MA 144 a/b: Probability

Oliver Knill

This introduction to probability theory was tought during the first and second term 1994 at Caltech.

The material which was distributed to the student has not been revised since then. All the included material has been covered during a time of 2 terms. It might have been at the upper limit what one can cover in 20 weeks for an introduction into probability.

Contents

0.1 Introduction

This course has the aim to get as fast as possible to the relevant results and to present short, intuitive proofs.

We assume no preknowledge in probability. Some measure theory and real analysis is needed but we will restate used theorems. Here is the plan

The topics might get permuted a bit. According to time more can be added or shortened

- Basic measure theory, random variables, independence
- Chebychev and Jensen inequalities, distributions and characteristic functions, Borel Cantelli lemma, 0-1 law.
- Limit theorems: Weak and strong law of large numbers, Three series thoerem, Central limit theorem.
- Martingales, convergence theorem (discrete time), optimal stopping time theorem.
- Random walks, the law of iterated logarithm.
- Markov chains, basic limit theorem.
- Brownian motion, Wiener Measure, Feynmann-Kac formula.

Literature
We do **not** follow a specific text book and the course is independent and can be followed without further reading. If the time allows it, we recommend to read beside in a textbook like for example the new and very advanced textbook **Stroock**, *Probability theory, An anlytic view.* It is concentrated and contains much material. For a slower and more detailed approach there are many older texted books like for example [4]. Also classical books like [1] are still good. For an introduction with use of Martingales see [16] or [6]. For Brownian motion, we refer to [8], for random walks to [14], for Markov chains to [2]. For applications in physics and chemistry, see [5]. Probability theory can be done (using nonstandard analysis) on finite probability spaces without loss of generality [11].

0.2 Some paradoxons in probability theory

Colloquial language is not precice enough to treat probabilistic problems. Paradoxons appear, when the definition of objects allows different interpretations which leads then to funny or wrong conclusions. We give three famous problems. These three examples should serve as a motivation to introduce probability theory on a precice footing.

1) **Bertrand's paradox (Bertrand 1889)**
Throw at random straight lines onto the unit disc. What is the probability that the straight line intersects the disc with a length $\geq \sqrt{3}$, the length of the equilateral triangle inscribed in the circle?

First answer: take an arbitrary point P in the disc and look at the set of all lines passing through that point parametrized by an angle ϕ. In order that the chord be longer than $\sqrt{3}$, the line has to lie within an angle of $60°$ out of $180°$. The probability is thus $1/3$.

Second answer: take all lines perpendicular to a fixed diameter. The chord is longer than $\sqrt{3}$, when the point of intersection lies on the middle half of the diameter. The probability is thus $1/2$.

Third answer: look at the midpoints of the chords. If they lie in the disc of radius $1/2$ which has area $1/4$ of the whole disc, the chord is longer than $\sqrt{3}$. The probability is thus $1/4$.

2) **Petersburger paradox (D.Bernoulli, 1738)**
In the casino, you pay an entrence fee c and you get the prize 2^T, where T is the number of times, one has to throw a coin until "head" appears. Fair would be an entrence fee of

$$c = \sum_{k=1}^{\infty} 2^k P[T = k] = \sum_{k=1}^{\infty} 1 = \infty .$$

The paradox is that nobody would agree to pay $c = 10$.

3) **The three door problem (1991)** Suppose you're on a game show and you are given a choice of three doors. Behind one door is a car and behind the others are goats. You pick a door-say No. 1 - and the host, who knows what's behind the doors, opens another door-say, No. 3-which has a goat. (In all games, he opens a door to reveal a goat). He then says to you, "Do you want to pick door No. 2?" (In all games he always offers an option to switch). Is it to your advantage to switch your choice?

1) **To Bertrand's paradox**:
Like most paradoxa in mathematics, there is a term in the question which is not well defined. Here it is "random throwing". The three answers depend on the choosen probability distributions.

2) **To the Petersburg paradox**:
the problem with the situation is that it is not quite clear, what is fair. For example, the situation $T = 20$ is so unprobably that it never occurs in the life time of a person so that one actually has not to worry about so big values of T. But how to solve the paradox? Bernoulli proposed to take not the the expectation value $E[G]$ of the profit $G = 2^T$ but an expectation value say $(E[\sqrt{G}])^2$ which would lead to a fair entrence

$$(E[\sqrt{G}])^2 = (\sum_{n=1}^{\infty} 2^{k/2} 2^{-k})^2 = (1/(\sqrt{2} - 1))^2 \sim 6.25 .$$

It is not so clear if that is the way out of the paradox. Such reasoning play however a role in economical and social sciences.

3) **To the three doors problem**:
The problem was discussed by Marilyn vos Savant in a "Parade" column in 1991 and provoked a big controversy in the next months. Thousends of letters were written. The problem is that intuitive argumentation can easily lead to the conclusion that it does not

matter whether to change the door or not. Switching the door doubles the chances to win:

No switching: you choose a door and win with probability 1/3. The opening of the host does not affect any more your choice.

Switching: you choose the door with the car. You loose since you switch. You choose a door with a goat. The host opens the other door with the goat. You win and there are two cases, where you win. There is one case, where you loose. The probability to win is 2/3.

0.3 Some applications of probability theory

Probability theory is a central topic in mathematics. There are close relations and intersections with other fields like computer science, ergodic theory and dynamical systems, cryptology, game theory, analysis, partial differential equation, mathematical physics, economical sciences or statistical mechanics. Again as a motivation for the course, we give some problems and topics which can be treated by probabilistic methods.

1)

Random walks

(statistical mechanics, gambling, stock markets, quantum field theory).

We walk randomly through a lattice by choosing at each vertex a direction at random. A problem is to determine the probability that the walk returns back to the origin.

2)

Percolation problems

(model of a porous medium, statistical mechanics).

To each bond in a lattice is attached the number 0 (closed) with probability $(1-p)$ or 1 (open) with probability p. Two sites x, y in the lattice are in the same cluster, if there is a path from x to y going though closed bonds. One says that percolation occurs if there is a positive probability that an infinite cluster appears. One problem is to find the critical probabily p_c, which is the infimum over all p, for which percolation occurs.

3)

Random Schrödinger operators.

(quantum mechanics, functional analysis, disordered systems, solid state physics)

Consider the linear map $Lu(n) = \sum_n u(n) + V(n)u(n)$, where $V(n)$ takes randomly values in $\{0, 1\}$. The problem is to determine the spectrum or spectral type of the infinite matrix L. The map L describes an electron in a one dimensional disordered crystal. The spectral properties of L have a relation with the conductivity of the crystal.

4)

Classical dynamical systems

(celestial mechanics, fluid dynamics, population models)

The study of deterministic dynamical systems like the map $x \mapsto 4ax(1-x)$ on the interval $[0, 1]$ or the three body problem in celestial mechancs has shown that such systems can behave like random systems. Many effects can be described by ergodic theory, which can be viewed as a brother of probability theory.

5)

Cryptology.

(computer science, coding theory, data encryption)

Data encryption like the DES are used in most computers or phones. Coding theory tries to find good codes which can repair loss of information due to bad channels. Public key systems use trapdoor algorithms like the problem to factor a large integer $N = pq$ with prime p, q. The number N can be public but only the one, who knows the factorisation can read the mails. Many algorithms need pseudo random number generators like the sequence generated by $m \mapsto m^2 + c(\mathrm{mod} p)$. Much probability theory is involved in designing, investigating and attacking data encryption, codes or random number generators.

0.4 Appendix: A simulation of the Petersburg game

The following Mathematica procedures allow to simulate the game of Petersburg.

```
Game[n_]:=Table[Module[{s=1},While[Random[Integer]==1,s=2*s];s],{k,n}];
AverageProfit[n_]:=N[Apply[Plus,Game[n]]/n];
ListPlot[Table[AverageProfit[k],{k,100}]]
```

This program produced the following graph:

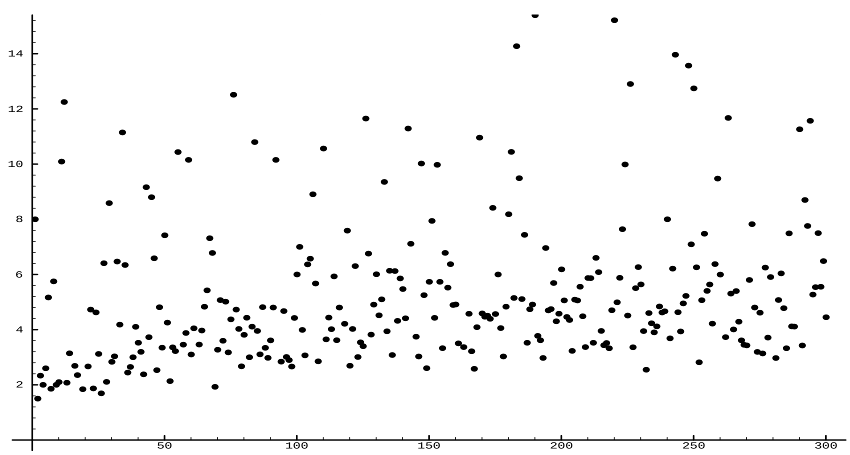


Figure 1. The average profit per game of the player in dependece on the number of games which were played.

Chapter 1

Limit theorems

1.1 Probability spaces, random variables, independence

Definition. $\mathcal{A}$ is a σ- **algebra** if $\mathcal{A}$ is a set of subsets of Ω satisfying

$$\boxed{\begin{array}{l} \Omega \in \mathcal{A}, \\ A \in \mathcal{A} \to A^c \in \mathcal{A}, \\ A_n \in \mathcal{A} \Rightarrow \bigcup_{n \in \mathbf{N}} A_n \in \mathcal{A} \end{array}}$$

The pair $(\Omega, \mathcal{A})$ is called a **measurable space**.

Properties. $\bigcap_{n \in \mathbf{N}} A_n \in \mathcal{A}$, $\limsup_n A_n := \bigcap_n^\infty \bigcup_{m=n}^\infty A_n \in \mathcal{A}$, $\liminf_n A_n := \bigcup_n^\infty \bigcap_{m=n}^\infty A_n \in \mathcal{A}$, $\{\mathcal{A}_\lambda\}_{i \in I}$ family of σ- subalgebras of $\mathcal{A} \Rightarrow \bigcap_{i \in I} \mathcal{A}_i$ σ-algebra.

Definition. Any set of subset $\mathcal{C}$ of Ω defines $\sigma(\mathcal{C})$, the smallest σ algebra generated by $\mathcal{C}$. If $(E, \mathcal{O})$ is a topological space, where $\mathcal{O}$ is the set of open sets in E, then $\sigma(\mathcal{O})$ is called the **Borel σ-algebra** of the topological space.

Definition. $P : \mathcal{A} \to \mathbf{R}$ is a **probability measure** and $(\Omega, \mathcal{A}, P)$ is a **probability space** if

$$\boxed{\begin{array}{l} P(A) \geq 0 \;\; \forall A \in \mathcal{A}, \\ P(\Omega) = 1, \\ A_n \in \mathcal{A} \;\text{ disjoint} \Rightarrow P(\bigcup_n A_n) = \sum_n P(A_n) \;(\sigma \text{ additivity}) \end{array}}$$

Properties. $P(\emptyset) = 0$. $A \subset B \Rightarrow P(A) \leq P(B)$. $P(\bigcup_n A_n) \leq \sum_n P(A_n)$. $P(A^c) = 1 - P(A)$. $A_1 \subset A_2, \subset \dots$, with $A_n \in \mathcal{A}$ then $P(\bigcup_{n \in \mathbf{N}} A_n) = \lim_{n \to \infty} P(A_n)$. The later statement is equivalent to σ-additivity if P is only assumed to be additive.

Definition. A map X from one measure space $(\Omega, \mathcal{A})$ to an other probabiltiy space $(\Delta, \mathcal{B})$ is called **measurable** if $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

Definition. $X : \Omega \to \mathbf{R}$ is a **random variable** if it is a measurable map from $(\Omega, \mathcal{A})$ to $(\mathbf{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbf{R}$. Denote with $\mathcal{L}$ the set of real random variables. $\mathcal{L}$ is an algebra under addition and multiplication.

We consider also random variables taking values in an arbitrary measurable space $(E, \mathcal{B})$. For example if $E = \mathbf{R}^d$, then a random variable X is called a $d-$ **dimensional random variable** which can be written as a sequence $X = (X_1, \dots, X_d)$.

Definition. Every random variable X defines a σ-algebra $X^{-1}(\mathcal{B})$, which we denote by $\sigma(X)$ and call it the **algebra generated by** X.

Definition. Given $B \in \mathcal{A}$ define $P(A|B) := P(A \cap B)/P(B)$, the **conditional probability**.

Properties. $P(\bigcap_{k=0}^n A_k) = P(A_0)P(A_1|A_0) \cdots P(A_n|\bigcap_{n=0}^{n-1} A_k)$.
If $\{B_k\}_{k=1}^n$ are disjoint, then $P(A) = \sum_{k=1}^n P(A|B_k)P(B_k)$.

Definition. A family $\{\mathcal{A}_i\}_{i \in I}$ of $\sigma-$ subalgebras of $\mathcal{A}$ is called **independent**, if for every finite subset J of I (we write $J \subset_f I$), and every choice $A_j \in \mathcal{A}_j$ $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$. A family $\{X_j\}_{j \in J}$ of random variables is called **independent**, if $\{\sigma(X_j)\}_{j \in J}$ are independent σ-algebras. A family of sets $\{A_j\}_{j \in I}$ is called **independent**, if the σ-algebras $\mathcal{A}_j = \{\emptyset, A_j, A_j^c, \Omega)$ are independent.

Properties. If a σ algebra $\mathcal{F} \subset \mathcal{A}$ is independent to itself, then $P(A \cap A) = P(A) = P(A)^2$ so that for every $A \in \mathcal{F}$, $P(A) \in \{0, 1\}$. Such a σ algebra is called $P-$ **trivial**.
Two sets $A, B \in \mathcal{A}$ are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$. Then, clearly also A, B^c are independent. It follows from the definition that if $P(B) > 0$, then $P(A|B) = P(A)$.

Definition. A family $\mathcal{I}$ of subsets of Ω is called a π-**system**, if $\mathcal{I}$ is closed under intersections. A σ- additive map from a π system $\mathcal{I}$ to $[0, \infty)$ is called a **measure**.

Lemma 1.1.1 (Extensions of measures from π-systems to σ algebras) *Given a π-system $\mathcal{I}$. If two measures μ, ν on $\sigma(\mathcal{I})$ satify $\mu(\Omega), \nu(\Omega) < \infty$ and agree on $\mathcal{I}$, then $\mu = \nu$.*

Proof. Definition. Use the notation $A_n \nearrow A$ if $A_n \subset A_{n+1}$ and $\bigcup_n A_n = A$. Let Ω be a set. $(\Omega, \mathcal{D})$ is a **Dynkin system** if $\mathcal{D}$ is a set of subsets of Ω satisfying

$$\begin{aligned} &\Omega \in \mathcal{A}, \\ &A, B \in \mathcal{D},\ A \subset B \Rightarrow B \setminus A \in \mathcal{D}. \\ &A_n \in \mathcal{D}, A_n \nearrow A \Rightarrow A \in \mathcal{D} \end{aligned}$$

Lemma 1.1.2 *$(\Omega, \mathcal{A})$ is a σ-algebra if and only if it is a π-system and a Dynkin system.*

Proof. If $\mathcal{A}$ is a σ-algebra, then it is clearly a π-system and a Dynkin system. Assume now, $\mathcal{A}$ is both a π-system and a Dynkin system. Given $A, B \in \mathcal{A}$. The Dynkin property implies that $A^c = \Omega \setminus A, B^c = \Omega \setminus B$ are in $\mathcal{A}$ and by the π-system property also $A \cup B = \Omega \setminus (A^c \cap B^c) \in \mathcal{A}$. Given a sequence $A_n \in \mathcal{A}$. Define $B_n = \bigcup_{k=1}^n A_k \in \mathcal{A}$ and $A = \bigcup_n A_n$. Then $A_n \nearrow A$ and by the Dynkin property $A \in \mathcal{A}$. Also $\bigcap_n A_n = \Omega \setminus \bigcup_n A_n^c \in \mathcal{A}$ so that $\mathcal{A}$ is a σ-algebra. □

Definition. If $\mathcal{I}$ is any set of subsets of Ω, we denote by $d(\mathcal{I})$ the smallest Dynkin system, which contains $\mathcal{I}$ and call it the **Dynkin system generated by** $\mathcal{I}$.

Lemma 1.1.3 (Dynkin lemma) *If $\mathcal{I}$ is a π- system, then $d(\mathcal{I}) = \sigma(\mathcal{I})$.*

Proof. By the last lemma, we need only to show that $d(\mathcal{I})$ is a $\pi-$ system.
(i) Define $\mathcal{D}_1 = \{B \in d(\mathcal{I}) \mid B \cap C \in d(\mathcal{I}), \forall C \in \mathcal{I}\}$. Because $\mathcal{I}$ is a π-system, we have $\mathcal{I} \subset \mathcal{D}_1$.
Claim. $\mathcal{D}_1$ is a Dynkin system.
Proof. Clearly $\Omega \in \mathcal{D}_1$. Given $A, B \in \mathcal{D} - 1$ with $A \subset B$. For $C \in \mathcal{I}$ we compute $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$ which is in $d(\mathcal{I})$. Therevore $A \setminus B \in \mathcal{D}_1$. Given $A_n \nearrow A$ with $A_n \in \mathcal{D}_1$ and $C \in \mathcal{I}$. Then $A_n \cap C \nearrow A \cap C$ so that $A \cap C \in d(\mathcal{I})$ and $A \in \mathcal{D}_1$.
(ii) Define $\mathcal{D}_2 = \{A \in d(\mathcal{I}) \mid B \cap A \in d(\mathcal{I}), \forall B \in d(\mathcal{I})\}$. From (i) we know that $\mathcal{I} \subset \mathcal{D}_2$. Like in (i), we show that $\mathcal{D}_2$ is a Dynkin-system. Therefore $\mathcal{D}_2 = d(\mathcal{I})$, which means that $d(\mathcal{I})$ is a π-system.
□

Proof of Lemma 1.1.4. The set $\mathcal{D} = \{A \in \sigma(\mathcal{I}) \mid \mu(A) = \nu(A)\}$ is Dynkin system: clearly $\Omega \in \mathcal{D}$. Given $A, B \in \mathcal{D}, A \subset B$. Then $\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$ so that $B \setminus A \in \mathcal{D}$. Given $A_n \in \mathcal{D}$ with $A_n \nearrow A$, then the σ additivity gives $\mu(A) = \limsup_n \mu(A_n) = \limsup_n \nu(A_n) = \nu(A)$, so that $A \in \mathcal{D}$. Since $\mathcal{D}$ is a Dynkin system containing the π-system $\mathcal{I}$, we know that $\sigma(\mathcal{I}) = d(\mathcal{I}) \subset \mathcal{D}$ which means that $\mu = \nu$ on $\sigma(\mathcal{I})$. □

Definition. Given a probability space $(\Omega, \mathcal{A}, P)$. Two π-systems $\mathcal{I}, \mathcal{J} \subset \mathcal{A}$ are called **P-independent**, if for all $A \in \mathcal{I}$ and $B \in \mathcal{J}$, $P(A \cap B) = P(A) \cdot P(B)$.

Lemma 1.1.4 *Let $\mathcal{G}, \mathcal{H}$ be two σ subalgebras of $\mathcal{A}$ and $\mathcal{I}$ and $\mathcal{J}$ be two π-systems satisfying $\sigma(\mathcal{I}) = \mathcal{G}$, $\sigma(\mathcal{J}) = \mathcal{H}$. Then $\mathcal{G}$ and $\mathcal{H}$ are independent if and only if $\mathcal{I}$ and $\mathcal{J}$ are independent.*

Proof. (i) Fix $I \in \mathcal{I}$ and define on $(\Omega, \mathcal{H})$ the measures $\mu(H) = P(I \cap H), \nu(H) = P(I)P(H)$ of total probabiliy $P(I)$. By the independence of $\mathcal{I}$ and $\mathcal{J}$, they agree on $\mathcal{J}$ and by the extension lemma, they agree on $\mathcal{H}$ and we have $P(I \cap H) = P(I)P(H)$ for all $I \in \mathcal{I}$ and $H \in \mathcal{H}$.
(ii) Define for fixed $H \in \mathcal{H}$ the measures $\mu(G) = P(G \cap H)$ and $\nu(G) = P(G)P(H)$ of total probabiliy $P(H)$ on $(\Omega, \mathcal{G})$. They agree on $\mathcal{I}$ and so on $\mathcal{G}$. We have shown that $P(G \cap H) = P(G)P(H)$ for all $G \in \mathcal{G}$ and all $H \in \mathcal{H}$. □

Definition. $\mathcal{A}$ is an **algebra** if $\mathcal{A}$ is a set of subsets of Ω satisfying

$$\begin{array}{l} \Omega \in \mathcal{A}, \\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \\ A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A} \end{array}$$

A σ-additive map from $\mathcal{A}$ to $[0, \infty)$ is called a **measure**.

Lemma 1.1.5 (Caratheodory continuation theorem) *Any measure on an algebra $\mathcal{R}$ can be continued uniquely to a measure on $\sigma(\mathcal{R})$.*

Proof. Definition. Let $\mathcal{A}$ be an algebra and $\lambda : \mathcal{A} \mapsto [0, \infty]$ with $\lambda(\emptyset) = 0$. A set $A \in \mathcal{A}$ is called a **λ-set**, if $\lambda(A \cap G) + \lambda(A^c \cap G) = \lambda(G)$ for all $G \in \mathcal{A}$.

Lemma 1.1.6 *The set $\mathcal{L}$ of λ sets of an algebra $\mathcal{A}$ is again an algebra and satisfies $\sum_{k=1}^n \lambda(A_k \cap G) = \lambda((\bigcup_{k=1}^n A_k) \cap G)$ for all finite disjoint families $\{A_k\}_{k=1}^n$ and all $G \in \mathcal{A}$.*

Proof. From the definition is clear that $\Omega \in \mathcal{L}$ and that if $B \in \mathcal{L}$, then $B^c \in \mathcal{L}$. Given $B, C \in \mathcal{L}$. Then $A = B \cap C \in \mathcal{L}$. Proof. Since $C \in \mathcal{L}$, we get

$$\lambda(C \cap A^c \cap G) + \lambda(C^c \cap A^c \cap G) = \lambda(A^c \cap G)$$

This can be rewritten with $C \cap A^c = C \cap B^c$ and $C^c \cap A^c = C^c$ as

$$\lambda(A^c \cap G) = \lambda(C \cap B^c \cap G) + \lambda(C^c \cap G) . \tag{1.1}$$

Since B is a λ-set, we get using $B \cap C = A$.

$$\lambda(A \cap G) + \lambda(B^c \cap C \cap G) = \lambda(C \cap G) . \tag{1.2}$$

Since C is a λ set, we have

$$\lambda(C \cap G) + \lambda(C^c \cap G) = \lambda(G) . \tag{1.3}$$

Adding up these three equations gives that $B \cap C$ is a λ set. We have shown that Λ is an algebra. If B and C are disjoint in $\mathcal{L}$ we get since B is a λ set

$$\lambda(B \cap (B \cup C) \cap G) + \lambda(B^c \cap (B \cup C) \cap G) = \lambda((B \cup C) \cap G) .$$

This can be rewritten as $\lambda(B \cap G) + \lambda(C \cap G) = \lambda((B \cup C) \cap G)$. The analogue statement for finitely many sets is obtained by induction. □

Definition. Let $\mathcal{A}$ be a $\sigma-$ algebra. A map $\lambda : \mathcal{A} \Rightarrow [0, \infty]$ is called an **outer measure**, if

$\lambda(\emptyset) = 1$,
$A, B \in \mathcal{A}$ with $A \subset B \Rightarrow \lambda(A) \leq \lambda(B)$.
$A_n \in \mathcal{A} \;\Rightarrow \lambda(\bigcup_n A_n) \leq \sum_n P(A_n)$ (σ subadditivity)

Lemma 1.1.7 (Carathéodory's lemma) *If λ is an outer measure on a measurable space $(\Omega, \mathcal{A})$, then the λ sets $\mathcal{L} \subset \mathcal{A}$ form a σ-algebra on which λ is countably additive.*

Proof. Given a disjoint sequence $A_n \in \mathcal{L}$. We have to show that $A = \bigcup_n A_n \in \mathcal{L}$ and $\lambda(A) = \sum_n \lambda(A_n)$. By the above lemma, $B_n = \bigcup_{k=1}^n A_k$ is in $\mathcal{L}$. We have therefore using the monotonicity, the additivity proved in the above lemma and the σ -subadditivity

$$\begin{aligned} \lambda(G) &= \lambda(B_n \cap G) + \lambda(B_n^c \cap G) \geq \lambda(B_n \cap G) + \lambda(A^c \cap G) \\ &= \sum_{k=1}^n \lambda(A_k \cap G) + \lambda(A^c \cap G) \geq \lambda(A \cap G) + \lambda(A^c \cap G) . \end{aligned}$$

Subadditivity for λ gives $\lambda(G) \leq \lambda(A \cap G) + \lambda(A^c \cap G)$. All the inequalities in this proof are therefore equalities. We conclude that $A \in \mathcal{L}$ and that λ is σ additive on $\mathcal{L}$. □

Proof of Caratheodory's continuation theorem. Given an algebra $\mathcal{R}$ with a measure μ. Define $\mathcal{A} = \sigma(\mathcal{R})$ and the σ-algebra $\mathcal{P}$ consisting of all subsets of Ω. Define on $\mathcal{P}$ the function

$$\lambda(A) = \inf\{\sum_{n \in \mathbf{N}} \mu(A_n) \mid \{A_n\}_{n \in \mathbf{N}} \text{ sequence in} \mathcal{R} \text{ satisfying } A \subset \bigcup_n A_n\} .$$

(i) λ is an outer measure on $\mathcal{P}$.
$\lambda(\emptyset) = 0$ and $\lambda(A) \geq \lambda(B)$ for $B \geq A$ are obvious. To see the σ subadditivity take a sequence $A_n \in \mathcal{P}$ with $\lambda(A_n) < \infty$ and fix $\epsilon > 0$. For all $n \in \mathbf{N}$, one can (by the definition of λ) find a sequence $\{B_{n,k}\}_{k \in \mathbf{N}}$ in $\mathcal{R}$ such that $A_n \subset \bigcup_{k \in \mathbf{N}} B_{n,k}$ and $\sum_{k \in \mathbf{N}} \mu(B_{n,k}) \leq \lambda(A_n) + \epsilon 2^{-n}$. Define $A = \bigcup_{n \in \mathbf{N}} A_n \subset \bigcup_{n,k \in \mathbf{N}} B_{n,k}$, so that $\lambda(A) \leq \sum_{n,k} \mu(B_{n,k}) \leq \sum_n \lambda(A_n) + \epsilon$. Since ϵ was arbitrary, the σ-subadditivity is proven.

(ii) $\lambda = \mu$ on $\mathcal{R}$.
Given $A \in \mathcal{R}$. Clearly $\lambda(A) \leq \mu(A)$. Suppose that $A \subset \bigcup_n A_n$, with $A_n \in \mathcal{R}$. Define a sequence $\{B_n\}_{n \in \mathbf{N}}$ of disjoint sets in $\mathcal{R}$ defined inductivly by $B_1 = A_1$, $B_n = A_n \cap (\bigcup_{k<n} A_k)^c$ such that $B_n \subset A_n$ and $\bigcup_n B_n = \bigcup_n A_n \supset A$. From the σ-additivity of μ on $\mathcal{R}$, we get

$$\mu(A) \leq \bigcup_n \mu(B_n) \leq \bigcup_n \mu(A_n)$$

so that $\mu(A) \geq \lambda(A)$.

(iii) Let $\mathcal{L}$ be the set of λ-sets in $\mathcal{P}$. Then $\mathcal{R} \subset \mathcal{L}$.
Given $A \in \mathcal{R}$ and $G \in \mathcal{P}$. There exists a sequence $\{B_n\}_{n \in \mathbf{N}}$ in $\mathcal{R}$ such that $G \subset \bigcup_n B_n$ and $\sum_n \mu(B_n) \leq \lambda(G) + \epsilon$. By the definition of λ

$$\sum_n \mu(B_n) = \sum_n \mu(A \cap B_n) + \sum_n \mu(A^c \cap B_n) \geq \lambda(A \cap G) + \lambda(A^c \cap G)$$

because $A \cap G \subset \bigcup_n A \cap B_n$ and $A^c \cap G \subset \bigcup_n A^c \cap B_n$. Since ϵ is arbitary, we get $\lambda(A) \geq \lambda(A \cap G) + \lambda(A^c \cap G)$. On the other hand, since λ is subadditive, we have also $\lambda(A) \leq \lambda(A \cap G) + \lambda(A^c \cap G)$ and A is a λ-set.

(iv) By Carathéodory's lemma, λ is a measure on $(\Omega, \mathcal{L})$. Since by step *(iii)* $\mathcal{R} \subset \mathcal{L}$, we know by Caratheodories lemma that $\mathcal{A} \subset \mathcal{L}$ so that we can define μ on $\mathcal{A}$ as the restriction of λ to $\mathcal{A}$. By step *(ii)*, this is an extension of the measure μ on $\mathcal{R}$. □

1.2 Kolmogorov's $0-1$ law, Borel-Cantelli lemma

Given a family $\{\mathcal{A}_i\}_{i\in I}$ of σ-subalgebras of $\mathcal{A}$. For any nonempty $J \subset I$, let $\mathcal{A}_J := \bigvee_{j\in J} \mathcal{A}_j$ be the σ-algebra generated by $\bigcup_{j\in J} \mathcal{A}_j$. Define also $\mathcal{A}_\emptyset = \{\emptyset, \Omega\}$. The **tail σ-algebra** $\mathcal{T}$ of $\{\mathcal{A}\}_{i\in I}$ is defined as $\mathcal{T} = \bigcap_{J\subset_f I} \mathcal{A}_{J^c}$, where $\mathcal{J}^c = \mathcal{I} \setminus \mathcal{I}$.

Theorem 1.2.1 (Kolmogorov's $0-1$ law) *If $\{\mathcal{A}_i\}_{i\in I}$ are independent, then the tail σ algebra $\mathcal{T}$ is P-trivial: $P(A) = 0$ or $P(A) = 1$ for every $A \in \mathcal{T}$.*

Proof. (i) The algebras $\mathcal{A}_F$ and $\mathcal{A}_G$ are independent, whenever $F, G \subset I$ are disjoint.
Proof. Define for $H \subset I$ the π-system

$$\mathcal{I}_H = \{A \in \mathcal{A} \mid A = \bigcap_{i\in K} A_i, K \subset_f H, A_i \in \mathcal{A}_i\} .$$

The π-systems $\mathcal{I}_F$ and $\mathcal{I}_G$ are independent and generate the σ algebras $\mathcal{A}_F$ and $\mathcal{A}_G$. Use Lemma 1.1.4.
(ii) Especially: $\mathcal{A}_J$ is independent of $\mathcal{A}_{J^c}$ for every $J \subset I$.
(iii) $\mathcal{T}$ is independent of $\mathcal{A}_I$.
Proof. $\mathcal{T} = \bigcap_{J\subset_f I} \mathcal{A}_{J^c}$ is independent of any $\mathcal{A}_K$ for $K \subset_f I$. It is therefore independent to the π-system $\mathcal{I}_I$ which generates $\mathcal{A}_I$. Use again Lemma 1.1.4.
(iv) $\mathcal{T}$ is a sub-σ algebra of $\mathcal{A}_I$. Therefore $\mathcal{T}$ is independent of itself which implies that it is P-trivial. □

Application. Let X_n be a sequence of real independent random variables and

$$A = \{\omega \in \Omega \mid \sum_{n=1}^{\infty} X_n \text{ converges}\} .$$

Then $P(A) = 0$ or $P(A) = 1$. Proof. Because $\sum_{n=1}^{\infty} X_n$ converges if and only if $\sum_{n=N}^{\infty} X_n$ converges, $A \in \sigma(A_N, A_{N+1} \dots)$ and so $A \in \mathcal{T}$, the tail σ- algebra defined by the independent σ-algebras $\mathcal{A}_n = \sigma(X_n)$. If for example, X_n takes values $\pm 1/n$ each with probability $1/2$, then $P(A) = 0$. If X_n takes values $\pm 1/n^2$ each with probability $1/2$, then $P(A) = 1$. As you might guess, this has to do with the convergence or divergence of series and we will come to that later.

Application. Let $\{A_n\}_{n\in\mathbf{N}}$ be a sequence of of subsets of Ω. The set $A_\infty := \limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n\geq m} A_n$ consists of the set $\{\omega \in \Omega\}$ such that $\omega \in A_n$ for infinitely many $n \in \mathbf{N}$. The set A_∞ is contained in the tail σ algebra of $\mathcal{A}_n = \{\emptyset, A, A^c, \Omega\}$. It follows from Kolmogorov's $0-1$ law that $P(A_\infty) \in \{0,1\}$ if $A_n \in \mathcal{A}$ and $\{A_n\}$ are P independent.

Remark. (Relation with ergodic theory). In the theory of dynamical systems, a measurable map T of a probability space $(\Omega, \mathcal{A}, P)$ is called a K-**system**, if there exists a $\sigma-$ subalgebra $\mathcal{F} \subset \mathcal{A}$ which satisfies $\mathcal{F} \subset \sigma(T(\mathcal{F}))$ and that the sequence $\mathcal{F}_n = \sigma(T^n(\mathcal{F}))$ satisfies $\mathcal{F}_{\mathbf{Z}} = \mathcal{A}$ and $\mathcal{T} = \{\emptyset, \Omega\}$. The fact that a Bernoulli automorphism $T(x)_n = x_{n+1}$ on $\Omega = \Delta^{\mathbf{Z}}$ is a K system is equivalent to Kolmogorov's $0-1$ law: take the algebra $\mathcal{F} = \bigvee_{k=1}^{\infty} T^k(\mathcal{F}_0)$, with $\mathcal{F}_0 = \{x \in \Omega = \Delta^{\mathbf{Z}} \mid x_0 = r \in \Delta\}$.

Theorem 1.2.2 (Borel-Cantelli-lemma) *1) Given a sequence $A_n \in \mathcal{A}$.*

$$\sum_{n \in \mathbf{N}} P(A_n) < \infty \Rightarrow P(A_\infty) = 0 .$$

2) Given a sequence $A_n \in \mathcal{A}$ of independent sets, then

$$\sum_{n \in \mathbf{N}} P(A_n) = \infty \Rightarrow P(A_\infty) = 1 .$$

Proof. (1) $P(A_\infty) = \lim_{n\to\infty} P(\bigcup_{k\geq n} A_k) \leq \lim_{n\to\infty} \sum_{k\geq n} P(A_k) = 0$.
(2) For every $n \in \mathbf{N}$, we have

$$P(\bigcap_{k\geq n} A_k^c) = \prod_{k\geq n} P(A_k^c) = \prod_{k\geq n}(1 - P(A_k)) \leq \prod_{k\geq n} \exp(-P(A_k)) \leq \exp(-\sum_{k\geq n} P(A_k)) = 0 .$$

From

$$P(A_\infty^c) = P(\bigcup_{n\in\mathbf{N}} \bigcap_{k\geq n} A_k^c) \leq \sum_{n\in\mathbf{N}} P(\bigcap A_k^c) = 0$$

follows $P(A_\infty^c) = 0$. □

Example. (The independence is necessary in 2): Take the probability space $([0,1], \mathcal{B}, P)$, where P is the Lebesgue measure on the Borel σ-algebra of $[0,1]$. For $A_n = [0, 1/n]$ we get $A_\infty = \emptyset$ and $P(A_n) = 1/n$.

Silly example. (Monkey typing Shakespeare) Writing the collected works of Shakespeare amounts to type a sequence of N symbols on a typewriter. A monkey types symbols at random, one per unit time, producing a random sequence X_n of identically distributed sequence of random variables in the set of all possible symbols. If each letter occurs with probability at least ϵ, then the probability that Shakespeare's work appears when typing the first N letters is ϵ^N. Call A_1 this event. Call A_k the event that this happens when typing the $(k-1)N+1$ until the kN'th letter. These sets A_k are all independent and have all equal probability. The Borel-Cantelli lemma, the events occur infinitely often. This means that Shakespeare is not only written once but infinitely many times.
To model this example precicely, we have to construct a probability space $(\Omega, \mathcal{A}, P)$. Take a finite alphabet Δ which is a compact topological space with the discrete topology. Form the product space $\Omega = \Delta^{\mathbf{Z}}$ which is by Tychonov compact and let $\mathcal{A}$ the Borel-σ-algebra on Ω. If we put on Δ a probability measure Q, the probability measure $Q^{\mathbf{Z}}$ exists on $(\Omega, \mathcal{A})$. It has the property that any zylinder set $Z(w) = \{\omega \in \Omega \mid \omega_k = r_k, \dots \omega_n = r_n\}$ defined by a word $w = [r_k, , \dots r_n]$ has the measure $P(Z(w)) = \prod_{i=k}^n P(\omega_i = r_i) = \prod_{i=k}^n Q(r_i)$. Finite unions of zylinder sets form an algebra $\mathcal{R}$ satisfying $\sigma(\mathcal{R}) = \mathcal{A}$. One can show that P is σ additive on this algebra. By Carathéodori's continuation theorem of measure theory, there exists a measure P on $(\Omega, \mathcal{A})$. The process of typing is given by the sequence of independent random variables $X_n(\omega) = \omega_n$. The event that Sheakespeare is written in the time interval $[Nk+1, N(k+1)]$ is given by a zylinder set A_k. They have all the same probability. Use Borel-Cantelli 2).

1.3 Appendix: Monkey typing Shakespeare

The following Mathematica procedures allow to simulate the typing of the Monkey:

```
Alphabet={a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z};
Monkey[n_]:=Table[Alphabet[[Random[Integer,{1,26}]]],{n}]
```

We let him type 700 letters on the computer and save it in a file Sheakspeare.

```
Shakespeare=Monkey[700];
Save["Shakespeare",Shakespeare];
```

and here is the output in the file Shakespeare:

```
Shakespeare = {x, t, h, i, s, p, r, o, b, a, b, i, l, i, t, y, c, o, u, r, s,
   e, i, s, g, r, e, a, t, x, i, a, j, b, z, v, j, j, v, n, s, z, g, e, x, v,
   x, f, b, i, u, h, a, f, c, o, z, t, m, b, x, u, v, o, a, n, b, d, t, r, s,
   d, u, c, c, m, y, f, g, e, j, n, k, f, g, u, x, c, o, f, a, w, k, o, k, b,
   e, l, o, z, g, d, q, p, j, v, b, e, x, n, e, f, p, n, o, n, a, m, f, z, h,
   v, z, q, f, r, k, q, i, k, r, f, j, w, l, w, u, n, x, q, m, m, x, m, j, z,
   b, p, b, h, n, y, g, m, p, q, v, x, k, v, c, p, m, z, b, m, d, w, e, u, i,
   t, h, i, s, p, r, o, b, a, b, i, l, i, t, y, c, o, u, r, s, e, i, s, g, r,
   e, a, t, y, d, a, p, k, j, c, m, o, q, k, r, r, k, f, d, a, j, s, f, t, f,
   o, v, x, w, o, r, u, j, i, s, p, q, o, v, a, h, i, o, x, v, j, q, z, o, t,
   v, m, p, q, x, a, s, n, y, c, z, j, g, v, j, w, m, n, x, a, n, x, l, s, b,
   i, o, i, g, s, v, j, n, a, o, f, a, w, l, m, f, e, t, y, y, j, n, r, i, d,
   m, j, g, r, o, q, s, u, i, k, z, e, g, a, k, d, c, k, o, p, i, a, n, o, k,
   s, n, y, x, b, e, a, t, x, w, y, i, k, m, w, n, j, h, f, j, p, i, c, j, c,
   a, h, o, p, o, a, u, q, g, k, v, j, z, r, y, x, m, t, o, h, m, v, m, q, q,
   d, v, g, c, v, x, q, f, w, q, q, t, n, u, g, w, g, n, g, n, k, f, v, m, b,
   d, x, u, r, t, y, y, g, g, e, a, b, o, n, v, i, j, q, c, q, l, h, a, c, e,
   s, n, r, d, v, o, v, m, u, z, f, p, g, i, y, w, p, x, i, x, l, q, s, x, a,
   w, i, q, e, k, o, e, k, b, z, i, f, v, o, i, i, z, b, p, k, v, l, i, c, k,
   y, b, k, i, l, i, g, n, s, y, a, b, o, n, a, q, k, e, z, s, d, b, k, o, b,
   i, x, u, a, l, e, t, f, w, l, f, c, h, j, g, i, p, r, m, u, i, w, r, o, k,
   u, z, f, r, p, j, k, x, k, h, h, d, b, c, l, m, m, w, a, o, m, a, r, l, h,
   d, d, p, h, r, q, c, h, f, i, c, b, u, e, a, i, k, k, j, s, l, g, u, o, m,
   d, h, d, v, u, z, i, m, s, b, l, z, b, q, p, t, i, v, d, p, k, f, i, u, a,
   u, w, r, i, b, s, v, e, b, r, w, q, i, j, t, c, o, y, k, n, k, l, x, s, f,
   t, h, i, s, p, r, o, b, a, b, i, l, i, t, y, c, o, u, r, s, e, i, s, g, r,
   e, a, t, y, d, a, p, k, j, c, m, o, q, k, r, r, k, f, d, a, j, s, f, t, f,
```

```
m, m, a, s, q, r, b, v, x, c, b, k, o, f, h, l, t, p, c, e, k, t, q, m, m,
z, b, t, n}
```

Exercice 1

Topic: Probability spaces, independence

1) Let $\Omega = \mathbf{N}$. We experiment with some measures on Ω.
a) The metric $d(n,m) = |n-m|$ defines a topology $\mathcal{O}$ on Ω. What is the Borel σ-algebra $\mathcal{A}$ generated by this topology?
b) Show that for every $\lambda > 0$

$$P(A) = \sum_{n \in A} e^{-\lambda} \frac{\lambda^n}{n!}$$

is a probability measure on the measurable space $(\Omega, \mathcal{A})$ considered in *b*). (It is not necessary to use here Caratheodories extension theorem).
c) Show that for every $s > 1$

$$P(A) = \sum_{n \in A} \zeta(s)^{-1} n^{-s}$$

is a probability measure on the measurable space $(\Omega, \mathcal{A})$. The function $s \mapsto \zeta(s) = \sum_{n \in \Omega} n^{-s}$ is the Riemann zeta function).
(i) Show that the sets $A_p = \{n \in \Omega | \ p \text{ divides } n\}$ with prime p are independent. What happens if p is not prime.
(ii) Give a probabilistic proof of Euler's formula

$$\frac{1}{\zeta(s)} = \prod_{p \text{ prime}} (1 - p^{-s}) \ .$$

(iii) Let A be the set of natural numbers which are not divisable by a square different from 1. Prove

$$P(A) = \zeta(2s)^{-1} \ .$$

d)[1] Let's try to define a "natural" measure on Ω. Denote with $|A| \in \mathbf{N} \cup \{\infty\}$ the cardinality of a set $A \in \Omega$. Define

$$\mathcal{B} = \{A \subset \Omega \mid P(A) = \lim_{n \to \infty} |A \cap \{1, 2, \dots, n\}|/n \text{ exists}\} \ .$$

Show that $\mathcal{B}$ is not a π-system and that P is therefore not such a good idea for a measure.

2) a) Given a probability space $(\Omega, \mathcal{A}, P)$. Given three π systems $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$ on Ω which all contain Ω. Show that

$$P(I_1 \cap I_2 \cap I_3) = P(I_1)P(I_2)P(I_3), \forall (I_1, I_2, I_3) \in (\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3)$$

implies that $\sigma(\mathcal{I}_1), \sigma(\mathcal{I}_2), \sigma(\mathcal{I}_3)$ are independent.
b) Find a counterexample to *a*) if $\mathcal{I}_k$ are not required to contain Ω.

[1] Might need some experimentation time

1.4 Integration, Expectation, Variance

We fix a probability space $(\Omega, \mathcal{A}, P)$.

Definition. A **statement** S about points $\omega \in \Omega$ is a map from Ω to $\{\text{true}, \text{false}\}$. A statement is said to hold **almost everywhere**, if the set $P(\{\omega \mid S(\omega) = \text{false}\}) = 0$. We use this in statements like for example "let $X_n \to X$ almost everywhere".

Definition. We denoted by $\mathcal{L}$ the algebra of all random variables. A **step function** is an element of $\mathcal{L}$ which is of the form $X = \sum_{i=1}^n \alpha_i \cdot 1_{A_i}$ with $\alpha_i \in \mathbf{R}$ and where $A_i \in \mathcal{A}$ are disjoint. Denote with $\mathcal{S}$ the algebra of step functions. For $X \in \mathcal{S}$ we can define the *integral*

$$E[X] := \int_\Omega X \, dP = \sum_{i=1}^n \alpha_i P(A_i) .$$

Definition. Define $\mathcal{L}^1 \subset \mathcal{L}$ the set of random variables X, for which

$$\sup_{Y \in \mathcal{S}, Y \leq |X|} \int Y \, dP$$

is finite. For $X \in \mathcal{L}^1$, we can define the **integral** or **expectation**

$$E[X] := \int X \, dP = \sup_{Y \in \mathcal{S}, Y \leq X^+} \int Y \, dP - \sup_{Y \in \mathcal{S}, Y \leq X^-} \int Y \, dP$$

where $X^+ = X \vee 0 = \max(X, 0)$ and $X^- = -X \vee 0 = \max(-X, 0)$. The vector space $\mathcal{L}^1$ is called the space of *integrable random variables.*

We recall some results of measure theory and outline their proofs. (The proofs are obtained from any book in measure theory by replacing $\int f(x) \, d\mu(x)$ by $E[X]$ and f, g, h by X, Y, Z ...)

Theorem 1.4.1 (Monotone convergence theorem, Beppo Levi 1906) *Let X_n be a sequence in $\mathcal{L}^1$ with $0 \leq X_1 \leq X_2, \dots$ and assume $X = \lim_{n\to\infty} X_n$ converges pointwise. If $\sup_n E[X_n] < \infty$, then $X \in \mathcal{L}^1$ and*

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

Proof. Assume $X_n \geq 0$ (replace else X_n by $X_n - X_1$). Find for each n a monotone sequence of step functions $X_{n,m} \in \mathcal{S}$ with $X_n = \sup_m X_{n,m}$. Consider the sequence of step functions

$$Y_n := \sup_{1 \leq k \leq n} X_{k,n} \leq \sup_{1 \leq k \leq n} X_{k,n+1} \leq \sup_{1 \leq k \leq n+1} X_{k,n+1} = Y_{n+1} .$$

Since $Y_n \leq \sup_{m=1}^n X_m = X_n$ also $E[Y_n] \leq E[X_n]$. One checks that $\sup_n Y_n = X$ implies $\sup_n E[Y_n] = \sup_{Y \in \mathcal{S}, Y \leq X} E[Y]$ and concludes

$$E[X] = \sup_{Y \in \mathcal{S}, Y \leq X} E[Y] = \sup_n E[Y_n] \leq \sup_n E[X_n] \leq E[\sup_n X_n] = E[X] .$$

□

Theorem 1.4.2 (Fatou lemma, 1906) *Let X_n be a sequence in $\mathcal{L}^1$ with $|X_n| \leq X$ for some $X \in \mathcal{L}^1$. Then*

$$E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n] \leq \limsup_{n\to\infty} E[X_n] \leq E[\limsup_{n\to\infty} X_n] .$$

Proof. Since for $p \geq n$

$$\inf_{m\geq n} X_m \leq X_p \leq \sup_{m\geq n} X_m$$

we get

$$E[\inf_{m\geq n} X_m] \leq E[X_p] \leq E[\sup_{m\geq n} X_m]$$

and since $p \geq n$ was arbitrary, we have also

$$E[\inf_{m\geq n} X_m] \leq \inf_{p\geq n} E[X_p] \leq \sup_{p\geq n} E[X_p] \leq E[\sup_{m\geq n} X_m] .$$

Since $Y_n = \inf_{m\geq n} X_m$ is increasing with $\sup_n E[Y_n] < \infty$ and $Z_n = \sup_{m\geq n} X_m$ is decreasing with $\inf_n E[Z_n] > -\infty$ we get from Beppo-Levi that $Y = \sup_n Y_n = \limsup_n X_n$ and $Z = \inf_n Z_n = \liminf_n X_n$ are in $\mathcal{L}^1$ and

$$\begin{aligned} E[\liminf_n X_n] &= \sup_n E[\inf_{m\geq n} X_m] \leq \sup_n \inf_{m\geq n} E[X_m] = \liminf_n E[X_n] \\ &\leq \limsup_n E[X_n] = \inf_n \sup_{m\geq n} E[X_m] \leq \inf_n E[\sup_{m\geq n} X_m] = E[\limsup_n X_n] . \end{aligned}$$

□

Theorem 1.4.3 (Lebesgue's dominated convergence theorem, 1902) *let X_n be a sequence in $\mathcal{L}^1$ with $|X_n| \leq X$ for some $X \in \mathcal{L}^1$. If $X_n \to X$ almost everywhere, then $E[X_n] \to E[X]$.*

Proof. Since $X = \liminf_n X_n = \limsup_n X_n$ we know that $X \in \mathcal{L}^1$ and from Fatou lemma

$$E[X] = E[\liminf_n X_n] \leq \liminf_n E[X_n] \leq \limsup_n E[X_n] \leq E[\limsup_n X_n] = E[X] .$$

□

A special case of Lebesgue's dominated convergence theorem is the bounded dominated convergence theorem, which says that $E[X_n] \to E[X]$ if $X_n \leq K$ and $X_n \to X$ almost everywhere.

Definition. We define also for $p \in [1, \infty)$ the vector spaces $\mathcal{L}^p = \{X \in \mathcal{L} \mid |X|^p \in \mathcal{L}^1$ and $\mathcal{L}^\infty = \{X \in \mathcal{L} \mid \exists K \in \mathbf{R}\ X \leq K$, almost everywhere$\}$.

For $X \in \mathcal{L}^2$, we can define the **variance** $\text{Var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$. The number $\sigma[X] = \text{Var}[X]^{1/2}$ is called the **standard deviation** of X.

1.5 Some inequalities

Definition. A function h is *convex*, if there exists for all $x_0 \in \mathbf{R}$ a linear map $l(x) = ax + b$ such that $l(x_0) = h(x_0)$ and for all $x \in \mathbf{R}$ the inequality $l(x) \leq h(x)$ holds.

Theorem 1.5.1 (Jensen inequality) *Given $X \in \mathcal{L}^1$. For any convex function $h : \mathbf{R} \to \mathbf{R}$, we have*

$$E[h(X)] \geq h(E[X])$$

where the left hand side can also be infinite.

Proof. Let l be the linear map at $x_0 = E[X]$. By the linearity and monoticity of the expectation, we get

$$h(E[X]) = l(E[X]) = E[l(X)] \leq E[h(X)] .$$

□

Example. Given $p \leq q$. Define $h(x) = |x|^{q/p}$. Jensen's inequality gives $E[|X|^q] = E[h(|X|^p)] \leq h(E[|X|^p] = E[|X|^p]^{q/p}$. This implies that $||X||_q := E[|X|^q]^{1/q} \leq E[|X|^p]^{1/p} = ||X||_p$ for $p \leq q$ and so $\mathcal{L}^p \subset \mathcal{L}^q$ for $p \geq q$.

We have defined $\mathcal{L}^p$ as the set of random variables which satisfy $E[|X|^p] < \infty$ for $p \in [1, \infty)$ and $|X| \leq K$ almost everywhere for $p = \infty$. The vector space $\mathcal{L}^p$ has the seminorm $||X||_p = E[|X|^p]^{1/p}$ rsp. $||X||_\infty = \inf\{K \in \mathbf{R} \mid |X| \leq K$ almost everywhere $\}$. The seminorm property follows from

Theorem 1.5.2 (Minkovski inequality) *Given $p \in [1, \infty]$ $X, Y \in \mathcal{L}^p$. Then $||X + Y||_p \leq ||X||_p + ||Y||_p$.*

Proof. We use Hölder's inequality from below to get

$$E[|X + Y|^p] = E[|X||X + Y|^{p-1}] + E[|Y||X + Y|^{p-1}] \leq ||X||_p C + ||Y||_p C ,$$

where $C = |||X + Y|^{p-1}||_q = E[|X + Y|^p]^{1/q}$ which leads to the claim. □

One constructs from $\mathcal{L}^p$ a real **Banach space** by defining the quotient $L^p = \mathcal{L}^p/\mathcal{N}$, where $\mathcal{N} = \{X \in \mathcal{L}_p \mid ||X||_p = 0\}$. For $p = 2$, one gets a real Hilbert space with scalar product $< X, Y >= E[XY]$. The finiteness of the scalar product follows from

Theorem 1.5.3 (Hölder inequality, Hölder 1889) *Given $p, q \in [1, \infty]$ with $p^{-1} + q^{-1} = 1$ and $X \in \mathcal{L}^p$ and $Y \in \mathcal{L}^q$. Then $XY \in \mathcal{L}^1$ and*

$$||XY||_1 \leq ||X||_p ||Y||_q .$$

Proof. Can restrict to $X, Y \geq 0$ and $||X||_p > 0$. Define the probabilty measure $Q = \frac{X^p P}{E[X^p]}$ and define $u = 1_{\{X>0\}} Y/X^{p-1}$. Jensen gives $Q(u)^q \leq Q(u^q)$ so that

$$E[|XY|] \leq ||X||_p ||1_{\{Z>0\}} Y||_q \leq ||X||_p ||Y||_q \ .$$

□

A special case of Hölder's inequality is the **Cauchy-Bunyakowsky-Schwarz** inequality

$$||XY||_1 \leq ||X||_2 ||Y||_2 \ .$$

Notation. We use the short-hand notation $P(X \geq c)$ for $P(\{\omega \in \Omega \mid X(\omega) \geq c\})$.

Theorem 1.5.4 (Chebychev-Markov inequality) *Let h be a monoton function on $\mathbf{R}$ with $h \geq 0$. For every $c > 0$, and $h(X) \in \mathcal{L}^1$ we have*

$$h(c) \cdot P(X \geq c) \leq E[h(X)] \ .$$

Proof. Integrate the inequality $h(c) 1_{X \geq c} \leq h(X)$ using the monotonicity and linearity of the expectation. □

Example. $h(x) = |x|$ leads to $P(|X| \geq c) \leq ||X||_1/c$ which implies for example that $E[|X|] = 0 \Rightarrow P(X = 0) = 1$.

Theorem 1.5.5 (Chebychev inequality) *If $X \in \mathcal{L}^2$, then*

$$P(|X - E[X]| \geq c) \leq \frac{\text{Var}[X]}{c^2} \ .$$

Proof. Take $h(x) = x^2$ and apply the Chebychev-Markov inequality to the random variable $Y = X - E[X] \in \mathcal{L}^2$ satisfying $h(Y) \in \mathcal{L}^1$. □

We have defined the variance $\text{Var}[X] = E[(X - E[X])^2]$ for random variables $X \in \mathcal{L}^2$.

Definition. For $X, Y \in \mathcal{L}^2$ we can define the **covariance**

$$\text{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \ .$$

Two random variables in $\mathcal{L}^2$ are called *uncorrelated* if $\text{Cov}[X, Y] = 0$.

Remark. The Cauchy-Schwarz-inequality can be restated as $|\text{Cov}[X, Y]| \leq \sigma[X]\sigma[Y]$.

Definition. If the standard deviations $\sigma[X], \sigma[Y]$ are different from zero, then one can define the **correlation coefficient**

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sigma(X)\sigma(Y)}$$

which is a number in $[-1, 1]$. Two random variables in $\mathcal{L}^2$ are called **uncorrelated** if $\text{Corr}[X, Y] = 0$. The other extreme is $|\text{Corr}[X, Y]| = 1$, then $Y = aX + b$ by Cauchy-Schwarz.

Lemma 1.5.6 *If two random variables* $X, Y \in \mathcal{L}^2$ *are independent, then* $\mathrm{Cov}[X,Y] = 0$. *If* X *and* Y *are uncorrelated, then* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Proof. Find monotone sequences of step functions $X_n = \sum_{i=1}^n \alpha_i 1_{A_i} \to X$ $Y_n = \sum_{j=1}^n \beta_j \cdot 1_{B_i} \to Y$. We can choose these function such that $A_i \in \sigma(X)$ and $B_j \in \sigma(Y)$. By Lebesgue's dominated convergence theorem, $E[X_n] \to E[X]$ and $E[Y_n] \to E[Y]$ almost everywhere. Compute $X_n \cdot Y_n = \sum_{i,j=1}^n \alpha_i \beta_j 1_{A_i \cap B_j}$. By Lebesgue's dominated convergence theorem again, $E[X_n Y_n] \to E[XY]$. By the independence of X, Y we have $E[X_n Y_n] = E[X_n] \cdot E[Y_n]$ and so $E[XY] = E[X]E[Y]$ which implies $\mathrm{Cov}[X,Y] = E[XY] - E[X] \cdot E[Y] = 0$.
The second statement follows from $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$. □

1.6 The weak laws of large numbers

Let $X_1, X_2, \dots$ a sequence of random variables on a probability space $(\Omega, \mathcal{A}, P)$. We are interested in the asymptotic behavior of the sums $S_n = X_1 + X_2 + \dots X_n$ for $n \to \infty$ and especially in the convergence of the averages S_n/n. This behavior is described by "laws of large numbers". Depending on the definition of convergence, one speaks of "weak" and "strong" laws of large numbers.

We first prove the weak law of large numbers. There exist different versions of this theorem since more assumptions on X_n can allow stronger statements.

Definition. A sequence of random variables Y_n converges **in probability** to a random variable Y, if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P[|Y_n - Y| \geq \epsilon] = 0 \; .$$

One speaks also from **stochastic convergence**.

Remark. If for some $p \in [1, \infty)$, $||X_n - X||_p \to 0$, then $X_n \to X$ in probability since by the Chebychev-Markov inequality, $P[|X_n - X| \geq \epsilon] \leq ||X - X_n||^p/\epsilon^p$.

Theorem 1.6.1 (Weak law of large numbers for uncorrelated $X_i \in \mathcal{L}^2$) *Assume $X_i \in \mathcal{L}^2$ have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n}\sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $S_n/n \to m$ in probability.*

Proof. Since $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\cdot\mathrm{Cov}[X,Y]$ and X_n are pairwise uncorrelated, we get $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_n]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\mathrm{Var}[S_n/n] = E[(S_n)^2/n^2] - E[S_n]^2/n^2 = \mathrm{Var}[S_n]/n^2 = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[X_n] \to 0 \; .$$

With Chebychev's inequality, we obtain

$$P[|S_n/n - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} \; .$$

□

As an application in analysis, this leads to a constructive proof of the theorem of Weierstrass which sais that polynomials are dense in $C[0,1]$:

Corollary 1.6.2 (Weierstrass with more information) *For every $f \in C[0,1]$, the* **Bernstein polynomials**

$$B_n(x) = \sum_{k=1}^n f(k/n) \begin{pmatrix} k \\ n \end{pmatrix} x^k (1-x)^{n-k}$$

converge uniformly to f. If $f(x) \geq 0$, then also $B_n(x) \geq 0$.

Proof. For $x \in [0,1]$, let X_n be a sequence of independent $\{0,1\}$- valued random variables with mean value x. In other words, we take the probability space $(\{0,1\}^{\mathbf{N}}, \mathcal{A}, P)$ defined by $P[\omega_n = 1] = x$. Since $P[S_n = k] = \left(\begin{array}{c} k \\ n \end{array} \right) p^k (1-p)^{n-k}$, we can write $B_n(x) = E[f(\frac{S_n}{n})]$. We estimate now

$$\begin{aligned} |B_n(x) - f(x)| &= |E[f(S_n/n)] - f(x)]| \le E[|f(S_n/n) - f(x)|] \\ &\le 2||f|| \cdot P[|S_n/n - x| \ge \delta] + \sup_{|x-y|\le\delta} |f(x) - f(y)| \cdot P[|S_n/n - x| < \delta] \\ &\le 2||f|| \cdot P[|S_n/n - x| \ge \delta] + \sup_{|x-y|\le\delta} |f(x) - f(y)| \; . \end{aligned}$$

The second term, the continuity module of f goes to zero for $\delta \to 0$. By Chebychev and the proof of the weak law, the first term can be estimated above by $2||f||\text{Var}[X_i]/(n\delta^2)$ which goes to zero because the variance satisfies $\text{Var}[X_i] = x(1-x) \le 1/4$. □

In the first version of the weak law of large numbers Theorem 1.6.1, we only assumed the random variables to be uncorrelated. Under the stronger condition of independece and a stronger conditions on the moments, the convergence can be accelerated:

Theorem 1.6.3 (Weak law of large numbers for independent $X_i \in \mathcal{L}^4$) *Assume $X_i \in \mathcal{L}^4$ have common expectation $E[X_i] = m$ and satisfy $M = \sup_n ||X||_4 < \infty$. If X_i are independent, then $S_n/n \to m$ in probability. More precicely, $\sum_n P[|S_n/n - m| \ge \epsilon]$ converges for all $\epsilon > 0$.*

Proof. We can assume without loss of generality that $m = 0$. Because the X_i are independent, we get

$$E[S_n^4] = \sum_{i_1,i_2,i_3,i_4=1}^{n} E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] \; .$$

Again by independence, a summand $E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$ is vanishing if an index $i = i_k$ occurs allone, is $E[X_i^4]$ if all indices are the same and $E[X_i^2]E[X_j^2]$, if there are two pairwise equal indices. Since by Jensen's inequality $E[X_i^2]^2 \le E[X_i^4] \le M$ we get

$$E[S_n^4] \le nM + n(n+1)M \; .$$

Use now the generalized Chebychev inequality with $h(x) = x^4$ to get

$$\begin{aligned} P[|S_n/n| \ge \epsilon] &\le \frac{E[(S_n/n)^4]}{\epsilon^4} \\ &\le M \frac{n + n^2}{\epsilon^4 n^4} \le 2M \frac{1}{\epsilon^4 n^2} \; . \end{aligned}$$

□

We can also weaken the moment assumption to treat $\mathcal{L}^1$ random variables. Of course, the assumptions have to be stronger at an other place.

Definiton: A family $\{X_i\}_{i \in I}$ of random variables is called **uniformly integrable**, if $\sup_{i\in I} E[1_{|X_i| \ge R}] \to 0$ for $R \to \infty$. A convenient notation is $E[1_A X] = E[X; A]$ for $X \in \mathcal{L}^1$ and $A \in \mathcal{A}$.

Theorem 1.6.4 (Weak law of large numbers for independent $X_i \in \mathcal{L}^1$) *Assume $X_i \in \mathcal{L}^1$ are uniformly integrable. If X_i are independent, then $\frac{1}{n}\sum_{i=1}^n (X_m - E[X_m]) \to 0$ in $\mathcal{L}^1$ and therefore in probability.*

Proof. Without loss of generality, we can assume that $E[X_n] = 0$ for all $n \in \mathbf{N}$ since we can replace else X_n by $Y_n = X_n - E[X_n]$.
Define $f_R(t) = t1_{[-R,R]}$, the random variables

$$X_n^{(R)} = f_R(X_n) - E[f_R(X_n)], Y_n^{(R)} = X_n - X_n^{(R)}$$

as well as

$$S_n^{(R)} = \frac{1}{n}\sum_{i=1}^n X_n^{(R)}, T_n^{(R)} = \frac{1}{n}\sum_{i=1}^n Y_n^{(R)} .$$

We estimate, using Minkowsky and Cauchy-Schwarz

$$\begin{aligned} ||S_n||_1 &\leq ||S_n^{(R)}||_1 + ||T_n^{(R)}||_1 \\ &\leq ||S_n^{(R)}||_2 + 2 \sup_{1 \leq l \leq n} E[|X_l|; |X_l| \geq R] \\ &\leq \frac{R}{\sqrt{n}} + 2 \sup_{l \in \mathbf{N}} E[|X_l|; |X_l| \geq R] . \end{aligned}$$

In the last step we have used the independence of the random variables and $E[X_n^{(R)}] = 0$ to get

$$||S_n^{(R)}||_2^2 = E[(S_n^{(R)})^2] = E[(X_n^{(R)})^2]/n \leq R^2/n .$$

The claim follows from the uniform integrabiltiy assumption
$\sup_{l \in \mathbf{N}} E[|X_l|; |X_l| \geq R] \to 0$ for $R \to \infty$ □

Definition: The **law** of a random variable X is the probability measure μ on $\mathbf{R}$ defined by $\mu(Y) = P(X^{-1}Y)$. The measure μ is also called the **push-forward measure** under the measurable map $X : \Omega \to \mathbf{R}$. The **distribution function** of X is defined as $F_X(x) = \mu((-\infty, x])$. A set of random variables is called **identically distributed** if they have all the same distribution function. A common abreviation for independent identically distributed random variables is IID.

A special case of the last "weak law of large numbers"-version is when the random variables are IID. To prove it we need first a little lemma:

Lemma 1.6.5 (Absolutely continuity property of a $X \in \mathcal{L}^1$) *Given $X \in \mathcal{L}^1$. For every $\epsilon > 0$ there exists $\delta > 0$ such that $A \in \mathcal{A}$ with $P(A) < \delta$ implies $E[|X|; A] = E[1_A|X|] < \epsilon$.*

Proof. If the conclusion is false, there exists $\epsilon > 0$ and a sequence $A_n \in \mathcal{A}$ with $P(A_n) \leq 2^{-n}$ and $E[|X|; A_n] \geq \epsilon$. Define $A_\infty = \limsup_n A_n$. By Borel-Cantelli, we have $P[A_\infty] = 0$. Define $Y_n = 1_{B_n}|X|, B_n = \bigcup_{k \geq n} A_k$ and $Y = 1_{A_\infty}|X|$. Then $Y_n \to Y$ pointwise a.e. and $|Y_n| \leq |X|$. By Lebesgue's dominated convergence theorem, $\epsilon \leq E[Y_n] \to E[Y]$ which implies $P[A_\infty] > 0$. Contradiction. □

Theorem 1.6.6 (Weak law of large numbers for IID $X_i \in \mathcal{L}^1$) *Assume $X_i \in \mathcal{L}^1$ are IID random variables with mean m. Then $S_n/n \to m$ in $\mathcal{L}^1$ and so in probability.*

Proof. We show that a set of IID $\mathcal{L}^1$ random variables is uniformly integrable: given $X \in \mathcal{L}^1$, we have $K \cdot P[|X| > K] \leq ||X||_1$ so that $P[|X| > K] \to 0$ for $K \to \infty$. By the above lemma, we have also $P[|X|; |X| \geq R] \to 0$. Since X_i are identically distributed, $P[|X_i|; |X_i| \geq R]$ is independent of i so that any set of IID random variables is uniformly integrable. □

Exercice 2

Topic: $\mathcal{L}^p$ random variables,Weak law

1) Show that if $X, Y \in \mathcal{L}^1$ are independent random variables, then $XY \in \mathcal{L}^1$. Find an example of two random variables $X, Y \in \mathcal{L}^1$ for which $XY \notin \mathcal{L}^1$.

