Lecture Notes on Probability

Taught winter term 1999/2000
By Peter Mörters

Universität Kaiserslautern

Contents

Chapter 1

Preface

The present volume represents most of my notes designed for the lecture "Probability Theory" taught in the winter term 99/00 at Universität Kaiserslautern to an audience of mixed background and previous knowledge and, as it turned out, also of mixed interests. The prerequisites I assumed for the lecture were knowledge of the course "Stochastische Methoden", a script of which is available at our department (in German). This previous knowledge of basic probability theory and, much more importantly, the many examples of random phenomena discussed in this course allowed me to start with a chapter on the construction of abstract probability spaces before embarking upon the more interesting subject of stochastic processes.

Although many important special processes of probability theory, like Galton-Watson processes or the Poisson process, appear in several places as examples, Brownian motion is at the heart of the material taught here and two of the main theorems, Theorem 3.3 and Theorem 6.6 are the best reasons why this is an inevitable choice. Existence of Brownian motion is shown following Lévy's elegant approach. The Markov property is discussed using the example of Brownian motion and without introducing the abstract concept of Markov processes. Instead the lecture contains many applications of the Markov property to the calculation of probabilities associated with Brownian motion. I hope that this leads to a better intuition. Martingales are discussed mainly in a discrete time setting, but an application to the exit probabilities of Brownian motion embeds the concept in the general flow of the lecture. The proof of Donsker's invariance principle by Skorokhod's embedding uses almost all the major results of the lecture in some way and is the climax of the development. The last lecture was spent with an outlook to the second part of the lecture, an introduction to stochastic calculus. The script also contains a collection of some of the accompanying exercises.

As usual, when writing these notes I have made frequent use of several good books on probability. A large part of the material is based on the excellent books by Rick Durrett and David Williams, which are much recommended to everyone who wants to get friendly with probability.

These notes do not include the first lectures, an overview of the material presented in the course "Stochastische Methoden", and —more importantly— the many pictures I drew during the lectures in order to explain theorems, illustrate ideas and give examples. As they were perhaps the most helpful contribution to the understanding of the subject I would recommend the readers of these notes not to try to learn the material from these notes without attending a lecture on probability theory. Hopefully these lectures will be offered annually.

Finally, I would like to thank the audience for their effort, Jochen Blath for his tutorials and several good discussions on the lectures and exercises and Christoph Lossen for gently forcing me to keep my notes in tidy LaTeX-format until the end of term.

Kaiserslautern, 3rd of February 2000.
Peter Mörters.

Chapter 2

Abstract probability spaces and random variables

2.1 Abstract probability spaces

In the previous course we have modelled phenomena involving a moment of uncertainty or randomness by means of the abstract concepts of a *probability space* and of *random variables* defined on this probability space. A *probability space* is a triple $(\Omega, \mathcal{A}, \mathbf{P})$ consisting of three objects:

- A set Ω consisting of all possible elementary events $\omega \in \Omega$, which might occur.
- A collection $\mathcal{A} \subset \mathcal{P}(\Omega)$ of subsets of Ω, the events we might observe. This collection in general need not consist of the whole collection $\mathcal{P}(\Omega)$ of subsets of Ω, but is a subcollection, a so-called *σ-field*. The pair $(\Omega, \mathcal{A})$ is called a *measurable space*.
- A probability measure $\mathbf{P}$ on the measurable space $(\Omega, \mathcal{A})$, which assigns to each event $A \in \mathcal{A}$ its probability $\mathbf{P}(A)$.

A collection $\mathcal{A} \subset \mathcal{P}(\Omega)$ of subsets of Ω is called a *σ-field* if the following three richness conditions are satisfied.

- $\emptyset \in \mathcal{A}$,
- if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$,
- if $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

A mapping $\mathbf{P} : \mathcal{A} \to [0,1]$ is called a *probability measure* if $\mathbf{P}(\Omega) = 1$ and for every countable collection $A_1, A_2, A_3, \ldots \in \mathcal{A}$ of events, which are pairwise disjoint, we have

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

A *random variable* is a measurable mapping $X : \Omega \to \Omega'$ from the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to a measurable space $(\Omega', \mathcal{A}')$. A mapping is called *measurable* if $X^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{A}'$. The probability measure $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ defined on the measurable space $(\Omega', \mathcal{A}')$ by

$$\mathbb{P}_X(A') = \mathbb{P}\{X \in A'\} = \mathbb{P}(X^{-1}(A')) \text{ for } A' \in \mathcal{A}'$$

is called the *distribution of X*.

Example The rabbit Bunny has a random number of children, each child again has — independently of its sisters and Bunny herself and with the same productivity as Bunny — a random number of children. The owner of the rabbit observes the number of grand children. Construct a probability space and a random variable, which appropriately describe the situation.

Solution: Ω should be

$$\Omega = \{(k; a_1, \ldots, a_k) \,:\, k \in \mathbb{N}, a_j \in \mathbb{N}\}.$$

Clearly, $\omega = (k; a_1, \ldots, a_k)$ means that Bunny has k children and the jth child has a_j children. A smaller Ω does not fully describe the situation, a larger Ω is possible, but makes the construction of the probability measure more complicated. We define $\mathcal{A} = \mathcal{P}(\Omega)$. Suppose that $p_0, p_1, \ldots$ is a sequence of nonnegative numbers with $\sum_{n=0}^{\infty} p_n = 1$, meaning that p_j is the probability that Bunny has j children. Then we define $\mathbb{P}$ first on the singletons by

$$\mathbb{P}\{(k; a_1, \ldots, a_k)\} = p_k p_{a_1} \cdots p_{a_k}\,.$$

Observe that Ω is a countable set and we can define $\mathbb{P}$ for any set $A \subset \Omega$ by

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

Then it is clear that $\mathbb{P}$ satisfies the additivity axiom and

$$\mathbb{P}(\Omega) = \sum_{k=1}^{\infty} \sum_{a_1=1}^{\infty} \cdots \sum_{a_k=1}^{\infty} p_k p_{a_1} \cdots p_{a_k} = 1.$$

The random variable, which describes the owner's observation is given by $X : \Omega \to \mathbb{N}$,

$$X(k; a_1, \ldots, a_k) = \sum_{j=1}^{k} a_j\,.$$

In the description of the example we have used the word independence naively. Heuristically, independence of two σ-fields means that each pair of events from the two σ-fields do not influence each other. We now recall the strict definition of independence.

Definition Suppose that $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space. Two σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$ are called *independent* if for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ we have

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\,.$$

Two random variables $X_1 : (\Omega, \mathcal{A}) \to (\Omega_1, \mathcal{A}_1)$ and $X_2 : (\Omega, \mathcal{A}) \to (\Omega_2, \mathcal{A}_2)$ are called *independent* if the σ-fields $X_1^{-1}(\mathcal{A}_1)$ and $X_2^{-1}(\mathcal{A}_2)$ of preimages are independent. Equivalently, for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ we require

$$\mathbb{P}(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) = \mathbb{P}\{X_1 \in A_1\}\mathbb{P}\{X_2 \in A_2\}.$$

More generally, a family $\{X_\lambda : \lambda \in \Lambda\}$ of random variables is independent, if for every finite subfamily $\{X_{\lambda_1}, \ldots, X_{\lambda_n}\}$ we have

$$\mathbf{P}\{X_{\lambda_1} \in A_1, \ldots, X_{\lambda_n} \in A_n\} = \prod_{k=1}^{n} \mathbf{P}\{X_{\lambda_k} \in A_k\}\,.$$

Example Show that in the example, if $p_0 \neq 0, 1$, the number of Bunny's children and the number of her grandchildren is dependent.

SOLUTION: On the probability space Ω we have to show the dependence of the random variables

$$Y : (k; a_1, \ldots, a_k) \mapsto k\,,$$

and

$$X : (k; a_1, \ldots, a_k) \mapsto \sum_{j=1}^{k} a_j\,.$$

Clearly, Y represents the number of children of Bunny herself and X the number of her grandchildren. For the proof it suffices to find two sets A, B of natural numbers such that

$$\mathbf{P}(\{X \in A\} \cap \{Y \in B\}) \neq \mathbf{P}\{X \in A\}\mathbf{P}\{Y \in B\}.$$

If $p_0 \neq 0, 1$, the easiest try is $B = \{0\}$ and $A = \{1, 2, \ldots\}$. Then $\{Y \in B\}$ implies $\{X \not\in A\}$ and we have $0 \neq \mathbf{P}\{X \neq 0\} \cdot p_0$.

Here is a **side remark**. One cannot check that in our model of Bunny's family the number of children of two different children is indeed independent (as was required in the naive formulation of the original exercise). This is due to the fact that the probability space Ω is too small to define a random variable Y_k which describes the number of children of Bunny's kth child (there is a problem if Bunny has less than k children). A solution to this problem is the concept of conditional probabilities, which we will discuss later. Alternatively, one can work on a larger probability space, as we will do in the next section.

2.2 Identification and construction of probability measures

In the example of the rabbit Bunny it would be useful to find a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and a family X_n of random variables on this space such that X_n is the number of rabbits alive in the nth generation. This is necessary, for example, to find out whether there is a population explosion, i.e. if $\lim_{n\to\infty} X_n = \infty$. In this case the set Ω has to be uncountable. Hence we cannot just define $\mathbf{P}$ on the singletons and extend this to arbitrary sets as before, but we need more complicated tools to define $\mathbf{P}$.

In this section we provide the essential measure theoretic tools needed to construct more complicated probability spaces. First we recall the simple method how to construct a σ-field.

Lemma 2.1 (Generation of σ-fields) *For every arbitrary collection $\mathcal{B} \subset \mathcal{P}(\Omega)$ of subsets of Ω there is a uniquely determined σ-field $\mathcal{A} = \sigma(\mathcal{B})$ such that $\mathcal{B} \subset \mathcal{A}$ and $\mathcal{A} \subset \mathcal{F}$ for every σ-field $\mathcal{F}$, which contains $\mathcal{B}$. In other words, the σ-field $\sigma(\mathcal{B})$ is the smallest σ-field, which contains $\mathcal{B}$ and is called* the σ-field generated by $\mathcal{B}$.

Proof: We have to show the existence of $\sigma(\mathcal{B})$. In fact $\sigma(\mathcal{B})$ is the intersection of all σ-fields, which contain $\mathcal{B}$. Then the two properties are clearly satisfied and for the proof one has to check that this intersection is indeed a σ-field, which is easy. Uniqueness is trivial. ■

Exercise: Find the σ-field on $\Omega = \{1,\dots,6\} \times \{1,\dots,6\} \times \{1,\dots,6\}$, which is generated by the events $\{(a,b,c) : a+b \text{ is even}\}$, $\{(a,b,c) : a+c \text{ is even}\}$, $\{(a,b,c) : b+c \text{ is even}\}$ and $\{a$ is larger than $4\}$.

The following lemma is sometimes useful and will be proved as an exercise.

Lemma 2.2 (Measurability criterion) *Suppose that $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ are measurable spaces and $X : \Omega \to \Omega'$ a mapping. If the σ-field $\mathcal{A}'$ is generated by a collection $\mathcal{B} \subset \mathcal{A}'$ and $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, then X is measurable.*

There are two important questions, the existence and the uniqueness of probability measures.

(1) if we have a mapping $P : \mathcal{B} \to [0,1]$, when can we find an extension $\mathbb{P} : \sigma(\mathcal{B}) \to [0,1]$, which is a probability measure?

