Probability Theory

Wilhelm Stannat

Technische Universität Darmstadt

Winter Term 2007/08

First part - 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. Sirjaev, 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

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

is the space of all possible outcomes.

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

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

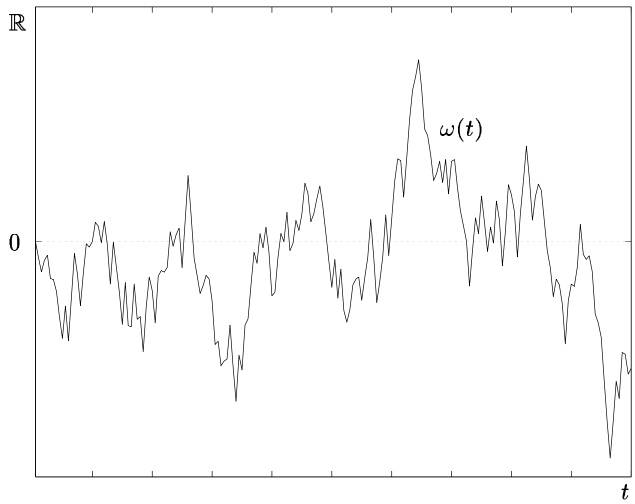
In this case Ω is uncountable in contrast to the previous examples. We can identify Ω with the set $[0, 1] \subset \mathbb{R}$ using the binary expansion

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

(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":

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

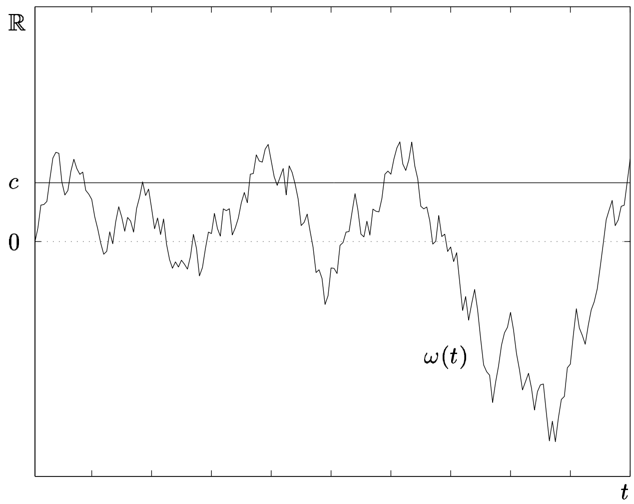
(iii) **Tossing a coin infinitely many times** "relative frequency of 1 equals p":

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

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

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

$$A = \{\omega \in \mathcal{C}([0,1]) \mid \max_{0 \leqslant t \leqslant 1} \omega(t) > c\}.$$



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*

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

(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*

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

(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

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

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

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

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!)

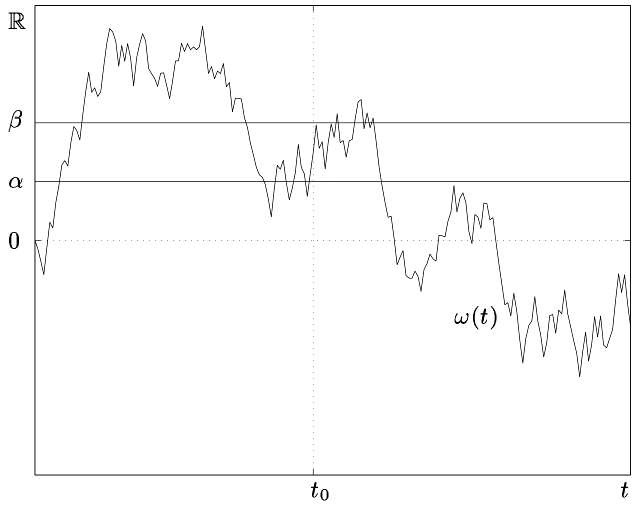
(v) **Continuous stochastic processes** $\Omega = \mathcal{C}([0,1])$, $P =$ Wiener measure ("Brownian motion") For fixed $t_0 \in \mathbb{R}_+$ and $\alpha, \beta \in \mathbb{R}$ we have that

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

(Gaussian or normal distribution).

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

Answer to this question: Later!



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

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

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* $P(A \cup B) = P(A) + P(B)$, *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 $(= \sigma(\{[a,b] \mid 0 \leqslant a \leqslant b \leqslant 1\}))$. 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 said to be the *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

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

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 $k \in \{0, \dots, n\}$:

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

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

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

(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,

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

(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

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

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

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

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

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

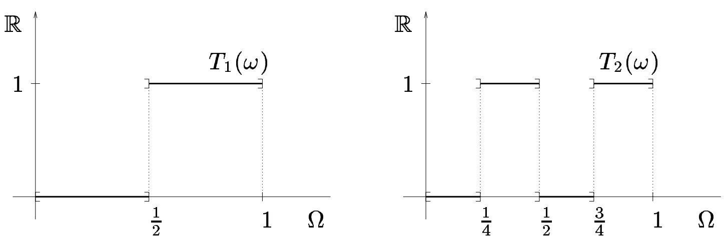
and let

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

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, $\dots$). 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: $\mathcal{B}(\bar{\mathbb{R}}) = \{B \subset \bar{\mathbb{R}} \mid B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\}$)

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 $X_1, X_2, \ldots$ 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

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

- sum

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

□

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

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

Then

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

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

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

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
$$\mathcal{N} := \{X \text{ r.v.} \,|\, X = 0 \ P\text{-a.s.}\}$$
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*

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

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

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

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

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

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* $0 \leqslant X_1 \leqslant X_2 \leqslant \dots$. *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.