2) Find the "theorem of Pythagoras" in the notes.

3) Show that if two random variables $X, Y \in \mathcal{L}^2$ have nonvanishing variance and satisfy $|\text{Corr}(X,Y)| = 1$, then $Y = aX + b$ for some real numbers a, b.

4) [2] a) Given a sequence $p_n \in [0,1]$. Given a sequence X_n of IID random variables taking values in $\{-1, 1\}$ such that $P[X_n = 1] = p_n$ and $P[X_n = -1] = 1 - p_n$. Show that

$$n^{-1} \sum_{k=1}^{n} (X_k - m_k) \to 0$$

in probability, where $m_k = 2p_k - 1$.

b) The same set up like in a) but this time, the sequence p_n is dependent on a parameter. Given a sequence X_n of independent random variables taking values in $\{-1, 1\}$ such that $P[X_n = 1] = p_n$ and $P[X_n = -1] = 1 - p_n$ with $p_n = (1 + \cos[\theta + n\alpha])/2$, where θ is a parameter. Prove that $\frac{1}{n} \sum_n X_n \to 0$ in $\mathcal{L}^1$ for almost all θ. (You can assume for granted the fact that $1/n \sum_{k=1}^{n} p_k \to 1/2$ for almost all θ.)

5) Prove that $X_n \to X$ in $\mathcal{L}^1$, then there exists of a subsequence $Y_n = X_{n_k}$ satisfying $Y_n \to X$ almost everywhere.

   Hint. Stare for a while at the proposition showing the relation between convergences.

6) Give an example of a sequence of random variables X_n which converges almost everywhere but not completely.

---

[2]The previous formulation of this question contained some mistakes

Appendix: An experiment

With the following Mathematica program we can simulate random walks, where the law p is depending on time also. This models a stock market with superimposed fluctuations of slower frequency like inflation. Also the average temperature of the earth shows such behavior (ice ages).

```
theta=Random[]; alpha=N[Sqrt[1/301]]; pi2=N[2Pi];
X[x_]:=x+Module[{},theta=Mod[theta+alpha,pi2];
  If[2Random[]<1+Cos[theta],1,-1]];
Y[x_]:=x+2 Random[]-1;
Display["!psfix -land -stretch > randomwalk.ps",
Show[{ListPlot[NestList[X,0,1000]],ListPlot[NestList[Y,0,1000]]}]];
```

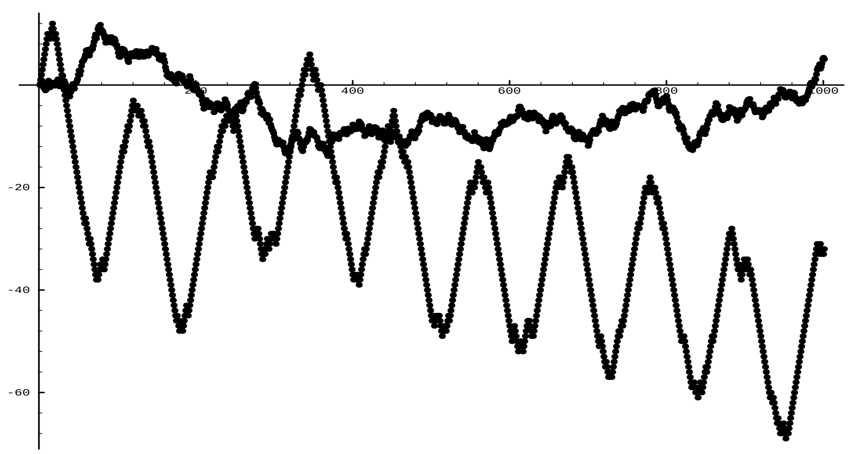


Figure 1. Comparison between the usual random walk and a random walk where the profit (win probability) p_n is depending on time. This plot is the outcome of the above program.

1.7 Convergence of random variables

In order to formulate the strong law of large numbers, we need some other notions of convergence. The first three of the following definitions we have met already.

Definition. A sequence of random variables X_n converges **in probability** to a random variable X, if $P[|X_n - X| \geq \epsilon] \to 0$ for all $\epsilon > 0$.

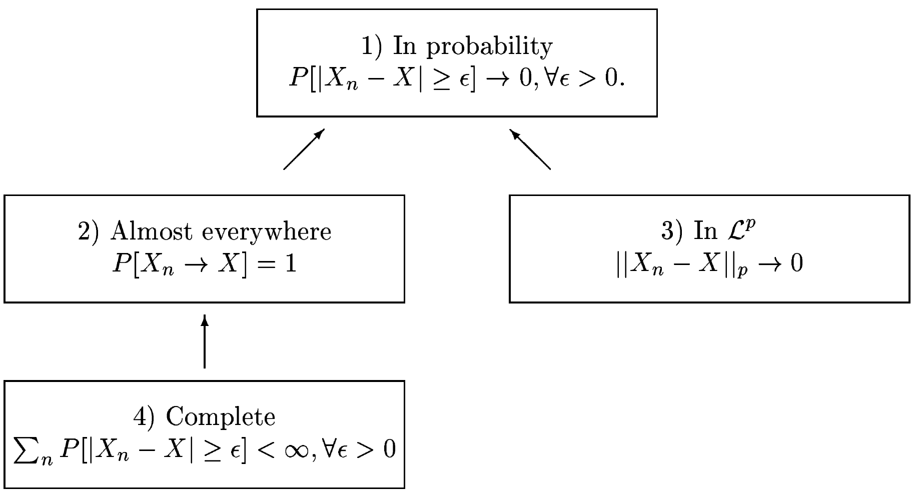
Definition. A sequence of random variables X_n converges **almost everywhere** to a random variable X, if $P[X_n \to X] = 1$.

Definition. A sequence of $\mathcal{L}^p$ random variables X_n **converges in** $\mathcal{L}^p$ to a random variable X, if $||X_n - X||_p \to 0$ for $n \to \infty$.

Definition. A sequence of random variables X_n converges **fast in probability**, or **completely** $\sum_n P[|X_n - X| \geq \epsilon] < \infty$ for all $\epsilon > 0$.

We have now four notions of convergence of random variables $X_n \to X$:

Proposition 1.7.1 (Relations between convergences) :



Proof. 2) $\Rightarrow$ 1): Since

$$\{X_n \to X\} = \bigcap_k \bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}$$

"almost everywhere convergence" is equivalent with

$$1 = P[\bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}] = \lim_{n \to \infty} P[\bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}]$$

for all k. Therefore

$$P[|X_m - X| \geq \epsilon] \leq P[\bigcap_{n \geq m} \{|X_n - X| \geq \epsilon\}] \to 0$$

for all $\epsilon > 0$.
4) $\Rightarrow$ 2): The Borel-Cantelli lemma implies that for all $\epsilon > 0$

$$P[|X_n - X| \geq \epsilon, \text{ infinitely often}] = 0$$

We get so for $\epsilon_n \to 0$

$$P[\bigcup_n |X_n - X| \geq \epsilon_k, \text{ infinitely often}] \leq \sum_n P[|X_n - X| \geq \epsilon_k, \text{ infinitely often}] = 0$$

from which we obtain $P[X_n \to X] = 1$.
3) $\Rightarrow$ 1): Use Chebychev-Markov inequality to get $P[|X_n - X| \geq \epsilon] \leq E[|X_n - X|^p]/\epsilon^p$. $\square$

Examples.
a) (Stochastic but not almost everywhere convergence). Let $([0,1], \mathcal{A}, P)$ be the Lebesgue measure space, where $\mathcal{A}$ is the Borel $\sigma-$ algebra on $[0,1]$. Define the random variables

$$X_{n,k} = 1_{[k2^{-n},(k+1)2^{-n}]},\ n = 1,2,\dots,\ k = 0,\dots,2^n - 1 .$$

By lexicographical ordering $X_1 = X_{1,1}, X_2 = X_{2,1}, X_3 = X_{2,2}, X_4 = X_{2,3}, \dots$ we get a sequence X_n satisfying

$$\liminf_{n\to\infty} X_n(\omega) = 0, \limsup_{n\to\infty} X_n(\omega) = 1$$

but $P[|X_{n,k} \geq \epsilon] \leq 2^{-n}$.
b) (Almost everywhere but not $\mathcal{L}^p$ convergence). Let $([0,1], \mathcal{A}, P)$ be the Lebesgue measure space. The random variables $X_n = 2^n 1_{[0,2^{-n}]}$ converge almost everywhere to $X = 0$ but not in $\mathcal{L}^p$.

Under more conditions, other implications can hold. We give two examples.

Proposition 1.7.2 *Given a sequence $X_n \in \mathcal{L}^\infty$ with $||X_n||_\infty \leq K$, then $X_n \to X$ in probability if and only if $X_n \to X$ in $\mathcal{L}^1$.*

Proof. (i) $P(|X| \leq K) = 1$. Proof. For $k \in \mathbf{N}$,

$$P(|X| > K + k^{-1}) \leq P(|X - X_n| > k^{-1}) \to 0, n \to \infty$$

so that $P(|X| > K + k^{-1}) = 0$. Therefore

$$P(|X| > K) = P(\bigcup_k \{|X| > K + k^{-1}\}) = 0 .$$

(ii) Given $\epsilon > 0$. Choose m such that for all $n > m$

$$P(|X_n - X| > \epsilon/3) < \epsilon/(3K) .$$

Then, using (i)

$$\begin{aligned} E[|X_n - X|) &= E[(|X_n - X|; |X_n - X| > \epsilon] + E[(|X_n - X|; |X_n - X| \leq \epsilon] \\ &\leq 2KP(|X_n - X| > \epsilon/3) + \epsilon/3 \leq \epsilon . \end{aligned}$$

$\square$

Remark. If the stochastic convergence in the above proposition would be replaced by almost sure convergence, this would be a reformulation of the bounded convergence theorem.

Recall that a family $\mathcal{C} \subset \mathcal{L}^1$ of random variables was called **uniformly integrable**, if

$$\lim_{R\to\infty} \sup_{X\in\mathcal{C}} E[1_{|X|>R}] \to 0$$

for all $X \in \mathcal{C}$. The next lemma was actually already proven in the proof of the IID-weak law. Nevertheless:

Lemma 1.7.3 *Given $X \in \mathcal{L}^1$ and $\epsilon > 0$. Then, there exists $K \geq 0$ with $E[|X|; |X| > K] < \epsilon$.*

Proof. Given $\epsilon > 0$. By the absolutely continuity property of $X \in \mathcal{L}^1$, we can find $\delta > 0$ such that if $P(A) < \delta$, then $E(|X|, A) < \epsilon$. Since $KP(|X| > K) \leq E(|X|)$, we can choose K such that $P(|X| > K) < \delta$. Therefore $E[|X|; |X| > K] < \epsilon$. □

The next proposition gives a neccessary and sufficient condition for $\mathcal{L}^1$ convergence.

Proposition 1.7.4 *Given a sequence random variables $X_n \in \mathcal{L}^1$. The following is equivalent:*
a) X_n converges in probability to X and $\{X_n\}_{n\in\mathbf{N}}$ is uniformly integrable.
b) X_n converges in $\mathcal{L}^1$ to X.

Proof. $a) \Rightarrow b)$. Define for $K \geq 0$ and a random variable X the bounded variable $X^{(K)} = X \cdot 1_{\{-K\leq X\leq K\}} + K \cdot 1_{\{X>K\}} - K \cdot 1_{\{X<-K\}}$. By the uniform integrability condition and the above lemma, we can choose K such that for all n,

$$E[|X_n^{(K)} - X_n|] < \epsilon/3, \; E[|X^{(K)} - X|] < \epsilon/3 \; .$$

Since $|X_n^{(K)} - X^{(K)}| \leq |X_n - X|$, we have $X_n^{(K)} \to X^{(K)}$ in probability. We have so by the last proposition $X_n^{(K)} \to X^{(K)}$ in $\mathcal{L}^1$ so that for $n > m$ $E[|X_n^{(K)} - X^{(K)}|] \leq \epsilon/3$. Therefore, for $n > m$ also

$$E[|X_n - X|] \leq E[|X_n - X_n^{(K)}|] + E[|X_n^{(K)} - X^{(K)}|] + E[|X^{(K)} - X|] \leq \epsilon \; .$$

$a) \Rightarrow b)$. We have seen already that $X_n \to X$ in probability if $||X_n - X||_1 \to 0$. We have to show that $X_n \to X$ in $\mathcal{L}^1$ implies that X_n is uniformly integrable.
Given $\epsilon > 0$. There exists m such that $E[|X_n - X|] < \epsilon/2$ for $n > m$. By the absolutely continuity property, we can choose $\delta > 0$ such that $P[A] < \epsilon$ implies

$$E[|X_n|; A] < \epsilon, 1 \leq n \leq m, E[|X|; A] < \epsilon/2 \; .$$

Since X_n is bounded in $\mathcal{L}^1$, we can choose K such that $K^{-1} \sup_n E[|X_n|] < \delta$ which implies $E[|X_n| > K] < \delta$. For $n \geq m$, we have therefore

$$E[|X_n|; |X_n| > K] \leq E[|X|; |X_n| > K] + E[|X - X_n|] < \epsilon \; .$$

□

1.8 The strong law of large numbers

The weak laws of large numbers made statements about stochastic convergence of sums $S_n/n = (X_1 + \ldots + X_n)/n$ of random variables X_n. The strong laws of large numbers make analogue statements about almost everywhere convergence.

Theorem 1.8.1 (Strong law for independent X_n, bounded in $\mathcal{L}^4$) *Assume X_n are independent random variables in $\mathcal{L}^4$ with common expectation $E[X_n] = m$ and satisfying $M = \sup_n ||X_n||_4^4 < \infty$. Then $S_n/n \to m$ almost everywhere.*

Proof. In the proof of the weak law of large numbers dealing with $\mathcal{L}^4$ random variables, we got

$$P[|S_n/n - m| \geq \epsilon] \leq 2M \frac{1}{\epsilon^4 n^2} .$$

This means that $S_n/n \to m$ fast in probability which implies convergence almost everywhere.
□

Theorem 1.8.2 (Strong law for pairwise independent $X_n \in \mathcal{L}^1$) *Assume $X_n \in \mathcal{L}^1$ are pairwise independent and identically distributed. Then $S_n/n \to E[X_1]$ almost everywhere.*

Proof. We can assume without loss of generality that $X_n \geq 0$ (since $X_n = X_n^+ + X_n^-$ with $X_n^+ = X_n \vee 0$ and knowing the result for $X_n^{\pm}$ implies the result for X_n.).
Define $f_R(t) = t \cdot 1_{[-R,R]}$, the random variables $X_n^{(R)} = f_R(X_n)$ and $Y_n = X_n^{(n)}$ as well as

$$S_n = \frac{1}{n}\sum_{i=1}^n X_i, T_n = \frac{1}{n}\sum_{i=1}^n Y_i .$$

(i) It is enough to show that $T_n - E[T_n] \to 0$.
Proof. Since $E[Y_n] \to E[X_1] = m$, we get $E[T_n] \to m$. Because

$$\begin{aligned}
\sum_{n\geq 1} P[Y_n \neq X_n] &\leq \sum_{n\geq 1} P[X_n \geq n] = \sum_{n\geq 1} P[X_1 \geq n] \\
&= \sum_{n\geq 1}\sum_{k\geq n} P[X_n \in [k, k+1]] \\
&= \sum_{k\geq 1} k \cdot P[X_1 \in [k, k+1]] \leq E[X_1] < \infty ,
\end{aligned}$$

we get by Borel-Cantelli that $P[Y_n \neq X_n, \text{ infinitely often}] = 1$. This means $T_n - S_n \to 0$ almost everywhere proving $E[S_n] \to m$.
(ii) Fix a real number $\alpha > 1$ and define the subsequence $k_n = [\alpha^n]$ the integer part of α^n. Denote with μ the law of the random variables X_n. For every $\epsilon > 0$, we get (using Chebychev, pairwise independece, $k_n = [\alpha^n]$)

$$\sum_{n=1}^{\infty} P[|T_{k_n} - E[T_{k_n}]| \geq \epsilon] \leq \sum_{n=1}^{\infty} \frac{\mathrm{Var}[T_{k_n}]}{\epsilon^2 k_n^2}$$

$$
\begin{aligned}
&= \frac{1}{\epsilon^2 k_n^2} \sum_{m=1}^{k_n} \mathrm{Var}[Y_m] = \frac{1}{\epsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}[Y_m] \sum_{n:k_n \geq m} \frac{1}{k_n^2} \\
&\leq^{(*)} \frac{1}{\epsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}[Y_m] \frac{\mathrm{const}}{m^2} \leq \mathrm{const} \sum_{m=1}^{\infty} \frac{1}{m^2} E[Y_m^2] \\
&= \mathrm{const} \sum_{m=1}^{\infty} \frac{1}{m^2} \sum_{l=0}^{m-1} \int_l^{l+1} x^2 \, d\mu(x) = \mathrm{const} \sum_{l=0}^{\infty} \sum_{m=l+1}^{\infty} \frac{1}{m^2} \int_l^{l+1} x^2 \, d\mu(x) \\
&\leq \mathrm{const} \sum_{l=0}^{\infty} \sum_{m=l+1}^{\infty} \frac{1}{m^2} (l+1) \int_l^{l+1} x \, d\mu(x) \leq^{(**)} \mathrm{const} \sum_{l=0}^{\infty} \int_l^{l+1} x \, d\mu(x) \\
&\leq \mathrm{const} \cdot E[X_1] < \infty \; .
\end{aligned}
$$

(*) We used that with $k_n = [\alpha^n]$ one has $\sum_{n:k_n \geq m} k_n^{-1} \leq \mathrm{const} \cdot m^{-2}$. (**) We used $\sum_{m=l+1}^{n} m^{-2} \leq \mathrm{const} \cdot (l+1)^{-1}$.

We have shown complete (=fast stochastic) convergence. This implies the a.e. convergence of $T_{k_n} - E[T_{k_n}] \to 0$.

(iii) The theorem is now shown only for a subsequenc k_n. Since we have assumed $X_n \geq 0$, the sequence $U_n = \sum_{i=1}^n Y_n = nT_n$ is monotone increasing. For $k \in [k_m, k_{m+1}]$, we get therefore

$$\frac{k_m}{k_{m+1}} \frac{U_{k_m}}{k_m} = \frac{U_{k_m}}{k_{m+1}} \leq \frac{U_n}{n} \leq \frac{U_{k_{m+1}}}{k_m} = \frac{k_{m+1}}{k_m} \frac{U_{k_{m+1}}}{k_{m+1}}$$

and from $\lim_{n \to \infty} T_n = E[X_1]$ a.e.,it follows that

$$\frac{1}{\alpha} E[X_1] \leq \liminf_n T_n \leq \limsup_n T_n \leq \alpha E[X_1] \; .$$

□

Remark. The strong law of large numbers can be interpreted as a statement about the growth of the sequence $\sum_{k=1}^n X_n$. For $E[X_1] = 0$, the convergence $\frac{1}{n} \sum_{k=1}^n X_n \to 0$ means that for all $\epsilon > 0$ there exists m such that for $n > m$

$$|\sum_{k=1}^{n} X_n| \leq \epsilon n \; .$$

This means that the trajectory $\sum_{k=1}^n X_n$ is finally contained in any arbitrary small cone which means that it grows slower than linear. The exact description for the growth of $\sum_{k=1}^n X_n$ is given by the **law of the iterated logarithm of Chintchine** which says that a sequence of IID random variables X_n with $E[X_n] = m$ and $\sigma(X_n) = \sigma \neq 0$ satisfies

$$\limsup_{n \to \infty} S_n / \Lambda_n = +1, \liminf_{n \to \infty} S_n / \Lambda_n = -1 \; ,$$

with $\Lambda_n = \sqrt{2\sigma^2 n \log \log n}$.

Historical remark. The strong law for IID random variables was first proven by Kolmogorov in 1930. Only 50 years later, (N.Etemadi, 1981) one has observed that one has only to assume the weaker pairwise independence.

Remark. Weakening of the assumption that the random variables have to be identically distributed can not be weakend without further restrictions. Take for example X_n satisfying $P[X_n = \pm 2^n] = 1/2$. We have $E[X_n] = 0$ but S_n/n does not converge.

1.9 Birkhoff's ergodic theorem

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

Definition. A measurable map $T : \Omega \to \Omega$ is *measure preserving*, if $P(T^{-1}(A)) = A$ for all $A \in \mathcal{A}$, *ergodic* if $T(A) = A$ implies $P(A) = 0$ or $P(A) = 1$. The map T is *invertible*, if there exists a measurable measure preserving inverse T^{-1}. A invertible measure preserving map T is called an **automorphism** of the probability space.

Remark. T is ergodic if and only if $X(T) = X$ implies that X is constant almost everywhere.

Given $X \in \mathcal{L}$, one has a sequence of random variables $X_n = X(T^n) \in \mathcal{L}$, where $X(T^n)\omega) = X(T^n\omega)$. Define $S_0 = 0$ and $S_n = \sum_{k=0}^{n} X(T^k)$.

Theorem 1.9.1 (Maximal ergodic theorem of Hopf)

Given $X \in \mathcal{L}^1$, then $E[X; A] \geq 0$ with $A = \{\sup_n S_n > 0\}$.

Proof. Define $Z_n = \max_{0 \leq k \leq n} S_k$ and the sets $A_n = \{Z_n > 0\} \subset A_{n+1}$. Then $A = \bigcup_n A_n$. Clearly $Z_n \in \mathcal{L}^1$. For $0 \leq k \leq n$ we have $Z_n \geq S_k$ and so $Z_n(T) \geq S_k(T)$ and hence

$$Z_n(T) + X \geq S_{k+1} \; .$$

By taking the maxima on both sides over $0 \leq k \leq n$ we get

$$Z_n(T) + X \geq \max_{1 \leq k \leq n} S_k$$

and if $Z_n(\omega) > 0$, then the right hand side is

$$Z_n = \max_{0 \leq k \leq n} S_k \; .$$

This means that on A_n

$$X \geq Z_n - Z_n(T) \; .$$

Integration over A_n gives

$$E[X; A_n] \geq E[Z_n; A_n] - E[Z_n(T); A_n] \; .$$

Using first $Z_n = 0$ on $X \setminus A_n$, and then $Z_n(T) \geq 0$, we get

$$E[X; A_n] \geq E[Z_n] - E[Z_n(T); A_n] = E[Z_n - Z_n(T)] = 0 \; .$$

□

Theorem 1.9.2 (Ergodic theorem of Birkhoff, 1931)

Given $X \in \mathcal{L}^1$, then $\overline{S}_n = S_n/n = \frac{1}{n}\sum_{i=0}^{n-1} X(T^i x)$ converges almost everywhere to a T-invariant random variable $\overline{X}$ satisfying $E[X] = E[\overline{X}]$.

Proof. Define $\overline{X} = \limsup_{n\to\infty} \overline{S}_n$, $\underline{X} = \liminf_{n\to\infty} \overline{S}_n$. We get $\overline{X} = \overline{X(T)}$ and $\underline{X} = \underline{X}(T)$ because

$$\frac{n+1}{n}\overline{S}_n - \overline{S}_n(T) = \frac{X}{n} \; .$$

(i) $\overline{X} = \underline{X}$.
Define for $\alpha < \beta \in \mathbf{R}$ the sets $A_{\alpha,\beta} = \{\underline{X} < \beta, \alpha < \overline{X}\}$. Because $\{\underline{X} < \overline{X}\} = \bigcup_{\alpha<\beta,\alpha,\beta\in\mathbf{Q}} A_{\alpha,\beta}$, it is enough to show that $P(A_{\alpha,\beta}) = 0$ for rational $\alpha < \beta$. Define

$$A = \{\sup_n (S_n - n\alpha) > 0\} = \{\sup_n (\overline{S}_n - \alpha) > 0\} .$$

Because $A_{\alpha,\beta} \subset A$ and $A_{\alpha,\beta}$ is T-invariant, we get from the maximal ergodic theorem $E[X - \alpha, A_{\alpha,\beta}] \geq 0$ and so

$$E[X, A_{\alpha,\beta}] \geq \alpha \cdot P(A_{\alpha,\beta}) .$$

Replacing X, α, β with $-X, -\beta, -\alpha$ and using $\overline{-X} = -\underline{X}$, $\underline{-X} = -\overline{X}$ gives $E[X; A_{\alpha,\beta}] \leq \beta \cdot P(A_{\alpha,\beta})$ and because $\beta < \alpha$, the claim $P(A_{\alpha,\beta}) = 0$ follows.

(ii) $\overline{X} \in \mathcal{L}^1$.
$|\overline{S}_n| \leq |X|$, and $\overline{S}_n$ converges pointwise to $\overline{X} = \underline{X}$ and $X \in \mathcal{L}^1$. Lebesgue's dominated convergence theorem gives $\overline{X} \in \mathcal{L}$.

(iii) $E[X] = E[\overline{X}]$.
Define the sets $B_{k,n} = \{\overline{X} \in [\frac{k}{n}, \frac{k+1}{n})\}$ for $k \in \mathbf{Z}, n \geq 1$. Define for $\epsilon > 0$, $Y = X - \frac{k}{n} + \epsilon$. Using the maximal ergodic theorem, we get $E[Y; B_{k,n}] \geq 0$. Because $\epsilon > 0$ was arbitrary,

$$E[X; B_{k,n}] \geq \frac{k}{n} .$$

With this inequality

$$E[\overline{X}, B_{k,n}] \leq \frac{k+1}{n} P(B_{k,n}) \leq \frac{1}{n} P(B_{k,n}) + E[X; B_{k,n}] .$$

Summing over k gives

$$E[\overline{X}] \leq \frac{1}{n} + E[X]$$

and because n was arbitrary, $E[\overline{X}] \leq E[X]$. Doing the same with $-X$ and using (i), we end with

$$E[-\overline{X}] = E[\underline{-X}] \leq E[\overline{-X}] \leq E[-X] .$$

□

Corollary 1.9.3 *The strong law of large numbers holds for IID random variables $X_n \in \mathcal{L}^1$.*

Proof. Given a sequence of IID random variables $X_n \in \mathcal{L}^1$. Let μ be the law of X_n. Define the probability space $\Omega = (\mathbf{R}^{\mathbf{Z}}, \mathcal{A}, P)$, where $P = \mu^{\mathbf{Z}}$ is the product measure. If $T : \Omega \to \Omega$, $T(\omega)_n = \omega_{n+1}$ denotes the shift on Ω, then $X_n = X(T^n)$ with with $X(\omega) = \omega_0$. Since every T-invariant function is constant almost everywhere, we must have $\overline{X} = E[X]$ a.e. so that $S_n/n \to E[X]$ almost everywhere. □

Remark. Evenso ergodic theory is very closely related to probability theory, the **notation** is different in almost all the literature. The reason is that the origin of the theories are different. One writes usually $(X, \mathcal{A}, m)$ for a probability space An example of different language is also that ergodic theorists do not use the word "random variables" X but speak of "functions" f.

1.10 Kolmogorov inequality, three series theorem, Levy's theorem

We mention now some results about the a.e. convergence of sums of random variables in contrast to the weak and strong laws which were dealing with averaged sums.

Theorem 1.10.1 (Kolmogorov's inequalities) *a) Assume $X_k \in \mathcal{L}^2$ are independent. Then*

$$P(\sup_{1 \le k \le n} |S_k - E[S_k]| \ge \epsilon) \le \frac{1}{\epsilon^2} \mathrm{Var}[S_n] .$$

b) Assume $X_k \in \mathcal{L}^\infty$ are independent and $||X_n||_\infty \le R$. Then

$$P(\sup_{1 \le k \le n} |S_k - E[S_k]| \ge \epsilon) \ge 1 - \frac{(R+\epsilon)^2}{\sum_{k=1}^n \mathrm{Var}[X_k]} .$$

Proof. We can assume $E[X_k] = 0$ without loss of generality.
a) For $1 \le k \le n$ we have

$$S_n^2 - S_k^2 = (S_n - S_k)^2 + 2(S_n - S_k)S_k = 2(S_n - S_k)S_k$$

and therefore $E[S_n^2; A_k] \ge E[S_k^2; A_k]$ for all $A_k \in \sigma(X_1, \dots, X_k)$ by the independence of $S_n - S_k$ and S_k. The sets $A_1 = \{|S_1| \ge \epsilon\}$, $A_{k+1} = \{|S_{k+1}| \ge \epsilon, \max_{i \le l \le k} |S_l| < \epsilon\}$ are mutually disjoint. We have to estimate the measure of

$$B_n = \{\max_{1 \le k \le n} |S_k| \ge \epsilon\} = \bigcup_{k=1}^n A_k .$$

We get

$$E[S_n^2] \ge E[S_n^2; B_n] = \sum_{k=1}^n E[S_n^2; A_k] \ge \sum_{k=1}^n E[S_k^2; A_k] \ge \epsilon^2 \sum_{k=1}^n P(A_k) = \epsilon^2 P(B_n) .$$

b)

$$E[S_k^2; B_n] = E[S_k^2] - E[S_k^2; B_n^c] \ge E[S_k^2] - \epsilon^2(1 - P(B_n)) .$$

On A_k, $|S_{k-1}| \le \epsilon$ and $|S_k| \le |S_{k-1}| + |X_k| \le \epsilon + R$ holds. Use that in the estimate

$$\begin{aligned}
E[S_n^2; B_n] &= \sum_{k=1}^n E[S_k^2 + (S_n - S_k)^2; A_k] \\
&= \sum_{k=1}^n E[S_k^2; A_k] + \sum_{k=1}^n E[(S_n - S_k)^2; A_k] \\
&\le (R+\epsilon)^2 \sum_{k=1}^n P(A_k) + \sum_{k=1}^n P(A_k) \sum_{j=k+1}^n \mathrm{Var}[X_j] \\
&\le P(B_n)((\epsilon + R)^2 + E[S_n^2])
\end{aligned}$$

so that

$$E[S_n^2] \le P(B_n)((\epsilon + R)^2 + E[S_n^2]) + \epsilon^2 - \epsilon^2 P(B_n) .$$

and so

$$P(B_n) \geq \frac{E[S_n^2] - \epsilon^2}{(\epsilon + R)^2 + E[S_n] - \epsilon^2} \geq 1 - \frac{(\epsilon + R)^2}{(\epsilon + R)^2 + E[S_n^2] - \epsilon^2} \geq 1 - \frac{(\epsilon + R)^2}{E[S_n^2]} .$$

□

Remark. The inequalities stay true in the limit $n \to \infty$. The first inequality is then

$$P(\sup_k |S_k - E[S_k]| \geq \epsilon) \leq \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} \mathrm{Var}[X_k] .$$

Of course, the statement in $a)$ is void, if the right hand side is infinite. In this case, however, the inequality in $b)$ states that $\sup_k |S_k - E[S_k]| \geq \epsilon$ almost surely for every $\epsilon > 0$.

Remark. For $n = 1$, Kolmogorov's inequality reduces to Chebychev's inequality.

Lemma 1.10.2 *A sequence X_n converges almost surely if and only if*

$$\lim_{n \to \infty} P(\sup_{k \geq 1} |X_{n+k} - X_n| > \epsilon) = 0$$

for all $\epsilon > 0$.

Proof. Exercice. □

Theorem 1.10.3 (Kolmogorov) *Assume $X_n \in \mathcal{L}^2$ are independent and $\sum_n \mathrm{Var}[X_n] < \infty$. Then $\sum_{n=1}^{\infty}(X_n - E[X_n])$ converges almost everywhere.*

Proof. Define $Y_n = X_n - E[X_n]$ and $S_n = \sum_{k=1}^n Y_k$. Given $m \in \mathbf{N}$. Apply Kolmogorov's inequality to the sequence Y_{m+k} to get

$$P(\sup_{n \geq m} |S_n - S_m| \geq \epsilon) \leq \frac{1}{\epsilon^2} \sum_{k=m+1}^{\infty} E[X_k^2] \to 0$$

for $m \to \infty$. The above lemma implies that $S_n(\omega)$ converges. □

The following theorem gives a necessary and sufficient condition that $S_n = \sum_{k=1}^n X_k$ converges for a sequence X_n of independent random variables. Given $R \in \mathbf{R}$ and a random variable X, we define the bounded random variable $X^{(R)} = 1_{|X|<R}X$.

Theorem 1.10.4 (Three series theorem) *Assume $X_n \in \mathcal{L}$ be independent. Then $\sum_{n=1}^\infty X_n$ converges almost everywhere if and only if for some (then for every) $R > 0$*

$$\sum_{k=1}^\infty P(|X_k| > R) \quad < \quad \infty \, , \tag{1.4}$$

$$\sum_{k=1}^\infty |E[X_k^{(R)}]| \quad < \quad \infty \, , \tag{1.5}$$

$$\sum_{k=1}^\infty \mathrm{Var}(X_k^{(R)}) \quad < \quad \infty \, . \tag{1.6}$$

Proof. "$\Rightarrow$" Assume first that the three series all converge. By (3) and Kolmogorov's theorem, we know that $\sum_{k=1}^\infty X_k^{(R)} - E[X_k^{(R)}]$ converges almost surely. Therefore, by (2), $\sum_{k=1}^\infty X_k^{(R)}$ converges almost surely. By (1) and Borel Cantelli, $P(X_k \neq X_k^{(R)}$ infinitely often$) = 0$. Since for almost all ω, $X_k^{(R)}(\omega) = X_k(\omega)$ for sufficiently large k and for almost all ω, $\sum_{k=1}^\infty X_k^{(R)}(\omega)$ converges, we get a set of measure one, where $\sum_{k=1}^\infty X_k$ converges.
"$\Leftarrow$" Assume now that $\sum_{n=1}^\infty X_n$ converges almost everywhere. Then $X_k \to 0$ almost everywhere and $P(|X_k| > R$, infinitely often$) = 0$ for every $R > 0$. By the second Borel-Cantelli lemma, the sum (1) converges.
The almost sure convergence of $\sum_{n=1}^\infty X_n$ implies the almost sure convergence of $\sum_{n=1}^\infty X_n^{(R)}$ since $P(|X_k| > R$, infinitely often$) = 0$.
Let $R > 0$ be fixed. Let Y_k be a sequence of independent random variables such that Y_k and $X_k^{(R)}$ have the same distribution and that all the random variables $X_k^{(R)}, Y_k$ are independent. The almost sure convergence of $\sum_{n=1}^\infty X_n^{(R)}$ implies that of $\sum_{n=1}^\infty X_n^{(R)} - Y_k$. Since $E[X_k^{(R)} - Y_k] = 0$ and $P(|X_k^{(R)} - Y_k| \leq 2R) = 1$, by Kolmogorov inequality b), the series $T_n = \sum_{k=1}^n X_k^{(R)} - Y_k$ satisfies for all $\epsilon > 0$

$$P(\sup_{k \geq 1} |T_{n+k} - T_n| > \epsilon) \geq 1 - \frac{(R+\epsilon)^2}{\sum_{k=n}^\infty \mathrm{Var}(X_k^{(R)} - Y_k)} \, .$$

Claim: $\sum_{k=1}^\infty \mathrm{Var}(X_k^{(R)} - Y_k) < \infty$.
Assume, the sum is infinite. Then the above inequality gives $P(\sup_{k\geq} |T_{n+k} - T_n| \geq \epsilon) = 1$. But this contradicts the almost sure convergence of $\sum_{k=1}^\infty X_k^{(R)} - Y_k$ because the latter implies by Kolmogorov inequality that $P(\sup_{k\geq 1} |S_{n+k} - S_n| > \epsilon) < 1/2$ for large enough n. Having shown that $\sum_{k=1}^\infty \mathrm{Var}(X_k^{(R)} - Y_k) < \infty$, we are done because then by Kolmogorov's theorem 1.10.3, the sum $\sum_{k=1}^\infty X_k^{(R)} - E[X_k^{(R)}]$ converges so that (2) holds.
□

Definition. $\alpha \in \mathbf{R}$ is called a **median** of $X \in \mathcal{L}$ if $P(X \leq \alpha) \geq 1/2$ and $P(X \geq \alpha) \geq 1/2$. Call $\mathrm{med}(X)$ the set of medians of X.

Note. The median is not unique and in general different from the mean. It is also defined for random variables for which the mean does not exist.

Remark. Given $Y \in \mathcal{L}$. Then every $\alpha \in \text{med}(Y)$ satisfies $|\alpha - E[Y]| \leq \sqrt{2\text{Var}[Y]}$. Proof. For every $\beta \in \mathbf{R}$, one has

$$\frac{|\alpha-\beta|^2}{2} \leq |\alpha-\beta|^2 \min(P(Y \geq \alpha), P(Y \leq \alpha)) \leq E[(Y-\beta)^2]$$

Put $\beta = E[Y]$.

Theorem 1.10.5 (Levy) *Given a sequence $X_n \in \mathcal{L}$ which is independent. Choose $\alpha_{l,k} \in \text{med}(S_l - S_k)$. Then, for all $n \in \mathbf{N}$ and all $\epsilon > 0$*

$$P(\max_{1 \leq k \leq n} |S_n + \alpha_{n,k}| \geq \epsilon) \leq 2P(S_n \geq \epsilon) .$$

Proof. Fix $n \in \mathbf{N}$ and $\epsilon > 0$. The sets

$$A_1 = \{S_1 + \alpha_{n,1} \geq \epsilon\}, A_{k+1} = \{\max_{1 \leq l \leq k}(S_n + \alpha_{n,l}) < \epsilon, S_{k+1} + \alpha_{n,k+1} \geq \epsilon\}$$

for $1 \leq k \leq n$ are disjoint and $\bigcup_{k=1}^n A_k = \{\max_{1 \leq k \leq n}(S_k + \alpha_{n,k}) \geq \epsilon\}$. Since $\{S_n \geq \epsilon\}$ contains A_k and $\{S_n - S_k \geq \alpha_{n,k}\}$ for $1 \leq k \leq n$, we have using the independence of $\sigma(A_k)$ and $\sigma(S_n - S_k)$

$$\begin{aligned} P(S_n \geq \epsilon) &\geq \sum_{k=1}^n P(A_k \cap \{S_n - S_k \geq \alpha_{n,k}\}) \\ &\geq \frac{1}{2}\sum_{k=1}^n P(A_k) = \frac{1}{2}P(\max_{1 \leq k \leq n}(S_n + \alpha_{n,k}) \geq \epsilon) . \end{aligned}$$

Applying this inequality to $-X_n$, we get also $P(-S_n - \alpha_{n,m} \geq -\epsilon\} \geq 2P(-S_n \geq -\epsilon)$ which gives

$$P(\max_{1 \leq k \leq n} |S_n + \alpha_{n,k}| \geq \epsilon) \leq 2P(S_n \geq \epsilon) .$$

□

Corollary 1.10.6 (Levy) *Given a sequence $X_n \in \mathcal{L}$ which is independent. If S_n converges in probability to S, then S_n converges almost everywhere to S.*

Proof. Exercice. □

Exercice 3

Topic: Levy's theorem, Strong law, Three series theorem

1) a) $P(\sup_{k \geq n} |X_k - X| > \epsilon) \to 0$ for $n \to \infty$ and all $\epsilon > 0$ if and only if $X_n \to X$ almost everywhere.
b) A sequence X_n converges almost surely if and only if

$$\lim_{n \to \infty} P(\sup_{k \geq 1} |X_{n+k} - X_n| > \epsilon) = 0$$

for all $\epsilon > 0$.

2) Prove the strong law of large numbers of independent but not necessarily identically distributed summands:

Given a sequence of independent random variables $X_n \in \mathcal{L}^2$ satisfying $E[X_n] = m$. If $\sum_{k=1}^{\infty} \mathrm{Var}[X_k]/k^2 < \infty$, then $S_n/n \to m$ almost everywhere.

3) Let X_n be an IID sequence of random variables with uniform distribution on $[0, 1]$. Prove that almost surely

$$\sum_{n=1}^{\infty} \prod_{i=1}^{n} X_i < \infty .$$

4) (Theorem of Levy:)

Given a sequence $X_n \in \mathcal{L}$ which is independent. If S_n converges in probability to S, then S_n converges almost everywhere to S.

Solutions

1) a) Define $A_k^{(\epsilon)} = \{|X_k - X| \geq \epsilon\}$. THen $A_\infty^{(\epsilon)} = \{|X_n - X| > 1/m, \text{ infinitely often}\}$.

$$P(A_\infty^{(\epsilon)}) = \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty} \{|X_k - X| > 1/m\}) = \lim_{n\to\infty} P(\{\sup_{k\geq n} |X_k - X| > 1/m\}) = 0$$

and so

$$P(X_n \not\to X) = P(\bigcup_{\epsilon>0} A_\infty^{(\epsilon)}) = P(\bigcup_m A_\infty^{(1/m)}) \leq \sum_m P(A_\infty^{(1/m)}) = 0 \; .$$

If $P(X_n \not\to X) \neq 0$, then $P(A_\infty^{(\epsilon)}) \geq \delta > 0$ for some $\epsilon > 0$ and therefore, there exists a subsequence such that $P(\sup_{k\geq n} |X_k - X| > \epsilon) \geq \delta > 0$.

b) Define $A_{n,k} = \{|X_{n+k} - X_n| > \epsilon\}$. Then $\bigcap_n \bigcup_{k=1}^{\infty} A_{n,k} = \{\sup_k |X_{n+k} - X - k| \geq \epsilon, i.o.\}$ has zero measure. so that for a set of measure one $\{\sup_k |X_{n+k} - X_n| \leq \epsilon$, large enough n There exists therefore a set of full measure, where X_n is a Cauchy sequence.

2) Proof of Strong Law: Use Kolmogorov's theorem for $Y_k = X_k/k$.

3) Use $\text{Var}(\prod_i X_i) = \prod E[X_i^2] - \prod E[X_i]^2$ and use the three series theorem.

4) Take $\alpha_{l,k} \in \text{med}(S_l - S_k)$. Since S_n converges in probability, there exists $m_1 \in \mathbf{N}$ such that $|\alpha_{l,k}| \leq \epsilon/2$ for all $m_1 \leq k \leq l$. In addition, there exists $m_2 \in \mathbf{N}$ such that $\sup_{n\geq 1} P(|S_{n+m} - S_m| \geq \epsilon/2) < \epsilon/2$ for all $m \geq m_2$. For $m = \max\{m_1, m_2\}$, we have for $n \geq 1$

$$P(\max_{1\leq l\leq n} |S_{l+m} - S_m| \geq \epsilon) \leq P(\max_{1\leq l\leq n} |S_{l+m} - S_m + \alpha_{n+m,l+m}| \geq \epsilon/2)$$

and the right hand side can be estimated by the above theorem applied to X_{n+m} by

$$\leq 2P(|S_{n+m} - S_m| \geq \frac{\epsilon}{2}) < \epsilon \; .$$

Apply the convergence lemma in the notes.

1.11 Distribution functions

Definition. Given a random variable X. The **distribution function** $F_X : \mathbf{R} \to [0,1]$ is defined as $F_X(x) = P[X \leq x]$. An equivalent definition which we have given already is as follows: $\mu_X = X^*P$ is the **law** of X, the from X "push-forward-measure" on $\mathbf{R}$, then $F_X(x) = \int_{-\infty}^{x} d\mu(x)$.

One reason to introduce distribution functions is that one can replace integrals on the probability space Ω by integrals on $\mathbf{R}$ which is more convenient.

Remark. The distribution function determines the law μ_X, since the measure $\nu((-\infty, a]) = F_X(a)$ on the π-system $\mathcal{I}$ given by the intervals $\{(-\infty, a]\}$ determines a unique measure on $\mathbf{R}$. Of course, the distribution does not determine the random variable.

> **Proposition 1.11.1 (Properties of distribution functions and existence)**
> *The distribution function F_X of a random variable is a) nondecreasing, b) $F_X(-\infty) = 0, F_X(\infty) = 1$ c) continuous from the right: $F_X(x+h) = F_X$. Given a function F with the properties $a), b), c)$, there exists a random variable X on a probability space $(\Omega, \mathcal{A}, P)$ satisfying $F_X = F$.*

Proof. a) follows from $\{X \leq x\} \subset \{X \leq y\}$ for $x \leq y$. b) $P[\{X \leq -n\}] \to 0$ and $P[\{X \leq n\}] \to 1$. c) $F_X(x+h) - F_X = P[x < X \leq x+h] \to 0$ for $h \to 0$.
Given F, define $\Omega = \mathbf{R}$ and $\mathcal{A}$ as the Borel σ-algebra on $\mathbf{R}$. The measure $P[(-\infty, a]] = F[a]$ on the π-system $\mathcal{I}$ defines a unique measure on $(\Omega, \mathcal{A})$. □

Remarks. 1) This proposition implies that every Borel probability measure μ on $\mathbf{R}$ determines a distribution function F_X of some random variable.
2) The proposition tells also that we can speak of **distribution functions**, real functions which satisfy $a), b), c)$ in the proposition.

Definition. Denote with $C_b(\mathbf{R})$ the vector space of bounded continuous functions on $\mathbf{R}$. A sequence of Borel probability measures μ_n on $\mathbf{R}$ **converges weakly** to a probability measure μ on $\mathbf{R}$ if for every $f \in C_b(\mathbf{R})$

$$\int_{\mathbf{R}} f \, d\mu_n \to \int_X f \, d\mu$$

holds.

Remark. For weak convergence, it is enough to test $\int_{\mathbf{R}} f \, d\mu_n \to \int_X f \, d\mu$ for a dense set (for example all polynomials) in $C_b(\mathbf{R})$.

Definition. A sequence of random variables X_n converges **weakly** or **in law** to a random variable X, if their laws μ_{X_n} converge weakly to the law of μ_X.

Definition. Given a distribution function F, we denote with $\text{Cont}(F)$ the set of continuity points of F.

Lemma 1.11.2 (Characterisation of weak convergence) *If a sequence X_n of random variables converges weakly to X then $F_{X_n}(x) \to F_X(x)$ pointwise for all $x \in \mathrm{Cont}(F)$.*

Proof. Given $x \in \mathrm{Cont}(f)$ and $\delta > 0$. Define a continuous function $1_{(-\infty,x]} \leq f \leq 1_{(-\infty,x+\delta]}$. Then

$$F_n(x) = \int_{\mathbf{R}} 1_{(-\infty,x]} \; d\mu_n \leq \int_{\mathbf{R}} f(x) \; d\mu_n(x) \leq \int_{\mathbf{R}} 1_{(-\infty,x+\delta]} \; d\mu_n = F_n(x+\delta) \; .$$

This gives

$$\limsup_{n\to\infty} F_n(x) \leq \lim_{n\to\infty} \int f \; d\mu_n = \int f \; d\mu \leq F(x+\delta) \; .$$

Similarily, we obtain

$$\liminf_{n\to\infty} F_n(x) \geq \lim_{n\to\infty} \int f \; d\mu_n = \int f \; d\mu \geq F(x-\delta) \; .$$

We get for $\delta \to 0$ since F is continuous at x

$$F(x) = \lim_{\delta\to 0} F(x-\delta) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x) \; .$$

□

Remark. The above lemma can be reversed: pointwise convergence of F_n to F on the continuity points is equivalent with weak convergence.

Definition. A distribution function F is called **absolutely continuous** (ac) if there exists a Borel measurable function f satisfying $F(x) = \int_{-\infty}^{x} f(x) \; dx$. The distribution function is called **discrete** (pp) if there exists a countable sequence of real numbers x_n and a sequence of positive numbers $p_n, \sum_n p_n = 1$ such that $F(x) = \sum_{n, x_n \leq x} p_n$. The distribution function F is called **singular continuous** (sc) if F is continuous and if there exists a Borel set S of zero Lebesgue measure such that $\mu_F(S) = 1$.

Remark. The definition of (ac),(pp) and (sc) distribution functions is compatible for the definition of (ac),(pp) and (sc) Borel measures on $\mathbf{R}$. A Borel measure is (pp), if $\mu(A) = \sum_{x \in A} \mu(\{a\})$. It is continuous, if it contains no atoms, points with positive measure. It is (ac), if there exists a measurable function f such that $\mu = f \; dx$. It is (sc), if it is continuous and if $\mu(S) = 1$ for some Borelset S of zero Lebesgue measure. That these classes of measures is natural shows the following theorem

Theorem 1.11.3 (Lebesgue decomposition theorem) *Every Borel measure μ can be decomposed uniquely as $\mu = \mu_{pp} + \mu_{ac} + \mu_{sc}$, where μ_{pp} is (pp), μ_{sc} is (sc) and μ_{ac} is (ac).*

Proof. Denote with λ the Lebesgue measure on $(\mathbf{R}, \mathcal{B})$. We first show that $\mu = \mu_{ac} + \mu_s$, where μ_{ac} is absolutely continuous with respect to λ and μ_s is singular. The decomposition is unique: $\mu = \mu_{ac}^{(1)} + \mu_s^{(2)} = \mu_{ac}^{(2)} + \mu_s^{(2)}$ implies that $\mu_{ac}^{(1)} - \mu_{ac}^{(2)} = \mu_s^{(2)} - \mu_s^{(2)}$ is both absolutely continuous and singular continuous with respect to μ which is only possible, if they are zero.

To get the existence of the decomposition, define $c = \sup_{A \in \mathcal{A}, \lambda(A)=0} \mu(A)$. If $c = 0$, then μ is absolutely continuous and we are done. If $c > 0$, take an increasing sequence $A_n \in \mathcal{B}$ with $\mu(A_n) \to c$. Define $A = \bigcup_{n \geq 1} A_n$ and μ_{ac} as $\mu_{ac}(B) = \mu(A \cap B)$.
To split the singular part μ_s into a singular continuous and pure point part, we have again uniqueness since $\mu_s = \mu_{sc}^{(1)} + \mu_{sc}^{(2)} = \mu_{pp}^{(2)} + \mu_{pp}^{(2)}$ implies that $\nu = \mu_{sc}^{(1)} - \mu_{sc}^{(2)} = \mu_{pp}^{(2)} - \mu_{pp}^{(2)}$ are both singular continuous and pure point which implies that $\nu = 0$.
To get existence define the finite or countable set $A = \{\omega \mid \mu(\omega) > 0\}$ and $\mu_{pp}(B) = \mu(A \cap B)$.
□

Examples of absolutely continuous distributions:

ac1) The **normal distribution** $N(m, \sigma^2)$ is given by the density function

$$f(x) = (2\pi\sigma^2)^{-1/2} e^{-\frac{(x-m)^2}{2\sigma^2}} .$$

ac2) The **Cauchy distribution** is given by the density function

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} .$$

ac3) The **uniform distribution** has the density function

$$f(x) = \frac{1}{b-a} 1_{[a,b]} .$$

ac4) The **exponential distribution** $\lambda > 0$ has the density function

$$f(x) = 1_{[0,\infty)}(x) \lambda e^{-\lambda x} .$$

Examples of discrete distributions:

pp1) The **Binomial distribution**

$$P[X = k] = \binom{n}{k} p^k (1 - p^{n-k}) .$$

pp2) The **Poisson distribution**

$$P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!} .$$

pp3) The **Discrete uniform distribution**

$$P[X = k] = 1_{\{1,\dots n\}}(k) 1/n .$$

pp4) The **geometric distribution**

$$P[X = k] = (1-p)^{k-1} p .$$

An example of a singular continuous distribution:

sc1) The **Cantor distribution** $C = \bigcap_n E_n$ Cantor set.

$$F(x) = \lim_{n \to \infty} F_n(x)$$

where $F_n(x)$ has the density $(3/2)^n \cdot 1_{E_n}$.

Lemma 1.11.4 *Given $X \in \mathcal{L}$ with law μ. For any measureable map $h : \mathbf{R}^1 \to [0, \infty)$, or any map h satisfying $h(X) \in \mathcal{L}^1$, one has $E[h(X)] = \int_{\mathbf{R}} h(x)\, d\mu(x)$. Especially, if $\mu = \mu_{ac} = f dx$ then*

$$E[h(X)] = \int_{\mathbf{R}} h(x) f(x)\, dx$$

If $\mu = \mu_{pp}$, then

$$E[h(X)] = \sum_{x, \mu(\{x\}) \neq 0} h(x) \mu(\{x\}) \; .$$

Proof. If h is nonnegative, show it first for $X = c1_{x \in A}$, then for step functions $X \in \mathcal{S}$ and then by the monotone convergence theorem for $X \in \mathcal{L}$.
If $h(X)$ is integrable, then $E[h(X)] = E[h^+(X)] - E[h^-(X)]$. □

Distribution	Parameters	Mean	Variance
ac1) Normal	$m \in \mathbf{R}$, $\sigma^2 > 0$	m	σ^2
ac2) Cauchy	-	"0"	∞
ac3) Uniform	$a < b$	$(a+b)/2$	$(b-a)^2/12$
ac4) Exponential	$\lambda > 0$	$1/\lambda$	$1/\lambda^2$
pp1) Binomial	$n \in \mathbf{N}$, $p \in [0,1]$	np	$np(1-p)$
pp2) Poisson	$\lambda > 0$	λ	λ
pp3) Uniform	$n \in \mathbf{N}$	$(1+n)/2$	$(4n)^{-1} \sum_{k=1}^{n} (2k-1+n)^2$
pp4) Geometric	$p \in (0,1)$	$1/p$	$1/p^2$
sc1) Cantor	-	0	$\int_{\mathbf{R}} x^2 \, d\mu(x)$

Proof. Some hints:
Exponential distribution:

$$E[X^p] = \int_0^\infty x^p \lambda e^{-\lambda x} \, dx = \frac{p}{\lambda} E[X^{p-1}] = p!/\alpha^p \; .$$

Poisson distribution:

$$E[X] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \; .$$

For calculating higher moments, it is useful to compute the the function

$$E[z^X] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda z)^k}{k!} = e^{-\lambda(1-z)}$$

and then to differentiate this identity with respect to z at the place $z = 0$. We get then

$$E[X] = \lambda, E[X(X-1)] = \lambda^2, E[X^3] = E[X(X-1)(X-2)], \ldots$$

so that $E[X^2] = \lambda + \lambda^2$ and $\mathrm{Var}[X] = \lambda$.

Geometric distribution.

$$E[X_p] = \sum_{k=1}^{\infty} k(1-p)^{(k-1)} p = \frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^k p + \sum_{k=1}^{\infty} (1-p)^{(k-1)} = 0 + 1/p \; .$$

For calculating the higher moments proceed as in the Poisson case. □

We see from the computation that calculating the moments can sometimes be done in an elegant way. Useful functions to compute are **characteristic functions** $f_X(z) = E[e^{izX}]$ to which we come later or **moment generating functions** $M(z) = E[e^{zX}]$. If one can compute the later, then one gets the moments immediately

$$E[X^n] = \int_{\mathbf{R}} x^n \, d\mu = \frac{d^n M}{dz^n}(0) \; .$$

For example, for the Normal distribution, one computes $M(z) = e^{zm + \sigma^2 z^2/2}$. The Taylor expansion leads then to the moments.

Remark. An interesting quantity for continuous distributions is the **entropy**

$$H(\mu) = -\int_X \log(f(x)) \; d\mu(x) \; .$$

It allows to distinguish several distributions from others by asking for the one with the highest entropy. For example, amoung all distribution functions on $[0, \infty)$ with fixed expectation $m = 1/\alpha$, the Poisson distribution with parameter α is the one with maximal entropy. Maybe we return to these interesting questions later.

An important fact in the proof of the central limit theorem: A sequence of random variables X_n converges in distribution to X if and only if $\mathrm{E}[f(X_n)] \to \mathrm{E}[f(X)]$ for all smooth f.
(see Karr p. 138-139). This implies that the weak convergence is holding if and only if the characteristic functions converge pointwise (see Karr p. 171).

Lemma 1.11.5 (Characterisation of weak convergence) *If a sequence X_n of random variables converges weakly to X then $F_{X_n}(x) \to F_X(x)$ pointwise for all $x \in$ Cont(F).*

Proof. Given $x \in \mathrm{Cont}(f)$ and $\delta > 0$. Define a continuous function $1_{(-\infty,x]} \leq f \leq 1_{(-\infty,x+\delta]}$. Then

$$F_n(x) = \int_{\mathbf{R}} 1_{(-\infty,x]} \; d\mu_n \leq \int_{\mathbf{R}} f(x) \; d\mu_n(x) \leq \int_{\mathbf{R}} 1_{(-\infty,x+\delta]} \; d\mu_n = F_n(x+\delta) \; .$$

This gives

$$\limsup_{n \to \infty} F_n(x) \leq \lim_{n \to \infty} \int f \; d\mu_n = \int f \; d\mu \leq F(x+\delta) \; .$$

Similarily, we obtain with a function $1_{(-\infty,x-\delta]} \leq f \leq 1_{(-\infty,x]}$.

$$\liminf_{n \to \infty} F_n(x) \geq \lim_{n \to \infty} \int f \; d\mu_n = \int f \; d\mu \geq F(x-\delta) \; .$$

We get for $\delta \to 0$ since F is continuous at x

$$F(x) = \lim_{\delta \to 0} F(x-\delta) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x) \; .$$

□

The proof shows that it is enough to assume $\int f\, d\mu_n \to \int f\, d\mu$ for smooth f.

1.12 The central limit theorem

Notation. For any random variable X with nonvanishing variance, we denote with

$$X^* = (X - E[X])/\sigma(X)$$

the **normalized random variable**, which has mean $E[X^*] = 0$ and variance $\sigma(X^*) = \sqrt{\text{Var}[X^*]} = 1$.

Given a sequence of random variables X_n, we define $S_n = \sum_{i=1}^n X_i$.

Theorem 1.12.1 (Central limit theorem for independent $X_i \in \mathcal{L}^3$) *Assume $X_i \in \mathcal{L}^3$ are independent and satisfy*

$$\sup_i ||X_i||_3 < \infty, \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \text{Var}[X_i] > 0 .$$

Then S_n^* converges in distribution to a random variable with standard normal distribution $N(0,1)$:

$$\lim_{n\to\infty} P[S_n^* \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2}\, dy, \quad \forall x \in \mathbf{R} .$$

Lemma 1.12.2 *A $N(0,\sigma^2)$ distributed random variable X satisfies $E[|X|^p] = \frac{1}{\sqrt{\pi}} 2^{p/2}\sigma^p\Gamma(\frac{1}{2}(p+1))$. Especially $E[|X|^3] = \sqrt{\frac{8}{\pi}}\sigma^3$.*

Proof. With the density function $f(x) = (2\pi\sigma^2)^{-1/2}e^{-\frac{x^2}{2\sigma^2}}$ we have $E[|X|^p] = 2\int_0^\infty x^p f(x)\, dx$ which is after a substitution $z = x^2/(2\sigma^2)$ equal to

$$\frac{1}{\sqrt{\pi}} 2^{p/2}\sigma^p \int_0^\infty z^{\frac{1}{2}(p+1)-1} e^{-z}\, dz .$$

The integral to the right is by definition equal to $\Gamma(\frac{1}{2}(p+1))$. □

After this preliminary computation to the proof of the central limit theorem.

Proof. Define for fixed $n \geq 1$ the random variables

$$Y_i = (X_i - E[X_i])/\sigma(S_n),\ 1 \leq i \leq n$$

so that $S_n^* = \sum_{i=1}^n Y_i$. Define $N(0, \sigma^2 = \text{Var}[Y_i])$-distributed random variables $\tilde{Y}_i$ having the property that $\{Y_1, \dots, Y_n, \tilde{Y}_1, \dots \tilde{Y}_n\}$ are independent. The distribution of $\tilde{S}_n = \sum_{i=1}^n \tilde{Y}_i$ is just the normal distribution $N(0,1)$. In order to show the theorem, we have to prove $E[f(S_n^*)] - E[f(\tilde{S}_n)] \to 0$ for any $f \in C_b(\mathbf{R})$ and it is enough to see it for smooth f. Define

$$Z_k = \tilde{Y}_1 + \dots \tilde{Y}_{k-1} + Y_{k+1} + \dots + Y_n .$$

Note that $Z_1 + Y_1 = S_n^*$ and $Z_n + \tilde{Y}_n = \tilde{S}_n$. Using a telescopic sum and Taylor's theorem, we can write

$$\begin{aligned} f(S_n^*) - f(\tilde{S}_n) &= \sum_{k=1}^n [f(Z_k + Y_k) - f(Z_k + \tilde{Y}_k)] \\ &= \sum_{k=1}^n [f'(Z_k)(Y_k - \tilde{Y}_k) + \frac{1}{2} f''(Z_k)(Y_k^2 - \tilde{Y}_k^2) \\ &\quad + R(Z_k, Y_k) + R(Z_k, \tilde{Y}_k)] \end{aligned}$$

with a Taylor rest term $R(Z, Y)$ depending on f. We get therefore

$$|E[f(S_n^*)] - E[f(\tilde{S}_n)]| \leq \sum_{k=1}^n E[|R(Z_k, Y_k)|] + E[|R(Z_k, \tilde{Y}_k)|] \; . \tag{1.7}$$

Since $\tilde{Y}_k$ are $N(0, \sigma^2)$-distributed we get using the computation of the moments of normal random variables and Jensen

$$E[|\tilde{Y}_k|^3] = \sqrt{\frac{8}{\pi}} \sigma^3 = \sqrt{\frac{8}{\pi}} E[|Y_k|^2]^{3/2} \leq \sqrt{\frac{8}{\pi}} E[|Y_k|^3] \; .$$

Taylor gives $|R(Z_k, Y_k)| \leq \text{const} \cdot |Y_k|^3$ so that

$$\begin{aligned} \sum_{k=1}^n E[|R(Z_k, Y_k)|] + E[|R(Z_k, \tilde{Y}_k)|] &\leq \text{const} \cdot \sum_{k=1}^n E[|Y_k|^3] \\ &\leq \text{const} \cdot n \cdot \sup_i ||X_i||_3 / \text{Var}[S_n]^{3/2} \\ &= \text{const} \cdot \frac{\sup_i ||X_i||_3}{(\text{Var}[S_n]/n)^{3/2}} \cdot \frac{1}{\sqrt{n}} \to 0 \; . \end{aligned}$$

□

Theorem 1.12.3 (Central limit theorem for IID $X_i \in \mathcal{L}^2$)
Assume $X_i \in \mathcal{L}^2$ are IID and satisfy $0 < \text{Var}[X_i]$, then S_n^* converges weakly to a random variable with standard normal distribution $N(0, 1)$.

Proof. The same proof gives Equation 1.7. We change the estimation of Taylor $|R(z, y)| \leq \delta(y) \cdot y^2$ with $\delta(y) \to 0$ for $|y| \to 0$. Using IID and dominated convergence

$$\begin{aligned} \sum_{k=1}^n E[|R(Z_k, Y_k)|] + E[|R(Z_k, \tilde{Y}_k)|] &\leq \sum_{k=1}^n E[\delta(Y_k) Y_k^2] + E[\delta(\tilde{Y}_k) \tilde{Y}_k^2] \\ &= n \cdot E[\delta(\frac{X_1}{\sigma \sqrt{n}}) \frac{X_1}{\sigma \sqrt{n}}] + n \cdot E[\delta(\frac{\tilde{X}_1^2}{\sigma^2 n}) \frac{\tilde{X}_1^2}{\sigma^2 n}] \to 0 \; . \end{aligned}$$

□

Remark. The central limit theorem can be interpreted as a solution to a certain fixed point problem.

Corollary 1.12.4 *Let $\mathcal{P}_{0,1}$ be the space of probability measure μ on $(\mathbf{R}, \mathcal{B}_\mathbf{R})$ which have the properties that $\int_\mathbf{R} x^2 \, d\mu(x) = 1, \int_\mathbf{R} x \, d\mu(x) = 0$. The map*

$$T\mu(A) = \int_\mathbf{R} \int_\mathbf{R} 1_A(\frac{x+y}{\sqrt{2}}) \mu(dx) \; \mu(dy)$$

maps $\mathcal{P}_{0,1}$ into itself. The only attracting fixed point is γ, the normal law.

Proof. If μ is the law of a random variable X with $\mathrm{Var}[X] = 1$ and $E[X] = 0$. Then $T(\mu)$ is the law of the normalized random variable $(X + X)/\sqrt{2}$ because the independent random variables X, Y can be realized on the probability space $(\mathbf{R}^2, \mathcal{B}, \mu \times \mu)$ as coordinate functions $X((x,y)) = x, Y((x,y)) = y$. Then $T(\mu)$ is obviously the law of $(X+Y)/\sqrt{2}$. Now note that $T^n(X) = (S_{2^n})^*$ which converges in distribution to $N(0,1)$. □

Remark. Note that the map $S(X) = (X + X)/2$ gives by the laws of large numbers also a fixed point in $\mathcal{P}$, the space of probability measure. The limit is a point mass located at the mean of X. Speaking dynamically, we can say that the Dirac masses on one point (which are all fixed points of S) attract the rest of $\mathcal{P}$. The remarkable thing is that if we scale differently, then the fixed point becomes unique and is independent of the initial conditions.

Remark. The global dynamics of the map T is completely understood since everything is attracted to a fixed point. There are other (so called renormalisation maps) on laws of independent random variables, which are believed to have a more interesting dynamics. An example is if we replace the arithmetic mean by the arithmetic mean of two harmonic means. We have to replace then $S(X) = (X + X)/2$ by

$$S(X) = (X^{-1} + X^{-1})^{-1} + (X^{-1} + X^{-1})^{-1} \; .$$

This is a hierarchichal network model. Analogue to the strong law of large numbers, it has been shown by J.Wehr (with Martingale methods) that $S^n(X)$ converges almost surely to a constant. There is strong evidence that with a suitable constant λ (playing the role of $\sqrt{2}$) the map T has a nontrivial fixed point, which is responsible for certain numerically observed universal behavior analogue to the Feigenbaum story in dynamical systems.

Historical remark. For independent $0-1$ experiments with win probability $p \in (0,1)$, the central limit theorem is quite old. One gets there

$$\lim_{n \to \infty} P(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-y^2/2} \, dy \; .$$

This has been shown by de Moivre in 1730 in the case $p = 1/2$ and for general $p \in (0,1)$ by Laplace in 1812. This result is then also called the **DeMoivre-Laplace global limit thereom**. Of course, it is a special case of the central limit theorem. There are known now many other (more general) versions of the central limit theorem which we do not treat here.

We mention now a limit theorem on $\mathbf{N}$ which shows that the Poisson distribution on $\mathbf{N}$ is natural. Denote with $B(n,p)$ the Binomial distribution on $\{1, \dots, n\}$ and with $P(\alpha)$ the Poisson distribution on $\mathbf{N} \setminus \{0\}$.

Theorem 1.12.5 (Poisson limit theorem) *Let X_n be a $B(n, p_n)$ distributed and suppose $np_n \to \alpha$. Then X_n converges in distribution to $P(\alpha)$.*

Proof. We have to show that $P(X_n = k) \to P(\alpha)$ for each fixed $k \in \mathbf{N}$.

