim\limits_{n\to\infty} X_n = X$ P-f.s. (or in probability). Then*

$$\begin{aligned} &\lim_{n\to\infty} X_n = X \quad \textit{in } \mathcal{L}^1 \\ \Leftrightarrow \quad &\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] \quad \textit{and} \quad \mathbb{E}[X] < \infty. \end{aligned}$$

Proof. **"‘$\Rightarrow$"’:** Obvious.

"‘$\Leftarrow$"’:

$$X + X_n = \underbrace{X \vee X_n}_{:=\sup\{X, X_n\}} + \underbrace{X \wedge X_n}_{:=\inf\{X, X_n\}}$$

Then

$$\lim_{n\to\infty} \mathbb{E}[X \wedge X_n] \overset{\text{Lebesgue}}{=} \mathbb{E}[X]$$

and thus

$$\lim_{n\to\infty} \mathbb{E}[X \vee X_n] = \mathbb{E}[X]\,.$$

Now $|X_n - X| = (X \vee X_n) - (X \wedge X_n)$ implies

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|\big] = \mathbb{E}[X] - \mathbb{E}[X] = 0.$$

□

$\mathcal{L}^p$-completeness

Proposition 8.14 (Riesz-Fischer)**.** *Let $1 \leqslant p < \infty$ and $X_n \in \mathcal{L}^p$ with*

$$\lim_{n,m\to\infty} \int |X_n - X_m|^p \, \mathrm{d}P = 0.$$

Then there exists a r.v. $X \in \mathcal{L}^p$ such that

(i) $\lim\limits_{k\to\infty} X_{n_k} = X$ *P-a.s. along some subsequence,*

(ii) $\lim\limits_{n\to\infty} X_n = X$ *in $\mathcal{L}^p$.*

Proof. See text books on measure theory. □

9 Distribution of random variables

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $X : \Omega \to \bar{\mathbb{R}}$ be a r.v.

Let μ be the *distribution of X (under P)*, i.e., $\mu(A) = P[X \in A]$ for all $A \in \mathcal{B}(\bar{\mathbb{R}})$.

Assume that $P[X \in \mathbb{R}] = 1$ (in particular, X P-a.s. finite, and μ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition 9.1. The function $F : \mathbb{R} \to [0,1]$, defined by

$$F(b) := P[X \leqslant b] = \mu\big(]-\infty, b]\big), \quad b \in \mathbb{R}, \tag{1.10}$$

is called the *distribution function* of X resp. μ.

Proposition 9.2. *(i) F is*

monotone increasing:	$a \leqslant b \quad \Rightarrow \quad F(a) \leqslant F(b)$
right continuous:	$F(a) = \lim_{b \searrow a} F(b)$
normalized:	$\lim_{a \searrow -\infty} F(a) = 0, \qquad \lim_{b \nearrow +\infty} F(b) = 1.$

(ii) To any such function there exists a unique probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with (1.10)*.*

Proof. (i) Monotonicity is obvious.

Right continuity: if $b \searrow a$ then $]-\infty, b] \searrow\,]-\infty, a]$, hence by continuity of μ from above (vgl. Proposition 1.9):

$$F(a) = \mu\big(]-\infty, a]\big) \overset{1.9}{=} \lim_{b \searrow a} \mu\big(]-\infty, b]\big) = \lim_{b \searrow a} F(b).$$

Similarly, $]-\infty, a] \searrow \emptyset$ if $a \searrow -\infty$ (resp. $]-\infty, b] \nearrow \mathbb{R}$ if $b \nearrow \infty$), and thus

$$\lim_{a \searrow -\infty} F(a) = \lim_{a \searrow -\infty} \mu\big(]-\infty, a]\big) = 0$$

(resp. $\lim_{b \nearrow \infty} F(b) = \lim_{b \nearrow \infty} \mu\big(]-\infty, b]\big) = 1$).

(ii) *Existence:* Let λ be the Lebesgue measure on $]0,1[$. Define the "inverse function" G of $F : \mathbb{R} \to [0,1]$ by

$$G :]0,1[\to \mathbb{R}$$
$$G(y) := \inf\big\{x \in \mathbb{R} \mid F(x) > y\big\}.$$

Note that $y < F(x) \quad \Rightarrow \quad G(y) \leqslant x$ implies

$$\big]0, F(x)\big[\subset \{G \leqslant x\}$$

and $G(y) \leqslant x \quad \Rightarrow \quad \exists\, x_n \searrow x$ with $F(x_n) > y$, hence $F(x) \geqslant y$, so that

$$\{G \leqslant x\} \subset \big]0, F(x)\big].$$

Combining both inclusions we obtain that

$$\big]0, F(x)\big[\subset \{G \leqslant x\} \subset \big]0, F(x)\big]\,.$$

so that G is measurable.
Let $\mu := G(\lambda) = \lambda \circ G^{-1}$ (probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). Then

$$\mu\big(]-\infty, x]\big) = \lambda(\{G \leqslant x\}) = \lambda\Big(\big]0, F(x)\big]\Big) = F(x) \qquad \forall\, x \in \mathbb{R}\,.$$

Uniqueness: later. □

Remark 9.3. *(i) Let Y be a r.v. with uniform distribution on $[0,1]$, then $X = G(Y)$ has distribution μ. In particular: simulating the uniform distribution on $[0,1]$ gives by transformation with G a simulation of μ.*

(ii) Some authors define the distribution function F by $F(x) := \mu\big(]-\infty, x[\big)$. In this case F is left continuous, not right continuous.

Remark 9.4. *(i) Let F be a distribution function and let $x \in \mathbb{R}$: Then*

$$F(x) - F(x-) = \lim_{n \nearrow \infty} \mu\left(\Big]x - \frac{1}{n}, x\Big]\right) = \mu(\{x\})$$

is called the step height of F in x. *In particular:*

$$F \text{ continuous} \quad \Leftrightarrow \quad \forall\, x \in \mathbb{R} : \mu(\{x\}) = 0 \quad \text{“}\mu \text{ is continuous”}.$$

(ii) Let F be monotone increasing and bounded, then F has at most countable many points of discontinuity.

Definition 9.5. (i) F (resp. μ) is called *discrete*, if there exists a countable set $S \subset \mathbb{R}$ with $\mu(S) = 1$. In this case, μ is uniquenely determined by the weights $\mu(\{x\})$, $x \in S$, and F is a *step function* of the following type:

$$F(x) = \sum_{\substack{y \in S,\\ y \leqslant x}} \mu(\{y\}).$$

(ii) F (resp. μ) is called *absolutely continuous*, if there exists a measurbale function $f \geqslant 0$ (called the "*density*"), such that

$$F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t, \tag{1.11}$$

resp., for all $A \in \mathcal{B}(\mathbb{R})$:

$$\mu(A) = \int_A f(t)\, \mathrm{d}t = \int_{-\infty}^{\infty} 1_A \cdot f\, \mathrm{d}t. \tag{1.12}$$

In particular $\displaystyle\int_{-\infty}^{+\infty} f(t)\, \mathrm{d}t = 1$.

Remark 9.6. *(i) Every measurable function $f \geqslant 0$ with $\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1$ defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by $A \mapsto \int_A f(t)\,\mathrm{d}t$.*

(ii) In the previous definition "(1.11)⇒(1.12)", because $A \mapsto \int_A f(t)\,\mathrm{d}t$ defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with distribution function F. Uniqueness in 9.2 implies the assertion.

Example 9.7. (i) **Uniform distribution on $[a, b]$.** Let $f := \frac{1}{b-a} \cdot 1_{[a,b]}$. The associated distribution function is given by

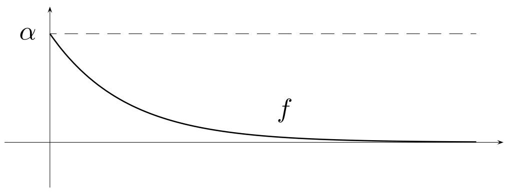
$$F(x) := \begin{cases} 0 & \text{if } x \leqslant a \\ \frac{1}{b-a} \cdot (x-a) & \text{if } x \in [a, b] \\ 1 & \text{if } x \geqslant b. \end{cases}$$

(continuous analogue to the dicrete uniform distribution on a finite set)

(ii) **Exponential distribution with parameter $\alpha > 0$.**

$$f(x) := \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0, \end{cases}$$

$$F(x) := \begin{cases} 1 - e^{-\alpha x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0. \end{cases}$$



(continuous analogue of the geometric distribution)

$$\int_k^{k+1} f(x)\,\mathrm{d}x = F(k+1) - F(k) = e^{-\alpha k}(1 - e^{-\alpha}) = (1-p)^k p \text{ with } p = 1 - e^{-\alpha}.$$

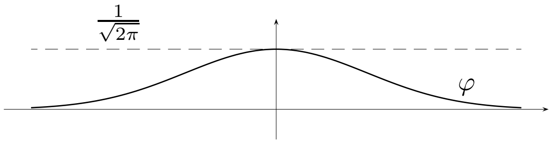
(iii) **Normal distribution $N(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$**

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

The associated distribution function is given by

$$\begin{aligned} F_{m,\sigma^2}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}}\,\mathrm{d}y \\ &\overset{z=\frac{y-m}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{z^2}{2}}\,\mathrm{d}z = F_{0,1}\Big(\frac{x-m}{\sigma}\Big). \end{aligned}$$

$\Phi := F_{0,1}$ is called the distribution function of the *standard normal distribution* $N(0, 1)$.



The expectation $E[X]$ (or more general $E[h(X)]$) can be calculated with the help of the distribution μ of X:

Proposition 9.8. *Let $h \geqslant 0$ be measurable, then*

$$\mathbb{E}\big[h(X)\big] = \int_{-\infty}^{+\infty} h(x)\ \mu(\mathrm{d}x)$$
$$= \begin{cases} \displaystyle\int_{-\infty}^{+\infty} h(x) \cdot f(x)\ \mathrm{d}x & \textit{if } \mu \textit{ absolutely continuous with density } f \\ \displaystyle\sum_{x \in S} h(x) \cdot \mu(\{x\}) & \textit{if } \mu \textit{ discrete, } \mu(S) = 1 \textit{ and } S \textit{ countable.} \end{cases}$$

Proof. See exercises. □

Example 9.9. Let X be $N(m, \sigma^2)$-distributed. Then

$$\mathbb{E}[X] = \int x \cdot f_{m,\sigma^2}(x)\ \mathrm{d}x = m + \underbrace{\int (x-m) \cdot f_{m,\sigma^2}(x)\ \mathrm{d}x}_{=0} = m\,.$$

The p^{th} *central moment of X* is given by

$$\begin{aligned}
\mathbb{E}\big[|X-m|^p\big] &= \int |x-m|^p \cdot f_{m,\sigma^2}(x)\ \mathrm{d}x, \\
&= \int |x|^p \cdot f_{0,\sigma^2}(x)\ \mathrm{d}x. \\
&= 2\int_0^\infty x^p \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}\ \mathrm{d}x, \\
&\underbrace{=}_{y=\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{\pi}} \cdot 2^{\frac{p}{2}} \cdot \sigma^p \underbrace{\int_0^\infty y^{\frac{p+1}{2}-1} \cdot e^{-y}\ \mathrm{d}y.}_{=\Gamma(\frac{p+1}{2})}
\end{aligned}$$

In particular:

$$p = 1 : \mathbb{E}\big[|X-m|\big] = \sigma \cdot \sqrt{\frac{2}{\pi}}$$
$$p = 2 : \mathbb{E}\big[|X-m|^2\big] = \sigma^2$$
$$p = 3 : \mathbb{E}\big[|X-m|^3\big] = 2^{\frac{3}{2}} \cdot \frac{\sigma^3}{\sqrt{\pi}}$$
$$p = 4 : \mathbb{E}\big[|X-m|^4\big] = 3\sigma^4.$$

10 Weak convergence of probability measures

Let S be a topological space and $\mathcal{S}$ be the Borel σ-algebra on S.

Let μ, μ_n, $n \in \mathbb{N}$, be probability measures on $(S, \mathcal{S})$.

What is a reasonable notion of convergence of the sequence μ_n towards μ? The notion of "pointwise convergence" in the sense that $\mu_n(A) \xrightarrow{n\to\infty} \mu(A)$ for all $A \in \mathcal{S}$ is too strong for many applications.

Definition 10.1. Let μ and μ_n, $n \in \mathbb{N}$, be probability measures on $(S, \mathcal{S})$. The sequence (μ_n) converges to μ *weakly* if for all $f \in C_b(S)$ (= the space of bounded continuous functions on S) it follows that

$$\int f \,\mathrm{d}\mu_n \xrightarrow{n\to\infty} \int f \,\mathrm{d}\mu.$$

Example 10.2. (i) $x_n \xrightarrow{n\to\infty} x$ in S implies $\delta_{x_n} \xrightarrow{n\to\infty} \delta_x$ weakly.

(ii) Let $S := \mathbb{R}^1$ and $\mu_n := N\big(0, \frac{1}{n}\big)$. Then $\mu_n \to \delta_0$ weakly, since for all $f \in \mathbb{C}_b(\mathbb{R})$

$$\begin{aligned}\int f \,\mathrm{d}\mu_n &= \int f(x) \cdot \frac{1}{\sqrt{2\pi \frac{1}{n}}} \cdot e^{-\frac{x^2}{2\cdot\frac{1}{n}}} \,\mathrm{d}x \\ &\overset{x=\frac{y}{\sqrt{n}}}{=} \int f\Big(\frac{y}{\sqrt{n}}\Big) \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}} \,\mathrm{d}y \\ &\overset{\text{Lebesgue}}{\xrightarrow{n\to\infty}} f(0) \quad = \quad \int f \,\mathrm{d}\delta_0.\end{aligned}$$

Proposition 10.3 (Portemanteau-Theorem)**.** *Let S be a metric space with metric d. Then the following statements are equivalent:*

(i) $\mu_n \to \mu$ *weakly*

(ii) $\int f \,\mathrm{d}\mu_n \xrightarrow{n\to\infty} \int f \,\mathrm{d}\mu$ *for all f bounded and uniformly continuous (w.r.t. d)*

(iii) $\limsup_{n\to\infty} \mu_n(F) \leqslant \mu(F)$ *for all $F \subset S$ closed*

(iv) $\liminf_{n\to\infty} \mu_n(G) \geqslant \mu(G)$ *for all $G \subset S$ open*

(v) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all $A \in \mathcal{S}$ with $\mu(\bar{A} \setminus \mathring{A}) = 0$.*

Proof. **(iii)⇔(iv):** Obvious by considering the complement.

(i)⇒(ii): Trivial.

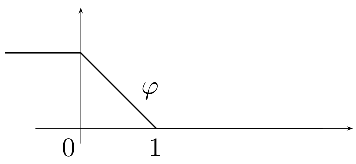
(ii)⇒(iii): Let $F \subset S$ be closed, let

$$G_m := \Big\{ x \in S \,\Big|\, d(x, F) < \frac{1}{m} \Big\}, \quad m \in \mathbb{N} \qquad \text{open!}$$

Then $G_m \searrow F$, hence $\mu(G_m) \searrow \mu(F)$.

If $\varepsilon > 0$ there exists some $m \in \mathbb{N}$ mit $\mu(G_m) < \mu(F) + \varepsilon$. Define

$$\varphi(x) := \begin{cases} 1 & \text{if } x \leqslant 0 \\ 1 - x & \text{if } x \in [0,1] \\ 0 & \text{if } x \geqslant 1. \end{cases}$$



and let $f := \varphi\big(m \cdot d(\,\cdot\,, F)\big)$.

f is Lipschitz, in particular uniformly continuous, $f = 0$ on G_m^c and $f = 1$ on F, and thus

$$\begin{aligned} \limsup_{n\to\infty} \mu_n(F) &\leqslant \limsup_{n\to\infty} \int f \, d\mu_n \overset{\text{(ii)}}{=} \int f \, d\mu \\ &\leqslant \mu(G_m) < \mu(F) + \varepsilon. \end{aligned}$$

(iii)⇒(v): Let A be such that $\mu(\bar{A} \setminus \mathring{A}) = 0$. Then

$$\begin{aligned} \mu(A) = \mu(\mathring{A}) &\overset{\text{(iv)}}{\leqslant} \liminf_{n\to\infty} \mu_n(\mathring{A}) \leqslant \liminf_{n\to\infty} \mu_n(A) \leqslant \limsup_{n\to\infty} \mu_n(A) \\ &\leqslant \limsup_{n\to\infty} \mu_n(\bar{A}) \overset{\text{(iii)}}{\leqslant} \mu(\bar{A}) = \mu(A). \end{aligned}$$

(v)⇒(iii): Let $F \subset S$ be closed. For all $\delta > 0$ we have that

$$\partial\big\{d(\,\cdot\,, F) \geqslant \delta\big\} \subset \big\{d(\,\cdot\,, F) = \delta\big\}.$$

Note The set

$$D := \Big\{\delta > 0 \,\Big|\, \mu\big(\{d(\,\cdot\,, F) = \delta\}\big) > 0\Big\}$$

is countable, since for all n the set

$$D_n := \Big\{\delta > 0 \,\Big|\, \mu\big(\underbrace{\{d(\,\cdot\,, F) = \delta\}}_{\text{disjoint!}}\big) > \frac{1}{n}\Big\}$$

is finite for any $n \in \mathbb{N}$. In particular, there exists a sequence $\delta_k \in \,]0, \infty[\, \setminus D$, $\delta_k \downarrow 0$ such that the set

$$F_k := \big\{d(\,\cdot\,, F) \leqslant \delta_k\big\}$$

satisfies $\mu(\bar{F}_k \setminus \mathring{F}_k) = 0$. $F_k \searrow F$ now implies that

$$\limsup_{n\to\infty} \mu_n(F) \leqslant \limsup_{n\to\infty} \mu_n(F_k) \overset{\text{(v)}}{=} \mu(F_k) \xrightarrow{k\to\infty} \mu(F).$$

(iii)⇒(i): Let $f \in C_b(S)$. It suffices to prove that

$$\limsup_{n\to\infty} \int f\,d\mu_n \leqslant \int f\,d\mu,$$

(since then

$$-\liminf \int f\,d\mu_n \leqslant \int (-f)\,\mathrm{d}\mu,$$

hence $\liminf \int f\,d\mu_n \geqslant \int f\,\mathrm{d}\mu$)

W.l.o.g. $0 \leqslant f \leqslant 1$

Fix $k \in \mathbb{N}$ and let $F_j := \left\{ f \geqslant \frac{j}{k} \right\}$, $j \in \mathbb{N}$ (F_j closed!)

Then

$$\frac{1}{k}\sum_{i=1}^{k} 1_{F_i} \leqslant f \leqslant \frac{1}{k} + \frac{1}{k}\sum_{i=1}^{k} 1_{F_i}$$

Hence for all probability measures ν on $(S, \mathcal{S})$:

$$\frac{1}{k}\sum_{i=1}^{k} \nu(F_i) \underset{\dagger}{\leqslant} \int f\,d\nu \underset{\ddagger}{\leqslant} \frac{1}{k} + \frac{1}{k}\sum_{i=1}^{k} \nu(F_i).$$

and

$$\begin{aligned}\limsup_{n\to\infty} \int f\,\mathrm{d}\mu_n - \frac{1}{k} &\overset{(\ddagger)}{\leqslant} \frac{1}{k} \cdot \limsup_{n\to\infty} \sum_{i=1}^{k} \mu_n(F_i) \\ &\leqslant \frac{1}{k}\sum_{i=1}^{k} \limsup_{n\to\infty} \mu_n(F_i) \overset{\text{(iii)}}{\leqslant} \frac{1}{k}\sum_{i=1}^{k} \mu(F_i) \overset{(\dagger)}{\leqslant} \int f\,d\mu\end{aligned}$$ □

Corollary 10.4. *Let X, X_n, $n \in \mathbb{N}$, be measurable mappings from $(\Omega, \mathcal{A}, P)$ to $(S, \mathcal{S})$ with distributions μ, μ_n, $n \in \mathbb{N}$. Then:*

$$X_n \xrightarrow{n\to\infty} X \quad \textit{in probability} \quad \Rightarrow \quad \mu_n \xrightarrow{n\to\infty} \mu \quad \textit{weakly}$$

Here, $\lim_{n\to\infty} X_n = X$ in probability, if $\lim_{n\to\infty} P(d(X, X_n) > \delta) = 0$ for all $\delta > 0$.

Proof. Let $f \in C_b(S)$ be uniformly continuous and $\varepsilon > 0$. Then there exists a $\delta = \delta(\varepsilon) > 0$ such that:

$x, y \in S$ with $d(x, y) \leqslant \delta$ implies $|f(x) - f(y)| < \varepsilon$

Hence

$$\begin{aligned}&\left| \int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu_n \right| = \left| \mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(X_n)\big] \right| \\ &\leqslant \int_{\{d(X,X_n)\leqslant\delta\}} \big|f(X) - f(X_n)\big|\,\mathrm{d}P + \int_{\{d(X,X_n)>\delta\}} \big|f(X) - f(X_n)\big|\,\mathrm{d}P \\ &\leqslant \varepsilon + 2\|f\|_\infty \cdot \underbrace{P\big[d(X_n, X) > \delta\big]}_{\xrightarrow{n\to\infty} 0}.\end{aligned}$$ □

Corollary 10.5. *Let $S = \mathbb{R}^1$ and let μ, μ_n, $n \in \mathbb{N}$, be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with distributions functions F, F_n. Then the following statements are equivalent:*

(i) $\mu_n \xrightarrow{n\to\infty} \mu$ *vaguely, i.e.* $\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$ *for all* $f \in \mathbb{C}_0(\mathbb{R}^1)$ *(= the space of continuous functions with compact support)*

(ii) $\mu_n \xrightarrow{n\to\infty} \mu$ *weakly*

(iii) $F_n(x) \xrightarrow{n\to\infty} F(x)$ *for all x where F is continuous.*

(iv) $\mu_n\big(]a,b]\big) \xrightarrow{n\to\infty} \mu\big(]a,b]\big)$ *for all $]a,b]$ with $\mu(\{a\}) = \mu(\{b\}) = 0$.*

Proof. **(i)⇒(ii):** Exercise.

(ii)⇒(iii): Let x be such that F is continuous in x. Then $\mu(\{x\}) = 0$, which implies by the Portmanteau theorem:

$$F_n(x) = \mu_n\big(]-\infty, x]\big) \xrightarrow{n\to\infty} \mu\big(]-\infty, x]\big) = F(x).$$

(iii)⇒(iv): Let $]a,b]$ be such that $\mu(\{a\}) = \mu(\{b\}) = 0$ then F is continuous in a and b and thus

$$\begin{aligned}\mu\big(]a,b]\big) = F(b) - F(a) &\overset{\text{(iii)}}{=} \lim_{n\to\infty} F_n(b) - \lim_{n\to\infty} F_n(a)\\ &= \lim_{n\to\infty} \mu_n\big(]a,b]\big).\end{aligned}$$

(iv)⇒(i): Let $D := \big\{x \in \mathbb{R} \bigm| \mu(\{x\}) = 0\big\}$. Then $\mathbb{R} \setminus D$ is countable, hence $D \subset \mathbb{R}$ dense. Let $f \in C_0(\mathbb{R})$, then f is uniformly continuous, hence for $\varepsilon > 0$ we find $c_0 < \cdots < c_m \in D$ such that

$$\Big\|\underbrace{f - \sum_{k=1}^m f(c_{k-1}) \cdot \mathbb{I}_{]c_{k-1}, c_k]}}_{=:g}\Big\|_\infty \leqslant \sup_k \sup_{x\in[c_{k-1},c_k]} \big|f(x) - f(c_{k-1})\big| < \varepsilon.$$

Then

$$\begin{aligned}&\left|\int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu_n\right|\\ &\leqslant \underbrace{\int |f-g|\,\mathrm{d}\mu}_{<\varepsilon} + \left|\int g\,\mathrm{d}\mu - \int g\,\mathrm{d}\mu_n\right| + \underbrace{\int |f-g|\,\mathrm{d}\mu_n}_{<\varepsilon}\\ &\leqslant 2\varepsilon + \sum_{k=1}^m f(c_{k-1}) \cdot \Big|\mu\big(]c_{k-1}, c_k]\big) - \mu_n\big(]c_{k-1}, c_k]\big)\Big| \xrightarrow[n\to\infty]{\text{(iv)}} 2\varepsilon.\end{aligned}$$

□

11 Dynkin-systems and Uniqueness of probability measures

Let $\Omega \neq \emptyset$.

Definition 11.1. A collection of subsets $\mathcal{D} \subset \mathcal{P}(\Omega)$ is called a *Dynkin-system*, if:

(i) $\Omega \in \mathcal{D}$.

(ii) $A \in \mathcal{D} \quad \Rightarrow \quad A^c \in \mathcal{D}$.