(2) if we have two probability measures $\mathbf{P}_1, \mathbf{P}_2 : \sigma(\mathcal{B}) \to [0,1]$, when does $\mathbf{P}_1(B) = \mathbf{P}_2(B)$ for all $B \in \mathcal{B}$ imply $\mathbf{P}_1(A) = \mathbf{P}_2(A)$ for all $A \in \sigma(\mathcal{B})$?

As these questions belong to the realm of measure theory rather than probability theory, we won't prove the two theorems, which give rather satisfactory answers to these questions. Instead, we will give nontrivial examples of their application.

Theorem 2.3 (Caratheodory Extension Theorem) *Suppose $\mathcal{B} \subset \mathcal{P}(\Omega)$ is a nonempty collection of subsets of Ω, such that whenever $A, B \in \mathcal{B}$, then $A^c \in \mathcal{B}$ and $A \cup B \in \mathcal{B}$. Such a collection is called a field. Suppose further that $P : \mathcal{B} \to [0,1]$ satisfies $P(\Omega) = 1$ and $P(A \cup B) = P(A) + P(B)$, whenever $A, B \in \mathcal{A}$ are disjoint. Then, if P satisfies the* continuity from above, *i.e. if*

$$\lim_{n\to\infty} P(A_n) = 0 \text{ for every sequence of sets } A_n \in \mathcal{B} \text{ with } A_n \downarrow \emptyset,$$

there exists a probability measure $\mathbf{P}$ on the σ-field $\sigma(\mathcal{B})$ such that $P(B) = \mathbf{P}(B)$ for all $B \in \mathcal{B}$.

The **proof** can be found in the book of Halmos on page 54 or in the book of Elstrodt, Chap. II, §4 or in the book of Durrett, Appendix A2. Note that in the definition of a field we could replace the condition on the finite union by the equivalent condition for finite intersections.

Theorem 2.4 (Uniqueness Theorem) *Suppose $\mathcal{B} \subset \mathcal{P}(\Omega)$ is a collection of subsets of Ω, such that whenever $A, B \in \mathcal{B}$, then $A \cap B \in \mathcal{B}$. Such a collection is called $\cap$-stable. Suppose further that $\mathbf{P}_1, \mathbf{P}_2 : \sigma(\mathcal{B}) \to [0,1]$ are probability measures on the σ-field generated by $\mathcal{B}$. Then $\mathbf{P}_1$ and $\mathbf{P}_2$ agree on $\sigma(\mathcal{B})$ if they agree on $\mathcal{B}$.*

Remark: By the uniqueness theorem, the extension constructed in the Caratheodory extension theorem is always uniquely determined. The **proof** of the theorem can be found in the book of Elstrodt, Chapter II, §5, or in Durrett's book, Appendix A2.

A useful consequence of the Uniqueness Theorem is that independence of random variables can be checked more easily.

Lemma 2.5 (Independence Criterion) *Suppose that* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space and* $\mathcal{A}_1 \subset \mathcal{A}$ *and* $\mathcal{A}_2 \subset \mathcal{A}$ *are* σ*-fields with* $\cap$*-stable generators* $\mathcal{B}_1$ *and* $\mathcal{B}_2$*. If*

$$\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\mathbb{P}(B_2)$$

for all $B_1 \in \mathcal{B}_1$ *and* $B_2 \in \mathcal{B}_2$*, then* $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *are independent.*

Proof: We fix a set $B_2 \in \mathcal{B}_2$ with $\mathbb{P}(B_2) > 0$ and look at the probability measures P_1 and P_2 on $(\Omega, \mathcal{A}_1)$ defined by $P_1(A) = \mathbb{P}(A)$ and $P_2(A) = \mathbb{P}(A \cap B_2)/\mathbb{P}(B_2)$ for all $A \in \mathcal{A}_1$. By assumption they agree on the $\cap$-stable generator $\mathcal{B}_1$ and hence $P_1 = P_2$. If we now let B_2 vary in $\mathcal{B}_2$ this means that for every fixed $A \in \mathcal{A}_1$ the probability measures $\mathbb{P}$ and P defined by $P(B) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ agree on $\mathcal{B}_1$ and hence they agree altogether on $\mathcal{A}_2$ and this means $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$. ∎

Our first and foremost application of the extension theorem is the existence of product spaces. Suppose we are given a sequence of probability measures P_n on some measurable space $(\Omega_n, \mathcal{A}_n)$. Is there a probability space Ω and random variables X_n such that X_i and X_j are independent for each $i \neq j$ and the distribution of X_j is P_j? We start with the existence of finite product spaces.

Theorem 2.6 (Existence of finite product spaces) *Suppose* $(\Omega_i, \mathcal{A}_i, P_i), 1 \leq j \leq n$ *is a family of probability spaces. Then we define,*

- *the product space* $\Omega = \prod_{j=1}^n \Omega_j$ *of all tuples* $\omega = (\omega_1, \dots, \omega_n)$ *with* $\omega_j \in \Omega_j$*,*
- *the product* σ*-field* $\mathcal{A} = \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$ *generated by the* rectangle events

  $$A_1 \times \cdots \times A_n = \{(\omega_1, \dots, \omega_n) \in \Omega \,:\, \omega_1 \in A_1, \dots, \omega_n \in A_n\}$$

  for all $A_1, \dots, A_n$ *with* $A_i \in \mathcal{A}_i$*,*
- *the mappings* $X_j : \Omega \to \Omega_j$ *as the projections* $X_j(\omega) = \omega_j$ *for* $1 \leq j \leq n$*.*

Then there is a unique probability measure $\mathbb{P}$ *(sometimes called* $P_1 \otimes \cdots \otimes P_n$*) on* $\mathcal{A}$ *such that*

$$\mathbb{P}(A_1 \times \cdots \times A_n) = P_1(A_1) \times \cdots \times P_n(A_n)\,.$$

On the space $(\Omega, \mathcal{A}, \mathbb{P})$ *the random variables* $X_1, \dots, X_n$ *are independent and the distribution of* X_j *is* P_j*.*

Proof: To simplify notation we assume $n = 2$. We denote by $\mathcal{B}$ the collection of all finite disjoint unions of rectangle events. Then, clearly, $\mathcal{B}$ is a generator of the product σ-field. To see that it is a field, observe that

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c)\,,$$

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)\,.$$

On $\mathcal{B}$ we define

$$P\Big((A_1^1 \times A_2^1) \cup \ldots \cup (A_1^n \times A_2^n)\Big) = \sum_{i=1}^{n} P_1(A_1^i)P_2(A_2^i)\,,$$

where the rectangle events have to be disjoint. We have to check that this mapping $P : \mathcal{B} \to [0,1]$ is well-defined. Assume that $A \times B = \bigcup A_i \times B_i$ with the union disjoint and countable or finite. Then, for each $x \in A$ let $I(x) = \{i : x \in A_i\}$. Then $B = \bigcup_{i \in I(x)} B_i$ and the union is disjoint, hence

$$1_A(x)P_2(B) = \sum_i 1_{A_i}(x)P_2(B_i)\,.$$

Integrating this with respect to P_1 and exchanging summation and limit, which is allowed by monotone convergence, yields

$$P(A \times B) = P_1(A)P_2(B) = \sum_i P_1(A_i)P_2(B_i)\,.$$

In particular, the definition of P is independent of the choice of the representation as a disjoint union of rectangle events. In order to apply the Extension Theorem, we have to check the *continuity from above*. So let $C_n \downarrow \emptyset$ where $C_n \in \mathcal{B}$. Then put $K_0 = \Omega \setminus C_1$ and $K_n = C_n \setminus C_{n+1}$. Each K_i is in $\mathcal{B}$, they are disjoint and their union is Ω. Hence, as we have seen,

$$\begin{aligned} 0 &= 1 - P(\Omega) = 1 - P\Big(\bigcup_{n=0}^{\infty} K_n\Big) = 1 - \sum_{n=0}^{\infty} P(K_n) \\ &= 1 - \lim_{N\to\infty} \Big[\sum_{n=1}^{N} P(K_n)\Big] = \lim_{N\to\infty} P(C_{N+1})\,, \end{aligned}$$

because $\bigcup_{n=0}^{N} K_n = \Omega \setminus C_{N+1}$. Hence P can be extended to a probability measure $\mathbf{P}$ on $\mathcal{A}$. Now the distribution of X_1 is P_1, because for $A \in \mathcal{A}_1$ we have

$$\mathbf{P}\{X_1 \in A\} = \mathbf{P}(A_1 \times \Omega) = P_1(A_1),$$

and analogously for X_2. Finally, the independence of X_1 and X_2 follows from

$$\mathbf{P}\{X_1 \in A, X_2 \in B\} = \mathbf{P}(A \times B) = P_1(A)P_2(\Omega_2)P_1(\Omega_1)P_2(B) = \mathbf{P}\{X_1 \in A\}\mathbf{P}\{X_2 \in B\}\,.$$

■

The main tool for the calculation of probabilities in product spaces is the following version of Fubini's Theorem.

Corollary 2.7 (Fubini's Theorem) *Suppose that* $(\Omega, \mathcal{A}, \mathcal{P})$ *is a product space as before and* $X : \Omega \to [0, \infty]$ *is measurable with respect to the product* σ*-field* $\mathcal{A}$*. Suppose that* $A \cup B \subset \{1, \dots, n\}$ *and* $A \cap B = \emptyset$ *and* $\mathbf{P}_A$ *and* $\mathbf{P}_B$ *the product measures on* $(\Omega_A = \prod_{a \in A} \Omega_a, , \mathcal{A}_A = \otimes_{a \in A} \mathcal{A}_a)$*, resp.* $(\Omega_B, \mathcal{A}_B)$*. Let* ω_a *be the projection of* $\omega \in \Omega$ *to* Ω_a *and analogously for* ω_b*, and write* $\omega = (\omega_a, \omega_b)$*. Then* Ω *is the product of* $(\Omega_A, \mathcal{A}_A, \mathbf{P}_A)$ *and* $(\Omega_B, \mathcal{A}_B, \mathbf{P}_B)$*.*

$$\mathbf{E}X = \int X \, d\mathbf{P} = \int_{\Omega_A} \int_{\Omega_B} X(\omega_a, \omega_b) \, d\mathbf{P}_B(\omega_b) \, d\mathbf{P}_A(\omega_a) \, .$$

In particular, by induction,

$$\mathbf{E}X = \int_{\Omega_1} dP_1(\omega_1) \cdots \int_{\Omega_n} dP_n(\omega_n) \, X(\omega_1, \dots, \omega_n) \, .$$

Proof: For the proof of the measurability of the integrands, see the homeworks or Durrett, Appendix A4. The fact that $(\Omega, \mathcal{A}, \mathbf{P})$ is the product of the product spaces Ω_A and Ω_B is clear from the definition. To check the formula, it suffices, by the standard measure theory trick, to prove the corollary for an indicator function 1_E with $E \in \mathcal{A}$ in place of a general X. Then the formula reads

$$\mathbf{P}(E) = \int_{\Omega_A} \int_{\Omega_B} 1_E(\omega_a, \omega_b) \, d\mathbf{P}_B(\omega_b) \, d\mathbf{P}_A(\omega_a) \, .$$

By definition of a product space this holds true if E is a rectangle event. Now, as a function of E, the left and right hand side of the formula both define probability measures. By the Uniqueness Theorem these measures agree, as they agree on the $\cap$-stable generator of rectangle events. This proves the corollary. ∎

Theorem 2.8 (Existence of infinite product spaces) *Suppose* $(\Omega_n, \mathcal{A}_n, P_n)$ *is a sequence of probability spaces. Then we define*