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k} \\ &\sim \frac{1}{k!}(np_n)^k (1 - \frac{np_n}{n})^{n-k} \to e^{-\alpha}\frac{\alpha^k}{k!} \ . \end{aligned}$$

□

1.13 Appendix: An experiment to the central limit theorem

```
X:=Random[]-1/2; Y[n_]:=Sum[X,{n}]/n;
Experiment=Sort[Table[Y[10],{10000}]];
Place[J_,x_]:=Module[{}, u=Sort[Join[J,{x}]]; Position[u,x][[1,1]]];
Density[n_]:=Table[Place[Experiment,N[k/(n)]]-Place[Experiment,N[(k-1)/(n)]],{k,-n/3,n/3}];
ListPlot[Density[500]]
```

Exercice 4

Topic: Distribution functions, Central limit theorem

1) Compute the mean and variance of the **Erlang distribution**

$$f(x) = 1_{0,\infty)} \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda x}$$

on $[0, \infty)$ with the help of the moment generating function.

2) Denote with $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy$ the distribution function of the standard normal distribution $N(0,1)$. Given a sequence of IID random variables X_n.
a) Justify that one can estimate for large n probabilities

$$P(a \le S_n^* \le b) \sim \Phi(b) - \Phi(a) \ .$$

b) Assume X_i are all uniformly distributed random variables in $[0,1]$.
Estimate for large n

$$P(|S_n/n - 0.5| \ge \epsilon)$$

in terms of Φ, ϵ and n.
c) Compaire the result in b) with the estimate obtained in the weak law of large numbers.

3) We define the transformation

$$T_\lambda(\mu)(A) = \int_{\mathbf{R}} \int_{\mathbf{R}} 1_A(\frac{x+y}{\lambda}) \ d\mu(x) \ d\mu(y)$$

in $\mathcal{P}$, the space of all Borel probability measures on $\mathbf{R}$.
For which λ can you describe the limit?

1.14 Entropy of distributions

Denote with ν a (not necessarily finite) measure on a measure space $(\Omega, \mathcal{A})$. An example is the Lebesgue measure on $\mathbf{R}$ or the counting measure on $\mathbf{N}$. Note that the measure is defined only on a δ-subring of $\mathcal{A}$ since we did not assume that ν is finite.

Definition. A probability measure μ on $\mathbf{R}$ is called ν **absolutely continuous**, if there exists a density $f \in \mathcal{L}^1(\nu)$ such that $\mu = f\nu$. One writes also $\mu << \nu$. Call $\mathcal{P}(\nu)$ the set of all ν absolutely continuous measures. Clearly $\mathcal{P}(\nu)$ is corresponging bijectively with densities in $\mathcal{L}^1(\nu)$.

Definition. For any probability measure $\mu \in \mathcal{P}(\nu)$ is defined the **Boltzmann-Gibbs entropy** (or simply entropy)

$$H(\mu) = \int_\Omega -f(\omega)\log(f(\omega))\, d\nu(\omega) \quad \in [0, \infty] .$$

Examples.
1) Let ν be the counting measure on a countable set Ω. ($\mathcal{A}$ is the σ-algebra of all subsets of Ω and the measure is defined on the δ-ring of all finite subsets.) In this case,

$$H(\mu) = \sum_{\omega \in \Omega} -f(\omega)\log(f(\omega)) .$$

2) Let ν be the Lebesgue measure on $\mathbf{R}$. For a density function f, we have

$$H(\mu) = \int_{\mathbf{R}} -f(x)\log(f(x))\, dx .$$

3) If ν is a probability measure and f a density. If we take the step function $\tilde{f} = \sum_{i=1}^n (\int_{A_i} f d\nu) 1_{A_i} \in \mathcal{S}(\nu)$, then $H(\tilde{f}\nu)$ is the number

$$H(\{A_i\}) = \sum_i -\nu(A_i)\log(\nu(A_i))$$

which is called the **entropy of the partition** $\{A_i\}$ used in ergodic theory. This approximation of the density by a step function is called **coarse graining** and the entropy of $\tilde{f}$ is called the **coarse grained entropy**. It has been considered first by Gibbs (1902).
Side remark. In **ergodic theory**, where one studies measure preserving transformations of probability spaces, one is interested in the growth rate of the entropy of a partition and leads to the **entropy of a measure preserving transformation**.

Interpretation. Assume for illustration that Ω is finite and that ν the counting measure and $\mu(\{\omega\}) = f(\omega)$ the probability distribution of random variable (describing a measurement of an experiment). If the event $\{\omega\}$ happens, then $-\log(f(\omega))$ is a measure for the **information or "surprise"** that the event happens. The averaged information or surprise is $H(\mu) = \sum_\omega -f(\omega)\log(f(\omega))$. If f is taking only values 0 or 1, which means that μ is **deterministic**, then $H(\mu) = 0$. There is no surprise and what the measurements show, is the reality. On the other hand, if f is constant on Ω (uniform distribution), then $H(\mu) = \log(|\Omega|)$. We will see in a moment that this is maximal entropy (maximal surprise).

Definition. Given two probability measures $\mu = f\nu$ $\tilde{\mu} = \tilde{f}\nu$ which are both absolutely continuous with respect to ν. Define the **relative entropy**

$$H(\tilde{\mu}|\mu) = \int \tilde{f}(\omega) \log(\frac{\tilde{f}(\omega)}{f(\omega)}) \, d\nu(x) \quad \in [0, \infty] \ .$$

It is the expectation $E_{\tilde{\mu}}[l]$ of the **Likelyhood coefficient** $l = \log(\frac{\tilde{f}(x)}{f(x)})$. The negative relative entropy $-H(\tilde{\mu}|\mu)$ is also called the **conditional entropy**. One writes also $H(f|\tilde{f})$ instead of $H(\tilde{\mu}|\mu)$.

Lemma 1.14.1 (Gibbs inequality) *$0 \leq H(\tilde{\mu}|\mu) \leq +\infty$ and $H(\tilde{\mu}|\mu) = 0$ if and only if $\mu = \tilde{\mu}$.*

Proof. Assume $H(\tilde{\mu}|\mu) < \infty$. The function $u(x) = x\log(x)$ is convex and satisfies $u(x) \geq x - 1$.

$$H(\tilde{\mu}|\mu) = \int_\Omega f(\omega) u(\frac{\tilde{f}(x)}{f(x)}) \, d\nu \geq \int_\Omega f(\omega)(\frac{\tilde{f}(x)}{f(x)} - 1) \, d\nu = 0 \ .$$

If $\mu = \tilde{\mu}$, then $f = \tilde{f}$ almost everywhere and $H(\tilde{\mu}|\mu) = 0$.
On the other hand, if $H(\tilde{\mu}|\mu) = 0$, then by Jensen's inequality

$$0 = E_\mu[u(\frac{\tilde{f}}{f})] \geq u(E_\mu[\frac{\tilde{f}}{f}]) = u(1) = 0 \ .$$

Therefore $E_\mu[u(\frac{\tilde{f}}{f})] = u(E_\mu[\frac{\tilde{f}}{f}])$. The strict convexity of u implies that $\frac{\tilde{f}}{f}$ must be a constant. Since both f and $\tilde{f}$ are densities $f = \tilde{f}$. □

Theorem 1.14.2 (Distributions with maximal entropy)

a) Ω finite with counting measure ν. The **uniform distribution** *on Ω has maximal entropy among all other distributions. It is unique with this property.*

b) $\Omega = \mathbf{N} \setminus \{0\}$ with counting measure ν. The **geometric distribution** *with parameter $p = c^{-1}$ has maximal entropy among all distributions with fixed mean $c = E[X]$ for $X(k) = k$. It is unique with this property.*

c) $\Omega = \{0,1\}^N$ with counting measure ν. The **product distribution** *η^N, where $\eta(1) = p, \eta(0) = 1 - p$ with $p = c/N$ has maximal entropy among all distributions satisfying $E[S_N] = c$, where $S_N(\omega) = \sum_{i=1}^N \omega_i$. It is unique with this property.*

d) $\Omega = [0, \infty)$ with Lebesgue measure ν. The **exponential distribution** *with parameter α on Ω has maximal entropy among all distributions with mean $c = 1/\alpha = E[X]$, where $X(x) = x$. It is unique with this property.*

e) $\Omega = \mathbf{R}$ with Lebesgue measure ν. The **normal distribution** *$N(m, \sigma^2)$ has maximal entropy among all distributions having mean $E[X] = m$ and variance $\mathrm{Var}[X] = \sigma^2$ with $X(x) = x$. It is unique with this property.*

f) **Finite measures**. *Let $(\Omega, \mathcal{A})$ be an arbitrary measure space and $\nu(\Omega) < \infty$. Then the measure ν with uniform distribution $f = 1$ has maximal entropy among all other measures on Ω. It is unique with this property.*

Proof. Let $\mu = f\nu$ be the measure of the distribution from which we want to prove maximal entropy and let $\tilde{\mu} = \tilde{f}\nu$ be any other measure. The aim is to show $H(\tilde{\mu}|\mu) = H(\mu) - H(\tilde{\mu})$ which implies the maximality since by the above lemma $H(\tilde{\mu}|\mu) \geq 0$.
In general

$$H(\tilde{\mu}|\mu) = -H(\tilde{\mu}) - \int_\Omega \tilde{f}(\omega)\log(f(\omega))\, d\nu$$

so that we have to show

$$H(\mu) = \int_\Omega \tilde{f}(\omega)\log(f(\omega))\, d\nu\ . \tag{1.8}$$

Having so shown

$$H(\tilde{\mu}|\mu) = H(\mu) - H(\tilde{\mu})$$

gives also uniqueness: if two measures $\tilde{\mu}, \mu$ have maximal entropy, then $H(\tilde{\mu}|\mu) = 0$ so that by the above lemma $\mu = \tilde{\mu}$.

a) The density $f = 1/|\Omega|$ is constant. Therefore $H(\omega) = \log(|\Omega|)$ and Equation (1.8) holds.
b) The density for the geometric distribution is $f(k) = (1-p)^{k-1}p$ so that

$$\int_\Omega \tilde{f}(\omega)\log(f(\omega))\, d\nu = \log(p) + \int \tilde{f}(\omega)(k-1)\log(1-p)\, d\nu = \log(p) - \log(1-p)(1+c)$$

which is also the entropy of μ since we fixed $E[X] = c$.
c) The density is $f(\omega) = p^{S_N}(1-p)^{N-S_N}$ so that

$$\log(f(k)) = S_N\log(p) + (N - S_N)\log(1-p)\ .$$

The claim follows since we fixed $E[S_N]$.
d) The density is $f(x) = \alpha e^{-\alpha x}$, so that $\log(f(x)) = \log(\alpha) - \alpha x$. The claim follows since we fixed $E[X] = \int x\, d\tilde{\mu}(x)$ for all distributions.
e) For the normal distribution $\log(f(x)) = a + b(x-m)^2$ with two real number a, b depending only on m, σ. The claim follows since we fixed $\text{Var}[X] = E[(x-m)^2]$ for all distributions.
f) The density $f = 1$ is constant. Therefore $H(\mu) = 0$ which is also on the right hand side of Equation (1.8). □

Remark. The origin of this results ly in the foundations of **thermodynamics**, where one consideres the phase space of N particles moving in a finite region in eucledian space. The energy surface is then a compact manifold Ω and the motion on this surface leaves a measure ν invariant which is induced from the flow invariant Lebesgue measure. The measure ν is called the **microcanonical ensemble**. According to f) in the above, it is the measure which maximizes entropy.

Let us try to get the maximal distribution using **calculus of variations**. In order to find the maximum of the **functional**

$$H(f) = -\int f\log(f)\, d\nu$$

on $\mathcal{L}^1(\nu)$ under the constraints

$$F_0(f) = \int_\Omega f\, d\nu = 1,\ F_1(f) = \int_\Omega Xf\, d\nu = c\ ,$$

we differentiate

$$\tilde{H} = H - \lambda_0 F_0 - \lambda_1 F_1$$

(in infinite dimensions this is called **functional derivative**:

$$D(\tilde{H}) = -(1 + \log(f)) - \lambda_0 - X\lambda_1$$

and solve $D(\tilde{H})(f) = 0, F_0(f) = 1, F_1(f) = c$ with respect to f by eliminating the Lagrange multipliers λ_0, λ_1. We determine f by

$$\begin{aligned} f &= e^{\lambda_1 X + \lambda_0 - 1}, \\ c &= \frac{\int X e^{\lambda_1 X} \, d\nu}{\int e^{\lambda_1 X} \, d\nu}, \\ e^{1-\lambda_0} &= \int e^{\lambda_1 X} \end{aligned}$$

by finding λ_1 from the second and λ_0 from the third equation. This variational approach gives critical points of the entropy. If the Hessian $D^2(H) = -1/f$ is negative definite, it is also negative definite when restricted to the manifold in $\mathcal{L}^1$ determined by the restrictions $F_0 = 1, F_1 = c$. This indicates again that we have found a global maximum.

Examples.
1) For $\Omega = \mathbf{R}$, $X(x) = x^2$, we get the normal distribution $N(0,1)$.
2) For $\Omega = \mathbf{N}$, $X(n) = \epsilon_n$, we get $f(n) = e^{-\epsilon_n \lambda_1}/Z(f)$ with $Z(f) = \sum_n e^{-\epsilon_n \lambda_1}$ and where λ_1 is determined by $\sum_n \epsilon_n e^{-\epsilon_n \lambda_1} = c$. This is called the **discrete Maxwell-Bolzmann distribution**. In physics, one writes $\lambda^{-1} = kT$ with the **Bolzmann constant** k, determining T, the **temperature**.

We give a little dictionnary giving the correspondence of some notions in probability theory and statistical physics. The physical chargon has an advantage that it is more intuitive. It is however possible to avoid it completely.

Probability theory	Statistical mechanics (thermodynamics)
Set Ω	Phase space
Measure space	Thermodynamic system
Random variable	Observable (for example energy)
Probability density	Theromodynamic state
Entropy	Boltzmann-Gibbs entropy
Densities of maximal entropy	Thermodynamic equlibria
Central limit theorem	Maximal entropy principle

1.15 Gibbs distributions

Repetition: distributions, which maximize the entropy under some constraint are mathematically natural since they are critical points of a variational principle. Physically, they are natural, simply because nature prefers them. From the statistical mecahnical point of view, the extremal properties of entropy offer insight into the basic foundation of thermodynamics.

Definition. Given a measure space $(\Omega, \mathcal{A})$ with a not necessarily finite measure ν and a random variable $X \in \mathcal{L}$. Given $f \in \mathcal{L}^1$ leading to the probability measure $\mu = f\nu$. Consider the moment generating function $Z(\lambda) = E_\mu[e^{\lambda X}]$ and define $\Lambda = \{\lambda \in \mathbf{R} \,|Z(\lambda) < \infty\}$ which is a (possibly degenerated) interval in $\mathbf{R}$. For every $\lambda \in \Lambda$ we can define a new probability measure

$$\mu_\lambda = f_\lambda \nu = \frac{e^{\lambda X}}{Z(\lambda)}\mu$$

on Ω. The set

$$\{\mu_\lambda \mid \lambda \in \Lambda\}$$

of measures on $(\Omega, \mathcal{A})$ is called the (to μ and X belonging) **exponential family**.

Theorem 1.15.1 *For all probability measures $\tilde{\mu}$ which are absolutely continuous with respect to ν, we have for all $\lambda \in \Lambda$*

$$H(\tilde{\mu}|\mu) - \lambda E_{\tilde{\mu}}[X] \geq -\log Z(\lambda) \ .$$

The minimum $-\log Z(\lambda)$ is obtained for μ_λ.

Proof. For every $\tilde{\mu} = \tilde{f}\nu$, we have

$$\begin{aligned} H(\tilde{\mu}|\mu) &= \int_\Omega \tilde{f} \log \frac{\tilde{f}}{\tilde{f}_\lambda}\frac{\tilde{f}_\lambda}{f} \, d\nu \\ &= -H(\tilde{\mu}|\mu_\lambda) + (-\log(Z(\lambda)) + \lambda E_{\tilde{\mu}}[X]) \ . \end{aligned}$$

For $\tilde{\mu} = \mu_\lambda$, we have

$$H(\mu_\lambda|\mu) = -\log(Z(\lambda)) + \lambda E_{\mu_\lambda}[X] \ .$$

Therefore

$$H(\tilde{\mu}|\mu) - \lambda E_{\tilde{\mu}}[X] = H(\tilde{\mu}|\mu_\lambda) - \log(Z(\lambda)) \geq -\log Z(\lambda) \ .$$

The minimum is obained for $\tilde{\mu} = \mu_\lambda$. □

Corollary 1.15.2 *a) μ_λ minimizes the relative entropy $\tilde{\mu} \mapsto H(\tilde{\mu}|\mu)$ among all ν-absolutely continuous measures $\tilde{\mu}$ with fixed $E_{\tilde{\mu}}[X]$.*
b) If we fix λ by requiring $E_{\mu_\lambda}[X] = c$, then μ_λ maximizes the entropy $H(\tilde{\mu})$ among all measures satisfying $E_{\tilde{\mu}}[X] = c$.

Proof. a) Minimizing $\tilde{\mu} \mapsto H(\tilde{\mu}|\mu)$ under the constraint $E_{\tilde{\mu}}[X] = c$ is equivalent to minimize

$$H(\tilde{\mu}|\mu) - \lambda E_{\tilde{\mu}}[X],$$

and to determine the Lagrange multiplier λ by $E_{\mu_\lambda}[X] = c$. The above theorem shows that μ_λ is minimizing that.
b) $\mu = f\nu, \mu_\lambda = e^{-\lambda X} f/Z$. From

$$0 \leq H(\tilde{\mu}, \mu_\lambda) = -H(\tilde{\mu}) + (-\log(Z)) - \lambda E_\mu[X] = -H(\tilde{\mu}) + H(\mu_\lambda) \ .$$

□

Corollary 1.15.3 *Assume $\nu = \mu$ is a probability measure. The measure μ_λ maximizes*

$$F(\mu) = H(\mu) + \lambda E_\mu[X]$$

among all measures $\tilde{\mu}$ absolutely continuous with respect to μ.

Proof. Take $\mu = \nu$. Since then $f = 1$, $H(\tilde{\mu}|\mu) = -H(\tilde{\mu})$. The claim follows now immediatly from the theorem since a minimum of $H(\tilde{\mu}|\mu) - \lambda E_{\tilde{\mu}}[X]$ corresponds to a maximum of $F(\mu)$.
□

This corollary can also be proven by calculus of variations, namely by finding the maximum of $F(f) = \int f \log(f) + X f d\nu$ under the condition $\int f \ d\nu = 1$.

Remark. In statistical mechanics, the measure μ_λ is called the **Gibbs distribution** or **Gibbs canonical ensemble** for the observable X and $Z(\lambda)$ is the **partition function**. One uses in physics the notation $\lambda = -(kT)^{-1}$, where T is the temperature. Maximizing $H(\mu) - (kT)^{-1} E_\mu[X]$ is the same than minimizing $E_\mu[X] - kTH(\mu)$ which is called the **free energy** if X is the **Hamiltonian** and $E_\mu[X]$ is the **energy**. The measure μ is the apriori model, the **microcanonical ensemble**. Adding the restriction that X has a specific expectation value $c = E_\mu[X]$ leads to the probability measure μ_λ, the **canonical ensemble**. We illustrated two physical principles: nature maximizes entropy when the energy is fixed and minimizes the free energy, when energy is not fixed.

Examples.
1) Take on the real line the Hamiltonian $X(x) = x^2$ and a measure $\mu = f dx$, we get the energy $\int x^2 \ d\mu$. Among all symmetric distributions fixing the energy, the Gaussian distribution maximizes the entropy.
2) Take $\Omega = \mathbf{N} = \{0, 1, 2, \ldots\}$ and $X(k) = k$, ν the counting measure on Ω and μ the Poisson measure with parameter 1. The partition function is

$$Z(\lambda) = \sum_k e^{\lambda k} \frac{e^{-1}}{k!} = \exp(e^\lambda - 1)$$

so that $\Lambda = \mathbf{R}$ and μ_λ is given by the weights

$$\mu_\lambda(k) = \exp(e^{-\lambda} + 1) e^{\lambda k} \frac{e^{-1}}{k!} = e^{-\alpha} \frac{\alpha^k}{k!} \ ,$$

where $\alpha = e^\lambda = 0$. The exponential family of the Poisson measure is the family of all Poisson measures (parametrized differently).
3) We know from the previous section that the geometric distribution on $\mathbf{N} \setminus \{0\} = \{1, 2, 3, \ldots\}$ is an exponential family. We saw also that the product measure on $\Omega = \{0, 1\}^N$ with win probability p is an exponential familiy with respect to $X(k) = k$.

4) $\Omega = \{1, \dots, N\}$, ν the counting measure and let μ_p be the Binomial distribution with p. Take $\mu = \mu_{1/2}$ and $X(k) = k$. Since

$$0 \leq H(\tilde{\mu}|\mu) = H(\tilde{\mu}|\mu_p) + \log(p)E[X] + \log(1-p)E[(N - E[X])] = -H(\tilde{\mu}|\mu_p) + H(\mu_p) ,$$

μ_p is an exponential family.

There is an obvious generalisation of the maximum entropy principle to the case, when we have finitely many random variables $\{X_i\}_{i=1}^n$. Given $\mu = f\nu$ we define the (n-dimensional) exponential family

$$\mu_\lambda = f_\lambda \nu = \frac{e^{\sum_{i=1}^n \lambda_i X_i}}{Z(\lambda)} \mu ,$$

where

$$Z(\lambda) = E_\mu[e^{\sum_{i=1}^n \lambda_i X_i}]$$

is the partition function defined on a subset Λ of $\mathbf{R}^n$.

Theorem 1.15.4 *For all probability measures $\tilde{\mu}$ which are absolutely continuous with respect to ν, we have for all $\lambda \in \Lambda$*

$$H(\tilde{\mu}|\mu) - \sum_i \lambda_i E_{\tilde{\mu}}[X_i] \geq -\log Z(\lambda) .$$

The minimum $-\log Z(\lambda)$ is obtained for μ_λ.
If we fix λ_i by requiring $E_{\mu_\lambda}[X_i] = c_i$, then μ_λ maximizes the entropy $H(\tilde{\mu})$ among all measures $\tilde{\mu}$ satisfying $E_{\tilde{\mu}}[X_i] = c_i$.
Assume $\nu = \mu$ is a probability measure. The measure μ_λ maximizes

$$F(\tilde{\mu}) = H(\tilde{\mu}) + \lambda E_{\tilde{\mu}}[X] .$$

Proof. Take the same proofs as before by replacing λX with $\lambda \cdot X = \sum_i \lambda_i X_i$. □

1.16 Markov operators

Definition. Given a (not necessarily finite) probability space $(\Omega, \mathcal{A}, \nu)$. A linear operator $P : \mathcal{L}^1(\Omega) \to \mathcal{L}^1(\Omega)$ is called a **Markov operator**, if

$$\boxed{\begin{array}{l} f \geq 0 \Rightarrow Pf \geq 0, \\ f \geq 0 \Rightarrow ||Pf||_1 = ||f||_1 . \end{array}}$$

In other works, a Markov operator P has to leave invariant the closed positive cone $\mathcal{L}^1_+ = \{f \in \mathcal{L}^1 \mid f \geq 0\}$ and preserve there the norm.

A Markov operator on $(\Omega, \mathcal{A}, \nu)$ leaves invariant the set $\mathcal{D}(\nu) = \{f \in \mathcal{L}^1 \mid f \geq 0, ||f||_1 = 0\}$ of **probability densities**. They correspond bijectively to the set $\mathcal{P}(\nu)$ of probability measures which are absolutely continuous with respect to ν. A Markov operator is (therefore) also called a **stochastic operator**.

Examples. Let T be a measure preserving transformation on $(\Omega, \mathcal{A}, \nu)$. It is called **nonsingular** if $T^*\nu$ is absolutely continuous with respect to ν. The unique operator $P : \mathcal{L}^1 \to \mathcal{L}^1$ satisfying

$$\int_A Pf \, d\nu = \int_{T^{-1}A} f \, d\nu$$

is called the **Perron-Frobenius operator** associated to T. Closely related is the operator $Pf(x) = f(Tx)$, the **Koopmann operator**. Both are Markov operators.

Exercice. Assume $\Omega = [0, 1]$ with Lebesgue measure μ. Compute the Perron-Frobenius operator for the tent map

$$T(x) = \left\{ \begin{array}{cl} 2x & , x \in [0, 1/2] \\ 2(1-x) & , x \in [1/2, 1] \end{array} \right. .$$

Answer. $Pf(x) = \frac{1}{2}(f(\frac{1}{2}x) + f(1 - \frac{1}{2}x))$.

Lemma 1.16.1 (Jensen Inequality for positive operators) *Given a convex function u and an operator $P : \mathcal{L}^1 \to \mathcal{L}^1$ mapping positive functions into positive functions, then*

$$u(Pf) \leq Pu(f)$$

for all $f \in \mathcal{L}^1_+$ for which $Pu(f)$ exists.

Proof. We have to show $u(Pf)(\omega) \leq Pu(f)(\omega)$ for almost all $\omega \in \Omega$. Given $x = (Pf)(\omega)$, there exists by definition of convexity a linear function $y \mapsto ay + b$ such that $u(x) = ax + b$ and $u(y) \geq ay + b$ for all $y \in \mathbf{R}$. Therefore, since $af + b \leq u(f)$ and P is positive

$$u(Pf)(\omega) = a(Pf)(\omega) + b = P(af + b)(\omega) \leq P(u(f))(\omega) .$$

□

The following theorem states that relative entropy is nonincreasing along orbits of Markov operators. The assumption that $\{f > 0\}$ is mapped into itself is actually not necessary but avoids some functional analytic gymnastic in the proof of Voigt.

Theorem 1.16.2 (Voigt, 1981) *Given a Markov operator P which maps $\{f > 0\}$ into itself. For all $f, g \in \mathcal{L}^1_+$,*

$$H(Pf|Pg) \leq H(f|g) .$$

Proof. We can assume that $\{g(\omega) = 0\} \subset A = \{f(\omega) = 0\}$ because nothing is to show in the case $H(f|g) = \infty$. By restriction to the measure space space $(A^c, \mathcal{A} \cap A^c, \nu(\cdot \cap A))$, we can assume $f > 0, g > 0$ so that by our assumption also $Pf > 0$ and $Pg > 0$.

(i) Assume first $(f/g)(\omega) \leq c$ for some constant $c \in \mathbf{R}$.
For fixed g, the linear operator $Rh = P(hg)/P(g)$ maps positive functions into positive functions. Take the convex function $u(x) = x \log(x)$ and put $h = f/g$. Using Jensen's inequality, we get

$$\frac{Pf}{Pg} \log \frac{Pf}{Pg} = u(Rh) \leq Ru(h) = \frac{P(f \log(f/g))}{Pg}$$

which is equivalent to $Pf \log \frac{Pf}{Pg} \leq P(f \log(f/g))$. Integration gives

$$\begin{aligned} H(Pf|Pg) &= \int Pf \log \frac{Pf}{Pg} \, d\nu \\ &\leq \int P(f \log(f/g)) \, d\nu = \int f \log(f/g) \, d\nu = \int H(f|g) . \end{aligned}$$

(ii) Define $f_k = \inf(f, kg)$ so that $f_k/g \leq k$. We have $f_k \subset f_{k+1}$ and $f_k \to f$ in $\mathcal{L}^1$. From (i) we know that $H(Pf_k|Pg) \leq H(f_k|g)$. We can assume $H(f|g) < \infty$ because the result is trivially true in the other case. Define $B = \{f \leq g\}$. On B, we have $f_k \log(f_k/g) = f \log(f/g)$ and on $\Omega \setminus B$ we have

$$f_k \log(f_k/g) \leq f_{k+1} \log(f_{k+1}/g) u \to f \log(f/g)$$

so that by Lebesgue dominated convergence theorem, $H(f|g) = \lim_{k \to \infty} H(f_k|g)$. As an increasing sequence, Pf_k converges to Pf almost everywhere. The elementary inequality $x \log(x) - x \geq x \log(y) - y$ for all $x, y \geq 0$ gives

$$Pf_k \log Pf_k - Pf_k \log Pg - Pf_k + Pg \geq 0 .$$

Integration gives with Fatou's lemma

$$H(Pf|Pg) - ||Pf|| + ||Pg|| \leq \liminf H(Pf_k|Pg) - ||Pf_k|| + ||Pg||$$

and so $H(Pf|Pg) \leq \liminf H(Pf_k|Pg)$. □

Corollary 1.16.3 *For a invertible Markov operator P, the relative entropy is constant: $H(Pf|Pg) = H(f|g)$.*

Proof. Since P and P^{-1} are both Markov operators

$$H(f|g) = H(PP^{-1}f|PP^{-1}g) \leq H(P^{-1}f|P^{-1}g) \leq H(f|g) .$$

□

Example. If a measure preserving transformation T is invertible, then the corresponding Koopman operator and Perron Frobenius operators preserve relative entropy.

Corollary 1.16.4 *The operator* $T(\mu)(A) = \int_{\mathbf{R}^2} 1_A(\frac{x+y}{\sqrt{2}}) \; d\mu(x) \; d\mu(y)$ *does never decrease entropy.*

Proof. Denote with X_μ a random variable having the law μ and with $\mu(X)$ the law of a random variable. For a fixed random variable Y, we define the Markov operator

$$P_Y(\mu) = \mu(\frac{X_\mu + Y}{\sqrt{2}}) .$$

(see the chapter about characteristic functions). Since the entropy is nondecreasing for each P_Y, we have this property also for $T(\mu) = P_{X_\mu}(\mu)$ which is a nonlinear operator. □

We have shown as a corollary of the central limit theorem that T has a unique fixed point attracting all of $\mathcal{P}_{0,1}$. The entropy is also strictly increasing at infinitely many points of the orbit $T^n(\mu)$ since it converges to the fixed point with maximal entropy. It follows that T is not invertible.

More generally: given a sequence X_n of IID random variables. For every n, the map P_n which maps the law of S_n^* into the law of S_{n+1}^* is a Markov operator which does not increase entropy. We can summarize: summing up IID random variables tends to increase the entropy of the distributions.

1.17 Characteristic functions

Since distribution functions are in general not easy to deal with, it is convenient to deal with its Fourier transforms, the characteristic functions.

Definition. Given a random variable X, its **characteristic function** is a real valued function on $\mathbf{R}$ defined as

$$f_X(u) = E[e^{iuX}] .$$

If F_X is the distribution function of X and μ_X its law, the characteristic function of X is the Fourier-Stieltjes transform

$$f(u) = \int_{\mathbf{R}} e^{iux} \, dF_X(x) = \int_{\mathbf{R}} e^{iux} \, \mu_X(dx) .$$

Remarks.
1) If F is a continuous distribution function $dF(x) = h(x)\, dx$, then f is the **Fourier transform** of h.
2) By definition, characteristic functions are Fourier transforms of probability measures.

Lemma 1.17.1 (Characteristic functions determine the distribution) *The characteristic function ϕ_X determines the distribution of X.*

Proof. (Proof by literature):
Since the Fourier transform of a measure determines the measure uniquely, the characteristic function of a distribution determines the distribution.
(Proof by hand):
Since a distribution function F has only countably many points of discontinuities, it is enough to determine $F(b) - F(a)$ in terms of ϕ if a and b are continuity points of F. A computation proves

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) \, dt .$$

□

Examples of characteristic functions.

Distribution	Parameter	Characteristic function
Normal	$m \in \mathbf{R}$, $\sigma^2 > 0$	$e^{mit - \sigma^2 t^2/2}$
St. normal		$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Binomial	$n \in \mathbf{N}$, $p \in [0, 1]$	$(1 - p + pe^{it})^n$
Poisson	$\lambda > 0$, λ	$e^{\lambda(e^{it} - 1)}$
Geometric	$p \in (0, 1)$	$\frac{pe^{it}}{(1-(1-p)e^{it}}$

Definition. Let F and G be two distribution functions. Their **convolution** $F \star G$ is defined by

$$F \star G(x) = \int_{\mathbf{R}} F(x-y)dG(y) \; .$$

Lemma 1.17.2 *If F and G are distribution functions, then $F \star G$ is a distribution function.*

Proof. We have to show properties a),b) and c) which characterizes distribution functions among real valued functions.
a) Since F is nondecreasing, also $F \star G$ is nondecreasing.
b) Since $F(-\infty) = 0$ also $F \star G(-\infty) = 0$. Since $F(\infty) = 1$ and dG is a probability measure, also $F \star G(\infty) = 1$.
c) Given a sequence $h_n \to 0$. Define $F_n(x) = F(x + h_n)$. Since F is continuous from the right, $F_n(x)$ converges pointwise to $F(x)$. Lebesgue dominated convergence theorem implies that $F_n \star G(x) = F \star G(x + h_n)$ converges to $F \star G(x)$. □

Example. Given two discrete distributions $F(x) = \sum_{n \leq x} p_n$, $G(x) = \sum_{n \leq x} q_n$. Then $F \star G(x) = \sum_{n \leq x} (p \star q)_n$, where $p \star q$ is the convolution of the sequences p, q defined by $(p \star q)_n = \sum_{k=0}^{n} p_k q_{n-k}$. We see that the convolution of discrete distributions gives again a discrete distribution.

Example. Given two continuous distributions F, G with densities h and k. Then the distribution of $F \star G$ is given by the convolution

$$h \star k(x) = \int_{\mathbf{R}} h(x-y)k(y) \; dy \; ,$$

since

$$F \star G(x) = \int_{\mathbf{R}} F(x-y)k(y) \; dy = \int_{\mathbf{R}} \int_{-\infty}^{x-y} h(z)dzk(y) \; dy$$

leads after a change of variables $x \mapsto x - y$, $z \mapsto z - y$ to

$$F \star G(x) = \int_{-\infty}^{x} \int_{\mathbf{R}} h(z-y)k(y) \; dydz \; .$$

Lemma 1.17.3 *If F and G are distribution functions with characteristic functions f and g, then $F \star G$ has the distribution function $f \cdot g$.*

Proof. One can assume this as a fact from the theory of Fourier transforms or prove it by hand: using an approximation of the integral by step functions:

$$\begin{aligned} & \int_{\mathbf{R}} e^{iux} d(F \star G)(x) \\ = & \lim_{N,n \to \infty} \sum_{k=-N2^n+1}^{N2^n} e^{iuk2^{-n}} \int_{\mathbf{R}} [F(\frac{k}{2^n} - y) - F(\frac{k-1}{2^n} - y)] \; dG(y) \end{aligned}$$

$$= \lim_{N,n\to\infty} \sum_{k=-N2^n+1}^{N2^n} \int_{\mathbf{R}} e^{iu\frac{k}{2^n}-y}[F(\frac{k}{2^n}-y) - F(\frac{k-1}{2^n}-y)] \cdot e^{iuy}\, dG(y)$$

$$= \int_{\mathbf{R}} [\lim_{N\to\infty} \int_{-N-y}^{N-y} e^{iux} dF(x)] e^{iuy}\, dG(y) = \int_{\mathbf{R}} f(u) e^{iuy} dG(y) = f(u)g(u)\ .$$

□

It follows that the set of distribution functions forms an associative commutative group with the convolution multiplication since its characteristic functions have these properties with pointwise multiplication.

Characteristic functions become especially useful, if one deals with independent random variables.

Proposition 1.17.4 *Given a sequence of independent random variables* X_j *with characteristic functions* ϕ_j*. The characteristic function of* $\sum_{j=1}^n X_j$ *is* $\phi = \prod_{j=1}^n \phi_j$.

Proof. Since X_j are independent, we get for any complex valued Borel-measureable functions g_j for which $E[g_j(X_j)]$ exists

$$E[\prod_{j=1}^n g_j(X_j)] = \prod_{j=1}^n E[g_j(X_j)]\ .$$

If we put $g_j(x) = \exp(ix)$, the proposition is proven. □

Corollary 1.17.5 *The distribution of the random variable* $\sum_{j=1}^n X_j$ *is given by* $\phi_1 \star \phi_2 \dots, \star\phi_n$*, if* X_j *have the distributions* f_j.

Proof. This follows immediately from the last proposition and the algebraic isomorphisms between the set of characteristic functions with convolution product and the set of distribution functions with pointwise multiplication. □

Exercice 5

Topic: Characteristic functions, Entropy, Markov operators

Characteristic functions.

1) Show that $X_n \to X$ in distribution if and only if $\phi_{X_n}(t) \to \phi_X(t)$ for all $t \in \mathbf{R}$.

2) a) The **characteristic function** of a vector valued random variable $X = (X_1, \dots, X_k)$ is defined as $\phi_X : \mathbf{R}^k \to \mathbf{R}$.
$$\phi_X(t) = E[e^{it \cdot X}] .$$
Two such random variables X, Y are **independent**, if the σ-algebras $X^{-1}(\mathcal{B})$ and $Y^{-1}(\mathcal{B})$ are independent, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbf{R}^k$. Show that if X and Y are independent then $\phi_{X+Y} = \phi_X \cdot \phi_Y$.
b) A given a real nonsingular $k \times k$ matrix A called the **covariance matrix** and a vector $m = (m_1, \dots, m_k)$ called the **mean**. A vector valued random variable X has a **Gaussian distribution with covariance** A **and mean** m, if
$$\phi_X(t) = e^{im \cdot t - \frac{1}{2}(t \cdot At)} .$$
Show that the sum $X + Y$ of two Gaussian distributed random variables is again Gaussian distributed.
c) Find the probability density of a Gaussian distributed random variable X with covariance matrix A and mean m.

Laplace transform, Moment generating function, generating function.

3) The **Laplace transform** of a positive random variable $X \geq 0$ is defined as $l_X(t) = E[e^{-tX}]$. The **moment generating function** is defined as $M(t) = E[e^{tX}]$ provided that the expectation exists in a neighborhood of 0. The **generating function** of an integer-valued random variable is defined as $\zeta(X) = E[u^X]$ for $u \in (0, 1)$. What does independence of two random variables X, Y mean in terms of (i) the Laplace transform, (ii) the moment generating function or (iii) the generating function?

Gibbs potential .

4) Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and let $U, V \in \mathcal{X}$ be random variables (describing the energy density and the mass density of a thermodynamical system). We have seen that the Helmholz free energy

$$E_{\tilde{\mu}}[U] - kTH[\tilde{\mu}]$$

(k is a physical constant), T is the temperature, is taking its minimum for the exponential family. Find the measure minimizing the **free enthalpy** or **Gibbs potential**

$$E_{\tilde{\mu}}[U] - kTH[\tilde{\mu}] - pE_{\mu}[V] ,$$

where p is the pressure.

5) Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and $X_i \in \mathcal{L}$ random variables. Compute $E_{\mu}[X_i]$ and the entropy of μ_{λ} in terms of the partition function $Z(\lambda)$.

Blackbody radiation

6) a) Given the discrete measure space $(\Omega = \{\epsilon_0 + n\delta\}, \nu)$, with $\epsilon_0 \in \mathbf{R}$ and $\delta > 0$ and where ν is the counting measure and let $X(k) = k$. Find the distribution f maximizing the entropy $H(f)$ among all measures $\tilde{\mu} = f\nu$ fixing $E_{\tilde{\mu}}[X] = \epsilon$.

b) The physical interpretation is as follows: Ω is the discrete set of energies of a harmonic oscillator, ϵ_0 is the ground state energy, $\delta = \hbar\omega$ is the incremental energy, where ω is the frequency of the oscillation and $\hbar$ is Plank's constant. $X(k) = k$ is the Hamiltonian and $E[X]$ is the energy. Put $\lambda = 1/kT$, where T is the temperature (in the answer of a), there appears a parameter λ, the Lagrange multiplier of the variational problem). Since can fix also the temperature T instead of the energy ϵ, the distribution in a) maximizing the entropy is determined by ω and T. Compute the spectrum $\epsilon(\omega, T)$ of the blackbody radiation defined by

$$\epsilon(\omega, T) = (E[X] - \epsilon_0)\frac{\omega^2}{\pi^2 c^3}$$

where c is the velocity of light. You have deduced then **Planck's blackbody radiation formula**.

Markov operators.

7) Compute the Markov operator of the tent map (verify the answer in the notes).

8) (*) Compute the Markov operator for the quadratic map $T(x) = 4x(1-x)$. Find the stationnary density of P, (the fixed point of P) and compute the entropy of this density.

1.18 The law of the iterated logarithm

We will give only a proof of the law of iterated logarithm in the special case, when the random variables X_n are IID with standard normal distribution. The proof of the theorem for general IID random variables X_n would take some more time. The full proof can be found for example in the book of Strook. The central limit theorem makes however the general result plausible from the special case, we are going to prove.

Definition. A random variable $X \in \mathcal{L}$ is called **symmetric** if its law μ_X is even: $\mu([-b,-a)) = \mu([a,b))$ for all $a < b$. A symmetric random variable $X \in \mathcal{L}^1$ has zero mean. We use the notation $S_n = \sum_{k=1}^n X_k$.

Lemma 1.18.1 *Let X_n by symmetric and independent. For every $\epsilon > 0$*

$$P(\max_{1 \leq k \leq n} S_k > \epsilon) \leq 2P(S_n > \epsilon) .$$

Proof. This is a direct consequence of Levy's theorem since we can take $m = 0$ as the median of a symmetric distribution. □

Define $\Lambda_n = \sqrt{2n \log\log n}$ (for $n \geq 2$).

Theorem 1.18.2 (Law of iterated logarithm for $N(0,1)$) *Let X_n be a sequence of IID $N(0,1)$-distributed random variables. Then*

$$\limsup_{n\to\infty} \frac{S_n}{\Lambda_n} = 1, \; \liminf_{n\to\infty} \frac{S_n}{\Lambda_n} = -1 .$$

Proof. Since the second statement follows obviously from the first one by replacing X_n by $-X_n$, we have only to prove $\limsup_{n\to\infty} S_n/\Lambda_n = 1$.

(i) $P(S_n > (1+\epsilon)\Lambda_n$, infinitely often$) = 0$ for all $\epsilon > 0$.

Define $n_k = [(1+\epsilon)^k] \in \mathbf{N}$ and the events

$$A_k = \{S_n > (1+\epsilon)\Lambda_n, \text{for some } n \in (n_k, n_{k+1}]\ \}.$$

Clearly $\limsup_k A_k = \{S_n > (1+\epsilon)\Lambda_n$, infinitely often$\}$. By the Borel Cantelli, it is enough to show that $\sum_k P(A_k) < \infty$. For each k, we get with the above lemma

$$\begin{aligned} P(A_k) &\leq P(\max_{n_k < n \leq n_{k+1}} S_n > (1+\epsilon)\Lambda_k) \\ &\leq P(\max_{1 \leq n \leq n_{k+1}} S_n > (1+\epsilon)\Lambda_k) \\ &\leq 2P(S_{n_{k+1}} > (1+\epsilon)\Lambda_k) . \end{aligned}$$

The right-hand side can be estimated further using that $S_{n_{k+1}}/\sqrt{n_{k+1}}$ is $N(0,1)$-distributed and that for a $N(0,1)$-distributed random variable $P(X > t) \leq \text{const} \cdot e^{-t^2/2}$

$$2P(S_{n_{k+1}} > \Lambda_k) \;\; = \;\; 2P\left(\frac{S_{n_{k+1}}}{\sqrt{n_{k+1}}} > (1+\epsilon)\frac{\sqrt{2n_k \log\log n_k}}{\sqrt{n_{k+1}}}\right)$$

$$
\begin{aligned}
&\leq C\exp(-\frac{1}{2}(1+\epsilon)^2)\frac{2n_k\log\log n_k}{n_{k+1}}) \\
&\leq C_1\exp(-(1+\epsilon)\log\log(n_k)) \\
&= C_1\log(n_k)^{-(1+\epsilon)} \leq C_2 k^{-(1+\epsilon)} .
\end{aligned}
$$

Having shown that $P(A_k) \leq \text{const}\cdot k^{-(1+\epsilon)}$ proves the claim $\sum_k P(A_k) < \infty$.

(ii) $P(S_n > (1-\epsilon)\Lambda_n,$ infinitely often$) = 1$ for all $\epsilon > 0$.

It suffices to show, that for all $\epsilon > 0$ there exists a subsequence n_k

$$P(S_{n_k} > (1-\epsilon)\Lambda_{n_k}, \text{ infinitely often}) = 1 .$$

Given $\epsilon > 0$. Choose $N > 1$ large enough and $c < 1$ near enough to 1 such that

$$c\sqrt{1-1/N} - 2/\sqrt{N} > 1-\epsilon . \tag{1.9}$$

Define $n_k = N^k$ and $\Delta n_k = n_k - n_{k-1}$. The sets

$$A_k = \{S_{n_k} - S_{n_{k-1}} > c\sqrt{2\Delta n_k \log\log \Delta n_k}\}$$

are independent. In the following estimate, we use the fact that $\int_t^\infty e^{-x^2/2}\, dx \geq C\cdot e^{-t^2/2}$ for some constant C.

$$
\begin{aligned}
P(A_k) &= P(\{S_{n_k} - S_{n_{k-1}} > c\sqrt{2\Delta n_k \log\log \Delta n_k}\}) \\
&= P(\{\frac{S_{n_k} - S_{n_{k-1}}}{\sqrt{\Delta n_k}} > c\frac{\sqrt{2\Delta n_k \log\log \Delta n_k}}{\sqrt{\Delta n_k}}\}) \\
&\geq C\cdot\exp(-c^2\log\log\Delta n_k) = C\cdot\exp(-c^2\log(k\log N)) \\
&= C_1\cdot\exp(-c^2\log k) = C_1 k^{-c^2}
\end{aligned}
$$

so that $\sum_k P(A_k) = \infty$. We have therefore by Borel-Cantelli a set A of full measure so that for $\omega \in A$

$$S_{n_k} - S_{n_{k-1}} > c\sqrt{2\Delta n_k \log\log\Delta n_k}$$

for infinitly many k. From (i), we know that

$$S_{n_k} > -2\sqrt{2n_k\log\log n_k}$$

for sufficiently large k. Both inequalites hold therefore for infinitely many values of k. We have for such k

$$
\begin{aligned}
S_{n_k}(\omega) &> S_{n_{k-1}}(\omega) + c\sqrt{2\Delta n_k\log\log\Delta n_k} \\
&\geq -2\sqrt{2n_{k-1}\log\log n_{k-1}} + c\sqrt{2\Delta n_k\log\log\Delta n_k} \\
&\geq (-2/\sqrt{N} + c\sqrt{1-1/N})\sqrt{2n_k\log\log n_k} \\
&\geq (1-\epsilon)\sqrt{2n_k\log\log n_k} ,
\end{aligned}
$$

where we have used assumption (1.9) in the last inequality. □

We know that $N(0,1)$ is the unique fixed point of the map T by the central limit theorem. The Law of iterated Logarithm is true for $T(X)$ implies that it is true for X. This shows that it would be enough to prove the theorem in the case when X has distribution in an arbitrary small neighborhood of $N(0,1)$. We would need however quite sharp estimates.

1.19 Use of characteristic functions

We present here a proof of the central limit theorem in the IID case to illustrate the use of characteristic functions.

> **Theorem 1.19.1 (Central limit theorem for IID random variables)** *Given $X_n \in \mathcal{L}^2$ which are IID with mean 0 and variance σ^2. Then $S_n/(\sigma\sqrt{n}) \to N(0,1)$ in distribution.*

Proof. Since the characteristic function of $N(0,1)$ is $e^{-t^2/2}$ we have to show that for all $t \in \mathbf{R}$

$$E[e^{it\frac{S_n}{\sigma\sqrt{n}}}] \to e^{-t^2/2} .$$

Denote with ϕ the characteristic function of X_n. Since by assumption $E[X_n] = 0, E[X_n^2] = \sigma^2$, we have

$$\phi(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2) .$$

Therefore

$$\begin{aligned} E[e^{it\frac{S_n}{\sigma\sqrt{n}}}] &= \phi(\frac{t}{\sigma\sqrt{n}})^n \\ &= (1 - \frac{1}{2}\frac{t^2}{n} + o(1/n))^n \\ &= e^{-t^2/2} + o(1) . \end{aligned}$$

□

This method can be adapted to more general situations: An example.

Given a sequence of independent events $A_n \subset \Omega$ with $P(A_n) = 1/n$. Define the random variables $X_n = 1_{A_n}$ and $S_n = \sum_{k=1}^n X_k$. Then

$$E[S_n] = \sum_{k=1}^n \frac{1}{k} = \log(n) + \gamma + o(1), \ \text{Var}[S_n] = \sum_{k=1}^n \frac{1}{k}(1 - 1/k) = \log(n) + \gamma - \pi^2/6 + o(1) .$$

The random variables $T_n = \frac{S_n - \log(n)}{\sqrt{\log(n)}}$ are not quite S_n^* but satisfy $E[T_n] \to 0$ and $\text{Var}[T_n] \to 1$. Compute $\phi_{X_n} = 1 - \frac{1}{n} + \frac{e^{it}}{n}$ so that $\phi_{S_n}(t) = \prod_{k=1}^n (1 - \frac{1}{k} + \frac{e^{it}}{k})$ and $\phi_{T_n}(t) = \phi_{S_n}(s(t))e^{-is\log(n)}$ where $s = t/\sqrt{\log(n)}$. For $n \to \infty$, we compute

$$\begin{aligned} \log \phi_{T_n}(t) &= -it\sqrt{\log(n)} + \sum_{k=1}^n \log(1 + \frac{1}{k}(e^{is} - 1)) \\ &= -it\sqrt{\log(n)} + \sum_{k=1}^n \log(1 + \frac{1}{k}(is - \frac{1}{2}s^2 + o(s^2))) \\ &= -it\sqrt{\log(n)} + \sum_{k=1}^n \frac{1}{k}(is + \frac{1}{2}s^2 + o(s^2)) + O(\sum_{k=1}^n \frac{s^2}{k^2}) \\ &= -it\sqrt{\log(n)} + (is - \frac{1}{2}s^2 + o(s^2))(\log(n) + O(1)) + t^2 O(1) \\ &= \frac{-1}{2}t^2 + o(1) \to -\frac{1}{2}t^2 . \end{aligned}$$

Chapter 2

Discrete Martingales

2.1 Conditional Expectation

Definition. Given a probability space $(\Omega, \mathcal{A}, P)$. A measure on $(\Omega, \mathcal{A})$ is called **absolutely continuous** with respect to P, if $P(A) = 0$ implies $P'(A) = 0$. One writes then $P' \ll P$.

Theorem 2.1.1 (Radon-Nykodym, 1913) *Given a signed measure P' which is absolutely continuous with respect to P, then there exists $Y \in \mathcal{L}^1(P)$ with $P' = YP$. The function Y is called the* **Radon-Nykodym derivative** *of P' with respect to P. It is unique in L^1.*

Proof. We can assume without loss of generality that P' is a positive measure (do else the Hahn decomposition $P = P^+ - P^-$), where P^+ and P^- are positive measures).
The set $\Gamma = \{Y \geq 0 \mid E[Y; A] \leq P'(A), \forall A \in \mathcal{A}\}$ is closed under formation of suprema

$$
\begin{aligned}
E[Y_1 \vee Y_2; A] &= E[Y_1; A \cap \{Y_1 > Y_2\}] + E[Y_2; A \cap \{Y_2 \geq Y_1\}] \\
&\leq P'(A \cap \{Y_1 > Y_2\}) + P'(A \cap \{Y_2 \geq Y_1\}) = P'(A)
\end{aligned}
$$

and contains a function Y different from 0 since else, P' would be singular with respect to P. We claim that the order supremum Y in Γ satisfies $YP = P'$: an application of Beppo-Levi's theorem shows that the supremum of Γ is in Γ. The measure $P'' = P' - YP$ is vanishing since we could do the same argument with a new set Γ for the absolutely continuous part of P''. To the uniqueness: assume there exist two derivatives Y, Y'. One has then $E[Y - Y'; \{Y \geq Y'\}] = 0$ and so $Y \geq Y'$ almost everywhere. A symmetric argument gives also $Y' \leq Y$ almost everywhere so that $Y = Y'$ almost everywhere. □

Lemma 2.1.2 (Conditional expectation, Kolmogorov 1933)
Given $X \in \mathcal{L}^1(\mathcal{A})$ and let $\mathcal{B} \subset \mathcal{A}$ be a sub σ-algebra. There exists $Y \in \mathcal{L}^1(\mathcal{B})$ with $\int_A Y \, dP = \int_A X \, dP$ for all $A \in \mathcal{B}$.

Proof. Define the measures $\tilde{P}[A] = P[A]$ and $P'[A] = \int_A X \, dP = E[X; A]$ on the probability space $(\Omega, \mathcal{B})$. Given a set $B \in \mathcal{B}$ with $\tilde{P}[B] = 0$, then $P'(B) = 0$ so that P' is absolutely continuous with respect to $\tilde{P}$. Radon-Nykodym's theorem gives a random variable $Y \in \mathcal{L}^1(\mathcal{B})$ with $P'[A] = \int_A X \, dP = \int_A Y \, dP$. □

Definition. The random variable Y in this theorem is denoted with $E[X|\mathcal{B}]$ and called the **conditional expectation of** X. The random variable $Y \in \mathcal{L}^1$ is unique in L^1. If Z is a random variable, then $E[X|Z]$ is defined as $E[X|\sigma(Z)]$. If $\{Z\}_{\mathcal{I}}$ is a family of random variables, then $E[X|\{Z\}_{\mathcal{I}}]$ is defined as $E[X|\sigma(\{Z\}_{\mathcal{I}})]$.

Example. If $\mathcal{B}$ is the trivial σ-algebra, then $E[X|\mathcal{B}] = E[X]$. If $\mathcal{B} = \mathcal{A}$, then $E[X|\mathcal{B}] = X$. If $\mathcal{B} = \{\emptyset, Y, Y^c, \Omega\}$ then $E[X|\mathcal{B}](\omega) = \int_Y X \; dP/m(Y)$ for $\omega \in Y$ and $E[X|\mathcal{B}](\omega) = \int_{Y^c} X \; dP/m(Y^c)$ for $\omega \in Y^c$.

The intuitive meaning: For an experiment, the possible outcomes are modeled by a probability space $(\Omega, \mathcal{A})$. The only information about the experiment are the events in a subalgebra $\mathcal{B}$ of $\mathcal{A}$. For example, $\mathcal{B}$ is generated by a set of random variables $\{Z_i\}_{i \in \mathcal{I}}$ obtained from some measure devices. A random variable X becomes the conditional expected random variable $E[X|\mathcal{B}]$ when having the information Z_i. This means that $E[X|\mathcal{B}](\omega)$ is the expected value of $X(\omega)$, then knowing $\mathcal{B}$.

Proposition 2.1.3 (Conditional expectation in $\mathcal{L}^2$ case) *The conditional expectation $X \mapsto E[X|\mathcal{B}]$ is the projection from $\mathcal{L}^2(\mathcal{A})$ to $\mathcal{L}^2(\mathcal{B})$.*

Proof. The space $\mathcal{L}^2(\mathcal{B})$ of square integrable $\mathcal{B}$-measurable functions is a closed subspace of $\mathcal{L}^2(\mathcal{A})$ since it is a Hilbert space with the induced topology. For any $X \in \mathcal{L}^2(\mathcal{A})$, there exists a unique projection $p(X) \in \mathcal{L}^2(\mathcal{B})$. The orthogonal complement $\mathcal{L}^2(\mathcal{B})^\perp$ is defined as

$$\mathcal{L}^2(\mathcal{B})^\perp = \{Z \in \mathcal{L}^2(\mathcal{A}) \mid (Z, Y) := E[Z, Y] = 0 \text{ for all } Y \in \mathcal{L}^2(\mathcal{B}) \} \; .$$

By the definition of the conditional expectation, we have for $A \in \mathcal{B}$

$$(X - E[X|\mathcal{B}], 1_A) = E[X - E[X|\mathcal{B}]; A] = 0 \; .$$

Therefore $X - E[X|\mathcal{B}] \in \mathcal{L}^2(\mathcal{B})^\perp$ and since the projection $p(X)$ and the conditional expectation are both unique, they coincide. □

Remark. This proposition means that Y is the least-squares-best $\mathcal{B}$ measurable square integrable predictor. This makes conditional expectation important for control of processes. If $\mathcal{B}$ is the σ-algebra describing the knowledge about a process (like the data a pilot knows about an aeroplane) and X is the random variable (the actual data of the flying aeroplane), we want to know, then $E[X|\mathcal{B}]$ is the best guess about this random variable we can make with our knowlege.

Theorem 2.1.4 (Properties of conditional expectation) *Given* $X, X_n, Y, \in \mathcal{L}$.

> **Linearity** *The map* $X \mapsto E[X|\mathcal{B}]$ *is linear.*
>
> **Positivity** $X \geq 0 \Rightarrow E[X|\mathcal{B}] \geq 0$.
>
> **Tower property** $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A} \Rightarrow E[E[X|\mathcal{B}]|\mathcal{C}] = E[X|\mathcal{C}]$.
>
> **Conditional Fatou** $|X_n| \leq X$, $E[\liminf_{n\to\infty} X_n|\mathcal{B}] \leq \liminf_{n\to\infty} E[X_n|\mathcal{B}]$.
>
> **Conditional dominated convergence** $|X_n| \leq X, X_n \to X$ *almost everywhere* $\Rightarrow E[X_n|\mathcal{B}] \to E[X|\mathcal{B}]$ *almost everywhere.*
>
> **Conditional Jensen** *Given* $c : \mathbf{R} \to \mathbf{R}$ *convex. Then* $E[c(X)|\mathcal{B}] \geq c(E[X|\mathcal{B}])$. *Especially* $||E[X|\mathcal{B}]||_p \leq ||X||_p$.
>
> **Taking knowledge out** $Z \in \mathcal{L}^\infty(\mathcal{B}) \Rightarrow E[ZX|\mathcal{B}] = ZE[X|\mathcal{B}]$.
>
> **Role of independence** *If* $\mathcal{C}$ *independent from* $\sigma(\sigma(X), \mathcal{B})$, *then* $E[X|\sigma(\mathcal{B}, \mathcal{C})] = E[X|\mathcal{B}]$. *Especially:* X *independent of* $\mathcal{C}$, *then* $E[X|\mathcal{C}] = E[X]$.

Proof. Linearity is clear. For positivity, note that if $Y = E[X|\mathcal{B}]$ would be negative on a set of postive measure, then $A = Y^{-1}([-1/n, 0]) \in \mathcal{B}$ would have positive probability for some n and leads to the contradiction $0 \leq E[1_A X] = E[1_A Y] \leq -n^{-1} m(A) < 0$.

The tower property follows from the fact that if $P'' \ll P' \ll P$ gives $P'' = Y'P' = Y'YP$ and $P'' \ll P$ gives $P'' = ZP$ so that $Z = Y'Y$ almost everywhere.

The tower property is especially useful when applied to the algebra $\mathcal{C}_Y = \{\emptyset, Y, Y^c, \Omega\}$. Because $X \leq Y$ *a.e.* if and only if $E[X|\mathcal{C}_Y] \leq E[Y|\mathcal{C}_Y]$ for all $Y \in \mathcal{B}$. The conditional versions of Fatou or domitated convergence are therefore true, if they are true conditioned with $\mathcal{C}_Y$ for each $Y \in \mathcal{B}$. The tower property reduces these statements to versions with $\mathcal{B} = \mathcal{C}_Y$ which are then on each of the sets Y, Y^c the usual theorems.

Conditional Jensen is proven directely: there exists a countable sequence $(a_n, b_n) \in \mathbf{R}^2$ such that $c(x) = \sup_n a_n x + b_n$ for all $x \in \mathbf{R}$. We get from $c(X) \geq a_n X + b_n$ that almost surely $E[c(X)|\mathcal{G}] \geq a_n E[X|\mathcal{G}] + b_n$. These inequalities hold therefore simultaneously for all n and we obtain almost surely

$$E[c(X)|\mathcal{G}] \geq \sup_n (a_n E[X|\mathcal{G}] + b_n) = c(E[X|\mathcal{G}]) .$$

The corollary is obtained with $c(x) = |x|^p$.

To show "taking knowledge out" it is enough to condition it with each algebra $\mathcal{C}_Y$ for $Y \in \mathcal{B}$. The tower property reduces these statements to linearity.

We come to the "role of independence": By linearity, we can assume $X \geq 0$. For $B \in \mathcal{B}$ and $C \in \mathcal{C}$, the random variables $1_B X$ and 1_C are independent so that

$$E[X 1_{B \cap C}] = E[X 1_B 1_C] = E[X 1_B] P[C] \, .$$

Every version of $Y = E[X|\mathcal{B}]$ is $\mathcal{B}$ measurable and since $Y 1_B$ is independent of $\mathcal{C}$ we get

$$E[(Y 1_B) 1_C] = E[Y 1_B] P[C]$$

so that $E[1_{B \cap C} X] = E[1_{B \cap C} Y]$. The measures on $\sigma(\mathcal{B}, \mathcal{C})$

$$\mu : A \mapsto E[1_A X], \nu : A \mapsto E[1_A Y]$$

agree therefore on the π-system of the form $B \cap C$ with $B \in \mathcal{B}$ and $C \in \mathcal{C}$ and therefore everywhere on $\sigma(\mathcal{B}, \mathcal{C})$. $\square$

Remarks.
1) From conditional Jensen, it follows that the operation of conditional expectation is contractive. It is therefore a linear, positive, continuous operation on $\mathcal{L}^p$.
2) Notice that conditional Fatou, Lebesgue and Jensen are statments about functions in $\mathcal{L}^1(\mathcal{B})$ and not about numbers like their usual theorems.

Question. Is there for almost all $\omega \in \Omega$ a probability measure P_ω such that

$$E[X|\mathcal{B}](\omega) = \int_\Omega X \, dP_\omega \ ?$$

If such a map exists which is $\mathcal{B}$ measurable, it is called a **regular conditional probability** given $\mathcal{B}$. In general such a map $\omega \mapsto P_\omega$ does not exist. However we state without proof

Theorem 2.1.5 (Existence of regular conditional probability) *Given a probability space $(\Omega, \mathcal{A}, P)$, where Ω is a complete seperable metric space and $\mathcal{A}$ is the Borel σ-algebra. For any sub σ-algebra $\mathcal{B}$ of $\mathcal{A}$, a regular conditional probability exists.*

We add a definition which is commonly used.

Definition. The **conditional probability** $P(\cdot|\mathcal{B})$ is a probability measure on $(\Omega, \mathcal{B})$ defined by

$$P(B \mid \mathcal{B}) = E[1_B|\mathcal{B}] \, .$$

2.2 Martingales

Definition. A sequence $\{\mathcal{A}_n\}_{n\in\mathbf{N}}$ of sub σ-algebras of $\mathcal{A}$ is called a **filtration**, if $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \dots \subset \mathcal{A}$. Given a filtration $\{\mathcal{A}_n\}_{n\in\mathbf{N}}$, one calls $(\Omega, \mathcal{A}, \{\mathcal{A}_n\}_{n\in\mathbf{N}}, P)$ a **filtered space**.

Definition. A sequence $X = \{X_n\}_{n\in\mathbf{N}}$ of random variables is also called a **(discrete stochastic) process**. We say, it is a $\mathcal{L}^p$-process if each X_n is in $\mathcal{L}^p$. A process is called **adapted to the filtration** $\{\mathcal{A}_n\}$ if X_n is $\mathcal{A}_n$-measurable for all $n \in \mathbf{N}$.

Definition. A $\mathcal{L}^1$-process which is adapted to a filtration $\{\mathcal{A}_n\}$ is called a **martingale** if

$$E[X_n|\mathcal{A}_{n-1}] = X_{n-1}$$

for all $n \geq 1$. It is called a **supermartingale** if $E[X_n|\mathcal{A}_{n-1}] \leq X_{n-1}$ and a **submartingale** if $E[X_n|\mathcal{A}_{n-1}] \geq X_{n-1}$. If we mean either submartingale or supermartingale (or martingale) we speak of a **semimartingale**.

Definition. If a semi-martingale X_n is given with respect to a filtered space defined by $\mathcal{A}_n = \sigma(Y_0, \dots, Y_n)$, where Y_n is also a given process, then we say X is a **martingale with respect** Y.

Remarks. If X is a supermartingale, then $-X$ is a submartingale and vice vera. A supermartingale, which is also a submartingale is a martingale. Since we can change X to $X - X_0$ without distroying any of the martingale properties, we could assume the process is **null at** 0 which means $X_0 = 0$. Given a martingale. From the tower property of conditional expectation follows that for $m < n$

$$E[X_n|\mathcal{A}_m] = E[E[X_n|\mathcal{A}_{n-1}]|\mathcal{A}_m] = E[X_{n-1}|\mathcal{A}_m] = \dots = E[X_m] \; .$$

Examples.

1) **Sum of independent random variables**
Let $X_i \in \mathcal{L}^1$ be a sequence of independent random variables with mean $E[X_i] = 0$. Define $S_n = \sum_{k=1}^n X_k$ and $\mathcal{A}_n = \sigma(X_1, \dots, X_n)$ with $\mathcal{A}_0 = \{\emptyset, \Omega\}$. Then S_n is a martingale since S_n is an $\{\mathcal{A}_n\}$-adapted $\mathcal{L}^1$-process and

$$E[S_n|\mathcal{A}_{n-1}] = E[S_{n-1}|\mathcal{A}_{n-1}] + E[X_n|\mathcal{A}_{n-1}] = S_{n-1} + E[X_n] = S_{n-1} \; .$$

We have used linearity, the independence property of the conditional expectation.

2) **Conditional expectation**
Given a random variable $X \in \mathcal{L}^1$ on a filtered space $(\Omega, \mathcal{A}, \{\mathcal{A}_n\}_{n\in\mathbf{N}}, P)$. Then $X_n = E[X|\mathcal{A}_n]$ is a martingale.
Especially: given a sequence Y_n of random variables. Then $\mathcal{A}_n = \sigma(Y_0, \dots, Y_n)$ is a filtered space and $X_n = E[X|Y_0, \dots, Y_n]$ is a martingale since by the tower property

$$E[X_n|Y_0, \dots, Y_{n-1}] = E[[X_n|Y_0, \dots, Y_n]|Y_0, \dots, Y_{n-1}] = E[X_n|Y_0, \dots, Y_{n-1}] = X_{n-1} \; .$$

We say X is a **martingale with respect to** Y. Note that since X_n is by definition $\sigma(Y_0, \dots, Y_n)$-measurable, there exist Borel measurable functions $h_n : \mathbf{R}^{n+1} \to \mathbf{R}$ such that $X_n = h_n(Y_0, \dots, Y_{n-1})$.

3) **Product of positive variables**
Given a sequence Y_n of independent random variables $Y_n \geq 0$ satisfying with $E[Y_n] = 1$. Define $X_0 = 1$ and $X_n = \prod_{i=0}^n Y_i$ and $\mathcal{A}_n = \sigma(Y_1, \dots, Y_n)$. Then X_n is a martingale. (The proof is an exercice. The martingale property does not follow from Example 1) by taking logarithms).