(iii) $A_i \in \mathcal{D}$, $i \in \mathbb{N}$, pairwise disjoint, then

$$\dot{\bigcup_{i\in\mathbb{N}}} A_i \in \mathcal{D}.$$

Example 11.2. (i) Every σ-Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ is a Dynkin-system

(ii) Let P_1, P_2 be probability measures on $(\Omega, \mathcal{A})$. Then

$$\mathcal{D} := \big\{A \in \mathcal{A} \mid P_1(A) = P_2(A)\big\}$$

is a Dynkin-system

Remark 11.3. *(i) Let $\mathcal{D}$ be a Dynkin-system. Then*

$$A, B \in \mathcal{D}\,, A \subset B \qquad \Rightarrow \qquad B \setminus A = (B^c \,\dot\cup\, A)^c \in \mathcal{D}$$

(ii) Every Dynkin-system which is closed under finite unions (short notation: $\cap$-stable), is a σ-algebra, because:

(a) $A, B \in \mathcal{D} \quad \Rightarrow \quad A \cup B = A \,\dot\cup\, \underbrace{\big(B \setminus \underbrace{(A \cap B)}_{\in \mathcal{D}}\big)}_{\overset{\textit{(i)}}{\in} \mathcal{D}} \in \mathcal{D}.$

(b) $A_i \in \mathcal{D},\ i \in \mathbb{N} \quad \Rightarrow \quad \bigcup_{i\in\mathbb{N}} A_i = \dot{\bigcup_{i\in\mathbb{N}}} \underbrace{\Big[A_i \cap \Big(\underbrace{\bigcup_{n=1}^{i-1} A_n}_{\overset{\textit{(a)}}{\in} \mathcal{D}}\Big)^c\Big]}_{\substack{\in\, \mathcal{D} \text{ by ass.,} \\ \text{pairwise disjoint}}} \in \mathcal{D}.$

Proposition 11.4. *Let $\mathcal{B} \subset \mathcal{P}(\Omega)$ be a $\cap$-stable collection of subsets. Then*

$$\sigma(\mathcal{B}) = \mathcal{D}(\mathcal{B})\,,$$

where

$$\mathcal{D}(\mathcal{B}) := \bigcap_{\substack{\mathcal{D} \text{ Dynkin-system} \\ \mathcal{B} \subset \mathcal{D}}} \mathcal{D}$$

is called the Dynkin-system generated by $\mathcal{B}$.

Proof. See text books on measure theory. □

Proposition 11.5 (Uniqueness of probability measures)**.** *Let P_1, P_2 be probability measures on $(\Omega, \mathcal{A})$, and $\mathcal{B} \subset \mathcal{A}$ be a $\cap$-stable collection of subsets. Then:*

$$P_1(A) = P_2(A) \text{ for all } A \in \mathcal{B} \quad \Rightarrow \quad P_1 = P_2 \text{ on } \sigma(\mathcal{B}).$$

Proof. The collection of subsets

$$\mathcal{D} := \{A \in \mathcal{A} \mid P_1(A) = P_2(A)\}$$

is a Dynkin-system containing $\mathcal{B}$. Consequently,

$$\sigma(\mathcal{B}) \overset{11.4}{=} \mathcal{D}(\mathcal{B}) \subset \mathcal{D}.$$ □

Example 11.6. (i) For $p \in \,]0,1[$ the probability measure P_p on $(\Omega := \{0,1\}^{\mathbb{N}}, \mathcal{A})$ is uniquely determined by

$$P_p[X_1 = x_1, \ldots, X_n = x_n] = p^k(1-p)^{n-k}, \quad \text{with } k := \sum_{i=1}^n x_i$$

for all $x_1, \ldots, x_n \in \{0,1\}$, $n \in \mathbb{N}$, because the collection of cylindrical sets

$$\{X_1 = x_1, \ldots, X_n = x_n\}, \quad n \in \mathbb{N}_0,\ x_1, \ldots, x_n \in \{0,1\}$$

is $\cap$-stable, generating $\mathcal{A}$ (cf. Example 1.7).

(Existence of P_p for $p = \frac{1}{2}$ see Example 3.6. Existence for $p \in \,]0,1[\, \setminus \{\frac{1}{2}\}$ later.)

(ii) A probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is uniquely determined through its distribution function F $(:= \mu(]-\infty, \cdot\,]))$, because

$$\mu(]a,b]) = F(b) - F(a),$$

and the collection of intervals $]a,b]$, $a, b \in \mathbb{R}$, is $\cap$-stable, generating $\mathcal{B}(\mathbb{R})$.

2 Independence

1 Independent events

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 1.1. A collection of events $A_i \in \mathcal{A}$, $i \in I$, are said to be *independent* (w.r.t. P), if for any finite subset $J \subset I$

$$P\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} P(A_j).$$

A family of collection of subsets $\mathcal{B}_i \subset \mathcal{A}$, $i \in I$, is said to be *independent*, if for all finite subsets $J \subset I$ and for all subsets $A_j \in \mathcal{B}_j$, $j \in J$

$$P\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} P(A_j).$$

Proposition 1.2. *Let $\mathcal{B}_i$, $i \in I$, be independent and closed under intersections. Then:*

(i) $\sigma(\mathcal{B}_i)$, $i \in I$, are independent.

(ii) Let J_k, $k \in K$, be a partition of the index set I. Then the σ-algebras

$$\sigma\Big(\bigcup_{i\in J_k} \mathcal{B}_i\Big), \quad k \in K,$$

are independent.

Proof. (i) Let $J \subset I$, J finite, be of the form $J = \{j_1, \dots, j_n\}$. Let $A_{j_1} \in \sigma(\mathcal{B}_{j_1}), \dots, A_{j_n} \in \sigma(\mathcal{B}_{j_n})$.

We have to show that

$$P(A_{j_1} \cap \dots \cap A_{j_n}) = P(A_{j_1}) \cdots P(A_{j_n}). \tag{2.1}$$

To this end suppose first that $A_{j_2} \in \mathcal{B}_{j_2}, \dots, A_{j_n} \in \mathcal{B}_{j_n}$, and define

$$\begin{aligned}\mathcal{D}_{j_1} := \big\{A \in \sigma(\mathcal{B}_{j_1}) \;\big|\; & P(A \cap A_{j_2} \cap \dots \cap A_{j_n}) \\ & = P(A) \cdot P(A_{j_2}) \cdots P(A_{j_n})\big\}.\end{aligned}$$

Then $\mathcal{D}_{j_1}$ is a Dynkin system (!) containing $\mathcal{B}_{j_1}$. Proposition 1.11.4 now implies

$$\sigma(\mathcal{B}_{j_1}) = \mathcal{D}(\mathcal{B}_{j_1}) \subset \mathcal{D}_{j_1}\ ,$$

hence $\sigma(\mathcal{B}_{j_1}) = \mathcal{D}_{j_1}$. Iterating the above argument for $\mathcal{D}_{j_2}$, $\mathcal{D}_{j_3}$, implies (2.1).

(ii) For $k \in K$ define

$$\mathcal{C}_k := \Big\{ \bigcap_{j \in J} A_j \;\Big|\; J \subset J_k,\ J \text{ finite},\ A_j \in \mathcal{B}_j \Big\}.$$

Then $\mathcal{C}_k$ is closed under intersections and the collection of subsets $\mathcal{C}_k$, $k \in K$, are still independent, because: given $k_1, \dots, k_n \in K$ and finite subsets $J^1 \subset J_{k_1}, \dots, J^n \subset J_{k_n}$, then

$$P\Big(\underbrace{\Big(\bigcap_{i \in J^1} A_i \Big)}_{\in \mathcal{C}_{k_1}} \cap \dots \cap \underbrace{\Big(\bigcap_{i \in J^n} A_i \Big)}_{\in \mathcal{C}_{k_n}} \Big) \overset{\substack{\mathcal{B}_i, i \in I \\ \text{ind.}}}{=} \prod_{j=1}^{n} P\Big(\bigcap_{i \in J^j} A_i \Big).$$

(i) now implies that

$$\sigma(\mathcal{C}_k) = \sigma\Big(\bigcup_{i \in J_k} \mathcal{B}_i \Big), \quad k \in K,$$

are independent too. □

Example 1.3. Let $A_i \in \mathcal{A}$, $i \in I$, be independent. Then A_i, A_i^c, $i \in I$, are independent too.

Remark 1.4. *Pairwise independence does not imply independence in general.*
Beispiel: *Consider two tosses with a fair coin, i.e.*

$$\Omega := \big\{ (i,k) \;\big|\; i,k \in \{0,1\} \big\}, \quad P := \textit{uniform distribution.}$$

Consider the events

$$A := \textit{"1. toss 1"} = \big\{ (1,0),\ (1,1) \big\}$$
$$B := \textit{"2. toss 1"} = \big\{ (0,1),\ (1,1) \big\}$$
$$C := \textit{"1. and 2. toss equal"} = \big\{ (0,0),\ (1,1) \big\}.$$

Then $P(A) = P(B) = P(C) = \frac{1}{2}$ *and* A, B, C *are pairwise independent*

$$P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{1}{4}.$$

But on the other hand

$$P(A \cap B \cap C) = 14 \neq P(A) \cdot P(B) \cdot P(C).$$

Example 1.5. Independent 0-1-experiments with success probability $p \in [0,1]$. Let $\Omega := \{0,1\}^{\mathbb{N}}$, $X_i(\omega) := x_i$ and $\omega := (x_i)_{i \in \mathbb{N}}$. Let P_p be a probability measure on $\mathcal{A} := \sigma\big(\{X_i = 1\},\ i = 1, 2, \dots \big)$, with

(i) $P_p[X_i = 1] = p$ (hence $P_p[X_i = 0] = P_p\big(\{X_i = 1\}^c \big) = 1 - p$).

(ii) $\{X_i = 1\}$, $i \in \mathbb{N}$, are *independent* w.r.t. P_p.

Existence of such a probability measure later! Then for any $x_1, \dots, x_n \in \{0,1\}$:

$$P_p[X_{i_1} = x_1, \dots, X_{i_n} = x_n] \overset{\text{(ii) and 1.3}}{=} \prod_{j=1}^{n} P_p[X_{i_j} = x_j] \overset{\text{(i)}}{=} p^k (1-p)^{n-k},$$

where $k := \sum_{i=1}^{n} x_i$ gilt. Hence P_p is uniquely determined by (i) and (ii).

Proposition 1.6 (Kolmogorov's Zero-One Law)**.** *Let $\mathcal{B}_n$, $n \in \mathbb{N}$, be independent σ-algebras, and*

$$\mathcal{B}_\infty := \bigcap_{n=1}^{\infty} \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{B}_m\Big)$$

be the tail-field *(resp.* σ-algebra of terminal events*). Then*

$$P(A) \in \{0,1\} \qquad \forall\, A \in \mathcal{B}_\infty$$

i.e., P is deterministic on $\mathcal{B}_\infty$.

Illustration: Independent 0-1-experiments
Let $\mathcal{B}_i = \sigma\big(\{X_i = 1\}\big)$. Then

$$\mathcal{B}_\infty = \bigcap_{n \in \mathbb{N}} \sigma\Big(\bigcup_{m \geqslant n} \mathcal{B}_m\Big)$$

is the σ-algebra containing the events of the remote future, e.g.

$$\limsup_{i \to \infty} \{X_i = 1\} = \{\text{“infinitely many ‘1’”}\}$$

$$\Big\{\omega \in \{0,1\}^{\mathbb{N}} \;\Big|\; \lim_{n \to \infty} \underbrace{\frac{1}{n} \sum_{i=1}^{n} X_i(\omega)}_{=: \frac{S_n(\omega)}{n}} \quad \text{exists}\Big\}$$

Proof of the Zero-One Law. Proposition 1.2 implies that for all n

$$\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_{n-1}, \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{B}_m\Big)$$

are independent. Since $\mathcal{B}_\infty \subset \sigma\Big(\bigcup_{m \geqslant n} \mathcal{B}_m\Big)$, this implies that for all n

$$\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_{n-1}, \mathcal{B}_\infty$$

are independent. By definition this implies that

$$\mathcal{B}_\infty\,, \mathcal{B}_n\,, n \in \mathbb{N} \text{ are independent}$$

and now Proposition 1.2(ii) implies that

$$\sigma\Big(\bigcup_{n\in\mathbb{N}} \mathcal{B}_n\Big) \quad \text{und} \quad \mathcal{B}_\infty$$

are idependent. Since $\mathcal{B}_\infty \subset \sigma\Big(\bigcup_{n\geqslant 1} \mathcal{B}_n\Big)$ we finally obtain that $\mathcal{B}_\infty$ and $\mathcal{B}_\infty$ are independent. The conclusion now follows from the next lemma. □

Lemma 1.7. *Let $\mathcal{B} \subset \mathcal{A}$ be a σ-algebra such that $\mathcal{B}$ is independent from $\mathcal{B}$. Then*

$$P(A) \in \{0, 1\} \qquad \forall\, A \in \mathcal{B}\,.$$

Proof. For all $A \in \mathcal{B}$

$$P(A) = P(A \cap A) = P(A) \cdot P(A) = P(A)^2.$$

Hence $P(A) = 0$ or $P(A) = 1$. □

For any sequence A_n, $n \in \mathbb{N}$, of independent events in $\mathcal{A}$, Kolmogorov's Zero-One Law implies in particular for

$$A_\infty := \bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m \quad (=: \limsup_{n\to\infty} A_n)$$

that $P(A_\infty) = 0 - 1$.
Proof: The σ-algebras $\mathcal{B}_n := \sigma\{A_n\} = \{\emptyset, \Omega, A, A^c\}$, $n \in \mathbb{N}$, are independent by Proposition 1.2 and $A_\infty \in \mathcal{B}_\infty$.

Lemma 1.8 (Borel-Cantelli). *(i) Let $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. Then*

$$\sum_{i=1}^\infty P(A_i) < \infty \quad \Rightarrow \quad P\big(\limsup_{i\to\infty} A_i\big) = 0.$$

(ii) Assume that $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, are independent. Then

$$\sum_{i=1}^\infty P(A_i) = \infty \quad \Rightarrow \quad P\big(\limsup_{i\to\infty} A_i\big) = 1.$$

Proof. (i) See Lemma 1.1.11.

(ii) It suffices to show that

$$P\Big(\bigcup_{m=n}^\infty A_m\Big) = 1 \quad \text{resp.} \quad P\Big(\bigcap_{m=n}^\infty A_m^c\Big) = 0 \qquad \forall\, n\,.$$

The last equality follows from the fact that

$$P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = \lim_{k\to\infty} \underbrace{P\Big(\bigcap_{m=n}^{n+k} A_m^c\Big)}_{=\prod_{m=n}^{n+k} P(A_m^c) \quad \text{ind.}}$$

$$= \prod_{m=n}^{n+k} (1 - P(A_m)) \leq \exp\left(\sum_{m=n}^{n+k} P(A_m)\right) = 0$$

where we used the inequality $1 - \alpha \leqslant e^{-\alpha}$ for all $\alpha \in \mathbb{R}$.

□

Example 1.9. Independent 0-1-**experiments with success probability** $p \in]0,1[$**.** Let $(x_1, \dots, x_N) \in \{0,1\}^N$ ("binary text of length N").

P_p["text occurs"] ?

To calculate this probability we partition the infinite sequence $\omega = (y_n) \in \{0,1\}^{\mathbb{N}}$ into blocks of length N

$$(\underbrace{y_1, y_2, \dots}_{\substack{\text{1. block} \\ \text{length} = N}} \quad \underbrace{\dots\dots}_{\substack{\text{2. block} \\ \text{length} = N}} \quad \dots) \in \Omega := \{0,1\}^{\mathbb{N}}.$$

and consider the events A_i = "text occurs in the i^{th} block". Clearly, A_i, $i \in \mathbb{N}$, are independent events (!) by Proposition 1.2(ii) with equal probability

$$P_p(A_i) = p^K (1-p)^{N-K} =: \alpha > 0.$$

where $K := \sum_{i=1}^N x_i$ is the total sum of ones. In particular, $\sum_{i=1}^\infty P_p(A_i) = \sum_{i=1}^\infty \alpha = \infty$, and now Borel-Cantelli implies $P_p(A_\infty) = 1$, where

$$A_\infty = \limsup_{i\to\infty} A_i := \text{"text occurs infinitely many times"}.$$

Moreover: since the indicator functions $1_{A_1}, 1_{A_2}, \dots$ are uncorrelated (since they are independent r.v. (see below)), the strong law of large numbers implies that

$$\frac{1}{n} \sum_{i=1}^n 1_{A_i} \xrightarrow{P_p\text{-a.s.}} \mathbb{E}[1_{A_i}] = \alpha\,,$$

i.e. the relative frequency of the given text in the infinite sequence is strictly positive.

2 Independent random variables

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 2.1. A family X_i, $i \in I$, of r.v. on $(\Omega, \mathcal{A}, P)$ is said to be *independent*, if the σ-algebras

$$\sigma(X_i) := X_i^{-1}\big(\mathcal{B}(\bar{\mathbb{R}})\big) \quad \Big(= \big\{\{X_i \in A\} \mid A \in \mathcal{B}(\bar{\mathbb{R}})\big\}\Big), \quad i \in I,$$

are independent, i.e. for all finite subsets $J \subset I$ and any Borel subsets $A_j \in \mathcal{B}(\bar{\mathbb{R}})$

$$P\Big(\bigcap_{j \in J} \{X_j \in A_j\}\Big) = \prod_{j \in J} P[X_j \in A_j].$$

Remark 2.2. *Let X_i, $i \in I$, be independent and $h_i : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$, $i \in I$, $\mathcal{B}(\bar{\mathbb{R}})/\mathcal{B}(\bar{\mathbb{R}})$-measurable. Then $Y_i := h_i(X_i)$, $i \in I$, are again independent, because $\sigma(Y_i) \subset \sigma(X_i)$ for all $i \in I$.*

Proposition 2.3. *Let $X_1, \dots, X_n$ be independent r.v., ≥ 0. Then*

$$\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n].$$

Proof. W.l.o.g. $n = 2$. (Proof of the general case by induction, using the fact that $X_1 \cdot \ldots \cdot X_{n-1}$ and X_n are independent , since $X_1 \cdot \ldots \cdot X_{n-1}$ is measurable w.r.t $\sigma\big(\sigma(X_1) \cup \cdots \cup \sigma(X_{n-1})\big)$ and $\sigma\big(\sigma(X_1) \cup \cdots \cup \sigma(X_{n-1})\big)$ and $\sigma(X_n)$ are independent by Proposition 1.2.)

It therefore suffices to consider two independent r.v. $X, Y, \geq 0$, and we have to show that

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \tag{2.2}$$

W.l.o.g. X, Y simple

(for general X and Y there exist increasing sequences of simple r.v. X_n (resp. Y_n), which are $\sigma(X)$-measurable (resp. $\sigma(Y)$-measurable), converging pointwise to X (resp. Y).

Then $\mathbb{E}[X_n Y_n] = \mathbb{E}[X_n] \cdot \mathbb{E}[Y_n]$ for all n implies (2.2) using monotone integration.)

But for X, Y simple, hence

$$X = \sum_{i=1}^{m} \alpha_i 1_{A_i} \quad \text{and} \quad Y = \sum_{j=1}^{n} \beta_j 1_{B_j},$$

with $\alpha_i, \beta_j \geqslant 0$ and $A_i \in \sigma(X)$ resp. $B_j \in \sigma(Y)$ it follows that

$$\mathbb{E}[XY] = \sum_{i,j} \alpha_i \beta_j \cdot P(A_i \cap B_j) = \sum_{i,j} \alpha_i \beta_j \cdot P(A_i) \cdot P(B_j) = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \qquad \square$$

Corollary 2.4. *X, Y independent, $X, Y \in \mathcal{L}^1$*

$$\Rightarrow \qquad XY \in \mathcal{L}^1 \qquad \textit{and} \qquad \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]\,.$$

Proof. Let $\varepsilon_1, \varepsilon_2 \in \{+,-\}$. Then X^{ε_1} and Y^{ε_2} are independent by Remark 2.2 and nonnegative. Proposition 2.3 implies

$$\mathbb{E}[X^{\varepsilon_1} \cdot Y^{\varepsilon_2}] = \mathbb{E}[X^{\varepsilon_1}] \cdot \mathbb{E}[Y^{\varepsilon_2}].$$

In particular $X^{\varepsilon_1} \cdot Y^{\varepsilon_2}$ in $\mathcal{L}^1$, because $\mathbb{E}[X^{\varepsilon_1}] \cdot \mathbb{E}[Y^{\varepsilon_2}] < \infty$. Hence

$$X \cdot Y = X^+ \cdot Y^+ + X^- \cdot Y^- - (X^+ \cdot Y^- + X^- \cdot Y^+) \quad \in \mathcal{L}^1,$$

and $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. □

Remark 2.5. *(i) In general the converse to the above corollary does not hold: For example let X be $N(0,1)$-distributed and $Y = X^2$. Then X and Y are not independent, but*

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0\,.$$

(ii)

$$X, Y \in \mathcal{L}^2 \qquad \textit{independent} \Rightarrow X, Y \qquad \textit{uncorelated}$$

because

$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0\,.$$

Corollary 2.6 (to the strong law of large numbers). *Let $X_1, X_2, \dots \in \mathcal{L}^2$ be independent with $\sup_{i \in \mathbb{N}} \operatorname{var}(X_i) < \infty$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \big(X_i(\omega) - \mathbb{E}[X_i]\big) = 0 \qquad \textit{P-a.s.}$$

If $\mathbb{E}[X_i] \equiv m$ then $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = m \quad P$-a.s.

3 Kolmogorov's law of large numbers

Proposition 3.1 (Kolmogorov, 1930). *Let $X_1, X_2, \dots \in \mathcal{L}^1$ be independent, identically distributed, $m = \mathbb{E}[X_i]$. Then*

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} X_i(\omega)}_{\substack{\textit{empirical} \\ \textit{mean}}} \xrightarrow{n\to\infty} m \quad \textit{P-a.s.}$$

Proposition 3.1 follows from the following more general result:

Proposition 3.2 (Etemadi, 1981). *Let $X_1, X_2, \dots \in \mathcal{L}^1$ be pairwise independent, identically distributed, $m = \mathbb{E}[X_i]$. Then*

$$\frac{1}{n} \sum_{i=1}^{n} X_i(\omega) \xrightarrow{n\to\infty} m \quad \textit{P-a.s.}$$

Proof. W.l.o.g. $X_i \geqslant 0$
(otherwise consider $X_1^+, X_2^+, \dots$ (pairwise independent, identically distributed)
and $X_1^-, X_2^-, \dots$ (pairwise independent, identically distributed))

1. Replace X_i by $\tilde{X}_i := 1_{\{X_i < i\}} X_i$.

   Clearly,

$$\tilde{X}_i = h_i(X_i) \quad \text{with} \quad h_i(x) := \begin{cases} x & \text{if } x < i \\ 0 & \text{if } x \geqslant i \end{cases}$$

   Then $\tilde{X}_1, \tilde{X}_2, \dots$ are pairwise independent by Remark 2.2. For the proof it is now sufficient to show that for $\tilde{S}_n := \sum_{i=1}^n \tilde{X}_i$ we have that

$$\frac{\tilde{S}_n}{n} \xrightarrow{n\to\infty} m \quad P\text{-a.s.}$$

   Indeed,

$$\begin{aligned} \sum_{n=1}^{\infty} P[X_n \neq \tilde{X}_n] &= \sum_{n=1}^{\infty} P[X_n \geqslant n] = \sum_{n=1}^{\infty} P[X_1 \geqslant n] \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\big[X_1 \in [k, k+1[\big] = \sum_{k=1}^{\infty} k \cdot P\big[X_1 \in [k, k+1[\big] \\ &= \sum_{k=1}^{\infty} \mathbb{E}\big[\underbrace{k \cdot 1_{\{X_1 \in [k,k+1[\}}}_{\leqslant X_1 \cdot 1_{\{X_1 \in [k,k+1[\}}}\big] \leqslant \mathbb{E}[X_1] < \infty \end{aligned}$$

   implies by the Borel-Cantelli lemma

$$P[X_n \neq \tilde{X}_n \text{ infinitely often}] = 0\,.$$

2. Reduce the proof to convergence along the subsequence $k_n = \lfloor \alpha^n \rfloor$ (= largest natural number $\leq \alpha^n$), $\alpha > 1$.