- *the product space* $\Omega = \prod_{j=1}^{\infty} \Omega_j$ *of all sequences* $\omega = (\omega_j)$ *with* $\omega_j \in \Omega_j$*,*
- *the* σ*-field generated by the* cylinder events

  $$C(A_1, \dots, A_n) = \{(\omega_j) \in \Omega \, : \, \omega_1 \in A_1, \dots, \omega_n \in A_n\}$$

  for all $A_1, \dots, A_n$ *with* $A_i \in \mathcal{A}_i$ *and* $n \in \mathbb{N}$*,*
- *the mappings* $X_j : \Omega \to \Omega_j$ *as the projections* $X_j(\omega) = \omega_j$*.*

Then there is exactly one probability measure $\mathbf{P}$ *on the measurable space* $(\Omega, \mathcal{A})$ *such that*

$$\mathbf{P}\{X_1 \in A_1, \dots, X_n \in A_n\} = P_1(A_1) \cdots P_n(A_n) \, ,$$

for all $A_j \in \mathcal{A}_j$ *and* $n \in \mathbb{N}$*. Moreover, on the probability space* $(\Omega, \mathcal{A}, \mathbf{P})$ *the random variables* $X_1, \dots, X_n$ *are independent and the distribution of* X_j *is* P_j*.*

Proof: We denote by $\mathcal{B}$ the collection of all finite disjoint unions of cylinder events. Then, clearly, $\mathcal{B}$ is a generator of the product σ-field and one can check easily that it is a field. On $\mathcal{B}$ we define

$$P\Big(C(A_1^1,\dots,A_{n_1}^1)\cup\ldots\cup C(A_1^n,\dots,A_{n_n}^n)\Big)=\sum_{i=1}^n \mathbf{P}_{\{1,\dots,n_i\}}(A_1^i\times\cdots\times A_{n_i}^i)\,,$$

where the cylinder sets have to be disjoint and $\mathbf{P}_A$ are the finite product measures on $\prod_{a\in A}\Omega_a$ for a finite subset $A\subset\mathbb{N}$. If $N=\max n_j$ we can define $A_k^j=\Omega_k$ for $k>n_j$ and continue the equation with

$$=P\Big(C(A_1^1,\dots,A_N^1)\cup\ldots\cup C(A_1^n,\dots,A_N^n)\Big)=\sum_{i=1}^n \mathbf{P}_{\{1,\dots,N\}}(A_1^i\times\cdots\times A_N^i)$$

$$=\mathbf{P}_{\{1,\dots,N\}}\Big(C(A_1^1,\dots,A_N^1)\cup\ldots\cup C(A_1^n,\dots,A_N^n)\Big)\,.$$

It is thus clear that the mapping $P:\mathcal{B}\to[0,1]$ is well-defined and finitely additive. In order to see that P can be continued to a probability measure, we have to check for every sequence $B_n\in\mathcal{B}$ with $B_n\downarrow\emptyset$ that $\lim P(B_n)=0$. Suppose to the contrary. As the sequence $P(B_n)$ is clearly decreasing, there exists a $\delta>0$ such that $P(B_n)>\delta$ for all n. Now we use Fubini's Theorem and dominated convergence to see that, for an appropriate $N=N(n)$,

$$\delta<\lim_{n\to\infty}P(B_n)=\lim_{n\to\infty}\int dP_1(x_1)\int 1_{B_n}(x_1,\dots,x_N)\,d\mathbf{P}_{\{2,\dots,N\}}$$

$$=\int dP_1(x_1)\lim_{n\to\infty}\int 1_{B_n}(x_1,\dots,x_N)\,d\mathbf{P}_{\{2,\dots,N\}}(x_2,\dots,x_N)\,.$$

Hence, there is $x_1\in\Omega_1$ such that

$$\delta<\lim_{n\to\infty}\int 1_{B_n}(x_1,\dots,x_N)\,d\mathbf{P}_{\{2,\dots,N\}}(x_2,\dots,x_N)\,.$$

We proceed inductively and thus construct a sequence $x=(x_1,x_2,\dots)\in\Omega$ such that, for all k,

$$\delta<\lim_{n\to\infty}\int 1_{B_n}(x_1,\dots,x_{k-1},y_k,\dots,y_N)\,d\mathbf{P}_{\{k,\dots,N\}}(y_k,\dots,y_N)\,.$$

However, if $k>N$ the latter integral is just $1_{B_n}(x)$ and, as the limit is decreasing, we must have $x\in\cap B_n=\emptyset$, which is a contradiction. Hence the condition of the Extension Theorem is satisfied and there exists a unique probability measure $\mathbf{P}$ extending P to the whole σ-field $\mathcal{A}$. Now the distribution of X_i is P_i because

$$\mathbf{P}\{X_i\in A\}=P(C(\Omega_1,\dots,\Omega_{i-1},A))=P_i(A).$$

Finally, the independence of $X_1,X_2,\dots$ follows, as

$$\begin{aligned}\mathbf{P}\{X_1\in A_1,\dots,X_n\in A_n\} &= \mathbf{P}(C(A_1,\dots,A_n)=P_1(A_1)\cdots P_n(A_n)\\ &= \mathbf{P}\{X_1\in A\}\cdots\mathbf{P}\{X_n\in A_n\}\,.\end{aligned}$$

■

Remark The σ-field generated by the cylinder events is called the *product σ-field*, the space is the *product space* and the measure the *product measure*, sometimes denoted by $\bigotimes_{j=1}^{\infty} P_j$.

From Fubini's theorem we get a formula for the expectation of random variables, which depend only on finitely many variables.

Corollary 2.9 *Suppose that $X^* : \Omega_1 \times \cdots \times \Omega_n \to [0,\infty]$ is measurable with respect to the (finite) product σ-field $\mathcal{A}^* = \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$. Then X can be defined on the infinite product space $(\Omega, \mathcal{A}, \mathbf{P})$ as $X(\omega) = X(\omega_1, \ldots, \omega_n)$ and we have*

$$\mathbf{E}X = \int_{\Omega_1} \cdots \int_{\Omega_n} X^*(\omega_1, \ldots, \omega_n) \, dP_n \cdots dP_1 \,.$$

EXAMPLE: THE GALTON-WATSON PROCESS AS A SIMPLE MODEL OF POPULATION GROWTH. Let us finish this section by solving the problem of a good probability space for the whole family of Bunny. Following an idea of Neveu we construct a probability space of random trees and a sequence (X_n) of random variables on this space such that X_n is the (random) number of rabbits in the nth generation. As before all rabbits breed independently and the distribution of the number of children of each rabbit is given by the sequence $(p_0, p_1, \ldots)$. Generally, a sequence (X_n) of random variables is sometimes called a *(discrete time) stochastic process*. The process we are going to construct is called the *Galton-Watson process* associated with the offspring distribution given by the probability sequence $(p_0, p_1, p_2, \ldots)$. We shall come back to this process throughout the course of the lecture.

For the construction of Ω we let $\mathbf{N} = \{1, 2, 3, \ldots\}$ and $\mathbf{N}_0 = \{0, 1, 2, 3, \ldots\}$. Define

$$K = \bigcup_{n=0}^{\infty} \mathbf{N}^n \,,$$

the *address space*, where by convention $\mathbf{N}^0 = \{\partial\}$. Observe that K is countable. An element $k \in K$ will be written $k = k_1 \cdots k_n$, where $k_j \in \mathbf{N}$ and in this case $|k| = n$. Formally, let $|\partial| = 0$. If $|k| > 1$ we set $\bar{k} = k_1 \cdots k_{n-1}$ and if $|k| = 1$ let $\bar{k} = \partial$. For $h = h_1 \cdots h_m$, $k = k_1 \cdots k_n$, let $hk = h_1 \cdots h_m k_1 \cdots k_n$.

A *tree* is a set of addresses, i.e. a subset $\omega \subset K$, such that

- $\partial \in \omega$ (i.e. the tree contains the root),
- $\bar{k} \in \omega$ whenever $k \in \omega$ and $|k| \geq 1$ (i.e. each element different from the root has a mother),
- if $k = k_1 \cdots k_n \in \omega$, then also $k_1 \cdots k_{n-1} l \in \omega$ for all $1 \leq l \leq k_n$.

We denote by Ω the set of all trees. In order to equip Ω with a σ-field, we define for every $k = k_1 \cdots k_n \in K$

$$A(k_1 \cdots k_n) = \{\omega \in \Omega \,:\, k_1 \cdots k_n \in \omega\}$$

and let $\mathcal{A} = \sigma(\mathcal{C})$ be the σ-field generated by the collection $\mathcal{C} = \{A(k) \,:\, k \in K\}$. A probability measure on the measurable space $(\Omega, \mathcal{A})$ defines a *random tree*. A first try to define such a probability measure would lead us to define for the set $A(k)$ the probabilities inductively by $P(A(\partial)) = 1$ and

$$P(A(k)) = \Big(\sum_{j=a_n}^{\infty} p_j \Big) P(A(\bar{k})) \,.$$

However, it is easy to see that $\mathcal{C}$ does not fulfill the requirements of the Caratheodory Extension Theorem or of the Uniqueness Theorem and we do not know how to extend P to a probability measure $\mathbf{P}$ on $\mathcal{A}$. We therefore have to follow a different route and infer the existence of the desired space from the Existence Theorem for product spaces.

Recall that K is countable and hence we can look at the product set $\Omega^* = \mathbb{N}_0^K$ endowed with the product σ-field $\mathcal{A}^*$ and the product measure $\mathbf{P}^* = \bigotimes_{k\in K} P_k$ where all P_k are the same and given by $P_k(j) = p_j$. We now define a tree-valued random variable $T : \Omega^* \to \Omega$ and the distribution of T defines the desired probability measure on $(\Omega, \mathcal{A})$. Each element of Ω^* is a K-tuple ω^* with entries ω_k^* from $\{0, 1, 2, \dots\}$ and we define projections $Y_k : \Omega^* \to \mathbb{N}_0$ by $Y_k(\omega^*) = \omega_k^*$, where $\omega^* = (\omega_k^* \,:\, k \in K)$. The idea is that Y_k describes the number of children of the rabbit with address k (but it is also defined if this rabbit was never born). By definition of the product space $(\Omega^*, \mathcal{A}^*, \mathbf{P}^*)$ the Y_k are independent random variables with distribution P_k, i.e. such that

$$\mathbf{P}^*\{Y_k = j\} = \mathbf{P}^*\{\omega^* \,:\, Y_k(\omega^*) = j\} = p_j \,.$$

Then define

$$T(\omega^*) = \{k = k_1 \cdots k_n \,:\, k_{j+1} \le Y_{k_1 \dots k_j}(\omega^*) \text{ for } 0 \le j < n\} \cup \{\partial\} \subset K \,.$$

It is easy to check that $T(\omega^*)$ is not only a subset of K but indeed a tree, i.e. an element of Ω^*. To see that T is measurable observe that

$$T^{-1}(A(k_1 \dots k_n)) = \{\omega^* \,:\, Y_{k_1 \dots k_j}(\omega^*) \ge k_{j+1} \text{ for all } 0 \le j < n\} = \bigcap_{j=0}^{n-1} \pi_{k_1 \cdots k_j}^{-1}\{n \in \mathbb{N}_0, n \ge k_{j+1}\}$$

and the set on the right hand side is in $\mathcal{A}^*$, because all Y_k are measurable. By the measurability criterion of Lemma 2.2 the mapping T is measurable and thus defines a tree-valued random variable. The random tree T is called the *Galton-Watson tree* and we interpret it as the family tree of our friend Bunny. We denote its distribution on the space Ω by $\mathbf{P}$. Y_k describes the number of children of the rabbit which is in position $k = k_1 \dots k_{n-1}$ in the family tree. Observe that, if $k \in T(\omega^*)$, then

$$Y_k(\omega^*) = \max\{k_n \,:\, kk_n \in T(\omega^*)\} \,.$$

We can now define random variables X_n, which describe the number of rabbits of the nth generation by $X_n : \Omega \to \mathbb{N}_0$,

$$X_n(\omega) = \sum_{k\in\omega, |k|=n-1} \max\{k_n \,:\, kk_n \in \omega\} \,,$$

which, implies that

$$X_n(T(\omega^*)) = \sum_{k\in T(\omega^*), |k|=n-1} Y_k(\omega^*) \,.$$

The fact that the Y_k are independent with the distribution given by the sequence $(p_0, p_1, \dots)$ confirms that this model is well fitted to our problem. The sequence of random variables

$X_1, X_2, X_3, \ldots$ define the Galton-Watson process with offspring distribution given by $(p_0, p_1, \ldots)$. As a side remark observe that the random variables Y_k have to be defined on the (larger) probability space $(\Omega^*, \mathcal{A}^*, \mathbf{P}^*)$, whereas the random variables X_n are defined on the space $(\Omega, \mathcal{A}, \mathbf{P})$ of trees.