4) **Product of matrix-valued random variables**
Given a sequence of independent random variables Z_n with values in the group of invertible $N \times N$ matrices and let $\mathcal{A}_n = \sigma(Z_1, \dots, Z_n)$. Assume $E[\log ||Z_n||] \leq 0$. Define the real valued random variables $X_n = \log ||Z_1 \cdot Z_2 \cdots Z_n||$, where $\cdot$ denotes matrix multiplication. Since $X_n \leq \log ||Z_n|| + X_{n-1}$, we get

$$E[X_n|\mathcal{A}_{n-1}] \leq E[\log ||Z_n|| \mid \mathcal{A}_{n-1}] + E[X_{n-1}|\mathcal{A}_{n-1}] = E[\log ||Z_n||] + X_{n-1} \leq X_{n-1}$$

so that X_n is a supermartingale. Remark. In ergodic theory, such a process X_n is called subadditive.
We can also look at the real random variable $Y_n = ||Z_1 \cdot Z_2 \cdots Z_n||$. If $E[||Z_n||] = 1$, then Y_n is a supermartingale (see also Exercice).

5) **Polya's urn scheme**
An urn contains initially a red and a black ball. At each time $n = 1, \dots$ a ball is taken randomly, its color noted, and both this ball and another ball of the same color are placed back into the urn. Like this, after n draws, the urn contains $n+2$ balls. Define Y_n as the number of black balls after n moves and $X_n = Y_n/(n+2)$, the fraction of black balls. We claim that X is a martingale with respect to Y: Y_n takes values in $\{1, \dots, n+1\}$. Clearly $P(Y_{n+1} = k+1|Y_n = k) = k/(n+2)$ and $P(Y_{n+1} = k|Y_n = k) = 1 - k/(n+2)$. Therefore

$$\begin{aligned} E[X_{n+1}|Y_1, \dots, Y_n] &= \frac{1}{n+3} E[Y_{n+1}|Y_1, \dots, Y_n] \\ &= \frac{1}{n+3} P(Y_{n+1} = k+1|Y_n = k) \cdot (Y_n + 1) + P(Y_{n+1} = k|Y_n = k) \cdot Y_n \\ &= \frac{1}{n+3}[(Y_n + 1)\frac{Y_n}{n+2} + Y_n(1 - \frac{Y_n}{n+2})] = \frac{Y_n}{n+2} = X_n . \end{aligned}$$

Note that X_n is not independent of X_{n-1}. In the appendix, you see some experiments with Mathematica. The process "learns" in the sense that if there is more success (black balls), then the winning chances are better. The process is very sensitive of what happens in the first few steps.

6) **Branching processes**
Let Z_{ni} be IID integer valued random variables with positive finite mean m. Define $Y_0 = 1$ and

$$Y_{n+1} = \sum_{k=1}^{Y_n} Z_{nk}$$

with the convention that for $Y_n = 0$ also the sum is zero. We claim that $X_n = Y_n/m^n$ is a martingale with respect to Y. By the independence of Y_n and $Z_{ni}, i \geq 1$, we have

$$E[Y_{n+1}|Y_0, \dots, Y_n] = E[\sum_{k=1}^{Y_n} Z_{nk}|Y_0, \dots, Y_n] = E[\sum_{k=1}^{Y_n} Z_{nk}] = mY_n$$

so that

$$E[X_{n+1}|Y_0, \dots, Y_n] = E[Y_{n+1}|Y_0, \dots Y_n]/m^{n+1} = mY_n/m^{n+1} = X_n .$$

(The branching process is used to model for example population growth, desease epidemis or nuclear reactions. In the first case, think of Y_n as the size of a population at time n and with Z_{ni} the number of progenies of the $i-th$ member of the population.)

Proposition 2.2.1 *Let $\mathcal{A}_n$ be a fixed filtered sequence of σ-algebras. Linear combinations of martingales over $\mathcal{A}_n$ are again martingales. Submartingales and supermartingales form cones: if for example X, Y are submartingales and $a, b > 0$, then $aX + bY$ is a submartingale.*

Proof. Use the linearity and positivity of the conditional expectation. □

Proposition 2.2.2 *a) If X is a martingale and u is convex such that $u(X_n) \in \mathcal{L}^1$, then $Y = u(X)$ is a submartingale. Especially, if X is a martingale, then $|X|$ is a submartingale.*
b) If u is monotone and convex and X is a submartingale such that $u(X_n) \in \mathcal{L}^1$, then $u(X)$ is a submartingale.

Proof. a) We have by conditional Jensen

$$Y_n = u(X_n) = u(E[X_{n+1}|\mathcal{A}_n]) \leq E[u(X_{n+1})|\mathcal{A}_n] = E[Y_{n+1}|\,|\mathcal{A}_n] .$$

b) Use conditional Jensen and the monotonicity of u to get

$$Y_n = u(X_n) \leq u(E[X_{n+1}|\mathcal{A}_n]) \leq E[u(X_{n+1})|\mathcal{A}_n] = E[Y_{n+1}|\,|\mathcal{A}_n] .$$

□

Definition. A process $C = \{C_n\}_{n\geq 1}$ is called **previsible** if C_n is $\mathcal{A}_{n-1}$-measurable. A process X is called **bounded**, if $X_n \in \mathcal{L}^\infty$ and if there exists $K \in \mathbf{R}$ such that $||X_n||_\infty \leq K$.

Definition. Given a semimartingale X and a previsible process C, the process

$$(\int C\ dX)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

is called a **discrete stochastic integral** or **martingale transform**. One writes also

$$(\int C\ dX) = C \cdot X .$$

Proposition 2.2.3 (The system can't be beaten) *If C is a bounded nonnegative previsible process and X is a supermartingale then $\int C\ dX$ is a supermartingale. The same statement is true for submartingales and martingales.*

Proof. Let $Y = \int C\ dX$. From the "taking out property", we get

$$E[Y_n - Y_{n-1}|\mathcal{A}_{n-1}] = E[C_n(X_n - X_{n-1})|\mathcal{A}_{n-1}] = C_n \cdot E[X_n - X_{n-1}|\mathcal{A}_{n-1}] \leq 0$$

since C_n is nonnegative. □

Remarks.
If one wants to relax the boundedness of C, then one has to strengthen the condition for X. The proposition stays true, if both C and X are $\mathcal{L}^2$-processes.

Interpretation. If X_n represents your **capital** in a game, then $X_n - X_{n-1}$ is the **net winnings** per unit stake. If C_n is the **stake** on game n, then

$$\int C \, dX = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

are the **total winnings** up to time n. A martingale represents a **fair game** since $E[X_n - X_{n-1}|\mathcal{A}_{n-1}] = 0$, whereas a supermartingale is a game which is **unfavourable** to you. The above proposition tells that you can not find a strategy for putting your stake to make the game fair.

Etymology. The word "martingale" means a gambling system in which losing bets are doubled, a part of a horse's harness or a belt on the back of a man's coat.

Exercice 6

Topic: Conditional expectation, Martingales

Conditional expectation

1) Given a probability space $(\Omega, \mathcal{A}, P)$ and a σ-algebra $\mathcal{B} \subset \mathcal{A}$.
a) Show that the map $\phi : X \in \mathcal{L}^1 \mapsto E[X|\mathcal{B}]$ is a Markov operator from $\mathcal{L}^1(\mathcal{A}, P)$ to $\mathcal{L}^1(\mathcal{B}, Q)$, where Q is the conditional probability measure on $(\Omega, \mathcal{B})$.
b) The map T can also be viewed as a map on the new probability space $(\Omega, \mathcal{B}, Q)$, where Q is the conditional probability. Denote this new map by S. Show that S is again measure preserving and invertible. (S is called **a factor** of T).
c) Given a measure preserving invertible map $T : \Omega \to \Omega$. A complex number λ is called an **eigenvalue of** T, if there exists $X \in \mathcal{L}^2$ such that $X(T) = \lambda X$. The map T is said to have pure point spectrum, if there exists a countable set of eigenvalues λ_i such that their eigenfuctions X_i span $\mathcal{L}^2$. Show that if T has pure point spectrum, then also S has pure point spectrum.
d) (requires a bit preknowledge in ergodic theory and harmonic analysis on locally compact groups). It is known that if T has pure point spectrum, then the system is isomorphic to a translation on the compact abelian group G which is the dual group of the discrete group formed by the spectrum $\sigma(T) \subset \mathbf{T}$. Describe the possible factors of T and their spectra.

2) a) Compute $E[Y|Y]$.
b) Show that if $E[X|\mathcal{A}] = 0$ and $E[X|\mathcal{B}] = 0$, then $E[X|\sigma(\mathcal{A}, \mathcal{B})] = 0$.
c) Given $X, Y \in \mathcal{L}^1$ satisfying $E[X|Y] = Y$ and $E[Y|X] = X$. Show that $X = Y$ almost everwhere.

3) Given $\Omega = \mathbf{T}^1$, the one dimensional circle. Let $\mathcal{A}$ be the Borel σ-algebra on $\mathbf{T}^1 = \mathbf{R}/(2\pi\mathbf{Z})$ and $P = dx$ the Lebesgue measure. Given $k \in \mathbf{N}$, denote with $\mathcal{B}_k$ the σ-algebra consisting of all $A \in \mathcal{A}$ such that $A + \frac{n2\pi}{k} = A$ (mod 2π for all $1 \leq n \leq k$. Compute the conditional expectation $E[X|\mathcal{B}_k]$ for a random variable $X \in \mathcal{L}^1$.

Martingales

4) a) (Example 2 in the notes). Let $Y_1, Y_2, \dots$ be a sequence of independent non-negative random variables satisfying $E[Y_k] = 1$ for all $k \in \mathbf{N}$. Define $X_0 = 1, X_n = Y_1 \cdots Y_n$ and $\mathcal{A}_n = \sigma(Y_1, Y_2, \dots, Y_n)$. Show that X_n is a martingale.
b) (Example 4 in the notes). Let Z_n be a sequence of independent random variables taking values in the set of $n \times n$ matrices satisfying $E[||Z_n||] = 1$. Define $X_0 = 1, X_n = ||Z_1 \cdots Z_n||$. Show that X_n is a supermartingale.

2.3 Stopping times

Definition. A random variable T with values in $\overline{\mathbf{N}} = \mathbf{N} \cup \{\infty\}$ is called a **random time**. Define $\mathcal{A}_\infty = \sigma(\bigcup_{n\geq 0} \mathcal{A}_n)$. A random time T is called a **stopping time** (with respect to a filtration $\mathcal{A}_n$) if $\{T \leq n\} \in \mathcal{A}_n$ for all $n \in \overline{\mathbf{N}}$.

Remark. A random time T is a stopping time if and only if $\{T = n\} \in \mathcal{A}_n$ for all $n \in \overline{\mathbf{N}}$ since $\{T \leq n\} = \bigcup_{0 \leq k \leq n} \{T = k\} \in \mathcal{A}_n$.

Remark. We allow the value ∞ in the definition of a random time since we want allow also to not to stop the process with with positive probability.

Interpretation. Stopping times are random times, whose occurrence can be determined without preknowledge of the future. The term comes from gambling since the rule whereby a gambler ceases to play must be a stopping time. Whether or not you stop after the $n-$th game depends only on the history up to and including time n.

Examples.

1) **First entry time.**
Let X_n be a $\mathcal{A}_n$-adapted process and given a Borel set $B \in \mathcal{B}_{\mathbf{R}}$. Define

$$T(\omega) = \inf\{n \geq 0 \mid X_n(\omega) \in B\}$$

which is the time of first entry of X_n into B. The set $\{T = \infty\}$ is the set which never enters into B. Obviously

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{X_k \in B\} \in \mathcal{A}_n$$

so that T is a stopping time.

2) **Last exit time.**
Same setup as in 1). But this time

$$T(\omega) = \sup\{n \geq 0 \mid X_n(\omega) \in B\}$$

is **not** a stopping time since it is impossible to know that X will return to B after some time k without knowing the whole future.

Proposition 2.3.1 *Let T_1, T_2 be two stopping times. The infimumum $T_1 \wedge T_2$, the maximum $T_1 \vee T_2$ as well as the sum $T_1 + T_2$ are stopping times.*

Proof. Obvious from the definition since $\mathcal{A}_n$-measurable functions are closed under taking minima, maxima and sums. □

Definition. Given a stochastic process X_n-adapted to a filtration $\mathcal{A}_n$ and let T be a stopping time with respect to $\mathcal{A}_n$, then define the random variable

$$X_T(\omega) = \begin{cases} X_{T(\omega)}(\omega) & , T(\omega) < \infty \\ 0 & , \text{else} \end{cases}$$

or equivalently $X_T = \sum_{n=0}^{\infty} X_n 1_{\{T=n\}}$. The process $X_n^T = X_{T \wedge n}$ is called the **stopped process**. It is equal to X_T for times $T \leq n$ but equal to X_n if $T \geq n$.

Proposition 2.3.2 *If X is a supermartingale and T is a stopping time, then the stopped process X^T is a supermartingale. In particular $E[X^T] \leq E[X_0]$. The same statement is true if supermartingale is replaced by martingale in which case $E[X^T] = E[X_0]$.*

Proof. Define the "stake process" $C^{(T)}$ by $C_n^{(T)} = 1_{n \leq T}$ (think of it as betting 1 unit and quit playing immediately after time T). Define then the "winning process"

$$(\int C^{(T)} \, dX)_n = \sum_{k=1}^n C_k(X_k - X_{k-1}) = X_{T \wedge n} - X_0 \; .$$

or shortly $\int C^{(T)} \, dX = X^T - X_0$. The process C is previsible, since it can only take values 0 and 1 and $\{C_n^{(T)} = 0\} = \{T \leq n-1\} \in \mathcal{A}_{n-1}$. The claim follows from the "system can't be beaten" theorem. □

Remark. It is important that we take the stopped process X^T and not X_T:
Example. Take the simple random walk X on $\mathbf{Z}$ starting at 0. Let T be the stopping time $T = \inf\{n \mid X_n = 1\}$ (this is the Martingale strategy in Casino which gave the name of these processes). We will see that $P(T < \infty) = 1$ (the walk is recurrent). However

$$1 = E[X_T] \neq E[X_0] = 0 \; .$$

The above theorem gives $E[X^T] = E[X_0]$.
When can we say $E[X_T] = E[X_0]$? The answer gives Doob's optimal stopping theorem:

Theorem 2.3.3 (Doob's optimal stopping time theorem) *Let X be a supermartingale and T be a stopping time. Assume, one of the following conditions is true*

(i) T is bounded.
(ii) X is bounded and T is almost everywhere finite.
(iii) $T \in \mathcal{L}^1$ and $|X_n - X_{n-1}|$ is bounded.
(iv) $X_T \in \mathcal{L}^1$ and $\lim_{k \to \infty} E[X_k; \{T > k\}] = 0$.
(v) X is uniformly integrable and T is almost everywhere finite.

then $E[X_T] \leq E[X_0]$.
If X is a martingale and any of the conditions is true, then $E[X_T] = E[X_0]$.

Proof. We know that $E[X_{T \wedge n} - X_0] \leq 0$.
(i) take $n = \sup T(\omega)$.
(ii) use the dominated convergence theorem to get $\lim_{n \to \infty} E[X_{T \wedge n} - X_0] \leq 0$.
(iii), we have a bound $|X_n - X_{n-1}| \leq K$ and so

$$|X_{T \wedge n} - X_0| = |\sum_{k=1}^{T \wedge n} X_k - X_{k-1}| \leq KT$$

since $KT \in \mathcal{L}^1$, we get the result by the dominated convergence theorem.
(iv) By (i), we get $E[X_0] = E[X_{T \wedge k}] = E[X_T; \{T \leq k\}] + E[X_k; \{T > k\}]$ and taking

the limit gives $E[X_0] = \lim_{n\to\infty} E[X_k; \{T \le k\}] \to E[X_T]$ by the dominated convergence theorem and the assumption.
(v) The uniformly integrability $E[|X_n|; |X_n| > R] \to 0$ for $R \to \infty$ assures that $X_T \in \mathcal{L}^1$ since $E[|X_T|] \le k \cdot \max_{1\le i\le n} E[|X_k|] + \sup_n E[|X_n|; \{T > k\}] < \infty$. Since $|E[X_k; \{T > k\}]| \le \sup_n E[|X_n|; \{T > k\}] \to 0$, we can apply (iv).

If X is a martingale, we use the supermartingale case for both X and $-X$. □

Interpretation. A fair game cannot be made unfair by sampling it at bounded stopping times.

Corollary 2.3.4 *Assume X is a martingale and suppose $|X_n - X_{n-1}|$ is bounded. Given a previsible process C which is bounded and let $T \in \mathcal{L}^1$ be a stopping time, then $E[(\int_C dX)_T] = 0$.*

Proof. We know that $\int_C dX$ is a martingale and since $(\int_C dX)_0 = 0$, the claim follows from the optimal stopping time theorem part (iii). □

Martingales can be characterized involving stopping times.

Proposition 2.3.5 (Komatsu's lemma) *Let X be an $\mathcal{A}_n-$ adapted integrable sequence such that for every bounded stopping time T of $\mathcal{A}_n$*

$$E[X_T] = E[X_0] \ ,$$

then X is a martingale with respect to $\mathcal{A}_n$.

Proof. Fix $n \in \mathbf{N}$ and $A \in \mathcal{A}_n$. The map $T = n + 1 - 1_A$ is a stopping time since $\sigma(T) = \{\emptyset, A, A^c, \Omega\} \subset \mathcal{A}_n$. Apply $E[X_T] = E[X_0]$ and $E[X_{T'}] = E[X_0]$ for the bounded constant stopping time $T' = n + 1$ to get

$$E[X_n; A] + E[X_{n+1}; A^c] = E[X_T] = E[X_0] = E[X_{n+1}] = E[X_{n+1}; A] + E[X_{n+1}; A^c]$$

so that $E[X_{n+1}; A] = E[X_n; A]$. Since this is true, for any $A \in \mathcal{A}_n$, we know that $E[X_{n+1}|\mathcal{A}_n] = E[X_n|\mathcal{A}_n] = X_n$ and X is a martingale. □

Applications.

1) **The gambler's ruin problem.**
Let Y_i be IID with $P(Y_i = \pm 1) = 1/2$ and let $X_n = \sum_{k=1}^n Y_i$ be the random walk with $X_0 = 0$. We know that X is a martingale with respect to Y. Given $a, b > 0$, we define the stopping time

$$T = \min\{n | X_n = b, \text{ or } X_n = -a\} \ .$$

We want to compute $P(X_T = -a)$ and $P(X_T = b)$ in dependence of a, b.

Interpretation. Y_i denotes the outcomes of a series of fair gambles between two players A and B. The random variable X_n is the net change in the fortune of the gamblers after n independent games. If at the beginning, A has fortune a and B has fortune b, then $P(X_T = -a)$ is the **ruin probability** of A and $P(X_T = b)$ is the **ruin probability** of B.

Proposition 2.3.6 (Calculation of the ruin probability)

$$P(X_T = -a) = 1 - P(X_T = b) = b/(a+b) .$$

Proof. T is finite almost everywhere since by the law of the iterated logarithm gives $\limsup_n X_n/\Lambda_n = 1$ and $\liminf_n X_n/\Lambda_n = -1$. (We will give later a direct proof the finiteness of T, when we treat the random walk in more detail.) It follows that $P(X_T = -a) = 1 - P(X_T = b)$. We check that X_k satisfies condition (iv) in Doob's stopping time theorem: Since X_T takes values in $\{a, b\}$ it is in $\mathcal{L}^1$ and since on the set $\{T > k\}$ the value of X_k is in $(-a, b)$, we have $|E[X_k; \{T > k\}]| \leq \max\{a, b\} P(T > k) \to 0$. □

Remark. The boundedness of T is necessary in Doob's stopping time theorem. Let $T = \inf\{n \mid X_n = 1\}$. Then $E[X_T] = 1$ but $E[X_0] = 0]$ which shows that some condition on T or X has to be imposed. This fact leads to the "martingale" gambling strategy defined by doubling the bet when loosing. If the Casinos would not impose a bound on the possible inputs, this gambling strategy would lead to wins. But you have to go there with enough money! One can see it also like this, If you are A and the casino is B and $b = 1$, $a = \infty$ then $P(X_T = b) = 1$, which means that the Casino is ruined with probability 1.

2) **Wald's identity**.

Proposition 2.3.7 (Wald's identity) *Assume T is a stopping time of a $\mathcal{L}^1$-process Y with Y_i IID and $T \in \mathcal{L}^1$. The process $S_n = \sum_{k=1}^n Y_k$ satisfies*

$$E[S_T] = E[T] \cdot E[Y_1] .$$

Proof. The process $X_n = S_n - nE[Y_1]$ is a martingale satisfying condition (iii) in Doob's stopping time theorem. Therefore

$$0 = E[X_0] = E[X_T] = E[S_T - TE[Y_1]] .$$

□

2.4 Doob's convergence theorem

Definition. Given a stochastic process X and two real numbers $a < b$, we define the random variable

$U_n[a,b](\omega) = \max\{k \in \mathbf{N} \mid \exists\, 0 \le s_1 < t_1 < \ldots < s_k < t_k \le n, X_{s_i}(\omega) < a, X_{t_i}(\omega) > b, 1 \le i \le k\}$

called the **number of upcrossings** of $[a,b]$. Denote with $U_\infty[a,b]$ the limit

$$U_\infty[a,b] = \lim_{n\to\infty} U_n[a,b]$$

since $n \mapsto U_n[a,b]$ is monotone, this limit exists in $\mathbf{N} \cup \{\infty\}$.

Remember the notation $Z^- = -Z \vee 0$ denoting the negative part of Z.

Lemma 2.4.1 (Doob's upcrossing lemma) *If X is a supermartingale. Then*

$$(b-a)E[U_n[a,b]] \le E[(X_n - a)^-]$$

Proof. Define inductively the process $C_1 = 1_{\{X_0 < a\}}$ and for $n \ge 2$

$$C_n := 1_{\{C_{n-1}=1\}} 1_{\{X_{n-1} \le b\}} + 1_{\{C_{n-1}=0\}} 1_{\{X_{n-1} < a\}} \; .$$

It is previsible. Define the winning process $Y = \int_C dX$ which satisfies by definition $Y_0 = 0$. We have the **winning inequality**

$$Y_n(\omega) \ge (b-a)U_n[a,b](\omega) - (X_n(\omega) - a)^- \; .$$

See the picture: every upcrossing of $[a,b]$ increases the Y-value (the winning) by at least $(b-a)$, while $(X_n - a)^-$ is essentially the loss during the last interval of play.
Since C is previsible, bounded and nonnegative, we know that Y_n is also a supermartingale (see "the system can't be beaten") and we have therefore $E[Y_n] \le 0$. Taking expectation of the winning inequality, gives the claim. □

Interpretation. The proof uses the following strategy for putting your stakes C: wait until X gets below a. Play then unit stakes until X gets above b and stop playing. Wait again until X gets below a, etc.

Definition. We say, a stochastic process X_n is bounded in $\mathcal{L}^p$ if there exists $M \in \mathbf{R}$ such that $||X_n||_p \le M$ for all $n \in \mathbf{N}$.

Corollary 2.4.2 *If X is a supermartingale which is bounded in $\mathcal{L}^1$. Then $P(U_\infty[a,b] = \infty) = 0$.*

Proof. By the upcrossing lemma, we have for each $n \in \mathbf{N}$

$$(b-a)E[U_n[a,b]] \le |a| + E[|X_n|] \le |a| + \sup_n ||X_n||_1 < \infty \; .$$

By the dominated convergence theorem

$$(b-a)E[U_\infty[a,b]] < \infty \; ,$$

which gives the claim. □

Theorem 2.4.3 (Doobs convergence theorem) *Let X be a supermartingale which is bounded in $\mathcal{L}^1$. Then $X_\infty = \lim_{n\to\infty} X_n$ exists almost everywhere and is finite.*

Proof.

$$\begin{aligned}\Lambda &= \{X_n \text{ has no limit in } [-\infty,\infty]\} \\ &= \{\liminf X_n < \limsup X_n\} \\ &= \bigcup_{a<b,a,b\in\mathbf{Q}} \{\liminf X_n < a < b < \limsup X_n\} \\ &=: \bigcup_{a<b,a,b\in\mathbf{Q}} \Lambda_{a,b}\end{aligned}$$

Since $\Lambda_{a,b} \subset \{U_\infty[a,b] = \infty\}$ we have $P(\Lambda_{a,b}) = 0$ and therefore also $P(\Lambda) = 0$. Therefore $X_\infty = \lim X_n$ exists almost surely in $[-\infty,\infty]$. By Fatou's lemma

$$E[|X_\infty|] = E[\liminf_{n\to\infty} |X_n|) \leq \liminf_{n\to\infty} E[|X_n|] \leq \sup_n E[|X_n|] < \infty$$

so that $P(X_\infty < \infty) = 1$. □

Remark. Of course, we can replace supermartingale by submartingale or martingale in the theorem.

Example.
(Polya's urn scheme. See section Martingales for the definition).
Since the process Y giving the fraction of black balls is a martingale and bounded $0 \leq Y \leq 1$, we can apply the convergence theorem to see that there exists Y_∞ with $Y_n \to Y_\infty$.

Corollary 2.4.4 *If X is a non-negative supermartingale, then $X_\infty = \lim_{n\to\infty} X_n$ exists almost everywhere and is finite.*

Proof. Since the supermartingale property gives $E[|X_n|] = E[X_n] \leq E[X_0]$, the process X_n is bounded in $\mathcal{L}^1$. Apply Doob's convergence theorem. □

Remark. This corollary is also true for nonpositive submartingales or martingales, which are either nonnegative or nonpositive.

Example.
(Branching process.).
Remember that we had IID random variables Z_{ni} with positive finite mean m and defined $Y_0 = 0$, $Y_{n+1} = \sum_{k=1}^{Y_n} Z_{nk}$. We saw that the process $X_n = Y_n/m^n$ is non-negative and a martingale. According to the above corollary, the limit X_∞ exists almost everywhere. It is an interesting problem to find the distribution of X_∞:

Assume Z_{ni} have the generating function $f(\theta) = E[\theta^{Z_{ni}}]$.

(i) Y_n has the generating function $f^n(\theta) = f(f^{n-1})(\theta)$.
We prove this by induction. For $n = 1$ this is trivial. Using the independence of Z_{nk} we have

$$E[\theta^{Y_{n+1}}|Y_n = k] = f(\theta)^k$$

and so

$$E[\theta^{Y_{n+1}}|Y_n] = f(\theta)^{Z_n} .$$

By the tower property, this leads to

$$E[\theta^{Y_{n+1}}] = E[f(\theta)^{Z_n}] .$$

Write $\alpha = f(\theta)$ and use induction to simplify the right hand side to

$$E[f(\theta)^{Y_n}] = E[\alpha^{Y_n}] = f^n(\alpha) = f^n(f(\theta)) = f^{n+1}(\theta) .$$

(ii) In order to find the distribution of X_∞ we calculate instead the characteristic function

$$L(\lambda) = L(X_\infty)(\lambda) = E[\exp(i\lambda X_\infty)] .$$

Since $X_n \to X_\infty$ almost everywhere, we have $L(X_n)(\lambda) \to L(X_\infty)(\lambda)$. Since $X_n = Y_n/m^n$ and $E[\theta^{Y_n}] = f^n(\theta)$, we have

$$L(X_n)(\lambda) = f^n(e^{i\lambda/m^n})$$

so that L satisfies the **functional equation**

$$L(\lambda m) = f(L(\lambda)) .$$

Proposition 2.4.5 (Limit distribution of the branching process) *For the branching process defined by IID random variables Z_{ni} having the generating function f, the Fourier transform $L(\lambda) = E[\exp(i\lambda X_\infty)]$ of the distribution of the limit martingale X_∞ can be computed by solving the functional equation*

$$L(\lambda \cdot m) = f(L(\lambda)) .$$

Remark. If f has no analytic extension to the complex plane, we have to replace the Fourier transform with the Laplace transform $L(\lambda) = E[\exp(-\lambda X_\infty)]$.

Remark. Somehow related to Doob's convergence theorem for supermartingales is Kingman's subadditive ergodic theorem, which generalizes Birkhoff's ergodic theorem and which we state without proof. Neither of the two theorems are however corollaries of each other.

Definition. A sequence of random variables X_n is called *subadditive* with respect to a measure preserving transformation T, if $X_{m+n} \leq X_m + X_n(T^m)$ almost everywhere.

Theorem 2.4.6 (The subadditive ergodic theorem of Kingmann) *Given a sequence of random variables, which $X_n : X \to \mathbf{R} \cup \{-\infty\}$ with $X_n^+ := \max(0, X_n) \in L^1(X)$ and which is subadditive with respect to a measure preserving transformation T. Then there exists a T-invariant integrable measurable function $X : \Omega \to \mathbf{R} \cup \{-\infty\}$ such that $\frac{1}{n}X_n(x) \to X(x)$ for almost all $x \in X$. Furthermore $\frac{1}{n}E[X_n] \to E[X]$.*

If the condition of boundedness of the process in Doob's convergence theorem is strengthened a bit by assuming that X_n is uniformly integrable, then one can reverse in some sense the convergence theorem:

Theorem 2.4.7 (Doob's convergence theorem and a kind of converse) *A supermartingale X_n is uniformly integrable if and only if there exists X such that $X_n \to X$ in $\mathcal{L}^1$.*

Proof. If X_n is uniformly integrable, then X_n is bounded in $\mathcal{L}^1$ and Doob's convergence theorem gives $X_n \to X$ almost everywhere. But a uniformly integable family X_n which converges almost everywhere converges in $\mathcal{L}^1$.
On the other hand, a sequence $X_n \in \mathcal{L}^1$ converging to $X \in \mathcal{L}^1$ is uniformly integrable. □

Theorem 2.4.8 (Characterisation of uniformly integrable martingales) *An $\mathcal{A}_n$-adapted process is an uniformly integrable martingale if and only if $X_n \to X$ in $\mathcal{L}^1$ and $X_n = E[X|\mathcal{A}_n]$.*

Proof. By the last theorem, we know the "if"-part. $X_n = E[X|\mathcal{A}_n]$ is a martingale. What we have to show is that it is uniformly integrable.
Given $\epsilon > 0$. Choose $\delta > 0$ such that for all $A \in \mathcal{A}$, $P(A) < \delta$ implies $E[|X|; A] < \epsilon$. Choose further $K \in \mathbf{R}$ such that $K^{-1} \cdot E[|X|] < \delta$. By Jensen

$$|X_n| = |E[X|\mathcal{A}_n]| \leq E[|X||\mathcal{A}_n] \leq E[|X|] \, .$$

Therefore

$$K \cdot P[|X_n| > K] \leq E[|X_n|] \leq E[|X|] \leq \delta \cdot K$$

so that $P[|X_n| > K] < \delta$. Now, by definition of conditional expectation , $|X_n| \leq E[|X||\mathcal{A}_n]$ and $\{|X_n| > K\} \in \mathcal{A}_n$

$$E[|X_n|; |X_n| > K] \leq E[|X|; |X_n| > K] < \epsilon \, .$$

□

Summary of Doob's convergence theorem

A supermargtingale X_n which is either bounded in $\mathcal{L}^1$ or nonnegative or uniformly integrable converges almost everywhere.

Exercice 7

Topic: Martingales

Stopping times

1) Let S and T be stopping times satisfying $S \leq T$.

a) Show that the process

$$C_n(\omega) = 1_{\{S(\omega) < n \leq T(\omega)\}}$$

is previsible.

b) Show that for every supermartingale X, the following inequality holds (if again $S \leq T$ are stopping times):

$$E[X^T] \leq E[X^S] \,.$$

Doob's convergence theorem

2) Recall the set-up in Polya's urn process. Define Y_n as the number of black balls after n moves and $X_n = Y_n/(n+2)$, the fraction of black balls. We have seen that X is a martingale.

a) Prove that $P(Y_n = k) = 1/(n+1)$ for every $1 \leq k \leq n+1$.

b) Compute the distribution of the limit X_∞.

3) a) Which polynomials f can you realise as generating functions of a probability distribution? Call this class of polynomials $\mathcal{P}$.

b) Design a martingale X_n, where the iteration of polynomials $P \in \mathcal{P}$ plays a role.

c) Use one of the consequences of Doob's convergence theorem to show that the dynamics of every polynomial $P \in \mathcal{P}$ on the positive axis can be conjugated to a linear map $T : z \mapsto mz$: there exists a map L such that

$$L \circ T(z) = P \circ L(z)$$

for every $z \in \mathbf{R}^+$.

2.5 Example of a computation of the limiting density

Repetition. A **branching process** $Y_{n+1} = \sum_{k=1}^{Y_n} Z_{nk}$ defined by random variables Z_{nk} having generating function f and mean m defines a martingale $X_n = Y_n/m^n$ and the Laplace transform $L(\lambda) = E[e^{-\lambda X_\infty}]$ of the limit X_∞ satisfies the the functional equation

$$L(m\lambda) = f(L(\lambda)) .$$

We consider an explicit example, where one can compute explicitely the Fourier transform of the limiting distribution L.

Example. We assume that the IID random variables Z have the geometric distribution $P[Z = n] = p(1-p)^n$ with parameter $0 < p < 1$. (Note that this is a geometric distribution on $\mathbf{N} = \{0, 1 \dots\}$ and not on $\mathbf{N} \setminus \{0\}$ as it appeared earlier in this course.)

(i) The generating function of this distribution is the **Möbius transformation**

$$f(\theta) = \frac{p}{1 - q\theta} ,$$

where $q = 1 - p$.

Proof.

$$f(\theta) = E[\theta^Z] = \sum_{k=1}^{\infty} pq^k\theta^k = p/(1-q\theta) .$$

□

(ii) The mean of Z is q/p.

Proof. The calculation is done as in the chapter about distribution functions

$$E[Z] = \sum_{k=1}^{\infty} pq^k k = q/p .$$

(or deduced from what we had there $\sum_{n=1}^{\infty} pq^{k-1}k = 1/p$). □

(iii) The function $f^n(\theta)$ can be computed as

$$f^n(\theta) = \frac{pm^n(1-\theta) + q\theta - p}{qm^n(1-\theta) + q\theta - p} .$$

Proof. Since f is a Möbius transformation, the iterates can be computed by computing

$$A^n = \begin{pmatrix} 0 & p \\ -q & 1 \end{pmatrix}^n .$$

This is done by a diagonalisation of A:

$$A^n = (q-p)^{-1} \begin{pmatrix} 1 & p \\ 1 & q \end{pmatrix} \begin{pmatrix} p^n & 0 \\ 0 & q^n \end{pmatrix} \begin{pmatrix} q & -p \\ -1 & 1 \end{pmatrix} .$$

□

(iv)

$$L(\lambda) = \frac{p\lambda + q - p}{q\lambda + q - p}$$

Proof.

$$\begin{aligned} L(\lambda) &= E[e^{-\lambda X_\infty}] \\ &= \lim_{n\to\infty} E[e^{-\lambda Y_n/m^n}] \\ &= \lim_{n\to\infty} f_n(\exp(\lambda/m^n)) \\ &= \frac{p\lambda + q - p}{q\lambda + q - p} \end{aligned}$$

□

(v) If $m \leq 1$, then the law of X_∞ is a Dirac mass at 0. which means that the process dies out.

Proof. We see that in this case directly that $\lim_{n\to\infty} f_n(\theta) = 1$. □

In the case $m > 1$, the law of X_∞ has a point mass at 0 of weight $p/q = 1/m$ and an absolutely continuous part $(1/m - 1)^2 e^{(1/m-1)x}\, dx$.

Proof. A sucessful "look up" in a table of Laplace transforms leads to

$$L(\lambda) = p/qe^{-\lambda 0} + \int_0^\infty (1-p/q)^2 e^{(p/q-1)x} \cdot e^{-\lambda x}\, dx .$$

□

2.6 Extinction probability for the branching process

Define $u_n = P[Y_n = 0]$, the probability that the process dies out until time n. Since $u_n = f^n(0)$ and therefore $u_{n+1} = f(u_n)$.

The finite time extinction probabilities are evolving along an orbit of the map f.

The generating function $f(\theta) = E[\theta^Z] = \sum_n p_n \theta^n$ is analytic in $[0,1]$. It is nondecreasing and satisfies $f(1) = 1$. If we assume that $P[Z = 0] > 0$, then $f(0) > 0$ and there exists then a unique solution of $f(x) = x$ satisfying $f'(x) < 1$. The orbit $f^n(u)$ converges to this fixed point for every $u \in (0,1)$ and this fixed point is the extinction probability of the process. The value of $f'(0) = E[Z]$ decides whether there exists an attracting fixed point in the interval $(0,1)$ or not. We have proven:

For $E[Z] \geq 1$, the extinction probability of the whole process is the unique root of the equation $f(x) = x$ in $(0,1)$. For $E[Z] \leq 1$, the extinction probability is 1.

2.7 Lévy's upward and downward theorems

Lemma 2.7.1 *Given $X \in \mathcal{L}^1$. Then the class*

$$\{E[X|\mathcal{B}] \mid \mathcal{B} \subset \mathcal{A}, \mathcal{B} \text{ is } \sigma - algebra\}$$

is uniformly integrable.

Proof. Given $\epsilon > 0$. Choose $\delta > 0$ such that for all $A \in \mathcal{A}$, $P(A) < \delta$ implies $E[|X|; A] < \epsilon$. Choose further $K \in \mathbf{R}$ such that $K^{-1} \cdot E[|X|] < \delta$. By Jensen

$$|X_{\mathcal{B}}| = |E[X|\mathcal{B}]| \leq E[|X||\mathcal{B}] \leq E[|X|] .$$

Therefore

$$K \cdot P[|X_n| > K] \leq E[|X_{\mathcal{B}}|] \leq E[|X|] \leq \delta \cdot K$$

so that $P[|X_{\mathcal{B}}| > K] < \delta$. Now, by definition of conditional expectation , $|X_{\mathcal{B}}| \leq E[|X||\mathcal{B}]$ and $\{|X_{\mathcal{B}}| > K\} \in \mathcal{B}$

$$E[|X_{\mathcal{B}}|; |X_{\mathcal{B}}| > K] \leq E[|X|; |X_{\mathcal{B}}| > K] < \epsilon .$$

□

Notation. Denote with $\mathcal{A}_\infty$ the σ-algebra generated by $\bigcup_n \mathcal{A}_n$.

We stated at the end of the last paragraph a theorem. When extended a slight little bit, it has a name.

Theorem 2.7.2 (Lévy's upward theorem) *Given $X \in \mathcal{L}^1$. Then $X_n = E[X|\mathcal{A}_n]$ is a uniformly integrable martingale and X_n converges in $\mathcal{L}^1$ to $X_\infty = E[X|\mathcal{A}_\infty]$.*

Proof. X is clearly a martingale. The sequenc X_n is uniformly integrable by the above lemma. Therefore X_∞ exists almost everywhere by Doob's convergence theorem for uniformly integrable martingales, and since the family X_n is uniformly integrable, the convergence is in $\mathcal{L}^1$.
We have to show that $X_\infty = Y := E[X|\mathcal{A}_\infty]$.
By proving the claim for the positive and negative part, we can assume that $X \geq 0$ (and so $Y \geq 0$). Consider the two measures

$$Q_1(A) = E[X; A],\ Q_2(A) = E[X_\infty; A] .$$

Since $E[X_\infty|\mathcal{A}_n] = E[X|\mathcal{A}_n]$, we know that Q_1 and Q_2 agree on the π-system $\bigcup_n \mathcal{A}_n$. They agree therefore everywhere on $\mathcal{A}_\infty$. Define $A = \{E[X|\mathcal{A}_\infty] >$

$X_\infty\} \in \mathcal{A}_\infty$. Since $Q_1(A) - Q_2(A) = E[E[X|\mathcal{A}_\infty] - X_\infty]; A] = 0$ we have $E[X|\mathcal{A}_\infty] \leq X_\infty$ almost everwhere. Similarily also $X_\infty \leq [X|\mathcal{A}_\infty]$ almost evereywhere. □

Application. A Martingale proof of Kolmogorov's zero-one law:

> Given a sequence $\mathcal{A}_n$ of independent σ-algebras. The tail σ-algebra $\mathcal{T} = \bigcap_n \mathcal{B}_n$ with $\mathcal{B}_n = \bigcup_{m>n} \mathcal{A}_m$ is trivial.

Proof. Given $A \in \mathcal{T}$, define $X = 1_A \in \mathcal{L}^\infty(\mathcal{T})$ and the $\sigma-$ algebras $\mathcal{C}_n = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_n)$. By Lévy's upward theorem $X = E[X|\mathcal{C}_\infty] = \lim_{n\to\infty} E[X|\mathcal{C}_n]$. But since $\mathcal{C}_n$ is independent of $\mathcal{A}_n$ and X is $\mathcal{C}_n$ measureable, we have

$$P(A) = E[X] = E[X|\mathcal{C}_n] \to X$$

and since X takes only the valus 0 or 1 and $X = P(A)$ shows that it must be constant, we get $P(A) = 1$ or $P(A) = 0$. □

Definition. A sequence $\mathcal{A}_{-n}$ of $\sigma-$ algebras $\mathcal{A}_{-n}$ satisfying

$$\dots \subset \mathcal{A}_{-n} \subset \mathcal{A}_{-(n-1)} \subset \dots \subset \mathcal{A}_{-1}$$

is called a **downward filtation**. Define $\mathcal{A}_{-\infty} = \bigcap_n \mathcal{A}_{-n}$.

> **Theorem 2.7.3 (Lévy's downward theorem)** *Given a downward filtration $\mathcal{A}_{-n}$ and $X \in \mathcal{L}^1$. Define $X_{-n} = E[X|\mathcal{A}_{-n}]$. Then $X_{-\infty} = \lim_{n\to\infty} X_{-n}$ converges in $\mathcal{L}^1$ and $X_{-\infty} = E[X|\mathcal{A}_{-\infty}]$.*

Proof. Apply Doob's upcrossing lemma to the uniformly integrable martingale

$$X_k, -n \leq k \leq -1 \ :$$

for all $a < b$, the number of upcrossings is bounded

$$U_k[a,b] \leq (|a| + ||X||_1)/(b-a) \ .$$

This implies in the same way as in the proof of Doob's convergence theorem that $\lim_{n\to\infty} Y_{-n}$ converges almost everywhere.
We show now that $X_{-\infty} = E[X|\mathcal{A}_{-\infty}]$: given $A \in \mathcal{A}_{-\infty}$. We have $E[X; A] = E[X_{-n}; A] \to E[X_{-\infty}; A]$. The same argument as before shows that $X_{-\infty} = E[X; \mathcal{A}_{-\infty}]$. □

Application. A martingale proof of the strong law of large numbers.

Given $X_n \in \mathcal{L}^1$ which are IID and have mean m. Then $S_n/n \to m$ in $\mathcal{L}^1$.

Proof. Define the downward filtration $\mathcal{A}_{-n} = \sigma(S_n, S_{n+1}, \ldots)$.
Since $E[X_1|\mathcal{A}_{-n}] = E[X_i|\mathcal{A}_{-n}] = E[X_i|S_n, S_{n+1}, \ldots] = X_i$, and $E[X_1|\mathcal{A}_n] = S_n/n$. We can apply Lévy's downward theorem to see that S_n/n converges in $\mathcal{L}^1$. Since the limit L is in $\mathcal{T}$, it is by Kolmogorov's 0-1 law a constant c and $c = E[L] = \lim_{n\to\infty} E[S_n/n] = m$. ☐

2.8 Doob's decomposition of a stochastic process

Definition. A process X_n is **increasing**, if $P[X_n \leq X_{n+1}] = 1$.

Theorem 2.8.1 (Doob's decomposition) *Let X_n be an $\mathcal{A}_n$-adapted $\mathcal{L}^1$-process. Then*

$$X = X_0 + N + A$$

where N is a martingale null at 0 and A is a previsible process null at 0. This decomposition is unique in L^1.
X is a submartingale if and only if A is increasing.

Proof. If X has a Doob decomposition $X = X_0 + N + A$, then

$$E[X_n - X_{n-1}|\mathcal{A}_{n-1}] = E[N_n - N_{n-1}|\mathcal{A}_n] + E[A_n - A_{n-1}|\mathcal{A}_{n-1}] = A_n - A_{n-1}$$

which means that

$$A_n = \sum_{k=1}^{n} E[X_k - X_{k-1}|\mathcal{A}_{n-1}] \; .$$

If we **define** A like this, we get the required decomposition and the submartingale characterisation is also obvious. □

Remark. The corresponding result for continuous time processes is deeper and called **Doob-Meyer decomposition**.

We look now at $\mathcal{L}^2$-martingales $X_n \in \mathcal{L}^2$.

Lemma 2.8.2 *Given $s, t, u, v \in \mathbf{N}$ with $s \leq t \leq u \leq v$. If X_n is a $\mathcal{L}^2$-martingale, then*

$$(X_t - X_s, X_v - X_u) = 0$$

and

$$E[X_n^2] = E[X_0^2] + \sum_{k=1}^{n} E[(X_k - X_{k-1})^2] \; .$$

Proof. Since $E[X_v - X_u|\mathcal{A}_u] = X_u - X_u = 0$, we know that $X_v - X_u$ is orthogonal to $\mathcal{L}(\mathcal{A}_u)$. The first claim follows since $X_t - X_x \in \mathcal{L}^2(\mathcal{A}_u)$. The formula

$$X_n = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1})$$

expresses X_n as a sum of orthogonal terms and Pythagoras theorem yealds the second claim. □

Corollary 2.8.3 *A $\mathcal{L}^2$-martingale X is bounded in $\mathcal{L}^2$ if and only if $\sum_k E[(X_k - X_{k-1})^2] < \infty$.*

Proof.

$$E[X_n^2] = E[X_0^2] + \sum_{k=1}^{n} E[(X_k - X_{k-1})^2] \leq E[X_0^2] + \sum_k E[(X_k - X_{k-1})^2] < \infty \, .$$

If on the other hand, X_n is bounded in $\mathcal{L}^2$, then $||X_n||_2 \leq K < \infty$ and $\sum_k E[(X_k - X_{k-1})^2] \leq K + E[X_0^2]$. □

Theorem 2.8.4 (Doob's convergence theorem of $\mathcal{L}^2$-martingales) *Let X_n be a $\mathcal{L}^2$-martingale which is bounded in $\mathcal{L}^2$, then there exists $X \in \mathcal{L}^2$ such that $X_n \to X$ in $\mathcal{L}^2$.*

Proof. If X is bounded in $\mathcal{L}^2$, then, by monotonicity of the norm, it is bounded in $\mathcal{L}^1$ so that $X_n \to X$ almost everywhere for some X.
By Pythagoras, we have

$$E[(X - X_n)^2] \leq \sum_{k \geq n+1} E[(X_k - X_{k-1})^2] \to 0$$

so that $X_n \to X$ in $\mathcal{L}^2$. □

Definition. Let now X_n be a martingale in $\mathcal{L}^2$ which is null at 0. Conditional Jensen's inequality shows that X_n^2 is a submartingale. Doob's decomposition allows to write $X^2 = N + A$, where N is a martingale and A is a previsible increasing process. Define $A_\infty = \lim_{n\to\infty} A_n$ pointwise, where the limit can take values ∞ also. One writes also $\langle X \rangle$ for A.

Lemma 2.8.5 *Assume X is a $\mathcal{L}^2$-martingale. X is bounded in $\mathcal{L}^2$ if and only if $E[\langle X \rangle_\infty] < \infty$.*

Proof. From $X^2 = N + A$, we get $E[X_n^2] = E[A_n]$ since for a martingale N, the equality $E[N_n] = E[N_0]$ holds and N is null at 0. Therefore, X is bounded in L^2 if and only if $E[A_\infty] < \infty$ since $E[X_n^2] = E[A_n]$ and A_n is increasing. □

We can now relate the convergence of X_n to the finiteness of $A_\infty = \langle X \rangle_\infty$.

Theorem 2.8.6 *a) If $\lim_{n\to\infty} X_n(\omega)$ converges then $A_\infty(\omega) < \infty$.*
b) On the other hand, if $||X_n - X_{n-1}||_\infty \leq K$, then $A_\infty(\omega) < \infty$ implies the convergence of $\lim_{n\to\infty} X_n(\omega)$.

Proof. a) Because A is previsible, we can define for every k a stopping time $S(k) = \inf\{n \in \mathbf{N} \mid A_{n+1} > k\}$. The stopped process $A^{S(k)}$ is previsible because for $B \in \mathcal{B}_{\mathbf{R}}$ and $n \in \mathbf{N}$,

$$\{A_{n \wedge S(k)} \in B\} = A_1 \cup A_2$$

with

$$\begin{aligned} A_1 &= \bigcup_{i=0}^{n-1} \{S(k) = i; A_i \in B\} \in \mathcal{A}_{n-1} \\ A_2 &= \{A_n \in B\} \cap \{S(k) \leq n-1\}^c \in \mathcal{A}_{n-1} . \end{aligned}$$

Since

$$(X^{S(k)})^2 - A^{S_k} = (X^2 - A)^{S(k)}$$

is a martingale, we see that $\langle X^{S(k)} \rangle = A^{S(k)}$. The later process $A^{S(k)}$ is bounded by k so that by the above lemma $X^{S(k)}$ is bounded in $\mathcal{L}^2$. Therefore $\lim_{n \to \infty} X_{n \wedge S(k)}$ exists almost surely. Combining this with

$$\{A_\infty < \infty\} = \bigcup_k \{S_k = \infty\}$$

proves the claim.
b) Suppose the claim is wrong and that

$$P[A_\infty = \infty, \sup_n |X_n| < \infty] > 0 .$$

Then,

$$P[T(c) = \infty; A_\infty = \infty] > 0$$

where $T(c)$ is the stopping time

$$T(c) = \inf\{n \mid |X_n| > c\} .$$

Now

$$E(X^2_{T(c) \wedge n} - A_{T(c) \wedge n}) = 0$$

and $X^{T(c)}$ is bounded by $c + K$. Thus

$$E[A_{T(c) \wedge n}] \leq (c+K)^2$$

for all n. This is a contradition to $P[A_\infty = \infty, \sup_n |X_n| < \infty] > 0$. □

Proposition 2.8.7 (A strong law for martingales) *Let X be a $\mathcal{L}^2$-martingale zero at 0 and let $A = \langle X \rangle$. Then*

$$X_n / A_n \to 0$$

almost surely on $\{A_\infty = \infty\}$.

Proof.

(i) **Césaro's lemma**:

Given $0 = b_0 < b_1 \leq \dots, b_n \leq b_{n+1} \to \infty$ and a sequence $v_n \in \mathbf{R}$ which converges $v_n \to v_\infty$, then $\frac{1}{b_n}\sum_{k=1}^n (b_k - b_{k-1})v_k \to v_\infty$.

Proof. Let $\epsilon > 0$. Choose m such that $v_k > v_\infty - \epsilon$ if $k \geq m$. Then

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{b_n}\sum_{k=1}^n (b_k - b_{k-1})v_k &\geq \liminf_{n\to\infty} \frac{1}{b_n}\sum_{k=1}^m (b_k - b_{k-1})v_k + \frac{b_n - b_m}{b_n}(v_\infty - \epsilon) \\ &\geq 0 + v_\infty - \epsilon\end{aligned}$$

Since this is true for every $\epsilon > 0$, we have $\liminf \geq v_\infty$. By a similar argument $\limsup \geq v_\infty$. □

(ii) **Kronecker's lemma**:

Given $0 = b_0 < b_1 \leq \dots, b_n \leq b_{n+1} \to \infty$ and a sequence x_n of real numbers. Define $s_n = x_1 + \dots + x_n$. Then the convergence of $u_n = \sum_{k=1}^n x_k/b_k$ implies that $s_n/b_n \to 0$.

Proof. We have $u_n - u_{n-1} = x_n/b_n$ and

$$s_n = \sum_{k=1}^n b_k(u_k - u_{k-1}) = b_n u_n - \sum_{k=1}^n (b_k - b_{k-1})u_{k-1} \; .$$

Césaro's lemma (i) implies that s_n/b_n converges to $u_\infty - u_\infty = 0$. □

(iii) Claim. Since A is increasing and null at 0, we have $A_n > 0$ and $1/(1+A_n)$ is bounded. Since A is previsible, also $1/(1+A_n)$ is previsible, we can define the martingale

$$W_n = (\int (1+A)^{-1} dX)_n = \sum_{k=1}^n \frac{X_k - X_{k-1}}{1 + A_k} \; .$$

Moreover, since $(1 + A_n)$ is $\mathcal{A}_{n-1}$-measureable, we have

$$E[(W_n - W_{n-1})^2 | \mathcal{A}_{n-1}] = (1+A_n)^{-2}(A_n - A_{n-1}) \leq (1+A_{n-1})^{-1} - (1+A_n)^{-1}$$

almost surely. This implies that $\langle W \rangle_\infty \leq 1$ so that $\lim_{n\to\infty} W_n$ exists almost surely. Kronecker's lemma (ii) applied pointwise implies that on $\{A_\infty = \infty\}$

$$\lim_{n\to\infty} X_n/(1+A_n) = \lim_{n\to\infty} X_n/A_n \to 0 \; .$$

□

2.9 Doob's submartingale inequality

Theorem 2.9.1 (Doob's submartingale inequality) *Let X be a non-negative submartingale. For any $\epsilon > 0$*

$$\epsilon \cdot P[\sup_{1 \leq k \leq n} X_k \geq \epsilon] \leq E[X_n; \{\sup_{1 \leq k \leq n} X_k \geq \epsilon\}] \leq E[X_n] .$$

Proof. The set $A = \{\sup_{1 \leq k \leq n} X_k \geq \epsilon\}$ is a disjoint union of the sets

$$\begin{aligned} A_0 &= \{X_0 \geq \epsilon\} \in \mathcal{A}_0 \\ A_k &= \{X_k \geq \epsilon\} \cap (\bigcup_{i=0}^{n} A_i^c) \in \mathcal{A}_k . \end{aligned}$$

Since X is a submartingale, and $X_k \geq \epsilon$ on A_k we have for $k \leq n$

$$E[X_n; A_k] \geq E[X_k; A_k] \geq \epsilon P[A_k] .$$

Summing up from $k = 0$ to n gives the result. □

Theorem 2.9.2 (Kolmogorov's inequality) *Given $X_n \in \mathcal{L}^2$ IID with $E[X_i] = 0$ and $S_n = \sum_{k=1}^n X_k$. Then for $\epsilon > 0$,*

$$P[\sup_{1 \leq k \leq n} |S_k| \geq \epsilon] \leq \mathrm{Var}[S_n]/\epsilon^2 .$$

Proof. S_n is a martingale with respect to $\mathcal{A}_n = \sigma(X_1, X_2, \dots, X_n)$. Since $u(x) = x^2$ is convex, then S_n^2 is a submartingale. Apply the submartingale inequality. □

Notation. For typographical reasons we will use in the following the notation $\Lambda(n) = \Lambda_n = \sqrt{2n \log\log(n)}$ or $S(n) = S_n = \sum_{k=1}^n S_k$. We use also for

$$1 - \Phi(x) = \int_x^\infty \phi(y)\, dy = \int_x^\infty (2\pi)^{-1/2} \exp(-y^2/2)\, dy$$

the elementary estimates

$$(x + x^{-1})^{-1} \phi(x) \leq 1 - \Phi(x) \leq x^{-1} \phi(x) .$$

Theorem 2.9.3 (Special case of law of iterated logarithm) *Given X_n IID with standard normal distribution $N(0,1)$. Then* $\limsup_{n \to \infty} S_n / \Lambda(n) = 1$.

Proof. (i) S_n is a martingale relative to $\mathcal{A}_n = \sigma(X_1, \dots, X_n)$. The function $x \mapsto e^{\theta x}$ is convex on $\mathbf{R}$ so that $e^{\theta S_n}$ is a submartingale. The submartingale inequality gives

$$P[\sup_{1 \le k \le n} S_k > \epsilon] = P[\sup_{1 \le k \le n} e^{\theta S_k} > e^{\theta \epsilon}] \le e^{-\theta \epsilon} E[e^{\theta S_n}] = e^{-\theta \epsilon} e^{\theta^2 \cdot n/2} .$$

For given $\epsilon > 0$, we get the best estimate for $\theta = \epsilon/n$ and obtain

$$P[\sup_{1 \le k \le n} S_k > \epsilon] \le e^{-\epsilon^2/(2n)} .$$

(ii) Given $K > 1$ (close to 1). Choose $\epsilon_n = K\Lambda(K^{n-1})$. The last inequality in (i) gives

$$P[\sup_{1 \le k \le K^n} S_k \ge \epsilon_n] \le \exp(-\epsilon_n^2/(2K^n)) = (n-1)^{-K} (\log K)^{-K} .$$

Borel Cantelli assures that for large enough n and $K^{n-1} \le k \le K^n$

$$S_k \le \sup_{1 \le k \le K^n} S_k \le \epsilon_n = K\Lambda(K^{n-1}) \le K\Lambda(k)$$

which means for $K > 1$ almost surely

$$\limsup_{k \to \infty} \Lambda(k)^{-1} S_k \le K .$$

By taking a sequence of K's converging down to 1, we obtain almost surely

$$\limsup_{k \to \infty} \Lambda(k)^{-1} S_k \le 1 .$$

(iii) Given $N > 1$ (large) and $\delta > 0$ (small). Define the independent sets

$$A_n = \{S(N^{n+1}) - S(N^n) > (1-\delta)\Lambda(N^{n+1} - N^n)\} .$$

Then

$$P[A_n] = 1 - \Phi(y) = (2\pi)^{-1/2} (y + y^{-1})^{-1} e^{-y^2/2}$$

with $y = (1-\delta)(2 \log\log(N^{n-1} - N^n))^{1/2}$. Since $P[A_n]$ is roughly (up to logarithmic terms) $(n \log N)^{-(1-\delta)^2}$, we have $\sum_n P[A_n] = \infty$. Borel Cantelly shows that $P[\limsup_n A_n] = 1$ so that

$$S(N^{n+1}) > (1-\delta)\Lambda(N^{n+1} - N^n) + S(N^n) .$$

By (ii), $S(N^n) > -2\Lambda(N^n)$ for large n so that for infinitely many n, we have

$$S(N^{n+1}) > (1-\delta)\Lambda(N^{n+1} - N^n) - 2\Lambda(N^n) .$$

It follows that

$$\limsup_n \frac{S_n}{\Lambda_n} \ge \limsup_n \frac{S(N^{n+1})}{\Lambda(N^{n+1})} \ge (1-\delta)(1 - N^{-1})^{1/2} - 2N^{1/2} .$$

$\square$

2.10 Doob's $\mathcal{L}^p$ inequality

Lemma 2.10.1 (Corollary of Hölder inequality) *Fix $p > 1$ and q satisfying $p^{-1} + q^{-1} = 1$. Given $X, Y \in \mathcal{L}^p$ satisfying*

$$\epsilon P[|X| \geq \epsilon] \leq E[|Y|; |X| \geq \epsilon], \forall \epsilon > 0 .$$

Then $||X||_p \leq q \cdot ||Y||_p$.

Proof. Integrating the assumption multiplied with $p\epsilon^{p-2}$ gives

$$L = \int_0^\infty p\epsilon^{p-1} P[|X| \geq \epsilon] \, d\epsilon \leq \int_0^\infty p\epsilon^{p-2} E[|Y|; |X| \geq \epsilon] \, d\epsilon =: R .$$

But Fubini gives that the left hand side is

$$L = \int_0^\infty E[p\epsilon^{p-1} 1_{\{|X| \geq \epsilon\}}] \, d\epsilon = E[\int_0^\infty p\epsilon^{p-1} 1_{\{|X| \geq \epsilon\}} d\epsilon] = E[|X|^p]$$

and in the same way that the right hand side $R = E[q \cdot |X|^{p-1}|Y|]$. With Hölder's inequality, we get

$$E[|X|^p] \leq E[q|X|^{p-1}|Y|] \leq q||Y||_p \cdot |||X|^{p-1}||_q .$$

Since $(p-1)q = p$, we can replace $|||X|^{p-1}||_q = E[|X|^p]^{1/q}$ on the right hand side which gives the claim. □

Theorem 2.10.2 (Doob's $\mathcal{L}^p$ inequality) *Given a non-negative submartingale X which is bounded in $\mathcal{L}^p$. Then $X^* = \sup_n X_k$ is in $\mathcal{L}^p$ and satisfies*

$$||X^*|| \leq q \cdot \sup_n ||X_n||_p .$$

Proof. Define $X_n^* = \sup_{1 \leq k \leq n} X_k$ for $n \in \mathbf{N}$. From Doob's submartingale inequality and the above lemma, we see that

$$||X_n^*||_p \leq q||X_n||_p \leq q \sup_n ||X_n||_p .$$

□

Corollary 2.10.3 *Given a non-negative submartingale X which is bounded in $\mathcal{L}^p$. Then $X_\infty = \lim_{n\to\infty} X_n$ exists in $\mathcal{L}_p$ and $||X_\infty||_p = \sup_n ||X_n||_p$.*

Proof. The submartingale X is dominated by the element X^* in the $\mathcal{L}^p$-inequality. The supermartingale $-X$ is bounded in $\mathcal{L}^p$ and so bounded in $\mathcal{L}^1$. We know therefore that $X_\infty = \lim_{n\to\infty} X_n$ exists almost everywhere. Since $|X_n - X_\infty|^p \leq (2X^*)^p \in \mathcal{L}^p$, the dominated convergence theorem shows that $X_n \to X_\infty$ in $\mathcal{L}^p$. $\square$

Corollary 2.10.4 *Given a martingale Y bounded in $\mathcal{L}^p$ and call $X = |Y|$. Then*

$$X_\infty = \lim_{n\to\infty} X_n$$

exists in $\mathcal{L}_p$ and $||X_\infty||_p = \sup_n ||X_n||_p$.

Proof. Use the above corollary for the submartingale $X = |Y|$. $\square$

Theorem 2.10.5 (Kakutani's theorem) *Let X_n be a non-negative independent $\mathcal{L}^1$ process with $E[X_n] = 1$. Define $S_0 = 1$ and $S_n = \prod_{k=1}^n X_k$. Then $S_\infty = \lim_n S_n$ exists since S_n is a non-negative $\mathcal{L}^1$ martingale. Then S_n is uniformly integrable if and only if $\prod_n E[X_n^{1/2}] > 0$.*

Proof. Define $a_n = E[X_n^{1/2}]$. The process

$$T_n = \frac{X_1^{1/2}}{a_1} \frac{X_2^{1/2}}{a_2} \cdots \frac{X_n^{1/2}}{a_n}$$

is a martingale. We have $E[T_n^2] = (a_1 a_2, \cdots a_n)^{-2} \leq (\prod_n a_n)^{-1} < \infty$ so that T is bounded in $\mathcal{L}^2$, By Doob's $\mathcal{L}^2$-inequality

$$E[\sup_n |S_n|] \leq E[\sup_n |T_n|^2] \leq 4 \sup_n E[|T_n|^2] < \infty$$

so that S is dominated by $S^* = \sup_n |S_n| \in \mathcal{L}^1$. This implies that S is uniformly integrable.

If S_n is uniformly integrable, then $S_n \to S_\infty$ in $\mathcal{L}^1$. We have to show that $\prod_n a_n > 0$. Aiming to a contradiction we assume that $\prod_n a_n = 0$. The martinagale T defined above is a nonnegative martingale which has a limit T_∞. But since $\prod_n a_n = 0$ we must then have that $S_\infty = 0$ and so $S_n \to 0$ in $\mathcal{L}^1$. This is not possible since $E[S_n] = 1$ by the independence of the X_n. $\square$

2.11 Random walks

Question. Consider the d-dimensional lattice $\mathbf{Z}^d$, where each point has $2d$ neighbors. A particle starts at the origin $0 \in \mathbf{Z}^d$ and makes in each time step a random step into one of the $2d$ directions. What is the probability that the particle returns back to the origin?

Setup. Define a sequence of IID random variables X_n which take values in

$$I = \{e \in \mathbf{Z}^d \mid |e| = \sum_{i=1}^d |e_i| = 1\}$$

and having the uniform distribution defined by $P[X_n = e] = (2d)^{-1}$ for all $e \in I$. The random variable $S_n = \sum_{i=1}^n X_i$ (with $S_0 = 0$) describes the position of the particle at time n. As a probability space, we can take $\Omega = I^{\mathbf{N}}$ with product measure $\nu^{\mathbf{N}}$, where ν is the measure on E, which assigns to each point e the probability $\nu(\{e\}) = (2d)^{-1}$. The random variables X_n are then defined by $X_n(\omega) = \omega_n$.

Define the sets $A_n = \{X_n = 0\}$ and the random variables

$$Y_n = 1_{A_n}$$

which tell, whether the particle has returned to position $0 \in \mathbf{Z}^d$ at time n. The sum $B_n = \sum_{i=0}^n Y_n$ counts the number of $0-$ visits of the particle up to time n and $B = \sum_{i=0}^\infty Y_n$ counts the total number of visits at the origin. The expectation

$$E[B] = \sum_{n=0}^\infty P[S_n = 0]$$

tells how many times a particle is expected to return to the origin. We write $E[B] = \infty$, if the sum diverges. We will see that in that case, the particle is returning back infinitely many times.

Theorem 2.11.1 (Polya) *$E[B] = \infty$ for $d = 1, 2$ and $E[B] < \infty$ for $d > 2$.*

Proof. Fix $n \in \mathbf{N}$ and define $a^{(n)}(k) = P[S_n = k]$ for $k \in \mathbf{Z}^d$. Since the particle can reach in time n only a bounded region, the function $a^{(n)} : \mathbf{Z}^d \to \mathbf{R}$ is vanishing outside a bounded set. We can therefore define its Fourier transform

$$\phi_{S_n} = \sum_{k \in \mathbf{Z}^d} a^{(n)}(k) e^{2\pi i k \cdot x}$$

which is smooth function on $\mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$ and is the characteristic function of S_n since

$$E[e^{ixS_n}] = \sum_k P[S_n = k] e^{ixk} \ .$$

(i) The characteristic function ϕ_X of X_k is

$$\phi_X(x) = \frac{1}{2d} \sum_{|j|=1} e^{2\pi i x_j} = \frac{1}{d} \sum_{i=1}^{d} \cos(2\pi x_i) .$$

Because the S_n is a sum of n independent random variables X_k

$$\phi_{S_n} = \phi_{X_k}^n = \frac{1}{d^n} (\sum_{i=1}^{d} \cos(2\pi x_i))^n .$$

Note that $\phi_{S_n}(0) = P[S_n = 0]$.