   We will show in Step 3. that

$$\frac{\tilde{S}_{k_n} - \mathbb{E}[\tilde{S}_{k_n}]}{k_n} \xrightarrow{n\to\infty} 0 \qquad \text{P-a.s.} \tag{2.3}$$

   This will imply the assertion of the Proposition, because

$$\mathbb{E}[\tilde{X}_i] = \mathbb{E}\big[1_{\{X_i < i\}} \cdot X_i\big] = \mathbb{E}\big[1_{\{X_1 < i\}} \cdot X_1\big] \overset{i\to\infty}{\nearrow} \mathbb{E}[X_1] (= m)$$

   hence

$$\frac{1}{k_n} \cdot \mathbb{E}[\tilde{S}_{k_n}] = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E}[\tilde{X}_i] \xrightarrow{n\to\infty} m,$$

and thus

$$\frac{1}{k_n} \cdot \tilde{S}_{k_n} \xrightarrow{n\to\infty} m \quad P\text{-a.s.}$$

If $l \in \mathbb{N} \cap [k_n, k_{n+1}[$, then

$$\underbrace{\frac{k_n}{k_{n+1}}}_{\xrightarrow{n\to\infty}\frac{1}{\alpha}} \cdot \underbrace{\frac{\tilde{S}_{k_n}}{k_n}}_{\xrightarrow[P\text{-a.s.}]{n\to\infty}m} \leqslant \frac{\tilde{S}_l}{l} \leqslant \underbrace{\frac{\tilde{S}_{k_{n+1}}}{k_{n+1}}}_{\xrightarrow[P\text{-a.s.}]{n\to\infty}m} \cdot \underbrace{\frac{k_{n+1}}{k_n}}_{\xrightarrow{n\to\infty}\alpha} .$$

Hence there exists a P-null set $N_\alpha \in \mathcal{A}$, such that for all $\omega \notin N_\alpha$

$$\frac{1}{\alpha} \cdot m \leqslant \liminf_{l\to\infty} \frac{\tilde{S}_l(\omega)}{l} \leqslant \limsup_{l\to\infty} \frac{\tilde{S}_l(\omega)}{l} \leqslant \alpha \cdot m.$$

Finally choose a subsequence $\alpha_n \searrow 1$. Then for all $\omega \notin N := \bigcup_{n\geqslant 1} N_{\alpha_n}$

$$\lim_{l\to\infty} \frac{\tilde{S}_l(\omega)}{l} = m.$$

3. Due to Lemma 1.7.7 it suffices for the proof of (2.3) to show that

$$\forall \varepsilon > 0 : \qquad \sum_{n=1}^{\infty} P\left[\left|\frac{\tilde{S}_{k_n} - \mathbb{E}[\tilde{S}_{k_n}]}{k_n}\right| \geqslant \varepsilon\right] < \infty$$

(fast convergence in probability towards 0)

Pairwise independence of $\tilde{X}_i$ implies $\tilde{X}_i$ pairwise uncorrelated, hence

$$P\left[\left|\frac{\tilde{S}_{k_n} - \mathbb{E}[\tilde{S}_{k_n}]}{k_n}\right| \geqslant \varepsilon\right] \leqslant \frac{1}{k_n^2\varepsilon^2} \cdot \mathrm{var}(\tilde{S}_{k_n}) = \frac{1}{k_n^2\varepsilon^2} \sum_{i=1}^{k_n} \mathrm{var}(\tilde{X}_i)$$
$$\leqslant \frac{1}{k_n^2\varepsilon^2} \sum_{i=1}^{k_n} \mathbb{E}\big[(\tilde{X}_i)^2\big].$$

It therefore suffices to show that

$$s := \sum_{n=1}^{\infty}\left(\frac{1}{k_n^2}\sum_{i=1}^{k_n}\mathbb{E}\big[(\tilde{X}_i)^2\big]\right) = \sum_{\substack{(i,n)\in\mathbb{N}^2,\\ i\leqslant k_n}} \frac{1}{k_n^2} \cdot \mathbb{E}\big[(\tilde{X}_i)^2\big] < \infty\,.$$

To this end note that

$$s = \sum_{i=1}^{\infty}\left(\sum_{n\,:\,k_n\geqslant i}\frac{1}{k_n^2}\right) \cdot \mathbb{E}\big[(\tilde{X}_i)^2\big].$$

We will show in the following that there exists a constant c such that

$$\sum_{n\,:\,k_n \geqslant i} \frac{1}{k_n^2} \leqslant \frac{c}{i^2}. \tag{2.4}$$

This will then imply that

$$\begin{aligned}
s &\overset{(2.4)}{\leqslant} c\sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathbb{E}\big[(\tilde{X}_i)^2\big] = c\sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathbb{E}\big[1_{\{X_1<i\}} \cdot X_1^2\big] \\
&\leqslant c\sum_{i=1}^{\infty}\left(\frac{1}{i^2}\sum_{l=1}^{i} l^2 \cdot P\big[X_1 \in [l-1,l[\big]\right) \\
&= c\sum_{l=1}^{\infty}\left(l^2 \cdot \underbrace{\left(\sum_{i=l}^{\infty}\frac{1}{i^2}\right)}_{\leqslant 2l^{-1}} \cdot P\big[X_1 \in [l-1,l[\big]\right) \\
&\leqslant 2c\sum_{l=1}^{\infty} l \cdot P\big[X_1 \in [l-1,l[\big] = 2c\sum_{l=1}^{\infty}\mathbb{E}\Big[\underbrace{l \cdot 1_{\{X_1\in[l-1,l[\}}}_{\leqslant (X_1+1)\cdot 1_{\{X_1\in[l-1,l[\}}}\Big] \\
&\leqslant 2c \cdot \big(\mathbb{E}[X_1]+1\big) < \infty,
\end{aligned}$$

where we used the fact that

$$\sum_{i=l}^{\infty}\frac{1}{i^2} \leqslant \frac{1}{l^2} + \sum_{i=l+1}^{\infty}\frac{1}{(i-1)i} = \frac{1}{l^2} + \sum_{i=l+1}^{\infty}\left(\frac{1}{i-1}-\frac{1}{i}\right) = \frac{1}{l^2}+\frac{1}{l} \leqslant \frac{2}{l}\,.$$

It remains to show (2.4). To this end note that

$$\begin{aligned}
&\lfloor \alpha^n \rfloor = k_n \leqslant \alpha^n < k_n + 1 \\
\Rightarrow\quad & k_n > \alpha^n - 1 \overset{\alpha>1}{\geqslant} \alpha^n - \alpha^{n-1} = \underbrace{\left(\frac{\alpha-1}{\alpha}\right)}_{=:c_\alpha}\alpha^n.
\end{aligned}$$

Let n_i be the smallest natural number satisfying $k_{n_i} = \lfloor \alpha^{n_i} \rfloor \geqslant i$, hence $\alpha^{n_i} \geqslant i$, then

$$\sum_{n\,:\,k_n\geqslant i}\frac{1}{k_n^2} \leqslant c_\alpha^{-2}\sum_{n\geqslant n_i}\frac{1}{\alpha^{2n}} = c_\alpha^{-2}\cdot\frac{1}{1-\alpha^{-2}}\cdot\alpha^{-2n_i} \leqslant \frac{c_\alpha^{-2}}{1-\alpha^{-2}}\cdot\frac{1}{i^2}. \qquad \square$$

Corollary 3.3. *Let $X_1, X_2, \ldots$ be pairwise independent, identically distributed (iid) with $X_i \geqslant 0$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = \mathbb{E}[X_1] \quad \big(\in [0,\infty]\big) \qquad \textit{P-a.s.}$$

Proof. W.l.o.g. $\mathbb{E}[X_1] = \infty$. Then $\frac{1}{n}\sum_{i=1}^n \big(X_i(\omega) \wedge N\big) \xrightarrow{n\to\infty} \mathbb{E}[X_1 \wedge N]$, P-a.s. for all N, hence

$$\frac{1}{n}\sum_{i=1}^n X_i(\omega) \geqslant \frac{1}{n}\sum_{i=1}^n \big(X_i(\omega) \wedge N\big) \xrightarrow{n\to\infty} \mathbb{E}[X_1 \wedge N] \overset{N\to\infty}{\nearrow} \mathbb{E}[X_1] \quad P\text{-a.s.} \qquad \square$$

Example 3.4. Growth in random media Let $Y_1, Y_2, \dots$ be i.i.d., $Y_i > 0$, with $m := \mathbb{E}[Y_i]$ (existence of such a sequence later!)

Define $X_0 = 1$ and inductively $X_n := X_{n-1} \cdot Y_n$

Clearly, $X_n = Y_1 \cdots Y_n$ and $\mathbb{E}[X_n] = \mathbb{E}[Y_1] \cdots \mathbb{E}[Y_n] = m^n$, hence

$$\mathbb{E}[X_n] \to \begin{cases} +\infty & \text{if } m > 1 \qquad \text{exponential growth (supercritical)} \\ 1 & \text{if } m = 1 \qquad \text{critical} \\ 0 & \text{if } m < 1 \qquad \text{exponential decay (subcritical)} \end{cases}$$

What will be the long-time behaviour of $X_n(\omega)$?

Surprisingly, in the supercritical case $m > 1$, one may observe that $\lim_{n\to\infty} X_n = 0$ with positive probability.

Explanation: Suppose that $\log Y_i \in \mathcal{L}^1$. Then

$$\frac{1}{n}\log X_n = \frac{1}{n}\sum_{i=1}^n \log Y_i \xrightarrow{n\to\infty} \mathbb{E}[\log Y_1] =: \alpha \quad P\text{-a.s.}$$

and

$\alpha < 0$: $\exists\, \varepsilon > 0$ with $\alpha + \varepsilon < 0$, so that $X_n(\omega) \leqslant e^{n(\alpha+\varepsilon)}$ $\forall\, n \geqslant n_0(\omega)$, hence P-a.s. exponential decay

$\alpha > 0$: $\exists\, \varepsilon > 0$ with $\alpha - \varepsilon > 0$, so that $X_n(\omega) \geqslant e^{n(\alpha-\varepsilon)}$ $\forall\, n \geqslant n_0(\omega)$, hence P-a.s. exponential growth

Note that Jensen's inequality

$$\alpha = \mathbb{E}[\log Y_1] \quad \leqslant \quad \log \underbrace{\mathbb{E}[Y_1]}_{=m},$$

and in general the inequality is strict, i.e. $\alpha < \log m$, so that it might happen that $\alpha < 0$ although $m > 1$ (!)

Illustration As a particular example let

$$Y_i := \begin{cases} \frac{1}{2}(1+c) & \text{with prob.} \frac{1}{2} \\ \frac{1}{2} & \text{with prob.} \frac{1}{2} \end{cases}$$

, so that $\mathbb{E}[Y_i] = \frac{1}{4}(1+c) + \frac{1}{4} = \frac{1}{2} + \frac{1}{4}c$ (supercritical if $c > 2$)

On the other hand

$$\mathbb{E}[\log Y_1] = \frac{1}{2} \cdot \left[\log\left(\frac{1}{2}(1+c)\right) + \log\frac{1}{2}\right] = \frac{1}{2} \cdot \log\frac{1+c}{4} \overset{c<3}{<} 0.$$

Hence $X_n \xrightarrow{n\to\infty} 0$ P-a.s. with exponential rate for $c < 3$, whereas at the same time for $c > 2$ $\mathbb{E}[X_n] = m^n \nearrow \infty$ with exponential rate.

Back to Kolmogorov's law of large numbers:
Let $X_1, X_2, \ldots \in \mathcal{L}^1$ i.i.d. with $m := \mathbb{E}[X_i]$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow{n\to\infty} \mathbb{E}[X_1] \quad P\text{-a.s.}$$

Define the "random measure"

$$\begin{aligned}\varrho_n(\omega, A) &:= \frac{1}{n}\sum_{i=1}^{n} 1_A\big(X_i(\omega)\big)\\ &= \text{"relative frequency of the event } X_i \in A\text{"}\end{aligned}$$

Then

$$\varrho_n(\omega,\,\cdot\,) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$$

is a probability measure on $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ for fixed ω and it is called the *empirical distribution of the first n observations*

Proposition 3.5. *For P-almost every $\omega \in \Omega$:*

$$\varrho_n(\omega,\,\cdot\,) \xrightarrow{n\to\infty} \mu := P \circ X_1^{-1} \quad \textit{weakly.}$$

Proof. Clearly, Kolmogorov's law of large numbers implies that for any $x \in \mathbb{R}$

$$\begin{aligned}F_n(\omega, x) := \varrho_n\big(\omega,]-\infty, x]\big) &= \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}\big(X_i(\omega)\big)\\ &\to \mathbb{E}[1_{]-\infty,x]}(X_1)] = P[X_1 \leq x] = \mu\big(]-\infty, x]\big) =: F(x)\end{aligned}$$

P-a.s., hence for every $\omega \notin N(x)$ for some P-null set $N(x)$.
Then

$$N := \bigcup_{r\in\mathbb{Q}} N(r).$$

is a P-null set too, and for all $x \in \mathbb{R}$ and all $s, r \in \mathbb{Q}$ with $s < x < r$ and $\omega \notin N$:

$$\begin{aligned}F(s) := \lim_{n\to\infty} F_n(\omega, s) &\leqslant \liminf_{n\to\infty} F_n(\omega, x)\\ &\leqslant \limsup_{n\to\infty} F_n(\omega, x) \leqslant \lim_{n\to\infty} F_n(\omega, r) = F(r).\end{aligned}$$

Hence, if F is continuous at x, then for $\omega \notin N$

$$\lim_{n\to\infty} F_n(\omega, x) = F(x).$$

Now the assertion follows from the Portmanteau theorem. $\square$

4 Joint distribution and convolution

Let $X_i \in \mathcal{L}^1$ i.i.d. Kolmogorov's law of large numbers implies that

$$\frac{1}{n}\underbrace{\sum_{i=1}^{n} X_i(\omega)}_{=:S_n} \xrightarrow{n\to\infty} \mathbb{E}[X_1] \quad P\text{-a.s.}$$

hence

$$\int f(x)\, d\Big(P \circ \Big(\frac{S_n}{n}\Big)^{-1}\Big)(x) = \mathbb{E}\Big[f\Big(\frac{S_n}{n}\Big)\Big]$$

$$\xrightarrow[n\to\infty]{\text{(Lebesgue)}} f\big(\mathbb{E}[X_1]\big) = \int f(x)\, d\delta_{\mathbb{E}[X_1]}(x) \qquad \forall f \in C_b(\mathbb{R})$$

i.e., the distribution of $\frac{S_n}{n}$ converges weakly to $\delta_{\mathbb{E}[X_1]}$. This is not surprising, because at least for $X_i \in \mathcal{L}^2$

$$\operatorname{var}\Big(\frac{S_n}{n}\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\operatorname{var}(X_i)}_{=\operatorname{var}(X_1)} \xrightarrow{n\to\infty} 0.$$

We will see later that if we rescale S_n appropriately, namely $\frac{1}{\sqrt{n}}S_n$, so that $\operatorname{var}\big(\frac{1}{\sqrt{n}}S_n\big) = \operatorname{var}(X_1)$), the sequence of distributions of $\frac{1}{\sqrt{n}}S_n$ is asymptotically distributed as a normal distribution.
One problem in this context is: How to calculate the distribution of S_n? Since S_n is a function of $X_1, \dots, X_n$, we need to consider their *joint* distribution in the sense of the following definition:

Definition 4.1. Let $X_1, \dots, X_n$ be real-valued r.v. on $(\Omega, \mathcal{A}, P)$. Then the distribution $\bar{\mu} := P \circ \bar{X}^{-1}$ of the transformation

$$\bar{X} : \Omega \to \mathbb{R}^n, \quad \omega \mapsto \big(X_1(\omega), \dots, X_n(\omega)\big)$$

under P is said to be the *joint distribution* of $X_1, \dots, X_n$.

Note that $\bar{\mu}$ is a probability measure on $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$ with $\bar{\mu}(\bar{A}) = P[\bar{X} \in \bar{A}]$ for all $\bar{A} \in \mathcal{B}(\mathbb{R}^n)$.

Remark 4.2. *(i) $\bar{\mu}$ is well-defined, because $\bar{X} : \Omega \to \mathbb{R}^n$ is $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable.*

Proof:

$$\begin{aligned}\mathcal{B}(\mathbb{R}^n) &= \sigma\big(\{A_1 \times \cdots \times A_n \mid A_i \in \mathcal{B}(\mathbb{R})\}\big)\\ &= \sigma\big(\{A_1 \times \cdots \times A_n \mid A_i =]-\infty, x_i]\,,\ x_i \in \mathbb{R}\}\big)\end{aligned}$$

and

$$\bar{X}^{-1}(A_1 \times \cdots \times A_n) = \bigcap_{i=1}^{n} \underbrace{\{X_i \in A_i\}}_{\in \mathcal{A}} \quad \in \mathcal{A}$$

which implies the measurability of the transformation $\bar{X}$ (see Remark 1.3.2 (ii))

(ii) *Proposition 1.11.5 implies that $\bar{\mu}$ is uniquely determined by*

$$\bar{\mu}(A_1 \times \cdots \times A_n) = P\Big(\bigcap_{i=1}^{n} \{X_i \in A_i\}\Big).$$

Example 4.3. (i) Let X, Y be r.v., uniformly distributed on $[0,1]$. Then

- X, Y independent $\Rightarrow$ joint distribution = uniform distribution on $[0,1]^2$
- $X = Y$ $\Rightarrow$ joint distribution = uniform distribution on the diagonal

(ii) Let X, Y be independent, $N(m, \sigma^2)$ distributed. The following Proposition shows that the joint distribution of X and Y has the density

$$f(x,y) = \frac{1}{2\pi\sigma^2} \cdot \exp\Big[-\frac{1}{2\sigma^2} \cdot \big((x-m)^2 + (y-m)^2\big)\Big]$$

which is a particular example of a *2-dimensional normal distribution*.

In the case $m = 0$ it follows that

$$R := \sqrt{X^2 + Y^2},$$
$$\Phi := \arctan \frac{Y}{X},$$

are independent and

Φ has a uniform distribution on $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$,

R has a density $\begin{cases} \frac{r}{\sigma^2} \exp\big(-\frac{r^2}{2\sigma^2}\big) & \text{if } r \geqslant 0 \\ 0 & \text{if } r < 0. \end{cases}$

Definition 4.4. (Products of probability spaces) The product of measurable spaces $(\Omega_i, \mathcal{A}_i)$, $i = 1, \dots n$, is defined as the measurable space

$$\Omega := \Omega_1 \times \ldots \times \Omega_n$$

endowed with the smallest σ-algebra

$$\mathcal{A} := \sigma\{A_1 \times \ldots \times A_n \mid A_i \in \mathcal{A}_i, 1 \leq i \leq n\}$$

generated by measurable cylindrical sets. $\mathcal{A}$ is said to be the *product σ-algebra* of $\mathcal{A}_i$ (notation: $\bigotimes_{i=1}^{n} \mathcal{A}_i$).

Let P_i, $i = 1, \dots, n$, be probability measures on $(\Omega_i, \mathcal{A}_i)$. Then there exists a unique probability measure P on the product space $(\Omega, \mathcal{A})$ satisfying

$$P(A_1 \times \cdots \times A_n) = P_1(A_1) \cdot \ldots \cdot P_n(A_n)$$

for every measurable cylindrical set. P is called the *product measure* of P_i (notation: $\bigotimes_{i=1}^{n} P_i$).

(Uniqueness of P follows from 1.11.5, existence later!)

Proposition 4.5. *Let $X_1, \dots, X_n$ be r.v. on $(\Omega, \mathcal{A}, P)$ with distributions $\mu_1, \dots, \mu_n$ and joint distribution $\bar{\mu}$. Then*

$$X_1, \dots, X_n \text{ independent} \quad \Leftrightarrow \quad \bar{\mu} = \bigotimes_{i=1}^{n} \mu_i,$$

(i.e., $\bar{\mu}(A_1 \times \cdots \times A_n) = \prod_i \mu_i(A_i)$ if $A_i \in \mathcal{B}(\mathbb{R})$).
In this case:

(i) $\bar{\mu}$ is uniquely determined by $\mu_1, \dots, \mu_n$.

(ii)

$$\begin{aligned}&\int \varphi(x_1, \dots, x_n)\, \mathrm{d}\bar{\mu}(x_1, \dots, x_n) \\ &\quad = \int \left(\cdots \left(\int \varphi(x_1, \dots, x_n)\, \mu_{i_1}(\mathrm{d}x_{i_1}) \right) \cdots \right) \mu_{i_n}(\mathrm{d}x_{i_n}).\end{aligned}$$

for all $\mathcal{B}(\mathbb{R}^n)$-measurable functions $\varphi : \mathbb{R}^n \to \bar{R}$ with $\varphi \geq 0$ or φ $\bar{\mu}$-integrable.

(iii) If μ_i is absolutely continuous with density f_i, $i = 1, \dots, n$, then $\bar{\mu}$ is absolutely continuous with density

$$\bar{f}(\bar{x}) := \prod_{i=1}^{n} f_i(x_i).$$

Proof. The equivalence is obvious.

(i) Obvious from part (ii) of the previous Remark 4.2.

(ii) See text books on measure theory.

(iii) $\bar{f}$ is nonnegative and measurable on $\mathbb{R}^n$ w.r.t. $\mathcal{B}(\mathbb{R}^n)$, and

$$\int_{\mathbb{R}^n} \bar{f}(\bar{x})\, \mathrm{d}\bar{x} = \prod_{i=1}^{n} \int_{\mathbb{R}} f_i(x_i)\, \mathrm{d}x_i = 1.$$

Hence,

$$\check{\mu}(A) := \int_A \bar{f}(\bar{x})\, \mathrm{d}\bar{x}, \quad A \in \mathcal{B}(\mathbb{R}^n),$$

defines a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. For $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ it follows that

$$\begin{aligned}\bar{\mu}(A_1 \times \cdots \times A_n) &= \prod_{i=1}^{n} \mu_i(A_i) = \prod_{i=1}^{n} \int_{A_i} f_i(x_i)\, \mathrm{d}x_i \\ &\overset{\text{(ii)}}{=} \int 1_{A_1 \times \cdots \times A_n}(\bar{x}) \cdot \bar{f}(\bar{x})\, \mathrm{d}\bar{x} = \check{\mu}(A_1 \times \cdots \times A_n).\end{aligned}$$

Hence $\bar{\mu} = \check{\mu}$ by 1.11.5. $\square$

Let $X_1, \dots, X_n$ be independent, $S_n := X_1 + \dots + X_n$
How to calculate the distribution of S_n with the help of the distribution of X_i?
In the following denote by $T_x : \mathbb{R}^1 \to \mathbb{R}^1$, $y \mapsto x + y$, the translation by $x \in \mathbb{R}$.

Proposition 4.6. *Let X_1, X_2 be independent r.v. with distributions μ_1, μ_2. Then:*

(i) The distribution of $X_1 + X_2$ is given by the convolution

$$\mu_1 * \mu_2 := \int \mu_1(\mathrm{d}x_1)\ \mu_2 \circ T_{x_1}^{-1}\,,\ \textit{i.e.}$$

$$\mu_1 * \mu_2(A) == \int 1_A(x_1 + x_2)\ \mu_1(\mathrm{d}x_1)\ \mu_2(\mathrm{d}x_2)$$

$$= \int \mu_1(\mathrm{d}x_1)\ \mu_2(A - x_1) \qquad \forall\, A \in \mathcal{B}(\mathbb{R}^1)\,.$$

(ii) If one of the distributions μ_1, μ_2 is absolutely continuous, e.g. μ_2 with density f_2, then $\mu_1 * \mu_2$ is absolutely continuous again with density

$$f(x) := \int \mu_1(\mathrm{d}x_1)\ f_2(x - x_1)$$

$$\Big(= \int f_1(x_1) \cdot f_2(x - x_1)\ \mathrm{d}x_1 =: (f_1 * f_2)(x) \quad \textit{if } \mu_1 = f_1\ \mathrm{d}x_1. \Big)$$

Proof. (i) Let $A \in \mathcal{B}(\mathbb{R})$, and define $\bar{A} := \big\{(x_1, x_2) \in \mathbb{R}^2 \bigm| x_1 + x_2 \in A\big\}$). Then

$$P[X_1 + X_2 \in A] = P\big[(X_1, X_2) \in \bar{A}\big] = (\mu_1 \otimes \mu_2)(\bar{A})$$

$$= \iint 1_{\bar{A}}(x_1, x_2)\ \mathrm{d}(\mu_1 \otimes \mu_2)(x_1, x_2)$$

$$= \iint 1_A(x_1 + x_2)\ \mathrm{d}(\mu_1 \otimes \mu_2)(x_1, x_2)$$

$$= \int \Big(\int 1_{A - x_1}(x_2)\ \mu_2(\mathrm{d}x_2) \Big)\ \mu_1(\mathrm{d}x_1)$$

$$= \int \mu_2(A - x_1)\ \mu_1(\mathrm{d}x_1) = (\mu_1 * \mu_2)(A).$$

(ii)

$$(\mu_1 * \mu_2)(A) = \int \mu_1(\mathrm{d}x_1)\ \mu_2(A - x_1) = \int \mu_1(\mathrm{d}x_1) \int_{A - x_1} f_2(x_2)\ \mathrm{d}x_2$$

$$\overset{\substack{\text{change of variable}\\ x - x_1 = x_2}}{=} \int \mu_1(\mathrm{d}x_1) \int_A f_2(x - x_1)\ \mathrm{d}x$$

$$\overset{4.5}{=} \int_A \Big(\int \mu_1(\mathrm{d}x_1)\ f_2(x - x_1) \Big)\ \mathrm{d}x. \qquad \square$$

Example 4.7.

(i) Let X_1, X_2 be independent r.v. with Poisson-distribution π_{λ_1} and π_{λ_2}. Then $X_1 + X_2$ has Poisson-distribution $\pi_{\lambda_1+\lambda_2}$, because

$$(\pi_{\lambda_1} * \pi_{\lambda_2})(n) = \sum_{k=0}^{n} \pi_{\lambda_1}(k) \cdot \pi_{\lambda_2}(n-k) = e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k}{k!} \cdot \frac{\lambda_2^{n-k}}{(n-k)!}$$
$$= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \cdot \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1+\lambda_2)^n}{n!}.$$

(ii) Let X_1, X_2 be independent r.v. with normal distributions $N(m_i, \sigma_i^2)$, $i = 1, 2$. Then X_1+X_2 has normal distribution $N(m_1+m_2, \sigma_1^2+\sigma_2^2)$, because $f_{m_1+m_2,\sigma_1^2+\sigma_2^2} = f_{m_1,\sigma_1^2} * f_{m_2,\sigma_2^2}$ (Exercise!)