Chapter 3

Brownian motion

3.1 Generalities on stochastic processes

A stochastic process is a phenomenon which evolves in time in a random manner. There are many examples of phenomena which can be thought of as a function both of time and of a random (or uncertainty) factor, think of the price of shares, the size of some populations, or the number of particles registered by a Geiger counter.

In the last chapter we have called any sequence of random variables defined on the same probability space a *(discrete time) stochastic process.* We know already mathematical examples, of such discrete processes: Sequences of independent, identically distributed random variables $Y_1, Y_2, \ldots$, their partial sums

$$X_n = Y_1 + \cdots + Y_n\,,$$

which are the *random walks* based on a sequence $Y_1, Y_2, \ldots$ and, defined in the last chapter, the *Galton-Watson process* with offspring distribution given by $(p_0, p_1, p_2, \ldots)$. In this chapter we start the discussion of *continuous time stochastic processes.* They are not so easy to construct, but of great practical and theoretical importance. In this chapter we get to know the most important example, the Brownian motion.

Definition: Suppose $I \subset \mathbf{R}$ is an interval. A *(continuous time) stochastic process* (with values in $\mathbf{R}^d$) is a family $\{X(t) : t \in I\}$ of random variables with values in $\mathbf{R}^d$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbf{P})$.

If we also include the variable ω in our consideration, we now write it in square brackets $X(t) : \Omega \to \mathbf{R}^d$ maps ω to $X(t)[\omega]$. Before even knowing an example we make an observation: we may consider every stochastic process as a random function, taking the parameter t as the variable of the function. Abstract measurability questions are settled in the following lemma.

Lemma 3.1 *Let $\{X(t) : t \in I\}$ be a stochastic process. Define by F the set of all functions $f : I \to \mathbf{R}^d$ and equip F with the σ-field $\mathcal{F}$ generated by the sets $\{f \in F : f(t) \in A\}$ where $t \in I$ and $A \subset \mathbf{R}^d$ is Borel. Then $X : \Omega \to F$ defines an F-valued random variable.*

Proof: By our measurability criterion it suffices to check that $X^{-1}(B) \in \mathcal{A}$ for each $B = \{f \in$

$F \,:\, f(t) \in A\}$ in the generator of $\mathcal{F}$. This is clear, as $X^{-1}(B) = \{\omega \in \Omega \,:\, X(t)[\omega] \in A\} \in \mathcal{A}$. ■

We say that two stochastic processes $\{X(t) \,:\, t \in I\}$ on $(\Omega, \mathcal{A}, \mathbf{P})$ and $\{Y(t) \,:\, t \in I\}$ on $(\Omega', \mathcal{A}', \mathbf{P}')$ are *equivalent* if the associated random functions X and Y have the same distribution. By the Uniqueness Theorem this is the same as requiring that for all $t_1 \le \ldots \le t_n \in I$ and $A_1, \ldots, A_n$ Borel,

$$\mathbf{P}\Big\{X(t_1) \in A_1, \ldots, X(t_n) \in A_n\Big\} = \mathbf{P}'\Big\{Y(t_1) \in A_1, \ldots, Y(t_n) \in A_n\Big\}.$$

We also say that X is a *version* of Y.

Our special interest focuses on stochastic processes, such that the random function $X[\omega]$ is *almost surely continuous*. This means that on the underlying probability space $(\Omega, \mathcal{A}, \mathbf{P})$ there is a set $A \in \mathcal{A}$ with $\mathbf{P}(A) = 1$, such that $t \mapsto X(t)[\omega]$ is a continuous function for each $\omega \in A$. On also says that X has *almost surely continuous paths*.

However, one can observe (this is an exercise) that for two equivalent processes X and Y, X may be almost surely continuous, but Y is almost surely not continuous. This implies that the set

$$\{f \in F \,:\, f \text{ is continuous }\}$$

is not in the σ-field $\mathcal{F}$. It is therefore useful to construct an almost surely continuous processes first by means of probability measures on the space of *continuous* functions and then define $X(t)$ afterwards as the value of the random function at time t.

3.2 Definition of Brownian motion

In 1827 the English botanist Brown observed that pollen particles suspended in a liquid perform irregular random movements. This is due to the hitting of pollen by the much smaller molecules of the liquid. These hits occur a large number of times in any small interval of time, independently of each other and hence a simple mathematical model of this process should be a stochastic process $\{B(t) \,:\, t \ge 0\}$ with the following features:

(1) for all times $0 \le t_1 \le t_2 \le \ldots \le t_n$ the displacements or increments $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), \ldots, B(t_2) - B(t_1)$ are independent, we say that $B(t)$ *is a process with independent increments*,

(2) the distribution of the displacement $B(t+h) - B(t)$ does not depend on t, we say that the process is *stationary* or has *stationary increments*.

(3) the process $\{B(t) \,:\, t \ge 0\}$ has almost surely continuous paths.

A real valued process with the first two properties is sometimes called a *Lévy Process*. It should be clear that processes with these features do come up naturally as a model in many other situations in nature, science and other fields of applications. The first question however must be: Do nontrivial processes with such features exist (in a mathematical sense)? In particular, it is not clear whether the randomness required in the first two conditions does not contradict the continuity required in the last condition. We approach this question first in the case of dimension $d = 1$.

A good starting point for an answer to this question is to find necessary consequences of these assumptions. Surprisingly, it turns out that the distributions of the displacements are almost completely determined by the three conditions. Here is our first major probabilistic theorem.

Theorem 3.2 *Suppose that $\{B(t) : t \geq 0\}$ is a real valued stationary process with independent increments and almost surely continuous paths. Then there are μ and $\sigma \geq 0$ such that, for each $t \geq 0$ and $h \geq 0$, the increment $B(t+h) - B(t)$ is normally distributed with expectation $h\mu$ and variance $h\sigma^2$.*

The theorem motivates the following definition.

Definition: A stochastic process $\{B(t) : t \geq 0\}$ is called a *Brownian motion with drift parameter μ, diffusion parameter σ^2 and start in $x \in \mathbb{R}$* if the following holds:

- $B(0) = x$,
- the process has independent increments,
- for all $t \geq 0$ and $h > 0$, the increments $B(t+h) - B(t)$ are normally distributed with expectation μh and variance $\sigma^2 h$,
- almost surely, the function $t \mapsto B(t)$ is continuous.

We say that $\{B(t) : t \geq 0\}$ is a *standard Brownian motion* if $\mu = 0$, $\sigma = 1$ and $x = 0$.

One can show easily that if $B(t)$ is a standard Brownian motion, then the process $Y(t) = x + \sigma B(t) + \mu t$ is a Brownian motion with start in x, drift parameter μ and variance parameter σ^2. Theorem 3.2 now has the following form:

Theorem 3.3 (Characterization of Brownian motion) *Suppose $\{B(t) : t \geq 0\}$ is a real valued stochastic process with stationary, independent increments and almost surely continuous paths. If $B(0) = x$, then there is μ and $\sigma \geq 0$ such that this process is a Brownian motion with start in x, drift parameter μ and variance parameter σ^2.*

We now come to the **proof of Theorem 3.2**. The proof uses the central limit theorem. We recall a formulation of the central limit theorem, which was proved (in a slightly more general form) in the lecture "Stochastische Methoden", see Satz 4.2.

Central Limit Theorem: For each $n \in \mathbb{N}$ let $X_1^n, \ldots, X_n^n$ be independent and identically distributed random variables with expectation μ_n, positive variance σ_n^2 and finite (centred) third moment $\gamma_n = \mathbf{E}|X_i^n - \mathbf{E}X_i^n|^3$ such that

$$\lim_{n\to\infty} \frac{\gamma_n}{\sigma_n^3 \sqrt{n}} = 0\,.$$

Then we have, for all $a \leq b \in [-\infty, \infty]$,

$$\lim_{n\to\infty} \mathbf{P}\Big\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i^n - \mu_n}{\sigma_n} \in (a,b) \Big\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx\,.$$

Proof of Theorem 3.2: We first fix $t = 0$ and $h \ge 0$ and show that the increment $B(h) - B(0)$ is normally distributed, before showing that expectation and variance have the given structure. To determine the distribution of the increments we fix $h > 0$. For each $n \in \mathbb{N}$ and $1 \le k \le n$ we define

$$Y_k^n = \begin{cases} B\left(\frac{hk}{n}\right) - B\left(\frac{h(k-1)}{n}\right) & \text{if this is less than 1 in absolute value,} \\ 0 & \text{otherwise.} \end{cases}$$

Step 1: Almost surely, $B(h) - B(0) = \lim_{n\to\infty} \sum_{k=1}^n Y_k^n$.

Proof: If we hadn't cut the values of the increments at 1 this would clearly hold without the limit. As it stands we have to show that for large n no cutting is needed, i.e. almost surely

$$\lim_{n\to\infty} \max_{k=1}^{n} \Big| B\Big(\frac{hk}{n}\Big) - B\Big(\frac{h(k-1)}{n}\Big) \Big| = 0 \,. \tag{3.1}$$

This holds simply because every continuous function on $[0, h]$ is uniformly continuous. ■

We now establish some facts about the distribution of the Y_k^n. By our assumptions, for a given n, all Y_k^n have the same distribution and are independent.

Step 2: For all $\delta > 0$, we have $\lim_{n\to\infty} n\mathbf{P}\{|Y_1^n| \ge \delta\} = 0$.