(ii) $E[B] = \sum_{n \geq 0} \phi_{S_n}(0)$ is finite if and only if $d < 3$. The Fourier inversion formula gives

$$\sum_n \phi_{S_n}(0) = \int_{\mathbf{T}^d} \sum_n \phi_X^n(x) \; dx = \int_{\mathbf{T}^d} \frac{1}{1 - \phi_X(x)} \; dx .$$

Since

$$1/2 \cdot \frac{(2\pi)^2}{2d} |x|^2 \leq 1 - \phi_X(x) \leq 2 \cdot \frac{(2\pi)^2}{2d} |x|^2 ,$$

(make a Taylor expansion $\phi_{X_k}(x) = 1 - \sum_j x_j^2/2(2\pi)^2 + \dots$) we get the claim of the theorem since $\int_{\{|x|<\epsilon\}} \frac{1}{|x|^2} \; dx$ is finite if and only if $d \geq 3$. □

> **Corollary 2.11.2** *The particle returns to the origin infinitely often almost surely if and only if $d \leq 2$. For $d \geq 3$, almost surely, the particle returns finitely many times to zero and $P[\lim_{n\to\infty} |S_n| = \infty] = 1$.*

Proof. (i) $d > 2$: define the event $A_n = \{S_n = 0\}$. Then $A_\infty = \limsup_n A_n$ is the subset of Ω, for which the particles returns to 0 infinitely many times. Since $E[B] = \sum_{n=0}^\infty P[A_n]$, Borel- Cantelli' lemma gives $P[A_\infty] = 0$ for $d > 2$. The particle returns therefore back to 0 only finitely many times and in the same way it visits each lattice point only finitely many times. This means that the particle eventually leaves every bounded set and converges to infinity.
(ii) $d \leq 2$: let p be the probability that the random walk returns to 0:

$$p = P[\bigcup_n A_n] .$$

Then p^{m-1} is the probability that there are at least m visits in 0 and the probability is $p^{m-1} - p^m = p^{m-1}(1-p)$ that there are exactly m visits. We can write

$$E[B] = \sum_{m \geq 1} m p^{m-1} (1-p) = 1/(1-p) .$$

Since $E[B] = \infty$, we know that $p = 1$. □

The use of characteristic functions allows to do combinatorics and to count the number of closed paths starting at zero in the graph.

Proposition 2.11.3 *There are* $\int_{\mathbf{T}^d} (\sum_{k=1}^d \cos(2\pi i x_k))^n$ *closed paths in* $\mathbf{Z}^d$ *of length* n*.*

Proof. If we know the probability $P(S_n = 0)$ that a path returns to 0 in n step, then $(2d)^n P(S_n = 0)$ is the number of closed paths in $\mathbf{Z}^d$ of length n. But $P(S_n = 0)$ is the zero'th Fouriercoefficient

$$\int_{\mathbf{T}^d} \frac{1}{d^n} (\sum_{k=1}^d \cos(2\pi i x_k))^n$$

of ϕ_{S_n}. □

For example: in the case $d = 1$, we have

$$\int_0^1 2^{2n} \cos^{2n}(2\pi x) \, dx = \binom{2n}{n}$$

closed paths of length $2n$ starting at 0.

Generalisations of symmetric random walks on $\mathbf{Z}^d$.

1) **More general graphs**. The lattice defined by $\mathbf{Z}^d$ by an arbitrary graph G which is regular (each vertex has the same number of neighbors). A convenient way is to take as the graph the Cayley graph of a discrete group G with generators $a_1, \dots a_d$.

Corollary 2.11.4 *If* G *is the Cayley graph of an abelian group* $\mathcal{G}$ *with generators* $a_1, \dots a_d$*, then the random walk on* G *is recurrent if and only maximal two of the generators have infinite order.*

Proof. The group has a presentation

$$\mathcal{G} = < a_1, \dots, a_d | [a_i, a_j] = 0, a_i^{n_i} = 0 > ,$$

where $n_i \in \mathbf{N} \cup \{\infty\}$ are the smallest positive numbers satisfying $a_i^{n_i} = e$ and $\mathcal{G}$ is isomorphic to $\prod_{i=1}^d \mathbf{Z}_{n_i}$. The characteristic function of X_n is a function on $\hat{\mathcal{G}}$.

$$\phi_{S_n}(0) = \int_{\hat{\mathcal{G}}} \phi_X^n(x) \, dx = \int_{\mathcal{G}^d} \frac{1}{1 - \phi_X(x)} \, dx$$

is finite if and only if $\mathcal{G}^d$ contains a three dimensional torus. □

The recurrence properties on nonabelian groups is more subtle, since characteristic functions loose then some of their good properties.

2) **Allow drift**. An other generalisation is to add a drift by changing the probability distribution ν on I. Given $p_j \in (0,1)$ with $\sum_{|j|=1} p_j = 1$. In this case

$$\phi_X(x) = \sum_{|j|=1} p_j e^{2\pi i x_j} .$$

We have recurrence if and only if

$$\int_{\mathbf{T}^d} \frac{1}{1-\phi_X(x)}\, dx = \infty .$$

Take for example the case $d = 1$ with drift parametrized by $p \in (0,1)$. Then

$$\phi_X(x) = pe^{2\pi i x} + (1-p)e^{-2\pi i x} = \cos(2\pi x) + i(2p-1)\sin(2\pi x) .$$

which shows that $\int_{\mathbf{T}^d} \frac{1}{1-\phi_X(x)}\, dx = \infty$ if $p \neq 1/2$.

3) **Relax IID**. An other generalisation of the random walk is to take identically distributed random variables X_n with values in I, which need not to be independent. An example which appears in number theory in the case $d = 1$: Take the probability space $\Omega = \mathbf{T}^1 = \mathbf{R}/\mathbf{Z}$, an irrational number α and a function f which takes each value in I on an interval $[k/2d, (k+1)/2d)$. Define then the random variable $X_n(\omega) = f(\omega + n\alpha)$.

In this case, the tool of characteristic functions is no more useful since we have no more $\phi_{S_n} = (\phi_X)^n$.

4) **Allow more general jumps**. Take any finite set $F \subset \mathbf{Z}^d$ with a probability distribution p_i discribing the probability distribution of X_n. We get

$$\phi_X(x) = \sum_{j \in F} p_j e^{2\pi i x_j} .$$

Again, we can decide about the recurrence by deciding whether $\int_{\mathbf{T}^d} \frac{1}{1-\phi_X(x)}\, dx$ is infinite.

2.12 The arc-sin law for the 1D random walk

Notation. Let X_n denote the $\{-1, 1\}$-valued random variable with $P(S_n = \pm 1) = 1/2$ and let $S_n = \sum_{k=1}^n X_k$ be the random walk. We have seen that it is a martingale with respect to X_n.
Given $a \in \mathbf{Z}$, we define the stopping time

$$T_a = \min\{n \in \mathbf{N} \mid S_n = a\} .$$

Lemma 2.12.1 (Reflexion principle) *For $a, b \in \mathbf{N} \setminus \{0\}$, one has*

$$P[a + S_n = b \mid T_{-a} \leq n] = P[S_n = a + b] .$$

Proof. The number of paths from a to b passing zero is equal to the number of paths from $-a$ to b which is the number of paths from zero to $a + b$. □

The reflection principle allows to compute the distribution of the random variable T_{-a}:

Theorem 2.12.2 (Distribution of stopping time (=ruin time))

a) $P[T_{-a} \leq n] = P[S_n \leq -a] + P[S_n > a]$.

b) $P[T_{-a} = n] = \frac{a}{n} P[S_n = a]$.

Proof. a) We use the reflection principle

$$\begin{aligned} P[T_{-a} \leq n] &= \sum_{b \in \mathbf{Z}} P[T_{-a} \leq n | a + S_n = b] \\ &= \sum_{b \leq 0} P[a + S_n = b] + \sum_{b > 0} P[S_n = a + b] \\ &= P[S_n \leq -a] + P[S_n > a] \end{aligned}$$

b) We have

$$P[S_n = a] = \left(\begin{array}{c} n \\ \frac{a+n}{2} \end{array} \right)$$

from which we get

$$\frac{a}{n} P[S_n = a] = \frac{1}{2}(P[S_{n-1} = a - 1] - P[S_{n-1} = a + 1]) .$$

We have also

$$
\begin{aligned}
P[S_n > a] - P[S_{n-1} > a] &= P[S_n > a \mid S_{n-1} \le a] + P[S_n > a \mid S_{n-1} > a] - P[S_{n-1} > a] \\
&= \frac{1}{2}(P[S_{n-1} = a] - P[S_{n-1} = a+1])
\end{aligned}
$$

and analoguously

$$
P[S_n \le -a] - P[S_{n-1} \le -a] = \frac{1}{2}(P[S_{n-1} = a-1] - P[S_{n-1} = a]) .
$$

Therefore, using a)

$$
\begin{aligned}
P[T_{-a} = n] &= P[T_{-a} \le n] - P[T_{-a} \le n-1] \\
&= P[S_n \le -a] - P[S_{n-1} \le -a] + P[S_n > a] - P[S_{n-1} > a] \\
&= \frac{1}{2}(P[S_{n-1} = a] - P[S_{n-1} = a+1]) + \frac{1}{2}(P[S_{n-1} = a-1] - P[S_{n-1} = a]) \\
&= \frac{1}{2}(P[S_{n-1} = a-1] - P[S_{n-1} = a+1]) = \frac{a}{n} P[S_n = a]
\end{aligned}
$$

□

Corollary 2.12.3 (Ballot theorem)

$$
P[S_n = a \mid S_1 > 0, \dots, S_{n-1} > 0] = \frac{a}{n} \cdot P[S_n = a]
$$

Proof. Reverse time: the number of paths from 0 to a of length n which do no more hit 0 is the number of paths of length n which start in a and for which $T_{-a} = n$. □

Corollary 2.12.4 (Distribution of the first return time)

$$
P[T_0 > 2n] = P[S_{2n} = 0] .
$$

Proof.

$$
\begin{aligned}
P[T_0 > 2n] &= \frac{1}{2} P[T_{-1} > 2n-1] + \frac{1}{2} P[T_1 > 2n-1] \\
&= P[T_{-1} > 2n-1] \qquad \text{(by symmetry)} \\
&= P[S_{2n-1} > -1 \text{ and } S_{2n-1} \le 1] \\
&= P[S_{2n-1} \in \{0, 1\}] \\
&= P[S_{2n-1} = 1] = P[S_{2n} = 0]
\end{aligned}
$$

□

Remark. We see that $\lim_{n\to\infty} P[T_0 > 2n] = 0$ which restates that the random walk is recurrent. However, the expected return time is very long:

$$E[T_0] = \sum_{n=0}^{\infty} nP[T_0 = n] = \sum_{n=0}^{\infty} P[T_0 > n] = \sum_{n=0}^{\infty} P[S_n = 0] = \infty$$

since by Stirling's formula

$$P[S_{2n} = 0] \sim (\pi n)^{-1/2} .$$

We are interested now in the random variable

$$L(\omega) = \max\{0 \leq n \leq 2N \mid S_n(\omega) = 0\}$$

which describes the **last visit of the random walk in** 0 **before time** $2N$.

Interpretation. If the random walk describes a game between two players, who play over a time $2N$, then L is the time when one of the two players does no more give up his leadership.

Theorem 2.12.5 (Arc Sin law) *a) L has the discrete arc-sin distribution:*

$$P[L = 2n] = 2^{-2N} \binom{2n}{n} \binom{2N-2n}{N-n}$$

b) For $N \to \infty$, we have

$$P[\frac{L}{2N} \leq z] \to \frac{2}{\pi} \arcsin(\sqrt{z}) .$$

Proof.

$$P[L = 2n] = P[S_{2n} = 0] \cdot P[T_0 > 2N - 2n] = P[S_{2n} = 0] \cdot P[S_{2N-2n} = 0]$$

which gives a).
Stirling gives $P[S_{2k} = 0] \sim \frac{1}{\sqrt{\pi k}}$ so that

$$P[L = 2k] = \frac{1}{\pi} \frac{1}{\sqrt{k(N-k)}} = \frac{1}{N} f(k/N)$$

with

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} .$$

It follows that

$$P[L/2N \leq z] \to \int_0^z f(x)\, dx = \frac{2}{\pi} \arcsin(\sqrt{z}) .$$

□

Interpretation. From the shape of the arc-sin distribution, One has to expect that the winner takes the final leading position either very early or very late.

Remark. The arc-sin distribution is a natural distribution on the interval $[0,1]$ from the different points of view. It belongs to a measure which is

• the Gibbs measure of the quadratic map $x \mapsto 4 \cdot x(1-x)$ on the unit interval maximizing the Bolzmann-Gibbs entropy (see Exercice 5). It is a **thermodynamic equilibrium measure** for this quadratic map.

• It is the measure μ on the interval $[0,1]$ which minimizes the energy

$$I(\mu) = -\int_0^1 \int_0^1 \log |E - E'| \; d\mu(E) \; d\mu(E') \; .$$

One calls such measures also **potential theoretical equilibrium measures**.

2.13 The random walk on the free group

Definition. The **free group** F_d with d generators is the set of finite words w written in the $2d$ letters

$$A = \{a_1, a_2, \dots, a_d, a_1^{-1}, a_2^{-1}, \dots, a_d^{-1}\}$$

modulo the identification $w = w_1 w_2 \cdots w_n \sim w^{-1} = w_n^{-1} \cdots w_2^{-1} 1_1^{-1}$. The group operation is concatenating words $v \circ w = vw$. The inverse of $w = w_1 w_2 \cdots w_n$ is $w^{-1} = w_n^{-1} \cdots w_2^{-1} w_1^{-1}$. Elements w in the group F_d can be uniquely represented by reduced words obtained by deleting all words vv^{-1} in w. The identity e in the group F_d is the empty word. We denote with $l(w)$ the length of the reduced word of w.

Definition. Given a free group G with generators A and let X_k be uniformly distributed random variables with values in A. The stochastic process $S_n = \prod_{k=1}^n X_k$ is called the **random walk on the group** G. Note that the product is not commutative.

Interpretation. The random walk on the free group can be interpreted as a walk on a **tree**, since the Cayley graph of the group F_d with generators A contains no noncontractible closed circles.

Notation. Define for $n \in \mathbf{N}$

$$r^{(n)} = P[S_n = e \mid S_1 \neq e, S_2 \neq e, \dots S_{n-1} \neq e]$$

which is the probability of returning for the first time to e if one starts at e. Define also for $n \in \mathbf{N}$

$$m^{(n)} = P[S_n = e] \ .$$

with the convention $m^{(0)} = 1$. Let r and m be the generating functions of the sequences $r^{(n)}$ and $m^{(n)}$:

$$m(x) = \sum_{n=0}^{\infty} m^{(n)} x^n, \ r(x) = \sum_{n=0}^{\infty} r^{(n)} x^n \ .$$

These sums converge for $|x| < 1$.

Lemma 2.13.1 (Feller)

$$m(x) = \frac{1}{1 - r(x)} \ .$$

Proof. Let T be the stopping time

$$T = \min\{n \in \mathbf{N} \mid S_n = e\} \; .$$

Since $P[T = n] = r^{(n)}$, the function $r(x)$ is the characteristic function of T. Let T_i be independent random variables with distribution T. Then $\sum_{i=1}^n T_i$ has the characteristic function $x \mapsto r^n(x)$. We have

$$\begin{aligned} m^{(n)} &= P[S_n = e] = \sum_{k=0}^{\infty} \sum_{0 \leq n_1 < n_2 < \ldots < n_k} P[S_{n_1} = e, S_{n_2} = e, \ldots S_{n_k} = e, S_n \neq e \text{ for } n \notin \{n_1, \ldots n_k\}] \\ &= \sum_{k=0}^{\infty} P[\sum_{k=1}^{n} T_k = n] = \sum_{k=0}^{\infty} r_n^k \; . \end{aligned}$$

This means that $m = \sum_{n=0}^{\infty} r^n = 1/(1-r)$. □

Remark. This lemma is true for any random walk on a Cayley graph of any finitely presented group.

The numbers $r^{(2n+1)}$ are clearly zero. The values of r^{2n} can be computed combinatorically:

Lemma 2.13.2 (Kesten)

$$r^{2n} = \frac{1}{(2d)^n} n^{-1} \binom{2n-2}{n-1} 2d(2d-1)^{n-1} \; .$$

Proof. Clearly

$$r^{2n} = (2d)^{-1} |\{w_1 w_2 \ldots w_n \in G, w^k = w_1 w_2 \ldots w_k \neq e\}| \; .$$

To count the number of such words, we map every word with $2n$ letters into a path in $\mathbf{Z}^2$ going from $(0,0)$ to (n,n) which is away from the diagonal (except at the beginning or the end). The mapping is constructed in the following way: for every letter we record a horizontal or vertical step of length 1. If $l(w^k) = l(w^{k-1}) + 1$, we record a horizontal step. In the other case, if $l(w^k) = l(w^{k-1}) - 1$, we record a vertical step. The first step is horizontal independent of the word. There are

$$n^{-1} \binom{2n-2}{n-1}$$

such paths since by the distribution of the stopping time in the one dimensional random walk

$$P[T_{-1} = 2n-1] = (2n-1)^{-1} \cdot P[S_{2n-1} = 1] = (2n-1)^{-1} \binom{2n-1}{n} = n^{-1} \binom{2n-2}{n-1} \; .$$

Counting the number of words mapped into the same path, we see that we have in the first step $2d$ possibilities and later $(2d-1)$ possibilities in each of the $n-1$ horizontal step and only 1 possibility in a vertical step. We have therefore to multiply the number of paths with $2d(2d-1)^{n-1}$. □

Proposition 2.13.3 (Kesten, 1959)

$$m(x) = \frac{2d-1}{(d-1)+(d^2-(2d-1)x^2)^{1/2}} .$$

Proof. Since we know r^{2n} we can compute

$$r(x) = \frac{d-(d^2-(2d-1)x^2)^{1/2}}{2d-1}$$

and get the claim with Feller's lemma $m(x) = 1/(1-r(x))$. □

Remark *. One can read of from this formula that the spectrum of the free Laplacian on the **Bethe lattice** (this is how the tree is called by physisics), $L : l^2(F_d) \to l^2(F_d)$

$$Lu(g) = \sum_{a \in A} u(g+a)$$

is the whole interval $[-a, a]$ with $a = 2\sqrt{2d-1}$.

Corollary 2.13.4 *The random walk on the free group F_d with d generators is recurrent if and only if $d = 1$.*

Proof. Denote as in the case of the random walk on $\mathbf{Z}^d$ with B the random variable counting the total number of visits of the origin. We have then again $E[B] = \sum_n P[S_n = e] = \sum_n m^{(n)} = m(1)$. We see that for $d = 1$ we have $m(1) = \infty$ and that $m(d) < \infty$ for $d > 1$. This establishes the analogue of Polya's result on $\mathbf{Z}^d$ and leads in the same way to the recurrence:
(i) $d = 1$: We know that $\mathbf{Z}_1 = F_1$, and that the walk in $\mathbf{Z}^1$ is recurrent.
(ii) $d \geq 2$: define the event $A_n = \{S_n = e\}$. Then $A_\infty = \limsup_n A_n$ is the subset of Ω, for which the walk returns to e infinitely many times. Since for $d \geq 2$,

$$E[B] = \sum_{n=0}^{\infty} P[A_n] m(d) < \infty ,$$

Borel-Cantelli's lemma gives $P[A_\infty] = 0$ for $d > 2$. The particle returns therefore to 0 only finitely many times and similiarly it visits each vetex in F_d only finitely many times. This means that the particle eventually leaves every bounded set and escapes to infinity. □

Remark. We could say that the problem of the random walk on a discrete group G is **solvable** if one can give an algebraic formula for the the function $m(x)$. We have seen that the classes of abelian finitely generated and free groups are solvable. Trying to extend the class of solvable random walks seems to be an interesting problem. It would also be interesting to know, whether there exists a group such that the function $m(x)$ is transcendental.

2.14 Repetition, distribution of first return time

Let S_n be the random walk on $\mathbf{Z}^1$. Given $a \in \mathbf{Z}$, we define the stopping time

$$T_a = \min\{n \in \mathbf{N} \mid S_n = a\} .$$

Lemma 2.14.1 (Reflexion principle) *For $a, b \in \mathbf{N} \setminus \{0\}$, one has*

$$P[a + S_n = b \mid T_{-a} \leq n] = P[S_n = a + b] .$$

Theorem 2.14.2 (Distribution of stopping time (=ruin time))

a) $P[T_{-a} \leq n] = P[S_n \leq -a] + P[S_n > a]$.

b) $P[T_{-a} = n] = \frac{a}{n} P[S_n = a]$.

Proof. a) We use the reflection principle in the second inequality

$$\begin{aligned} P[T_{-a} \leq n] &= \sum_{b \in \mathbf{Z}} P[T_{-a} \leq n | a + S_n = b] \\ &= \sum_{b \leq 0} P[a + S_n = b] + \sum_{b > 0} P[S_n = a + b] \\ &= P[S_n \leq -a] + P[S_n > a] \end{aligned}$$

b) We have

$$P[S_n = a] = \begin{pmatrix} n \\ \frac{a+n}{2} \end{pmatrix}$$

from which we get

$$\frac{a}{n} P[S_n = a] = \frac{1}{2}(P[S_{n-1} = a - 1] - P[S_{n-1} = a + 1]) .$$

We have also

$$\begin{aligned} P[S_n > a] - P[S_{n-1} > a] &= P[S_n > a \mid S_{n-1} \leq a] + P[S_n > a \mid S_{n-1} > a] - P[S_{n-1} > a] \\ &= P[S_{n-1} = a, x_n = 1] + P[S_{n-1} > a | S_n \leq a] \\ &= P[S_{n-1} = a, x_n = 1] + P[S_{n-1} = a + 1 | X_n = -1] \\ &= \frac{1}{2}(P[S_{n-1} = a] - P[S_{n-1} = a + 1]) \end{aligned}$$

and analoguously

$$P[S_n \leq -a] - P[S_{n-1} \leq -a] = \frac{1}{2}(P[S_{n-1} = a - 1] - P[S_{n-1} = a]) .$$

Therefore, using a)

$$\begin{aligned}
P[T_{-a} = n] &= P[T_{-a} \leq n] - P[T_{-a} \leq n-1] \\
&= P[S_n \leq -a] - P[S_{n-1} \leq -a] + P[S_n > a] - P[S_{n-1} > a] \\
&= \frac{1}{2}(P[S_{n-1} = a] - P[S_{n-1} = a+1]) + \frac{1}{2}(P[S_{n-1} = a-1] - P[S_{n-1} = a]) \\
&= \frac{1}{2}(P[S_{n-1} = a-1] - P[S_{n-1} = a+1]) = \frac{a}{n}P[S_n = a]
\end{aligned}$$

□

2.15 The free Laplacian on a discrete group

Definition. Let G be a countable discrete group and $A \subset G$ a finite set which generates G. The **Cayley graph** Γ of (G, A) is the graph with edges G and sites (i, j) satisfying $i - j \in A$ or $j - i \in A$.

Notice. We write the composition in G additively evenso we do not assume that G is abelian. We allow A to contain also the identity $e \in G$. In this case, the Cayley graph contains two closed loops of length 1 at each site.

Definition. The **symmetric random walk** on $\Gamma(G, A)$ is the process obtained by summing up independent uniformly distributed $(A \cup A^{-1})$-valued random variables X_n. More generally, we can allow the random variables X_n to be independent but have any distribution on $A \cup A^{-1}$. This distribution is given by numbers $p_a = p_a^{-1} \in [0, 1]$ satisfying $\sum_{a \in A} p_a + \sum_{a \in A^{-1}} p_a = 1$. It is convenient to extend the probability distribution p onto G by putting it to zero on $G \setminus A$.

Definition. The **free Laplacian** for the random walk given by (G, A, p) is the linear operator on $l^2(G)$ defined by

$$L_{gh} = p_{g-h} \ .$$

Since we assumed $p_a = p_{a^{-1}}$, the matrix L is symmetric: $L_{gh} = L_{hg}$ and the spectrum $\sigma(L) = \{E \in \mathbf{C} \mid (L - E) \text{ invertible}\}$ is a compact subset of the real line.

Remark. One can look at L as the transition probability matrix of the random walk which is a Markov chain (see later).

Example. $G = \mathbf{Z}$, $A = \{1\}$. $p = p_a = 1/2$ for $a = 1, -1$ and $p_a = 0$ for $a \notin \{1, -1\}$. The matrix

$$L = \begin{pmatrix} \cdot & \cdot & & & & & & \\ \cdot & 0 & p & & & & & \\ & p & 0 & p & & & & \\ & & p & 0 & p & & & \\ & & & p & 0 & p & & \\ & & & & p & 0 & p & \\ & & & & & p & 0 & \cdot \\ & & & & & & \cdot & \cdot \end{pmatrix}$$

is also called a **Jacobi matrix**. It acts on $l^2(\mathbf{Z})$ by $(Lu)_n = p(u_{n+1} + u_{n-1})$.

A basic question. What is the relation between the spectrum of L and the structure of the group G?

Definition. Define $m^{(n)}$ be the probabiliy that the random walk starting in e returns in n steps to e and let

$$m(x) = \sum_{n \in G} m^{(n)} x^n$$

be the generating function of the sequence $m^{(n)}$.

Proposition 2.15.1 *The norm of L (which is equal to the spectral radius of L), is equal to $\limsup_{n\to\infty}(m^{(n)})^{1/n}$, (the inverse of the radius of convergence of $m(x)$).*

Proof. Since L is symmetric and real, it is selfadjoint and the spectrum of L is a subset of $\mathbf{R}$ and the spectral radius is equal to the norm.

We have $[L^n]_{ee} = m^{(n)}$ since $[L^n]_{ee}$ is the sum of products $\prod_{j=1}^n p_{a_j}$ each of which is the probability that a specific path of length n starting and landing at e occurs.

It rests therefore to show that

$$\limsup_{n\to\infty} ||L^n||^{1/n} = \limsup_{n\to\infty} [L^n]_{ee}^{1/n}$$

and since the $\geq$ direction is trivial we have only to show that $\leq$ direction.

Denote with $E(\lambda)$ the spectral projection matrix of L. $dE(\lambda)$ is a projection valued measure on the spectrum and the spectral theorem says that L can be written as $L = \int \lambda \, dE(\lambda)$. The measure dE_{ee} is called the **density of states** of L. Since $E(\lambda)$ is a projection, we have

$$\begin{aligned} E_{ee}(\lambda) - E_{ee}(\mu) &= (E(\lambda) - E(\mu))^2_{ee} \\ &= \sum_g (E(\lambda) - E(\mu))_{eg}(E(\lambda) - E(\mu))_{ge} \\ &\geq |(E(\lambda) - E(\mu))_{eg}|^2 \end{aligned}$$

we see that the support of the density of states dk is equal to the spectrum of L because $E(\lambda) - E(\mu)$ is nonzero iff there exists some spectrum of L in $[\lambda, \mu)$. Since

$$\frac{(-1)}{\lambda} \sum_n \frac{[L^n]_{ee}}{\lambda^n} = \int (E - \lambda)^{-1} \, dk(E)$$

can't be analytic in λ in a point λ_0 of the support of dk which is the spectrum of L, the claim follows. $\square$

Example.
The Fourier transform $F : l^2(\mathbf{Z}^1) \to L^2(\mathbf{T}^1)$:

$$\hat{u}(x) = (Fu)(x) = \sum_{n \in \mathbf{Z}} u_n e^{inx}$$

diagonalises the matrix

$$\begin{aligned}(FLF^*)\hat{u}(x) &= ((FL)(u_n)(x) = pF(u_{n+1} + u_{n-1})(x) \\ &= p \sum_{n \in \mathbf{Z}} (u_{n+1} + u_{n-1}) e^{inx} \\ &= p \sum_{n \in \mathbf{Z}} u_n (e^{i(n-1)x} + e^{i(n+1)x}) \\ &= p \sum_{n \in \mathbf{Z}} u_n (e^{ix} + e^{-ix}) e^{inx} = p \sum_{n \in \mathbf{Z}} u_n 2\cos(x) e^{inx} = \cos(x) \cdot \hat{u}(x)\end{aligned}$$

This shows that the spectrum of FLF^* is $[-1, 1]$ and because F is an unitary transformation, also the spectrum of L is $[-1, 1]$.

Example. $G = \mathbf{Z}^d$, $A = \{e_i\}_{i=1}^d$ the standard bases. $p = p_a = 1/(2d)$. The analoguous Fourier transform $F : l^2(\mathbf{Z}^d) \to L^2(\mathbf{T}^d)$ shows that FLF^* is the multiplication with $\frac{1}{d} \sum_{j=1}^d \cos(x_j)$. The spectrum is again the interval $[-1, 1]$.

Example. The Fourier diagonalisation works for any discrete abelian group with finitely many generators.

Example. $G = F_d$ the free group with the natural d generators. The spectrum of L is

$$[-\sqrt{2d-1}/d . \sqrt{2d-1}/d]$$

which is strictly contained in $[-1, 1]$ if $d > 1$.

Kesten has shown that the spectral radius of L is equal to 1 if and only if the group G has an invariant mean.

Generalisations. Random walks and Laplacians can also be defined on any graph. The spectrum of the Laplacian on a finite graph is an invariant of the graph but there are nonisomorphic graphs with the same spectrum. There are known selfsimilar graphs, for which the Laplacian has pure point spectrum. There are also known infinite graphs, such that the Laplacian has pure singular continuous spectrum. (For more information see the review article B.Hohar,W.Woess, A survey on spectra of infinite graphs, Bull. London. Math. Soc.21 (1981(209-234).)

2.16 A discrete Feynmann-Kac formula

Definition. A **discrete Schrödinger operator** is a bounded linear operator L on $l^2(\mathbf{Z}^d)$ of the form

$$(Lu)(n) = u_{n+1} + u_{n-1} + v_n u_n \ ,$$

where v_n are two real bounded sequences. They are discrete versions of operators $L = -\Delta + V(x)$ on $L^2(\mathbf{R}^d)$, where Δ is the free Laplacian.

Definition. The **Schrödinger equation**

$$i\hbar\dot{u} = Lu, \ u(0) = u_0$$

is a differential equaiton in $l^2(\mathbf{Z}^d)$ which describes the motion of the wave function $u \in l^2(\mathbf{Z}^d)$ of a classical quantum mechanical system.

Remark. The solution of the Schrödinger equation can be given by

$$u(t) = e^{\frac{t}{i\hbar}L} u(0) \ .$$

The solution exists for all times.

It is an acchievment of Feynmann to see that the evolution as a **path integral**. In the case of differential operators L, where this idea can be made rigorous by going to imaginary time and one can write for $L = -\Delta + V$

$$e^{-tH}u(x) = E_x[e^{\int_0^t V(\gamma(s))\, ds} u_0(\gamma(t))] \ ,$$

where E_x is the expectation value with respect to the measure P_x given by the Brownian motion starting at x. We derive here a discrete version of that **Feynmann-Kac formula**.

Definition. The Schrödinger equation with **discrete time** is given by

$$i\hbar(u_{t+\epsilon} - u_t) = \epsilon L u_t$$

where $\epsilon > 0$ is fixed. We get then

$$u_{t+n\epsilon} = (1 - \frac{i\epsilon}{\hbar} L)^n u_t$$

and we denote the right hand side with $\tilde{L}^n u_t$.

Notation. Denote by $\Gamma_n(i,j)$ the set of paths of length n in the graph G having as edges $\mathbf{Z}^d$ and sites pairs $[i,j]$ with $|i-j| \leq 1$. The graph G can be considered as the Cayley graph of the group $\mathbf{Z}^d$ with the generators $A \cup A^{-1} \cup \{e\}$, where

$A = \{e_1, \dots, e_d,\}$ is the set of natural generators and where e is the identity.

Notation. Given a path γ of finite length n, we use the notation

$$\exp(\int_\gamma L) = \prod_{i=1}^{n} L_{\gamma(i),\gamma(i+1)} \ .$$

Notation. Let Ω is the set of all paths on G and E denotes the expectation with respect to a measure P of the random walk on G starting at 0.

Theorem 2.16.1 (Discrete Feynmann-Kac formula) *Given a discrete Schrödinger operator L. Then*

$$(L^n u)(0) = E_0[\exp(\int_0^n L)\ u(\gamma(n))] \ .$$

Proof.

$$(L^n u)(0) = \sum_j (L^n)_{0j} u(j) = \sum_j \sum_{\gamma \in \Gamma_n(0,j)} \exp(\int_0^n L)\ u(j) = \sum_{\gamma \in \Gamma_n} \exp(\int_0^n L) u(\gamma(n)) \ .$$

□

Remark. This discrete random walk expansion corresponds clearly to the Feynmann-Kac formula. If we extend the potential to all the sites of the Cayley graph by putting $V([[i,i]) = v(i)$ and $V([i,j]) = 0$ else, we can define $\exp(\int_\gamma V)$ as the product $\prod_{i=1}^n V([\gamma(i), \gamma(i+1)])$. Then

$$(L^n u)(0) = E[\exp(\int_0^n V) u(\gamma(n))]$$

which is formally the Feynman-Kac formula.

In order to compute $(\tilde{L}^n u)(i)$ with $\tilde{L} = (1 - \frac{i\epsilon}{\hbar} L)$ we have to take the potential $\tilde{v}$

$$\tilde{v}([i,i]) = 1 - \frac{i\epsilon}{\hbar} v(\gamma(i)) \ .$$

Remark. Looking at the Schrödinger equation with discrete time has the disadvantage that the time evolution of the system is no more unitary. This drawback could be overcome however by considering also $i\hbar(u_t - u_{t-\epsilon}) = \epsilon L u_t$ so that the propagator from $u_{t-\epsilon}$ to $u_{t+\epsilon}$ is given by the unitary operator

$$U = (1 - \frac{i\epsilon}{\hbar} L)(1 + \frac{i\epsilon}{\hbar} L)^{-1}$$

which is a **Cayley transform** of L.

2.17 Markov chains

Definition. Given a measurable space $(S, \mathcal{B})$ called **state space**, where S is a set and $\mathcal{B}$ is a σ-algebra on S. A function $\mathcal{P} : S \times \mathcal{B} \to \mathbf{R}$ is called a **transition probability function** if $\mathcal{P}(x, \cdot)$ is a probability measure on $(S, \mathcal{B})$ for all $x \in S$ and if for every $B \in \mathcal{B}$, the map $s \to \mathcal{P}(s, B)$ is $\mathcal{B}$-measurable. Define $\mathcal{P}^1(x, B) = \mathcal{P}(x, B)$ and inductively the measures $\mathcal{P}^{n+1}(x, B) = \int_S \mathcal{P}^n(y, B)\mathcal{P}(x, dy)$, where we write $\int \mathcal{P}(x, dy)$ for the integration on S with respect to the measure $\mathcal{P}(x) = \mathcal{P}(x, \cdot)$.

Remark. The transition probability functions are elements in $\mathcal{L}(S, M(S))$, where $M(S)$ is the set of Borel probability measures on S. The multiplication

$$(\mathcal{P} \circ \mathcal{Q})(x, B) = \int_S \mathcal{P}(y, B) \; d\mathcal{P}(x)$$

makes them into a semigroup. Given one element $\mathcal{P}$, we have defined the semigroup $\mathcal{P}^n$ which is commutative. The relation $\mathcal{P}^{n+m} = \mathcal{P}^n \circ \mathcal{P}^m$ is calso called **Chapmann-Kolmogorov equation**.

Definition. Given a probability space $(\Omega, \mathcal{A}, P)$ with a filtration $\mathcal{A}_n$ of σ-algebras. An $\mathcal{A}_n$-adapted process X_n with values in S is called a **Markov process** or **Markov chain**, if there exists a transition probability function $\mathcal{P}$ such that

$$P[X_n \in B \mid \mathcal{A}_k](\omega) = \mathcal{P}^{n-k}(X_k(\omega), B) \; .$$

Definition. A Markov chain is called a **denumerable Markov chain**, if the state space S is countable, a **finite Markov chain**, if the state space is finite.

Remark. It follows that a Markov process X_n satisfies the **elementary Markov property**

$$P[X_n \in B \mid X_1, \dots, X_k] = P[X_n \in B \mid X_k] \; .$$

This means that the probability distribution of X_n is determined by knowing the one of X_{n-1}.

> **Lemma 2.17.1 (Markov processes exist)** *Given any state space $(S, \mathcal{B})$ and a transition probability function $\mathcal{P}$, there exists a corresponding Markov process.*

Proof. Choose a probability measure μ on $(S, \mathcal{B})$ and define on the product space $(\Omega, \mathcal{A}) = (S^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}})$ the π-system $\mathcal{C}$ consisting of of zylinder-sets $\prod_{n \in \mathbf{N}} B_n$ given by a sequence $B_n \in \mathcal{B}$ such that $B_n = S$ except for finitely many n.

Define a measure $P = P_\mu$ on $(\Omega, \mathcal{C})$ by requiring

$$P[\omega_k \in B_k, k = 1, \dots n] = \int_{B_0} \mu(dx_0) \int_{B_1} \mathcal{P}(x_0, dx_1) \dots \int_{B_n} \mathcal{P}(x_{n-1}, dx_n) \; .$$

This measure has a unique extension to the σ-algebra $\mathcal{A}$.
Define the increasing sequence of σ-algebras $\mathcal{A}_n = \mathcal{B}^n \times \prod_{i=1}^n \{\emptyset, \Omega\}$ containing zylinder sets. The random variables $X_n(\omega) = x_n$ are clearly $\mathcal{A}^n$-adapted. In order to see that it is a Markov process, we have to check that

$$P[X_n \in B_n \mid \mathcal{A}_{n-1}](\omega) = \mathcal{P}(X_{n-1}(\omega), B_n)$$

which is a special case of the above requirement by taking $B_k = S$ for $k \neq n$.
□

Examples.
1) **Independent S-valued random variables**
Assume the measures $\mathcal{P}(x, \cdot)$ are independent of x. Denote this measure by $\mathcal{P}$. In this case

$$P[X_n \in B_n \mid \mathcal{A}_{n-1}](\omega) = \mathcal{P}(B_n)$$

which means that $P[X_n \in B_n \mid \mathcal{A}_{n-1}] = P[X_n \in B_n]$ and so X_n are IID S-valued random variables and $\mathcal{P}$ is the law of X_n. Every sequence of IID random variables is a Markov process.

2) **Countable and finite state Markov chains**.
Given a Markov process with finite or countable state space S. We define the transition matrix $\mathcal{P}_{ij}$ on the Hilbert space $l^2(S)$ by

$$P_{ij} = \mathcal{P}(i, \{j\}) \; .$$

The matrix $\mathcal{P}$ transports the law of X_n into the law of X_{n+1}.
The transition matrix $\mathcal{P}_{ij}$ has the property to be a stochastic matrix: $\sum_j \mathcal{P}_{ij} = 1$ with $\mathcal{P}_{ij} \geq 0$. Every measure on S can be given by a vector $\pi \in l^2(S)$ and $\mathcal{P}\pi$ is again a measure. If X_0 is constant equal to i and X_n is a Markov process with transition probabilty $\mathcal{P}$, then $\mathcal{P}^n_{ij} = P[X_n = j]$.

3) **Sum of independent S-valued random variables** Let S be a countable abelian group and π a probability distribution on S assiging to each $j \in S$ the weight π_j. Define $\mathcal{P}_{ij} = \pi_{j-i}$. Now X_n is the sum of n independent random variables with law π. The sum changes from i to j with probability $P_{ij} = p_{i-j}$.

4) **Branching processes** Given $S = \{0, 1, 2 \dots\} = \mathbf{N}$ with fixed probability distribution π. If X is a S-valued random variable with distribution π then $\sum_{k=1}^n X_k$ has a distribution which we denote with $\pi^{(n)}$. Define the matrix $P_{ij} = \pi^{(i)}_j$. The Markov chain with this transition probability matrix on S is

called a **branching process**.

Definition. The transition probability function $\mathcal{P}$ acts also on measures of S by $\mathcal{P}(\pi)(B) = \int_S P(x,B)\ d\pi(x)$. A probability measure π is called **invariant** if $\mathcal{P}\pi = \pi$.

We can assign a **Markov operator** to a tansition probability function.

Lemma 2.17.2 *There exists a probability measure ν on $(S, \mathcal{B})$, such that every measure μ which is abslutely continuous with respect to ν has the property that also $\mathcal{P}\mu$ is absolutely continuous with respect to μ.*

Proof. Choose $x \in S$ and define the probability measure

$$\nu = \sum_{n=0}^{\infty} 2^{-n} \mathcal{P}^n(x, \cdot) \ .$$

Given $\mu = f \cdot \nu$ with $f \in L^1(S)$. Then

$$\mathcal{P}\mu = \int_S P(x,B) f(x)\ d\nu(x)$$

is absolutely continuous with respect to μ since $\mathcal{P}\mu(B) = 0$ implies $\mathcal{P}(x,B) = 0$ for almost all x with $f(x) > 0$ and so $f\nu(B) = 0$. □

Corollary 2.17.3 *To each transition probability function can be assigned a Markov operator $\mathcal{P} : L^1(S) \to L^1(S)$.*

Proof. Choose ν as above and define

$$\mathcal{P} f_1 = f_2$$

if $\mathcal{P}\mu_1 = \mu_2$ with $\mu_i = f_i \nu_i$. □

We see that the abstract approach with studiing Markov operators on $L^1(S)$ is more general then looking at transition probability measures.

Chapter 3

Stochastic calculus

3.1 Brownian motion

Definition. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A collection of random variables X_t, $t \in \mathbf{T}$ with values in a separable metric space S is called a **stochastic process**. The time $\mathbf{T}$ can be a discrete subset of $\mathbf{R}$ which is then called a **discrete time stochastic process** or an interval or $\mathbf{R}$ which is called a **stochastic process with continuous parameter**. When a sample $\omega \in \Omega$ is fixed, we can regard $X_t(\omega)$ as a function of t called a **sample function** of the stochastic process. We mainly will consider the case, when $S = \mathbf{R}^d$.

Definition. A stochastic process is called **measurable** if $X : \mathbf{T} \times \Omega \to S$ is measureable with respect to the product σ-algebra $\mathcal{B}(\mathbf{T}) \times \mathcal{A}$. It is called **continuous** if $X : \mathbf{T} \times \Omega \to S$ is continuous. In the case $S = \mathbf{R}$, one sais X is **continuous in probability** if for any $t \in \mathbf{R}$ and $X_{t+h} \to X_t$ in probabiliy for $h \to 0$. If the sample function $X_t(\omega)$ is a continuous function of t for almost all ω, then X_t is called a **continuous stochastic process**. If the sample function is **a right continuous** function in t for almost all $\omega \in \Omega$, X_t is called a **right continuous stochastic process**. Two stochastic process X_t and Y_t satisfying $P[X_t - Y_t = 0] = 1$ for all $t \in \mathbf{T}$ are called **modifications of each other**. They are called **indistinguishable** if for almost all ω, the sample functions coincide $X_t(\omega) = Y_t(\omega)$.

Definition. A continuous process X_t with values in $\mathbf{R}^d$ is having the mean vector $m_t = E[X_t]$ and the covariance matrix $V(s,t) = \text{Cov}[X_s, X_t] = E[(X_s - m_s)(X_t - m_t)^*]$ is called **Brownian motion** if for any $0 \leq t_0 < t_1 < \ldots < t_n$, the random variables $X_{t_0}, X_{t_{i+1}} - X_{t_i}$ are independent and the covariance matrix V satisfies $V(s,t) = V(r,r)$, where $r = \min(s,t)$ and $s \mapsto V(s,s)$ is increasing. It is called the **standard Brownian motion** if $m_t = 0$ for all t and $V(s,t) = \min\{s,t\}$. We will in the following mainly consider the case of standard Brownian motion.

Definition. A $\mathbf{R}^n$-valued random variable X is called **Gaussian** if

$$E[e^{iu\cdot X}] = e^{-(u,Vu)/2+imu}$$

for some nonsingular symmetric covariance matrix V and a vector $m = E[X]$.
A **Gaussian processs** is a real valued stochastic process such that $(X_{t_0}, X_{t_1}, \dots X_{t_n})$ has a Gaussian distribution for any $t_0 \leq t_1 < \dots < t_n$. It is called **centered** if $m_t = E[X_t] = 0$ for all t.

Remark. By Fourier transform, we compute the law of a centered Gaussian random variable X by

$$d\mu_X(x) = (2\pi)^{-n/2}\det(V)^{-1/2}e^{-\frac{1}{2}(x,V^{-1}x)} .$$

(This computation is done in a bases, where V is diagonal, where it reduces to one dimensional Fourier integrals).

Lemma 3.1.1 *Two Gaussian random variables X, Y are independent if and only if they are uncorrelated:* $\text{Cov}[X,Y] = 0$.

Proof. We can assume without loss of generality that the random variables X, Y are centered.
Two $\mathbf{R}^n$-valued Gaussian random variables X and Y are independent if and only if

$$\phi_{(X,Y)}(s,t) = \phi_X(s)\cdot\phi_Y(t), \forall s,t \in \mathbf{R}^n .$$

Proof. we have only to show that X_i and Y_j are independent. But we have Lévy's inversion formula

$$\begin{aligned} P[a < X < b, c < Y < d] &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{isa}-e^{-isb}}{is}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{ita}-e^{-itb}}{it}\cdot\phi_{(X,Y)}(s,t)\,ds\,dt \\ &= P[a < X < b]\cdot P[c < Y < d] . \end{aligned}$$

The claim of the lemma follows since for Gaussian random variable X, Y, the characteristic function of (X, Y) is

$$\begin{aligned} \phi_{(X,Y)}(s,t) &= E[e^{isX+itY}] = e^{-\frac{1}{2}(sX+tY)^2} \\ &= E[e^{isX}]E[e^{itY}]e^{-ts\text{Cov}[X,Y]} . \end{aligned}$$

□

Proposition 3.1.2 *If X_t is a Gaussian process with covariance $V(s,t) = V(r,r)$ with $r = \min(s,t)$, then it is a Brownian motion.*

Proof. By the above lemma, we have to check only that for all $i < j$

$$\mathrm{Cov}[X_{t_0}, X_{t_{j+1}} - X_{t_j}] = 0,\ \mathrm{Cov}[X_{t_{i+1}} - X_{t_i}, X_{t_{j+1}} - X_{t_j}] = 0 \ .$$

But by assumption

$$\mathrm{Cov}[X_{t_0}, X_{t_{j+1}} - X_{t_j}] = V(t_0, t_{j+1}) - V(t_0, t_j) = V(t_0, t_0) - V(t_0, t_0) = 0$$

and

$$\begin{aligned}\mathrm{Cov}[X_{t_{i+1}} - X_{t_i}, X_{t_{j+1}} - X_{t_j}] &= V(t_{i+1}, t_{j+1}) - V(t_{i+1}, t_j) - V(t_i, t_{j+1}) + V(t_i, t_j) \\ &= V(t_{i+1}, t_{i+1}) - V(t_{i+1}, t_{i+1}) - V(t_i, t_i) + V(t_i, t_i) = 0 \ .\end{aligned}$$

□

3.2 History of Brownian motion

Anyone looking at water through a microscope sees little things moving around. The first dynamical theory of that was that these particles were alive. **Brown** was studying the fertilization process in a species of flowers. Looking at the pollen in water through a microscope, he observed small particles in "rapid oscillatory motion". Brown's explanation to this was that matter is composed of small particles, which he calls active molecules, which exhibit a rapid, irregular motion having its origin in the particles themselves and not in the surrounding fluid. Brown's contribution was to establish Brownian motion as an important phenomenon, to demonstrate its presence in inorganic as well as organic matter and to refute by experiment wrong mechanical explanations of the phenomenon.

The topic seemed to have been neglected in the first part of the 19'th century but awareness of the pheomenon remained widespread. From 1860 on, many scientists worked on the phenomenon. The first one, to express a notion close to the modern theory of Brownian motion was **Wiener** in 1863. Careful experiments and arguments lead to the kinetic theory that Brownian motion is caused by bombardment by the molecules of fluid. But the results failed the theory by a factor of about 100'000. The difficulty was the fact that the motion is very irregular composed of translations and rotations and that the trajectory appears to have no tangent. So, any attempt to determine the velocity of the particles failed. The success of **Einstein**'s theory of Brownian motion was largely due to his go around this question.

Einstein himself was unaware of the phenomenon Brownian motion. He predicted it on theoretical grounds and formulated a correct quantitative theory of it. Einsteins's arguments do not give a dynamical theory of Brownian motion. It only determines the nature of the motion and the value of the diffusion coefficients on the basis of some assumptions.

A modern probabilistic treatment of Brownian motion became possible in this century with an axiomatically developed theory of probability and stochastic processes.

For details about the history of Brownian motion, see the first chapter of Nelson's monograph.

3.3 The existence proof

The construction of Brownian motion goes in two steps: construct first a Gaussian process which has the desired properties and show then that it has a modifiction which is a continuous process.

Proposition 3.3.1 *Given a separable real Hilbert space $(H, ||\cdot||)$. There exists a probability space $(\Omega, \mathcal{A}, P)$ and a family $X(h), h \in H$ of real valued random variables on Ω such that $h \mapsto X(h)$ is linear, and $X(h)$ is Gaussian, centered and $E[X(h)^2] = ||h||^2$.*

Proof. Pick an orthonormal basis e_n in H and attach to each e_n a centered Gaussian IID random variable g_n satisfying $||g_n||_2 = 1$. Given a general $h = \sum h_n e_n \in H$, define

$$X(h) = \sum_n h_n g_n$$

which converges in $\mathcal{L}^2$.
Since g_n are independent, they are orthonormal in $\mathcal{L}^2$ so that

$$||X(h)||_2^2 = \sum_n h_n^2 ||g_n||_2 = \sum_n h_n^2 = ||h||_2^2 \ .$$

□

Definition. If we choose $H = L^2(\mathbf{R}^+, dx)$, the map $X : H \mapsto \mathcal{L}^2$ is also called a **Gaussian measure**. For a Borel set $A \subset \mathbf{R}^+$ we define then $X(A) = X(1_A)$. The term "measure" is warranted from the fact that $X(A) = \sum_n X(A(n))$ if A is a countable disjoint union of Borel sets A_n.

Remark. The space $X(H) \subset \mathcal{L}^2$ is a Hilbert space isomorphic to H and in particular

$$E[X(h)X(h')] = (h, h') \ .$$

We know from the lemma that h and h' are orthogonal if and only if $X(h)$ and $X(h')$ are independent and that

$$E[X(A)X(B)] = \text{Cov}[X(A), X(B)] = (1_A, 1_B) = |A \cap B| \ .$$

Especially $X(A)$ and $X(B)$ are independent if and only if A and B are disjoint.

Definition. Define the process $B_t = X([0, t])$. This process has independent increments $B_{t_i} - B_{t_{i-1}}$ and is a Gaussian process. For each t, we have $E[B_t^2] = t$ and for $s < t$, the increment $B_t - B_s$ has variance $t - s$ so that

$$E[B_s B_t] = E[B_s^2] + E[B_s(B_t - B_s)] = E[B_s^2] = s \ .$$

This model of Brownian motion has therefore everything except continuity!

Theorem 3.3.2 (Kolmogorov's lemma) *Given a real-valued process $X_t, t \in [a,b]$ for which there exist three constants $p > r, K$ such that*

$$E[|X_{t+h} - X_t|^p] \leq K \cdot h^{1+r}$$

for every $t, t+h \in [a,b]$, then X_t has a modification Y_t which is almost surely continuous:

$$|Y_t(\omega) - Y_s(\omega)| \leq C(\omega)\ |t-s|^\alpha, 0 < \alpha < r/p\ .$$

Proof. Assume without loss of generality that $a = 0, b = 1$. Define $\epsilon = r - \alpha p$. We get by Chebychev-Markov inequality

$$P[|X_{t+h} - X_t|] \geq |h|^\alpha] \leq |h|^{-\alpha p} E[|X_{t+h} - X_t|^p] \leq K|h|^{1+\epsilon}$$

so that

$$P[|X_{(k+1)/2^n} - X_{k/2^n}| \geq 2^{-n\alpha}] \leq K 2^{-n(1+\epsilon)}\ .$$

Therefore

$$\sum_{n=1}^\infty \sum_{k=0}^{2^n-1} P[|X_{(k+1)/2^n} - X_{k/2^n}| \geq 2^{-n\alpha}] < \infty\ .$$

By Borel-Cantelli's lemma, there exists $n(\omega) < \infty$ almost everywhere such that for all $n \geq n(\omega)$ and $k = 0, \dots, 2^n - 1$

$$|X_{(k+1)/2^n}(\omega) - X_{k/2^n}(\omega)| < 2^{-n\alpha}\ .$$

Let $n \geq n(\omega)$ and $t \in [k/2^n, (k+1)/2^n]$ of the form $t = k/2^n + \sum_{i=1}^m \gamma_i/2^{n+i}$ with $\gamma_i \in \{0,1\}$. Then

$$|X_t(\omega) - X_{k2^{-n}}(\omega)| \leq \sum_{i=1}^m \gamma_i 2^{-\alpha(n+i)} \leq d 2^{-n\alpha}$$

with $d = (1 - 2^{-\alpha})^{-1}$. Similarly

$$|X_t - X_{(k+1)2^{-n}}| \leq d 2^{-n\alpha}\ .$$

Given $t, t+h \in D = \{k2^{-n} \mid n \in \mathbf{N}, k = 0, \dots n-1\}$. Take n so that $2^{-n-1} \leq h < 2^{-n}$ and k so that $k/2^{n+1} \leq t < (k+1)/2^{n+1}$. Then $(k+1)/2^{n+1} \leq t+h \leq (k+3)/2^{n+1}$ and

$$|X_{t+h} - X_t| \leq 2d2^{-(n+1)\alpha} \leq 2dh^\alpha\ .$$

For almost all ω, this holds for sufficiently small h.

We know now that for almost all ω, the path $X_t(\omega)$ is uniformly continuous on the dense set of dyadic numbers D. Such a function can be extended to a continuous function on $[0,1]$ by defining

$$Y_t(\omega) = \lim_{s \in D \to t} X_s(\omega)\ .$$

Since the inequality in the assumption of the theorem implies $E[X_t(\omega) - \lim_{s \in D \to t} X_s(\omega)] = 0$ and by Fatou's lemma $E[Y_t(\omega) - \lim_{s \in D \to t} X_s(\omega)] = 0$ we know that $X_t = Y_t$ almost everywhere. Y is therefore a modification of X. Moreover, Y satisfies

$$|Y_t(\omega) - Y_s(\omega)| \leq C(\omega) \, |t-s|^\alpha$$

for all s, t. □

Corollary 3.3.3 *Brownian motion exists.*

Proof. In one dimension, take the process B_t from above. Since $X_h = B_{t+h} - B_t$ is centered with variance h, the four'th moment is $E[X_h^4] = \frac{d^4}{dx^4} \exp(-x^2 h/2)_{|x=0} = 3h^2$, so that

$$E[(B_{t+h} - B_t)^4] = 3h^2 \; .$$

Kolmogorov's lemma gives the existence of a continuous modification of B.

To define standard Brownian motion in n dimension, we take the joint motion $B_t = (B_t^{(1)}, \dots, B_t^{(n)})$ of n independent one- dimensional Brownian motions. □

Definition. Let B_t be the standard Brownian motion. For any $x \in \mathbf{R}^n$, the process $X_t^x = x + B_t$ is called **Brownian motion started at** x.

3.4 Overview over other existence proofs

The first rigorous construction of Brownian motion was given by Wiener in 1923.

By construction of the Wiener measure on $C[0,1]$, one has a construction of Brownian motion, where the probability space is directly the set of paths. One has then the process $X_t(\omega) = \omega(t)$. We will come to this later. A general construction of such measures is possible given a Markov transition proability function (see for example Strook: Lectures on Stoachastic Analysis and Diffusion Theory).

The construction given in this course is due to Neveu going back to Kakutani. It can be found in the book of Simon about functional integration or in the book of Revuz and Yor about continuous Martingales and Brownian motion. The construction has the advantage that it can be applied to more general situations.

In McKean's book ("Stochastic integrals") one can find Lévy's (modified by Ciesielski) direct proof of the existence of Brownian motion. The proof is also nice since it gives an explicit formula for the Brownian motion process B_t. We outline it shortly:

1) Take as a basis in $L^2([0,1]$ the *Haar functions*

$$f_{k,n} := 2^{(n-1)/2}(1_{[(k-1)2^{-n},k2^{-n}]} - 1_{[k2^{-n},(k+1)2^{-n}]})$$

for $\{(k,n)|n \geq 1, k < 2^n,\}$ and $f_{0,0} = 1$.

2) Take a family $g_\iota = g_{k,n}$ for $(k,n) \in I = \{(k,n)|n \geq 1, k < 2^n, k \text{ odd}\} \cup \{(0,0\}$ of independent Gaussian random variables.

3) Define

$$B_t = \sum_{\iota \in I} g_\iota \int_0^t f_\iota \ .$$

4) Prove convergence of the above series. Here, one has to do some work.

5) Check

$$E[B_s B_t] = \sum_{\iota \in I} \int_0^s \int_0^t f_\iota f_\iota = \int_0^1 1_{[0,s]} 1_{[0,t]} = \inf\{s,t\} \ .$$

6) Extend the definition from $t \in [0,1]$ to $t \in [0,\infty)$ by taking independent Brownian motions $B_t^{(i)}$ and defining $B_t = \sum_{n<[t]} B_{t-n}^{(n)}$, where $[t]$ is the largerst integer smaller or equal to t.

3.5 Some properties of Brownian motion

We consider now some symmetries of Brownian motion.

Proposition 3.5.1 *(i)* **Time-homogeneity***: For any $s > 0$, the process $\tilde{B}_t = B_{t+s} - B_s$ is a Brownian motion indpendent of $\sigma(B_u, u \leq s)$.*

(ii) **Reflection symmetry***: The process $\tilde{B}_t = -B_t$ is a Brownian motion.*

(iii) **Brownian scaling***: For every $c > 0$, the process $\tilde{B}_t = cB_{t/c^2}$ is a Brownian motion.*

(iv) **Time inversion***: The process $\tilde{B}_0 = 0, \tilde{B}_t = tB_{1/t}, t > 0$ is a Brownian motion.*

Proof.
(i),(ii),(iii) In each case, $\tilde{B}_t$ is a continuous centered Gaussian process with continuous paths, independent increments and variance t.

(iv) $\tilde{B}$ is a centered Gaussian process with covariance

$$\mathrm{Cov}[\tilde{B}_s, \tilde{B}_t] = E[\tilde{B}_s, \tilde{B}_t] = st \cdot E[B_{1/s}, B_{1/t}] = st \cdot \inf(1/s, 1/t) = \inf(s,t) \; .$$

Continuity of $\tilde{B}_t$ is clear for $t > 0$ and we have to check continuity only for $t = 0$: but since $E[\tilde{B}_s^2] = s \to 0$ for $s \to 0$, we know that $\tilde{B}_s \to 0$ almost everywhere. □

It follows the **strong law of large numbers for Brownian motion**:

Corollary 3.5.2 (SLLN for Brownian motion)
$\lim_{t\to\infty} t^{-1} B_t = 0$ *almost surely.*

Proof. From the time inversion property (iv), we see that $t^{-1}B_t = B_{1/t}$ which converges for $t \to \infty$ to 0 almost everywhere, because of the a.e. continuity of B_t. □

An other property of Brownian motion one has to mention is its uniqueness. Two processes X_t on $(\Omega, \mathcal{A}, P)$ and X'_t on $(\Omega', \mathcal{A}', P')$ are **indistinguishable**, if there exists an isomorphism $\phi : \Omega \to \Omega'$ of probability spaces, such that $X'_t(\phi\omega) = X_t(\omega)$. Indistinguishable processes are considered the same. A special case is if the two processes are defined on the same probability space $(\Omega, \mathcal{A}, P)$ and $X_t(\omega) = Y_t(\omega)$ for almost all ω.

Proposition 3.5.3 *Brownian motion is unique in the sense that two Brownian motions are indistinguishable.*

Proof. The construction of the map $H \to \mathcal{L}^2$ was unique in the sense that if we construct two different processes $X(h)$ and $Y(h)$, then there exists an isomorphism ϕ of the probability space such that $X(h) = Y(h)(\phi)$. The continuity of X_t and Y_t implies then that for almost all ω, $X_t(\omega) = Y_t(\phi\omega)$. In other words, they are indistinguishable. □

Definition. A curve $t \in [0, \infty) \mapsto X_t \in \mathbf{R}^n$ is called **Hölder continuous of order** α if there exists a constant C such that

$$||X_{t+h} - X_t|| \leq C \cdot h^\alpha$$

for all $h > 0$ and all t. It is called **locally Hölder continuous of order** α if there exists for each t a constant $C = C(t)$ such that

$$||X_{t+h} - X_t|| \leq C \cdot h^\alpha$$

for all small enough h. For a $\mathbf{R}^d$-valued stochastic process, Hölder continuity (or local Hölder continuity holds) if for almost all $\omega \in \Omega$ the sample path $X_t(\omega)$ is Hölder (rsp. locally Hölder) continuous for almost all $\omega \in \Omega$.

Proposition 3.5.4 *Brownian motion has a modifiction which is locally Hölder continuous of order α for every $\alpha < 1/2$.*

Proof. It is enough to show it in one dimension since a vector consisting of locally Hölder continuous functions is locally Hölder continuous. Since increments of Brownian motion are Gaussian, we have

$$E[(B_t - B_s)^{2p}] = C_p \cdot |t - s|^p$$

for some constant C_p. Kolmogorov's lemma assures the existence of a modifiction satisfying locally

$$|B_t - B_s| \leq C\ |t - s|^\alpha, 0 < \alpha < (p-1)/(2p)\ .$$

Since we can take p arbitrary large, the result follows. □

Proposition 3.5.5 (Wiener) *Brownian motion is nowhere differentiable: for almost all ω, $t \mapsto X_t(\omega)$ is nowhere differentiable.*

Proof. It is enough to show it in one dimensions.
Suppose B_t is differentiable at some point $0 \leq s \leq 1$. There exists then an integer l such that $|B_t - B_s| \leq l(t-s)$ for $t - s > 0$ small enough. But this means that

$$|B_{j/n} - B_{(j-1)/n}| \leq 7l/n$$

for $i = [ns] + 1 \leq j \leq [ns] + 4 = i + 3$ and sufficiently large n so that the set of differentiable paths is included in the set

$$B = \bigcup_{l \geq 1} \bigcup_{m \geq 1} \bigcap_{n \geq m} \bigcup_{0 < i \leq n+1} \bigcap_{i < j \leq i+3} \{|B_{j/n} - B_{(j-1)/n}| < 7\frac{l}{n}\} .$$

Using Brownian scaling, we show that $P(B) = 0$ as follows

$$\begin{aligned}
& P[\bigcap_{n \geq m} \bigcup_{0 < i \leq n+1} \bigcap_{i < j \leq i+3} \{|B_{j/n} - B_{(j-1)/n}| < 7\frac{l}{n}\}] \\
\leq \ & \liminf_{n \to \infty} nP[|B_{1/n}| < 7l/n]^3 \\
= \ & \liminf_{n \to \infty} nP[|B_1| < 7l/\sqrt{n}]^3 \\
\leq \ & \lim_{n \to \infty} Cn^{-1/2} = 0 .
\end{aligned}$$

□

Remark. This proposition shows especially that we have no Lipshitz continuity. A slight generalisation shows that Brownian motion is not Hölder continuous for $\alpha \geq 1/2$. One has just to do the same trick with k instead of 3 steps where $k(\alpha - 1/2) > 1$.
The right modulus of continuity is very near to $\alpha = 1/2$: $|B_t - B_{t+\epsilon}|$ is of the order

$$h(\epsilon) = \sqrt{2\epsilon \log(1/\epsilon)} .$$

More precicely (without proof).

Theorem 3.5.6 (Lévy's modulus of continuity)
$P[\limsup_{\epsilon \to 0} \sup_{|s-t| \leq \epsilon} \frac{|B_s - B_t|}{h(\epsilon)} = 1] = 1$.

Convention. We will assume from now on that all the paths of Brownian motion are locally Hölder continuous of order $\alpha < 1/2$.

3.6 Other Brownian processes

The covariance of standard Brownian motion was given by $E[B_s B_t] = \min\{s,t\}$. We constructed it by implementing the Hilbert space $L^2([0,\infty)$ as a Gaussian subspace of $\mathcal{L}^2(\Omega,\mathcal{A},P)$. We construct in this section a more general class of Gaussian processes.

Definition. A real-valued process $X_t, t \in \mathbf{T}$ is called a **Gaussian process** if for any finite set $\{t_1,\dots,t_n\} \subset \mathbf{T}$, the vector $(X_{t_1},\dots,X_{t_n})$ is Gaussian. The process is **centered**, if $E[X_t] = 0$ for all $t \in \mathbf{T}$. The **covariance** of a centered Gaussian process is $V(s,t) = E[X_s X_t]$.

Definition. A function $V : \mathbf{T} \times \mathbf{T}$ is called **positive semidefinit**, if for all finite sets $\{t_1,\dots,t_d\} \subset \mathbf{T}$, the matrix $V_{ij} = V(t_i,t_j)$ satisfies $(u,Vu) \geq 0$ for all vectors $u = (u_1,\dots,u_n)$.

Proposition 3.6.1 *The covariance of a centered Gaussian process is positive semidefinit. Any positive semidefinite function V on $\mathbf{T} \times \mathbf{T}$ is the covariance of a centered Gaussian process X_t.*

Proof. The first statement follows from the fact that for all $u = (u_1,\dots,u_n)$

$$\sum_{i,j} V(t_i,t_j) u_i u_j = E[(\sum_{i=1}^n u_i X_{t_i})^2] \geq 0 \; .$$

Introduce for $t \in \mathbf{T}$ a formal symbol δ_t. Consider the vector space of finite sums $\sum_{i=1}^n a_i \delta_{t_i}$ with inner product

$$(\sum_{i=1}^d a_i \delta_{t_i}, \sum_{j=1}^d b_j \delta_{t_j}) = \sum_{i,j} a_i b_j V(t_i,t_j) \; .$$

This is a positive semidefinite inner product. Quotiening out the null vectors $\{||v|| = 0\}$ and completion gives a separable Hilbert space H. Define now as in step 1 of the construction of Brownian motion the process $X_t = X(\delta_t)$. Since the map $X : H \to \mathcal{L}^2$ is preserving the scalar product, we have $E[X_t, X_s] = (\delta_s, \delta_t) = V(s,t)$. □

A large class of Gaussian processes can now be obtained. We give some examples.

1) Ornstein-Uhlenbeck (osciallator) process
This is a one dimensional process used to describe the quantum mechanical oscillator. Let $\mathbf{T} = \mathbf{R}$ and take the function $V(s,t) = \frac{1}{2} e^{-|t-s|}$ on $\mathbf{T} \times \mathbf{T}$.

Claim: V is positive semidefinit.

Proof. Calculate the Fourier transform $\int_{\mathbf{R}} e^{ikt} e^{-|t|} dt = \frac{1}{2\pi(k^2+1)}$. By Fourier inversion, we get

$$(2\pi)^{-1} \int_{\mathbf{R}} (k^2+1)^{-1} e^{ik(t-s)} \, dk = \frac{1}{2} e^{-|t-s|} ,$$

and so

$$\begin{aligned} 0 &\leq (2\pi)^{-1} \int_{\mathbf{R}} (k^2+1)^{-1} \sum_j |u_j e^{ikt_j}|^2 \, dk \\ &= \sum_{j,k=1}^{n} u_j u_k \frac{1}{2} e^{-|t_j - t_k|} . \end{aligned}$$

□

Claim. This process has a continuous modification.