(iii) The *Gamma distribution* $\Gamma_{\alpha,p}$ is defined through its density $\gamma_{\alpha,p}$ given by

$$\gamma_{\alpha,p}(x) = \begin{cases} \frac{1}{\Gamma(p)} \cdot \alpha^p x^{p-1} e^{-\alpha x} & \text{if } x > 0 \\ 0 & \text{if } x \leqslant 0 \end{cases}$$

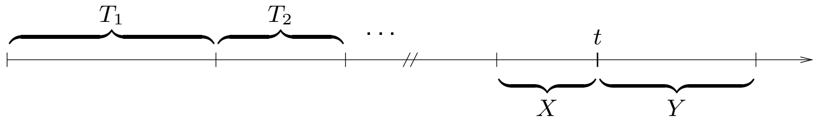
If X_1, X_2 are independent with distribution Γ_{α,p_i}, $i = 1, 2$, then $X_1 + X_2$ has distribution Γ_{α,p_1+p_2}. (Exercise!)

In the particular case $p_i = 1$: The sum $S_n = T_1 + \ldots + T_n$ of independent r.v. T_i with exponential distribution with parameter α has Gamma-distribution $\Gamma_{n,\alpha}$, i.e.

$$\gamma_{\alpha,n}(x) = \begin{cases} \frac{\alpha^n}{(n-1)!} \cdot e^{-\alpha x} x^{n-1} & \text{if } x > 0 \\ 0 & \text{if } x \leqslant 0. \end{cases}$$

Example 4.8 (The waiting time paradox)**.** Let $T_1, T_2, \ldots$ be independent, exponentially distributed waiting times with parameter $\alpha > 0$, so that in particular

$$\mathbb{E}[T_i] = \int_0^\infty x \cdot \alpha e^{-\alpha x} \, \mathrm{d}x = \cdots = \frac{1}{\alpha}.$$



Question: Fix some time t. How long on average is the remaining waiting time until the next event, i.e., how big is $\mathbb{E}[Y]$?
Answer: $\mathbb{E}[Y] = \frac{1}{\alpha}$, and

$$\mathbb{E}[X+Y] = \frac{1}{\alpha}(1 - e^{-\alpha t}) \approx \frac{1}{\alpha} \text{ for large } t\,.$$

More precisely:

(i) X, Y are independent.

(ii) Y has exponential distribution with parameter α.

(iii) X has exponential distribution with parameter α, "conditioned on" $[0,t]$, i.e.:

$$P[X \geqslant s] = e^{-\alpha s} \quad \forall\, 0 \leqslant s \leqslant t,$$
$$P[X = t] = e^{-\alpha t};$$

In particular,

$$\mathbb{E}[X] = \int_0^t s \cdot \alpha e^{-\alpha s}\ \mathrm{d}s + t \cdot e^{-\alpha t} = \cdots = \frac{1}{\alpha}(1 - e^{-\alpha t}).$$

Proof. Let us first determine the joint distribution of X and Y: Fix $0 \leqslant x \leqslant t$ and $y \geqslant 0$. Then for $S_n := T_1 + \cdots + T_n$, $S_0 := 0$:

$$\begin{aligned}
&P[X \geqslant x,\ Y \geqslant y] \\
&\quad = P\Big(\dot{\bigcup_{n\in\mathbb{N}}} \{S_n + x \leqslant t,\ S_{n+1} - y \geqslant t\}\Big) \\
&\quad = P[T_1 \geqslant y + t] + \sum_{n=1}^{\infty} P\big[S_n \leqslant t - x,\ T_{n+1} \geqslant y + t - S_n\big] \\
&\quad = e^{-\alpha(t+y)} + \sum_{n=1}^{\infty} \iint_{\substack{\{(s,r)\,|\, s\leqslant t-x,\\ r\geqslant y+t-s\}}} \gamma_{\alpha,n}(s) \cdot \alpha e^{-\alpha r}\ \mathrm{d}s\ \mathrm{d}r \\
&\quad = e^{-\alpha(t+y)} + \sum_{n=1}^{\infty} \int_{\{s\,|\, s\leqslant t-x\}} \gamma_{\alpha,n}(s) \cdot e^{-\alpha(y+t-s)}\ \mathrm{d}s \\
&\quad = e^{-\alpha(t+y)} \Big(1 + \int_0^{t-x} e^{\alpha s} \underbrace{\sum_{n=1}^{\infty} \gamma_{\alpha,n}(s)}_{\overset{(*)}{=}\alpha}\ \mathrm{d}s\Big) \\
&\quad = e^{-\alpha(t+y)} \Big(1 + \int_0^{t-x} \alpha e^{\alpha s}\ \mathrm{d}s\Big) \\
&\quad = e^{-\alpha(t+y)} \cdot e^{\alpha(t-x)} = e^{-\alpha y} \cdot e^{-\alpha x}.
\end{aligned}$$

Consequently:

(i) If $x = 0$: Y ist exponentially distributed with parameter α.

(ii) If $y = 0$: X ist exponentially distributed with parameter α, conditioned on $[0,t]$.

(iii) X, Y are independent.

We have used in line six that:

$$\sum_{n=1}^{\infty} \gamma_{\alpha,n}(s) = \sum_{n=1}^{\infty} \frac{\alpha^n}{(n-1)!} \cdot s^{n-1} e^{-\alpha s} = \alpha e^{-\alpha s} \sum_{n=1}^{\infty} \frac{(\alpha s)^{n-1}}{(n-1)!} = \alpha e^{-\alpha s} e^{\alpha s} = \alpha. \ \square$$

5 Characteristic functions

Let $\mathcal{M}_+^1(\mathbb{R}^n)$ be the set of all probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

For given $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$ define its *characteristic function* as the complex-valued function $\hat{\mu} : \mathbb{R}^n \to \mathbb{C}$ defined by

$$\hat{\mu}(u) := \int e^{i\langle u,y\rangle}\,\mu(dy) := \int \cos(\langle u,y\rangle)\,\mu(dy) + i\int \sin(\langle u,y\rangle)\,\mu(dy)\,.$$

Proposition 5.1. *Let $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$. Then*

(i) $\hat{\mu}(0) = 1$.

(ii) $|\hat{\mu}| \leqslant 1$.

(iii) $\hat{\mu}(-u) = \overline{\hat{\mu}(u)}$.

(iv) $\hat{\mu}$ *is uniformly continuous.*

(v) $\hat{\mu}$ *is* positive definite, *i.e. for all* $c_1, \dots, c_m \in \mathbb{C}$, $u_1, \dots, u_m \in \mathbb{R}^n$, $m \geqslant 1$:

$$\sum_{j,k=1}^{m} c_j \bar{c}_k \cdot \hat{\mu}(u_j - u_k) \geqslant 0.$$

Proof. Exercise. □

Proposition 5.2 (Uniqueness theorem)**.** *Let $\mu_1, \mu_2 \in \mathcal{M}_+^1(\mathbb{R}^n)$ with $\hat{\mu}_1 = \hat{\mu}_2$. Then $\mu_1 = \mu_2$.*

Proposition 5.3 (Bochner's theorem)**.** *Let $\varphi : \mathbb{R}^n \to \mathbb{C}$ be a continuous, positive definite function with $\varphi(0) = 1$. Then there exists one (and only one) $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$ with $\hat{\mu} = \varphi$.*

Proposition 5.4 (Lévy's continuity theorem)**.** *Let $(\mu_m)_{m\in\mathbb{N}}$ be a sequence in $\mathcal{M}_+^1(\mathbb{R}^n)$. Then*

(i) $\lim_{m\to\infty} \mu_m = \mu$ *weakly implies* $\lim_{m\to\infty} \hat{\mu}_m = \hat{\mu}$ *uniformly on every compact subset of* $\mathbb{R}^n$.

(ii) Conversely, if $(\hat{\mu}_m)_{m\in\mathbb{N}}$ converges pointwise to some function $\varphi : \mathbb{R}^n \to \mathbb{C}$ which is continuous in $u = 0$, then there exists a unique $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$ such that $\hat{\mu} = \varphi$ and $\lim_{m\to\infty} \mu_m = \mu$ weakly.

Proof. See Satz 15.23 in Klenke. □

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : \Omega \to \mathbb{R}^n$ be $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable. Let P_X $(:= P \circ X^{-1})$ be the distribution of X. Then

$$\varphi_X(u) := \hat{P}_X(u) = \int e^{i\langle u,y\rangle}\,P_X(dy) = \int e^{i\langle u,X\rangle} dP = E\left[e^{i\langle u,X\rangle}\right]$$

is said to be the *characteristic function of X*.

Remark 5.5. *$X_1, \ldots, X_n$ are independent if and only if*

$$\underbrace{\hat{P}_{(X_1,\ldots,X_n)}}_{=\varphi_{(X_1,\ldots,X_n)}}(u_1, \ldots, u_n) = \prod_{j=1}^{n} \underbrace{\hat{P}_{X_j}(u_j)}_{=\varphi_{X_j}(u_j)} \quad \big(= (P_{X_1} \otimes \hat{\cdots} \otimes P_{X_n})(u_1, \ldots, u_n)\big),$$

i.e.: $\hat{P}_{(X_1,\ldots,X_n)} = \prod_{j=1}^{n} \hat{P}_{X_j}$.

Proposition 5.6. *Let $X_1, \ldots, X_n$ be independent r.v., $\alpha \in \mathbb{R}$ and $S := \alpha \sum_{k=1}^{n} X_k$. Then for all $u \in \mathbb{R}$:*

$$\varphi_S(u) = \prod_{k=1}^{n} \varphi_{X_k}(\alpha u).$$

Proof.

$$\varphi_S(u) = \int e^{iuS} \, \mathrm{d}P = \int \prod_{k=1}^{n} e^{i\alpha u X_k} \, \mathrm{d}P \overset{\text{Indep.}}{=} \prod_{k=1}^{n} \int e^{i\alpha u X_k} \, \mathrm{d}P = \prod_{k=1}^{n} \varphi_{X_k}(\alpha u) \square$$

Proposition 5.7. *For all $u \in \mathbb{R}^n$:*

$$\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int e^{i\langle u,y\rangle} e^{-\frac{1}{2}|y|^2} \, \mathrm{d}y = e^{-\frac{1}{2}\|u\|^2}.$$

Proof. See Satz 15.12 in Klenke. □

Example 5.8. (i) $\hat{\delta}_a(u) = e^{iua}$.

(ii) Let $\mu := \sum_{i=1}^{\infty} \alpha_i \delta_{a_i}$ ($\alpha_i \geq 0$, $\sum_{i=1}^{\infty} \alpha_i = 1$). Then

$$\hat{\mu}(u) = \sum_{i=1}^{\infty} \alpha_i e^{iua_i}, \quad u \in \mathbb{R}^n.$$

Examples:

a) **Binomial distribution** $\beta_n^p = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} \, \delta_k$ Then for all $u \in \mathbb{R}$:

$$\hat{\beta}_n^p(u) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \cdot e^{iuk} = (q + pe^{iu})^n.$$

b) **Poisson distribution** $\pi_\alpha = \sum_{n=0}^{\infty} e^{-\alpha} \, \frac{\alpha^n}{n!} \, \delta_n$. Then for all $u \in \mathbb{R}$:

$$\hat{\pi}_\alpha(u) = e^{-\alpha} \sum_{n=0}^{\infty} \underbrace{\frac{\alpha^n}{n!} \cdot e^{iun}}_{=\frac{(\alpha e^{iu})^n}{n!}} = e^{\alpha(e^{iu}-1)}.$$

6 Central limit theorem

Definition 6.1. Let $X_1, X_2, \ldots \in \mathcal{L}^2$ be independent r.v., $S_n := X_1 + \cdots + X_n$ and

$$S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\operatorname{var}(S_n)}} \qquad \text{("standardization")}$$

(so that in particular $\mathbb{E}[S_n^*] = 0$ and $\operatorname{var}(S_n^* = 1)$). The sequence $X_1, X_2, \ldots$ of r.v. is said to have the *central limit property* (CLP), if

$$\lim_{n\to\infty} P_{S_n^*} = N(0,1) \quad \text{weakly,}$$

or equivalently (by the Portmanteau theorem)

$$\lim_{n\to\infty} P[S_n^* \leqslant b] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-\frac{x^2}{2}} \, \mathrm{d}x = \Phi(b) \quad \forall\, b \in \mathbb{R}.$$

Proposition 6.2. *(Central limit theorem) Let $X_1, X_2, \ldots \in \mathcal{L}^2$ be independent r.v., $\sigma_n^2 := \operatorname{var}(X_n) > 0$ and*

$$s_n := \Big(\sum_{k=1}^{n} \sigma_k^2\Big)^{\frac{1}{2}}.$$

Assume that $(X_n)_{n\in\mathbb{N}}$ satisfies Lindeberg's condition

$$\lim_{n\to\infty} \underbrace{\sum_{k=1}^{n} \int_{\left\{\frac{|X_k - \mathbb{E}[X_k]|}{s_n} \geqslant \varepsilon\right\}} \left(\frac{X_k - \mathbb{E}[X_k]}{s_n}\right)^2 \mathrm{d}P}_{=:L_n(\varepsilon)} = 0 \qquad \forall\, \varepsilon > 0\,. \tag{L}$$

Then $(X_n)_{n\in\mathbb{N}}$ has the CLP.

Remark 6.3. *(i) $(X_n)_{n\in\mathbb{N}}$ i.i.d. $\Rightarrow$ $(X_n)_{n\in\mathbb{N}}$ satisfies* (L).

Proof: *Let $m := \mathbb{E}[X_n]$, $\sigma^2 := \operatorname{var}(X_n)$. Then $s_n^2 = n\sigma^2$, so that*

$$L_n(\varepsilon) = \sigma^{-2} \int_{\{|X_1 - m| \geqslant \varepsilon\sqrt{n}\sigma\}} (X_1 - m)^2 \, \mathrm{d}P \xrightarrow[n\to\infty]{\text{Lebesgue}} 0.$$

(ii) The following stronger condition, known as Lyapunov's condition, *is often easier to check in applications:*

$$\exists\, \delta > 0 : \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta}\Big]}{s_n^{2+\delta}} = 0\,. \tag{Lya}$$

To see that Lyapunov's condition implies Lindeberg's condition note that for all $\varepsilon > 0$:

$$\begin{aligned} & \big|X_k - \mathbb{E}[X_k]\big| && \geqslant \varepsilon s_n \\ \Rightarrow \quad & \big|X_k - \mathbb{E}[X_k]\big|^{2+\delta} && \geqslant \big|X_k - \mathbb{E}[X_k]\big|^2 \cdot (\varepsilon s_n)^\delta \end{aligned}$$

and therefore

$$L_n(\varepsilon) \leqslant \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta}\Big].$$

(iii) *Let (X_n) be bounded and suppose that $s_n \to \infty$. Then (X_n) satisfies Lyapunov's condition for any $\delta > 0$, because*

$$\begin{aligned}
& |X_k| \leqslant \frac{\alpha}{2} \\
\Rightarrow \quad & \big|X_k - \mathbb{E}[X_k]\big| \leqslant \alpha \\
\Rightarrow \quad & \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta}\Big] \\
& \quad \leqslant \frac{\alpha^\delta}{s_n^{2+\delta}} \underbrace{\sum_{k=1}^n \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^2\Big]}_{=s_n^2} = \left(\frac{\alpha}{s_n}\right)^\delta.
\end{aligned}$$

Lemma 6.4. *Suppose that (X_n) satisfies Lindeberg's condition. Then*

$$\lim_{n\to\infty} \max_{1 \leqslant k \leqslant n} \frac{\sigma_k}{s_n} = 0\,. \tag{2.5}$$

Proof. For all $1 \leqslant k \leqslant n$

$$\begin{aligned}
\left(\frac{\sigma_k}{s_n}\right)^2 &= \int \left(\frac{X_k - \mathbb{E}[X_k]}{s_n}\right)^2 \mathrm{d}P \leqslant \int_{\left\{\frac{|X_k - \mathbb{E}[X_k]|}{s_n} \geqslant \varepsilon\right\}} \left(\frac{X_k - \mathbb{E}[X_k]}{s_n}\right)^2 \mathrm{d}P + \varepsilon^2 \\
&\leqslant L_n(\varepsilon) + \varepsilon^2.
\end{aligned}$$

□

The proof of Proposition 6.2 requires some further preparations.

Lemma 6.5. *For all $t \in \mathbb{R}$ and $n \in \mathbb{N}$:*

$$\left| e^{it} - 1 - \frac{it}{1!} - \frac{(it)^2}{2!} - \cdots - \frac{(it)^{n-1}}{(n-1)!} \right| \leqslant \frac{|t|^n}{n!}.$$

Proof. Define $f(t) := e^{it}$, then $f^{(k)}(t) = i^k e^{it}$, and the Taylor series expansion, applied to real and imaginary part, implies that

$$\left| e^{it} - 1 - \cdots - \frac{(it)^{n-1}}{(n-1)!} \right| = \big|R_n(t)\big|$$

with

$$\big|R_n(t)\big| = \left| \frac{1}{(n-1)!} \int_0^t (t-s)^{n-1} i^n e^{is} \, \mathrm{d}s \right| \leqslant \frac{1}{(n-1)!} \int_0^{|t|} s^{n-1} \, \mathrm{d}s = \frac{|t|^n}{n!}.$$

□

Proposition 6.6. *Let $X \in \mathcal{L}^2$. Then $\varphi_X(u) = \int e^{iuX}\, dP$ is two times continuously differentiable with*

$$\varphi_X'(u) = i \int X e^{iuX} \,\mathrm{d}P\,, \qquad \varphi_X''(u) = -\int X^2 e^{iuX}\,\mathrm{d}P\,.$$

In particular

$$\varphi_X'(0) = i\cdot \mathbb{E}[X], \quad \varphi_X''(0) = -\,\mathbb{E}[X^2], \quad |\varphi_X''| \leqslant \mathbb{E}[X^2].$$

Moreover, for all $u \in \mathbb{R}$

$$\varphi_X(u) = 1 + iu\cdot \mathbb{E}[X] + \frac{1}{2}\cdot \theta(u) u^2 \cdot \mathbb{E}[X^2]$$

with $\big|\theta(u)\big| \leqslant 1$ and $\theta(u) \in \mathbb{C}$.

Proof. Clearly,

$$(e^{iuX})' = iX \cdot e^{iuX}, \quad (e^{iuX})'' = -X^2 e^{iuX}, \quad |e^{iuX}| = 1\,.$$

Now, Lebesgue's dominated convergence theorem implies all assertions up to the last one. For the proof of the last assertion note that the previous lemma implies in the case $n = 2$ that

$$|e^{iuX} - 1 - iuX| \leqslant \frac{1}{2}\cdot u^2 X^2.$$

Integration w.r.t. P now implies that

$$\left|\int e^{iuX} - 1 - iuX \,\mathrm{d}P\right| = \big|\varphi_X(u) - 1 - iu\cdot \mathbb{E}[X]\big| \leqslant \frac{1}{2}\cdot u^2 \cdot \mathbb{E}[X^2]. \qquad \square$$

From now on assume that $X_1, X_2, \dots \in \mathcal{L}^2$ are independent and

$$\mathbb{E}[X_n] = 0 \;\forall\, n, \quad \sigma_n^2 := \operatorname{var}(X_n) > 0, \quad s_n = \Big(\sum_{k=1}^n \sigma_k^2\Big)^{\frac{1}{2}}.$$

Proposition 6.7. *Suppose that*

$$(a) \qquad \lim_{n\to\infty} \max_{1\le k\le n} \frac{\sigma_k}{s_n} = 0 \qquad \text{and}$$

$$(b) \qquad \lim_{n\to\infty} \sum_{k=1}^n \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right) = -\frac{1}{2}u^2 \qquad \forall\, u \in \mathbb{R}\,.$$

Then (X_n) has the CLP.

Proof. It is sufficient to show that

$$\lim_{n\to\infty} \prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right) = e^{-\frac{1}{2}u^2}. \tag{2.6}$$

because for $S_n^* = \frac{1}{s_n}\sum_{k=1}^n X_k$ we have that

$$\varphi_{S_n^*}(u) = \prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right),$$

and $\varphi_{S_n^*}(u) \xrightarrow{n\to\infty} e^{-\frac{1}{2}u^2} = \widehat{N(0,1)}(u)$ pointwise, implies by Lévy's continuity theorem that $\lim_{n\to\infty} P_{S_n^*} = N(0,1)$ weakly.

For the proof of (2.6) we need to show that for all $u \in \mathbb{R}$

$$\lim_{n\to\infty}\left(\prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right) - \underbrace{\prod_{k=1}^n \exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]}_{=\exp[\sum\cdots]\to\exp[-\frac{1}{2}u^2]}\right) = 0\,.$$

To this end fix $u \in \mathbb{R}$ and note that $|\varphi_{X_k}| \leqslant 1$, hence

$$\left|\exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]\right| = \exp\left[\mathsf{Re}\,\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right] \leqslant 1.$$

Note that for $a_1,\dots,a_n,b_1,\dots,b_n \in \big\{z \in \mathbb{C} \;\big|\; |z| \leqslant 1\big\}$

$$\begin{aligned}\Big|\prod_{k=1}^n a_k - \prod_{k=1}^n b_k\Big| &= \big|(a_1-b_1)\cdot a_2\cdots a_n + b_1\cdot(a_2-b_2)\cdot a_3\cdots a_n + \dots \\ &\quad + b_1\cdots b_{n-1}\cdot(a_n-b_n)\big| \\ &\leqslant \sum_{k=1}^n |a_k - b_k|.\end{aligned}$$

Consequently,

$$\begin{aligned}&\left|\prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right) - \prod_{k=1}^n \exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]\right| \\ &\quad\leqslant \sum_{k=1}^n\left|\varphi_{X_k}\left(\frac{u}{s_n}\right) - \exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]\right| =: D_n.\end{aligned}$$

If we define $z_k := \varphi_{X_k}(\frac{u}{s_n}) - 1$, we can write

$$D_n = \sum_{k=1}^n |z_k + 1 - e^{z_k}|.$$

Note that $\mathbb{E}[X_k] = 0$ and $\mathbb{E}[X_k^2] = \sigma_k^2$. The previous proposition now implies that for all k

$$|z_k| \leqslant \frac{1}{2}\left(\frac{u}{s_n}\right)^2 \sigma_k^2.$$

For $\varepsilon > 0$ we can find $\delta > 0$ such that

$$|z + 1 - e^z| \leqslant \varepsilon |z| \qquad \forall\, |z| < \delta\,.$$

Let $n_0 \in \mathbb{N}$ be such that for all $n \geqslant n_0$

$$\frac{u^2}{2}\left(\frac{\sigma_k}{s_n}\right)^2 < \delta$$

for all $n \geq n_0$. Then, for all $n \geqslant n_0$

$$D_n \leqslant \varepsilon \sum_{k=1}^n |z_k| \leqslant \varepsilon \frac{u^2}{2} \sum_{k=1}^n \frac{\sigma_k^2}{s_n^2} = \varepsilon \cdot \frac{u^2}{2}.$$

Consequently, $\lim_{n\to\infty} D_n = 0$. □

Proof of Proposition 6.2. It remains to show that Lindeberg's condition implies

$$\lim_{n\to\infty} \sum_{k=1}^n \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 \right) = -\frac{1}{2} \cdot u^2.$$

W.l.o.g. assume that $\mathbb{E}[X_n] = 0$ for all $n \in \mathbb{N}$. Let $u \in \mathbb{R}$, $n \in \mathbb{N}$, $1 \leqslant k \leqslant n$. Lemma 6.5 implies that

$$Y_k := \left| \exp\left(i \cdot \frac{u}{s_n} \cdot X_k \right) - 1 - \underbrace{i \cdot \frac{u}{s_n} \cdot X_k}_{\mathbb{E}[\dots]=0} + \frac{1}{2} \cdot \frac{u^2}{s_n^2} \cdot X_k^2 \right| \leqslant \frac{1}{6} \left| \frac{u}{s_n} \cdot X_k \right|^3,$$

and

$$Y_k \leqslant \left| \exp\left(i \cdot \frac{u}{s_n} \cdot X_k \right) - 1 - i \cdot \frac{u}{s_n} \cdot X_k \right| + \frac{1}{2} \cdot \frac{u^2}{s_n^2} \cdot X_k^2 \leqslant \frac{u^2}{s_n^2} \cdot X_k^2.$$

With these notations

$$\begin{aligned} \left| \sum_{k=1}^n \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 \right) + \frac{1}{2} \cdot u^2 \right| &= \left| \sum_{k=1}^n \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 + \frac{1}{2} \cdot \frac{u^2}{s_n^2} \cdot \sigma_k^2 \right) \right| \\ &\leqslant \sum_{k=1}^n \mathbb{E}[Y_k]. \end{aligned}$$

For $\delta > 0$

$$\begin{aligned} \mathbb{E}[Y_k] &= \int_{\{|X_k| \geqslant \delta s_n\}} Y_k \,\mathrm{d}P + \int_{\{|X_k| < \delta s_n\}} Y_k \,\mathrm{d}P \\ &\leqslant \frac{u^2}{s_n^2} \int_{\{|X_k| \geqslant \delta s_n\}} X_k^2 \,\mathrm{d}P + \frac{|u|^3}{6 s_n^3} \int_{\{|X_k| < \delta s_n\}} |X_k|^3 \,\mathrm{d}P. \end{aligned}$$

Note that

$$\frac{1}{s_n^3}\int_{\{|X_k|<\delta s_n\}}|X_k|^3\,\mathrm{d}P \leqslant \frac{\delta}{s_n^2}\int X_k^2\,\mathrm{d}P = \delta\cdot\frac{\sigma_k^2}{s_n^2},$$

so that for $\varepsilon>0$ and $\delta>0$ with $\frac{|u|^3}{6}\delta<\frac{\varepsilon}{2}$, we obtain that

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}[Y_k] &\leqslant u^2\sum_{k=1}^n\int_{\{|\frac{X_k}{s_n}|\geqslant\delta\}}\left(\frac{X_k}{s_n}\right)^2\mathrm{d}P + \frac{|u|^3}{6}\cdot\delta\underbrace{\sum_{k=1}^n\frac{\sigma_k^2}{s_n^2}}_{=1}\\ &\leqslant u^2L_n(\delta)+\frac{\varepsilon}{2}\,.\end{aligned}$$

Note that $u^2L_n(\delta)<\frac{\varepsilon}{2}$ for large n, so that

$$\lim_{n\to\infty}\sum_{k=1}^n E[Y_k]=0\,,$$

and thus

$$\lim_{n\to\infty}\left(\sum_{k=1}^n\left[\varphi_{X_k}\left(\frac{u}{s_n}\right)-1\right]+\frac{1}{2}\cdot u^2\right)=0.$$

Now the assertion follows from Proposition 6.7. □

Example 6.8 (Applications). (i) **"Ruin probability"**

Consider a portfolio of n contracts of a risc insurance (e.g. car insurance, fire insurance, health insurance, ...). Let $X_i\geqslant 0$ be the claim size (or claim severity) of the ith contract, $1\leqslant i\leqslant n$. We assume that $X_1,\ldots,X_n\in\mathcal{L}^2$ are i.i.d. with $m:=\mathbb{E}[X_i]$ and $\sigma^2:=\mathrm{var}(X_i)$.