Proof: Because almost sure convergence implies convergence in probability we infer from (3.1)

$$\begin{aligned} 1 &= \lim_{n\to\infty} \mathbf{P}\Big\{ \max_{k=1}^{n} \Big| B\Big(\frac{hk}{n}\Big) - B\Big(\frac{h(k-1)}{n}\Big) \Big| < \delta \Big\} \\ &= \lim_{n\to\infty} \mathbf{P} \bigcap_{k=1}^{n} \Big(\Big\{ \Big| B\Big(\frac{hk}{n}\Big) - B\Big(\frac{h(k-1)}{n}\Big) \Big| < \delta \Big\} \Big) \\ &= \lim_{n\to\infty} \mathbf{P}\Big\{ |B(h/n) - B(0)| < \delta \Big\}^n \\ &= \lim_{n\to\infty} \Big(1 - \frac{n\mathbf{P}\{|B(h/n) - B(0)| \ge \delta\}}{n} \Big)^n \\ &= \exp\Big(-\lim_{n\to\infty} n\mathbf{P}\{|B(h/n) - B(0)| \ge \delta\} \Big) , \end{aligned}$$

using Euler's formula. Hence $\lim_{n\to\infty} n\mathbf{P}\{|B(h/n) - B(0)| \ge \delta\} = 0$. As $|Y_1^n| \le |B(h/n) - B(0)|$ the statement follows. ■

Note that in particular,

$$\lim_{n\to\infty} \mathbf{E}|Y_1^n| = \lim_{n\to\infty} \int_0^1 \mathbf{P}\{|Y_1^n| \ge \delta\}\, d\delta = \int_0^1 \lim_{n\to\infty} \mathbf{P}\{|Y_1^n| \ge \delta\}\, d\delta = 0.$$

This implies that also

$$\lim_{n\to\infty} n\mathbf{P}\{|Y_1^n - \mathbf{E}Y_1^n| > \delta\} = 0 \,. \tag{3.2}$$

Step 3: If $\liminf_{n\to\infty} n\mathrm{Var}(Y_1^n) = 0$, then, almost surely $B(h) - B(0)$ is constant.

Proof: We have, by Bienaymé's equality,

$$\mathrm{Var}\Big(\sum_{k=1}^{n} Y_k^n \Big) = n\mathrm{Var}(Y_1^n) \,,$$

and a subsequence of this converges to 0 as $n \to \infty$. In particular, a subsequence of $\sum_{k=1}^n Y_k^n - \sum_{k=1}^n \mathbf{E} Y_k^n$ converges in L^2 to 0. Convergence in L^2 implies convergence in probability, at the same time, by Step 1, $\lim_{n\to\infty} \sum_{k=1}^n Y_k^n = B(h) - B(0)$ in probability. This implies

$$B(h) - B(0) = \lim_{n\to\infty} \sum_{k=1}^n \mathbf{E} Y_k^n \,,$$

which is deterministic and, in particular, normally distributed. ■

Now we can concentrate on the nontrivial case $\liminf_{n\to\infty} n\mathrm{Var}(Y_1^n) > 0$. Define $X_k^n = \sqrt{n} Y_k^n$. Then

$$B(h) - B(0) = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k^n \tag{3.3}$$

In order to apply the central limit theorem to this expression we have to check the moment conditions. Let μ_n be the expectation, σ_n^2 be the variance and γ_n the third (centred) moment of X_1^n. They all exist because $|X_k^n|$ is bounded by $\sqrt{n}$.

Step 4: $\lim_{n\to\infty} \frac{\gamma_n}{\sigma_n^3 \sqrt{n}} = 0.$

Proof: We have, by Step 3,

$$\liminf_{n\to\infty} \sigma_n^2 = \liminf_{n\to\infty} n\mathrm{Var}(Y_k^n) > 0 \,,$$

Choose a small $\delta > 0$. Let $Z_1^n = X_1^n - \mathbf{E} X_1^n$. Because $|Y_1^n - \mathbf{E} Y_1^n|$ is bounded by 2 we infer from (3.2),

$$\begin{aligned}
\gamma_n &= \mathbf{E}[|Z_1^n|^3] \\
&\le n^{3/2}\Big[\mathbf{E}\Big(|Y_1^n - \mathbf{E}Y_1^n|^3 \mathbf{1}_{\{|Y_1^n - EY_1^n| \le \delta\}}\Big) + 8\mathbf{P}\{|Y_1^n - \mathbf{E}Y_1^n| > \delta\}\Big] \\
&\le \sqrt{n}\delta n \mathbf{E}\Big(|Y_1^n - \mathbf{E}Y_1^n|^2\Big) + o(\sqrt{n}) \\
&\le \sqrt{n}\delta \mathbf{E}[(Z_1^n)^2] + o(\sqrt{n}) \,.
\end{aligned}$$

Hence

$$\limsup_{n\to\infty} \frac{\gamma_n}{\sigma_n^3\sqrt{n}} \le \delta \limsup_{n\to\infty} \frac{1}{\sigma_n} \,.$$

As δ can be chosen arbitrarily small and $\limsup 1/\sigma_n < \infty$ the statement follows. ■

Step 5: The increment $B(h) - B(0)$ is normally distributed.

Proof: Observe that by the Central Limit Theorem, for all a, b,

$$\lim_{n\to\infty} \mathbf{P}\Big\{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i^n - \mu_n}{\sigma_n} \in (a,b)\Big\} = \frac{1}{2\pi} \int_a^b e^{-x^2/2}\, dx \,.$$

At the same time, by (3.3),

$$\lim_{n\to\infty} \mathbf{P}\Big\{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i^n - \mu_n}{\sigma_n} \in (a,b)\Big\} = \lim_{n\to\infty} \mathbf{P}\Big\{\frac{B(h) - B(0)}{\sigma_n} - \frac{\sqrt{n}\mu_n}{\sigma_n} \in (a,b)\Big\} \tag{3.4}$$

$$= \lim_{n\to\infty} \mathbf{P}\Big\{B(h) - B(0) \in (a\sigma_n + \sqrt{n}\mu_n, b\sigma_n + \sqrt{n}\mu_n)\Big\}.$$

This can only hold true if (σ_n) and $(\sqrt{n}\mu_n)$ stay bounded. Then we can pick convergent subsequence with limits $\overline{\sigma}$ and $\overline{\mu}$ and infer that the distributions of $B(h)-B(0)$ and $\overline{\sigma}X+\overline{\mu}$ coincide for a standard normally distributed random variable X. As $\overline{\sigma}X+\overline{\mu}$ is normally distributed with expectation $\overline{\mu}$ and variance $\overline{\sigma}^2$ the proof is finished. ■

We can now **finish the proof of Theorem 3.2** by showing the special structure of the expectations and variances.

Step 6: There is μ and $\sigma \geq 0$ such that, for all $t, h \geq 0$, $\mathbf{E}(B(t+h) - B(t)) = h\mu$ and $\mathrm{Var}(B(t+h) - B(t)) = h\sigma^2$.

Proof: By stationarity it suffices to show this for $t = 0$. Define the function $f, g : [0,\infty) \to \mathbf{R}$ by $f(h) = \mathrm{Var}(B(h)-B(0))$ and $g(h) = \mathbf{E}(B(h)-B(0))$. For all $h, k \geq 0$ we have by stationarity,

$$g(h+k) = \mathbf{E}\Big(B(h+k) - B(h)\Big) + \mathbf{E}\Big(B(h) - B(0)\Big) = g(k) + g(h).$$

and, by Bienaymé's equality and stationarity,

$$f(h+k) = \mathrm{Var}(B(h+k) - B(0)) = \mathrm{Var}(B(h+k) - B(k)) + \mathrm{Var}(B(k) - B(0)) = f(h) + f(k).$$

Hence there is a μ with $g(h) = \mu h$ and a $\sigma^2 \geq 0$ such that $f(h) = \sigma^2 h$. ■

We finish this section with an example, which shows that there are other processes, which have stationary, independent increments but fail to have continuous paths.

Example Consider $X(t)$ the number of customers arriving at a store by time $t \geq 0$. Intuitively the process $\{X(t) : t \geq 0\}$ has to satisfy the following assumptions

- The number of customers arriving during one time interval does not affect the number of customers arriving in another disjoint time interval. Mathematically, this means that the process has independent increments.
- The rate at which customers arrive should be constant, more precisely, there is some $\lambda \geq 0$ such that $\mathbf{E}[X(t)] = \lambda t$.
- Customers arrive one at a time. To make this precise we assume that $X(t)$ is increasing, takes values in $\mathbb{N}$ and we have

$$\begin{aligned} \mathbf{P}\{X(t+h) = X(t) + 1\} &= \lambda h + o(h), \\ \mathbf{P}\{X(t+h) \geq X(t) + 2\} &= o(h). \end{aligned}$$

Brownian motion satisfies the first two, but not the last assumption. A stochastic process fulfilling these assumptions is called a *Poisson process with rate (or intensity)* λ. We shall see as an exercise that it is uniquely determined up to equivalence. Here is one way to construct it: Let S be a Poisson distributed random variable with parameter λ and $Y_1, Y_2, Y_3, \ldots$ independent random variables with uniform distribution on $[0,1)$. For $0 \leq t \leq 1$ let

$$X(t) = \#\{Y_i : Y_i \leq t \text{ and } i \leq S\}.$$

Then X satisfies the assumptions on the interval $[0,1)$ and we extend X to $[0,\infty)$ by glueing together independent copies of X.

3.3 Gaussian random variables and processes

In this section we prepare the existence proof of Brownian motion with some lemmas about Gaussian random variables. This and the following two sections are essentially taken from Peres (1998).

We have seen that the normal distribution comes up naturally in the study of stochastic processes. One of the drawbacks of this distribution is that its distribution function can not be expressed in terms of classical functions. Therefore the following (quite precise) estimate will later be useful.

Lemma 3.4 *Suppose X is standard normally distributed. Then, for all $x \geq 0$,*

$$\frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \leq \mathsf{P}\{X > x\} \leq \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Proof: The right inequality is obtained by the estimate

$$\mathsf{P}\{X > x\} \leq \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{u}{x}e^{-u^2/2}\,du = \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,.$$

For the left inequality we define

$$f(x) = xe^{-x^2/2} - (x^2+1)\int_x^\infty e^{-u^2/2}\,du\,.$$

Remark that $f(0) < 0$ and $\lim_{x\to\infty} f(x) = 0$. Moreover,

$$f'(x) = (1 - x^2 + x^2 + 1)e^{-x^2/2} - 2x\int_x^\infty e^{-u^2/2}\,du = -2x\Big(\int_x^\infty e^{-u^2/2}\,du - \frac{e^{-x^2/2}}{x}\Big)\,,$$

which is positive for $x \geq 0$, by the first part. Hence $f(x) \leq 0$, proving the lemma. ■

We now look more closely at random vectors with normally distributed components. Our motivation is that they arise, for example, as vectors consisting of the increments of a Brownian motion. Let us clarify some terminology.

Definition: A random variable $X = (X_1, \ldots, X_d)^T$ with values in $\mathbf{R}^d$ has the *d-dimensional standard Gaussian distribution* if its d coordinates are standard normally distributed and independent (independent whether we write it as row– or column). A random variable Y with values in $\mathbf{R}^n$ is called *Gaussian* if there exists an $n \times d$ matrix A and an n dimensional vector b such that $Y^T = AX + b$. The *covariance matrix* of the vector Y is given by

$$\mathrm{Cov}(Y) = \mathbf{E}\Big[(Y - \mathbf{E}Y)(Y - \mathbf{E}Y)^T\Big] = AA^T\,,$$

here the expectations are defined componentwise.

Our first lemma shows that applying an orthogonal $d\times d$ matrix does not change the distribution of a standard Gaussian random vector.

Lemma 3.5 *If A is an orthogonal $d \times d$ matrix, i.e. $AA^T = I_d$, and X is a d-dimensional standard Gaussian vector, then AX is also a d-dimensional standard Gaussian vector.*

Proof: As the coordinates of X are independent, standard normally distributed, X has a density

$$f(x_1, \dots, x_d) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{d/2}} e^{-\|x\|^2/2},$$

where $\|\cdot\|$ is the Euclidean norm. The density of AX is (by the transformation rule) $f(A^T x)\sqrt{AA^T}$. The determinant is 1 and hence, since orthogonal matrices preserve the Euclidean norm, the density of X is invariant under A. ■

Corollary 3.6 *Let X_1 and X_2 be independent and normally distributed with expectation 0 and variance $\sigma^2 > 0$. Then $X_1 + X_2$ and $X_1 - X_2$ are independent and normally distributed with expectation 0 and variance $2\sigma^2$.*

Proof: $(X_1/\sigma, X_2/\sigma)^T$ is standard Gaussian by assumption. Look at

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

This is an orthogonal matrix and applying it to our vector yields $((X_1 + X_2)/(\sqrt{2}\sigma), (X_1 - X_2)/(\sqrt{2}\sigma))$, which thus must have independent standard normal coordinates. ■

The next lemma shows that the distribution of a Gaussian random vector is determined by its expectation and covariance.