Proof. Compute

$$E[(X_t - X_s)^2] = (e^{-|t-t|} + e^{-|s-s|} - 2e^{-|t-s|})/2 = (1 - e^{-|t-s|}) \leq |t-s| .$$

We use Kolmogorov's criterium. □

There is a relation between the Brownian motion B_t and the Ornstein-Uhlenbeck process. We claim that

$$X_t = 2^{-1/2} e^{-t} B_{e^{2t}}$$

for $t \geq 0$ and $X_{-t} = X_t$ is the Ornstein-Uhlenbeck process Y_t.

Proof. Both X and Y are centered Gaussian, continuous processes with independent increments. We show that they have the same covariance

$$E[X_t X_s](1/2) e^{-t} e^{-s} \min\{e^{2t}, e^{2s}\} = e^{|s-t|}/2 .$$

□

It follows from this relation that also the Ornstein-Uhlenbeck process is not differentiable almost everywhere.

2) Generalized Ornstein-Uhlenbeck processes.
In 1), we had $V(s,t) = \int_{\mathbf{R}} e^{-ik(t-s)} \, d\mu(k) = \hat{\mu}(t-s)$ with the Cauchy measure $\mu = \frac{1}{2\pi(k^2+1)} dx$ on $\mathbf{R}$. This can be generalized. Take any symmetric measure μ on $\mathbf{R}$ and let $\hat{\mu}$ be its Fourier transform $\int_{\mathbf{R}} e^{-ikt} \, d\mu(k)$. The same calculation as above shows that the function $V(s,t) = \hat{\mu}(t-s)$ is positive semidefinite.

3) Brownian bridge (tied down process) .
This is a one dimensional process with time $\mathbf{T} = [0,1]$ and $V(s,t) = s(1-t)$ for $1 \leq s \leq t \leq 1$ and $V(s,t) = V(t,s)$ else. In order to show that V is positive semidefinite, one observes that $X_t = B_s - sB_1$ is a Gaussian process, which has the covariance

$$E[X_s X_t] = E[(B_s - sB_1)(B_t - tB_1)] = s + st - 2st = s(1-t) .$$

Since $E[X_1^2] = 0$, we have $X_1 = 0$ which means that all paths go from 0 at time 0 to 1 at time 1.
The realisation $X_t = B_s - sB_1$ shows also that X_t has a continuous realisation.
Let X_t be the Brownian bridge and let y be a point in $\mathbf{R}^d$. We can consider the Gaussian process $Y_t = ty + X_t$ which discribes paths going from 0 at time 0 to y at time 1. The process Y has however no more zero mean.
Brownian motion B and Brownian bridge X can be translated into each other by

$$B_t = \tilde{B}_t := (t+1)X_{t/(t+1)}, \ , X_t = \tilde{X}_t := (1-t)B_{t/(1-t)} .$$

These identities follow from the fact that both are continuous centered Gaussian processes having the right covariance

$$\begin{aligned} E[\tilde{B}_s \tilde{B}_t] &= (t+1)(s+1)\min\{t/(t+1), s/(s+1)\} = \min\{s,t\} = E[B_s B_t] , \\ E[\tilde{X}_s \tilde{X}_t] &= (1-t)(1-s)\min\{s/(1-s), t/(1-t)\} = s(1-t) = E[X_s X_t] . \end{aligned}$$

4) Great disorder.
Take $V(s,t) = 1_{\{s=t\}}$. This Gaussian process has the property that X_s and X_t are independent if $s \neq t$. This process can however not be modified so that $(t,\omega) \mapsto X_t(\omega)$ is measurable.

Proof. Assume, $(t,\omega) \mapsto X_t(\omega)$ is measureable. Then $Y_t = \int_0^t X_s \ ds$ is measurable. But then

$$E[Y_t^2] = E[(\int_0^t X_s)^2] = \int_0^t ds \int_0^{t'} ds' E[X_{s'} X_{t'}] = 0$$

which gives $Y_t = 0$ almost everywhere so that the measure $d\mu(\omega) = X_s(\omega) ds$ is zero for almost all ω.

$$t = E[\int_0^t X_s^2] = E[\int_0^t X_s X_s \ ds] = E[\int_0^t X_s \ d\mu(s)] = 0 .$$

□

5) Brownian sheet.
This is not a stochastic process but a random field since the time $\mathbf{T} = \mathbf{R}_+^2$ is two dimensional. As long as we deal only with Gaussian random variables and not with regularity questions, the time $\mathbf{T}$ can be quite arbitrary and the proposition 3.6.1 at the beginning of this section holds true. The Gaussian process with

$$V((s_1,s_2),(t_1,t_2)) = \min(s_1,t_1) \cdot \min(s_2,t_2)$$

is called Brownian sheet. It has similar scaling properties as Brownian motion.

3.7 The Wiener measure

Let $(E,\mathcal{E})$ be a measurable space and $\mathbf{T}$ a set. A stochastic process on a probablity space $(\Omega,\mathcal{A},P)$ indexed by $\mathbf{T}$ and with values in E defines a map

$$\phi : \Omega \to E^{\mathbf{T}}, \omega \mapsto X_t(\omega) .$$

Let $E^{\mathbf{T}}$ be equiped with the product σ-algebra $\mathcal{E}^{\mathbf{T}}$, which is the smallest algebra for which all the functions X_t are measureable which is the σ-algebra generated by the π-system

$$\{ \prod_{t_1,\dots,t_n}^{n} A_{t_i} = \{x \in E^{\mathbf{T}}, x_{t_i} \in A_{t_i}\} \mid A_{t_i} \in \mathcal{E}\}$$

consisting of zylinder sets. Denote with $Y_t(w) = w(t)$ the coordianate mappings on $E^{\mathbf{T}}$. Since $Y_t \circ \phi$ is measurable for all t, also ϕ is measurable. Denote with P_X the push forward measure of ϕ from $(\Omega,\mathcal{A})$ to $(E^{\mathbf{T}},\mathcal{E}^{\mathbf{T}})$. For any finite set $(t_1,\dots,t_n) \subset \mathbf{T}$ and all sets $A_i \in \mathcal{E}$, we have

$$P[X_{t_i} \in A_i, i = 1,\dots,n] = P_X[Y_{t_i} \in A_i, 1 = 1,\dots n] .$$

One says, the two processes X and Y are **versions of each other**.

Definition. Y is called the **coordinate process** of X and the probability measure P_X is called the **law of** X.

Two processes X, X' possibly defined on different probability spaces are versions of each other if and only if they have the same law $P_X = P_{X'}$.

One usually does not work with the coordinate process but prefers to work with processes which have some continuity properties. many processes have versions which are right continuous and have left hand limits at every point.

Let D be a subset of $E^{\mathbf{T}}$ and assume the process has a version X such that almost all paths $X(\omega)$ are in D. Define the probability space $(D,\mathcal{E}^{\mathbf{T}} \cap D, Q)$, where Q is the measure $Q = \phi^* P$ is the law of X. Obviously, the process Y defined on $(D,\mathcal{E}^{\mathbf{T}} \cap D, Q)$ is another version of X. If D is right continuous with left hand limits, the process is called the **canonical version** of X.

Corollary 3.7.1 *Let $E = \mathbf{R}^d$ and $\mathbf{T} = \mathbf{R}^+$. There exists a unique probability measure W on $C(\mathbf{T},E)$ for which the coordinate process Y is the Brownian motion B.*

Proof. Let $D = C(\mathbf{T},E) \subset E^{\mathbf{T}}$. Define the measure $W = \phi^* P_X$ and let Y be the coordinate process of B. Uniqueness: assume we have two such measures

W, W' and let Y, Y' be the coordinate processes of B on D with respect to W and W'. Since both Y and Y' are versions of X and "being a version" is an equivalence relation, they are also versions of each other. This means that W and W' coincide on a π- system and are therefore the same. $\square$

Definition. If $E = \mathbf{R}^d$ and $\mathbf{T} = [0, \infty)$, the measure W on $C(\mathbf{T}, E)$ is called the **Wiener measure** and the probability space $(C(\mathbf{T}, E), \mathcal{E}^{\mathbf{T}} \cap C(\mathbf{T}, E), W)$ is called the **Wiener space**.

Let $\mathcal{B}'$ be the σ-algebra $\mathcal{E}^{\mathbf{T}} \cap C(\mathbf{T}, E)$, which is the Borel σ-algebra restricted to $C(\mathbf{T}, E)$. The space $C(\mathbf{T}, E)$ carries an other σ-algebra, namely the Borel $\sigma-$ algebra $\mathcal{B}$ generated by its own topology. We have $\mathcal{B} \subset \mathcal{B}'$, since all closed balls $\{f \in C(\mathbf{T}, E) \mid |f - f_0| \leq r\} \in \mathcal{B}$ are in $\mathcal{B}'$. The other relation $\mathcal{B}' \subset \mathcal{B}$ is clear so that $\mathcal{B} = \mathcal{B}'$. The Wiener measure is therefore a **Borel measure**.

Remark. The Wiener measure can also be constructed without having Brownian motion and can be used to define Brownian motion. We sketch the idea. Let $S = \dot{\mathbf{R}}^n$ denote the one point compactification of $\mathbf{R}^n$. Define $\Omega = S^{[0,t]}$ be the set of functions from $[0, t]$ to S which is also the set of paths in $\overline{\mathbf{R}}^n$. It is by Tychonov a compact space with the product topology. Define

$$C_{fin}(\Omega) = \{\phi \in C(\Omega, \mathbf{R}) \mid \phi(\omega) = F(\omega(t_1), \dots, \omega(t_n))\} .$$

Define also the **Gauss kernel** $p(x, y, t) = (4\pi t)^{-n/2} \exp(-|x - y|^2/4t)$. Define on $C_{fin}(\Omega)$ the functional

$$L\phi = \int_{(\mathbf{R}^n)^m} F(x_1, x_2, \dots, x_m) p(0, x_1, s_1) p(x_1, x_2, s_2) \dots p(x_{m-1}, x_m, s_m) \, dx$$

with $s_1 = t_1$ and $s_k = t_k - t_{k-1}$ for $k \geq 2$. Since $L(\phi) \leq |\phi(\omega)|_\infty$, it is a bounded linear functional on the dense linear subspace $C_{fin}(\Omega) \subset C(\Omega)$. It is nonnegative and $L(1) = 1$. By Hahn Banach, it extends uniquely to a bounded linear functional on $C(\Omega)$. By Riesz representation theorem, there exists a unique measure μ on $C(\Omega)$ such that $L(\phi) = \int \phi(\omega) \, d\mu(\omega)$. This is the Wiener measure on Ω.

3.8 Lévy's modulus of continuity

Remember the elementary inequalities

Lemma 3.8.1

$$a^{-1}e^{-a^2/2} > \int_a^\infty e^{-b^2/2} \, db > \frac{a}{a^2+1}e^{-a^2/2} \, .$$

Proof.

$$\int_a^\infty e^{-b^2/2} \, db < \int_a^\infty e^{-b^2/2}(b/a) \, db = a^{-1}e^{-a^2/2} \, .$$

For the right inequality consider

$$\int_a^\infty b^{-2}e^{-b^2/2} \, db < a^{-2}\int_a^\infty e^{-b^2/2} \, db \, .$$

Integrating by parts of the left hand side of this gives

$$a^{-1}e^{-a^2/2} - \int_a^\infty e^{-b^2/2} \, db < a^{-2}\int_a^\infty e^{-b^2/2} \, db \, .$$

□

Theorem 3.8.2 (Lévy's modulus of continuity)

$$P[\limsup_{\epsilon \to 0} \sup_{|s-t| \leq \epsilon} \frac{|B_s - B_t|}{h(\epsilon)} = 1] = 1 \, ,$$

where $h(\epsilon) = \sqrt{2\epsilon \log(1/\epsilon)}$.

Proof. (i) $\limsup \geq 1$.
Take $0 < \delta < 1$. Define $a_n = (1-\delta)h(2^{-n}) = (1-\delta)\sqrt{n 2 \log 2}$. Consider

$$P[A_n] = P[\max_{1 \leq k \leq 2^n} |B_{k2^{-n}} - B_{(k-1)2^{-n}}| \leq a_n] \, .$$

Since $B_{k2^{-n}} - B_{(k-1)2^n}$ are Gaussian and independent, we compute, using the above lemma and $1 - s < e^{-s}$

$$\begin{aligned} P[A_n] &\leq (1 - 2\int_{a_n}^\infty \frac{1}{2\pi}e^{-x^2/2} \, dx)^{2^n} \\ &\leq (1 - 2\frac{a_n}{a_n^2+1}e^{-a_n^2/2})^{2^n} \\ &\leq \exp(-2^n \frac{2a_n}{a_n^2+1}e^{-a_n^2/2}) \leq e^{-Ce^{n(1-(1-\delta)^2)/\sqrt{n}}} \, , \end{aligned}$$

where C is a constant independent of n. Since $\sum_n P[A_n] < \infty$, we get by Borel-Cantelli that $P[\limsup_n A_n] = 0$ so that

$$P[\lim_{n\to\infty} \max_{1\leq k\leq 2^n} |B_{k2^{-n}} - B_{(k-1)2^n}| \geq h(2^{-n})] = 1 \ .$$

(ii) $\limsup \leq 1$.
Take again $0 < \delta < 1$ and pick $\epsilon > 0$ such that $(1+\epsilon)(1-\delta) > (1+\delta)$. Define

$$\begin{aligned} P[A_n] &= P[\max_{k=j-i\in K} |B_{j2^{-n}} - B_{i2^{-n}}|/h(k2^{-n}) \geq (1+\epsilon)] \\ &= P[\bigcup_{k\in K} \{|B_{j2^{-n}} - B_{i2^{-n}}| \geq a_{n,k}\} \ , \end{aligned}$$

where

$$K = \{0 < k = j - i \leq 2^{n\delta}\}$$

and $a_{n,k} = h(k2^{-n})(1+\epsilon)$.
Using the above lemma, we get with some constants C (which may vary from line to line)

$$\begin{aligned} P[A_n] &\leq \sum_{k\in K} a_{n,k}^{-1} e^{-a_{n,k}^2/2} \\ &\leq C \cdot \sum_{k\in K} \log(k^{-1}2^n)^{-1/2} e^{-(1+\epsilon)^2 \log(k^{-1}2^n)} \\ &\leq C \cdot 2^{-n(1-\delta)(1+\epsilon)^2} \sum_{k\in K} (\log(k^{-1}2^n))^{-1/2} \qquad (\text{ since } \ k^{-1} > 2^{-n\delta}) \\ &\leq C \cdot n^{-1/2} 2^{n(\delta-(1-\delta)(1+\epsilon)^2)} \ . \end{aligned}$$

We have used in the last step that there are at most $2^{n\delta}$ points in K and for each of them $\log(k^{-1}2^n) > \log(2^n(1-\delta))$.
We see that $\sum_n P[A_n]$ converges and by Borel-Cantelli, we get for almost every ω an integer $n(\omega)$ such that for $n > n(\omega)$

$$|B_{j2^{-n}} - B_{i2^{-n}}| < (1+\epsilon) \cdot h(k2^{-n}) \ ,$$

where $k = j - i \in K$. Increase eventually $n(\omega)$ so that for $n > n(\omega)$

$$\sum_{m>n} h(2^{-m}) < \epsilon \cdot h(2^{-(n+1)(1-\delta)}) \ .$$

Pick $0 \leq t_1 < t_2 \leq 1$ such that $t = t_2 - t_1 < 2^{-n(\omega)(1-\delta)}$. Take next $n > n(\omega)$ such that $2^{-(n+1)(1-\delta)} \leq t < 2^{-n(1-\delta)}$ and write the dyadic developement of t_1, t_2:

$$t_1 = i2^{-n} - 2^{-p_1} - 2^{-p_2} \dots \ , t_2 = j2^{-n} + 2^{-q_1} + 2^{-q_2} \dots$$

with $t_1 \leq i2^{-n} < j2^{-n} \leq t_2$ and $0 < k = j - i \leq t2^n < 2^{n\delta}$. We get

$$\begin{aligned} |B_{t_2}(\omega) - B_{t_2}(\omega)| &\leq |B_{t_1} - B_{i2^{-n}}(\omega)| + |B_{i2^{-n}}(\omega) - B_{j2^{-n}}(\omega)| + |B_{j2^{-n}}(\omega) - B_{t_2}| \\ &\leq 2\sum_{p>n} (1+\epsilon)h(2^{-p}) + (1+\epsilon)h(k2^{-n}) \\ &\leq (1 + 3\epsilon + 2\epsilon^2)h(t) \ . \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the proof is complete. □

3.9 Stopping times

We repeat the notion of stopping times for continuous time stochastic processes and martingales. Stopping times are useful; for example for the construction of new processes, in proofs of inequalities and convergence theorems as well as in the study of return time results.

Definition. A **filtration** of a measurable space $(\Omega, \mathcal{A})$ is an increasing family $(\mathcal{A}_t)_{t\geq 0}$ of sub-σ algebras of $\mathcal{A}$. A measurable space endowed with a filtration $(\mathcal{A}_t)_{t\geq 0}$ is called a **filtered space**. A process X is called **adapted** to the filtration $\mathcal{A}_t$, if X_t is $\mathcal{A}_t$-measurable for all t.

A process X on $(\Omega, \mathcal{A}, P)$ defines a natural filtration $\mathcal{A}_t = \sigma(X_s \mid s \leq t)$, the **minimal filtration** of X for which X is adapted.

Interpretation. Heuristically, $\mathcal{A}_t$ is the set of events, which may occur up to time t.

Definition. With a filtration we can associate two other filtration by setting for $t > 0$

$$\mathcal{A}_{t^-} = \sigma(\mathcal{A}_s, s < t), \mathcal{A}_{t^+} = \bigcap_{s>t} \mathcal{A}_s \ .$$

For $t = 0$ we can still define $\mathcal{A}_{0^+} = \bigcap_{s>0} \mathcal{A}_s$ and define $\mathcal{A}_{0^-} = \mathcal{A}_0$. Define also $\mathcal{A}_\infty = \sigma(\mathcal{A}_s, s \geq 0)$.

Remark. We have always $\mathcal{A}_{t^-} \subset \mathcal{A}_t \subset \mathcal{A}_{t^+}$ and both inclusions may be strict.

Definition. If $\mathcal{A}_t = \mathcal{A}_{t^+}$ then the filtration $\mathcal{A}_t$ is called **right continuous**. If $\mathcal{A}_t = \mathcal{A}_{t^-}$, then $\mathcal{A}_t$ is **left continuous**. As an example, the filtration $\mathcal{A}_{t^+}$ of any filtration is right continuous.

Definition. A **stopping time** relative to a filtration $\mathcal{A}_t$ is a map $T : \Omega \to [0, \infty]$ such that $\{T \leq t\} \in \mathcal{A}_t$.

Remark. If $\mathcal{A}_t$ is right continuous, then T is a stopping time if and only if $\{T < t\} \in \mathcal{A}_t$. Also T is a stopping time if and only if $X_t = 1_{(0,T]}(t)$ is adapted. X is then a left continuous adapted process.

Definition. If T is a stopping time, define

$$\mathcal{A}_T = \{A \in \mathcal{A}_\infty \mid A \cap \{T \leq t\} \in \mathcal{A}_t, \forall t\} \ .$$

It is a σ-algebra. As an example, if $T = s$ is constant, then $\mathcal{A}_T = \mathcal{A}_s$. Note also that

$$\mathcal{A}_{T^+} = \{A \in \mathcal{A}_\infty \mid A \cap \{T < t\} \in \mathcal{A}_t, \forall t\} \ .$$

We give examples of stopping times.

Proposition 3.9.1 *Let X be the coordinate process on $C(\mathbf{R}_+, E)$, where E is a metric space. Let A be a closed set in E. Then the so called* **entry time**

$$T_A(\omega) = \inf\{t \geq 0 \mid X_t(\omega) \in A\}$$

is a stopping time relative to the filtration $\mathcal{A}_t = \sigma(\{X_s\}_{s \leq t})$.

Proof. Let d be the metric on E. We have

$$\{T_A \leq t\} = \{\inf_{s \in \mathbf{Q}, s \leq t} d(X_s(\omega), A) = 0\}$$

which is in $\mathcal{A}_t = \sigma(X_s, x \leq t)$. □

Proposition 3.9.2 *Let X be the coordinate process on $D(\mathbf{R}_+, E)$, the space of right continuous functions, where E is a metric space. Let A be an open subset of E. Then the* **hitting time**

$$T_A(\omega) = \inf\{t > 0 \mid X_t(\omega) \in A\}$$

is a stopping time with respect to the filtration $\mathcal{A}_{t+}$.

Proof. T_A is a $\mathcal{A}_{t+}$ stopping time if and only if $\{T_A < t\} \in \mathcal{A}_t$ for all t. If A is open and $X_s(\omega) \in A$, we know by the right-continuity of the paths that $X_t(\omega) \in A$ for every $t \in [s, s + \epsilon)$ for some $\epsilon > 0$. Therefore

$$\{T_A < t\} = \{\inf_{s \in \mathbf{Q}, s < t} X_s \in A\} \in \mathcal{A}_t \; .$$

□

Definition. Let $\mathcal{A}_t$ be a filtration on $(\Omega, \mathcal{A})$ and let T be a stopping time. For a process X, we define a new mapping X_T on the set $\{T < \infty\}$ by

$$X_T(\omega) = X_{T(\omega)}(\omega) \; .$$

We have met this definition already in the case of discrete time but here it is not clear whether X_T is measurable! It turns out that this is true for many processes.

Definition. A process X is called **progressively measurable** with respect to a filtration $\mathcal{A}_t$ if for all t, the map $(s, \omega) \mapsto X_s(\omega)$ from $([0, t] \times \Omega, \mathcal{B}([0, t] \times \mathcal{A}_t)$ into $(E, \mathcal{E})$ is measurable.

A progressively measurable process is adapted. For some processes, the inverse holds:

Lemma 3.9.3 *An adapted process with right or left continuous paths is progressively measurable.*

Proof. Assume right continuity (the argument is similar in the case of left continuity). Write X as the coordinate process $D([0,t],E)$. Denote the map $(s,\omega) \mapsto X_s(\omega)$ with $Y = Y(s,\omega)$. Given a closed ball $U \in \mathcal{E}$. We have to show that $Y^{-1}(U) = \{(s,\omega) \mid Y(s,\omega) \in U\} \in \mathcal{B}([0,t]) \otimes \mathcal{A}_t$. Given $k = \mathbf{N}$, we define $E_{0,U} = 0$ and inductively for $k \geq 1$ the k'th hitting time (a stopping time)

$$H_{k,U}(\omega) = \inf\{s \in \mathbf{Q} | E_{k-1,U}(\omega) < s < t,\ X_s \in U\}$$

as well as the k'th exit time (not necessarily a stopping time)

$$E_{k,U}(\omega) = \inf\{s \in \mathbf{Q} | H_{k,U}(\omega) < s < t,\ X_s \notin U\} \ .$$

These are countably many measurable maps from $D([0,t],E)$ to $[0,t]$. Then by the right-continuity

$$Y^{-1}(U) = \bigcup_{k=1}^{\infty} \{(s,\omega) | H_{k,U}(\omega) \leq s \leq E_{k,U}(\omega)\}$$

which is in $\mathcal{B}([0,t]) \otimes \mathcal{A}_t$. □

Proposition 3.9.4 *If X is progressively measurable and T is a stopping time, then X_T is $\mathcal{A}_T$-measurable on the set $\{T < \infty\}$.*

Proof. The set $\{T < \infty\}$ is itself in $\mathcal{A}_T$. To say that X_T is $\mathcal{A}_T$- measurable on this set is equivalent with $X_T \cdot 1_{\{T \leq t\}} \in \mathcal{A}_t$ for every t. But the map

$$T : (\{T \leq t\}, \mathcal{A}_t \cap \{T \leq t\}) \to ([0,t], \mathcal{B}[0,t])$$

is measurable because T is a stopping time. This means that the map $\omega \mapsto (T(\omega),\omega)$ from $(\Omega, \mathcal{A}_t)$ to $([0,t] \times \Omega, \mathcal{B}[0,t] \otimes \mathcal{A}_t)$ is measurable and X_T is the composition of this map with X which is $\mathcal{B}[0,t] \otimes \mathcal{A}_t$ measurable by hypothesis.
□

Definition. Given a stopping time T and a process X, we define the **stopped process** $X^T_t(\omega) = X_{T \wedge t}(\omega)$.

If $\mathcal{A}_t$ is a filtration then $\mathcal{A}_{t \wedge T}$ is a filtration since if T_1 and T_2 are stopping times, then $T_1 \wedge T_2$ is a stopping time.

Corollary 3.9.5 *If X is progressively measurable with respect to $\mathcal{A}_t$ and T is a stopping time, then $X^T = X_{t \wedge T}$ is progressively measurable with respect to $\mathcal{A}_{t \wedge T}$.*

Proof. Since $t \wedge T$ is a stopping time, we have from the previous proposition that X^T is $\mathcal{A}_{t\wedge T}$ measurable.
We know by assumption that $\phi : (s, \omega) \mapsto X_s(\omega)$ is measurable. Since also $\psi : (s, \omega) \mapsto (s \wedge T)(\omega)$ is measurable, we know also that the composition $(s, \omega) \mapsto X_T(\omega) = X_{\psi(s,\omega)}(\omega) = \phi(\psi(s, \omega), \omega)$ is measurable. □

Proposition 3.9.6 *Every stopping time is the decreasing limit of a sequence of stopping times taking only finitely many values.*

Proof. Given a stopping time T. Define $T_k = +\infty$ if $T \geq k$ and $T_k = q2^{-k}$ if $(q-1)2^{-k} \leq T < q2^{-k}, q < 2^k k$.
Each T_k is a stopping time and T_k decreases to T. □

3.10 Overview: relation with potential theory

Many concepts of classical potential theory can be expressed in an elegant form in a probabilistic language. We give very briefly some examples without proofs, but some hints to the literature.

Let B_t be Brownian motion in $\mathbf{R}^d$ and T_A the hitting time of a set $A \subset \mathbf{R}^d$. Let D be a domain in $\mathbf{R}^d$ with boundary $\delta(D)$ such that the Green function $G(x,y)$ exists in D (such a domain is then called a **Green domain**).

1) The Green function of a domain D is defined as the fundamental solution satisfying $\Delta G(x,y) = \delta(x-y)$, where $\delta(x-y)$ is the Dirac measure at $y \in D$. Having the fundamental solution G, we can solve $\Delta u = v$ for a given function v by

$$u = \int_D G(x,y) \cdot v(y) \, dy .$$

The Green function can be computed using Brownian motion as follows:

$$G(x,y) = \int_0^\infty g(t,x,y) \, dt ,$$

where

$$\int_C g(x,y) \, dy = P_x[B_t \in C, T_{\delta D} > t], x, y \in D$$

with P_x, the Wiener measure of B_t starting at the point x. We can interpret that as follows. To determine $G(x,y)$, consider the killed Brownian motion B^T starting at x, where T is the hitting time of the boundary. $G(x,y)$ is then the probability density, of the particles discribed by the Brownian motion.

2) The classical Dirichlet problem for a bounded Green domain D. Given a function $f \in C(\delta(D))$, find $u \in C(\overline{D})$ such that $\Delta u = 0$ inside D and $\lim_{x \to y, x \in D} u(x) = f(y)$ for every $y \in \delta D$. This problem can not be solved in general if $d \geq 3$ (an example to this is Lebesgue's thorn found by Lebesgue in 1913). There is a nice description of this problem in the book K.Ito, H.P.McKean, p 262. take a spherical surface and push a sharp thorn into its side. Think of the inside as a chamber D, where the wall δD is held on constant temperature f, where $f = 1$ at the tip of the thorn y and zero except in a small neighborhood of y. The temperature u inside D is a solution of the Dirichlet problem $\Delta_D u = 0$ satisfying the boundary condition $u = f$ on the boundary δD. But the heat radiated from the thorn is proportinal to its surface area. if the tip is sharp enough, a person sitting in the chamber will be cold no matter how close he (she) is to the heater. This means $\liminf_{x \to y, x \in D} u(x) < 1 = f(y)$.

Because of this problem, one has to modify the problem and one says, u is a **solution of a modified Dirichlet problem**, if u satisfies $\Delta_D u = 0$ inside D and $\lim_{x \to y, x \in D} u(x) = f(y)$ for all nonsingular points y in the boundary δD. Irregularity of a point y can be defined analytically but it is equivalent with

$P_y[T_{D^c} > 0] = 1$, which means that almost every Brownian particle starting at $y \in \delta D$ will return to δD after positive time.
The solution of the regularized Dirichlet problem can be expressed with Brownian motion B_t and the hitting time T of the boundary. The solution is

$$u(x) = E_x[f(B_T)] \ .$$

Ikeda has discoved that there exists also a probabilistic method for solving the classical von Neumann problem in the case $d = 2$. For more information about this consult Ito, McKean, p. 264 or the book S. Port, C.Stone, "Brownian motion and classical potential theory".

3) Given the Dirichlet Laplacian Δ_D of a bounded domain D. One can compute the heat flow $e^{-tH_D}u$ by the following formula

$$(e^{-tH_D}u)(x) = E_x[u(B_t); t < T] \ ,$$

where T is the hitting time of δD for Brownian motion B_t starting at x. Also for the von Neumann Laplacian, there exists such a path integral formula. The process is then not killed Brownian motion, but reflected Brownian motion.

4) Let B be a compact subset of a Green domain D. The hitting probability

$$p(x) = P_x[T_B < T_{\delta D}]$$

is the equilibrium potential of B relative to D.
We give a definition of the equilibrium potential later. Physically, the equilibrium potential is obtained by measuring the electrostatic potential, if one is grounding the conducting boundary and charging the conducting set B with a unit amount of charge.

3.11 Martingales

Definition. Given a filtration $\mathcal{A}_t$ of the probability space $(\Omega, \mathcal{A}, P)$. A real-valued process $X_t \in \mathcal{L}^1$ which is $\mathcal{A}_t$ adapted is called a **submartingale**, if $E[X_t|\mathcal{A}_s] \geq X_s$, a **supermartingale** if $-X$ is a submartingale and a **martingale**, if it is both a super and sub-martingale. If additionally $X_t \in \mathcal{L}^p$ for all t, we speak of $\mathcal{L}^p$ super or sub-martingales.

We have seen martingales for discrete time already in the last term. Brownian motion gives examples with continuous time.

Proposition 3.11.1 *Let B_t be standard Brownian motion. Then B_t, $B_t^2 - t$ and $e^{\alpha B_t - \alpha^2 t/2}$ are martingales.*

Proof. $B_t - B_s$ is independent of B_s. Therefore

$$E[B_t \mid \mathcal{A}_s] - B_s = E[B_t - B_s|\mathcal{A}_s] = E[B_t - B_s] = 0 \; .$$

Since by the taking out property $E[B_t B_s \mid \mathcal{A}_s] = B_s \cdot E[B_t \mid \mathcal{A}_s] = 0$, we get

$$\begin{aligned} E[B_t^2 - t \mid \mathcal{A}_s] - (B_s^2 - s) &= E[B_t^2 - B_s^2 \mid \mathcal{A}_s] - (t-s) \\ &= E[(B_t - B_s)^2 \mid \mathcal{A}_s] - (t-s) = 0 \; . \end{aligned}$$

Since Brownian motion begins at any time s new, we have

$$E[e^{\alpha(B_t - B_s)}|\mathcal{A}_s] = E[e^{\alpha B_{t-s}}] = e^{\alpha^2(t-s)/2}$$

from which

$$E[e^{\alpha B_t}|\mathcal{A}_s]e^{-\alpha^2 t/2} = E[e^{\alpha B_s}]e^{-\alpha^2 s/2}$$

follows. □

As in the discrete case, we remark:

Proposition 3.11.2 *If X_t is a $\mathcal{L}^p$-martingale, then $|X_t|^p$ is a submartingale for $p \geq 1$.*

Proof. Conditional Jensen gives

$$E[|X_t|^p|\mathcal{A}_s] \geq |E[X_t|\mathcal{A}_s]|^p = |X_s|^p \; .$$

□

Recall from the last term the notion

$$(\int C\, dX)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

of a **discrete stochastic integral** for a discrete stochastic process X and a previsible process C. We had the "system can't be beaten" result:

If X_n is a submartingale with respect to $\mathcal{A}_n$ and If C_n is a bounded nonnegative previsible process then $\int C\, dX$ is a supermartingale. The same statement is true for submartingales and martingales.

from which followed the result

If X is a submartingale (supermartingale) and T is a stopping time, then the stopped process X^T is a submartingale (supermartingale).

We give an other example of a Martingale which is not continuous, the **Poisson process**.

Let X_n be a sequence of IID exponential distributed random variables The probability density of X_n is $f(x) = e^{-cx}c$. Let $S_n = \sum_{k=1}^{n} X_k$. Define the process N_t with time $\mathbf{T} = \mathbf{R}^+$ as

$$N_t = \sum_{k=1}^{\infty} 1_{S_k \leq t} \; .$$

This process takes values in $\mathbf{N}$ and measures, how many jumps are necessary to reach t. This process has independent increments

$$N_t - N_s = \sum_{n=1}^{\infty} 1_{s < S_n \leq t} \; .$$

Moreover, N_t is Poisson distributed with parameter tc:

$$P[N_t = k] = \frac{(tc)^k}{k!} e^{-tc} \; .$$

The proof is done by starting with a Poisson distributed process N_t. Define then $S_n(\omega) = \{t\ N_t = n, N_{t-0} = n-1\}$ and show that $X_k = S_k - S_{k-1}$ are independent with exponential distribution. This fact is called the **interval theorem** and we don't prove it here.

Since $E[N_t] = ct$, it follows that $N_t - ct$ is a martingale with respect to the filtration $\mathcal{A}_t = \sigma(N_s, s \leq t)$. It is a right continuous process. We know therefore that it is progressively measurable and that for each stopping time T, also N^T is progressively measurable. The book of Kingman " Poisson processes" has more information about Poisson processes and the proof of the interval theorem (p. 39 in that book).

Poisson processes on the lattice $\mathbf{Z}^d$ are called **Brownian motion on the lattice** and can be used to discribe Feynmann-Kac formulas for **discrete Schrödinger operators**. The process is defined as follows: take X_t as above and define

$$Y_t = \sum_{k=1}^{\infty} Z_k 1_{S_k \leq t} ,$$

where Z_n are IID random variables taking values in $\{m \in \mathbf{Z}^d | |m| = 1\}$. This means that a particle stays at a lattice site for an exponential time and jumps then to one of the neighbors of n with equal probability.
In the book of Carmona-Lacroix about random Schrödinger operators you can read about the **Feynman-Kac formula** for discrete Schrödinger operators $H = H_0 + V$. Let P_n be the analogue of the Wiener measure on right continuous paths on the lattice and denote with E_n the expectation. The Feynman-Kac formula is

$$(e^{-itH} u)(n) = e^{2dt} E_n[u(X_t) i^{N_t} e^{-i \int_0^t V(X_s) \, ds}] .$$

This is exciting since one can deal in the discrete case with the evolution of the Schrödinger equation $i\dot{u} = Hu$ and has not to go to imaginary time and take the heat equation $\dot{u} = Hu$.

3.12 Doob inequalities

We have already proven in the last term the inequalities of Doob for discrete times $\mathbf{T} = \mathbf{N}$. By a limiting argument, they hold also for right contininuous submartingales.

Theorem 3.12.1 (Doob's submartingale inequality) *Let X be a non-negative right continuous submartingale with time $\mathbf{T} = [a,b]$. For any $\epsilon > 0$*

$$\epsilon \cdot P[\sup_{a \leq t \leq b} X_t \geq \epsilon] \leq E[X_b; \{\sup_{a \leq t \leq b} X_t \geq \epsilon\}] \leq E[X_b] .$$

Proof. Take a countable subset D of $\mathbf{T}$ and choose an increasing sequence D_n of finite sets such that $\bigcup_n D_n = D$. We know now that for all n

$$\epsilon \cdot P[\sup_{t \in D_n} X_t \geq \epsilon] \leq E[X_b; \{\sup_{t \in D_n} X_t \geq \epsilon\}] \leq E[X_b] .$$

since $E[X_t]$ is nondecreasing in t. Going to the limit $n \to \infty$ gives the claim with $\mathbf{T} = D$. Since X is right continuous, we get the claim for $\mathbf{T} = [a,b]$. □

One often applies this inequality to the non-negative submartingale $|X|$ if X is a martingale.

Theorem 3.12.2 (Doob's $\mathcal{L}^p$ inequality) *Fix $p > 1$ and q satisfying $p^{-1} + q^{-1} = 1$. Given a non-negative right-continuous submartingale X with time $\mathbf{T} = [a,b]$ which is bounded in $\mathcal{L}^p$. Then $X^* = \sup_{t \in \mathbf{T}} X_t$ is in $\mathcal{L}^p$ and satisfies*

$$||X^*||_p \leq q \cdot \sup_{t \in \mathbf{T}} ||X_t||_p .$$

Proof. Take a countable subset D of $\mathbf{T}$ and choose an increasing sequence D_n of finite sets such that $\bigcup_n D_n = D$.
We had

$$|| \sup_{t \in D_n} X_t|| \leq q \cdot \sup_{t \in D_n} ||X_t||_p .$$

Going to the limit gives

$$|| \sup_{t \in D} X_t|| \leq q \cdot \sup_{t \in D} ||X_t||_p .$$

Since D is dense and X is right continuous we can replace D by $\mathbf{T}$. □

The following inequality measures, how big is the probability that one dimensional Brownian motion will leave the cone $\{(t,x), |x| \leq a \cdot t\}$.

Corollary 3.12.3 (Exponential inequality) *Define* $S_t = \sup_{0 \leq s \leq t} B_s$. *We have for any* $a > 0$

$$P[S_t \geq a \cdot t] \leq e^{-a^2 t/2} .$$

Proof. We have seen in the previous section about martingales that $M_t = e^{\alpha B_t - \frac{\alpha^2 t}{2}}$ is a martingale. It is nonnegative. Since

$$\exp(\alpha S_t - \frac{\alpha^2 t}{2}) \leq \exp(\sup_{s \leq t} B_s - \frac{\alpha^2 t}{2}) \leq \sup_{s \leq t} \exp(B_s - \frac{\alpha^2 s}{2}) = \sup_{s \leq t} M_s ,$$

we get with Doob's submartingale inequality

$$\begin{aligned} P[S_t \geq at] &\leq P[\sup_{s \leq t} M_s \geq e^{\alpha a t - \frac{\alpha^2 t}{2}}] \\ &\leq \exp(-\alpha a t + \frac{\alpha^2 t}{2}) E[M_t] . \end{aligned}$$

Since $E[B_t] = E[B_0] = 1$ and $\inf_{\alpha > 0} \exp(-\alpha a t + \frac{\alpha^2 t}{2}) = \exp(-\frac{a^2 t}{2})$, the result follows. □

An other corollary of Doob's maximal inequality will also be useful.

Corollary 3.12.4 *For* $a, b > 0$,

$$P[\sup_{s \in [0,1]} (B_s - \frac{\alpha s}{2}) \geq \beta] \leq e^{-\alpha\beta} .$$

Proof.

$$\begin{aligned} P[\sup_{s \in [0,1]} (B_s - \frac{\alpha s}{2}) \geq \beta] &\leq P[\sup_{s \in [0,1]} (B_s - \frac{\alpha t}{2}) \geq \beta] \\ &= P[\sup_{s \in [0,1]} (e^{\alpha B_s - \frac{\alpha^2 t}{2}}) \geq e^{\beta\alpha}] \\ &= P[\sup_{s \in [0,1]} M_s \geq e^{\beta\alpha}] \\ &\leq e^{-\beta\alpha} \sup_{s \in [0,1]} E[M_s] = e^{-\beta\alpha} \end{aligned}$$

since $E[M_s] = 1$ for all s. □

3.13 Kintchine's law of the iterated logarithm

Khinchine's law of the iterated logarithm for Brownian motion is a very precise statement about how one dimensional Brownian motion oscillates in a neighborhood of the origin. Define

$$\Lambda(t) = \sqrt{2t \log|\log t|} \; .$$

Theorem 3.13.1 (Khinchine) $$P[\limsup_{t\to 0} \frac{B_t}{\Lambda(t)} = 1] = 1, \quad P[\liminf_{t\to 0} \frac{B_t}{\Lambda(t)} = -1] = 1$$

Proof. The second statement folllows from the first one by changing B_t to $-B_t$.

(i) $\limsup_{s\to 0} \frac{B_s}{\Lambda(s)} \leq 1$ a.e

Take $\theta, \delta \in (0,1)$ and define

$$\alpha_n = (1+\delta)\theta^{-n}\Lambda(\theta^n), \beta_n = \Lambda(\theta^n)/2 \; .$$

We have $\alpha_n\beta_n = \log\log(\theta^n)(1+\delta) = \log(n)\log(\theta)$. From Corollary 3.12.4, we get

$$P[\sup_{s\leq 1}(B_s - \frac{\alpha_n s}{2}) \geq \beta_n] \leq e^{-\alpha_n\beta_n} = Kn^{(-1+\delta)} \; .$$

Borel-Cantelli gives

$$P[\liminf_{n\to\infty} \sup_{s\leq 1}(B_s - \frac{\alpha_n s}{2}) < \beta_n] = 1$$

which means that for almost every ω, there is $n_0(\omega)$ such that for $n > n_0(\omega)$ and $s \in [0, \theta^{n-1})$,

$$B_s(\omega) \leq \alpha_n \frac{s}{2} + \beta_n \leq \alpha_n \frac{\theta^{n-1}}{2} + \beta_n = (\frac{(1+\delta)}{2\theta} + \frac{1}{2})\Lambda(\theta^n) \; .$$

Since Λ is increasing on a sufficiently small interval $[0, a)$, we have for sufficiently large n and $s \in (\theta^n, \theta^{n-1}]$

$$B_s(\omega) \leq (\frac{(1+\delta)}{2\theta} + \frac{1}{2})\Lambda(s) \; .$$

Letting θ tend to 1 and δ tend to zero, we get the claim.

(ii) $\limsup_{s \to 0} \frac{B_s}{\Lambda(s)} \geq 1$ a.e.

For $\theta \in (0,1)$, the sets

$$A_n = \{B_{\theta^n} - B_{\theta^{n+1}} \geq (1-\sqrt{\theta})\Lambda(\theta^n)\}$$

are independent and since $B_{\theta^n} - B_{\theta^{n+1}}$ is Gaussian we have

$$P[A_n] = \int_a^\infty e^{-u^2/2} \frac{du}{\sqrt{2\pi}} > \frac{a}{a^2+1} e^{-a^2/2}$$

with $a = (1-\sqrt{\theta})\Lambda(\theta^n) \leq K n^{-\alpha}$ with some constants K and $\alpha < 1$. Therefore $\sum_n P[A_n] = \infty$ and by the Borel-Cantelli lemma,

$$B_{\theta^n} \geq (1-\sqrt{\theta})\Lambda(\theta^n) + B_{\theta^{n+1}} \tag{3.1}$$

for infinitely many n. Since $-B$ is also Brownian motion, we know from (i) that

$$-B_{\theta^{n+1}} < 2\Lambda(\theta^{n+1}) \tag{3.2}$$

for sufficiently large n. Using these two inequalities (3.15.2) and (3.15.3) and $\Lambda(\theta^{n+1}) \leq 2\sqrt{\theta}\Lambda(\theta^n)$ for large enough n, we get

$$B_{\theta^n} > (1-\sqrt{\theta})\Lambda(\theta^n) - 4\Lambda(\theta^{n+1}) > \Lambda(\theta^n)(1-\sqrt{\theta}-4\sqrt{\theta})$$

for infinitely many n and therefore

$$\liminf_{t \to 0} \frac{B_t}{\Lambda(t)} \geq \limsup_{n \to \infty} \frac{B_{\theta^n}}{\Lambda(\theta^n)} > 1 - 5\sqrt{\theta} \; .$$

The claim follows for $\theta \to 0$. □

Remark. This statement shows that B_t is changing sign infinitely often for $t \to 0$ and that Brownian motion is recurrent in one dimension. One could show more, namely that the set $\{B_t = 0\}$ is a nonempty perfect set with Hausdorff dimension $1/2$ which is in particularily uncountable.

By time inversion, one gets the law of iterated logarithm near infinity:

Corollary 3.13.2

$$P[\limsup_{t \to \infty} \frac{B_t}{\Lambda(t)} = 1] = 1, \quad P[\liminf_{t \to 0} \frac{B_t}{\Lambda(t)} = -1] = 1 \; .$$

Proof. Since $\tilde{B}_t = tB_{1/t}$ (with $\tilde{B}_0 = 0$) is a Brownian motion, we have with $s = 1/t$

$$\begin{aligned} 1 = \limsup_{s \to 0} \frac{\tilde{B}_s}{\Lambda(s)} &= \limsup_{s \to 0} s \frac{B_{1/s}}{\Lambda(s)} \\ = \limsup_{t \to \infty} \frac{B_t}{t\Lambda(1/t)} &= \limsup_{t \to \infty} \frac{B_t}{\Lambda(t)} \; . \end{aligned}$$

The other statement follows again by reflection. □

Corollary 3.13.3 *For d dimensional Brownian motion, one has*

$$P[\limsup_{t \to 0} \frac{|B_t|}{\Lambda(t)} = 1] = 1, \quad P[\liminf_{t \to 0} \frac{|B_t|}{\Lambda(t)} = -1] = 1$$

Proof. Let e be a unitvector in $\mathbf{R}^d$. Then $B_t \cdot e$ is a 1-dimensional Brownian motion since B_t was defined as the product of d orthogonal Brownian motions. From Khinchine's theorem, we have

$$P[\limsup_{t \to 0} \frac{B_t \cdot e}{\Lambda(t)} = 1] = 1 \; .$$

Since $B_t \cdot e \leq |B_t|$, we know that the $\limsup$ is ≥ 1. This is true for all unitvectors and we can even get it simultaneously for a dense set $\{e_n\}_{n \in \mathbf{N}}$ of unitvectors in the unitsphere. Assume the $\limsup$ is $1 + \epsilon > 1$. Then, there exists e_n such that

$$P[\limsup_{t \to 0} \frac{B_t \cdot e_n}{\Lambda(t)} \geq 1 + \frac{\epsilon}{2}] = 1$$

in contradiction to Khinchin's theorem. Therefore $\limsup = 1$. By reflection symmetry, $\liminf = -1$. □

Remark. It follows that in d dimensions, the set of limit points of $B_t/\Lambda(t)$ for $t \to 0$ is the whole unit ball $\{|v| \leq 1\}$.

3.14 The theorem of Dynkin-Hunt

Definition. Given a stopping time T, we denote with $T^{(n)}$ its discretisation

$$T^{(n)}(\omega) = k/2^n, T(\omega) \in 2^{-n}[k-1, k)$$

which is a stopping time. We have defined

$$\mathcal{A}_{T+} = \{A \in \mathcal{A}_\infty \mid A \cap \{T < t\} \in \mathcal{A}_t, \forall t\} \ .$$

Theorem 3.14.1 (Dynkin-Hunt) *Let T be a stopping time for Brownian motion, then $\tilde{B}_t = B_{t+T} - B_T$ is Brownian motion when conditioned to $\{T < \infty\}$ and $\tilde{B}_t$ is independent of $\mathcal{A}_{T+}$ when conditioned to $\{T < \infty\}$.*

Remark. If $T < \infty$ almost everywhere, no conditioning is necessary and $B_{t+T} - B_T$ is again Brownian motion.

Proof. Let A be the set $\{T < \infty\}$. The theorem says that for every function $f(B_t) = g(B_{t+t_1}, B_{t+t_2}, \dots, B_{t+t_n})$ with $g \in C(\mathbf{R}^n)$

$$E[f(\tilde{B}_t)1_A] = E[f(B_t)] \cdot P[A]$$

and that for every set $C \in \mathcal{A}_{T+}$

$$E[f(\tilde{B}_t)1_{A\cap C}] \cdot P[A] = E[f(\tilde{B}_t)1_A] \cdot P[A \cap C] \ .$$

This two statments are equivalent to the statement that for every $C \in \mathcal{A}_{T+}$

$$E[f(\tilde{B}_t) \cdot 1_{A\cap C}] = E[f(B_t)] \cdot P[A \cap C] \ .$$

Let $T^{(n)}$ be the discretisations of the stopping time T and $A_n = \{T^{(n)} < \infty\}$ as well as $A_{n,k} = \{T^{(n)} = k/2^n\}$. We compute

$$\begin{aligned}
E[f(\tilde{B}_t)1_{A\cap C}] &= \lim_{n\to\infty} E[f(B_{T^{(n)}})1_{A_n\cap C}] \\
&= \lim_{n\to\infty} \sum_{k=0}^{\infty} E[f(B_{k/2^n})1_{A_{n,k}\cap C}] \\
&= \lim_{n\to\infty} \sum_{k=0}^{\infty} E[f(B_0)] \cdot P[A_{n,k} \cap C] \\
&= E[f(B_0)] \lim_{n\to\infty} P[\bigcup_{k=1}^{\infty} A_{n,k} \cap C] \\
&= E[f(B_0)1_{A\cap C}] = E[f(B_0)] \cdot P[A \cap C] = E[f(B_t)] \cdot P[A \cap C] \ .
\end{aligned}$$

□

Corollary 3.14.2 (Blumental's zero-one law) *For every set $A \in \mathcal{A}_{0+}$ we have $P[A] = 0$ or $P[A] = 1$.*

Proof. Take the stopping time T which is identically 0. Now $\tilde{B} = B_{t+T} - B_t = B$. By Dynkin-Hunt's result, we know that $\tilde{B} = B$ is independent of $\mathcal{B}_{T+} = \mathcal{A}_{0+}$.
Since every $C \in \mathcal{A}_{0+}$ is $\{B_s, s > 0\}$ measurable, we know that $\mathcal{A}_{0+}$ is independent to itself. □

Remark. This zero-one law can be used to define regular points on the boundary of a domain $D \in \mathbf{R}^d$. Given a point $y \in \delta D$. We say it is **regular**, if $P_y[T_{\delta D} > 0] = 0$ and **irregular** $P_y[T_{\delta D} > 0] = 1$.
(This definition turns out to be equivalent to the classical definition in potential theory: a point $y \in \delta D$ is irregular if and only if there exists a barrier function $f : N \to \mathbf{R}$ in a neighborhood N of y. A **barrier function** is defined as a negative subharmonic function on $\text{int}(N \cap D)$ satisfying $f(x) \to 0$ for $x \to y$ within D.)

3.15 Selfintersection of Brownian motion

Our aim is to prove there

Theorem 3.15.1 *For $d \leq 3$, Brownian motion has infinitely many self intersections.*

Remark. Kakutani, Dvoretsky and Erdös have shown that for $d > 3$, there are no selfintersections. For $d \leq 2$, there are infinitely many $n-$fold points and for $d \geq 3$, there are no triple points. There exists now by Peres (in preprint) a simpler proof of these results.

Proposition 3.15.2 *Let K be a compact subset of $\mathbf{R}^d$ and T the hitting time of K with respect to Brownian motion starting at y. The hitting probability $h(y) = P[y + B_s \in K, T \leq s < \infty]$ is a harmonic function on $\mathbf{R}^d \setminus K$.*

Proof. Let T_δ be the hitting time of $S_\delta = \{|x-y| = \delta\}$. By the law of iterated logarithm, we have $T_\delta < \infty$ almost everywhere. By Dynkin-Hunt, we know that $\tilde{B}_t = B_{t+T_\delta} - B_t$ is again Brownian motion.

If δ is small enough, then $y + B_s \notin K$ for $t \leq T_\delta$. The random variable $B_{T_\delta} \in S_\delta$ has a uniform distribution on S_δ since Brownian motion is rotational symmetric. We have therefore

$$\begin{aligned} h(y) &= P[y + B_s \in K, s \geq T_\delta] \\ &= P[y + B_{T_\delta} + \tilde{B} \in K] \\ &= \int_{S_\delta} h(y+x) \, d\mu(x) \, , \end{aligned}$$

where μ is the normalized Lebesgue measure on S_δ. This equality for δ small enough is the definition of harmonicity. □

Proposition 3.15.3 *Let K be a countable union of closed balls. Then $h(K, y) \to 1$ for $y \to K$.*

Proof. (i) We show the claim first for one ball $K = B_r(z)$ and let $R = |z - y|$. By Brownian scaling $B_t \sim c \cdot B_{t/c^2}$, the hitting probability of K can only be a function $f(r/R)$ of r/R:

$$h(y, K) = P[y + B_s \in K, T \leq s] \;=\; P[cy + B_{s/c^2} \in cK, T_K \leq s]$$

$$\begin{aligned} &= P[cy + B_{s/c^2} \in cK, T_{cK} \leq s/c^2] \\ &= P[cy + B_{\tilde{s}}, T_{cK} \leq \tilde{s}] \\ &= h(cy, cK) \ . \end{aligned}$$

We have to show therefore that $f(x) \to 1$ as $x \to 1$.
Since the problem is translation invariant, we can fix $y = y_0 = (1, 0, \ldots, 0)$ and change K_α, which is a ball of radius α around $(-\alpha, 0, \ldots)$. We have

$$h(y_0, K_\alpha) = f(\alpha/(1+\alpha))$$

and take therefore the limit $\alpha \to \infty$

$$\begin{aligned} \lim_{x \to 1} f(x) &= \lim_{\alpha \to \infty} h(y_0, K_\alpha) = h(y_0, \bigcup K_\alpha) \\ &= E[\inf_{s \geq 0} (B_s)_1 < -1] = 1 \end{aligned}$$

because of the law of iterated logarithm.
(ii) Given $y_n \to y_0 \in K$. Then $y_0 \in K_0$ for some ball K_0.

$$\liminf_{n \to \infty} h(y_n, K) \geq \lim_{n \to \infty} h(y_n, K_0) = 1$$

by (i). □

Definition. Let μ be a probability measure on $\mathbf{R}^3$. Define the **potential theoretical energy** of μ as

$$I(\mu) = \int \int |x - y|^{-1} \, d\mu(x) \, d\mu(y) \ .$$

Given a compact set K, the **capacity** of K is defined as

$$(\inf_{\mu \in M(K)} I(\mu))^{-1} \ ,$$

where $M(K)$ is the set of probability measures on K. A measure on K minimizing the energy is called an **equilibrium measure**.

Remark. Analogues definitions can be done in any dimension. In the case $d = 2$, one replaces $|x - y|^{-1}$ by $\log |x - y|^{-1}$. In the case $d \geq 3$, one takes $|x - y|^{-(d-2)}$. The capacity is for $d = 2$ defined as $\exp(-\inf_\mu I(\mu))$ and for $d \geq 3$ as $(\inf_\mu I(\mu))^{-(d-2)}$.

We use now some results of potential theory. The next proposition is part of Frostman's fundamental theorem of potential theory. We do not give the detailed proofs (see the book of Hayman and Kennedy about subharmonic functions).

Definition. We say a measure μ_n on $\mathbf{R}^d$ **converges weakly** to μ, if for all continuous functions f, $\int f \, d\mu_n \to \int f \, d\mu$. The set of all probability measures on a comact subset E of $\mathbf{R}^d$ is known to be compact.

Proposition 3.15.4 *For every compact set $K \subset \mathbf{R}^d$, there exists an equilibrium measure μ on K and the equilibrium potential $\int |x-y|^{-(d-2)}\, d\mu(y)$ rsp. $\int \log(|x-y|^{-1})\, d\mu(y)$ takes the value $C(K)^{-1}$ on the support K^* of μ.*

Proof. (i) (Lower semicontinuity of energy) If μ_n converges to μ, then

$$\liminf_{n\to\infty} I(\mu_n) \geq I(\mu) .$$

(ii) (Existence of equilirium measure) The existence of an equilibrium measure μ follows from the compactness of the set of probability measures on K and the lower semicontinuity of the energy since a lower semicontinuous function takes a minimum on a compact space. Take a sequence μ_n such that

$$I(\mu_n) \to \inf_{\mu \in M(K)} I(\mu) .$$

Then μ_n has an accumulation point μ and $I(\mu) \leq \inf_{\mu\in M(K)} I(\mu)$.
(iii) (Value of capacity) If the potential $\phi(x)$ belonging to μ is constant on K, then it must take the value $C(K)^{-1}$ since

$$\int \phi(x)\, d\mu(x) = I(\mu) .$$

(iv) (Constancy of capacity) Assume the potential is not constant $C(K)^{-1}$ on K^*. By constructing a new measure on K^* one shows then that one can stricly decrease the energy. This is physically evident if we think of ϕ as the potential of a charge distribution μ on the set K. ☐

Corollary 3.15.5 *Let μ be the equilibrium distribution on K. Then*

$$h(y,K) = \phi_\mu \cdot C(K)$$

and therefore $h(y,K) \geq C(K) \cdot \inf_{x\in K} |x-y|^{-1}$.

Proof. Assume first K is a countable union of balls. According to Proposition 3.15.2 and Proposition 3.15.3, both functions h and $\phi_\mu \cdot C(K)$ are harmonic, vanishing at ∞ and equal to 1 on $\delta(K)$. They must therefore be equal.
In general, let $\{y_n\}$ be a dense set in K and let $K_\epsilon = \bigcup_n B_\epsilon(y_n)$. One can pass to the limit $\epsilon \to 0$. Both $h(y,K_\epsilon) \to h(y,K)$ and $\inf_{x\in K_\epsilon} |x-y|^{-1} \to \inf_{x\in K} |x-y|^{-1}$ are clear. The statment $C(K_\epsilon) \to C(K)$ follows from the upper semicontinuity of the capacity: if G_n is a sequence of open sets with $\cap G_n = E$, then $C(G_n) \to C(E)$.
The upper semicontinuity of the capacity follows essentially from the lower semicontinuity of the energy. ☐

Proposition 3.15.6 *Case $d = 3$. For any interval $J = [a, b]$, the set*

$$B_J(\omega) = \{B_t(\omega) \mid t \in [a, b]\}$$

has positive capacity for almost all ω.

Proof. We have to find a probability measure $\mu(\omega)$ on $B_I(\omega)$ such that its energy $I(\mu(\omega))$ is finite almost everywhere. Define such a measure by

$$d\mu(A) = |\frac{\{s \in [a, b] \mid B_s \in A\}}{(b - a)}| .$$

Then

$$I(\mu) = \int\int |x - y|^{-1} \; d\mu(x) d\mu(y) = \int_a^b \int_a^b (b - a)^{-1} |B_s - B_t|^{-1} \; dsdt .$$

To see the claim we have to show that this is finite almost everywhere, we integrate over Ω which is by Fubini

$$E[I(\mu)] = \int_a^b \int_a^b (b - a)^{-1} E[|B_s - B_t|^{-1}] \; dsdt$$

which is finite since $B_s - B_t$ has the same distribution as $\sqrt{s - t} B_1$ by Brownian scaling and since $E[|B_1|^{-1}] = \int |x|^{-1} e^{-|x|^2/2} \; dx < \infty$ in dimension $d \geq 2$ and $\int_a^b \int_a^b \sqrt{s - t} \; ds \; dt < \infty$. □

Now we prove the theorem

Proof. We have only to show that in the case $d = 3$, Brownian motion has infinitely many selfintersections since we get the result in smaller dimensions by projection.

(i) $\alpha = P[\bigcup_{t \in [0,1], s \geq 2} B_t = B_s] > 0$.
Proof. Let K be the set $\bigcup_{t \in [0,1]} B_t$. We know that it has positive capacity almost everywhere and that therefore $h(B_s, K) > 0$ almost everywhere. But $h(B_s, K) = \alpha$ since $B_{s+2} - B_s$ is Brownian motion independent of $B_s, 0 \leq s \leq 1$.

(ii) $\alpha_T = P[\bigcup_{t \in [0,1], 2 \leq T} B_t = B_s] > 0$ for some $T > 0$. Proof. Clear since $\alpha_T \to \alpha$ for $T \to \infty$.
(iii) The claim. Proof. Define the random variables $X_n = 1_{C_n}$ with

$$C_n = \{\omega \mid B_t = B_s, \text{ for some } t \in [nT, nT + 1], s \in [nT + 2, (n + 1)T] \} .$$

They are independent and by the strong law of large numbers $\sum_n X_n = \infty$ almost everywhere. □

Corollary 3.15.7 *Any point $B_s(\omega)$ is an accumulation point of selfcrossings of $\{B_t(\omega)\}_{t\geq 0}$.*

Proof. Again, we have only to treat the three dimensional case. Let $T > 0$ be such that

$$\alpha_T = P[\bigcup_{t\in[0,1],2\leq T} B_t = B_s] > 0$$

in the proof of the theorem. By scaling,

$$P[B_t = B_s \mid t \in [0,\beta], s \in [2\beta, T\beta]\,]$$

is independent of β. We have thus selfintersections of the random walk in any interval $[0,b]$ and by translation in any interval $[a,b]$. □

3.16 Recurrence of Brownian motion

We show here that like its discrete brother, the random walk, Brownian motion is transient in dimensions $d \geq 3$ and recurrent in dimensions $d \leq 2$.

Lemma 3.16.1 *Let T be a finite stopping time and $R_T(\omega)$ be a rotation in $\mathbf{R}^d$ which turns $B_T(\omega)$ onto the first coordinate axis*

$$R_T(\omega)B_T(\omega) = (|B_T(\omega)|, 0, \dots 0) .$$

Then $\tilde{B}_t = R_T(B_{t+T} - B_T)$ is again Brownian motion.

Proof. By Dynkin-Hunt, $\tilde{B}_t = B_{t+T} - B_T$ is Brownian motion and independent of $\mathcal{A}_T$. By checking the definitions of Brownian motion, it follows that if B is Brownian motion, also $R(x)B_t$ is Brownian motion, if $R(x)$ is a random rotation on $\mathbf{R}^d$ independent of B_t. Since R_T is $\mathcal{A}_T$ measureable and $\tilde{B}_t$ is independent of $\mathcal{A}_T$, the claim follows. □

Lemma 3.16.2 *Let K_r be the ball of radius r centered at $0 \in \mathbf{R}^d$ with $d \geq 3$. We have for $y \notin K_r$*

$$h(y, K_r) = (r/|y|)^{d-2} .$$

Proof. Both $h(y, K_r)$ and $(r/|y|)^{d-2}$ are harmonic functions which are 1 at δK_r and vanishing at infinity. They are the same. □

Theorem 3.16.3 *For $d \geq 3$, we have $\lim_{t \to \infty} |B_t| = \infty$ almost surely.*

Proof. Define a sequence of stopping times T_n by

$$T_n = \inf\{s > 0 \mid |B_s| = 2^n\} ,$$

which is finite almost everywhere because of the law of iterated lograrithm. We know from the Lemma 3.16.1 that

$$\tilde{B}_t = R_{T_n}(B_{t+T_n} - B_{T_n})$$

is a copy of Brownian motion. Clearly also $|B_{T_n}| = 2^n$.
We have $B_s \in K_r(0) = \{|x| < r\}$ for some $s > T_n$ if and only if $\tilde{B}_t \in$

$(2^n, 0 \dots, 0) + K_r(0)$ for some $t > 0$.
Therefore using the previous lemma

$$P[B_s \in K_r(0); s > T_n] = P[\tilde{B}_t \in (2^n, 0 \dots, 0) + K_r(0); t > 0] = (\frac{r}{2^n})^{d-2}$$

which implies in the case $r2^{-n} < 1$ by Borel-Cantelli that for almost all ω, $B_s(\omega) \geq r$ for $s > T_n$. Since T_n is finite almost everywhere, we get $\liminf_s |B_s| \geq r$. Since r is arbitrary, the claim follows. □

Brownian motion is recurrent in dimensions $d \leq 2$. In the case $d = 1$, this follows readily from the law of iterated logarithm. First a lemma

Lemma 3.16.4 *In dimensions $d = 2$, almost every path of Brownian motion hits a ball K_r if $r > 0$: one has $h(y, K) = 1$.*

Proof. We know that $h(y) = h(y, K)$ is harmonic and equal to 1 on δK. It is also rotational invariant and therefore $h(y) = a + b \log |y|$. Since $h \in [0, 1]$ we have $h(y) = a$ and so $a = 1$. □

Theorem 3.16.5 *Let $d \leq 2$ and S be an open nonempty set in $\mathbf{R}^d$. Then the Lebesque measure of $\{t \mid B_t \in S\}$ is infinite.*

Proof. It suffices to take $S = K_r(x_0)$, a ball of radius r around x_0. Since by the previous lemma, Brownian motion hits every ball almost surely, we can assume that $x_0 = 0$ and by scaling that $r = 1$.
Define inductively a sequence of hitting rsp. leaving times T_n, S_n of the annulus $\{1/2 < |x| < 2\}$, where $T_1 = \inf\{t \mid |B_t| = 2\}$ and

$$\begin{aligned} S_n &= \inf\{t > T_n \mid |B_t| = 1/2\} \\ T_n &= \inf\{t > S_{n-1} \mid |B_t| = 2\} . \end{aligned}$$

These are finite stopping times. The Dynkin-Hunt theorem shows that $S_n - T_n$ and $T_n - S_{n-1}$ are two mutually independent families of IID random variables. The measures $Y_n = |I_n|$ of the time intervals

$$I_n = \{t, |B_t| \leq 1 m T_n \leq t \leq T_{n+1}\} ,$$

where B_t is in the unitball are independent random variables. Therefore, also $X_n = \min(1, Y_n)$ are independent bounded IID random variables and by the law of large numbers, $\sum_n X_n = \infty$ which implies $\sum_n Y_n = \infty$ and the claim follows since $|\{t \in [0, \infty) \mid |B_t| \leq 1\}| \geq \sum_n T_n$. □

3.14 Feynman-Kac formula

In quantum mechanics, the Schrödinger equation $i\hbar\dot{u} = Hu$ defines the evolution of the wave function $u(t) = e^{-itH/\hbar}u(0)$ in a Hilbert space $\mathcal{H}$. The operator H is the **Hamiltonian** of the system. We assume, it is a **Schrödinger operator** $H = H_0 + V$, where $H_0 = -\Delta/2$ is the Hamiltonian of a free particle and $V : \mathbf{R}^d \to \mathbf{R}$ is the potential. The free operator H_0 already is not defined on the whole Hilbert space $\mathcal{H} = L^2(\mathbf{R}^d)$ and one restricts H to a vector space $D(H)$ called **domain** containing the in $\mathcal{H}$ dense set $C_0^\infty(\mathbf{R}^d)$ of all smooth functions vanishing at infinity. Define

$$D(A^*) = \{u \in \mathcal{H} \mid v \mapsto (Av, u) \text{ is a bounded linear functional on } D(A)\} .$$

If $u \in D(A^*)$, then there exists a unique function $w = A^*u \in \mathcal{H}$ such that $(Av, u) = (v, w)$ for all $u \in D(A)$. This defines the **adjoint** A^* of A with domain $D(A^*)$.

Definition. A linear operator $A : D(A) \subset \mathcal{H} \to \mathcal{H}$ is **symmetric** if $(Au, v) = (u, Av)$ for all $u, v \in D(A)$ and **selfadjoint**, if it is symmetric and $D(A) = D(A^*)$.