Suppose the insurance holder has to pay he following premium

$$\begin{aligned}\Pi &:= m+\lambda\sigma^2\\ &= \text{average claim size} + \text{safety loading.}\end{aligned}$$

After some fixed amount of time:

Income: $n\Pi$

Expenditures: $S_n=\sum_{i=1}^n X_i$.

Suppose that K is the initial capital of the insurance company. What is the probability $P(R)$, where

$R:=\{S_n>K+n\Pi\}$ denotes the ruin ?

We assume here that:

- No interest rate.
- Payments due only at the end of the time period.

Let

$$S_n^* := \frac{S_n - nm}{\sqrt{n}\sigma}.$$

The central limit theorem implies for large n that $S_n^* \sim N(0,1)$, so that

$$\begin{aligned} P(R) &= P\left[S_n^* > \frac{K + n\Pi - nm}{\sqrt{n}\sigma}\right] = P\left[S_n^* > \frac{K + n\lambda\sigma^2}{\sqrt{n}\sigma}\right] \\ &\overset{\text{CLT}}{\approx} 1 - \Phi\Bigg(\underbrace{\frac{K + n\lambda\sigma^2}{\sqrt{n}\sigma}}_{\xrightarrow{n\to\infty}\infty}\Bigg), \end{aligned}$$

where Φ denotes the distribution of the standard normal distribution. Note that the ruin probability decreases with an increasing number of contracts.

Example
Assume that $n = 2000$, $\sigma = 60$, $\lambda = 0.5‰$.

(a) $K = 0 \Rightarrow P(R) \approx 1 - \Phi(1.342) \approx 9\%$.

(b) $K = 1500 \Rightarrow P(R) \approx 3\%$.

How large do we have to choose n in order to let the probability of ruin $P(R)$ fall below $1‰$?
Answer: $\Phi(\ldots) \geqslant 0.999$, hence $n \geqslant 10\,611$.

(ii) **Stirling's formula**
Remark: Stirling proved the following formula

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \tag{2.7}$$

in the year 1730 and De Moivre used it in his proof of the CLT for Bernoulli experiments.

Conversely, in 1977, Weng provided an independent proof of the formula, using the CLT (note that we did not use Stirling's formula in our proof of the CLT). Here is Weng's proof:

Let $X_1, X_2, \ldots$ be i.i.d. with distribution π_1, i.e.,

$$P_{X_n} = e^{-1} \sum_{k=0}^{\infty} \frac{1}{k!} \delta_k\,.$$

Then $S_n := X_1 + \cdots + X_n$ has Poisson distribution π_n, i.e.,

$$P_{S_n} = e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} \delta_k,$$

and in particular $\mathbb{E}[S_n] = \mathrm{var}(S_n) = n$. As usual, define

$$S_n^* := \frac{S_n - n}{\sqrt{n}},$$

so that $S_n^* = t_n \circ S_n$ for $t_n(x) := \frac{x-n}{\sqrt{n}}$. Then

$$\int f \,\mathrm{d}P_{S_n^*} = \mathbb{E}\big[f(S_n^*)\big] = \mathbb{E}\big[(f \circ t_n)(S_n)\big] = \int f \circ t_n \,\mathrm{d}\underbrace{P_{S_n}}_{=\pi_n}$$

In particular, for

$$f_\infty(x) := x^- \quad \big(= (-x) \vee 0\big)$$

it follows that

$$\begin{aligned}
\int f_\infty \,\mathrm{d}P_{S_n^*} &= \int_{\mathbb{R}} \underbrace{f_\infty\Big(\frac{x-n}{\sqrt{n}}\Big)}_{\begin{cases} = 0 \text{ if } x>n \\ = \frac{n-x}{\sqrt{n}} \text{ if } x \leqslant n\end{cases}} \pi_n(\mathrm{d}x) = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \cdot \underbrace{\frac{n-k}{\sqrt{n}}}_{=f_\infty(k)} \\
&= \frac{e^{-n}}{\sqrt{n}} \cdot \bigg(n + \sum_{k=1}^{n} \frac{n^k(n-k)}{k!}\bigg) \\
&= \frac{e^{-n}}{\sqrt{n}} \cdot \bigg(n + \underbrace{\sum_{k=1}^{n} \frac{n^{k+1}}{k!} - \frac{n^k}{(k-1)!}}_{=\frac{n^{n+1}}{n!} - \frac{n^1}{0!}}\bigg) = \frac{e^{-n} \cdot n^{n+\frac{1}{2}}}{n!}.
\end{aligned}$$

Moreover,

$$\int f_\infty \,\mathrm{d}N(0,1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} (-x) \cdot e^{-\frac{x^2}{2}} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \bigg|_{-\infty}^{0} = \frac{1}{\sqrt{2\pi}}.$$

Hence, Stirling's formula (2.7) would follow, once we have shown that

$$\int f_\infty \,\mathrm{d}P_{S_n^*} \xrightarrow{n\to\infty} \int f_\infty \,\mathrm{d}N(0,1). \tag{2.8}$$

Note that this is not implied by the weak convergence in the CLT since f_∞ is continuous but unbounded. Hence, we consider for given $m \in \mathbb{N}$

$$f_m := f_\infty \wedge m \quad \big(\in C_b(\mathbb{R})\big).$$

The CLT now implies that

$$\int f_m \,\mathrm{d}P_{S_n^*} \xrightarrow{n\to\infty} \int f_m \,\mathrm{d}N(0,1).$$

Define $g_m := f_\infty - f_m \quad (\geq 0)$. (2.8) then follows from a "3ε-argument", once we have shown that

$$(0 \leqslant) \quad \int g_m \,\mathrm{d}P_{S_n^*} \leqslant \frac{1}{m} \qquad \forall\, m,$$

$$(0 \leqslant) \quad \int g_m \,\mathrm{d}N(0,1) \leqslant \frac{1}{m} \qquad \forall m.$$

The first inequality follows from

$$\begin{aligned}\int g_m \,\mathrm{d}P_{S_n^*} &= \int_{]-\infty,-m[} \big(|x| - m\big) \,\mathrm{d}P_{S_n^*} \leqslant \int_{]-\infty,-m]} |x| \,\mathrm{d}P_{S_n^*} \\ &\overset{\frac{|x|}{m} \geqslant 1}{\leqslant} \frac{1}{m} \int_{]-\infty,-m]} x^2 \,\mathrm{d}P_{S_n^*} \leqslant \frac{1}{m} \cdot \underbrace{\mathrm{var}(S_n^*)}_{=1},\end{aligned}$$

the second inequality can be shown similarly.

3 Conditional probabilities

1 Elementary definitions

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 1.1. Let $B \in \mathcal{A}$ with $P(B) > 0$. Then

$$P[A \,|\, B] := \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{A},$$

is said to be the *conditional probability of A given B*. In the case $P(B) = 0$ we simply define $P[A \,|\, B] := 0$. The probability measure

$$P_B := P[\,\cdot\,|\, B]$$

on $(\Omega, \mathcal{A})$ is said to be the *conditional distribution given B*.

Remark 1.2. *(i) $P(A)$ is called the a priori probability of A.*

$P[A \,|\, B]$ is called the a posteriori probability of A, given the information that B occurred.

(ii) In the case of Laplace experiments

$$P[A \,|\, B] = \frac{|A \cap B|}{|B|} = \textit{fraction of all outcomes in } A \textit{ that are contained in } B.$$

(iii) If A and B are disjoint (hence $A \cap B = \emptyset$), then $P[A \,|\, B] = 0$.

(iv) If A and B are independent, then

$$P[A \,|\, B] = \frac{P(A) \cdot P(B)}{P(B)} = P(A).$$

Example 1.3. (i) Suppose that a family has two children. Consider the following two events: $B :=$ "at least one boy" and $A :=$ "two boys". Then $P[A \,|\, B] = \frac{1}{3}$, because

$$\Omega = \{(J, J), (M, J), (J, M), (M, M)\},$$
$$P = \text{uniform distribution},$$

and thus

$$P[A \,|\, B] = \frac{|A \cap B|}{|B|} = \frac{1}{3}.$$

(ii) Let X_1, X_2 be independent r.v. with Poisson distribution with parameters λ_1, λ_2. Then

$$P[X_1 = k \,|\, X_1 + X_2 = n] = \begin{cases} 0 & \text{if } k > n \\ ? & \text{if } 0 \leqslant k \leqslant n. \end{cases}$$

According to Example 4.7 $X_1 + X_2$ has Poisson distribution with parameter $\lambda := \lambda_1 + \lambda_2$. Consequently,

$$\begin{aligned} P[X_1 = k \,|\, X_1 + X_2 = n] &= \frac{P[X_1 = k,\ X_2 = n-k]}{P[X_1 + X_2 = n]} \\ &= \frac{e^{-\lambda_1} \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-\lambda} \frac{\lambda^n}{n!}} = \binom{n}{k} \cdot \left(\frac{\lambda_1}{\lambda}\right)^k \left(\frac{\lambda_2}{\lambda}\right)^{n-k}, \end{aligned}$$

i.e., $P[\,\cdot \mid X_1 + X_2 = n]$ is the binomial distribution with parameters n and $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$.

(iii) Consider n independent 0-1-experiments $X_1, \dots, X_n$ with success probability $p \in]0,1[$. Let

$$S_n := X_1 + \ldots + X_n$$

and

$$\begin{array}{llll} X_i : & \Omega := \{0,1\}^n & \to & \{0,1\}, \\ & (x_1, \dots, x_n) & \mapsto & x_i. \end{array}$$

For given $(x_1, \dots, x_n) \in \{0,1\}^n$ and fixed $k \in \{0, \dots, n\}$

$$\begin{aligned} & P[X_1 = x_1, \dots, X_n = x_n \,|\, S_n = k] \\ &= \begin{cases} 0 & \text{if } \sum_i x_i \neq k \\ \frac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \binom{n}{k}^{-1} & \text{otherwise} \end{cases} \end{aligned}$$

It follows that the conditional distribution $P[\,\cdot \,|\, S_n = k]$ is the uniform distribution on

$$\Omega_k := \left\{ (x_1, \dots, x_n) \,\middle|\, \sum_{i=1}^n x_i = k \right\}.$$

Proposition 1.4. *(**Formula for total probability**) Let $B_1, \dots, B_n$ be disjoint, $B_i \in \mathcal{A}$ $\forall\, 1 \leq i \leq n$. Then for all $A \subset \bigcup_{i=1}^n B_i$, $A \in \mathcal{A}$:*

$$P(A) = \sum_{i=1}^n P[A \,|\, B_i] \cdot P(B_i).$$

Proof. Clearly, $A = \dot\cup_{i\leqslant n}(A\cap B_i)$. Consequently,

$$P(A) = \sum_{i=1}^{n} P(A\cap B_i) = \sum_{i=1}^{n} P(A\cap B_i) = \sum_{i=1}^{n} P(A\,|\,B_i)P(B_i)\,.$$

□

Example 1.5. (Simpson's paradox)
Consider applications of male (M) and female (W) students at a university in the United States

	Applications	accepted	
M	2084	1036	$P[A\,|\,M] \approx 0.49$
W	1067	349	$P[A\,|\,W] \approx 0.33$

Is this an example for discrimination of female students? A closer look to the biggest four faculties $B_1, \dots, B_4$:

	male			female		
	Appl.	acc.	$P_M[A\,|\,B_i]$	Appl.	acc.	$P_W[A\,|\,B_i]$
B_1	826	551	0.67	108	89	0.82
B_2	560	353	0.63	25	17	0.68
B_3	325	110	0.34	593	219	0.37
B_4	373	22	0.06	341	24	0.07
	2084	1036		1067	349	

It follows that for all four faculties the probability of being accepted was higher for female students than it was for male students:

$$P_M[A\,|\,B_i] < P_W[A\,|\,B_i].$$

Nevertheless, the preference turns into its opposite if looking at the total probability of admission:

$$P_W(A) := P[A\,|\,W] = \sum_{i=1}^{4} P_W[A\,|\,B_i]\cdot P_W(B_i)$$

$$< P_M(A) := P[A\,|\,M] = \sum_{i=1}^{4} P_M[A\,|\,B_i]\cdot P_M(B_i).$$

For an explanation consider the distributions of applications:

$$P_M(B_1) = \frac{|B_1\cap M|}{|M|} = \frac{826}{2084} \approx \frac{4}{10}, \quad P_W(B_1) = \frac{|B_1\cap W|}{|W|} = \frac{108}{1067} \approx \frac{1}{10},$$

etc. and observe that male students mainly applied at faculties with a high probability of admission, whereas female students mainly applied at faculties with a low probability of admission.

Proposition 1.6 (Bayes' theorem)**.** *Let $B_1, \ldots, B_n \in \mathcal{A}$ be disjoint with $P(B_i) > 0$ for $i = 1, \ldots, n$. Let $A \in \mathcal{A}$, $A \subset \bigcup_{i=1}^n B_i$ with $P(A) > 0$. Then:*

$$P[B_i \,|\, A] = \frac{P[A \,|\, B_i] \cdot P(B_i)}{\sum_{j=1}^{n} P[A \,|\, B_j] \cdot P(B_j)}.$$

Proof.

$$P[B_i \,|\, A] = \frac{P(A \cap B_i)}{P(A)} \overset{1.4}{=} \frac{P[A \mid B_i] \cdot P(B_i)}{\sum_{j=1}^{n} P[A \mid B_j] \cdot P(B_j)}. \qquad \square$$

Example 1.7 (A posteriori probabilities in medical tests)**.** Suppose that one out of 145 persons of the same age have the disease K, i.e. the a priori probability of having K is $P[\mathsf{K}] = \frac{1}{145}$.

Suppose now that a medical test for K is given which detects K in 96 % of all cases, i.e.

$$P[\texttt{positive} \,|\, K] = 0.96\,.$$

However, the test also is positive in 6% af the cases, where the person does not have K, i.e.

$$P[\texttt{positive} \,|\, K^c] = 0.06\,.$$

Suppose now that the test is positive. What is the a posteriori probability of actually having K?

So we are interested in the conditional probability $P[K \,|\, \texttt{positive}]$:

$$\begin{aligned} P[K \,|\, \texttt{positive}] &\overset{1.6}{=} \frac{P[\texttt{positive} \,|\, K] \cdot P[K]}{P[\texttt{positive} \,|\, K] \cdot P[K] + P[\texttt{positive} \,|\, K^c] \cdot P[K^c]} \\ &= \frac{0.96 \cdot \frac{1}{145}}{0.96 \cdot \frac{1}{145} + 0.06 \cdot \frac{144}{145}} = \frac{1}{1 + \frac{6}{96} \cdot 144} = \frac{1}{10}. \end{aligned}$$

Note: in only one out of ten cases, a person with a positive result actually has K.

Another conditional probability of interest in this context is the probability of not having K, once the test is negative, i.e., $P[K^c \,|\, \texttt{negative}]$:

$$\begin{aligned} P[K^c \,|\, \texttt{negative}] &= \frac{P[\texttt{negative} \,|\, K^c] \cdot P[K^c]}{P[\texttt{negative} \,|\, K] \cdot P[K] + P[\texttt{negative} \,|\, K^c] \cdot P[K^c]} \\ &= \frac{0.94 \cdot \frac{144}{145}}{0.04 \cdot \frac{1}{145} + 0.94 \cdot \frac{144}{145}} = \frac{94 \cdot 144}{4 + 94 \cdot 144} \approx 0.9997. \end{aligned}$$

Note: The two conditional probabilities interchange, if the a priori probability of not having K is low (e.g. $\frac{1}{145}$). If the risc of having K is high and one wants to test whether or not having K, the a posteriori probability of not having K, given that the test was negative, is only 0.1.

Example 1.8 (computing total probabilities with conditional probabilities)**.** Let S be a finite set, $\Omega := S^{n+1}$, $n \in \mathbb{N}$, and P be a probability measure on Ω. Let $X_i : \Omega \to S$, $i = 0, \dots, n$, be the canonical projections $X_i(\omega) := x_i$ for $\omega = (x_0, \dots, x_n)$.

If we interpret $0, 1, \dots, n$ as time points, then $(X_i)_{0 \leqslant i \leqslant n}$ may be seen as a *stochastic process* and $\big(X_0(\omega), \dots, X_n(\omega)\big)$ is said to be a *sample path* (or a *trajectory*) of the process.

For all $\omega \in \Omega$ we either have $P(\{\omega\}) = 0$ or

$$\begin{aligned}
P(\{\omega\}) &= P[X_0 = x_0, \dots, X_n = x_n] \\
&= P[X_0 = x_0, \dots, X_{n-1} = x_{n-1}] \\
&\qquad \cdot P[X_n = x_n \,|\, X_0 = x_0, \dots, X_{n-1} = x_{n-1}] \\
&\vdots \\
&= P[X_0 = x_0] \\
&\qquad \cdot P[X_1 = x_1 \,|\, X_0 = x_0] \\
&\qquad \cdot P[X_2 = x_2 \,|\, X_0 = x_0, X_1 = x_1] \\
&\qquad \cdots \\
&\qquad \cdot P[X_n = x_n \,|\, X_0 = x_0, \dots, X_{n-1} = x_{n-1}].
\end{aligned}$$

Note: $P(\{\omega\}) \neq 0$ implies $P[X_0 = x_0, \dots, X_k = x_k] \neq 0$ for all $k \in \{0, \dots, n\}$.

Conclusion: A probability measure P on Ω is uniquely determined by the following:

Initial distribution: $\mu := P \circ X_0^{-1}$

Transition probabilities: the conditional distributions

$$P[X_k = x_k \,|\, X_0 = x_0, \dots, X_{k-1} = x_{k-1}]$$

for any $k \in \{1, \dots, n\}$ and $(x_0, \dots, x_k) \in S^{(k+1)}$.

Existence of P for given initial distribution and given transition probabilities is shown in Section 3.3.

Example 1.9. A stochastic process is called a *Markov chain*, if $P[X_k = x_k \,|\, X_0 = x_0, \dots, X_{k-1} = x_{k-1}] = P[X_k = x_k \,|\, X_{k-1} = x_{k-1}]$, i.e., if the transition probabilities for X_k only depend on X_{k-1}.

If we denote by X_{k-1} the "present", by X_k the "future" and by "$X_0, \dots, X_{k-2}$" the past, then we can state the Markov property as: given the "present", the "future" of the Markov chain is independent of the "past".

2 Transition probabilities and Fubini's theorem

Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be measurable spaces.

Definition 2.1. A mapping

$$\begin{aligned} K: \quad S_1 \times \mathcal{S}_2 &\to [0,1] \\ (x_1, A_2) &\mapsto K(x_1, A_2) \end{aligned}$$

is said to be a *transition probabilities* (from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$), if

(i) $\forall x_1 \in S_1$: $K(x_1, \cdot\,)$ is a probability measure on $(S_2, \mathcal{S}_2)$.

(ii) $\forall A_2 \in \mathcal{S}_2$: $K(\,\cdot\,, A_2)$ is $\mathcal{S}_1$-measurable.

Example 2.2. (i) For given probability measure μ on $(S_2, \mathcal{S}_2)$ define

$$K(x_1, \cdot\,) := \mu \qquad \forall\, x_1 \in S_1 \ \text{ no coupling!}$$

(ii) Let $T : S_1 \to S_2$ be a $\mathcal{S}_1/\mathcal{S}_2$-measurable mapping, and

$$K(x_1, \cdot\,) := \delta_{T(x_1)} \qquad \forall x_1 \in S_1\,.$$

(iii) **Stochastic matrices** Let S_1, S_2 be countable and $\mathcal{S}_i = \mathcal{P}(S_i)$, $i = 1, 2$. In this case, any transition probability from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$ is given by

$$K(x_1, x_2) := K\big(x_1, \{x_2\}\big), \quad x_1 \in S_1,\ x_2 \in S_2,$$

where $K : S_1 \times \mathcal{S}_2 \to [0,1]$ is a mapping, such that for all $x_1 \in S_1$ $\sum_{x_2 \in S_2} K(x_1, x_2) = 1$. Consequently, K can be identified with a stochastic matrix, or a transition matrix, i.e. a matrix with nonnegative entries and row sums equal to one.

Example 2.3. (i) **Transition probabilities of the random walk on $\mathbb{Z}^d$**

$S_1 = S_2 = S := \mathbb{Z}^d$ with $\mathcal{S} := \mathcal{P}(\mathbb{Z}^d)$

$$K(x, \cdot\,) := \frac{1}{2d} \sum_{y \in N(x)} \delta_y \,, x \in \mathbb{Z}^d\,,$$

with

$$N(x) := \big\{ y \in \mathbb{Z}^d \;\big|\; \|x - y\| = 1 \big\}$$

denotes the set of nearest neighbours of x.

(ii) **Ehrenfest model** Consider a box containing N balls. The box is divided into two parts ("left" and "right"). A ball is selected randomly and put into the other half.

"microscopic level" the state space is $S := \{0,1\}^N$ with $x = (x_1, \dots, x_N) \in S$ defined by

$$x_i := \begin{cases} 1 & \text{if the } i^{th} \text{ ball is contained in the "left" half} \\ 0 & \text{if the } i^{th} \text{ ball is contained in the "right" half} \end{cases}$$

the transition probability is given by

$$K(x, \cdot\,) := \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_1, \dots, x_{i-1}, 1 - x_i, x_{i+1}, \dots, x_N)}.$$

"macroscopic level" the state space is $S := \{0, \ldots, N\}$, where $j \in S$ denotes the number of balls contained in the left half. The transition probabilities are given by

$$K(j, \cdot\,) := \frac{N-j}{N} \cdot \delta_{j+1} + \frac{j}{N} \cdot \delta_{j-1}.$$

(iii) **Transition probabilities of the Ornstein-Uhlenbeck process** $S = S_1 = S_2 = \mathbb{R}$, $K(x, \cdot\,) := N(\alpha x, \sigma^2)$ with $\alpha \in \mathbb{R}$, $\sigma^2 > 0$.

We now turn to Fubini's theorem. To this end, let μ_1 be a probability measure on $(S_1, \mathcal{S}_1)$ and $K(\,\cdot\,,\cdot\,)$ be a transition probability from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$.
Our aim is to construct a probability measure P $(:= \mu_1 \otimes K)$ on the product space $(\Omega, \mathcal{A})$, where

$$\Omega := S_1 \times S_2$$

$$\mathcal{A} := \mathcal{S}_1 \otimes \mathcal{S}_2 := \sigma(X_1, X_2) \overset{!}{=} \sigma\big(\{A_1 \times A_2 \,|\, A_1 \in \mathcal{S}_1,\ A_2 \in \mathcal{S}_2\}\big)\,,$$

and

$$\begin{array}{rrcl} X_i: & \Omega = S_1 \times S_2 & \to & S_i, \quad i = 1, 2, \\ & (x_1, x_2) & \mapsto & x_i, \end{array}$$

satisfying

$$P(A_1 \times A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(\mathrm{d}x_1)$$

for all $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$.