Lemma 3.7 *If X and Y are d-dimensional Gaussian vectors with $\mathbf{E}X = \mathbf{E}Y$ and $\mathrm{Cov}(X) = \mathrm{Cov}(Y)$, then X and Y have the same distribution.*

Proof: It is sufficient to consider the case $\mathbf{E}X = \mathbf{E}Y = 0$. By definition, there are standard Gaussian random vectors X_1 and X_2 and matrices A and B with $X = AX_1$ and $Y = BX_2$. By adding columns of zeroes to A or B, if necessary, we can assume that X_1 and X_2 are both k-vectors and A, B are both $d \times k$ matrices. Let $\mathcal{A}$ and $\mathcal{B}$ be the vector subspaces of $\mathbb{R}^k$ generated by the row vectors of A resp. B. To simplify notation assume that the first l row vectors of A form a basis of $\mathcal{A}$. Define the linear map $L : \mathcal{A} \to \mathcal{B}$ by

$$L(A_i) = B_i \text{ for } i = 1, \dots, l.$$

Here A_i is the ith row vector of A. Our aim is to show that L is an orthogonal isomorphism and then use the previous lemma. Let us first show that L is an isomorphism. Our covariance assumption gives that $AA^T = BB^T$. Assume there is a vector $v_1 A_1 + \dots v_l A_l$ whose image is 0. Then the d-vector

$$v = (v_1, \dots, v_l, 0, \dots, 0)$$

satisfies $vB = 0$. Hence

$$\|vA\|^2 = vAA^T v^T = vBB^T v^T = 0.$$

We conclude that $vA = 0$. Hence L is injective and $\dim \mathcal{A} \leq \dim \mathcal{B}$. Interchanging the roles of A and B gives that L is an isomorphism. As the entry (i, j) of $AA^T = BB^T$ is the scalar product of A_i and A_j as well as B_i and B_j, the mapping L is orthogonal. We can extend it on the orthocomplement of $\mathcal{A}$ to an orthogonal map $L : \mathbb{R}^k \to \mathbb{R}^k$ (resp. orthogonal $k \times k$-matrix). Then $X = AX_1$ and $Y = BX_2 = ALX_2$. As LX_2 is standard Gaussian, by the previous lemma, X and Y have the same distribution . ■

Corollary 3.8 *A Gaussian random vector X has independent entries if and only if its covariance matrix is diagonal.*

We end this section with the definition of an important class of processes.

Definition
A stochastic process $\{Y(t) : t \in I\}$ is called **a Gaussian process**, if for all $t_1 \le t_2 \le \dots t_n$ the vector $(Y(t_1), \dots, Y(t_n))$ is a Gaussian random vector.

Show as an EXERCISE that every Brownian motion is a Gaussian process, i.e. given times $t_1 \le \dots \le t_n$ find a matrix A and a vector b such for a standard Gaussian vector X, we have

$$(B(t_1), \dots, B(t_n)) = AX + b\,.$$

3.4 Existence of Brownian motion and Wiener measure

The central result of this section is the following theorem, which establishes the existence of Brownian motions. Recall that we need only construct a standard Brownian motion B, as $X(t) = x + \sigma B(t) + \mu t$ is a Brownian motion with arbitrary starting point, drift and diffusion constant.

Theorem 3.9 (Wiener 1923) *Standard Brownian motion exists.*

We **prove** this using a construction of Paul Lévy (1948). We will use the existence theorem for countable products. We first construct Brownian motion on the interval $[0,1]$ as a random element on the space $\mathcal{C}[0,1]$ of continuous functions. The idea is to construct the right values of Brownian motion step by step on the finite sets

$$\mathcal{D}_n = \left\{ \frac{k}{2^n} : 0 \le k \le 2^n \right\}$$

of dyadic points. If the values on $\mathcal{D}_n$ are constructed we interpolate them linearly and later we define Brownian motion as the uniform limit of these continuous functions, which is automatically continuous.

To do this let $\mathcal{D} = \bigcup_{n=0}^{\infty} \mathcal{D}_n$ and let $(\Omega, \mathcal{A}, \mathbb{P})$ be a countable product space such that a collection $\{Z_d : d \in \mathcal{D}\}$ of independent, standard normally distributed random variables can be defined on the space. Let $B(0) = 0$ and $B(1) = Z_1$. For each $n \in \mathbb{N}$ we construct random variables $B(d)$, $d \in \mathcal{D}_n$ such that

- for all $r < s < t$ in $\mathcal{D}_n$ the random variable $B(t) - B(s)$ has $\mathcal{N}(0, t-s)$-distribution and is independent of $B(s) - B(r)$,

- the vectors $(B(d) : d \in \mathcal{D}_n)$ and $(Z_d : d \in \mathcal{D} \setminus \mathcal{D}_n)$ are independent.

Note that we have already done this for $\mathcal{D}_0 = \{0,1\}$. If we have succeeded in doing it for some $n-1$, we proceed by defining the $B(d)$ for $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ by

$$B(d) = \frac{B(d-2^{-n}) + B(d+2^{-n})}{2} + \frac{Z_d}{2^{(n-1)/2}}\,.$$

The values of B used in this definition are both in $\mathcal{D}_{n-1}$ and the second property is clearly fulfilled. Since $(1/2)[B(d+2^{-n}) - B(d-2^{-n})]$ is (by induction) normally distributed with expectation 0 and variance $1/2^{n+1}$ and $Z_d/2^{(n+1)/2}$ is independent with the same distribution, their sum $B(d) - B(d-2^{-n})$ and their difference $B(d+2^{n}) - B(d)$ are independent with the same distribution by Corollary 3.6. Now if $d \in \mathcal{D}_j$ for some $1 \le j \le n$ every increment of length 2^{-n} with right endpoint a dyadic point in $(d-2^{-j}, d]$ is independent from all increments of length 2^{-n} with left endpoint a dyadic point in $[d, d+2^{-j})$, by construction and independence of the Z_d for different d. The first of our properties follows from this.

Having thus chosen the values of the process on all dyadic points, we interpolate between them. Formally, define

$$F_0(x) = \begin{cases} Z_1 & \text{for } x = 1 \\ 0 & \text{for } x = 0 \\ \text{linear} & \text{in between.} \end{cases}$$

and, for each $n \ge 0$,

$$F_n(x) = \begin{cases} 2^{-(n+1)/2} Z_x & \text{for } x \in \mathcal{D}_n \setminus \mathcal{D}_{n-1} \\ 0 & \text{for } x \in \mathcal{D}_{n-1} \\ \text{linear} & \text{between consecutive points in } \mathcal{D}_n. \end{cases}$$

These functions are continuous on $[0,1]$ and for all n and $d \in \mathcal{D}_n$

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d). \tag{3.5}$$

This can be seen by induction. It holds for $n = 0$. Suppose that it holds for $n-1$. Let $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$. Since for $0 \le i \le n-1$ the function F_i is linear on $[d-2^{-n}, d+2^{-n}]$, we get

$$\sum_{i=0}^{n-1} F_i(d) = \sum_{i=1}^{n-1} \frac{F_i(d-2^{-n}) + F_i(d+2^{-n})}{2} = \frac{B(d-2^{-n}) + B(d+2^{-n})}{2}.$$

Since $F_n(d) = 1/2^{(n-1)/2} Z_d$, this gives (3.5).

On the other hand, we have, by definition of Z_d and by Lemma 3.4, for large n,

$$\mathbf{P}\{|Z_d| \ge c\sqrt{n}\} \le \exp\left(\frac{-c^2 n}{2}\right),$$

so that the series

$$\sum_{n=0}^{\infty} \mathbf{P}\{\exists d \in \mathcal{D}_n \text{ with } |Z_d| \ge c\sqrt{n}\} \le \sum_{n=0}^{\infty} \sum_{d \in \mathcal{D}_n} \mathbf{P}\{|Z_d| \ge c\sqrt{n}\}$$

converges as soon as $c > \sqrt{2 \log 2}$. Fix such a c. By the Borel-Cantelli Lemma there exists a random (but finite) N such that for all $n \geq N$ and $d \in \mathcal{D}_n$ we have $|Z_d| < c\sqrt{n}$. Hence,

$$\|F_n\|_\infty < c\sqrt{n} 2^{-n/2}\,. \tag{3.6}$$

This upper bound implies that the series

$$B(t) = \sum_{n=0}^{\infty} F_n(t)$$

is uniformly convergent on $[0,1]$ and we denote the continuous limit by $\{B(t) : t \in [0,1]\}$.

It remains to check that the increments of this process have the right finite-dimensional joint distributions. This follows directly from the properties of B on the dense set $\mathcal{D} \subset [0,1]$ and the continuity of the paths. Indeed, suppose that $t_1 > t_2 > t_3$ are in $[0,1]$. We find sequences $t_{i,n}$ in $\mathcal{D}$ converging to t_i and infer from the continuity of B that

$$B(t_3) - B(t_2) = \lim_{n\to\infty} B(t_{3,n}) - B(t_{2,n})\,.$$

As a limit of normally distributed random variables, this increment is itself normally distributed with mean 0 and variance

$$\lim_{n\to\infty} t_{3,n} - t_{2,n} = t_3 - t_2\,.$$

The analogous fact holds for $B(t_2) - B(t_1)$. Moreover, the random variables entering in the limits are independent if n is large enough that $t_{1,n} > t_{2,n} > t_{3,n}$. Hence, the increments must be independent, too. From Corollary 3.8 we infer that the process has independent increments on $[0,1]$.

We have thus constructed a Brownian motion B on $[0,1]$. By the existence of product spaces there exists a probability space on which a sequence $B_1, B_2, \ldots$ of independent $\mathcal{C}[0,1]$-valued random variables with the properties of our B can be defined. We glue them together by letting

$$B(t) = B_{[t]}(t - [t]) + \sum_{i=0}^{[t]-1} B_i(1)\,.$$

This defines a continuous random function in $\mathcal{C}[0,\infty)$ and one can see easily from what we have shown so far that the requirements of a standard Brownian motion are fulfilled.

We end this chapter by noting that our procedure defines a natural probability measure $\mathbf{W}$ on the space $\{f \in \mathcal{C}[0,\infty) : f(0) = 0\}$ of continuous real valued functions starting in 0. This probability measure is called the *Wiener measure.*

3.5 Basic path properties of Brownian motion

The Brownian motion has two very useful invariance properties. The first of them is the *scaling invariance*, the second is the invariance under *time-inversion.* In each case there is a transformation on the space of functions, which changes the individual Brownian random functions but leaves their distribution unchanged.

Lemma 3.10 (Scaling invariance) *Suppose* $\{B(t) : t \geq 0\}$ *is a standard Brownian motion and* $a > 0$*. Then the process* $\{X(t) : t \geq 0\}$ *defined by*

$$X(t) = \frac{B(a^2 t)}{a}$$

is also a standard Brownian motion.