Definition. A sequence of bounded operators A_n converges **strongly** to A, if $A_n u \to Au$ for all $u \in \mathcal{H}$. One writes $A = s - \lim_{n\to\infty} A_n$.

The only thing, we actually have to know is the fact that a selfadjoint operator defines a one parameter family of unitary operators $t \mapsto e^{itA}$ which is strongly continuous. Moreover e^{itA} leaves the domain $D(A)$ of A invariant.

Proposition 3.14.1 (Trotter product formula) *Given selfadjoint operators A, B defined on $D(A), D(B) \subset \mathcal{H}$. Assume $A + B$ is selfadjoint on $D = D(A) \cap D(B)$, then*

$$e^{it(A+B)} = s - \lim_{n\to\infty} (e^{itA/n} e^{itB/n})^n .$$

If A, B are bounded from below, then

$$e^{-t(A+B)} = s - \lim_{n\to\infty} (e^{-tA/n} e^{-tB/n})^n .$$

Proof. Define

$$S_t = e^{it(A+B)}, V_t = e^{itA}, W_t = e^{itB}, U_t = V_t W_t$$

and $v_t = S_t v$ for $v \in D$. Since $A + B$ is selfadjoint on D, $v_t \in D$. Use a telescopic sum to estimate

$$
\begin{aligned}
||(S_t - U_{t/n}^n)v|| &= ||\sum_{j=0}^{n-1} U_{t/n}^j (S_{t/n} - U_{t/n}) S_{t/n}^{n-j-1} v|| \\
&\leq n \sup_{0 \leq s \leq t} ||(S_{t/n} - U_{t/n}) v_s||
\end{aligned}
$$

and we have to show that this goes to zero for $n \to \infty$. Given $u \in D = D(A) \cap D(B)$. We have

$$
\lim_{s \to 0} \frac{S_s - 1}{s} u = i(A+B)u = \lim_{s \to 0} \frac{U_s - 1}{s} u
$$

so that for each $u \in D$

$$
\lim_{n \to \infty} n \cdot ||(S_{t/n} - U_{t/n})u|| = 0 \; . \tag{3.3}
$$

The linear space D with norm $|||u||| = ||(A+B)u|| + ||u||$ is a Banach space since $A+B$ is selfadjoint on D and therefore closed. We have a bounded family $\{n(S_{t/n} - U_{t/n})\}_{n \in \mathbf{N}}$ of bounded operators from D to $\mathcal{H}$. The principle of uniform boundedness states that

$$
||n(S_{t/n} - U_{t/n})u|| \leq C \cdot |||u||| \; .
$$

An $\epsilon/3$ argument shows that the limit 3.3 exists uniformly on compact subsets of D and especially on $\{v_s\}_{s \in [0,t]} \subset D$ and so $n \sup_{0 \leq s \leq t} ||(S_{t/n} - U_{t/n})v_s|| = 0$. The second statement is proven in exactly the same way. □

Remark. Trotter's product formula generalizes the Lie product formula

$$
\lim_{n \to \infty} (\exp(A/n) \exp(B/n))^n = \exp(A+B)
$$

for finite dimensional matrices A, B, which is a special case.

Corollary 3.14.2 (Feynman 1948) *Assume $H = H_0 + V$ is selfadjoint on $D(H)$. Then*

$$
e^{-itH} u(x_0) = \lim_{n \to \infty} (\frac{2\pi i t}{n})^{-d/2} \int_{(\mathbf{R}^d)^n} e^{iS_n(x_0, x_1, x_2, \dots, x_n, t)} u(x_n) \; dx_1 \dots dx_n
$$

where

$$
S_n(x_0, x_1, \dots, x_n, t) = \frac{t}{n} \sum_{i=1}^{n} \frac{1}{2} (\frac{|x_i - x_{i-1}|}{t/n})^2 - V(x_i) \; .
$$

Proof. (Nelson) From $\dot{u} = -iH_0 u$, we get by Fourier transform $\dot{\hat{u}} = i\frac{|k|^2}{2}\hat{u}$ which gives $\hat{u}_t(k) = \exp(i\frac{|k|^2|}{2}t)\hat{u}_0(k)$ and by inverse Fourier transform

$$e^{-itH_0}u(x) = u_t(x) = (2\pi it)^{-d/2}\int_{\mathbf{R}^d} e^{i\frac{|x-y|^2}{2t}}u(y)\, dy\ .$$

The Trotter product formula

$$e^{-it(H_0+V)} = s - \lim_{n\to\infty}(e^{itH_0/n}e^{itV/n})^n$$

gives now the claim. □

Remark. We did not specify the set of potentials, for which H_0+V can be made selfadjoint. For example, $V \in C_0^\infty(\mathbf{R}^\nu)$ is enough or $V \in L^2(\mathbf{R}^3)\cap L^\infty(\mathbf{R}^3)$ in three dimensions.

We have seen in the above proof that e^{-itH_0} has the integral kernel $\tilde{P}_t(x,y) = (2\pi it)^{-d/2}e^{i\frac{|x-y|^2}{2t}}$. The same Fourier calculation shows that e^{-tH_0} has the integral kernel

$$P_t(x,y) = (2\pi t)^{-d/2}e^{-\frac{|x-y|^2}{2t}}\ ,$$

where g_t is the density of a Gaussian random variable with variance t. Note that even if $u \in L^2(\mathbf{R}^d)$ is only defined almost everywhere, the function $u_t(x) = e^{-tH_0}u(x) = \int P_t(x-y)u(y)dy$ is continuous and defined everywhere.

Lemma 3.14.3 *Given $f_1,\dots,f_n \in L^\infty(\mathbf{R}^d)\cap L^2(\mathbf{R}^d)$ and $0 < s_1 < \dots < s_n$. Then*

$$(e^{-t_1H_0}f_1\cdots e^{-t_nH_0}f_n)(0) = \int f_1(B_{s_1})\cdots f_n(B_{s_n})\, dB,$$

where $t_1 = s_1, t_i = s_i - s_{i-1}, i \geq 2$ and the f_i on the left hand side are understood as multiplication operators on $L^2(\mathbf{R}^d)$.

Proof. Since $B_{s_1}, B_{s_2}-B_{s_1},\dots B_{s_n}-B_{s_{n-1}}$ are mutually independent Gaussian random variables of variance $t_1,t_2,\dots,t_n$, their joint distribution is

$$P_{t_1}(0,y_1)P_{t_2}(0,y_2)\dots P_{t_n}(0,y_n)\, dy$$

which is after a change of variables $y_1 = x_1, y_i = x_i - x_{i-1}$

$$P_{t_1}(0,x_1)P_{t_2}(x_1,x_2)\dots P_{t_n}(x_{n-1},x_n)\, dx\ .$$

Therefore

$$\int f_1(B_{s_1})\cdots f_n(B_{s_n})\, dB \quad = \quad \int_{(\mathbf{R}^d)^n} P_{t_1}(0,y_1)P_{t_2}(0,y_2)\dots P_{t_n}(0,y_n)f_1(y_1)\dots f_n(y_n)\, dy$$

$$
\begin{aligned}
&= \int_{(\mathbf{R}^d)^n} P_{t_1}(0,x_1)P_{t_2}(x_1,x_2)\dots P_{t_n}(x_{n-1},x_n)f_1(x_1)\dots f_n(x_n)\;dx \\
&= (e^{-t_1 H_0}f_1 \cdots e^{-t_n H_0} f_n)(0)\;.
\end{aligned}
$$

□

Denote with dB the Wiener measure on $C([0,\infty),\mathbf{R}^d)$ and with dx the Lebesgue measure on $\mathbf{R}^d$. We define also an extended Wiener measure $dW = dx \otimes dB$ on $C([0,\infty),\mathbf{R}^d)$ on all paths $s \mapsto W_s = x + B_s$ starting at $x \in \mathbf{R}^d$.

Corollary 3.14.4 *Given $f_0, f_1, \dots, f_n \in L^\infty(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$ and $0 < s_1 < \dots < s_n$. Then*

$$
\int f_0(W_{s_0}) \cdots f_n(W_{s_n})\;dW = (\overline{f_0}, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n)\;.
$$

Proof. (i) Case $s_0 = 0$. From the above lemma, we have after the dB integration that

$$
\begin{aligned}
\int f_0(W_{s_0}) \cdots f_n(W_{s_n})\;dW &= \int_{\mathbf{R}^d} f_0(x) e^{-t_1 H_0} f_1(x) \cdots e^{-t_n H_0} f_n(x)\;dx \\
&= (\overline{f}_0, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n)\;.
\end{aligned}
$$

(ii) In the case $s_0 > 0$ we have from (i) convergence theorem

$$
\begin{aligned}
\int f_0(W_{s_0}) \cdots f_n(W_{s_n})\;dW &= \lim_{R\to\infty} \int_{\mathbf{R}^d} 1_{\{|x|<R\}}(W_0) f_0(W_{s_0}) \cdots f_n(W_{s_n})\;dW \\
&= \lim_{R\to\infty} (\overline{f}_0 e^{-s_0 H_0} 1_{\{|x|<R\}}, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n(x)) \\
&= (\overline{f}_0, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n)\;.
\end{aligned}
$$

□

We prove now the Feynman-Kac formula for Schrödinger operators of the form $H = H_0 + V$ with $V \in C_0^\infty(\mathbf{R}^d)$. Since V is continuous, the integral $\int_0^t V(W_s(\omega))\;ds$ can be taken for each ω as a limit of Rieman sums and $\int_0^t V(W_s)\;ds$ is therefore a random variable.

Theorem 3.14.5 (Feynman-Kac formula) *Given $H = H_0 + V$ with $V \in C_0^\infty(\mathbf{R}^d)$, then*

$$
(f, e^{-tH} g) = \int \overline{f}(W_0) g(W_t) e^{-\int_0^t V(W_s))\;ds}\;dW\;.
$$

Proof. (Nelson) By the Trotter product formula

$$(f, e^{-tH} g) = \lim_{n\to\infty} (f, (e^{-tH_0/n} e^{-tV/n})^n g)$$

so that by the last Corrollary

$$(f, e^{-tH} g) = \lim_{n\to\infty} \int \overline{f}(W_0) g(W_t) \exp(-\frac{t}{n} \sum_{j=0}^{n-1} V(W_{tj/n}))\, dW \; . \tag{3.4}$$

and since $s \mapsto W_s$ is continuous, we have almost everywhere

$$\frac{t}{n} \sum_{j=0}^{n-1} V(W_{tj/n}) \to \int_0^t V(W_s)\, ds \; .$$

The integrand on the right hand side of (3.4) is dominated by $|f(W_0)| \cdot |g(W_t)| \cdot e^{t||V||_\infty}$ which is in $L^1(dW)$ since again by the Corollary

$$\int |f(W_0)| \cdot |g(W_t)|\, dW = (|f|, e^{-tH_0} |g|) < \infty \; .$$

The dominated convergence theorem gives the claim. □

Remark. The formula can be extended to larger classes of potentials like potentials V which are locally in L^1. Selfadjointness, needed in Trotter's product formula is assured if $V \in L^2 \cap L^p$ with $p > d/2$. But also Trotter's product formula allows further generalizations. See the book of Simon about functional integration or the book of Glimm and Jaffe about Quantum physics.

Why is Feynman-Kac useful?

* Can use Brownian motion to study Schrödinger semigroups and vice vera: for example an easy proof of the ArcSin law for Brownian motion.

* Can treat operators with magnetic fields in a uniform way.

* Functional integration is a way of quantisation which generalizes to more situations.

* Useful to study groundstates and groundstate energies under perturbations.

* Study of the classical limit $\hbar \to 0$.

3.15 The quantum mechanical oscillator

The one dimensional Schrödinger operator

$$H = H_0 + U = -\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2 - \frac{1}{2}$$

is the Hamiltonian of the **quantum mechanical oscillator**. It is a quantum mechanical system which can be solved explicitely like his brother in classical mechanics with Hamiltonian $H = -\frac{1}{2}\dot{x}^2 + \frac{1}{2}x^2$. One has

$$H = A^* A \ ,$$

with

$$A^* = 2^{-1/2}(\frac{d}{dx} + x),\ A = 2^{-1/2}(-\frac{d}{dx} + x) \ .$$

The first order operator A^* is also called **particle creation operator** and its adjoint A, the **particle annihilation operator**. Since for all $u, v \in C_0^\infty(\mathbf{R})$, we have $(Au, v) = (u, A^*v)$ they are adjoint to each other. The unitvector

$$\Omega_0 = \pi^{-1/4}\ e^{-x^2/2}$$

satisfies $A\Omega_0 = 0$ and is therefore an eigenvector of $H = AA^*$ with eigenvalue 0. It is called the **ground state** or **vacuum state** discribing the system with no particle. Define inductively the n-particle states

$$\Omega_n = n^{-1/2} \cdot A^*\Omega_{n-1}$$

by creating an additional particle from the $(n-1)$-particle state. The functions Ω_n are called **Hermite functions** in $L^2(\mathbf{R})$.

Proposition 3.15.1 (Quantum mechanical oscillator)

a) $(\Omega_n, \Omega_m) = \delta_{n,m}$.

b) $A\Omega_n = \sqrt{n}\Omega_{n-1}, A^*\Omega_n = \sqrt{n+1}\Omega_{n+1}$.

c) $H = A^*A\Omega_n = n\Omega_n$.

d) Ω_n *form a basis in* $L^2(\mathbf{R})$.

Proof. Denote with $[A, B] = AB - BA$ the commutator of two operators. We check first by induction the formula

$$[A, (A^*)^n] = n \cdot (A^*)^{n-1} \ .$$

For $n = 1$, this means $[A, A^*] = I$. The induction step is

$$[A, (A^*)^n] = [A, (A^*)^{n-1}]A^* + (A^*)^{n-1}[A, A^*] = (n-1)(A^*)^{n-1} + (A^*)^{n-1} = n(A^*)^{n-1} .$$

a) We check by induction

$$((A^*)^n\Omega_0, (A^*)^m\Omega_0) = n!\delta_{mn} .$$

It is clear for $n = 0$ since Ω_0 is normalized. The induction step is using $[A, (A^*)^n] = n \cdot (A^*)^{n-1}$ and $A\Omega_0 = 0$

$$\begin{aligned} ((A^*)^n\Omega_0, (A^*)^m\Omega_0) &= (A(A^*)^n\Omega_0 (A^*)^{m-1}\Omega_0) \\ &= ([A, (A^*)^n]\Omega_0 (A^*)^{m-1}\Omega_0) \\ &= n((A^*)^{n-1}\Omega_0, (A^*)^{m-1}\Omega_0) . \end{aligned}$$

If $n < m$, then we get from this zero after n steps, while in the case $n = m$, we get $((A^*)^n\Omega_0, (A^*)^n\Omega_0) = n \cdot ((A^*)^{n-1}\Omega_0, (A^*)^{n-1}\Omega_0)$.

b) $A^*\Omega_n = \sqrt{n+1} \cdot \Omega_{n+1}$ is the definition of Ω_n.

$$A\Omega_n = (n!)^{-1/2} A(A^*)^n\Omega_n = (n!)^{-1/2} n\Omega_0 = \sqrt{n}\Omega_{n-1} .$$

c) Follows readily from b).

d) a) shows that $\{\Omega_n\}_{n\in\mathbf{N}}$ it is an orthonormal set. In order to show that they span $L^2(\mathbf{R})$ we have only to show that they span the dense set

$$\mathcal{S} = \{f \in C_0^\infty(\mathbf{R}) \mid x^m f^{(n)}(x) \to 0, |x| \to \infty, \forall m, n \in \mathbf{N}\}$$

called **Schwarz space**, since a function f must vanish in $L^2(\mathbf{R})$ if it is orthogonal to a dense set. Assume $(f, \Omega_n) = 0$ for all n. Then, since $A^* + A = x$

$$0 = \sqrt{n!}\,(f, \Omega_n) = (f, (A^*)^n\Omega_0) = (f, (A^* + A)^n\Omega_0) = 2^{n/2}\,(f, x^n\Omega_0)$$

and so

$$\begin{aligned} (f\Omega_0)\hat{}(k) &= \int_{-\infty}^{\infty} f(x)\Omega_0(x)e^{ikx}\,dx \\ &= (f, \Omega_0 e^{ikx}) = (f, \sum_{n\geq 0} \frac{(ikx)^n}{n!}\Omega_0) \\ &= \sum_{n\geq 0} \frac{(ik)^n}{n!}(f, x^n\Omega_0) = 0 . \end{aligned}$$

From this follows $f\Omega_0 = 0$ and since Ω_0 is nowhere vanishing, we have $f = 0$.
□

Remark. This formalism of particle creation and annihilation operators works for any potential of the form $U(x) = q^2(x) - q'(x)$ and the operator $H = H_0 + U/2$ can then be written as $H = A^*A$, where

$$A^* = 2^{-1/2}(\frac{d}{dx} + q(x)),\ A = 2^{-1/2}(-\frac{d}{dx} + q(x)) .$$

The oscillator is the special case $q(x) = x$. If $U(x)$ has a lowest eigenvalue, then it can indeed be written as $U(x) = q^2(x) - q'(x)$: Let E_0 be the lowest eigenvalue with eigenfunction Ω_0. Define $m = \partial_x(\log(\Omega(x)))$ and $A = m\partial_x m^{-1} = \partial_x + q(x)$ which gives with a little calculation $H = A^*A$ and $U(x) = q^2(x) - q'(x)$. On can then consider the **Bäcklund transformation** $H \mapsto \tilde{H} = AA^* = H$ which is in the case of the harmonic oscillator the map $H \mapsto H + 1$ and in general given by replacing U with $\tilde{U} = U - \partial_x^2 \log \Omega_0$. The new operator $\tilde{H}$ has the same spectrum then H except that the lowest eigenvalue can be removed (Deift 1978). This procedure can be inversed and can be used to create potentials (solitons) out of the vacuum.

Remark. It is very natural to use the language of **supersymmetry** (Witten): take two copies $\mathcal{H}_f \oplus \mathcal{H}_b$ of the Hilbert space where "f" stands for Fermion and "b" for Boson. With

$$Q = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix},\ P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

one can write $H \oplus \tilde{H} = Q^2$, $P^2 = 1$, $QP + PQ = 0$ and one says (H, P, Q) has supersymmetry. The operator Q is also called a **Dirac operator**. A supersymmetric system has the property that nonzero eigenvalues have the same number of bosonic and fermionic eigenstates. This implies that $\tilde{H}$ has the same spectrum as H except that lowest eigenvalue can disappear.

Remark. In quantum field theory, there exists a process called canonical quantisation, where a quantum mechanical system is extended to a quantum field. Particle annihilation and creation operators play an important role.

3.16 Feynman-Kac for the oscillator

Let

$$L_0 = H_0 + U = -\frac{1}{2}\dot{x}^2 + \frac{1}{2}x^2 - \frac{1}{2}$$

We want to treat perturbations $L = L_0 + V$ of the osciallator with an analogue Feynman-Kac formula. The calculation of the integral kernel $P_t(x,y)$ of e^{-tL_0} satisfying

$$(e^{-tL_0}f)(x) = \int_{\mathbf{R}} P_t(x,y)f(y)\,dy\ .$$

is slightly more involved than in the case of the free Laplacian. Let Ω_0 be the ground state of L_0 as discussed in the last section.

Lemma 3.16.1 *Given $f_0, f_1, \dots, f_n \in L^\infty(\mathbf{R})$ and $-\infty < s_0 < s_1 < \dots < s_n < \infty$. Then*

$$(\Omega_0, f_0 e^{-t_1 L_0} f_1 \cdots e^{-t_n L_0} f_n \Omega_0) = \int f_0(Q_{s_0}) \cdots f_n(Q_{s_n})\,dQ\ ,$$

where $t_0 = s_0, t_i = s_i - s_{i-1}, i \geq 1$.

Proof. The Trotter product formula for $L_0 = H_0 + U$ gives

$$\begin{aligned} & (\Omega_0, f_0 e^{-t_1 L_0} f_1 \cdots e^{-t_n L_0} f_n \Omega_0) \\ = & \lim_{m=(m_1,\dots,m_n), m_i \to \infty} (\Omega_0, f_0 (e^{-t_1 H_0/m_1} e^{-t_1 U/m_1})^{m_1} f_1 \cdots e^{-t_n H_0} f_n \Omega_0) \\ = & \int f_0(x_0) \cdots f_n(x_n)\,dG_m(x,y) \end{aligned}$$

and G_m is a measure. Since e^{-tH_0} has a Gaussian kernel and e^{-tU} is a multiple of a Gaussian density and integrals are Gaussians, the measure dG_m is Gaussian converging to a Gaussian measure dG. Since $L_0(x\Omega_0) = x\Omega_0$ and $(x\Omega_0, x\Omega_0) = 1/2$ we have

$$\int x_i x_j\,dG = (x\Omega_0, e^{-(s_j - s_i)} L_0 x\Omega_0) = \frac{1}{2}e^{-(s_j - s_i)}$$

which shows that dG is the joint probability distribution of $Q_{s_0}, \dots Q_{s_n}$. The claim follows. □

Corollary 3.16.2 *The kernel $P_t(x,y)$ of L_0 is given by the Mehler formula*

$$P_t(x,y) = \frac{1}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{(x^2+y^2)(1+e^{-2t}) - 4xye^{-t}}{2\sigma^2}\right)\ .$$

with $\sigma^2 = (1 - e^{-2t})$.

Proof. We have

$$(f, e^{-tL_0} g) = \int f(y)\Omega_0^{-1}(y) g(x)\Omega_0^{-1}(x)\ dG(x,y) = \int f(y) P_t(x,y)\ dy$$

with the Gaussian measure dG having covariance

$$A = \frac{1}{2}\begin{pmatrix} 1 & e^{-t} \\ e^{-t} & 1 \end{pmatrix} .$$

We get Mehler's formula by inverting this matrix and using that the density is

$$(2\pi)\det(A)^{-1/2} e^{-((x,y),A(x,y))} .$$

□

Let dQ be the Wiener measure on $C(\mathbf{R})$ belonging to the oscillator process.

Theorem 3.16.3 (Feynman-Kac for oscillator process)
Given $L = L_0 + V$ with $V \in C_0^\infty(\mathbf{R})$, then

$$(f\Omega_0, e^{-iL} g\Omega_0) = \int \overline{f}(Q_0) g(Q_t) e^{-\int_0^t V(Q_s))\ ds}\ dQ$$

for all $f, g \in L^2(\mathbf{R}, \Omega_0^2 dx)$.

Proof. By the Trotter product formula

$$(f\Omega_0, e^{-iL} g\Omega_0) = \lim_{n\to\infty} (f\Omega_0, (e^{-tL_0/n} e^{-tV/n})^n g\Omega_0)$$

so that

$$(f\Omega_0, e^{-iL} g\Omega_0) = \lim_{n\to\infty} \int \overline{f}(Q_0) g(Q_t) \exp(-\frac{t}{n}\sum_{j=0}^{n-1} V(Q_{tj/n}))\ dQ . \qquad (3.5)$$

and since Q is continuous, we have almost everywhere

$$\frac{t}{n}\sum_{j=0}^{n-1} V(Q_{tj/n}) \to \int_0^t V(Q_s)\ ds .$$

The integrand on the right hand side of (3.5) is dominated by $|f(Q_0)||g(Q_t)|e^{t||V||_\infty}$ which is in $L^1(dQ)$ since

$$\int |f(Q_0)||g(Q_t)|\ dQ = (\Omega_0|f|, e^{-tL_0}\Omega_0|g|) < \infty .$$

The dominated convergence theorem gives the claim. □

3.17 Wiener sausage

The Feynman-Kac formula can be used to understand the Dirichlet Laplacian of a domain $\Omega \in \mathbf{R}^d$. We give two examples without proofs, and and example with proof.

1) **Theorem of Weyl**.
Let Ω be an open set in $\mathbf{R}^d$ such that the Lebesgue measure $|\Omega|$ is finite and the Lebesgue measure of the boundary $|\delta\Omega|$ is vanishing. Denote with H_Ω the Dirichlet Laplacian $-\Delta/2$. Denote with $k_\Omega(E)$ the number of eigenvalues of H_Ω below E. This function is also called the integrated density of states. Denote with K_d the unit ball in $\mathbf{R}^d$ and with $|K_d| = \pi^{d/2}\Gamma(\frac{d}{2}+1)^{-1}$ its Lebesgue measure. **Weyl's formula** describing the asymptotic behavior of $k_\Omega(E)$ for large E:

$$\lim_{E\to\infty} E^{-d/2}k_\Omega(E) = 2^{-d/2}\pi^{-d} \cdot |K_d| \cdot |\Omega| \; .$$

2) **Problem of crushed ice**.
Put n ice balls $K_{j,n}, 1 \leq j \leq n$ of radius r_n into a glass of water so that $n \cdot r_n = \alpha$. In order to know, how good this ice cools the water it is good to know the lowest eigenvalue E_1 of the Dirichlet Laplacian H_Ω since the motion of the temperature distribution u by the heat equation $\dot{u} = H_\Omega u$ is dominated by e^{-tE_1}. This motivates to compute the lowest eigenvalue of the domain $\Omega \setminus \bigcup_{j=1}^n K_{j,n}$. This can be done exactly in the limit $n \to \infty$ and when ice $K_{j,n}$ is randomly distributed in the glass. Mathematically, this is described as follows:
Let Ω be an open bounded domain in $\mathbf{R}^d$. Given a sequence $x = (x_1, x_2, \dots) \in \Omega^{\mathbf{N}}$ and a sequence of radii $r_1, r_2, \dots$, we define

$$\Omega_n = \Omega \setminus \bigcup_{i=1}^n \{|x - x_i| \leq r_n\}$$

which is the domain Ω with n points balls $K_{j,n}$ with center $x_1, \dots x_n$ and radius r_n removed. Let $H(x,n)$ be the Dirichlet Laplacian on Ω_n and $E_k(x,n)$ the k-th eigenvalue of $H(x,n)$ which are random variable $E_k(n)$ in x, if we take on $\Omega^{\mathbf{N}}$ the product Lebesgue measure. One can show that in the case $nr_n \to \alpha$

$$E_k(n) \to E_k(0) + 2\pi\alpha|\Omega|^{-1}$$

in probability. Random impurities produce a constant shift in the spectrum. For the physical system with the crushed ice, where the crushing makes $nr_n \to \infty$, there is much better cooling as one thinks.

3) **Wiener sausage**.
In the proof of the crushed ice problem appears an interesting problem in Brownian motion, which uses the Feynman-Kac formula in the form

$$e^{-t(H_0+V)}u(x) = \int e^{-\int_0^t V(x+B_s)\, ds} u(x+B_s) \, dB \; .$$

Definition. Let $W_\delta(t)$ be the set

$$\{x \in \mathbf{R}^d \mid |x - B_t(\omega)| \le \delta, \text{ for some } s \in [0,t]\} .$$

It is of course dependent on ω and just a δ-neighborhood of the Brownian path $B_{[0,t]}(\omega)$. This set is called **Wiener sausage** and one is interested in the expected volume $|W_\delta(t)|$ of this set as $\delta \to 0$.

First a lemma, which relates the Dirichlet Laplacian $H_\Omega = -\Delta/2$ on Ω with Brownian motion.

Lemma 3.17.1 *Let Ω be a bounded domain in $\mathbf{R}^d$ containing 0 and $P_\Omega(x,y,t)$, the integral kernel of e^{-tH}, where H is the Dirichlet Laplacian on Ω. Then*

$$E[B_s \in \Omega; 0 \le s \le t] = 1 - \int P_\Omega(0,x,t)\; dx .$$

Proof. (i) It is known that the Dirichlet Laplacian can be approximated in the strong resolvent sense by operators $H_0 + \lambda \cdot V$, where $V = 1_{\Omega^c}$ is the characteristic function of the exterior of Ω. This means that

$$(H_0 + \lambda \cdot V)^{-1} u \to (H_\Omega - z)^{-1} u, \lambda \to \infty$$

for z outside $[0,\infty)$ and all $u \in C_c^\infty(\mathbf{R}^d)$.

(ii) Since Brownian paths are continuous, we have $\int_0^t V(B_s)\; ds > 0$ if and only if $B_s \in \Omega^c$ for some $s \in [0,t]$. We get therefore $e^{-\lambda \int_0^t V(B_s)\; ds} \to 1_{\{B_s \in \Omega^c\}}$ pointwise almost everywhere.

Let u_n be a sequence in C_c^∞ converging pointwise to 1. We get with the dominated convergence theorem, using (i) and (ii) and Feynman-Kac

$$\begin{aligned}
E[B_s \in \Omega^c; 0 \le s \le t] &= \lim_{n\to\infty} E[u_n(B_s) \in \Omega^c; 0 \le s \le t] \\
&= \lim_{n\to\infty} \lim_{\lambda\to\infty} E[e^{-\lambda \int_0^t V(B_s)\; ds} u_n(B_t)] \\
&= \lim_{n\to\infty} \lim_{\lambda\to\infty} e^{-t(H_0+\lambda\cdot V)} u_n(0) \\
&= \lim_{n\to\infty} e^{-tH_\Omega} u_n(0) \\
&= \lim_{n\to\infty} \int P_\Omega(0,x,t) u_n(0)\; dx = \int P_\Omega(0,x,t)\; dx .
\end{aligned}$$

□

Theorem 3.17.2 (Spitzer) *Assume $d = 3$.*

$$E[|W_\delta(t)|] = 2\pi\delta t + 4\delta^2\sqrt{2\pi t} + \frac{4\pi}{3}\delta^3 .$$

Proof. Using Brownian scaling,

$$\begin{aligned}
E[|W_{\lambda\delta}(\lambda^2 t)|] &= E[|\{|x - B_s| \leq \lambda\delta, 0 \leq s \leq \lambda^2 t\}|] \\
&= E[|\{|\frac{x}{\lambda} - \frac{B_{\tilde{s}\lambda^2}}{\lambda}| \leq \delta, 0 \leq \tilde{s} = s/\lambda^2 \leq t\}|] \\
&= E[|\{|\frac{x}{\lambda} - B_{\tilde{s}}| \leq \delta, 0 \leq \tilde{s} \leq t\}|] \\
&= \lambda^3 \cdot E[|W_\delta(t)|] ,
\end{aligned}$$

so that one assume without loss of generality that $\delta = 1$: knowing $E[|W_1(t)|$, we get the general case with the formula $E[|W_\delta(t)|] = \delta^3 \cdot E[|W_1(\delta^{-2}t)|]$.

Let K be the closed unit ball in $\mathbf{R}^d$. Define the hitting probability

$$f(x,t) = P[x + B_s \in K; 0 \leq s \leq t] .$$

We have

$$E[|W_1(t)|] = \int f(x,t)\ dx .$$

Proof.

$$\begin{aligned}
E[|W_1(t)|] &= \int \int P[x \in W_1(t)]\ dx\ dB \\
&= \int \int P[B_s - x \in K; 0 \leq s \leq t]\ dx\ dB \\
&= \int \int P[B_s - x \in K; 0 \leq s \leq t]\ dB\ dx \\
&= \int f(x,t)\ dx .
\end{aligned}$$

The hitting probability is radially symmetric and can be computed explicitely in terms of $r = |x|$: for $|x| \geq 1$, one has

$$f(x,t) = \frac{2}{r\sqrt{2\pi t}} \int_0^\infty e^{-\frac{(|x|+z-1)^2}{2t}}\ dz .$$

Proof. The kernel of e^{-tH} satisfies the heat equation $\partial_t P(x,0,t) = (\Delta/2)P(x,0,t)$ insided Ω. From the previous lemma follows that $\dot{f} = (\Delta/2)f$ so that the function $g(r,t) = rf(x,t)$ satisfies $\dot{g} = \frac{\partial^2}{2(\partial r)^2}g(r,t)$ with boundary condition $g(r,0) = 0, g(1,t) = 1$.
We compute

$$\int_{|x|\geq 1} f(x,t)\ dx = 2\pi t + 4\sqrt{2\pi t}$$

and $\int_{|x|\leq 1} f(x,t)\, dx = 4\pi/3$ so that

$$E[|W_1(t)| = 2\pi t + 4\sqrt{2\pi t} + 4\pi/3 \; .$$

□

Corollary 3.17.3

$$\lim_{\delta \to 0} \delta^{-1} \cdot E[|W_\delta(t)|] = 2\pi t$$

$$\lim_{t \to \infty} t^{-1} \cdot E[|W_\delta(t)|] = 2\pi\delta \; .$$

Proof. Immediate from Spitzer's result. □

Remarks.

1) If Brownian motion were one dimensional, then $\delta^{-2} E[|W_\delta(t)|]$ would be bounded as $\delta \to 0$. The corollary shows that the Wiener sausage is quite fat and indicates that Brownian motion is two dimensional.

2) Kesten, Spitzer and Wightman have got stronger results. It is even true that $\lim_{\delta \to 0} |W_\delta(t)|/t = 2\pi\delta$ and $\lim_{t \to \infty} |W_\delta(t)|/t = 2\pi\delta$ for almost all paths.

3.18 The Ito integral for Brownian motion

We start now to develop stochastic integration first for Brownian motion and then more generally for continuous martingales.

Motivation. We know that almost all paths of Brownian motion are not differentiable. The usual Stieljes integral

$$\int f(B_s)\dot{B}_s \, ds$$

can therefore not be defined and we are first going to see, how Palay defined a stochastic integral. Actually, we were already dealing with a special case of stochstic integrals, namely with Wiener integrals $\int f(B) \, dB$, where f is a function on $C([0,\infty],\mathbf{R}^d)$ which can contain for example $\int_0^t V(B_s) \, ds$ as in the Feynman Kac formula. But this was a number while the stochastic integral, we are going to define will be a random variable!

Definition. Let B_t be the one-dimensional Brownian motion process and let f be a function $f : \mathbf{R} \to \mathbf{R}$. Define for $n \in \mathbf{N}$ the random variable

$$J_n(f) = \sum_{m=1}^{2^n} f(B_{(m-1)2^{-n}})(B_{m2^{-n}} - B_{(m-1)2^{-n}}) =: \sum_{m=1}^{2^n} J_{n,m}(f) \; .$$

We will use later for $J_{n,m}(f)$ also the notation $f(B_{t_{m-1}})\delta_n B_{t_m}$.

Remark. In the last term, we have defined the discrete stochastic integral for a previsible process C and a martingale X

$$(\int C \, dX)_n = \sum_{m=1}^{n} C_m(X_m - X_{m-1}) \; .$$

If we want to take for C a function of X, then we have to take $C_m = f(X_{m-1})$. This is the reason, why we have to take the differentials $\delta_n B_{t_m}$ to "stick out into future".

The stochastic integral is a limit of discrete stochastic integrals:

Lemma 3.18.1 *If $f \in C^1(\mathbf{R})$ such that f, f' are bounded on $\mathbf{R}$, then J_n converges in $\mathcal{L}^2$ to a random variable*

$$\int_0^1 f(B_s)\, dB = \lim_{n \to \infty} J_n$$

satisfying

$$|| \int_0^1 f(B_s)\, dB||_2^2 = E[\int_0^1 f(B_s)^2\, ds] .$$

Proof.
(i) For $i \neq j$ we have $E[J_{n,i}(f) J_{n,j}(f)] = 0$.
Proof. For $j > i$, there is a factor $B_{j2^{-n}} - B_{(j-1)2^{-n}}$ of $J_{n,i}(f)J_{n,j}(f)$ independent of the rest of $J_{n,i}(f)J_{n,j}(f)$ and the claim follows from $E[B_{j2^{-n}} - B_{(j-1)2^{-n}}] = 0$.

(ii) $E[J_{n,m}(f)^2] = E[f(B_{(m-1)2^{-n}})^2]2^{-n}$.
Proof. $f(B_{(m-1)/2^n})$ is independent of $(B_{m2^{-n}} - B_{(m-1)2^{-n}})^2$ which has expectation 2^{-n}.

(iii) From (ii) follows

$$||J_n(f)||_2 = \sum_{m=1}^{2^n} E[f(B_{(m-1)2^{-n}})^2]2^{-n} .$$

(iv) The claim: J_n converges in $\mathcal{L}^2$.
Since $f \in C^1$, there exists $C = ||f'||_\infty^2$ and this gives $|f(x) - f(y)|^2 \leq C \cdot |x-y|^2$. We get

$$\begin{aligned} ||J_{n+1}(f) - J_n(f)||_2^2 &= \sum_{m=1}^{2^n-1} E[(f(B_{(2m+1)2^{-(n+1)}}) - f(B_{(2m)2^{-(n+1)}}))^2]2^{-(n+1)} \\ &\leq C \sum_{m=1}^{2^n-1} E[(B_{(2m+1)2^{-(n+1)}} - B_{(2m)2^{-(n+1)}})^2]2^{-(n+1)} \\ &= C \cdot 2^{-n-2} , \end{aligned}$$

where the last equality followed from the fact that $E[(B_{(2m+1)2^{-(n+1)}} - B_{(2m)2^{-(n+1)}})^2] = 2^{-n}$ since B is Gaussian. We see that J_n is Cauchy in $\mathcal{L}^2$ and has therefore a limit.

(v) The claim $|| \int_0^1 f(B_s)\, dB||_2^2 = E[\int_0^1 f(B_s)^2\, ds]$.
Proof. Since $\sum_m f(B_{(m-1)2^{-n}})^2 2^{-n}$ converges pointwise to $\int_0^1 f(B_s)^2\, ds$, (which exists because f and B_s are continuous), and is dominated by $||f||_\infty^2$, the claim follows since J_n converges in $\mathcal{L}^2$. $\square$

We can extend the integral to functions f, which are locally L^1 and bounded near 0. We write $L^p_{loc}(\mathbf{R})$ for funcions f which are in $L^p(I)$ when restricted to any finite interval I.

Corollary 3.18.2 *$\int_0^1 f(B_s)\, dB$ exists as a $\mathcal{L}^2$ random variable for $f \in L^1_{loc}(\mathbf{R}) \cap L^\infty(-\epsilon, \epsilon)$.*

Proof. (i) If $f \in L^1_{loc}(\mathbf{R}) \cap L^\infty(-\epsilon, \epsilon)$ for some $\epsilon > 0$, then

$$E[\int_0^1 f(B_s)^2\, ds] = \int_0^1 \int_{\mathbf{R}} \frac{f(x)^2}{\sqrt{2\pi s}} e^{-x^2/2s}\, dx\, ds \ < \infty \ .$$

(ii) If $f \in L^1_{loc}(\mathbf{R}) \cap L^\infty(-\epsilon, \epsilon)$, then for almost every $B(\omega)$, the limit

$$\lim_{a \to \infty} \int_0^1 1_{[-a,a]}(B_s) f(B_s)^2\, ds$$

exists pointwise and is finite.
Proof. B_s is continuous for almost all ω so that $1_{[-a,a]}(B_s)f(B)$ is independent of a for large a. The integral $E[\int_0^1 1_{[-a,a]}(B_s) f(B_s)^2\, ds]$ is bounded by $E[f(B_s)^2\, ds] < \infty$ by (i).

(iii) The claim.
Proof. Assume $f \in L^1_{loc}(\mathbf{R}) \cap L^\infty(-\epsilon, \epsilon)$. Given $f_n \in C^1(\mathbf{R})$ with $1_{[-a,a]} f_n \to f$ in $L^2(\mathbf{R})$.
By the domitated convergence theorem, we have

$$\int 1_[- a, a] f_n(B_s) dB \to \int 1_[- a, a] f(B_s)\, dB$$

in $\mathcal{L}^2$. Since by (ii), the $\mathcal{L}^2$ bound is independent of a, we can also pass to the limit $a \to \infty$. □

Definition. This integral is called an **Ito integral**. Having the one dimensional integral allows also for the d-dimensional Brownian motion and $f \in L^2_{loc}(\mathbf{R}^d)$ to define the integral $\int_0^1 f(B_s)\, dB_s$ and we get easily also $\int_s^t$ instead of $\int_0^1$.

Lemma 3.18.3

$$\sum_{j=1}^{2^n} J_{n,j}(1)^2 = \sum_{j=1}^{2^n} (B_{j/2^n} - B_{(j-1)/2^n})^2 \to 1 \ .$$

Proof. By definition of Brownian motion, we know that for fixed n, $J_{n,j}$ are $N(0, 2^{-n})$-distributed random variables and so

$$E[\sum_{j=1}^{2^n} J_{n,j}(1)^2] = 2^n \cdot \mathrm{Var}[B_{j/2^n} - B_{(j-1)/2^n}] = 1 \; .$$

Now, $X_j = 2^n J_{n,j}$ are IID $N(0,1)$-distributed random variables so that by the law of large numbers

$$\frac{1}{2^n} \sum_{j=1}^{2^n} X_j \to 1$$

for $n \to \infty$. □

The formal rules of integration do not hold for this integral. We have for example in one dimension:

$$\int_0^1 B_s \, dB = \frac{1}{2}(B_1^2 - 1) \neq \frac{1}{2}(B_1^2 - B_0^2) \; .$$

Proof. Define

$$\begin{aligned} J_n^- &= \sum_{m=1}^{2^n} f(B_{(m-1)2^{-n}})(B_{m2^{-n}} - B_{(m-1)2^{-n}}) \; , \\ J_n^+ &= \sum_{m=1}^{2^n} f(B_{m2^{-n}})(B_{m2^{-n}} - B_{(m-1)2^{-n}}) \; . \end{aligned}$$

The above lemma implies that $J_n^+ - J_n^- \to 1$ almost everywhere for $n \to \infty$ and we check also $J_n^+ + J_n^- = B_1^2$. Both of these identities come from cancellations in the sum and imply together the claim. □

We mention now some trivial properties of the stochastic integral.

Proposition 3.18.4
(1) $\int_0^t f(B_s) + g(B_s) \; dB = \int_0^t f(B_s) \; dB + \int_0^t g(B_s) \; dB$.
(2) $\int_0^t \lambda \cdot f(B_s) \; dB = \lambda \cdot \int_0^t f(B_s) \; dB$.
(3) $t \mapsto \int_0^t f(B_s) \; dB$ *is a continuous map from* $[0, \mathbf{R})$ *to* $\mathcal{L}^2$.

Proof. (1) and (2) follow from the definition of the integral.
For (3) define $X_t = \int_0^t f(B_s)\, dB$. Since

$$\begin{aligned} ||X_t - X_{t+\epsilon}||_2^2 &= E[\int_t^{t+\epsilon} f(B_s)^2\, ds] \\ &= \int_t^{t+\epsilon} \int_{\mathbf{R}} \frac{f(x)^2}{\sqrt{2\pi s}} e^{-x^2/2s}\, dx\, ds \to 0 \end{aligned}$$

for $\epsilon \to 0$, the claim follows. □

It will be useful to consider an other generalisations of the integral.

Definition.
Let $dW = dxdB$ be the Wiener measure on $\mathbf{R}^d \times C([0,\infty)$. We define also

$$\int_0^t f(W_s)\, dW_s = \int_{\mathbf{R}^d} \int_0^t f(x + B_s)\, dB_s\, dx\ .$$

Definition.
Assume f is also time dependent so that it is a function on $\mathbf{R}^d \times \mathbf{R}$. As long as $E[\int_0^1 |f(B_s,s)|^2\, ds] < \infty$, we can also define the integral

$$\int_0^t f(B_s, s)\, ds\ .$$

3.19 Ito's formula

Ito's formula is very useful for understanding and calculating stochastic integrals. It is the fundamental theorem of calculus for stochastic integrals.

Theorem 3.19.1 (Ito's formula) *Given a function $f(x)$ on $\mathbf{R}^d$, which is twice differentiable in x. Then*

$$f(B_t) - f(B_0) = \int_0^t \nabla f(B_s) \cdot \, dB_s + \frac{1}{2} \int_0^t \Delta f(B_s) \, ds .$$

If B_s would be an ordinary path in $\mathbf{R}^d$ with velocity vector $dB_s = \dot{B}_s \, ds$, then we had

$$f(B_t) - f(B_0) = \int_0^t \nabla f(B_s) \cdot \, \dot{B}_s .$$

It is a bit surprising that in the stochastic setup, a second derivative Δf appears in a first order differential. One writes sometimes the formula also in the differential form

$$df = \nabla f \, dB + \frac{1}{2} \Delta f \, dt .$$

We will prove a slightly more general formula for functions f, which can also be time dependent.

Theorem 3.19.2 (Generalized Ito's formula) *Given a function $f(x,t)$ on $\mathbf{R}^d \times [0,t]$ which is twice differentiable in x and differentiable in t. Then*

$$f(B_t, t) - f(B_0, 0) = \int_0^t \nabla f(B_s, s) \cdot \, dB_s + \frac{1}{2} \int_0^t \Delta f(B_s, s) \, ds + \int_0^t \dot{f}(B_s, s) \, ds .$$

In differential notation, this means

$$df = \nabla f \, dB + (\frac{1}{2} \Delta f + \dot{f}) \, dt .$$

Proof. By a change of variables, we can assume $t = 1$. For each n, we discretize time $\{0 < 2^{-n} < \dots, t_k = k \cdot 2^{-n}, \cdots, 1\}$ and define $\delta_n B_{t_k} = B_{t_k} - B_{t_{k-1}}$. We write

$$
\begin{aligned}
f(B_1,1) - f(B_0,0) &= \sum_{k=1}^{2^n} (\nabla f)(B_{t_{k-1}}, t_{k-1}) \Delta_n B_{t_k} \\
&+ \sum_{k=1}^{2^n} f(B_{t_k}, t_{k-1}) - f(B_{t_{k-1}}, t_{k-1}) - (\nabla f)(B_{t_{k-1}}, t_{k-1}) \delta_n B_{t_k} \\
&+ \sum_{k=1}^{2^n} f(B_{t_k}, t_k) - f(B_{t_k}, t_{k-1}) \\
&= I_n + II_n + III_n \; .
\end{aligned}
$$

(i) By definition of the Ito integral, the first sum I_n converges in $\mathcal{L}^2$ to $\int_0^1 (\nabla f)(B_s, s) \; dB_s$.

(ii) If $p > 2$, we have $\sum_{k=1}^{2^n} |\delta_n B_{t_k}|^p \to 0$ for $n \to \infty$.
Proof. $\delta_n B_{t_k}$ is a $N(0, 2^{-n})$-distributed random variable so that

$$
E[|\delta_n B_{t_k}|^p] = (2\pi)^{-1/2} 2^{-(np)/2)} \int |x|^p e^{-x^2/2} \; dx = C 2^{-(np)/2} \; .
$$

This means

$$
E[\sum_{k=1}^{2^n} |\delta_n B_{t_k}|^p] = C 2^n 2^{-(np)/2}
$$

which goes to zero for $n \to \infty$.

(iii) $\sum_{k=1}^{2^n} E[g(B_{t_k}, t_k)^2 (B_{t_k} - B_{t_{k-1}})^p] \to 0$ for $p > 2$ follows from ii. We have alsousing (ii)

$$
\begin{aligned}
\sum_{k=1}^{2^n} E[g(B_{t_k}, t_k)^2 ((B_{t_k} - B_{t_{k-1}})^2 - 2^{-n})^2] \quad leq & \quad C \sum_{k=1}^{2^n} \mathrm{Var}[(B_{t_k} - B_{t_{k-1}})^2] \\
\leq & \quad C \sum_{k=1}^{2^n} E[(B_{t_k} - B_{t_{k-1}})^4] \; .
\end{aligned}
$$

(iv) Using Taylor

$$
|f(x) - f(y) - \nabla f(y)(x-y) - \frac{1}{2} \partial_{x_i x_j} f(y)(x-y)_i (x-y)_j + O(|x-y|^3
$$

we get for $n \to \infty$

$$
II_n - \sum_{k=1}^{2^n} \frac{1}{2} \sum_{i,j} \partial_{x_i x_j} f(B_{t_{k-1}}, t_{k-1}) (\delta_n B_{t_k})_i (\delta_n B_{t_k})_j \to 0
$$

in $\mathcal{L}^2$. Since

$$
\sum_{k=1}^{2^n} \frac{1}{2} \partial_{x_i x_j} f(B_{t_{k-1}}, t_{k-1}) [(\delta_n B_{t_k})_i (\delta_n B_{t_k})_j - \delta_{ij} 2^{-n}]
$$

goes to zero in $\mathcal{L}^2$, we have

$$II_n \to \frac{1}{2} \int_0^t \Delta f(B_s, s) \; ds$$

in $\mathcal{L}^2$.

(v) Taylor with respect to t

$$f(x,t) - f(x,s) - \dot{f}(x,s)(t-s) + O((t-s)^2)$$

gives

$$III_n \to \int_0^t \dot{f}(B_s, s) \; ds$$

in $\mathcal{L}^1$. □

Examples.

1) Consider the function

$$f(x,t) = e^{\alpha x - \alpha^2 t/2} \; .$$

Since this function satisfies $\dot{f} + f''/2 = 0$, we get from Ito's formula

$$f(B_t, t) - f(B_0, t) = \alpha \int_0^t f(B_s, s) \cdot \; dB_s \; .$$

We see that for functions satisfying $\dot{f} + f''/2 = 0$ Ito's formula reduces to the usual rule of calculus.

2) If we make a power expansion in α of the in 1) proven formula

$$\int_0^t e^{\alpha B_s - \alpha^2 s/2} \; dB = \alpha^{-1} e^{\alpha B_s - \alpha^2 s/2} - \alpha^{-1} \; ,$$

we get other formulas like for example

$$\int_0^t B_s \; dB = \frac{1}{2}(B_t^2 - t) \; .$$

Wick ordering.
There is a notation used in quantum field theory developed by Wick at about the same time as Ito's invented the integral. This **Wick ordering** is a map on polynomials $\sum_{i=1}^n a_i x^i$ which leave monomials (polynomials of the form $x^n + a_{n-1} x^{n-1} \cdots$) invariant.

Definition. Let $\Omega_n = (n!)^{-1/2} H_n(x/\sqrt{2}) \Omega_0(y)$ be the n'-th eigenfunction of the quantum mechanical oscillator. H_n is called the the n-th **Hermite polynomial**. Define

$$: x^n := 2^{-n/2} H_n(x/\sqrt{2})$$

and extend the definition to all polynomials by linearity. The Polynomials $: x^n :$ are orthogonal with respect to the measure $\Omega_0^2 dy = \pi^{-1/2} e^{-y^2}\, dy$ since we have seen that the Ω_n are orthnormal.

Examples computed with the mathematica line

```
Wick[n_]:=Simplify[N[HermiteH[n,x/N[Sqrt[2]]] N[2^(-n/2)]]]
```

$$\begin{aligned} : x : &= x \\ : x^2 : &= x^2 - 1 \\ : x^3 : &= x^3 - 3x \\ : x^4 : &= x^4 - 6x^2 + 3 \\ : x^5 : &= x^5 - 10x^3 + 15x \ . \end{aligned}$$

The variable x is an operator $f \mapsto xf$ and called the **position operator** denoted with Q. We can write it by definition of the creation and annihilation operator as $Q = \sqrt{2}(A + A^*)$. The following formula indicates, why Wick ordering has its name and why it is useful in quanum mechanics: as operators, we have the identity

$$: Q^n := 2^{-n/2} : (A + A^*)^n := 2^{-n/2} \sum_{j=0}^{n} \binom{n}{j} (A^*)^j A^{n-j} =: 2^{-n/2} L \ .$$

The new ordering made the operators A, A^* behave as if they were commutative (they are not, since we have seen the commutation relations $[A, A^*] = 1$)!

Proof. Since we know that Ω_n forms a basis in L^2, we have only to show that $: Q^n : \Omega_k = 2^{-n/2} L \Omega_k$ for all k. From

$$\begin{aligned} 2^{-1/2}[Q, L] &= [A + A^*, \sum_{j=0}^{n} \binom{n}{j} (A^*)^j A^{n-j}] \\ &= \sum_{j=0}^{n} \binom{n}{j} j (A^*)^{j-1} A^{n-j} - (n-j)(A^*)^j A^{n-j-1} \\ &= 0 \end{aligned}$$

we obtain by linearity $[H_k(\sqrt{2}Q), L]$. Because $: Q^n : \Omega_0 = 2^{-n/2}(n!)^{1/2}\Omega_n = 2^{-n/2}(A^*)^n \Omega_0 = 2^{-n/2} L \Omega_0$, we get

$$\begin{aligned} 0 &= (: Q^n : -2^{-n/2} L)\Omega_0 \\ &= (k!)^{-1/2} H_k(\sqrt{s}Q)(: Q^n : -2^{-n/2} L)\Omega_0 \\ &= (: Q^n : -2^{-n/2} L)(k!)^{-1/2} H_k(\sqrt{s}Q)\Omega_0 \\ &= (: Q^n : -2^{-n/2} L)\Omega_k \ . \end{aligned}$$

$\square$

That stochastic integration has to do with quantum mechanics can be seen from the following formula for the Ito integral:

$$\boxed{\int_0^t :B_s^n: \; dB = \frac{1}{n+1} :B_t^{n+1}:}$$

Therefore, the Wick ordering makes the Ito integral behave like an ordinary integral.

Proof. By rescaling, we can assume that $t = 1$.

We prove all these equalities simultaneously by showing

$$\int_0^1 :e^{\alpha B_s}: \; dB = \alpha^{-1} :e^{\alpha B_1}: - \alpha^{-1} .$$

The generating function for the Hermite polynomials is known to be

$$\sum_{n=0}^{\infty} H_n(x) \frac{\alpha^n}{n!} = e^{\alpha \sqrt{2} x - \frac{\alpha^2}{2}} .$$

(We can check this formula by multiplying it with Ω_0, replacing x with $x/\sqrt{2}$ so that we have

$$\sum_{n=0}^{\infty} \frac{\Omega_n(x) \alpha^n}{(n!)^{1/2}} = e^{\alpha x - \frac{\alpha^2}{2} - \frac{x^2}{2}} .$$

If we apply A^* on both sides, the equation goes onto itself and we get after k such applications of A^* that that the scalar product with Ω_k is the same on both sides. Therefore the functions must be the same.)
This means

$$:e^{\alpha x} := \sum_{j=0}^{\infty} \frac{\alpha^n :x^n:}{n!} = e^{\alpha x - \frac{1}{2}\alpha^2} .$$

Since the right hand side satisfies $\dot{f} + f''/2 = 2$, the claim follows from the Ito formula for such functions. $\square$

We can now determine all the integrals $\int B_s^n \, dB$:

$$\begin{aligned}
\int_0^t B_s \, dB &= \frac{1}{2}(B_t^2 - 1) \\
\int_0^t B_s^2 \, dB &= \int_0^t :B_s^2: +1 \, dB = B_t + \frac{1}{3}(:B_t:^3) = B_t + \frac{1}{3}(B_t^3 - 3B_t)
\end{aligned}$$

and so on.

Stochastic integrals for the oscillator and the Bridge process.
Let $Q_t = e^{-t} B_{e^{2t}}/\sqrt{2}$ the oscillator process and $A_t = (1-t) B_{t/(1-t)}$ the Brownian bridge. If we define new discrete differentials

$$\begin{aligned} \delta_n Q_{t_k} &= Q_{t_{k+1}} - e^{-(t_{k+1}-t_k)} Q_{t_k} \\ \delta_n A_{t_k} &= A_{t_{k+1}} - A_{t_k} + \frac{t_{k+1}-t_k}{(1-t)} A_{t_k} \end{aligned}$$

the stochastic integrals can be defined as in the case of Brownian motion as a limit of discrete integrals.

Feynman-Kac formula for Schrödinger operators with magnetic fields.
Stochastic integrals appear in the Feynman-Kac formula for particles moving in a magnetic field. Let $a(x)$ be a vector potential in $\mathbf{R}^3$ which gives the magnetic field $B(x) = \mathrm{rot} a$. Quantum mechanically, a particle moving in an magnetic field together with an external field is described by the Hamiltonian

$$H = (i\nabla + a)^2 + V \; .$$

In the case $a = 0$, we get the usual Schrödinger operator. The Feynman-Kac formula is the Wiener integral

$$e^{-tH} u(0) = \int e^{-F(B,t)} u(B_t) \; dB \; ,$$

where $F(B,t)$ is a stochastic integral.

$$F(B,t) = i \int a(B_s) \; dB + \frac{i}{2} \int_0^t \mathrm{div}(a) \; ds + \int_0^t V(B_s) \; ds \; .$$

3.20 Processes of bounded quadratic variation

We develop now the stochastic Ito integral with respect to general martingales. Brownian motion B will be replaced by a martingale M which are in $\mathcal{L}^2$. The aim will be to define an integral

$$\int_0^t K_s \, dM_s \,,$$

where K is a progressively measurable process which satisfies some boundedness condition.

Definition. Given a right continuous function $f : [0, \infty) \to \mathbf{R}$. For each finite subdivision $\Delta = \{0 = t_0, t_1, \dots, t = t_n\}$ of the interval $[0, t]$ we define $|\Delta| = \sup_{i=1}^r |t_{i+1} - t_i|$ called the **modulus** of Δ. Define

$$||f||_\Delta = \sum |f_{t_{i+1}} - f_{t_i}| \,.$$

A function with finite total variation $||f||_t = \sup_\Delta ||f||_\Delta < \infty$ is called a function of **finite variation**. If $\sup_t |f|_t < \infty$, then f is called **of bounded variation** (BV).

Examples. C^1 functions are of finite variation. Monotone finite functions are of finite variation, sums of BV functions are of bounded variation.

Property. Every function of finite variation can be written as $f = f^+ - f^-$, where $f^\pm$ are positive and increasing. Take just $f^\pm = (\pm f_t + ||f||_t)/2$. Functions of bounded variation are in one to one correspondence to Borel measures on $[0, \infty)$ by the Stieltjes integral $\int_0^t |df| = f_t^+ + f_t^-$.

Definition. A process X_t is called **increasing** if the paths $X_t(\omega)$ are finite, right-continuous and increasing for almost all $\omega \in \Omega$. A process X_t is called of **finite variation**, if the paths $X_t(\omega)$ are finite, right-continuous and of finite variation for almost all $\omega \in \Omega$.

Remark. Every bounded variation process A can be written as $A_t = A_t^+ - A_t^-$, where $A_t^\pm$ are increasing. The process $V_t = \int_0^t |dA|_s = A_t^+ + A_t^-$ is increasing and we get for almost all $\omega \in \Omega$ a measure called the **variation** of A.

If X_t is a bounded $\mathcal{A}_t$-adapted process and A is a process of bounded variation, we can form the Stieltjes integral

$$(X \cdot A)_t(\omega) = \int_0^t X_s(\omega) dA_s(\omega) \,.$$

We would like to define such an integral for martingales. The problem is:

Proposition 3.20.1 *A continuous martingale M is never of finite variation unless it is constant.*

Proof. Assume M is of finite variation. We show that it is constant.

(i) We can assume without loss of generality that M is of bounded variation. Proof. Consider else the martingale M^{S_n}, where S_n is the stopping time $S_n = \inf\{s \mid V_s \geq n\}$ and V_t is the variation of M on $[0,t]$.
(ii) We can also assume also without loss of generality that $M_0 = 0$.

(iii) Let $\Delta = \{t_0 = 0, t_1, \dots, t_n = t\}$ be a subdivision of $[0,t]$. Since M is a martingale, we have (use Pytagoras)

$$\begin{aligned} E[M_t^2] &= E[\sum_{i=0}^{k-1}(M_{t_{i+1}}^2 - M_{t_i}^2)] \\ &= E[\sum_{i=1}^{k-1}(M_{t_{i+1}} - M_{t_i})(M_{t_{i+1}} + M_{t_i})] \\ &= E[\sum_{i=1}^{k-1}(M_{t_{i+1}} - M_{t_i})^2] \end{aligned}$$

and so

$$E[M_t^2] \leq E[V_t(\sup_i |M_{t_{i+1}} - M_{t_i}|] \leq K \cdot E[\sup_i |M_{t_{i+1}} - M_{t_i}|] .$$

If the modulus $|\Delta|$ goes to zero, then the right hand side goes to zero since M is continuous. Therefore $M = 0$. □

Remark. This applies especially for Brownian motion and underlines again the fact that the stochastic integral could not be defined pointwise by a Stieljes integral.

Definition. If $\Delta = \{t_0 = 0 < t_1 < \dots ...\}$ is a subdivision of $\mathbf{R}^+$ with only finitely many points $\{t_0, t_1, \dots, t_k\}$ in each interval $[0,t]$ we define for a process X

$$T_t^\Delta = T_t^\Delta(X) = (\sum_{i=0}^{k-1}(X_{t_{i+1}} - X_{t_i})^2) + (X_t - X_{t_k})^2 .$$

The process X is called of **finite quadratic variation**, if there exists a process $< X, X >$ such that for each t, T_t^Δ converges in probability to $< X, X >$ as $|\Delta| \to 0$.

Theorem 3.20.2 (Doob-Meyer decomposition in special case) *Given a continuous and bounded martingale M of finite quadratic variation.*
$< M, M >$ is the unique continuous increasing adapted process vanishing at zero such that $M^2 - < M, M >$ is a martingale.

Remark. Before we go to the not so easy proof, let us compaire that with the discrete case, where we remarked that M^2 was a submartingale and that M^2 could be written uniquely as a sum of a martingale and an increasing previsible process.

Proof. Uniqueness follows from the previous proposition: if there would be two such processes A, B, then $A - B$ would be a continuous martingale with bounded variation (since A and B are increasing and so of bounded variation) which vanishes at 0. Therefore $A = B$.

(i) $M_t^2 - T_t^\Delta(M)$ is a continuous martingale.
Proof. For $t_i < s < t_{i+1}$, we have from the martingale property using that $(M_{t_{i+1}} - M_s)^2$ and $(M_s - M_{t_i})^2$ are independent,

$$E[(M_{t_{i+1}} - M_{t_i})^2 \mid \mathcal{A}_s] = E[(M_{t_{i+1}} - M_s)^2 | \mathcal{A}_s] + (M_s - M_{t_i})^2 .$$

This implies with $0 = t_0 < t_1 < \ldots < t_l < s < t_{l+1} < \ldots < t_k < t$ and using orthogonality

$$\begin{aligned} E[T_t^\Delta(M) - T_s^\Delta(M) | \mathcal{A}_s] &= E[\sum_{j=l}^{k} (M_{t_{j+1}} - M_{t_j})^2 | \mathcal{A}_s] \\ &+ E[(M_t - M_{t_k})^2 | \mathcal{A}_s] + E[(M_s - M_{t_l})^2 | \mathcal{A}_s] \\ &= E[(M_t - M_s)^2 | \mathcal{A}_s] = E[M_t^2 - M_s^2 | \mathcal{A}_s] . \end{aligned}$$

This implies that $M_t^2 - T_t^\Delta(M)$ is a continuous martingale.

(ii) Let C be a constant such that $|M| \leq C$ in $[0, a]$. Then $E[T_a^\Delta \leq 4C^2]$, independent of the subdivision $\Delta = \{t_0, \ldots, t_n\}$ of $[0, a]$.
Proof. We can assume $t_n = a$. We write

$$\begin{aligned} (T_a^\Delta)^2 &= (\sum_{k=1}^{n} (M_{t_k} - M_{t_{k-1}})^2)^2 \\ &= 2 \sum_{k=1}^{n} (T_a^\Delta - T_{t_k}^\Delta)(T_{t_k}^\Delta - T_{t_{k-1}}^\Delta) + \sum_{k=1}^{n} (M_{t_k} - M_{t_{k-1}})^4 . \end{aligned}$$

From (i), we have

$$E[T_a^\Delta - T_{t_k}^\Delta | \mathcal{A}_{t_k}] = E[(M_a - M_{t_k})^2 \mid \mathcal{A}_{t_k}]$$

and consequently

$$\begin{aligned} E[(T_a^\Delta)^2] &= 2\sum_{k=1}^n E[(M_a - M_{t_k})^2(T_{t_k}^\Delta - T_{t_{k+1}}^\Delta)] + \sum_{k=1}^n E[(M_{t_k} - M_{t_{k-1}})^4] \\ &\leq E[(2\sup_k |M_a - M_{t_k}|^2 + \sup_k |M_{t_k} - M_{t_{k-1}}|^2) T_a^\Delta] \; . \end{aligned}$$

(iii) For fixed $a > 0$ and subdivisions Δ_n of $[0, a]$, the sequence $T_a^{\Delta_n}$ has a limit in $\mathcal{L}^2$ for $|\Delta_n| \to 0$.
Proof. Given two subdivisions Δ, Δ' of $[0, a]$, let $\Delta\Delta'$ be the subdivision obtained by taking all the point of Δ and Δ'. By (i), the process $X = T^\Delta - T^{\Delta'}$ is a martingale and by (i) again, applied to X instead of M we have using $(x + y)^2 \leq 2(x^2 + y^2)$

$$E[X_a^2] = E[(T_a^\Delta - T_a^{\Delta'})^2] = E[T_a^{\Delta\Delta'}(X)] \leq 2(E[T_a^{\Delta\Delta'}(T^\Delta)] + E[T_a^{\Delta\Delta'}(T^{\Delta'})]) \; .$$

We have therefore only to show that $E[T_a^{\Delta\Delta'}(T^\Delta)] \to 0$ for $|\Delta| + |\Delta'| \to 0$. Let s_k be in $\Delta\Delta'$ and t_m the rightmost point in Δ such that $t_m \leq s_k < s_{k+1} \leq t_{m+1}$. We have

$$\begin{aligned} T_{s_{k+1}}^\Delta - T_{s_k}^\Delta &= (M_{s_{k+1}} - M_{t_m})^2 - (M_{s_k} - M_{t_m})^2 \\ &= (M_{s_{k+1}} - M_{s_k})(M_{s_{k+1}} + M_{s_k} - 2M_{t_m}) \end{aligned}$$

and so

$$T_a^{\Delta\Delta'} \leq (\sup_k \; |M_{s_{k+1}} + M_{s_k} - 2M_{t_m}|^2) T_a^{\Delta\Delta'} \; .$$

By Schwarz-inequality

$$E[T_a^{\Delta\Delta'}] \leq (E[\sup_k \; |M_{s_{k+1}} + M_{s_k} - 2M_{t_1}|^4)^{1/2} E[(T_a^{\Delta\Delta'})^2]^{1/2}$$

and the first factor goes to 0 as $|\Delta| + |\Delta'| \to 0$ and the second factor is bounded since by (ii) $E[(T_a^{\Delta\Delta'})^2]^{1/2} \leq 12 \cdot C^2 E[T_a^\Delta] \leq 48C^4$.

(iv) There exists a sequence of $\Delta_n \subset \Delta_{n+1}$ such that $T_t^{\Delta_n}$ converges uniformly to a limit $< M, M >$ on $[0, a]$.
Proof. Doob's inequality applied to the discrete time martingale $T^{\Delta_n} - T^{\Delta_m}$ gives

$$E[\sup_{t \leq a} |T_t^{\Delta_n} - T_t^{\Delta_m}|^2 \leq 4E[(T_a^{\Delta_n} - T_a^{\Delta_m})^2] \; .$$

Choose the sequence Δ_n such that Δ_{n+1} is a refinement of Δ_n and such that $\bigcup_n \Delta_n$ is dense in $[0, a]$, we can acchieve that the convergence is uniform. The limit $< M, M >$ is therefore continuous.