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

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.

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

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

$$\mathbb{E}\big[\lim_{n\to\infty} X_n\big] = \mathbb{E}\big[\liminf_{n\to\infty} X_n\big] \overset{\text{Fatou}}{\leqslant} \liminf_{n\to\infty} \mathbb{E}[X_n] \leqslant \limsup_{n\to\infty} \mathbb{E}[X_n]$$
$$\leqslant \mathbb{E}\big[\limsup_{n\to\infty} X_n\big] = \mathbb{E}\big[\lim_{n\to\infty} X_n\big] \qquad \square$$

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

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

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

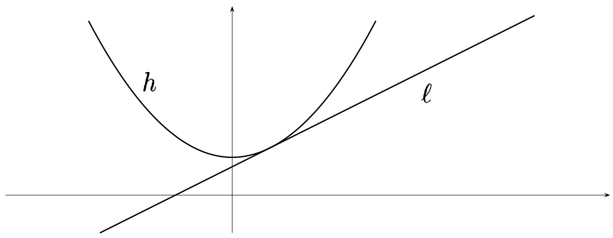
5 Inequalities

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

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

$\mathcal{L}^p$ 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 $\Omega = \{\omega_1, \ldots, \omega_n\}$, then*

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

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

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

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 said to be 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

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

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

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

(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

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

then

$$\mathbb{E}[T] = \mathbb{E}\Big[\sum_{n=1}^{\infty} n \cdot 1_{\{T=n\}}\Big] = \sum_{n=1}^{\infty} n \cdot P_p[T=n] = \sum_{n=1}^{\infty} n \cdot (1-p)^{n-1}p = \frac{1}{p},$$

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"*

- $\text{cov}(\tilde{X}_i, \tilde{X}_j) = \text{cov}(X_i, X_j) = 0$ *for* $i \neq j$
- $\text{var}(\tilde{X}_i) = \text{var}(X_j)$.

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.

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

□

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

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

□

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 $\text{var}(X_i) = p(1-p)$ $(\leqslant \frac{1}{4})$

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 $f \in \mathcal{C}([0,1])$ with Bernstein polynomials

Let

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

$\varepsilon > 0$: 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,

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

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 $\varepsilon > 0$

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

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 $\varepsilon > 0$

$$P\bigg[\Big|\frac{S_{k^2}}{k^2}\Big| > \varepsilon\bigg] \overset{\text{Chebychev}}{\leqslant} \frac{1}{\varepsilon^2 k^4}\,\mathrm{var}(S_{k^2}) = \frac{1}{\varepsilon^2 k^4}\sum_{l=1}^{k^2}\mathrm{var}(X_i) \leq \frac{c}{\varepsilon^2 k^2}$$

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 $\varepsilon > 0$:

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

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

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

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

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

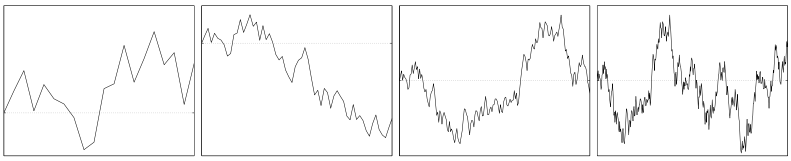
Example 7.9. Bernoulli experiments with $p \in [0,1]$

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

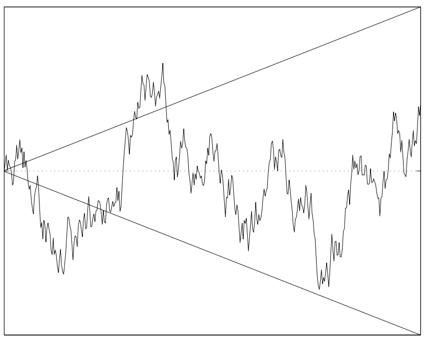
Consider the experiment of tossing a fair coin ($p = \frac{1}{2}$), $Y_i := 2X_i - 1$

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

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 $X, X_1, X_2, \dots$ 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 $\lim_{n\to\infty} \|X_n - X\|_p = 0$)

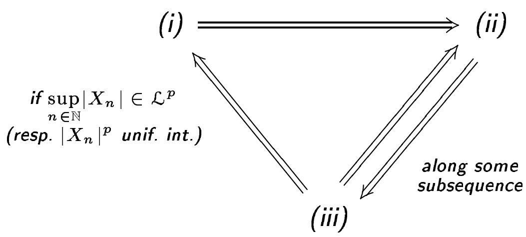
(ii) **Convergence in probability**

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

(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

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

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

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

(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

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

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

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

Lemma 7.7 now implies

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

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,

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

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

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

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

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

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,* $\alpha, \beta \in \mathbb{R}$

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

(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 $P\big[|X_n| \geqslant \frac{\varepsilon}{2}\big] < \delta \ \forall n \ge n_0$. Hence, for $n \ge n_0$

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

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

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

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

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*

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

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

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

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")

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

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 7.8 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 7.8 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 Lebesgue's dominated convergence theorem.

Proposition 8.13. *Let* $X_n \geq 0$, $\lim_{n\to\infty} X_n = X$ *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. **"'⇒"':** Obvious.

"'⇐"':

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

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

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

Proof. See text books on measure theory. □