Proof: Continuity of the paths, independence and stationarity of the increments remain unchanged under the rescaling. It remains to observe that

$$X(t) - X(s) = \frac{1}{a}\Big(B(a^2 t) - B(a^2 s)\Big)$$

is normally distributed with expectation 0 and variance $(1/a^2)(a^2 t - a^2 s) = t - s$. ■

Lemma 3.11 (Time inversion) *Suppose* $\{B(t) : t \geq 0\}$ *is a standard Brownian motion. Then the process* $\{X(t) : t \geq 0\}$ *defined by*

$$X(t) = \begin{cases} 0 & t = 0 \\ tB(1/t) & t > 0 \end{cases}$$

is also a standard Brownian motion.

Proof: Like Brownian motion $\{X(t)\}$ is a Gaussian process such that the Gaussian random vectors $(X(t_1), \ldots, X(t_n))$ have expectation 0. The covariances, for $t, h \geq 0$, are given by

$$\mathrm{Cov}(X(t+h), X(t)) = (t+h)t\,\mathrm{Cov}(B(1/t+h), B(1/t)) = t(t+h)\frac{1}{t+h} = t\,.$$

Hence X is a variant of Brownian motion. Its paths are clearly continuous for all $t > 0$ and in $t = 0$ we use the following two facts: First, the distribution of X on the rationals Q is the same as for a Brownian motion, hence

$$\lim_{t \to 0, t \in Q} X(t) = 0 \text{ almost surely.}$$

And second, X is almost surely continuous on $(0, \infty)$, so that

$$\lim_{t \to 0, t \in Q} X(t) = \lim_{t \to 0} X(t)\,.$$

■

Corollary 3.12 (Law of large numbers) *Almost surely,* $\lim_{t\to\infty} B(t)/t = 0$.

Proof: Using the time-inversion we see that $\lim_{t\to\infty} B(t)/t = \lim_{t\to\infty} X(1/t) = X(0) = 0$. ■

We now prove two theorems that make the degree of continuity of the paths of Brownian motion more precise. Pay attention to the order of the *almost surely* and the *for each* in the following theorem and note that a change of this order would give a (correct, but) much weaker statement.

Theorem 3.13 *There exists a random variable C such that, almost surely, for each $0 \le t \le t+h \le 1$,*

$$\Big|B(t+h) - B(t)\Big| \le C\sqrt{h\log(1/h)}\,.$$

Proof: We go back to the construction of the Brownian motion. We have represented Brownian motion as a series

$$B(t) = \sum_{n=0}^{\infty} F_n(t)\,,$$

where each F_n is a piecewise linear function. Its derivative exists almost everywhere, and by definition and (3.6),

$$\|F_n'\|_\infty \le \frac{2\|F_n\|_\infty}{2^{-n}} \le C_1(\omega) + 2c\sqrt{n}2^{n/2}\,.$$

The random constant C_1 is here to deal with the finitely many exceptions to (3.6). Now for each $t, t+h \in [0,1]$,

$$|B(t+h) - B(t)| \le \sum_{n=0}^{\infty} |F_n(t+h) - F_n(t)| \le \sum_{n=0}^{l} h\|F_n'\|_\infty + \sum_{n=l+1}^{\infty} 2\|F_n\|_\infty\,.$$

Hence, using (3.6) again, if $l > N$ for a random N, this is bounded by

$$h\sum_{n=0}^{l}\Big(C_1(\omega) + 2c\sqrt{n}2^{n/2}\Big) + 2\sum_{n=l+1}^{\infty} c\sqrt{n}2^{-n/2} \le C_2(\omega)h\sqrt{l}2^{l/2} + C_3(\omega)\sqrt{l}2^{-l/2}.$$

Choosing $l = \log_2(1/h)$ and $C(\omega)$ sufficiently large to take care of the cases $l \le N$ we get

$$\Big|B(t+h) - B(t)\Big| \le C(\omega)\sqrt{h\log(1/h)}\,.$$

∎

This theorem is sharp in the sense that the function $\sqrt{h\log(1/h)}$ cannot be replaced by any function which decreases faster as $h \downarrow 0$.

Theorem 3.14 *There exists a constant $c > 0$ such that, almost surely, for every $\varepsilon > 0$ there exist $t \ge 0$ and $0 < h < \varepsilon$ with*

$$B(t+h) - B(t) > c\sqrt{h\log(1/h)}\,.$$

Proof: Let $c < \sqrt{2\log 2}$ and define the events

$$A_{k,n} = \Big\{B((k+1)2^{-n}) - B(k2^{-n}) > c\sqrt{n}2^{-n/2}\Big\}.$$

Then, using Lemma 3.4,

$$\mathbf{P}(A_{k,n}) = \mathbf{P}\{B(2^{-n})) > c\sqrt{n}2^{-n/2}\} = \mathbf{P}\{B(1) > c\sqrt{n}\} \ge \frac{c\sqrt{n}}{c^2n+1}e^{-c^2n/2}\,.$$

By our assumption on c, we have

$$2^n \mathsf{P}(A_{k,n}) \to \infty\,.$$

Therefore, using $1-x \le e^{-x}$ for all x,

$$\mathsf{P}\Big(\bigcap_{k=0}^{2^n-1} A_{k,n}^c\Big) = (1-\mathsf{P}(A_{k,n}))^{2^n} \le \exp(-2^n \mathsf{P}(A_{k,n})) \to 0\,.$$

by considering $h = 2^{-n}$ one can now see that

$$\mathbb{P}\Big\{\forall h < \varepsilon \forall t \text{ with } h+t \le 1: \quad B(t+h)-B(t) \le c\sqrt{h \log_2(1/h)}\Big\} = 0\,.$$

■

For Hölder-continuity of Brownian motion we have the following consequence.

Corollary 3.15 *For every $\alpha < 1/2$ Brownian motion is almost surely α-Hölder continuous, but not $1/2$-Hölder continuous.*

Proof: Observe that, for every $\alpha > 0$, there is $C_1 > 0$ such that $\sqrt{h\log(1/h)} < C_1 h^\alpha$ and, for every $C > 0$, and sufficiently small h we have $\sqrt{h \log(1/h)} > C\sqrt{h}$ ■

We shall come back to a longer discussion of the properties of Brownian motions later in the lecture, in particular we will then prove that Brownian motion is nowhere differentiable.

3.6 Higher dimensional Brownian motion

We define *(standard) d-dimensional Brownian motion* as the process $\{B(t) : t \ge 0\}$ defined by $B(t) = (B_1(t), \dots, B_d(t))$, where $B_1, \dots, B_d$ are d independent Brownian motions. Note that we are able to construct a finite product space, on which d independent Brownian motions can be defined. The d-dimensional Brownian motion is an $\mathbb{R}^d$-valued process with stationary, independent increments and almost surely continuous paths.

3.7 Geometric Brownian motion

In order to model a time homogenous stock price in the simplest manner one would require a process $\{X(t) : t \ge 0\}$ with the following features:

- the process X is almost surely continuous and has positive values,
- for all $0 \le t_1 \le t_2 \le \dots \le t_n$ the *returns* $X(t_n)/X(t_{n-1}), X(t_{n-1})/X(t_{n-2}), \dots, X(t_2)/X(t_1)$ are independent.
- the distribution of the returns $X(t+h)/X(t)$ does not depend on t.

These assumption already determine the model up to two constants:

Theorem 3.16 *Suppose that the process $\{X(t) : t \geq 0\}$ satisfies the assumptions above and $X(0) = x > 0$ is the initial price. Then there is a μ and a σ such that $\log X(t)$ is a Brownian motion with start in $\log x$, drift μ and variance σ^2. The process X is called geometric Brownian motion with trend parameter μ and volatility parameter σ.*

Proof: $\log X$ satisfies the conditions of the Characterization Theorem for Brownian motion. ■

Of course, this model is not the final word about modelling a stock price. For example, in practice, trend and volatility parameters will depend on t. The definition of such more complicated processes has to be deferred to the second part of the lecture, when methods of stochastic analysis are available.

Chapter 4

The strong Markov property of Brownian motion

We start the discussion of the Markov property with a thorough introduction of the term *conditional expectation*, which is essential to the concepts of the Markov property and also of martingales.

4.1 Conditional expectations

Suppose that $(\Omega, \mathcal{A}, \mathbf{P})$ is a probability space and X and Z are random variables on this space. We assume that both random variables take on finitely many real values $\{x_1, \ldots, x_m\}$ resp. $\{z_1, \ldots, z_n\}$ each with positive probability. We suppose that X and Z are *not* independent. How does knowledge about the outcome of Z influence the outcome of X? This can be described by means of *conditional probabilities*

$$\mathbf{P}\Big\{X = x_i \,|\, Z = z_j\Big\} := \frac{\mathbf{P}\Big\{X = x_i \text{ and } Z = z_j\Big\}}{\mathbf{P}\{Z = z_j\}}.$$

One way to look at this is the following: if we have observed the event $\{Z = z_j\}$ this changes our perception of the random variable X, it is now defined *on a different probability space*, on which only those ω can occur, which satisfy $Z(\omega) = z_j$. The new space still consists of the set Ω and the σ-field $\mathcal{A}$, but the probability measure is now concentrated on the set

$$Z^{-1}(z_j) = \{\omega \in \Omega \,:\, Z(\omega) = z_j\}.$$

The new probability measure $\mathbf{P}\{\cdot|Z = z_j\}$ is given by

$$\mathbf{P}\Big\{A \,|\, Z = z_j\Big\} := \frac{\mathbf{P}\Big\{A \text{ and } Z = z_j\Big\}}{\mathbf{P}\{Z = z_j\}}.$$

The random variable X may still be defined on the space, but its distribution has changed and is now the conditional distribution given $Z = z_j$. Its expectation is now

$$\mathbf{E}\Big\{X \,|\, Z = z_j\Big\} := \sum_{k=1}^{m} x_k \mathbf{P}\Big\{X = x_k | Z = z_j\Big\}.$$

This expectation depends on z_j and can be interpreted as a mapping (or random variable) on Ω namely by

$$\mathbf{E}\Big\{X \mid Z\Big\} : \Omega \to \mathbf{R}\,, \qquad \mathbf{E}\Big\{X \mid Z\Big\}(\omega) = \mathbf{E}\Big\{X \mid Z = Z(\omega)\Big\}\,.$$

The mapping $\mathbf{E}\{X|Z\}$ has the following properties :

1) it is measurable with respect to the σ-field generated by the sets $Z^{-1}\{z_j\}$ for $j = 1, \ldots, n$.

2) for every $A \subset \{z_1, \ldots, z_n\}$ we have $\int_{Z^{-1}(A)} \mathbf{E}\{X|Z\}(\omega) d\mathbf{P}(\omega) = \int_{Z^{-1}(A)} X(\omega) d\mathbf{P}(\omega)$.

The first property just means that $\mathbf{E}\{X|Z\}$ is constant on each set $Z^{-1}\{z_j\}$ and the second property follows from the calculation

$$\int_{Z^{-1}(A)} \mathbf{E}\{X|Z\}(\omega)\, d\mathbf{P}(\omega) = \sum_{z\in A} \mathbf{E}\{X|Z = z\}\mathbf{P}\{Z = z\}$$

$$= \sum_{z\in A}\sum_{k=1}^{m} x_k \mathbf{P}\Big\{X = x_k | Z = z\Big\}\mathbf{P}\{Z = z\}\,,$$

the last term equals, by the definition of conditional probabilities,

$$\sum_{z\in A}\sum_{k=1}^{m} x_k \mathbf{P}\Big\{X = x_k \text{ and } Z = z\Big\} = \int \sum_{z\in A}\sum_{k=1}^{m} x_k \mathbb{1}_{\{X(\omega)=x_k\}}\mathbb{1}_{\{Z(\omega)=z\}}\, d\mathbf{P}(\omega)$$

$$= \int X(\omega) \sum_{k=1}^{m} \mathbb{1}_{\{X(\omega)=x_k\}} \sum_{z\in A} \mathbb{1}_{\{Z(\omega)=z\}}\, d\mathbf{P}(\omega) = \int_{Z^{-1}(A)} X(\omega)\, d\mathbf{P}(\omega)\,.$$

One could say that we have decomposed the probability space according to the *information coming from* Z, each probability measure $\mathbf{P}\{\cdot|Z = z_j\}$ is concentrated on a different region of the space Ω, because the additional information we had consisted of the knowledge of the cell in which the random ω was to be found.