(v) $< M, M >$ is increasing.
Proof. Take $\Delta_n \subset \Delta_{n+1}$. For any pair $s < t$ in $\bigcup_n \Delta_n$, we have $T_s^{\Delta_n} \leq T_t^{\Delta_n}$ if n is so large that Δ_n contains both s and t. Therefore $< M, M >$ is increasing on $\bigcup_n \Delta_n$ which can be choosen to be dense. The continuity implies that $< M, M >$ is increasing everywhere. □

Remark. The fact that this result is for bounded martingales is not essential. It holds for general martingales and more generally for so called local martingales. One takes a sequence of bounded stopping times T_n increasing to ∞ and approximates a general martingale M by bounded martingales $M^{T_n} 1_{\{T_n>0\}}$ (if this is possible, then M is called a local martingale).

Corollary 3.20.3 *Let M, N be two continuous martingales with the same filtration. There exists a unique continuous adapted process $< M, N >$ of finite variation which is vanishing at zero and such that $MN- < M, N >$ is a martingale.*

Proof. Uniqueness follows again from the fact that a finite variation martingale must vanish.
To get existence, use the parallelogram law

$$< M, N >= \frac{1}{4}(< M + N, M + N > - < M - N, M - N >) .$$

This is vanishing at zero and of finite variation since it is a sum of two processes with this property.
We know that $M^2- < M, M >, N^2- < N, N >$ and so that $(M \pm N)^2- < M \pm N, M \pm N >$ are martingales. Therefore

$$\begin{aligned} & (M + N)^2- < M + N, M + N > -(M - N)^2- < M - N, M - N > \\ = \ & 4MN- < M + N, M + N > - < M - N, M - N > \ . \end{aligned}$$

and $MN- < M, N >$ is a martingale. $\square$

Definition. The process $< M, N >$ is called the **bracket** of M and N and $< M, M >$ the increasing process of M.

Example. If $B = (B^{(1)}, \dots, B^{(d)})$ is Brownian motion, then $< B^{(i)}, B^{(j)} >= \delta_{ij} \cdot t$ as we have computed in the proof of the Ito formula in the case $t = 1$.
It can be shown that every martingale M which has the property $< M^{(i)}, M^{(j)} >= \delta_{ij} \cdot t$ must be Brownian motion. This is **Lévy's characterisation of Brownian motion**.

Remark. If M is a martingale vanishing at zero and $< M, M >= 0$, then $M = 0$. Since $M_t^2- < M, M >_t$ is a martingale vanishing at zero, we have $E[M_t^2] = E[< M, M >_t]$.

Remark. Since we have got $< M, M >$ as a limit of processes T_t^{Δ}, we could also write $< M, N >$ as such a limit.

3.21 The Ito integral for martingales

In the last section, we have defined for two continuous martingales M,N, the bracket process $< M, N >$. Since $< M, M >$ was increasing, it was of finite variation and therefore also $< M, N >$ is of finite variation. It defines a random measure $d < M, N >$.

Theorem 3.21.1 (Kunita-Watanabe inequality) *Let M, N be two continuous margtingales and H, K two measurable processes. Then for all $p, q \geq 1$ satisfying $1/p + 1/q = 1$, we have for all $t \leq \infty$*

$$\begin{aligned} E[\int_0^t |H_s|\, |K_s|\, |d < M, N >_s|] &\leq ||(\int_0^t H_s^2 d < M, M >)^{1/2}||_p \\ &\cdot\ ||(\int_0^t K_s^2 d < N, N >)^{1/2}||_q \,. \end{aligned}$$

Proof. (i) Define $< M, N >_s^t = < M, N >_t - < M, N >_s$. Claim: almost surely

$$| < M, N >_s^t | \leq (< M, M >_s^t)^{1/2} (< N, N >_s^t)^{1/2} \,.$$

Proof. For fixed r, the random variable

$$< M, M >_s^t + 2r < M, N >_s^t + r^2 < N, N >_s^t = < M + rN, M + rN >_s^t$$

is positive almost everywhere and this stays true simultaneously for a dense set of $r \in \mathbf{R}$. Since M, N are continuous, it holds for all r. The claim follows, since $a + 2rb + cr^2 \geq 0$ for all $r \geq 0$ with nonnegative a, c implies $b \leq \sqrt{a}\sqrt{c}$.

(ii) To prove the claim, it is, using Hölder's inequality, enough to show almost everywhere, the inequality

$$\int_0^t |H_s|\, |K_s|\, d| < M, N > |_s \leq (\int_0^t H_s^2 d < M, M >)^{1/2} \cdot (\int_0^t K_s^2 d < N, N >)^{1/2}$$

holds. By taking limits, it is enough to prove this for $t < \infty$ and bounded K, H. By a density argument, we can also assume the both K and H are step functions $H = \sum_{i=1}^n H_i 1_{J_i}$ and $K = \sum_{i=1}^n K_i 1_{J_i}$, where $J_i = [t_i, t_{i+1})$.

(iii) We get from (i) for step functions H, K as in (ii)

$$\begin{aligned} |\int_0^t H_s K_s d < M, N >_s | &\leq \sum_i |H_i K_i| | < M, N >_{t_i}^{t_{i+1}} | \\ &\leq \sum_i |H_i K_i| (< M, M >_{t_i}^{t_{i+1}})^{1/2} (< M, M >_{t_i}^{t_{i+1}})^{1/2} \end{aligned}$$

$$\leq \quad (\sum_i H_i^2 < M, M >_{t_i}^{t_{i+1}})^{1/2} (\sum_i K_i^2 < N, N >_{t_i}^{t_{i+1}})^{1/2}$$
$$= \quad (\int_0^t H_s^2 d < M, M >)^{1/2} \cdot (\int_0^t K_s^2 d < N, N >)^{1/2} ,$$

where we have used Cauchy-Schwarz inequality for the summation over i. □

Definition. Denote with $\mathcal{H}^2$ the set of $\mathcal{L}^2$-martingales which are $\mathcal{A}_t$-adapted and satisfy

$$||M||_{\mathcal{H}_2} = (\sup_t E[M_t^2])^{1/2} < \infty .$$

Call H^2 the subset of continuous martingales in $\mathcal{H}^2$ and with H_0^2 the subset of continuous martingales which are vanishing at zero.
Given a martingale $M \in \mathcal{H}^2$, we define $\mathcal{L}^2(M)$ the space of progressively measurable processes K such that

$$||K||^2_{\mathcal{L}^2(M)} = E[\int_0^\infty K_s^2 d < M, M >_s] < \infty .$$

Both $\mathcal{H}^2$ and $\mathcal{L}^2(M)$ are Hilbert spaces.

Lemma 3.21.2 *The space H^2 of continuous martingales is closed in $\mathcal{H}^2$ and so a Hilbert space. Also H_0^2 is closed in H^2 and is therefore a Hilbert space.*

Proof. Take a sequence $M^{(n)}$ in H^2 converging to $M \in \mathcal{H}^2$. By Doob's inequality

$$E[(\sup_t |M_t^n - M_t|)^2] \leq 4||M^{(n)} - M||^2_{\mathcal{H}^2} .$$

We can extract a subsequence, for which $\sup_t |M_t^{(n_k)} - M_t|$ converges pointwise to zero almost everywhere. Therefore $M \in H^2$. The same argument shows also that H_0^2 is closed. □

Proposition 3.21.3 *Given $M \in H^2$ and $K \in \mathcal{L}^2(M)$. There exists a unique element $\int_0^t KdM \in H_0^2$ such that*

$$< \int_0^t KdM, N >= \int_0^t Kd < M, N >$$

for every $N \in H^2$. The map $K \mapsto \int_0^t KdM$ is an isometry form $\mathcal{L}^2(M)$ to H_0^2.

Proof. We can assume $M \in H_0$ since in general, we define $\int_0^t K \ dM = \int_0^t K \ d(M - M_0)$.

(i) By the Kunita-Watanabe inequality, we have for every $N \in H_0^2$

$$|E[\int_0^t K_s d < M, N >_s]| \leq ||N||_{\mathcal{H}^2} \cdot ||K||_{\mathcal{L}^2(M)} .$$

The map

$$N \mapsto E[(\int_0^t K_s) \; d < M, N >_s]$$

is therefore a linear continuous functional on the Hilbert space H_0^2. By Riesz representation theorem, there is an element $\int K \; dM \in H_0^2$ such that

$$E[(\int_0^t K_s \; dM_s) N_t] = E[\int_0^t K_s d < M, N >_s]$$

for every $N \in H_0^2$.

(ii) Uniqueness. Assume there exist two martingales $L, L' \in H_0^2$ such that $< L, N > = < L', N >$ for all $N \in H_0^2$. Then, in particular, $< L - L', L - L' > = 0$, from which $L = L'$ follows.

(iii) The integral $K \mapsto \int_0^t K \; dM$ is an isometry because

$$\begin{aligned} || \int_0^t K \; dM ||_{\mathcal{H}_0}^2 &= E[(\int_0^\infty K_s \; dM_s)^2] \\ &= E[\int_0^\infty K_s^2 \; d < M, M >] \\ &= ||K||_{\mathcal{L}^2(M)}^2 . \end{aligned}$$

□

Definition. The martingale $\int_0^t K_s \; dM_s$ is called the **Ito integral** of the progressively measureable process K with respect to the martingale M. We can take especially, $K = f(M)$, since continuous processes are progressively measurable. If we take $M = B$, Brownian motion, we get the previous integral.

Definition. An $\mathcal{A}_t$ adapted right-continuous process is called a **local martingale** if there exists a sequence T_n of increasing stopping times with $T_n \to \infty$ almost everywhere, such that for every n, the process $X^{T_n} 1_{\{T_n > 0\}}$ is a uniformly integrable $\mathcal{A}_t$-martingale. Local martingales are more general than martingales.

Remark. As indicated below, stochastic integration can be defined more generally for local martingales.

We show quickly, that Ito's formula holds also for general martingales. First, a special case, the integration by parts formula.

Theorem 3.21.4 (Integration by parts) *Let X, Y be two continuous martingales. Then*

$$X_tY_t - X_0Y_0 = \int_0^t X_s \, dY_s + \int_0^t Y_s dX_s + < X, Y >_t$$

and especially

$$X_t^2 - X_0^2 = 2\int_0^t X_s \, dX_s + < X, X >_t \quad .$$

Proof. The general case follows from the special case by polarisation: use the special case for $X \pm Y$ as well as X and Y.
The special case is proven by discretisation: let $\Delta = \{t_0, t_1, \dots, t_n\}$ be a finite discretisation of $[0, t]$. Then

$$\sum_{i=1}^n (X_{t_{i+1}} - X_{t_i})^2 = X_t^2 - X_0^2 - 2\sum_{i=1}^n X_{t_i}(X_{t_{i+1}} - X_{t_i}) \ .$$

Letting $|\Delta|$ going to zero, we get the claim. □

Theorem 3.21.5 (Ito formula) *Given a vector martingale $M = (M^{(1)}, M^{(2)}, \dots, M^{(d)})$ and a function $f \in C^2(\mathbf{R}^d, \mathbf{R})$. Then*

$$f(X_t) - f(X_0) = \int_0^t \nabla f(X) \, dM + \frac{1}{2}\sum_{ij}\int_0^t \delta_{x_i}\delta_{x_j} f_{x_i x_j}(X_s) \, d < M^{(i)}, M^{(j)} > \ .$$

Proof. It is enough to prove the formula for polynomials. By the integration by parts formula, we get the result for functions $f(x) = x_i g(x)$, if it is established for a function g. Since it is true for constant functions, we are done by induction. □

3.22 Stochastic differential equations

In the section about martingales, we have shown that if B is Brownian motion, then $X = f(B,t) = e^{\alpha B_t - \alpha^2 t/2}$ is a martingale and we have seen in the last section using Ito's formula and the fact that $\frac{1}{2}\Delta f + \dot{f} = 0$ that

$$\int_0^t \alpha X_s \, dM_s = X_t - 1 \ .$$

We can write this in differential form as

$$dX_t = \alpha X_t \, dM_t, X_0 = 1 \ .$$

This is an example of a stochastic differential equation and one would use the notation

$$\frac{dX}{dM} = \alpha X$$

if it would not lead to confusion with the corresponding ordinary differential equation, where B is not a stochastic process but a variable and where the solution would be $X = e^{\alpha B}$. Here, the solution is the stochastic process $X_t = e^{\alpha B_t - \alpha^2 t/2}$.

Definition. Let B_t be Brownian motion in $\mathbf{R}^d$. A **solution of a stochastic differential equation SDE**

$$dX_s = f(X_s, s) \ \cdot dB_s + g(X_s, s) \, ds \ ,$$

is a $\mathbf{R}^d$-valued process X satisfying

$$X_t = \int_0^t f(X_s, s) \ \cdot dB_s + \int_0^t g(X_s, s) \, ds \ ,$$

where $f : \mathbf{R}^d \times \mathbf{R}^+ \to \mathbf{R}^d \times \mathbf{R}^d$ and $g : \mathbf{R}^d \times \mathbf{R}^+ \to \mathbf{R}^d$.
Brownian motion can be replaced by a continuous martingale M. A solution must then satisfy

$$X_t = \int_0^t f(X_s, s) \ \cdot dM_s + \int_0^t g(X_s, s) \, ds \ .$$

Example.

$$X_t = e^{\alpha M_t - \alpha^2 < X,X >_t /2}$$

is a solution of $dM_t = \alpha M_t dX_t, M_0 = 1$. One can also allow the functions f and g to depend on the whole process X and not only $f(X,s) - f(X_s,s)$ or one can allow the solution X to be a $\mathbf{R}^r$-valued process, where r is different from d. The definition, we have given, is also called a **SDE of the simple kind**.

Historical remark. Stochastic differential equations were invented by Ito in 1951. Differential equations with a different integral came from Stratonovich but there are formulas relating them with each other, it is enough to consider the equations with the Ito integral.

For ODE's $\dot{x} = f(x,t)$, one knows that unique solutions exist locally if f is Lipschitz continuous in x and continuous in t. The proof given for 1-dimensional systems generalizes to differential equations in arbitary Banach spaces. The idea of the proof is a Picard iteration of an operator which is a contraction. In the appendix, we give a detailed proof of this for ODE's. For SDE's, one can do the same. We will do such an iteration on the Hilbert space $H^2_{[0,t]}$ of $\mathcal{L}^2$ martingales X having finite norm

$$||X||_T = E[\sup_{t \leq T} X_t^2] .$$

We will need the following version of Doob's inequality:

Lemma 3.22.1 *Let X be a $\mathcal{L}^p$ martingale with $p \geq 1$. Then*

$$E[\sup_{s \leq t} |X_s|^p] \leq (\frac{p}{p-1})^p \cdot E[|X_t|^p] .$$

Proof. We can assume without loss of generality that X is bounded. The general result follows by approximating X by $X \wedge k$ with $k \to \infty$.
Define $X^* = \sup_{s \leq t} |X_s|^p$. From Doob's inequality

$$P[X \geq \lambda] \leq E[|X_t| \cdot 1_{X^* \geq \lambda}]$$

we get

$$\begin{aligned}
E[|X^*|^p] &= E[\int_0^{X^*} p\lambda^{p-1} \, d\lambda] \\
&= E[\int_0^{\infty} p\lambda^{p-1} 1_{\{X^* \geq \lambda\}} \, d\lambda] \\
&= E[\int_0^{\infty} p\lambda^{p-1} P[X^* \geq \lambda] \, d\lambda] \\
&\leq E[\int_0^{\infty} p\lambda^{p-1} E[|X_t| \cdot 1_{X^* \geq \lambda}] \, d\lambda] \\
&= pE[|X_t| \int_0^{X^*} \lambda^{p-2} \, d\lambda \\
&= \frac{p}{p-1} E[|X_t| \cdot (X^*)^{p-1}] .
\end{aligned}$$

Hölder's inequality gives

$$E[|X^*|^p] \leq \frac{p}{p-1} E[(X^*)^p]^{(p-1)/p} E[|X_t|^p]^{1/p}$$

and the claim follows. □

Theorem 3.22.2 (Local existence and uniqueness of solutions) *Let M be a continuous martingale. Assume f and g are continuous in t and Lipschitz continuous in x. Then there exists a unique solution X_t of the SDE*

$$dX = f(x,t)\; dM + g(x,t)\; ds$$

on some interval $[0,T)$ with given initial condition $X_0 = X_0$.

Proof. We define the operator

$$\mathcal{S}(X) = \int_0^t f(s, X_s)\; dM_s + \int_0^t g(s, X_s)\; ds$$

on $\mathcal{L}^2$-processes. Write $\mathcal{S}(X) = \mathcal{S}_1(X) + \mathcal{S}_2(X)$. We will show that on some time interval $(0,T]$, the map $\mathcal{S}$ is a contraction and that $\mathcal{S}^n(X)$ converges in the metric $|||X-Y|||_T = E[\sup_{s \leq T}(X_s - Y_s)^2]$, if T is small enough to a unique fixed point. It is enough that for $i = 1,2$

$$|||\mathcal{S}_i(X) - \mathcal{S}_i(Y)|||_T \leq (1/4) \cdot ||X - Y||_T$$

then $\mathcal{S}$ is a contraction

$$|||\mathcal{S}(X) - \mathcal{S}(Y)|||_T \leq (1/2) \cdot ||X - Y||_T \; .$$

By assumption, there exists a constant K, such that

$$|f(t,w) - f(t,w')| \leq K \cdot \sup_{s \leq 1} |w - w'| \; .$$

(i) $|||\mathcal{S}_1(X) - \mathcal{S}_1(Y)|||_T = ||| \int_0^t f(s,X_s) - f(s,Y_s)\; dM_s|||_T \leq (1/4) \cdot |||X-Y|||_T$ for T small enough.
Proof. By the above lemma for $p = 2$, we have

$$\begin{aligned}
|||\mathcal{S}_1(X) - \mathcal{S}_1(Y)|||_T &= E[\sup_{t \leq T}(\int_0^t f(s,X) - f(s,Y)\; dM_s)^2] \\
&\leq 4E[(\int_0^T f(t,X) - f(t,Y)\; dM_t)^2] \\
&= 4E[(\int_0^T f(t,X) - f(t,Y))^2\; d < M,M >_t] \\
&\leq 4K^2 E[\int_0^T \sup_{s \leq t} |X_s - Y_s|^2\; dt] \\
&= 4K^2 \int_0^T |||X-Y|||_s\; ds \\
&\leq (1/4) \cdot |||X - Y|||_T \; ,
\end{aligned}$$

where the last inequality holds for T small enough.

(ii) $|||\mathcal{S}_2(X) - \mathcal{S}_2(Y)|||_T = ||| \int_0^t g(s, X_s) - g(s, Y_s) \, ds|||_T \leq (1/4) \cdot |||X - Y|||_T$ for T small enough. This is proven for differential equations in Banach spaces (see the Appendix).

The two estimates (i) and (ii) prove the claim in the same way as in the classical Cauchy-Picard existence theorem. □

Appendix: Cauchy-Picard existence theorem for ODE's.

Let $\mathcal{X}$ be a Banach space and I an interval in $\mathbf{R}$. The following lemma is useful for proving existence of fixed points of maps.

Lemma 3.22.3 *Let $X = \overline{B_r(x_0)} \subset \mathcal{X}$ and assume ϕ is a differentiable map $\mathcal{X} \to \mathcal{X}$. If for all $x \in X$, $||D\phi(x)|| \leq |\lambda| < 1$ and*

$$||\phi(x_0) - x_0|| \leq (1 - \lambda) \cdot r$$

then ϕ has exactly one fixed point in X.

Proof. The condition $||x - x_0|| < r$ implies that

$$||\phi(x) - x_0|| \leq ||\phi(x) - \phi(x_0)|| + ||\phi(x_0) - x_0|| \leq \lambda r + (1 - \lambda) r = r \; .$$

The map ϕ maps therefore the ball X into itself. Banach's fixed point theorem applied to the complete metric space X and the contraction ϕ implies the result. □

Let f be a mapping $I \times \mathcal{X} \to \mathcal{X}$. A differentiable mapping $u : \mathcal{J} \to \mathcal{X}$ of an open ball $J \subset I$ in $\mathcal{X}$ is called a *solution of the differential equation*

$$\dot{x} = f(t, x)$$

if we have for all $t \in J$ the relation

$$\dot{u}(t) = f(t, u(t)) \; .$$

Theorem 3.22.4 (Cauchy-Picard Existence theorem)

Let $f : I \times \mathcal{X} \to \mathcal{X}$ be continuous in the first coordinate and locally Lipschitz continuous in the second. Then the following holds: for every $(t_0, x_0) \in I \times \mathcal{X}$, there exists an open interval $J \subset I$ with midpoint t_0, such that on J, there exists exactly one solution of the differential equation $\dot{x} = f(t, x)$.

Proof. There exists an interval $J(t_0, a) = (t_0 - a, t_0 + a) \subset I$ and a ball $B(x_0, b)$, such that

$$M = \sup\{||f(t, x)|| \mid (t, x) \in J(t_0, a) \times B(x_0, b)\}$$

as well as

$$k = \sup\{\frac{||f(t,x_1) - f(t,x_2)||}{||x_1 - x_2||} \mid (t,x_1),(t,x_2) \in J(t_0,a) \times B(x_0,b), x_1 \neq x_2\}$$

are finite. Define for $r < a$ the Banach space

$$\mathcal{X}_r = C(\overline{J}(t_0,r),\mathcal{X}) = \{y : \overline{J}(t_0,r) \to \mathcal{X},\ y \text{ continuous}\}$$

with norm

$$||y|| = \sup_{t \in \overline{J}(t_0,r)} ||y(t)||$$

Let $V_{r,b}$ be the open ball in $\mathcal{X}_r$ with radius b around the constant mapping $t \mapsto x_0$. For every $y \in V_{r,b}$ we define

$$\phi(y) : t \mapsto x_0 + \int_{t_0}^{t} f(s,y(s))ds$$

which is again an element in $\mathcal{X}_r$. We prove now, that for r small enough, ϕ is a contraction. A fixed point of ϕ is then a solution of the differential equation $\dot{x} = f(t,x)$, which exists on $J = J_r(t_0)$. For two points $y_1, y_2 \in V_r$, we have by assumption

$$||f(s,y_1(s)) - f(s,y_2(s))|| \leq k \cdot ||y_1(s) - y_2(s)|| \leq k \cdot ||y_1 - y_2||$$

for every $s \in \overline{J}_r$. Thus, we have

$$\begin{aligned} ||\phi(y_1) - \phi(y_2)|| &= ||\int_{t_0}^{t} f(s,y_1(s)) - f(s,y_2(s))\ ds|| \\ &\leq \int_{t_0}^{t} ||f(s,y_1(s)) - f(s,y_2(s))||\ ds \\ &\leq kr \cdot ||y_1 - y_2||\ . \end{aligned}$$

On the other hand, we have for every $s \in \overline{J}_r$

$$||f(s,y(s))|| \leq M$$

and so

$$||\phi(x_0) - x_0|| = ||\int_{t_0}^{t} f(s,x_0(s))\ ds|| \leq \int_{t_0}^{t} ||f(s,x_0(s))||\ ds \leq M \cdot r\ .$$

We can apply the above lemma, if $kr < 1$ and $Mr < b(1-kr)$. This is the case, if $r < b/(M+kb)$. By choosing r small enough, we can get the contraction rate as small as we wish. □

Chapter 4

Selected topics

4.1 Percolation

Notation. Denote with e_i the standard basis in $\mathbf{Z}^d$. Denote with $\mathbf{L}^d$ the Cayley graph of $\mathbf{Z}^d$ with the generators $A = \{e_1, \dots, e_d\}$. This graph $\mathbf{L}^d = (V, E)$ has $\mathbf{Z}^d$ as vertices and as edges straight line segment connecting points with $|x - y| = \sum_{i=1}^d |x_i - y_i| = 1$.

Definition. Declare each edge of $\mathbf{L}^d$ to be **open** with probability $p \in [0, 1]$ and **closed** otherwise, independenty of all other edges. This defines a probability measure $P = P_p$ on $\Omega = \prod_{e \in E}\{0, 1\}$ of **configurations**. We denote expectation with respect to P with $E_p[\cdot]$.

Definition. A **path** in $\mathbf{L}^d$ is a sequence of vertices $(x_0, x_1, \dots, x_n)$ such that $(x_i, x_{i+1}) = e_i$ are edges of $\mathbf{L}^d$. Such a path has **length** n and **connects** x_0 with x_n. A path is called **open** if all its edges are open and **closed** if all its edges are closed. Two subgraphs of $\mathbf{L}^d$ are **disjoint** if they have neither edges nor vertices in common.

Definition. Consider the random subgraph of $\mathbf{L}^d$ containing the vertex set $\mathbf{Z}^d$ and only open edges. The conneced components of this graph are called **open clusters**. If it is finite, it is also called a **lattice animal**. Call $C(x)$ the open cluster containing the vertex x. By translation invariance, the distribution of $C(x)$ is independent of x and we can take $x = 0$ for which we write $C(0) = C$.

Definition. Define the **percolation probability** $\theta(p)$ being the probability that a given vertex belongs to an infinite open cluster.

$$\theta(p) = P[|C| = \infty] = 1 - \sum_{n=1}^{\infty} P[|C| = n] .$$

It is one of the problems of **bond percolation** to learn about the function $\theta(p)$.

Lemma 4.1.1 *There exists a critical value $p_c = p_c(d)$ such that $\theta(p) = 0$ for $p < p_c$ and $\theta(p) > 0$ for $p > p_c$. The value $d \mapsto p_c(d)$ is non-increasing: $p_c(d+1) \leq p_c(d)$.*

Proof. Clearly $p \mapsto \theta(p)$ is non-decreasing and $\theta(0) = 0, \theta(1) = 1$. We can define therefore

$$p_c = \inf\{p \in [0,1] \mid \theta(p) > 0\ \}.$$

We can embed the graph $\mathbf{Z}^d$ into the graph $\mathbf{Z}^{d'}$ for $d < d'$ by realising $\mathbf{Z}^d$ as a linear subspace of $\mathbf{Z}^{d'}$. Any configuration in $\mathbf{L}^{d'}$ projects into a configuration in $\mathbf{L}^d$. If the origin is in an infinite cluster of $\mathbf{Z}^d$, then it is also in an infinite cluster of $\mathbf{Z}^{d'}$. □

Remark. The one dimensional case $d = 1$ is not interesting since $p_c = 1$ there. The existence of interesting phenomena is therefore only possible for $d > 1$.

Definition. A **selfavoiding random walk** in $\mathbf{L}^d$ is the process S_T obtained by stopping the ordinary random walk S_n with stopping time

$$T(\omega) = \inf\{n \in \mathbf{N} \mid \omega(n) = \omega(m), m < n\}\ .$$

Let $\sigma(n)$ be the number of selfavoiding paths in $\mathbf{L}^d$ which have length n. The **connective constant** of $\mathbf{L}^d$ is defined as

$$\lambda(d) = \lim_{n\to\infty} \sigma(n)^{1/n}\ .$$

Remark. The exact value of $\lambda(d)$ is not known but one has the elementary estimate $d < \lambda(d) < 2d - 1$ since a selfavoiding walk can not reverse direction and so $\sigma(n) \leq 2d(2d-1)^{n-1}$ and a walk going only forward in each direction is selfavoiding. For example, it is known that $\lambda(2) \in [2.62002, 2.69576]$ and numerical estimates makes one believe that the real value is 2.6381585. The number c_n of selfavoiding walks in $\mathbf{L}^2$ is for small values $c_1 = 4, c_2 = 12, c_3 = 36, c_4 = 100, c5 = 284, c6 = 780, c7 = 2172, \dots$.

Theorem 4.1.2 (Broadbent-Hammersley) *If $d > 1$, then*

$$0 < \lambda(d)^{-1} \leq p_c(d) \leq p_c(2) < 1\ .$$

Proof. (i) $p_c(d) \geq \lambda(d)^{-1}$.
Let $N(n) \leq \sigma(n)$ be the number of open selfavoiding paths of length n in $\mathbf{L}^n$.

Since any such path is open with probability p^n, we have

$$E_p[N(n)] = p^n \sigma(n) .$$

If the origin is in an infinite open cluster, there must exist open paths of all lengths beginning at the origin so that

$$\theta(p) \leq P_p[N(n) \geq 1] \leq E_p[N(n)] = p^n \sigma(n) = (p\lambda(d) + o(1))^n$$

which goes to zero for $p < \lambda(p)^{-1}$. This shows that $p_c(d) \geq \lambda(d)^{-1}$.

(ii) $p_c(2) < 1$.
Denote with $\mathbf{L}^2_*$ the dual graph of $\mathbf{L}^2$ which has as vertices the faces of $\mathbf{L}^2$ and as vertices pairs of faces which are adjacent. We can realize the vertices as $\mathbf{Z}^2 + (1/2, 1/2)$. Since there is a bijective relation between the edges of $\mathbf{L}^2$ and $\mathbf{L}^2_*$ and we declare an edge of $\mathbf{L}^2_*$ to be open (rsp. closed) if it crosses an open (rsp. closed) edge in $\mathbf{L}^2$. This defines a bond percolation on $\mathbf{L}^2_*$.

Claim. The origin is in the interior of a closed circuit of the dual lattice if and only if the open cluster at the origin is finite.
This follows essentially from Jordan's curve theorem saying that a closed path in the plane divides the plane into two disjoint subsets.

Let $\rho(n)$ denote the number of closed circuits in the dual which have length n and which contain in their interiors the origin of $\mathbf{L}^2$. Each such circuit contains a self-avoiding walk of lenth $n-1$ starting from a vertex of the form $(k + 1/2, 1/2)$, where $0 \leq k < n$. Since the number of such paths γ is at most $n\sigma(n-1)$, we have

$$\rho(n) \leq n\sigma(n-1)$$

and with $q = 1 - p$

$$\sum_\gamma P[\gamma \text{ is closed}] \leq \sum_{n=1}^\infty q^n n\sigma(n-1) = \sum_{n=1}^\infty qn(q\lambda(2) + o(1))^{n-1}$$

which is finite if $q\lambda(2) < 1$. Furthermore, this sum goes to zero if $q \to 0$ so that we can find $0 < \delta < 1$ such that for $p > \delta$

$$\sum_\gamma P[\gamma \text{ is closed}] \leq 1/2.$$

We have therefore

$$P[|C| = \infty] = P[\text{no } \gamma \text{ is closed}] \geq 1 - \sum_\gamma P[\gamma \text{ is closed}] \geq 1/2$$

so that $p_c(2) \leq \delta < 1$. □

Remark.
We will see below that even $p_c(2) < 1 - \lambda(2)^{-1}$. It is however known that $p_c(2) = 1/2$.

Definition. The parameter set $p < p_c$ is called the **subcritical phase**, the set $p > p_c$ is the **supercritical phase**.

Other questions. For $p < p_c$, one is interested in the **mean size** of the open cluster

$$\chi(p) = E_p[|C|] .$$

For $p > p_c$, one would like to know the **mean size** of the finite clusters

$$\chi^f(p) = E_p[|C| \mid |C| < \infty] .$$

It is known that $\chi(p) < \infty$ for $p < p_c$ but only conjectured that $\chi^f(p) < \infty$ for $p > p_c$.
An interesting question is whether there exists an open cluster at the critical point $p = p_c$. The answer is known to be no in the case $d = 2$ and generally believed to be no for $d \geq 3$. For p near p_c it is believed that the percolation probability $\theta(p)$ and the mean size $\chi(p)$ behave as powers of $|p - p_c|$. It is conjectured that the following critical exponents

$$\begin{aligned}
\gamma &= -\lim_{p \nearrow p_c} \frac{\log \chi(p)}{\log |p - p_c|} \\
\beta &= \lim_{p \searrow p_c} \frac{\log \theta(p)}{\log |p - p_c|} \\
\delta^{-1} &= -\lim_{n \to \infty} \frac{\log P_{p_c}[|C| \geq n}{\log n}] .
\end{aligned}$$

exist.

4.2 The FKG correlation inequality

We recall that percolation deals with a **family** of probability spaces $(\Omega, \mathcal{A}, P_p)$, where $\Omega = \{0,1\}^E$ is the set of configurations with product σ algebra $\mathcal{A}$ and product measure $P_p = (p, 1-p)^E$.

Definition. There exists a natural partial ordering in Ω coming from the ordering on $\{0,1\}$: we say $\omega \leq \omega'$, if $\omega(e) \leq \omega'(e)$ for all $e \in E$. We call a random variable X on $(\Omega, \mathcal{A}, P)$ **increasing** if $\omega \leq \omega'$ implies $X(\omega) \leq X(\omega')$. It is called **decreasing** if $-X$ is increasing. As usual, this notion can also be defined for measurable sets $A \in \mathcal{A}$: a set A is **increasing** if 1_A is increasing.

Lemma 4.2.1 *If X is a increasing random variable in $\mathcal{L}^1(\Omega, P_q) \cap \mathcal{L}^1(\Omega, P_p)$, then*

$$E_p[X] \leq E_q[X]$$

if $p \leq q$.

Proof. If X depends only on a single bond e, we can write $E_p[X] = pX(1) + (1-p)X(0)$. Since X is increasing, we have $\frac{d}{dp} E_p[X] = X(1) - X(0) \geq 0$ which gives $E_p[X] \leq E_q[X]$ for $p \leq q$. If X depends only on finitely many bonds, we can write it as a sum $X = \sum_{i=1}^d X_i$ of variables X_i which depend only on one bond and get again

$$\frac{d}{dp} E_p[X] = \sum_{i=1}^n (X_i(1) - X_i(0)) \geq 0 \; .$$

In general we approximate every random variable in $\mathcal{L}^1(\Omega, P_p) \cap \mathcal{L}^1(\Omega, P_q)$ by step functions which depend only on finitely many coordinates X_i. Since $E_p[X_i] \to E_p[X]$ and $E_q[X_i] \to E_q[X]$, the claim follows. □

The following **correlation inequality** is named after Fortuin, Kasterleyn and Ginibre (1971).

Theorem 4.2.2 (FKG inequality) *For increasing random variables $X, Y \in \mathcal{L}^2(\Omega, P_p)$, we have*

$$E_p[XY] \geq E_p[X] \cdot E_p[Y] \; .$$

Proof. As in the proof of the above lemma, we prove the claim first for random variables X which depend only on n edges $e_1, e_2, \dots, e_n$. We proceed by induction.

(i) X and Y depend only on one edge e.
We have

$$(X(\omega) - X(\omega')(Y(\omega) - Y(\omega')) \geq 0$$

since the left hand side is 0 if $\omega(e) = \omega'(e)$ and if $1 = \omega(e) = \omega'(e) = 0$, both factors are nonnegative since X, Y are increasing, if $0 = \omega(e) = \omega'(e) = 1$ both factors are nonpositive since X, Y are increasing. Therefore

$$\begin{aligned} 0 &\leq \sum_{\sigma,\sigma' \in \{0,1\}} (X(\omega) - X(\omega'))(Y(\omega) - Y(\omega'))P_p[\omega(e) = \sigma]P_p[\omega(e) = \sigma'] \\ &= 2(E_p[XY] - E_p[X]E_p[Y]) \ . \end{aligned}$$

(ii) Assume the claim is known for all functions which depend on k edges with $k < n$. We claim that it holds also for X, Y depending on n edges $e_1, e_2, \dots, e_n$. Let $\mathcal{A}_k = \mathcal{A}(e_1, \dots e_k)$ be the σ-algebra generated by functions depending only on the edges e_k. The random variables

$$X_k = E_p[X|\mathcal{A}_k], Y_k = E_p[Y|\mathcal{A}_k]$$

depend only on the $e_1, \dots, e_k$ and are increasing. By induction,

$$E_p[X_{n-1}Y_{n-1}] \geq E_p[X_{n-1}]E_p[Y_{n-1}] \ .$$

By the tower property of conditional expectation, the right hand side is $E_p[X]E_p[Y]$. For fixed $e_1, \dots, e_{n-1}$, we have $(XY)_{n-1} \geq X_{n-1}Y_{n-1}$ and so

$$E_p[XY] = E_p[(XY)_{n-1}] \geq E_p[X_{n-1}Y_{n-1}] \ .$$

(iii) Let X, Y be arbitrary and define $X_n = E_p[X|\mathcal{A}_n]$, $Y_n = E_p[Y|\mathcal{A}_n]$. We know from (ii) that $E_p[X_nY_n] \geq E_p[X_n]E_p[Y_n]$. Since $X_n = E[X|\mathcal{A}_n]$ and $Y_n = E[X|\mathcal{A}_n]$ are martingales which are bounded in $\mathcal{L}^2(\Omega, P_p)$, Doob's convergence theorem (see the section "Doob's decomposition of a stochastic process" in the last term) implies that $X_n \to X$ and $Y_n \to Y$ in $\mathcal{L}^2$ and therefore $E[X_n] \to E[X]$ and $E[Y_n] \to E[Y]$. By the Schwarz inequality, we get also in $\mathcal{L}^1$ rep. $\mathcal{L}^2$ norm in $(\Omega, \mathcal{A}, P_p)$

$$\begin{aligned} ||X_nY_n - XY||_1 &\leq ||(X_n - X)Y_n||_1 + ||X(Y_n - Y)||_1 \\ &\leq ||X_n - X||_2||Y_n||_2 + ||X||_2||Y_n - Y||_2 \\ &\leq C(||X_n - X||_2 + ||Y_n - Y||_2) \to 0 \end{aligned}$$

where $C = \max(||X||_2, ||Y||_2)$ is a constant. This means $E_p[X_nY_n] \to E_p[XY]$.
□

Remark. It follows immediatly that if A, B are increasing events in Ω, then $P_p[A \cap B] \geq P_p[A] \cdot P_p[B]$.

Example. Let Γ_i be families of paths in $\mathbf{L}_*^d$ and let A_i be the event that some path in Γ_i is open. Then A_i are increasing events and so after applying the inequality k times, we get

$$P_p[\bigcap_{i=1}^k A_i] \geq \prod_{i=1}^k P_p[A_i] \; .$$

We show now, how this inequality can be used to give an explicit bound for the critical percolation probability p_c in $\mathbf{L}^2$. The following corollary belongs still to the theorem of Broadbent-Hammersley.

Corollary 4.2.3

$$p_c(2) \leq (1 - \lambda(2)^{-1}) \; .$$

Proof. Given $N \in \mathbf{N}$, define the events

$$\begin{aligned} F_N &= \{\exists \text{ no closed path of length} \leq N \text{ in } \mathbf{L}_*^{\mathrm{d}}\} \\ G_N &= \{\exists \text{ no closed path of length} > N \text{ in } \mathbf{L}_*^{\mathrm{d}}\} \; . \end{aligned}$$

We know that $F_N \cap G_N \subset \{|C| = \infty\}$. Since F_N and G_N are both increasing, the correlation inequality says $P_p[F_N \cap G_N] \geq P_p[F_N] \cdot P_p[G_N]$. We deduce

$$\theta(p) = P_p[|C| = \infty] = P_p[F_N \cap G_N] \geq P_p[F_N] \cdot P_p[G_N] \; .$$

If $(1-p)\lambda(2) < 1$, then we know (see the proof in the last section) that

$$P_p[G_N^c] \leq \sum_{n=N}^{\infty} (1-p)^n n \sigma(n-1)$$

which goes to zero for $N \to \infty$. For N large enough, we have therefore $P_p[G_N] \geq 1/2$. Since also $P_p[F_N] > 0$, it follows that $\theta_p > 0$, if $(1-p)\lambda(2) < 1$ or $p < (1 - \lambda(2)^{-1})$ which proves the claim. $\square$

4.3 Russo's formula

Definition. Given $A \in \mathcal{A}$ and $\omega \in \Omega$. We say, an edge $e \in E$ is **pivotal** for the pair (A, ω) if $1_A(\omega) \neq 1_A(\omega_e)$, where ω_e is the unique configuration which agrees with ω except at the edge e.

Theorem 4.3.1 (Russo's formula) *Let A be an increasing event depending only on finitely many edges of $\mathbf{L}^d$. Then*

$$\frac{d}{dp} P_p[A] = E_p[N(A)] ,$$

where $N(A)$ is the number of edges which are pivotal for A.

Proof. (i) A new probability space.
The family of probability spaces $(\Omega, \mathcal{A}, P_p)$, can be embedded in one probability space $([0,1]^E, \mathcal{B}([0,1]^E), P)$, where P is the product measure dx^E. Given a configuration $\eta \in [0,1]^E$ and $p \in [0,1]$, we get a configuration in Ω by defining $\eta_p(e) = 1$ if $\eta(e) < p$ and $\eta_p = 0$ else. More generally, given $\mathbf{p} \in [0,1]^E$, we get configurations $\eta_{\mathbf{p}}(e) = 1$ if $\eta(e) < \mathbf{p}(e)$ and $\eta_{\mathbf{p}} = 0$ else. Like this, we can define configurations with a large class of probability measures $P_{\mathbf{p}} = \prod_{e \in E}(p(e), 1 - p(e))$ with **one** probabiliy space and we have

$$P_{\mathbf{p}}[A] = P[\eta_{\mathbf{p}} \in A] .$$

(ii) Derivative with respect to one $p(f)$.
Assume $\mathbf{p}$ and $\mathbf{p}'$ differ only at an edge f such that $p(f) \leq p'(f)$. Then $\{\eta_{\mathbf{p}} \in A\} \subset \{\eta_{\mathbf{p}'} \in A\}$ so that

$$\begin{aligned} P_{\mathbf{p}'}[A] - P_{\mathbf{p}}[A] &= P[\eta_{\mathbf{p}'} \in A] - P[\eta_{\mathbf{p}} \in A] \\ &= P[\eta_{\mathbf{p}'} \in A; \eta_{\mathbf{p}} \notin A] \\ &= (p'(f) - p(f)) P_p[f \text{ pivotal for } A] . \end{aligned}$$

Divide both sides by $(p'(f) - p(f))$ and let $p'(f) \to p(f)$. This gives

$$\frac{\partial}{\partial p(f)} P_{\mathbf{p}}[A] = P_{\mathbf{p}}[f \text{ pivotal for } A] .$$

(iii) The claim, if A depends on finitely many edges.
If A depends on finitely many edges, then $P_{\mathbf{p}}[A]$ is a function of a finite set $\{p(f_i)\}_{i=1}^m$ of edge probabilities. The chain rule gives then

$$\begin{aligned} \frac{d}{dp} P_p[A] &= \sum_{i=1}^m \frac{\partial}{\partial p(f_i)} P_{\mathbf{p}}[A]|_{\mathbf{p}=(p,p,p,\ldots,p)} \\ &= \sum_{i=1}^m P_{\mathbf{p}}[f_i \text{ pivotal for } A] \\ &= E_p[N(A)] . \end{aligned}$$

(iv) The general claim.
In general, define for every finite set $F \subset E$

$$\mathbf{p}_F(e) = p + 1_{\{e \in F\}}\delta$$

where $0 \leq p \leq p + \delta \leq 1$. Since A is increasing, we have

$$P_{p+\delta}[A] \geq P_{\mathbf{p}_F}[A]$$

and therefore

$$\frac{1}{\delta}(P_{p+\delta}[A] - P_p[A]) \geq \frac{1}{\delta}(P_{\mathbf{p}_F}[A] - P_p[A]) \to \sum_{e \in F} P_p[e \text{ pivotal for } A]$$

as $\delta \to 0$. The claim is obtained by making F larger and larger filling out E.
□

Example. Let $F = \{e_1, e_2, \dots, e_m\} \subset E$ be a finite set in of edges.

$$A = \{\text{the number of open edges in } F \text{ is} \geq k\} .$$

An edge $e \in F$ is pivotal for A if and only if $A \setminus \{e\}$ has exactly $k-1$ open edges. We have

$$P_p[e \text{ is pivotal}] = \begin{pmatrix} m-1 \\ k-1 \end{pmatrix} p^{k-1}(1-p)^{m-k}$$

so that by Russo's formula

$$\frac{d}{dp}P_p[A] = \sum_{e \in F} P_p[e \text{ is pivotal}] = m \begin{pmatrix} m-1 \\ k-1 \end{pmatrix} p^{k-1}(1-p)^{m-k} .$$

Since we know $P_0[A] = 0$, we obtain by integration

$$P_p[A] = \sum_{l=k}^{m} \begin{pmatrix} m \\ l \end{pmatrix} p^l (1-p)^{m-1} .$$

Remarks. If A is no more depending on finitely many edges, then $P_p[A]$ need no more be differentiable for all values of p.

4.4 The mean size of the open cluster

We are interested in this section in the **mean size of the open cluster** $\chi(p) = E_p[|C|]$ and the aim is to prove

Theorem 4.4.1 (Uniqueness) *For $p < p_c$, we have $\chi(p) < \infty$.*

The proof of this theorem is quite involved and we will not give the whole argument.

Let $S(n,x) = \{y \in \mathbf{Z}^d \mid |x-y| = \sum_{i=1}^d |x_i| \leq n\}$ be the ball of radius n around x in $\mathbf{Z}^d$ and let A_n be the event that there exists an open path joining the origin with some vertex in $\delta S(n,0)$. The above theorem folllows from the following fact

Theorem 4.4.2 (Exponential decay of radius of open cluster) *If $p < p_c$, there exists ψ_p such that $P_p[A_n] < e^{-n\psi_p}$.*

Proof of Theorem **??**:

Proof. Clearly, $|S(n,0)| \leq C_d \cdot (n+1)^d$ with some constant C_d. Let $M = \max\{n \mid A_n \text{ occurs}\}$. By definition of p_c, if $p < p_c$, then $P_p[M < \infty] = 1$. We get

$$\begin{aligned} E_p[|C|] &\leq \sum_n E_p[|C| \mid M = n] \cdot P_p[M = n] \\ &\leq \sum_n |S(n)| P_p[A_n] \\ &\leq \sum_n C_d (n+1)^d e^{-n\psi_p} < \infty \; . \end{aligned}$$

□

Proof. We are concerned with the probabilities $g_p(n) = P_p[A_n]$. Sine A_n are increasing events, Russo's formula gives

$$g_p'(n) = E_p[N(A_n)] \; ,$$

where $N(A_n)$ is the number of pivotal edges in A_n. We have

$$g_p'(n) = \sum_e P_p[\text{e pivotal for } A]$$

$$
\begin{aligned}
&= \sum_e \frac{1}{p} P_p[e \text{ open and pivotal for } A] \\
&= \sum_e \frac{1}{p} P_p[A \cap \{e \text{ pivotal for } A\}] \\
&= \sum_e \frac{1}{p} P_p[A \cap \{e \text{ pivotal for } A\} | A] \cdot P_p[A] \\
&= \sum_e \frac{1}{p} E_p[N(A) \mid A] \cdot P_p[A] \\
&= \sum_e \frac{1}{p} E_p[N(A) \mid A] \cdot g_p(n)
\end{aligned}
$$

so that

$$
\frac{g_p'(n)}{g_p(n)} = \frac{1}{p} E_p[N(A_n) \mid A_n] \; .
$$

By integrating up from α to β, we get

$$
\begin{aligned}
g_\alpha(n) &= g_\beta(n) \exp(-\int_\alpha^\beta \frac{1}{p} E_p[N(A_n) \mid A_n] \, dp) \\
&\leq g_\beta(n) \exp(-\int_\alpha^\beta E_p[N(A_n) \mid A_n] \, dp) \\
&\leq \exp(-\int_\alpha^\beta E_p[N(A_n) \mid A_n] \, dp) \; .
\end{aligned}
$$

One needs to show then that $E_p[N(A_n) \ |A_n]$ grows roughly linearly when $p < p_c$. This is quite technical and we skip it. $\square$

4.5 The average number of open clusters

Definition. The **number of open clusters per vertex** is defined as

$$\kappa(p) = E_p[|C|^{-1}] = \sum_{n=1}^{\infty} \frac{1}{n} P_p[|C| = n] \; .$$

Let B_n the box with side length $2n$ and centre at the origin and let K_n be the number of open clusters in B_n. The following theorem explains the name of κ.

Theorem 4.5.1 *In* $\mathcal{L}^1(\Omega, \mathcal{A}, P_p)$ *we have*

$$K_n/|B_n| \to \kappa(p) \; .$$

Proof. Let $C_n(x)$ be the connected component of the open cluster in B_n which contains $x \in \mathbf{Z}^d$. Define $\Gamma(x) = |C(x)|^{-1}$.

(i) $\sum_{x \in B_n} \Gamma_n(x) = K_n$.
Proof. If Σ is an open cluster of B_n, then each vertex $x \in \Sigma$ contributes $|\Sigma|^{-1}$ to the left hand side. Thus, each open cluster contributes 1 to the left hand side.

(ii) $\frac{K_n}{|B_n|} \geq \frac{1}{|B_n|} \sum_{x \in B_n} \Gamma(x)$ where $\Gamma(x) = |C(x)|^{-1}$.
Proof. Follows from (i) and the trivial fact $\Gamma(x) \leq \Gamma_n(x)$.

(iii) $\frac{1}{|B_n|} \sum_{x \in B_n} \Gamma(x) \to E_p[\Gamma(0)] = \kappa(p)$.
Proof. $\Gamma(x)$ are bounded random variables which have a distribution which is invariant under the ergodic group of translations in $\mathbf{Z}^d$. The claim follows from the ergodic theorem.

(iv) $\liminf_{n \to \infty} \frac{K_n}{|B_n|} \geq \kappa(p)$ almost everywhere.
Proof. Follows from (ii) and (iii).

(v) $\sum_{x \in B(n)} \Gamma_n(x) \leq \sum_{x \in B(n)} \Gamma(x) + \sum_{x \sim \delta B_n} \Gamma_n(x)$, where $x \sim Y$ means that x is in the same cluster than one of the elements $y \in Y \subset \mathbf{Z}^d$.

(vi) $\frac{1}{|B_n|} \sum_{x \in B_n} \Gamma_n(x) \leq \frac{1}{|B_n|} \sum_{x \in B_n} \Gamma(x) + \frac{|\delta B_n|}{|B_n|}$. □

Without proof we mention

Theorem 4.5.2 *The function $\kappa(p)$ is continuously differentiable on $[0,1]$.*

Remark. It is even known that κ and the mean size of the open cluster $\chi(p)$ are real analytic functions on the interval $[0, p_c)$. There would be much more to say in percolation theory. We mention:
The uniqueness of the infinite open cluster:
For $p > p_c$ and if $\theta(p_c) > 0$ also for $p = p_c$, there exists a unique infinite open cluster.
Regularity of some functions $\theta(p)$
For $p > p_c$, the functions $\theta(p), \chi^f(p), \kappa(p)$ are differentiable. In general $\theta(p)$ is continuous from the right.
The critical probability in two dimensions is $1/2$.

4.6 Localisation of random Jacobi matrices

Definition. A **Jacobi matrix** with IID potential $V_\omega(n)$ is a bounded selfadjoint operator on $l^2(\mathbf{Z}^d) = \{n \mapsto x_n \mid \sum_{i=1}^\infty x_i^2 = 1\}$ of the form

$$L_\omega u(n) = \sum_{|m-n|=1} u(m) + V_\omega(n)u(n) = (\Delta + V_\omega)(u)(n) ,$$

where $V(n)$ are IID random variables. These **discrete Schrödinger operators** are used to model disordered systems.

Definition. A bounded linear operator L has **pure point spectrum**, if there exists a countable set of eigenvalues λ_i with eigenfunctions ϕ_i such that $L\phi_i = \lambda_i\phi_i$ and ϕ_i span the Hilbert space $l^2(\mathbf{Z}^d)$.

Theorem 4.6.1 (Fröhlich-Spencer) *If $V(n)$ are IID random variables with uniform distribution on $[0,1]$, then $L_\omega = \Delta + \lambda \cdot V_\omega$ has pure point spectrum for almost all ω if λ is large enough.*

Definition. Given $E \in \mathbf{C} \setminus \mathbf{R}$, define $G_\omega(m,n,E) = [(L-E)^{-1}]_{mn}$, the **Green function** of L. Let $\mu = \mu_\omega$ be the spectral measure of the vector e_0, (this measure is defined by the map $C(\mathbf{R}) \to \mathbf{R}, f \mapsto f(L_\omega)_{00}$ by $f(L_\omega)_{00} = \int f(L_\omega)_{00}\, d\mu$), define the functions

$$F(x) = \int \frac{d\mu(y)}{y-x}$$

$$G(z) = \int \frac{d\mu(\lambda)}{(y-x)^2} .$$

The function F is called the **Borel transform** of the measure μ.

Definition. Given any Jacobi matrix L, let L_α be the operator $L + \alpha P_0$, where P_0 is the projection onto the one dimensional space spanned by δ_0. One calls L_α a **rank-one perturbation** of L.

Lemma 4.6.2 (Integral formula of Javrjan-Kotani)

$$\int_{\mathbf{R}} d\mu_\alpha \, d\alpha = dE .$$

Proof. The second resolvent formula gives

$$(L_\alpha - z)^{-1} - (L-z)^{-1} = -\alpha(L_\alpha - z)^{-1} P_0 (L-z)^{-1} .$$

Looking at the 00-matrix elements, we obtain

$$F_\alpha(z) - F(z) = -\alpha F_\alpha(z) F(z)$$

which gives, when solved for F_α, the **Aronzajn-Krein formula**

$$F_\alpha(z) = \frac{F(z)}{1 + \alpha F(z)} .$$

We have to show that for any continuous function $f \in \mathbf{C} \to \mathbf{C}$

$$\int_{\mathbf{R}} \int_{\mathbf{R}} f(x) d\mu_\alpha(x) \; d\alpha = \int f(x) dE(x)$$

and it is enough to verify this for the dense set of functions

$$\{ f_z(x) = (x-z)^{-1} - (x+i)^{-1} \; | z \in \mathbf{C} \setminus \mathbf{R} \} .$$

Contour integration in the upper half plane gives $\int_{\mathbf{R}} f_z(x) \; dx = 0$ for $\mathrm{Im}(z) < 0$ and $2\pi i$ for $\mathrm{Im}(z) > 0$. On the other hand

$$\int f_z(x) d\mu_\alpha(x) = F_\alpha(z) - F_\alpha(-i)$$

which is by the Aronzajn-Krain formula equal to

$$h_z(\alpha) := \frac{1}{\alpha + F(z)^{-1}} - \frac{1}{\alpha + F(-i)^{-1}} .$$

Now, if $\pm\mathrm{Im}(z) > 0$, then $\pm\mathrm{Im}F(z) > 0$ so that $\pm\mathrm{Im}F(z)^{-1} < 0$. This means that $h_z(\alpha)$ has either two poles in the lower half plane if $\mathrm{Im}(z) < 0$ or one in each half plane if $\mathrm{Im}(z) > 0$. Contour integration in the upper half plane (now with α) implies that $\int_{\mathbf{R}} h_z(\alpha) \; d\alpha = 0$ for $\mathrm{Im}(z) < 0$ and $2\pi i$ for $\mathrm{Im}(z) > 0$. □

Definition. As we have proven in chapter 1, that any Borel measure μ has the **Lebesgue decomposition** $d\mu = d\mu_{ac} + d\mu_{sing} = d\mu_{ac} + d\mu_{sc} + d\mu_{pp}$.

Proposition 4.6.3 (Facts about Borel transform) *For $\epsilon \to 0$, the measures $\pi^{-1}\mathrm{Im}F(E + i\epsilon)dE$ converge weakly to μ.*
$d\mu_{sing}(\{E \mid \mathrm{Im}F(E + i0) = \infty\}) = 1$,
$d\mu(\{E_0\}) = \lim_{\epsilon \to 0} \mathrm{Im}F(E_0 + i\epsilon)\epsilon$,
$d\mu_{ac}(E) = \pi^{-1}\mathrm{Im}F(E + iI) \; dE$.

Definition. Define for $\alpha \neq 0$ the sets

$$\begin{aligned} S_\alpha &= \{ x \in \mathbf{R} \mid F(x + i0) = -\alpha^{-1}, \; G(x) = \infty \} \\ P_\alpha &= \{ x \in \mathbf{R} \mid F(x + i0) = -\alpha^{-1}, \; G(x) < \infty \} \\ L &= \{ x \in \mathbf{R} \mid \mathrm{Im}F(x + i0) \neq 0 \} \end{aligned}$$

Proposition 4.6.4 (Aronszajn-Donoghue) *The set P_α is the set of eigenvalues of L_α. One has $(d\mu_\alpha)_{sc}(S_\alpha) = 1$ and $(d\mu_\alpha)_{ac}(L) = 1$. The sets P_α, S_α, L are mutually disjoint.*

Proof. If $F(E + i0) = -1/\alpha$, then

$$\lim_{\epsilon \to 0} \epsilon \, \mathrm{Im} F_\alpha(E + i\epsilon) = (\alpha^2 G(E))^{-2}$$

since $F(E+i\epsilon) = -1/\alpha + i\epsilon G(x) + o(\epsilon)$ if $G(E) < \infty$ and $\epsilon^{-1}\mathrm{Im}(1+\alpha F) \to \infty$ if $G(E) = \infty$ which means $\epsilon|1 + \alpha F|^{-1} \to 0$ and since $F \to -1/\alpha$, one gets $\epsilon|F/(1 + \alpha F)| \to 0$.

The theorem of de la Vallée Poussin (see Rudin: Real and complex analysis) states that the set $\{E \mid \ |F_\alpha(E + i0)| = \infty\}$ has full $(d\mu_\alpha)_{sing}$ measure. But since $F_\alpha = F/(1 + \alpha F)$, we know that $|F_\alpha(E + i0)| = \infty$ if and only if $F(E + i0) = -1/\alpha$. □

Important will be the following criterium of Simon-Wolff. In the case of IID potentials with absolutely continuous distribution, a spectral averaging argument will then lead to pure point spectrum also for $\alpha = 0$.

Theorem 4.6.5 (Simon-Wolff criterium) *For any interval $[a, b] \subset \mathbf{R}$, the operator L_α has pure point spectrum for Lebesgue all α if $G(E) < \infty$ for Lebesgue almost all $E \in [a, b]$.*

Proof. By hypothesis, the Lebesgue measure of $S = \{E \mid G(E) = \infty\}$ is zero. This means by the integral formula that $d\mu_\alpha(S) = 0$ for almost all α. By Aronszajn-Donoghue, this means $\mu_\alpha(S_\alpha \cap [a, b]) = \mu_\alpha(L \cap [a, b]) = 0$ so that μ_α has only point spectrum. □

Lemma 4.6.6 (Formula of Simon-Wolff) *For each $E \in \mathbf{R}$, the sum $\sum_{n \in \mathbf{Z}^d} |(L - E - i\epsilon)_{0n}|^2$ is monotonically increasing as $\epsilon \searrow 0$ and converges pointwise to $G(E)$.*

Proof. For $\epsilon > 0$, we have

$$\begin{aligned}
\sum_{n \in \mathbf{Z}^d} |(L - E - i\epsilon)_{0n}|^2 &= ||(L - E - i\epsilon)\delta_0||^2 = |[(L - E - i\epsilon)(L - E + i\epsilon)]_{00}| \\
&= \int_{\mathbf{R}} \frac{d\mu(x)}{(x - E)^2 + \epsilon^2}
\end{aligned}$$

from which the monotonicity and the limit follow. □

Lemma 4.6.7 (An elementary integral estimate) *There exists a constant C, such that for all $\alpha, \beta \in \mathbf{C}$*

$$\int_0^1 |x-\alpha|^{1/2}|x-\beta|^{-1/2} \, dx \geq C \int_0^1 |x-\beta|^{-1/2} \, dx \; .$$

Proof. We can suppose that $\alpha \in [0,1]$, since replacing a general $\alpha \in \mathbf{C}$ with the nearest point in $[0,1]$ only decreases the left hand side. Since the symmetry $\alpha \mapsto 1/-\alpha$ leaves the claim invariant, we can also suppose that $\alpha \in [0,1/2]$. But then

$$\int_0^1 |x-\alpha|^{1/2}|x-\beta|^{-1/2} \, dx \geq (1/4)^{1/2} \int_{3/4}^1 |x-\beta|^{-1/2} \, dx \; .$$

The function

$$h(\beta) = \frac{\int_{3/4}^1 |x-\beta|^{-1/2} \, dx}{\int_0^1 |x-\alpha|^{1/2}|x-\beta|^{-1/2} \, dx}$$

is nonvanishing, continuous and satisfies $h(\infty) = 1/4$. Therefore $C := \inf_{\beta \in \mathbf{C}} h(\beta) > 0$. □

Lemma 4.6.8 (An estimate for the free Laplacian) *Let $f, g \in l^\infty(\mathbf{Z}^d)$ be nonnegative and let $0 < a < (2d)^{-1}$.*

$$(1-a\Delta)f \leq g \Rightarrow f \leq (1-a\Delta)^{-1} g \; .$$

$$[(1-a\Delta)^{-1}]_{ij} \leq (2da)^{|j-i|}(1-2da)^{-1} \; .$$

Proof. Since $||\Delta|| < 2d$, we can write $(1-a\Delta)^{-1} = \sum_{m=0}^\infty (a\Delta)^m$ which is preserving positivity. Since $[(a\Delta)^m]_{ij} = 0$ for $m < |i-j|$ we have

$$[(a\Delta)^m]_{ij} = \sum_{m=|i-j|}^\infty [(a\Delta)^m]_{ij} \leq \sum_{m=|i-j|}^\infty (2da)^m \; .$$

□

Proof of theorem 4.6.1:

Proof. In order to prove 4.6.1, we have by Simon-Wolff only to show that $G(E) < \infty$ for almost all E. This will be achieved by proving $E[G(E)^{1/4}] < \infty$. By the formula of Simon-Wolff, we have therefore to show that

$$\sup_{z \in \mathbf{C}} E[(\sum_n |G(n,0,z)|^2)^{1/4}] < \infty \; .$$

Since $(\sum_n |G(n,0,z)|^2)^{1/4}) \leq \sum_n |G(n,0,z)|^{1/2}$, we have only to control the later. Define $g_z(n) = G(n,0,z)$ and $k_z(n) = E[|g_z(n)|^{1/2}]$. The aim is now to give an estimate for

$$\sum_{n \in \mathbf{Z}^d} k_z(n)$$

which holds uniformly for $\mathrm{Im}(z) \neq 0$.

(i)

$$E[|\lambda V(n) - z|^{1/2} |g_z(n)|^{1/2}] \leq \delta_{n,0} + \sum_{|j|=1} k_z(n+j) \ .$$

Proof. $(L-z)g_z(n) = \delta_{n0}$ means

$$(\lambda V(n) - z) g_z(n) = \delta_{n0} - \sum_{|j|=1} g_z(n+j) \ .$$

Jensen's inequality gives

$$E[|\lambda V(n) - z|^{1/2} |g_z(n)|^{1/2}] \leq \delta_{n0} + \sum_{|j|=1} k_z(n+j) \ .$$

(ii)

$$E[|\lambda V(n) - z|^{1/2} |g_z(n)|^{1/2}] \geq C\lambda^{1/2} k(n) \ .$$

Proof. We can write $g_z(n) = A/(\lambda V(n) + B)$, where A, B are functions of $\{V(l)\}_{l \neq n}$. The independent random variables $V(k)$ can be realized over the probability space $\Omega = [0,1]^{\mathbf{Z}^d} = \prod_{k \in \mathbf{Z}^d} \Omega(k)$. We average now $|\lambda V(n) - z|^{1/2} |g_z(n)|^{1/2}$ over $\Omega(n)$ and use the elementary integral estimate lemma:

$$\begin{aligned}
\int_{\Omega(n)} |\lambda v - z|^{1/2} |A|^{1/2} |\lambda v + B|^{-1/2} \, dv &= |A|^{1/2} \int_0^1 |v - z\lambda^{-1}| |v + B\lambda^{-1}|^{-1/2} \, dv \\
&\geq C|A|^{1/2} \int_0^1 |v + B\lambda^{-1}|^{-1/2} \, dv \\
&= C\lambda^{1/2} \int_0^1 |A/(\lambda v + B)|^{1/2} = E[g_z(n)^{1/2}] = k_z(n) \ .
\end{aligned}$$

(iii)

$$k_z(n) \leq (C\lambda^{1/2})^{-1} \left(\sum_{|j|=1} k_z(n+j) + \delta_{n0} \right) \ .$$

Proof. Follows directly from (i) and (ii).

(iv)

$$(1 - C\lambda^{1/2}\Delta) k \leq \delta_{n0} \ .$$

Proof. Rewriting of (iii).

(v) Define $\alpha = C\lambda^{1/2}$.

$$k_z(n) \leq \alpha^{-1}(2d/\alpha)^{|n|}(1 - 2d/\alpha)^{-1} .$$

Proof. For $\mathrm{Im}(z) \neq 0$, we have $k_z \in l^\infty(\mathbf{Z}^d)$. From Lemma 4.6.8 and (iv), we have

$$k(n) \leq \alpha^{-1}[(1 - \Delta/\alpha)^{-1}]_{0n} \leq \alpha^{-1}(2d/\alpha)^{|n|}(1 - 2d/\alpha)^{-1} .$$

(vi) For $\lambda > 4C^{-2}d^2$, we get a uniform bound for $\sum_n k_z(n)$.
Proof. Since $C\lambda^{1/2} < (2d)^{-1}$, we get the estimate from (v).

(vii) Pure point spectrum.
Proof. By Simon-Wolff, we have pure point spectrum for L_α for almost all α. Since the set of random operators of L_α and L_0 coincide on a set of measure $\geq 1 - 2\alpha$, we get also pure point spectrum of L for almost all ω. □

Bibliography

[1] W. Feller. *An introduction to probability theory and its applications.* John Wiley and Sons, 1968.

[2] D. Freedman. *Markov Chains.* Springer Verlag, New York Heidelberg, Berlin, 1983.

[3] G. Grimmet. *Percolation.* Springer Verlat, 1989.

[4] H.G.Tucker. *A graduate course in probability.* Probability and Mathematical Statistics. Academic Press, 1967.

[5] N.van Kampen. *Stochastic processes in physics and chemistry.* North-Holland Personal Library, 1992.

[6] A.F. Karr. *Probability.* Springer texts in statistics. Springer-Verlag, 1993.

[7] H. Kunita. *Stochastic flows and stochastic differential equations.* Cambridge University Press, 1990.

[8] H.P. McKean. *Stochastic integrals.* Academic Press, 1969.

[9] M. Nagasava. *Schrödinger equations and diffusion theory.* Monographs in Mathematics, Vol. 86. 1983.

[10] E. Nelson. *Dynamical theories of brownian motion.* Princeton university press, 1967.

[11] E. Nelson. *Radically elementary probability theory.* Princeton university text, 1987.

[12] D. Revuz and M.Yor. *Continuous Martingales and Brownian Motion.* Springer Verlag, 1991. Grundlehren der mathmatischen Wissenschaften, 293.

[13] B. Simon. *Functional Integration and Quantum Physics.* Academic Press, 1979. Pure and applied mathematics.

[14] F. Spitzer. *Principles of Random walk.* Graduate texts in mathematics. Springer-Verlag, New York Heidelberg Berlin, 1976.

[15] D.W. Stroock. *Probability theory, an analytic view.* Cambridge University Press, 1993.

[16] D. Williams. *Probability with Martingales.* Cambridge mathematical Texbooks, 1991.