Proposition 2.4 (Fubini)**.** *Let μ_1 be a probability measure on $(S_1, \mathcal{S}_1)$, K a transition probability from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$, and*

$$\Omega := S_1 \times S_2, \tag{3.1}$$

$$\mathcal{A} := \sigma\big(\{A_1 \times A_2 \,|\, A_i \in \mathcal{S}_i\}\big) =: \mathcal{S}_1 \otimes \mathcal{S}_2\,. \tag{3.2}$$

Then there exists a probability measure P $(=: \mu_1 \otimes K)$ on $(\Omega, \mathcal{A})$, such that for all $\mathcal{A}$-measurable functions $f \geqslant 0$

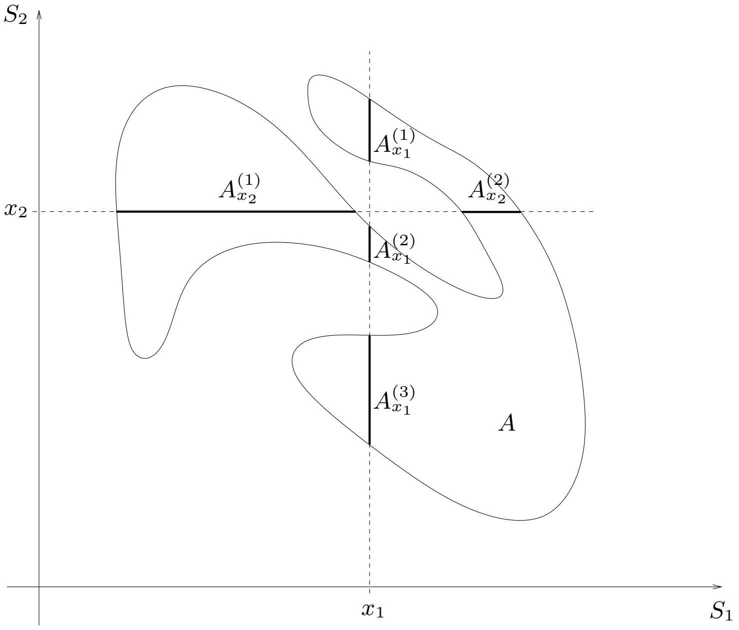
$$\int_\Omega f\, \mathrm{d}P = \int \left(\int f(x_1, x_2)\, K(x_1, \mathrm{d}x_2) \right) \mu_1(\mathrm{d}x_1), \tag{3.3}$$

in particular, for all $A \in \mathcal{A}$

$$P(A) = \int K(x_1, A_{x_1})\, \mu_1(\mathrm{d}x_1)\,. \tag{3.4}$$

Here

$$A_{x_1} = \{x_2 \in S_2 \,|\, (x_1, x_2) \in A\}$$



Note

$$A_{x_1} = A_{x_1}^{(1)} \cup A_{x_1}^{(2)} \cup A_{x_1}^{(3)} \quad \text{and} \quad A_{x_2} = A_{x_2}^{(1)} \cup A_{x_2}^{(2)}.$$

is called the section of A by x_1*. In particular, for* $A_1 \in \mathcal{S}_1$*,* $A_2 \in \mathcal{S}_2$*:*

$$P(A_1 \times A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(\mathrm{d}x_1). \tag{3.5}$$

P *is uniquely determined by* (3.5)*.*

Proof. Uniqueness: Clearly, the collection of cylindrical sets $A_1 \times A_2$ with $A_i \in \mathcal{S}_i$ is stable under intersections and generates $\mathcal{A}$, so that the uniqueness now follows from Proposition 1.11.5.

Existence: For given $x_1 \in S_1$ let

$$\varphi_{x_1}(x_2) := (x_1, x_2)\,.$$

$\varphi_{x_1} : S_2 \to \Omega$ is measurable, because for $A_1 \in \mathcal{S}_1$, $A_2 \in \mathcal{S}_2$

$$\varphi_{x_1}^{-1}(A_1 \times A_2) = \begin{cases} \emptyset & \text{if } x_1 \notin A_1 \\ A_2 & \text{if } x_1 \in A_1\ . \end{cases}$$

It follows that for any $f : \Omega \to \mathbb{R}$ $\mathcal{A}$-measurable and any $x_1 \in S_1$, the mapping

$$f_{x_1} := f \circ \varphi_{x_1} : S_2 \to \mathbb{R}\,, x_2 \mapsto f(x_1, x_2)$$

is $\mathcal{S}_2/\mathcal{B}(\mathbb{R})$-measurable.

Suppose now that $f \geqslant 0$ or bounded. Then

$$x_1 \mapsto \int f(x_1, x_2)\, K(x_1, \mathrm{d}x_2) \left(= \int f_{x_1}(x_2)\, K(x_1, \mathrm{d}x_2)\right) \tag{3.6}$$

is well-defined.

We will show in the following that this function is $\mathcal{S}_1$-measurable. We will prove the assertion for $f = 1_A$, $A \in \mathcal{A}$ first. For general f the measurability then follows by measure-theoretic induction.

Note that for $f = 1_A$ we have that

$$\int \underbrace{1_A(x_1, x_2)}_{=1_{A_{x_1}}(x_2)}\, K(x_1, \mathrm{d}x_2) = K(x_1, A_{x_1}).$$

Hence, in the following we consider

$$\mathcal{D} := \big\{A \in \mathcal{A} \mid x_1 \mapsto K(x_1, A_{x_1})\ \mathcal{S}_1\text{-measurable}\big\}.$$

$\mathcal{D}$ is a Dynkin system (!) and contains all cylindrical sets $A = A_1 \times A_2$ with $A_i \in \mathcal{S}_i$, because

$$K\big(x_1, (A_1 \times A_2)_{x_1}\big) = 1_{A_1}(x_1) \cdot K(x_1, A_2)\,.$$

Since measurable cylindrical sets are stable under intersections, we conclude that $\mathcal{D} = \mathcal{A}$.

It follows that for all nonnegative or bounded $\mathcal{A}$-measurable functions $f : \Omega \to \mathbb{R}$, the integral

$$\int \left(\int f(x_1, x_2)\, K(x_1, \mathrm{d}x_2)\right) \mu(\mathrm{d}x_1)$$

is well-defined.

For all $A \in \mathcal{A}$ we can now define

$$P(A) := \int \left(\int \underbrace{1_A(x_1, x_2)}_{=1_{A_{x_1}}(x_2)}\, K(x_1, \mathrm{d}x_2)\right) \mu(\mathrm{d}x_1) = \int K(x_1, A_{x_1})\, \mu(\mathrm{d}x_1).$$

P is a probability measure on $(\Omega, \mathcal{A})$, because

$$P(\Omega) = \int K(x_1, S_2)\, \mu(\mathrm{d}x_1) = \int 1\, \mu(\mathrm{d}x_1) = 1.$$

For the proof of the σ-additivity, let $A_1, A_2, \dots$ be pairwise disjoint subsets in $\mathcal{A}$. It follows that for all $x_1 \in S_1$ the subsets $(A_1)_{x_1}, (A_2)_{x_1}, \dots$ are pairwise disjoint too, hence

$$\begin{aligned} P\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) &= \int K\bigg(x_1, \Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big)_{x_1}\bigg)\ \mu(\mathrm{d}x_1) \\ &= \int \sum_{n=1}^{\infty} K\big(x_1, (A_n)_{x_1}\big)\ \mu(\mathrm{d}x_1) \\ &= \sum_{n=1}^{\infty} \int K\big(x_1, (A_n)_{x_1}\big)\ \mu(\mathrm{d}x_1) = \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

In the second equality we used that $K(x_1, \cdot)$ is a probability measure for all x_1 and in the third equality we used monotone integration.

Finally, (3.3) follows from measure-theoretic induction. □

2.1 Examples and Applications

Remark 2.5. *The classical Fubini theorem is a particular case of Proposition 2.4: $K(x_1, \cdot\,) = \mu_2$. In this case, the measure $\mu_1 \otimes K$, constructed in Fubini's theorem, is called the* product measure *of μ_1 and μ_2 and is denoted by $\mu_1 \otimes \mu_2$. Moreover, in this case*

$$\int f \,\mathrm{d}P = \int \bigg(\int f(x_1, x_2)\ \mu_2(\mathrm{d}x_2)\bigg)\ \mu_1(\mathrm{d}x_1).$$

Remark 2.6 (Marginal distributions)**.** *Let $X_i : \Omega \to S_i$, $i = 1, 2$, be the natural projections $X_i\big((x_1, x_2)\big) := x_i$. The distributions of X_i under the measure $\mu_1 \otimes K$ are called the* marginal distributions *and they are given by*

$$\begin{aligned} (P \circ X_1^{-1})(A_1) &= P[X_1 \in A_1] = P(A_1 \times S_2) \\ &= \int_{A_1} \underbrace{K(x_1, S_2)}_{=1}\ \mu_1(\mathrm{d}x_1) = \mu_1(A_1) \end{aligned}$$

and

$$\begin{aligned} (P \circ X_2^{-1})(A_2) &= P[X_2 \in A_2] = P(S_1 \times A_2) \\ &= \int K(x_1, A_2)\ \mu_1(\mathrm{d}x_1) =: (\mu_1 K)(A_2). \end{aligned}$$

So, the marginal distributions are

$$P \circ X_1^{-1} = \mu_1 \qquad P \circ X_2^{-1} = \mu_1 K\,.$$

Definition 2.7. Let $S_1 = S_2 = S$ and $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$. A probability measure μ on $(S, \mathcal{S})$ is said to be an *equilibrium distribution for K* (or *invariant distribution under K*) if $\mu = \mu K$.

Example 2.8. (i) **Ehrenfest model** (macroscopic) Let $S = \{0, 1, \ldots, N\}$ and

$$K(y, \cdot\,) = \frac{y}{N} \cdot \delta_{y-1} + \frac{N-y}{N} \cdot \delta_{y+1}.$$

In this case, the binomial distribution μ with parameter $N, \frac{1}{2}$ is an equilibirum distribution, because

$$\begin{aligned}
(\mu K)(\{x\}) &= \sum_{y \in S} \mu(\{y\}) \cdot K(y, x) \\
&= \mu(\{x+1\}) \cdot \frac{x+1}{N} + \mu(\{x-1\}) \cdot \frac{N-(x-1)}{N} \\
&= 2^{-N} \binom{N}{x+1} \cdot \frac{x+1}{N} + 2^{-N} \binom{N}{x-1} \cdot \frac{N-(x-1)}{N} \\
&= 2^{-N} \left[\binom{N-1}{x} + \binom{N-1}{x-1} \right] = 2^{-N} \cdot \binom{N}{x} = \mu(\{x\}).
\end{aligned}$$

(ii) **Ornstein-Uhlenbeck process** Let $S = \mathbb{R}$ and $K(x, \cdot\,) = N(\alpha x, \sigma^2)$ with $|\alpha| < 1$. Then

$$\mu = N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$$

is an equilibrium distribution. (Exercise.)

We now turn to the converse problem: Given a probability measure P on the product space $(\Omega, \mathcal{A})$. Can we "disintegrate" P, i.e., can we find a probability measure μ_1 on $(S_1, \mathcal{S}_1)$ and a transition probability from S_1 to S_2 such that

$$P = \mu_1 \otimes K\,?$$

Answer: In most cases - yes, e.g. if S_1 and S_2 are Polish spaces (i.e., a topological space having a countable basis, whose topology is induced by some complete metric), using conditional expectations (see below).

Example 2.9. In the particular case, when S_1 is countable (and $\mathcal{S}_1 = \mathcal{P}(S_1)$), we can disintegrate P explicitly as follows: Necessarily, μ_1 has to be the distribution of the projection X_1 onto the first coordinate. To define the kernel K, let ν be any probability measure on $(S_2, \mathcal{S}_2)$ and define

$$K(x_1, A_2) := \begin{cases} P[X_2 \in A_2 \,|\, X_1 = x_1] & \text{if } \underbrace{\mu_1(\{x_1\})}_{=P[X_1 = x_1]} > 0 \\ \nu(A_2) & \text{if } \mu_1(\{x_1\}) = 0. \end{cases}$$

Then

$$\begin{aligned}
P(A_1 \times A_2) = P[X_1 \in A_1,\ X_2 \in A_2] &= \sum_{x_1 \in A_1} P[X_1 = x_1,\ X_2 \in A_2] \\
&= \sum_{\substack{x_1 \in A_1, \\ \mu_1(\{x_1\})>0}} P[X_1 = x_1] \cdot P[X_2 \in A_2 \,|\, X_1 = x_1] \\
&= \sum_{x_1 \in A_1} \mu_1(\{x_1\}) \cdot K(x_1, A_2) = \int_{A_1} K(x_1, A_2)\ \mu_1(\mathrm{d}x_1) \\
&= (\mu_1 \otimes K)(A_1 \times A_2),
\end{aligned}$$

hence $P = \mu_1 \otimes K$.

In the next proposition we are interested in an explicit formula for the disintegration in the case of absolutely continuous probability measures.

Note: If P is a probability measure on $(\Omega, \mathcal{A})$ and $\varphi : \Omega \to \mathbb{R}_+$ is $\mathcal{A}$-measurable with $\int \varphi \,\mathrm{d}P = 1$. Then

$$(\varphi P)(A) := \int_A \varphi \,\mathrm{d}P$$

defines another probability measure on $(\Omega, \mathcal{A})$.

For a given transition probability K from S_1 to S_2 and a function $\varphi : \Omega \to \mathbb{R}_+$, $\mathcal{A}$-measurable, let

$$K\varphi(x) := \int K(x, \mathrm{d}y)\ \varphi(x, y)\,, x \in S_1\,.$$

Proposition 2.10. *Let $P = \mu \otimes K$ and $\tilde{P} := \varphi P$. Then $\tilde{P} = \tilde{\mu} \otimes \tilde{K}$ with*

$$\tilde{\mu} = (K\varphi)\mu \quad \textit{und} \quad \tilde{K}(x, \mathrm{d}y) := \frac{\varphi(x, y)}{K\varphi(x)} \cdot K(x, \mathrm{d}y)$$

for all $x \in S_1$ with $K\varphi(x) > 0$ (and $\tilde{K}(x, \cdot\,) = \nu$, ν any probability measure on $(S_2, \mathcal{S}_2)$ if $x \in S_2$ is such that $K\varphi(x) = 0$).

Proof. (i) Let $\tilde{\mu}$ be the distribution of X_1 under $\tilde{P}$. Then for all $A \in \mathcal{S}_1$

$$\begin{aligned}
\tilde{\mu}(A) = \tilde{P}(A \times S_2) &= \int_{A \times S_2} \varphi(x, y)\ \mathrm{d}P \\
&= \int 1_A(x) \Big(\int \varphi(x, y)\ K(x, \mathrm{d}y) \Big)\ \mu(\mathrm{d}x) = \int_A (K\varphi)(x)\ \mu(\mathrm{d}x),
\end{aligned}$$

hence $\tilde{\mu} = (K\varphi)\mu$. In particular, $\tilde{\mu}$-a.s. $K\varphi > 0$, because

$$\tilde{\mu}(K\varphi = 0) = \int_{\{K\varphi = 0\}} (K\varphi)(x)\, \mu(dx) = 0\,.$$

(ii) Let $\tilde{K}$ be as above. Clearly, $\tilde{K}$ is a transition probability, because

$$\int \varphi(x,y)\; K(x,\mathrm{d}y) = K\varphi(x)\,, \text{ so that } \tilde{K}(x,S_2) = 1 \qquad \forall x \in S_1\,.$$

For all $A \in \mathcal{S}_1$ and $B \in \mathcal{S}_2$ we then have

$$\begin{aligned}\tilde{P}(A\times B) &= \int \Bigl(\underbrace{\int 1_{A\times B}(x,y)\cdot\varphi(x,y)\; K(x,\mathrm{d}y)}_{\leqslant K\varphi(x)}\Bigr)\; \mu(\mathrm{d}x)\\ &= \int_{\{K\varphi>0\}} \Bigl(\int 1_{A\times B}(x,y)\cdot\varphi(x,y)\; K(x,\mathrm{d}y)\Bigr)\; \mu(\mathrm{d}x)\\ &= \int_A K\varphi(x)\; \tilde{K}(x,B)\; \mu(\mathrm{d}x) = \int_A \tilde{K}(x,B)\; \tilde{\mu}(\mathrm{d}x)\\ &= (\tilde{\mu}\otimes\tilde{K})(A\times B).\end{aligned}$$

□

3 The canonical model for the evolution of a stochastic system in discrete time

Consider the following situation: suppose we are given

- measurable spaces $(S_i, \mathcal{S}_1)$, $i = 0,1,2,\dots$ and we define

  $$\begin{aligned}S^n &:= S_0\times S_1\times\cdots\times S_n\\ \mathcal{S}^n &:= \mathcal{S}_0\otimes\mathcal{S}_1\otimes\cdots\otimes\mathcal{S}_n = \sigma\bigl(\{A_0\times\cdots\times A_n \mid A_i\in\mathcal{S}_i\}\bigr).\end{aligned}$$

-
  - an initial distribution μ_0 on $(S_0, \mathcal{S}_0)$
  - transition probabilities

    $$K_n\bigl((x_0,\dots,x_{n-1}),\mathrm{d}x_n\bigr)$$

    from $(S^{n-1}, \mathcal{S}^{n-1})$ to $(S_n, \mathcal{S}_n)$, $n = 1,2,\dots$.

Using Fubini's theorem, we can then define probability measures P^n on S^n, $n = 0,1,2,\dots$ as follows:

$$\begin{aligned}P^0 &:= \mu_0 && \text{on } S_0,\\ P^n &:= P^{n-1}\otimes K_n && \text{on } S^n = S^{n-1}\times S_n\end{aligned}$$

Note that Fubini's theorem (see Proposition 2.4) implies that for any $\mathcal{S}^n$-measurable function $f : S^n \to \mathbb{R}_+$:

$$\begin{aligned}
&\int f \, \mathrm{d}P^n \\
&= \int P^{n-1}\big(\mathrm{d}(x_0, \ldots, x_{n-1})\big) \int K_n\big((x_0, \ldots, x_{n-1}), \mathrm{d}x_n\big) \; f(x_0, \ldots, x_{n-1}, x_n) \\
&= \cdots \\
&= \int \mu_0(\mathrm{d}x_0) \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_n\big((x_0, \ldots, x_{n-1}), \mathrm{d}x_n\big) \; f(x_0, \ldots, x_n).
\end{aligned}$$

3.1 The canonical model

Let $\Omega := S_0 \times S_1 \times \ldots$ be the set of all paths (or trajectories) $\omega = (x_0, x_1, \ldots)$ with $x_i \in S_i$, and

$$\begin{aligned}
&X_n(\omega) := x_n \qquad \text{(projection onto } n^{th}\text{-coordinate)}, \\
&\mathcal{A}_n := \sigma(X_0, \ldots, X_n) \quad (\subset \mathcal{A}), \\
&\mathcal{A} := \sigma(X_0, X_1, \ldots) = \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{A}_n\Big).
\end{aligned}$$

Our main goal in this section is to construct a probability measure P on $(\Omega, \mathcal{A})$ satisfying

$$\int f(X_0, \ldots, X_n) \, dP = \int f \, dP^n \qquad \forall \, n = 1, 2, \ldots$$

In other words, the "finite dimensional distributions" of P, i.e., the joint distributions of $(X_0, \ldots, X_n)$ under P, are given by P^n.

Proposition 3.1 (Ionescu-Tulcea)**.** *There exists a unique probability measure P on $(\Omega, \mathcal{A})$ such that for all $n \geqslant 0$ und all $\mathcal{S}^n$-measurable functions $f : S^n \to \mathbb{R}_+$:*

$$\int_{S^n} f \, \mathrm{d}P_{(X_0, \ldots, X_n)} = \int_{\Omega} f(X_0, \ldots, X_n) \, \mathrm{d}P = \int_{S^n} f \, \mathrm{d}P^n. \tag{3.7}$$

In other words: there exists a unique P such that $P^n = P \circ (X_0, \ldots, X_n)^{-1}$.

Proof. Uniqueness: Obvious, because the collection of finite cylindrical subsets

$$\mathcal{E} := \Big\{ \bigcap_{i=0}^{n} \{X_i \in A_i\} \;\Big|\; n \geqslant 0, \; A_i \in \mathcal{S}_i \Big\}$$

is closed under intersections and generates $\mathcal{A}$.
Existence: Let $A \in \mathcal{A}_n$, hence

$$A = (X_0, \ldots, X_n)^{-1}(A^n) \quad \text{for some } A^n \in \mathcal{S}^n, \; 1_A = 1_{A^n}(X_0, \ldots, X_n)$$

In order to have (3.7) we thus have to define

$$P(A) := P^n(A^n)\,. \tag{3.8}$$

We have to check that P is well-defined. To this end note that $A \in \mathcal{A}_n \subset \mathcal{A}_{n+1}$ implies

$$A = A^n \times S_{n+1} \times S_{n+2} \times \cdots \quad = A^{n+1} \times S_{n+2} \times \cdots\,,$$

hence $A^{n+1} = A^n \times S_{n+1}$. Consequently,

$$\begin{aligned} P^{n+1}(A^{n+1}) &= P^{n+1}(A^n \times S_{n+1}) \\ &= \int_{A^n} \underbrace{K_{n+1}\big((x_0, \ldots, x_n), S_{n+1}\big)}_{=1}\ \mathrm{d}P^n = P^n(A^n)\,. \end{aligned}$$

It follows that P is well-defined by (3.8) on $\mathcal{B} = \bigcup_{n+1}^{\infty} \mathcal{A}$. $\mathcal{B}$ is an algebra (i.e., a collection of subsets of Ω containing Ω, that is closed under complements and finite (!) unions), and P is finitely additive on $\mathcal{B}$, since P is (σ-) additive on $\mathcal{A}_n$ for every n. To extend P to a σ-additive probability measure on $\mathcal{A} = \sigma(\mathcal{B})$ with the help of Caratheodory's extension theorem, it suffices now to show that P is $\emptyset$-continuous, i.e., the following condition is satisfied:

$$B_n \in \mathcal{B}\,, B_n \searrow \emptyset \quad \Rightarrow \quad P(B_n) \xrightarrow{n\to\infty} 0\,.$$

(For Caratheodory's extension theorem see text books on measure theory, or Satz 1.41 in Klenke.)

W.l.o.g. $B_0 = \Omega$ and $B_n \in \mathcal{A}_n$ (if $B_n \in \mathcal{A}_m$, just repeat B_{n-1} m-times!). Then

$$B_n = A^n \times S_{n+1} \times S_{n+2} \times \ldots$$

with

$$A^{n+1} \subset A^n \times S_{n+1}$$

and we have to show that

$$\big(P(B_n) =\big) \quad P^n(A^n) \xrightarrow{n\to\infty} 0$$

(i.e., $\inf_n P^n(A^n) = 0$).

Suppose on the contrary that

$$\inf_{n\in\mathbb{N}} P^n(A^n) > 0\,.$$

We have to show that this implies

$$\bigcap_{n=0}^{\infty} B_n \neq \emptyset\,.$$

Note that

$$P^n(A^n) = \int \mu_0(\mathrm{d}x_0)\ f_{0,n}(x_0)$$

with

$$f_{0,n}(x_0) := \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), \mathrm{d}x_n\big)\ 1_{A^n}(x_0, \dots, x_n).$$

It is easy to see that the sequence $(f_{0,n})_{n\in\mathbb{N}}$ is decreasing, because

$$\begin{aligned}\int K_{n+1}\big((x_0, \dots, x_n), \mathrm{d}x_{n+1}\big)\ 1_{A^{n+1}}(x_0, \dots, x_{n+1})&\\ \leqslant \int K_{n+1}\big((x_0, \dots, x_n), \mathrm{d}x_{n+1}\big)\ 1_{A^n\times S_{n+1}}(x_0, \dots, x_{n+1})&\\ = 1_{A^n}(x_0, \dots, x_n)\,,&\end{aligned}$$

hence

$$\begin{aligned}f_{0,n+1}(x_0) &= \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_{n+1}\big((x_0, \dots, x_n), \mathrm{d}x_{n+1}\big)\ 1_{A^{n+1}}(x_0, \dots, x_{n+1})\\ &\leq \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), \mathrm{d}x_n\big)\ 1_{A^n}(x_0, \dots, x_n) = f_{0,n}(x_0)\,.\end{aligned}$$

In particular,

$$\int \inf_{n\in\mathbb{N}} f_{0,n}\ \mathrm{d}\mu_0 = \inf_{n\in\mathbb{N}} \int f_{0,n}\ \mathrm{d}\mu_0 = \inf_{n\in\mathbb{N}} P^n(A^n) > 0\,.$$

Therefore we can find some $\bar{x}_0 \in S_0$ with

$$\inf_{n\in\mathbb{N}} f_{0,n}(\bar{x}_0) > 0\,.$$

On the other hand we can write

$$f_{0,n}(\bar{x}_0) = \int K_1(\bar{x}_0, \mathrm{d}x_1)\ f_{1,n}(x_1)$$

with

$$\begin{aligned}f_{1,n}(x_1) := \int K_2\big((\bar{x}_0, x_1), \mathrm{d}x_2\big)&\\ \cdots \int K_n\big((\bar{x}_0, x_1, \dots, x_{n-1}), \mathrm{d}x_n\big)\ \mathbb{I}_{A^n}(\bar{x}_0, x_1, \dots, x_n)\,.&\end{aligned}$$

Using the same argument as above (with $\mu_1 = K_1(\bar{x}_0, \cdot)$) we can find some $\bar{x}_1 \in S_1$ with

$$\inf_{n\in\mathbb{N}} f_{1,n}(\bar{x}_1) > 0\,.$$

Iterating this procedure, we find for any $i = 0, 1, \dots$ some $\bar{x}_i \in S_i$ such that for all

$m \geqslant 0$

$$\inf_{n\in\mathbb{N}} \int K_m\big((\bar{x}_0,\dots,\bar{x}_{m-1}),\mathrm{d}x_m\big) \cdots \int K_n\big((\bar{x}_0,\dots,\bar{x}_{m-1},x_m,\dots,x_{n-1}),\mathrm{d}x_n\big) 1_{A^n}(\bar{x}_0,\dots,\bar{x}_{m-1},x_m,\dots,x_n) > 0.$$

In particular, if $m = n$

$$0 < \int K_m\big((\bar{x}_0,\dots,\bar{x}_{m-1}),\mathrm{d}x_m\big)\ 1_{A^m}(\bar{x}_0,\dots,\bar{x}_{m-1},x_m) \leqslant 1_{A^{m-1}}(\bar{x}_0,\dots\bar{x}_{m-1}),$$

so that

$$(\bar{x}_0,\dots,\bar{x}_{m-1}) \in A^{m-1} \quad \text{and} \quad \bar{\omega} := (\bar{x}_0,\bar{x}_1,\dots) \in \underbrace{B_{m-1}}_{=A^{m-1}\times S_m\times S_{m+1}\times\cdots}$$

for all $m \geq 1$, i.e.,

$$\bar{\omega} \in \bigcap_{m=0}^{\infty} B_m .$$

Hence the assertion is proven. □

Definition 3.2. Suppose that $(S_i,\mathcal{S}_i) = (S,\mathcal{S})$ for all $i = 0,1,2,\dots$. Then $(X_n)_{n\geqslant 0}$ on $(\Omega,\mathcal{A},P)$ (with P as in the previous proposition is said to be a *stochastic process* (in discrete time) with *state space* $(S,\mathcal{S})$, *initial distribution* μ_0 and *transition probabilities* $\big(K_n(\,\cdot\,,\,\cdot\,)\big)_{n\in\mathbb{N}}$.

3.2 Examples

1) Infinite product measures

Let

$$K_n\big((x_0,\dots,x_{n-1}),\,\cdot\,\big) = \mu_n ,$$

independent of $(x_0,\dots,x_{n-1})$: Then

$$P =: \bigotimes_{n=0}^{\infty} \mu_n$$

is said to be the *product measure associated with* $\mu_0,\mu_1,\dots$.

For all $n \geqslant 0$ and $A_0, \ldots, A_n \in \mathcal{S}$ we have that

$$\begin{aligned} P[X_0 \in A_0, \ldots, X_n \in A_n] &\overset{\text{I.-T.}}{=} P^n(A_0 \times \cdots \times A_n) \\ &= \int \mu_0(\mathrm{d}x_0) \int \mu_1(\mathrm{d}x_1) \cdots \int \mu_n(\mathrm{d}x_n)\, \mathbb{I}_{A_0 \times \cdots \times A_n}(x_0, \ldots, x_n) \\ &= \mu_0(A_0) \cdot \mu_1(A_1) \cdots \mu_n(A_n). \end{aligned}$$

In particular, $P_{X_n} = \mu_n$ for all n, and the natural projections $X_0, X_1, \ldots$ are independent. We thus have the following:

Proposition 3.3. *Let (μ_n) be a sequence of probability measures on a measurable space $(S, \mathcal{S})$. Then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a sequence (X_n) of independent r.v. with $P_{X_n} = \mu_n$ for all n.*

We have thus proven in particular the existence of a probability space modelling infinitely many independent $0-1$-experiments!

2) Markov chains

$$K_n\big((x_0, \ldots, x_{n-1}), \cdot\,\big) = \tilde{K}_n(x_{n-1}, \cdot\,)$$

time-homogeneous, if $\tilde{K}_n = K$ for all n.

For given initial distribution μ and transition probabilities K there exists a unique probability measure P on $(\Omega, \mathcal{A})$, which is said to be the canonical model for the time evolution of a *Markov chain*.

Example 3.4. Let $S = \mathbb{R}$, $\beta > 0$, $x_0 \in \mathbb{R} \setminus \{0\}$, $\mu_0 = \delta_{x_0}$ and $K(x, \cdot\,) = N(0, \beta x^2)$ $(K(0, \cdot\,) = \delta_0)$
For which β does the sequence (X_n) converge and what is its limit?
For $n \geqslant 1$

$$\begin{aligned} \mathbb{E}[X_n^2] &\overset{\text{I.-T.}}{=} \int x_n^2\, P^n\big(\mathrm{d}(x_0, \ldots, x_n)\big) \\ &= \int \Bigg(\underbrace{\int x_n^2\, K(x_{n-1}, \mathrm{d}x_n)}_{\substack{=\beta x_{n-1}^2, \\ K(x_{n-1}, \mathrm{d}x_n) = N(0, \beta x_{n-1}^2)}} \Bigg) P^{n-1}(\mathrm{d}x_0, \ldots, \mathrm{d}x_{n-1}) \\ &= \beta \cdot \mathbb{E}[X_{n-1}^2] = \cdots = \beta^n x_0^2\,. \end{aligned}$$

If $\beta < 1$ it follows that

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n^2\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n^2] = \sum_{n=1}^{\infty} \beta^n x_0^2 < \infty,$$

hence $\sum\limits_{n=1}^{\infty} X_n^2 < \infty$ P-a.s., and therefore

$$\lim_{n \to \infty} X_n = 0 \quad P\text{-a.s.}$$

A similar calculation as above for the first absolute moment yields

$$\mathbb{E}\big[|X_n|\big] = \cdots = \sqrt{\frac{2}{\pi}\cdot\beta}\cdot\mathbb{E}\big[|X_{n-1}|\big] = \cdots = \Big(\frac{2}{\pi}\cdot\beta\Big)^{\frac{n}{2}}\cdot\underbrace{\mathbb{E}\big(|X_0|\big)}_{=|x_0|},$$

because

$$\int |X_n|\; K(x_{n-1}, \mathrm{d}x_n) = \sqrt{\frac{2}{\pi}}\cdot\sigma = \sqrt{\frac{2}{\pi}\cdot\beta\,|x_{n-1}|}\,.$$

Consequently,

$$\mathbb{E}\Big[\sum_{n=1}^{\infty}|X_n|\Big] = \sum_{n=1}^{\infty}\Big(\frac{2}{\pi}\cdot\beta\Big)^{\frac{n}{2}}\cdot|x_0|\,,$$

so that also for $\beta < \frac{\pi}{2}$:

$$\lim_{n\to\infty} X_n = 0 \quad P\text{-a.s.}$$

In fact, if we define

$$\begin{aligned}\beta_0 &:= \exp\Big(-\frac{4}{\sqrt{2\pi}}\int_0^{\infty}\log x\cdot e^{-\frac{x^2}{2}}\;\mathrm{d}x\Big)\\ &= 2e^C \approx 3.56,\end{aligned}$$

where

$$C := \lim_{n\to\infty}\Big(\sum_{k=1}^{n}\frac{1}{k} - \log n\Big) \approx 0.577$$

denotes the Euler-Mascheroni constant, it follows that

$$\forall\,\beta < \beta_0 : X_n \xrightarrow{n\to\infty} 0 \quad P\text{-a.s. with exp. rate}$$

$$\forall\,\beta > \beta_0 : |X_n| \xrightarrow{n\to\infty} \infty \quad P\text{-a.s. with exp. rate.}$$

Proof. It is easy to see that for all n: $X_n \neq 0$ P-a.s. For $n \in \mathbb{N}$ we can then define

$$Y_n := \begin{cases}\frac{X_n}{X_{n-1}} & \text{on } \{X_{n-1}\neq 0\}\\ 0 & \text{on } \{X_{n-1} = 0\}.\end{cases}$$

Then $Y_1, Y_2, \ldots$ are independent r.v. with distribution $N(0,\beta)$, because for all measurable functions $f : \mathbb{R}^n \to \mathbb{R}_+$

$$\begin{aligned}\int f(Y_1,\ldots,Y_n)\;\mathrm{d}P &\overset{\text{I.-T.}}{=} \int f\Big(\frac{x_1}{x_0},\ldots,\frac{x_n}{x_{n-1}}\Big)\cdot\Big(\frac{1}{2\pi\beta}\Big)^{\frac{n}{2}}\cdot\Big(\frac{1}{x_0^2\cdots x_{n-1}^2}\Big)^{\frac{1}{2}}\\ &\qquad\cdot\exp\Big(-\frac{x_1^2}{2\beta x_0^2} - \cdots - \frac{x_n^2}{2\beta x_{n-1}^2}\Big)\;\mathrm{d}x_1\ldots\mathrm{d}x_n\\ &= \int f(y_1,\ldots,y_n)\cdot\Big(\frac{1}{2\pi\beta}\Big)^{\frac{n}{2}}\cdot\exp\Big(-\frac{y_1^2+\cdots+y_n^2}{2\beta}\Big)\;\mathrm{d}y_1\ldots\mathrm{d}y_n\,.\end{aligned}$$

Note that

$$|X_n| = |x_0| \cdot |Y_1| \cdots |Y_n|$$

and thus

$$\frac{1}{n} \cdot \log|X_n| = \frac{1}{n} \cdot \log|x_0| + \frac{1}{n} \sum_{i=1}^{n} \log|Y_i| \,.$$

Note that $(\log|Y_i|)_{i\in\mathbb{N}}$ are independent and identically distributed with

$$\mathbb{E}\big[\log|Y_i|\big] = 2 \cdot \frac{1}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}} \,\mathrm{d}x \,.$$

Kolmogorov's law of large numbers now implies that

$$\lim_{n\to\infty} \frac{1}{n} \cdot \log|X_n| = \frac{2}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}} \,\mathrm{d}x \quad P\text{-a.s.}$$

Consequently,

$$|X_n| \xrightarrow{n\to\infty} 0 \quad \text{with exp. rate, if} \int \cdots < 0,$$

$$|X_n| \xrightarrow{n\to\infty} \infty \quad \text{with exp. rate, if} \int \cdots > 0.$$

Note that

$$\begin{aligned}
&\frac{2}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}} \,\mathrm{d}x \overset{y=\frac{x}{\sqrt{\beta}}}{=} \frac{2}{\sqrt{2\pi}} \int_0^\infty \log(\sqrt{\beta}y) \cdot e^{-\frac{y^2}{2}} \,\mathrm{d}y \\
&= \frac{1}{2} \cdot \log\beta + \frac{2}{\sqrt{2\pi}} \int_0^\infty \log y \cdot e^{-\frac{y^2}{2}} \,\mathrm{d}y \\
&< 0 \quad \Leftrightarrow \quad \beta < \beta_0.
\end{aligned}$$

□

It remains to check that

$$-\frac{4}{\sqrt{2\pi}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2}} \,\mathrm{d}x = \log 2 + C$$

where C is the Euler-Mascheroni constant (Exercise!)

Example 3.5. Consider independent 0-1-experiments with success probability $p \in [0,1]$ but suppose that p ist unknown. In the canonical model:

$$S_i := \{0,1\},\ i \in \mathbb{N}; \quad \Omega := \{0,1\}^{\mathbb{N}},$$

$$X_i : \Omega \to \{0,1\},\ i \in \mathbb{N}, \quad \text{projections,}$$

$$\mu_i := p\varepsilon_1 + (1-p)\varepsilon_0,\ i \in \mathbb{N}; \quad P_p := \bigotimes_{i=1}^{\infty} \mu_i$$

$\mathcal{A}_n$ and $\mathcal{A}$ are defined as above.

Since p is unknown, we choose an a priori distribution μ on $\big([0,1], \mathcal{B}([0,1])\big)$ (as a distribution for the unknown parameter p).

Claim: $K(p, \cdot\,) := P_p(\,\cdot\,)$ is a transition probability from $\big([0,1], \mathcal{B}([0,1])\big)$ to $(\Omega, \mathcal{A})$.

Proof. We only need to show that for given $A \in \mathcal{A}$ the mapping $p \mapsto P_p(A)$ is measurable on $[0,1]$. To this end define

$$\mathcal{D} := \big\{A \in \mathcal{A} \;\big|\; p \mapsto P_p(A) \text{ is } \mathcal{B}([0,1])\text{-measurable}\big\}$$

Then $\mathcal{D}$ is a Dynkin system and contains all finite cylindrical sets

$$\{X_1 = x_1, \dots, X_n = x_n\}, \quad n \in \mathbb{N},\ x_1, \dots, x_n \in \{0,1\}\,,$$

because

$$P_p(X_1 = x_1, \dots, X_n = x_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n - \sum_{i=1}^n x_i}$$

is measurable (even continuous) in p!

The claim now follows from the fact, that the finite cylindrical sets are closed under intersections and generate $\mathcal{A}$. $\square$

Let $\bar{P} := \mu \otimes K$ on $\bar{\Omega} := [0,1] \times \Omega$ with $\mathcal{B}([0,1]) \otimes \mathcal{A}$. Using Remark 2.6 it follows that $\bar{P}$ has marginal distributions μ and

$$P(\,\cdot\,) := \int P_p(\,\cdot\,)\ \mu(\mathrm{d}p) \tag{3.9}$$

on $(\Omega, \mathcal{A})$. The integral can be seen as mixture of P_p according to the a priori distribution μ.

Note: The X_i are no longer independent under P!

We now calculate the initial distribution P_{X_1} and the transition probabilities in the particular case where μ is the Lebesgue measure (i.e., the uniform distribution on the unknown parameter p):

$$\begin{aligned} P \circ X_1^{-1} &= \int \big(p\varepsilon_1 + (1-p)\varepsilon_0\big)(\,\cdot\,)\ \mu(\mathrm{d}p) \\ &= \int p\ \mu(\mathrm{d}p) \cdot \varepsilon_1 + \int (1-p)\ \mu(\mathrm{d}p) \cdot \varepsilon_0 = \frac{1}{2} \cdot \varepsilon_1 + \frac{1}{2} \cdot \varepsilon_0. \end{aligned}$$

For given $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \{0,1\}$ with $k := \sum_{i=1}^n x_i$ it follows that

$$\begin{aligned} &P[X_{n+1} = 1 \,|\, X_1 = x_1, \dots, X_n = x_n] \\ &\qquad = \frac{P[X_{n+1} = 1, X_n = x_n, \dots, X_1 = x_1]}{P[X_n = x_n, \dots, X_1 = x_1]} \\ &\qquad \overset{(3.9)}{=} \frac{\displaystyle\int p^{k+1}(1-p)^{n-k}\ \mu(\mathrm{d}p)}{\displaystyle\int p^{k}(1-p)^{n-k}\ \mu(\mathrm{d}p)} \\ &\qquad = \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(n+3)} \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} = \frac{k+1}{n+2} \\ &\qquad = \underbrace{\left(1 - \frac{n}{n+2}\right) \cdot \frac{1}{2} + \frac{n}{n+2} \cdot \frac{k}{n}}_{\text{convex combination}}. \end{aligned}$$

Proposition 3.6. *Let P be a probability measure on $(\Omega, \mathcal{A})$ ("canonical model"), and*

$$\mu_n := P \circ X_n^{-1}, \quad n \in \mathbb{N}_0.$$

Then:

$$X_n,\ n \in \mathbb{N},\ \textit{independent} \quad \Leftrightarrow \quad P = \bigotimes_{n=0}^{\infty} \mu_n.$$

Proof. Let $\tilde{P} := \bigotimes_{n=0}^{\infty} \mu_n$. Then

$$P = \tilde{P}$$

if and only if for all $n \in \mathbb{N}_0$ and all $A_0 \in \mathcal{S}_0, \dots, A_n \in \mathcal{S}_n$

$$\begin{aligned} &P[X_0 \in A_0, \dots, X_n \in A_n] = \tilde{P}[X_0 \in A_0, \dots, X_n \in A_n] \\ &\quad = \prod_{i=0}^{n} \mu_i(A_i) = \prod_{i=0}^{n} P[X_i \in A_i], \end{aligned}$$

which is the case if and only if X_n, $n \in \mathbb{N}_0$, are independent. □

Definition 3.7. Let $S_i := S$, $i \in \mathbb{N}_0$, $(\Omega, \mathcal{A})$ be the canonical model and P be a probability measure on $(\Omega, \mathcal{A})$. In particular, $(X_n)_{n \geqslant 0}$ is a stochastich process in the sense of Definition 3.2. Let $J \subset \mathbb{N}_0$, $|J| < \infty$. Then the distribution of $(X_j)_{j \in J}$ under P

$$\mu_J := P \circ (X_i)_{i \in J}^{-1}$$

is said to be the *finite dimensional distribution* (w.r.t. J) on $(S^J, \mathcal{S}^J)$.

Remark 3.8. *P is uniquely determined by its finite-dimensional distributions resp. by*

$$\mu_{\{0,\dots,n\}}, \quad n \in \mathbb{N}\,.$$

4 Stationarity

Let $(S, \mathcal{S})$ be a measurable space, $\Omega = S^{\mathbb{N}_0}$ and $(\Omega, \mathcal{A})$ be the associated canonical model. Let P be a probability measure on $(\Omega, \mathcal{A})$.

Definition 4.1. The mapping $T : \Omega \to \Omega$, defined by

$$\omega = (x_0, x_1, \dots) \mapsto T\omega := (x_1, x_2, \dots)$$

is called the *shift-operator* on Ω.

Remark 4.2. *For all $n \in \mathbb{N}_0$, $A_0, \dots, A_n \in \mathcal{S}$*

$$T^{-1}\big(\{X_0 \in A_0, \dots, X_n \in A_n\}\big) = \{X_1 \in A_0, \dots, X_{n+1} \in A_n\}.$$

In particular: T is $\mathcal{A}/\mathcal{A}$-measurable

Definition 4.3. The measure P is said to be *stationary* (or *shift-invariant*) if

$$P \circ T^{-1} = P\,.$$

Proposition 4.4. *The measure P is stationary if and only if for all $k, n \in \mathbb{N}_0$:*

$$\mu_{\{0,\dots,n\}} = \mu_{\{k,\dots,k+n\}}.$$

Proof.

$$\begin{aligned} & P \circ T^{-1} = P \\ & \quad \Leftrightarrow P \circ T^{-k} = P \quad \forall k \in \mathbb{N}_0 \\ & \quad \overset{3.8}{\Leftrightarrow} (P \circ T^{-k}) \circ (X_0, \dots, X_n)^{-1} = P \circ (X_0, \dots, X_n)^{-1} \quad \forall\, k, n \in \mathbb{N}_0 \\ & \quad \Leftrightarrow \mu_{\{k,\dots,n+k\}} = \mu_{\{0,\dots,n\}}. \end{aligned}$$

□

Remark 4.5. (i) *The last proposition implies in the particular case*

$$P = \bigotimes_{i=1}^{\infty} \mu_n \quad \textit{with} \quad \mu_n := P \circ X_n^{-1}$$

that

$$P \textit{ stationary} \quad \Leftrightarrow \quad \mu_n = \mu_0 \;\; \forall\, n \in \mathbb{N}.$$

(ii) *If $P = \bigotimes \mu_n$ as in (i), hence $X_0, X_1, X_2, \dots$ independent, Kolmogorov's zero-one law implies that $P = 0 - 1$ on the tail-field*

$$\mathcal{A}^* := \bigcap_{n \geqslant 0} \sigma(X_n, X_{n+1}, \dots)$$

Proposition 4.6. *Let* $P = \bigotimes_{n=0}^{\infty} \mu_n$*,* $\mu_n := P \circ X_n^{-1}$*,* $n \in \mathbb{N}_0$*. Then* P *is* ergodic*, i.e.*

$$P = 0 - 1 \text{ on } \mathcal{I} := \{A \in \mathcal{A} \,|\, T^{-1}(A) = A\}\,.$$

$\mathcal{I}$ *is called the* σ*-algebra of shift-invariant sets.*

Proof. Using part (ii) of the previous remark, it suffices to show that $\mathcal{I} \subset \mathcal{A}^*$. But

$$\begin{aligned} A \in \mathcal{I} &\Rightarrow A = T^{-n}(A) \in \sigma(X_n, X_{n+1}, \dots) \quad \forall\, n \in \mathbb{N} \\ &\Rightarrow A \in \mathcal{A}^*. \end{aligned}$$

□

4 Conditional expectations

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{A}_0 \subset \mathcal{A}$ be a sub-σ-algebra.

1 Definitions

Definition 1.1. Let $X \geqslant 0$ be a r.v. A r.v. $X_0 \geqslant 0$ is said to be (a version of) the *conditional expectation of X given $\mathcal{A}_0$* if

(i) X_0 is $\mathcal{A}_0$-measurable.

(ii) $\mathbb{E}[Y_0 \cdot X] = \mathbb{E}[Y_0 \cdot X_0]$ for all $\mathcal{A}_0$-measurable r.v. $Y_0 \geqslant 0$.

Proposition 1.2. *Let X and $\mathcal{A}_0$ be as above. Then*

(i) A r.v. X_0 satisfying (i) and (ii) of the previous definition exists. (see Subsection 3 below).

(ii) Any two random variables satisfying (i) and (ii) coincide P-a.s.

Notation:

$$X_0 =: \mathbb{E}[X \,|\, \mathcal{A}_0].$$

Remark 1.3. *(i) "extreme cases"*

$$\mathbb{E}\big[X \;\big|\; \{\emptyset, \Omega\}\big] = \mathbb{E}[X] \qquad \mathbb{E}[X \,|\, \mathcal{A}] = X$$

(ii) Let X be a r.v., not necessarily nonnegative. We can decompose $X = X^+ - X^-$. If

$$\min\big(\mathbb{E}[X^+ \,|\, \mathcal{A}_0]\,,\, \mathbb{E}[X^- \,|\, \mathcal{A}_0]\big) < \infty \quad P-a.s.$$

we define

$$\mathbb{E}[X \,|\, \mathcal{A}_0] := \mathbb{E}[X^+ \,|\, \mathcal{A}_0] - \mathbb{E}[X^- \,|\, \mathcal{A}_0]\,.$$

Note that

$$\begin{aligned} X \in \mathcal{L}^1 \quad &\Leftrightarrow \quad \mathbb{E}[X \,|\, \mathcal{A}_0] \in \mathcal{L}^1 \\ &\Leftrightarrow \quad \begin{cases} \mathbb{E}[X^+] = \mathbb{E}\big[\mathbb{E}[X^+ \mid \mathcal{A}_0]\big] < \infty \\ \mathbb{E}[X^-] = \mathbb{E}\big[\mathbb{E}[X^- \mid \mathcal{A}_0]\big] < \infty, \end{cases} \end{aligned}$$

(iii) For any $A \in \mathcal{A}$ *let*

$$P[A \,|\, \mathcal{A}_0] := \mathbb{E}[1_A \,|\, \mathcal{A}_0]\,.$$

$P[A \,|\, \mathcal{A}_0]$ *is said to be the* conditional probability given $\mathcal{A}_0$.

(iv) discrete case *Let* $B_i \in \mathcal{A}$, $i \in \mathbb{N}$, *be pairwise disjoint with* $\Omega = \bigcup_{i\in\mathbb{N}} B_i$ *such that* $\mathcal{A}_0 = \sigma\{B_i \,|\, i \in \mathbb{N}\}$. *Then for any r.v.* $X \geqslant 0$*:*

$$\mathbb{E}[X \,|\, \mathcal{A}_0] = \sum_{i\in\mathbb{N}: P(B_i)>0} \underbrace{\mathbb{E}[X \,|\, B_i]}_{:=\frac{1}{P(B_i)}\cdot\mathbb{E}[X\cdot 1_{B_i}]} \cdot 1_{B_i}.$$

where

$$\mathbb{E}[X \,|\, B_i] = \frac{1}{P[B_i]}\int_{B_i} X\,dP$$

denotes the elementary *conditional expectation of* X *given* B_i*.*

(v) Let $(\Omega', \mathcal{A}')$ *be a measurable space and* $Y : \Omega \to \Omega'$ *be* $\mathcal{A}/\mathcal{A}'$*-measurable. Let* $\mathcal{A}_0 := \sigma(Y)$ *and* $X \geqslant 0$ *be a r.v. on* Ω*. The factorization lemma then implies that there exists a function* $f_X : \Omega' \to \bar{\mathbb{R}}_+$*, such that* $P - a.s.$

$$\mathbb{E}[X \,|\, Y] := \mathbb{E}\big[X \;\big|\; \sigma(Y)\big] = f_X \circ Y$$

Notation:

$$\mathbb{E}[X \,|\, Y = \omega'] := f_X(\omega') \qquad \omega' \in \Omega'\,.$$

$$\begin{tikzcd}
 & \begin{matrix}(\Omega,\mathcal{A}) \\ \mathcal{A}_0 := \sigma(Y) \subset \mathcal{A}\end{matrix} \arrow[r, "Y"] \arrow[d, "\mathbb{E}[X|Y]"] & (\Omega',\mathcal{A}') \arrow[dl, "f_X"] \\
(\Omega,\mathcal{A}) \arrow[r, "X"] & \mathbb{R}_+ &
\end{tikzcd}$$

In particular, $Y^{-1}(A') \in \mathcal{A}_0 = \sigma(Y)$ *for all* $A' \in \mathcal{A}'$ *and*

$$\int_{Y^{-1}(A')} X \,\mathrm{d}P \overset{\textit{1.1(ii)}}{=} \int_{Y^{-1}(A')} f_X(Y)\,\mathrm{d}P = \int_\Omega 1_{A'}(Y)\, f_X(Y)\,\mathrm{d}P$$
$$= \int_{A'} f_X \,\mathrm{d}(P \circ Y^{-1}),$$

Hence, f_X *is* $P \circ Y^{-1}$*-a.s. unique.*

2 Properties of the conditional expectation

(a) Linearity and monotonicity

$$\mathbb{E}[c_1X_1 + c_2X_2 \,|\, \mathcal{A}_0] = c_1 \cdot \mathbb{E}[X_1 \,|\, \mathcal{A}_0] + c_2 \cdot \mathbb{E}[X_2 \,|\, \mathcal{A}_0]$$

$$X \leqslant Y \quad P-a.s. \quad \Rightarrow \quad \mathbb{E}[X \,|\, \mathcal{A}_0] \leqslant \mathbb{E}[Y \,|\, \mathcal{A}_0]\,.$$

(b) Convergence theorems *B. Levi, monotone convergence*

$$0 \leqslant X_1 \leqslant X_2 \leqslant \dots \quad P-a.s. \quad \Rightarrow \quad \mathbb{E}\big[\lim_{n\to\infty} X_n \,\big|\, \mathcal{A}_0\big] = \lim_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{A}_0]\,.$$

Fatou

$$X_n \geq 0 \quad \forall\, n \in \mathbb{N} \quad \Rightarrow \quad \mathbb{E}\big[\liminf_{n\to\infty} X_n \,\big|\, \mathcal{A}_0\big] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{A}_0].$$

Lebesgue, dominated convergence $|X_n| \leqslant Y \in \mathcal{L}^1$ for all $n \in \mathbb{N}$ and $X_n \to X$ P-a.s. Then

$$\mathbb{E}\big[\lim_{n\to\infty} X_n \,\big|\, \mathcal{A}_0\big] = \lim_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{A}_0].$$

(c) contraction properties *Jensen's inequality* $X \in \mathcal{L}^1$ and u concave (!) function on $\mathbb{R}$. Then

$$\mathbb{E}\big[u(X) \,\big|\, \mathcal{A}_0\big] \leqslant u\big(\mathbb{E}[X \,|\, \mathcal{A}_0]\big) \quad P-a.s.$$

contraction on $\mathcal{L}^p$
In particular, for $p \geqslant 1$ and $X \in \mathcal{L}^p$

$$\big\|\mathbb{E}[X \,|\, \mathcal{A}_0]\big\|_p \leqslant \|X\|_p.$$

It follows that the mapping

$$X \mapsto \mathbb{E}[X \,|\, \mathcal{A}_0]$$

is continuous on $\big(L^p, \|\cdot\|_p\big)$

(d) smoothing properties Let $X \geqslant 0$, and $Y_0 \geqslant 0$ be $\mathcal{A}_0$-measurable. Then

$$\mathbb{E}[Y_0 \cdot X \,|\, \mathcal{A}_0] = Y_0 \cdot \mathbb{E}[X \,|\, \mathcal{A}_0] \quad P-a.s.$$

Tower property (in german: Projektivität)
in particular: let $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \mathcal{A}$ be σ-algebras

$$\Rightarrow \qquad \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{A}_1] \,\big|\, \mathcal{A}_0\big] = \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{A}_0] \,\big|\, \mathcal{A}_1\big] = \mathbb{E}[X \,|\, \mathcal{A}_0] \quad P-a.s.$$

(e) conditional expectation and independence

Proposition 2.1. *Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ be σ-algebras and $X \in \mathcal{L}^1$. Let $\sigma(\mathcal{A}_1, \mathcal{A}_2)$ (resp. $\sigma(\mathcal{A}_1, X)$) be the σ-algebra generated by $\mathcal{A}_1$ and $\mathcal{A}_2$ (resp. $\mathcal{A}_1$ and X). Then*

$$\sigma(\mathcal{A}_1, X) \textit{ independent of } \mathcal{A}_2$$
$$\Rightarrow \quad \mathbb{E}\big[X \bigm| \sigma(\mathcal{A}_1, \mathcal{A}_2)\big] = \mathbb{E}[X \,|\, \mathcal{A}_1] \quad P\textit{-a.s.}$$

In particular

$$X \textit{ independent of } \mathcal{A}_0 \quad \Rightarrow \quad \mathbb{E}[X \,|\, \mathcal{A}_0] = \mathbb{E}[X].$$

The proof of the proposition follows from the next proposition.

Proposition 2.2. *Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ be σ-algebras and $X \in \mathcal{L}^1$, $X \geq 0$. Then the following statements are equivalent:*

(i) $\mathbb{E}\big[X \bigm| \sigma(\mathcal{A}_1, \mathcal{A}_2)\big] = \mathbb{E}[X \,|\, \mathcal{A}_1]$.

(ii)

$$\mathbb{E}[X \cdot Y \,|\, \mathcal{A}_1] = \mathbb{E}[X \,|\, \mathcal{A}_1] \cdot \mathbb{E}[Y \,|\, \mathcal{A}_1].$$

for all $Y \geq 0$ $\sigma(\mathcal{A}_1, \mathcal{A}_2)$-measurable.

(iii)

$$\mathbb{E}[X \cdot X_2 \,|\, \mathcal{A}_1] = \mathbb{E}[X \,|\, \mathcal{A}_1] \cdot \mathbb{E}[X_2 \,|\, \mathcal{A}_1].$$

for all $X_2 \geq 0$ $\mathcal{A}_2$-measurable.

Proof. Exercise. □

Example 2.3 (Markov chain)**.** Let P_μ be the distribution of a Markov chain $X_0, X_1, \ldots$ on $\Omega = S^{\{0,1,\ldots\}}$ with initial distribution μ and transition probabilities $p(x, \mathrm{d}y)$ on $(S, \mathcal{S})$. Let $\mathcal{A}_n := \sigma(X_0, \ldots, X_n)$, $\hat{\mathcal{A}}_n := \sigma(X_i \,|\, i \geqslant n)$ and ϑ^n be the shift by n, i.e.

$$\vartheta^n((x_0, x_1, x_2, \ldots)) = (x_n, x_{n+1}, \ldots)$$

so that in particular $X_k \circ \vartheta^n = X_{n+k}$. Then the Markov property, applied at time n, implies

$$\mathbb{E}_\mu[\psi \circ \vartheta^n \,|\, \mathcal{A}_n] = \mathbb{E}_{X_n}[\psi] \tag{4.1}$$

for all $\psi \geq 0$ $\mathcal{A}$-measurable. It follows for $X \geqslant 0$, $\hat{\mathcal{A}}_n$-measurable, that

$$\mathbb{E}_\mu[X \,|\, \mathcal{A}_n] = \mathbb{E}_\mu[X \,|\, X_n]\,.$$

According to Proposition 2.2 this is equivalent to say that for all $\mathcal{A}_n$-measurable r.v. $Y \geqslant 0$

$$\mathbb{E}_\mu[Y \cdot X \,|\, X_n] = \mathbb{E}_\mu[Y \,|\, X_n] \cdot \mathbb{E}_\mu[X \,|\, X_n],$$

i.e., given the present $\sigma(X_n)$, the "future" $\hat{\mathcal{A}}_n$ is independent from the past $\mathcal{A}_n$. This is called the *elementary Markov property*.

(f) Best approximation in $\mathcal{L}^2$. Let $X \in \mathcal{L}^2$ be a r.v. Then

$$\mathbb{E}\Big[\big(X - \mathbb{E}[X \,|\, \mathcal{A}_0]\big)^2\Big] \leqslant \mathbb{E}\big[(X - Y_0)^2\big]$$

for all $Y_0 \in \mathcal{L}^2$, $\mathcal{A}_0$-measurable.

3 Existence

(a) Hilbert space method. Let $L^2 := \mathcal{L}^2/\sim$, with equivalence relation $X \sim Y$ meaning that $X = Y$ P-a.s. For given $X \in \mathcal{L}^2$, let $\bar{X}$ denote the corresponding equivalence class, i.e. $Y \in \bar{X}$ if and only if $X = Y$ P-a.s. Any $Y \in \bar{X}$ is called a representative of the equivalence class $\bar{X}$. Given two equivalence classes $\bar{X}, \bar{Y} \in L^2$ define its scalar product by

$$(\bar{X}, \bar{Y}) := \mathbb{E}[X \cdot Y],$$

where X (resp. Y) is a representative of $\bar{X}$ (resp. $\bar{Y}$). Then $\big(L^2, (\ ,\ )\big)$ is a *Hilbert space*.

Let $\mathcal{L}_0^2 := \mathcal{L}^2(\Omega, \mathcal{A}_0, P)$ $(\subset \mathcal{L}^2)$ be the subspace of square-integrable r.v. that are measurable w.r.t. the smaller σ-algebra $\mathcal{A}_0$. and let $L_0^2 := \mathcal{L}_0^2/\sim$. Then L_0^2 is a closed subspace of L^2 (by Riesz-Fisher (see Proposition 1.8.14), because any L^2-Cauchy sequence (X_n) has a subsequence converging P-a.s., so that X_n $\mathcal{A}_0$-measurable for all n implies that its L^2-limit X is $\mathcal{A}_0$-measurable too.) According to paragraph (f) of the preceeding subsection, we have that $\mathbb{E}[X \,|\, \mathcal{A}_0]$ for $X \in \mathcal{L}^2$ is a representative of the orthogonal projection $\bar{\pi}(\bar{X})$ of $\bar{X} \in L^2$ onto L_0^2. Using the existence of the orthogonal projection, we can now define the conditional expectation $\mathbb{E}[X \,|\, \mathcal{A}_0]$ as follows:

Step 1: For $X \in \mathcal{L}^2$ define

$$X_0 := \pi(\bar{X}) \quad \big(:= \mathcal{A}_0\text{-measurable representative } \bar{\pi}(\bar{X})\big)$$

It follows for all $Y_0 \in \mathcal{L}_0^2$ that

$$\begin{aligned} \mathbb{E}[Y_0 \cdot X] &= \mathbb{E}[Y_0 \cdot X_0] + \mathbb{E}\big[Y_0 \cdot (X - X_0)\big] \\ &= \mathbb{E}[Y_0 \cdot X_0] + \underbrace{(\bar{Y}_0, \overline{X - X_0})}_{=0}. \end{aligned} \tag{4.2}$$

Hence

$$\mathbb{E}[Y_0 \cdot X] = \mathbb{E}[Y_0 \cdot X_0]$$

for all $Y_0 \geq 0$ $\mathcal{A}_0$-measurable, so that

$$X_0 = \mathbb{E}[X \,|\, \mathcal{A}_0]$$

Similar to the last subsection

$$X \leqslant Y \quad P\text{-a.s.} \quad \Rightarrow \quad \pi(\bar{X}) \leqslant \pi(\bar{Y}) \quad P\text{-a.s.}$$

Step 2: For general $X \geqslant 0$, not necessarily in $\mathcal{L}^2$, consider $X \wedge n \in \mathcal{L}^2$. Monotonicity implies that

$$Z_0 := \lim_{n\to\infty} \pi(\overline{X \wedge n})$$

exists P-a.s. ($\mathcal{A}_0$-measurable !) Monotone convergence implies that for any $Y_0 \geqslant 0$, $\mathcal{A}_0$-measurable,

$$\mathbb{E}[Y_0 \cdot Z_0] = \lim_{n\to\infty} \mathbb{E}\big[Y_0 \cdot \pi(\overline{X \wedge n})\big] = \lim_{n\to\infty} \mathbb{E}[Y_0 \cdot X \wedge n] = \mathbb{E}[Y_0 \cdot X].$$

It follows that $Z_0 = \mathbb{E}[X \,|\, \mathcal{A}_0]$, hence the existence of the conditional expectation.

(b) Radon-Nikodym theorem Throughout the whole paragraph let $(\Omega, \mathcal{A})$ be a measurable space.

Definition 3.1. Let μ, ν be two finite measures on $(\Omega, \mathcal{A})$. Then ν is said to be *absolutely continuous w.r.t.* μ (notation: $\nu \ll \mu$), if

$$\mu(N) = 0\,, N \in \mathcal{A} \quad \Rightarrow \quad \nu(N) = 0\,.$$

In other words: every μ-null set is a ν-null set (but not necessarily conversely!).

Example 3.2. Let μ be a finite measure on Ω and $f \in \mathcal{L}^1$, $f \geq 0$. Define the (finite) measure

$$\nu(A) := \int_A f\, d\mu := \int 1_A f\, d\mu\,, \quad A \in \mathcal{A}\,. \tag{4.3}$$

Then $\nu \ll \mu$.

The theorem of Radon-Nikodym (see Proposition 3.4 below) tells us that conversely, if $\nu \ll \mu$ there exists an $\mathcal{A}$-measurable nonnegative function $f : \Omega \to \mathbb{R}_+$ satisfying (4.3). f is μ-a.s. uniquely determined and called the *density* of ν w.r.t. μ (Notation: $\frac{d\nu}{d\mu}$).

The Radon-Nikodym theorem will be used to obtain a second, independent proof for the existence of the conditional expectation. We will prove the theorem in the case of finite measures.

Lemma 3.3. *Let σ and τ be finite (positive) measures on a measurable space $(\Omega, \mathcal{A})$ with $\sigma(\Omega) < \tau(\Omega)$. Then there exists a measurable set $\Omega' \in \mathcal{A}$ satisfying:*

(i) $\sigma(\Omega') < \tau(\Omega')$.

(ii) $\sigma(A) \leqslant \tau(A)$ *for all* $A \in \Omega' \cap \mathcal{A} := \{A \subset \Omega' \,|\, A \in \mathcal{A}\}$.

Proof. (i) Let $\delta := \tau - \sigma$ (i.e., $\delta(A) := \tau(A) - \sigma(A)$ for all $A \in \mathcal{A}$). δ is bounded on $\mathcal{A}$, since

$$-\sigma(\Omega) \leqslant \delta(A) \leqslant \tau(\Omega)\,.$$

Define inductively sequences

$$(A_n)_{n\in\mathbb{N}\cup\{0\}}\ ,\quad (\Omega_n)_{n\in\mathbb{N}\cup\{0\}}$$

as follows:

Let $A_0 := \emptyset$, $\Omega_0 := \Omega\setminus A_0$ $(= \Omega$, and, given $A_0,\dots,A_n$ and $\Omega_0,\dots,\Omega_n$, we have that

$$\alpha_n := \inf_{A\in\Omega_n\cap\mathcal{A}} \delta(A) \leqslant 0\,(\text{ since } \delta(\emptyset)=0).$$

If $\alpha_n = 0$, let $A_{n+1} := \emptyset$ and $\Omega_{n+1} := \Omega_n\setminus A_{n+1}$ $(=\Omega_n)$.

If $\alpha_n < 0$ choose $A_{n+1}\in\Omega_n\cap\mathcal{A}$ with $\delta(A_{n+1})\leqslant\frac{\alpha_n}{2}$ and let $\Omega_{n+1} := \Omega_n\setminus A_{n+1}$.

It follows that the A_n, $n\in\mathbb{N}$, are pairwise disjoint, hence

$$\sum_{n=0}^{\infty}\delta(A_n)\quad\Big(=\tau\Big(\bigcup_{n\geq 0}A_n\Big)-\sigma\Big(\bigcup_{n\geq 0}A_n\Big)\Big)$$

is convergent, so that

$$\lim_{n\to\infty}\delta(A_n)=0\quad\Rightarrow\quad\lim_{n\to\infty}\alpha_n=0.$$

Let

$$\Omega' := \bigcap_{n\geqslant 0}\Omega_n.$$

Since (Ω_n) is decreasing, it follows that

$$\delta(\Omega')=\lim_{n\to\infty}\tau(\Omega_n)-\lim_{n\to\infty}\sigma(\Omega_n)=\lim_{n\to\infty}\delta(\Omega_n)>\delta(\Omega)$$

because

$$\delta(\Omega_{n+1})\geq\delta(\Omega_n)-\delta(A_{n+1})\geq\delta(\Omega_n)\geq\delta(\Omega_0)=\delta(\Omega)\,.$$

This proves (i).

(ii) Let $A\in\Omega'\cap\mathcal{A}$. Then $A\in\Omega_n\cap\mathcal{A}$ for all n, hence $\delta(A)\geqslant\alpha_n$ for all n, which implies $\delta(A)\geqslant\lim\limits_{n\to\infty}\alpha_n=0$. ☐

Proposition 3.4 (Radon-Nikodym). *Let μ and ν be finite (positive) measures on the measurable space $(\Omega,\mathcal{A})$. Then the following statements are equivalent:*

(i) There exists $f\geq 0$ $\mathcal{A}$-measurable (μ-a.s. uniquely determined) such that $\nu = f\cdot\mu$ (i.e., $\nu(A)=\int_A f\,d\mu$ for all $A\in\mathcal{A}$).

(ii) $\nu\ll\mu$ (i.e., $\mu(N)=0$ for $N\in\mathcal{A}$ implies $\nu(N)=0$).

Proof. (i)⇒(ii) obvious.
(ii)⇒(i)
Let G be the collection of all $\mathcal{A}$-measurable numerical functions $g \geqslant 0$ on Ω satisfying

$$g \cdot \mu \leqslant \nu,$$

i.e., $\nu(A) \geqslant \int_A g \, \mathrm{d}\mu$ for all $A \in \mathcal{A}$. Note that $g \equiv 0 \in G$. Note that G is stable under taking sup, because for $g, h \in G$

$$\begin{aligned}\int_A \sup(g,h) \, \mathrm{d}\mu &= \int_{A\cap\{g\geqslant h\}} g \, \mathrm{d}\mu + \int_{A\cap\{g<h\}} h \, \mathrm{d}\mu \\ &\leqslant \nu\big(A \cap \{g \geqslant h\}\big) + \nu\big(A \cap \{g < h\}\big) = \nu(A) \qquad \forall\, A \in \mathcal{A}\,.\end{aligned}$$

Let

$$\gamma := \sup_{g\in G} \int g \, \mathrm{d}\mu \quad \big(\leqslant \nu(\Omega) < \infty\big).$$

Since G is sup-stable, there exists an increasing sequence (g_n) of functions in G such that (by montone integration)

$$\gamma = \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int \lim_{n\to\infty} g_n \, \mathrm{d}\mu\,.$$

Let $f := \lim\limits_{n\to\infty} g_n$. Then

$$\int_A f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_A g_n \, \mathrm{d}\mu \leqslant \nu(A) \quad \forall\, A \in \mathcal{A}\,.$$

Consequently, $f \in G$. In other words: f is a maximum of

$$g \mapsto \int g \, \mathrm{d}\mu \text{ on } G\,.$$

We will show next that $f \cdot \mu = \nu$. Clearly, $f \cdot \mu \leqslant \nu$, since $f \in G$. Define

$$\tau := \nu - f \cdot \mu. \tag{4.4}$$

τ is a positive, finite measure on $\mathcal{A}$ and it remains to show that $\tau \equiv 0$.
Suppose on the contrary that $\tau(\Omega) > 0$ (so that $\mu(\Omega) > 0$). Let

$$\beta := \frac{1}{2} \cdot \frac{\tau(\Omega)}{\mu(\Omega)} > 0.$$

Then

$$\tau(\Omega) = 2\beta \cdot \mu(\Omega) > \beta \cdot \mu(\Omega).$$

Due to the previous lemma there exists a set $\Omega' \in \mathcal{A}$ satisfying

$$\tau(\Omega') > \beta \cdot \mu(\Omega') \quad \text{and} \quad \tau(A) \geqslant \beta \cdot \mu(A) \tag{4.5}$$

for all $A \in \Omega' \cap \mathcal{A}$. Define $f_0 := f + \beta \cdot 1_{\Omega'}$. Then f_0 is $\mathcal{A}$-measurable and for all $A \in \mathcal{A}$

$$\int_A f_0 \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu + \beta \cdot \mu(A \cap \Omega') \leqslant \int_A f \, \mathrm{d}\mu + \tau(A) = \nu(A).$$

It follows that $f_0 \in G$. On the other hand

$$\int f_0 \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \beta \cdot \mu(\Omega') = \gamma + \beta \cdot \mu(\Omega') > \gamma \,,$$

where we used the fact that $\nu \ll \mu$ implies $\mu(\Omega') > 0$. This is a contradiction to the fact that f is a maximum of

$$g \mapsto \int g \, \mathrm{d}\mu$$

on G. Consequently, $\tau \equiv 0$. □

Remark 3.5. *(i) Let μ, ν be finite measures, $\mu \ll \nu$ and $\frac{d\mu}{d\nu}$ be the density. Then*

$$\nu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$$

but in general not $\mu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$.

(ii) Let μ and ν be finite measures. μ and ν are said to be equivalent (notation: $\mu \sim \nu$) if $\mu \ll \nu$ and $\nu \ll \mu$. It is easy to see that $\mu \sim \nu$ if and only if $\nu \ll \mu$, hence $\nu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$, and in addition $\mu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$. In this case

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1} .$$

(iii) Let ν, μ and λ be finite measures, $\nu \ll \mu$ and $\mu \ll \lambda$. Then

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda} \quad \lambda - a.s.$$

Application to the construction of the conditional expectation

Let $\mathcal{A}_0 \subset \mathcal{A}$ be a sub-σ-algebra, $X \geq 0$, $X \in \mathcal{L}^1(P)$

Then

$$Q(A) := \int_A X \, dP = \int 1_A X \, dP \,, \quad A \in \mathcal{A}_0$$

defines a finite measure on $(\Omega, \mathcal{A}_0)$. Clearly, $Q \ll P_{|\mathcal{A}_0}$, hence

$$\exists \quad X_0 := \frac{dQ}{dP_{|\mathcal{A}_0}} \qquad \mathcal{A}_0\text{- measurable}\,.$$

Note that for $A \in \mathcal{A}_0$, by definition of the density,

$$\int 1_A X \, dP = Q(A) = \int 1_A \frac{dQ}{dP} \, dP = \int 1_A X_0 \, dP \,. \tag{4.6}$$

Clearly, (4.6) extends to

a) simple functions $Y_0 = \sum_{k=1}^n a_k 1_{A_k}$ by linearity

b) general $Y_0 \geq 0$, $\mathcal{A}_0$-measurable, by taking pointwise limits of increasing simple functions $Y_n \uparrow Y_0$

It follows that $X_0 = \frac{dQ}{dP_{|\mathcal{A}_0}}$ is a version of $\mathbb{E}[X|\mathcal{A}_0]$. For general $X \geq 0$ consider the approximation $X \wedge n \uparrow X$ and use monotonicity.

4 Regular conditional probabilities

Consider the mapping

$$A \mapsto P[A \,|\, \mathcal{A}_0] \quad \big(:= \mathbb{E}[1_A \,|\, \mathcal{A}_0]\big).$$

The following properties haven been shown in Subsection 2:

- $0 \leqslant P[A \,|\, \mathcal{A}_0] \leqslant 1$ P-a.s.
- $P[\emptyset \,|\, \mathcal{A}_0] = 0$ and $P[\Omega \,|\, \mathcal{A}_0] = 1$ P-a.s.
- $A_1 \subset A_2$ implies $P[A_1 \,|\, \mathcal{A}_0] \leqslant P[A_2 \,|\, \mathcal{A}_0]$ P-a.s.
- A_n, $n \in \mathbb{N}$, pairwise disjoint
  $\Rightarrow \quad P\Big[\bigcup_{n=1}^\infty A_n \,\Big|\, \mathcal{A}_0\Big] = \sum_{n=1}^\infty P[A_n \,|\, \mathcal{A}_0]$ P-a.s.

Note that this does not yet imply that

$$A \mapsto P[A \,|\, \mathcal{A}_0](\omega), \quad A \in \mathcal{A}, \tag{4.7}$$

defines a probability measure on $\mathcal{A}$ for P-a.e. $\omega \in \Omega$.

However, this is true in the discrete case and we may ask under what assumptions this is true in the general case, i.e., under what assumptions is it possible to choose "good" versions of $P[A \,|\, \mathcal{A}_0]$, $A \in \mathcal{A}$, such that (4.7) in fact defines a probability measure on $(\Omega, \mathcal{A})$ at least for P-a.e. $\omega \in \Omega$.

Definition 4.1. A measurable space $(\Omega, \mathcal{A})$ is said to be a *Borel space*, if there exists a Borel subset $U \in \mathcal{B}(\mathbb{R})$ and a bijection $\varphi : \Omega \to U$ such that both φ and φ^{-1}, are measurable.

Proposition 4.2. *Assume that $(\Omega, \mathcal{A})$ is a Borel space and let P be a probability measure on $(\Omega, \mathcal{A})$. Let $\mathcal{A}_0 \subset \mathcal{A}$ be a sub σ-algebra. Then there exists a transition probability from $(\Omega, \mathcal{A}_0)$ to $(\Omega, \mathcal{A})$ such that for all $A \in \mathcal{A}$*

$$K_{\mathcal{A}_0}(\omega, A) = P[A \,|\, \mathcal{A}_0](\omega) \quad P - a.s.$$

In other words: $K_{\mathcal{A}_0}(\cdot, A)$ is a version of the conditional probability $P[A \,|\, \mathcal{A}_0]$ for all $A \in \mathcal{A}$. The transition probability $K_{\mathcal{A}_0}$ is called a regular conditional probability given *$\mathcal{A}_0$.*

$K_{\mathcal{A}_0}$ is uniquely determined in the following sense: if $\tilde{K}_{\mathcal{A}_0}$ is a second transition probability from $(\Omega, \mathcal{A}_0)$ to $(\Omega, \mathcal{A})$ having these properties, it follows that there exists a P-null set $N \in \mathcal{A}$, such that for all $\omega \in \Omega \setminus N$ and all $A \in \mathcal{A}$

$$K_{\mathcal{A}_0}(\omega, A) = \tilde{K}_{\mathcal{A}_0}(\omega, A).$$

For a proof see Klenke, Wahrscheinlichkeitstheorie, Satz 8.36.