What can we say if additional information is coming into our experiment from other sources than the observation of the values of a discrete random variable such as Z? Can we get a similar decomposition? How can we describe *information*?

Information from our point of view is a subcollection $\mathcal{B} \subset \mathcal{A}$ of events, about which we have already the knowledge whether they occur or not. Such a subcollection, by simple considerations is necessarily a σ-field. Hence we adopt the point of view that information can be modelled by means of (sub)-σ-fields $\mathcal{B} \subset \mathcal{A}$. Let us give some EXAMPLES:

- Suppose we have observed the value of a random variable $Z' : (\Omega, \mathcal{A}, \mathbf{P}) \to (\Omega', \mathcal{A}')$. Then we have knowledge about all the events $\{Z \in A\}$ if A runs through the σ-field $\mathcal{A}'$. These sets form a σ-field $Z^{-1}(\mathcal{A}')$. In the case that Z takes on only finitely many values (as before) the σ-field consists of all finite unions of the cells (or atoms) $Z^{-1}(z_j)$.

- Suppose we have observed the path of a stochastic process $\{X(t)\,:\, t \geq 0\}$ up to time T. The information gained by this is encoded in the σ-field $\mathcal{F}_T$ generated by all the increments $X(t) - X(s)$ for $s \leq t \leq T$, equivalently generated by

$$\{X(t_n) - X(t_{n-1}) \in A_n, \ldots, X(t_1) - X(t_0) \in A_1\}$$

for all A_i Borel and $t_0 \le t_1 \le \ldots \le t_n \le T$. Increasing the T gives us a whole nested sequence of σ-fields $\mathcal{F}_T, T \ge 0$, which model the increase of knowledge when we observe a longer and longer part of the process.

We now know what we need: an extension of the notion of conditional expectation from finitely valued random variables to σ-fields. Observing that we formulated our properties of the conditional expectation in the special case in such a way that we can use them to formulate a new definition.

Theorem 4.1 *Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ a sub-σ-field and X a random variable with $\mathbf{E}|X| < \infty$. Then there is a random variable $Y = \mathbf{E}\{X|\mathcal{F}\}$ with $\mathbf{E}|Y| < \infty$ called* the conditional expectation of X given $\mathcal{F}$ *with the following two properties:*

1) $\mathbf{E}\{X|\mathcal{F}\}$ is $\mathcal{F}$-measurable,

2) for all $F \in \mathcal{F}$, $\int_F \mathbf{E}\{X|\mathcal{F}\}\, d\mathbf{P} = \int_F X\, d\mathbf{P}$.

Any two random variables Y satisfying these two conditions coincide almost surely. If $\mathcal{F} = Z^{-1}(\mathcal{A}')$ is the σ-field of preimages of all Borel sets under a random variable Z, one also writes $\mathbf{E}\{X|Z\}$ for $\mathbf{E}\{X|\mathcal{F}\}$ and says this is the conditional expectation of X given Z.

Remark: By chosing $F = \Omega$ in the last property, we see that

$$\mathbf{E}\Big\{\mathbf{E}\{X|\mathcal{F}\}\Big\} = \mathbf{E}X\,.$$

We start with the **proof of the uniqueness**. Suppose Y_1 and Y_2 both satisfy the definition of a conditional expectation. Suppose further that $\mathbf{P}\{Y_1 > Y_2\} > 0$. Then there is a natural number n such that $\mathbf{P}\{Y_1 - Y_2 > 1/n\} > 1/n$. Note that the event $F = \{Y_1 - Y_2 > 1/n\}$ is in $\mathcal{F}$. Hence

$$1/n^2 \le \int_F (Y_1 - Y_2)\, d\mathbf{P} = \int_F X\, d\mathbf{P} - \int_F X\, d\mathbf{P} = 0\,,$$

a contradiction. Therefore $\mathbf{P}\{Y_1 > Y_2\} = 0$. Exchanging the role of Y_1 and Y_2 yields $\mathbf{P}\{Y_1 = Y_2\} = 1$. ■

Before we prove existence, we formulate some properties of conditional expectations. Note first that uniqueness implies the following property called *linearity*: For all a, b real, almost surely,

$$\mathbf{E}\Big\{aX + bY|\mathcal{F}\} = a\mathbf{E}\{X|\mathcal{F}\} + b\mathbf{E}\{Y|\mathcal{F}\}\,.$$

Pay attention to the order of the quantors and of the almost surely. The following theorem is to be proved as an EXERCISE. It describes the convergence properties of conditional expectations.

Theorem 4.2 *Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ a sub-σ-field and X a random variable with $\mathbf{E}|X| < \infty$. Then every conditional probability $\mathbf{E}\{X|\mathcal{F}\}$ has the following properties.*

Positivity *If $X \ge 0$, then $\mathbf{E}\{X|\mathcal{F}\} \ge 0$ almost surely.*

Monotone Convergence *If* $0 \le X_n \uparrow X$*, then* $\mathbf{E}\{X_n|\mathcal{F}\} \uparrow \mathbf{E}\{X|\mathcal{F}\}$ *almost surely.*

Fatou *If* $0 \le X_n$ *and* $\mathbf{E}\{X_n|\mathcal{F}\} < \infty$*, then*

$$\mathbf{E}\{\liminf_{n\to\infty} X_n \,|\, \mathcal{F}\} \le \liminf_{n\to\infty} \mathbf{E}\{X_n \,|\, \mathcal{F}\} \text{ almost surely.}$$

Dominated Convergence *If there is a random variable* Z *such that* $\mathbf{E}Z < \infty$ *and* $|X_n| \le Z$ *for all* n*, and if* $X_n \to X$ *almost surely, then* $\mathbf{E}\{X_n|\mathcal{F}\} \to \mathbf{E}\{X|\mathcal{F}\}$*.*

The next set of properties is of more probabilistic nature. If $\mathcal{F}$ and $\mathcal{G}$ are σ-fields we denote by $\mathcal{F} \vee \mathcal{G}$ the σ-field generated by their union.

Theorem 4.3 *Let* $(\Omega, \mathcal{A}, \mathbf{P})$ *be a probability space,* $\mathcal{F} \subset \mathcal{A}$ *a sub-*σ*-field and* X *a random variable with* $\mathbf{E}|X| < \infty$*. Then every conditional probability* $\mathbf{E}\{X|\mathcal{F}\}$ *has the following properties.*

Tower property. *If* $\mathcal{G} \subset \mathcal{F}$ *is a sub-*σ*-field, then* $\mathbf{E}\Big\{\mathbf{E}\{X|\mathcal{F}\}\Big|\mathcal{G}\Big\} = \mathbf{E}\{X|\mathcal{G}\}$ *almost surely.*

Taking out what is known. *If* Z *is* $\mathcal{F}$*-measurable and bounded, then* $\mathbf{E}\{ZX|\mathcal{F}\} = Z\mathbf{E}\{X|\mathcal{F}\}$ *almost surely. If* X *itself is* $\mathcal{F}$*-measurable, then* $\mathbf{E}\{X|\mathcal{F}\} = X$ *almost surely.*

Independence. *If* X *is independent of* $\mathcal{F}$*, then* $\mathbf{E}\{X \,|\, \mathcal{F}\} = \mathbf{E}\{X\}$ *almost surely.*

Independence (Advanced case). *Suppose that* $\mathcal{G}, \mathcal{H} \subset \mathcal{A}$ *are sub-*σ*-fields, such that* $\mathcal{H}$ *and* $\mathcal{G} \vee X^{-1}(\mathcal{A}')$ *are independent. Then, almost surely,*

$$\mathbf{E}\{X|\mathcal{G} \vee \mathcal{H}\} = \mathbf{E}\{X|\mathcal{G}\}\,.$$

Proof: By linearity we can assume that $X \ge 0$ in all parts of the proof. For the tower property one first observes that the left hand side is $\mathcal{G}$ measurable and, for every $G \in \mathcal{G} \subset \mathcal{F}$,

$$\int_G \mathbf{E}\Big\{\mathbf{E}\{X|\mathcal{F}\}\Big|\mathcal{G}\Big\} d\mathbf{P} = \int_G \mathbf{E}\{X|\mathcal{F}\} d\mathbf{P} = \int_G X \, d\mathbf{P}\,.$$

Hence the integrand on the left hand side satisfies all the properties of conditional expectation $\mathbf{E}\{X|\mathcal{G}\}$ and must be a conditional expectation of X given $\mathcal{G}$.

To see the second property, we observe that $Z\mathbf{E}\{X|\mathcal{F}\}$ is $\mathcal{F}$-measurable and, because Z is bounded, say by $C > 0$, satisfies $\mathbf{E}|Z\mathbf{E}\{X|\mathcal{F}\}| \le C\mathbf{E}|\mathbf{E}\{X|\mathcal{F}\}| < \infty$. We just have to show, for every $F \in \mathcal{F}$,

$$\int_F ZX \, d\mathbf{P} = \int_F Z\mathbf{E}\{X|\mathcal{F}\} \, d\mathbf{P}\,,$$

then it follows that $Z\mathbf{E}\{X|\mathcal{F}\}$ is a conditional expectation of ZX given $\mathcal{F}$. The given equality holds for $Z = 1_H$, $H \in \mathcal{F}$, by definition and follows for bounded $\mathcal{F}$-measurable Z by the standard measure theory machinery. The particular cases mentioned at the end holds, because in the given situation, X itself satifies the conditions of a conditional expectation of X given $\mathcal{F}$.

For the last property observe that the constant function $\mathbf{E}\{X\}$ is $\mathcal{F}$-measurable and we just have to show, for all $F \in \mathcal{F}$,

$$\int_F X(\omega) \, d\mathbf{P}(\omega) = \mathbf{P}(F)\mathbf{E}\{X\}\,.$$

We show the more general statement that, for all $h : \mathbb{R} \to [0, \infty)$,

$$\int_F h(X(\omega))\, d\mathbf{P}(\omega) = \mathbf{P}(F)\mathbf{E}\{h(X)\}\,.$$

Note that, if $h = 1_A$, the left hand side is $\mathbf{P}(F \cap \{X \in A\})$ and the right hand side is $\mathbf{P}(F)\mathbf{P}\{X \in A\}$, so that our statement is the definition of independence of the events $\{X \in A\}$ and F. By the standard measure theory machinery this can be extended to arbitrary h and plugging in $h(x) = x$ gives the desired result. The last property is similar and an EXERCISE. ■

The **proof of the existence** illustrates that $\mathbf{E}\{X|\mathcal{F}\}$ is in a sense a projection of X on the space of $\mathcal{F}$-measurable random variables. Our strategy is to show this first for random variables X satisfying $\mathbf{E}|X|^2 < \infty$. The idea is that then X is an element of the Hilbert space

$$\mathcal{L}^2(\Omega, \mathcal{A}, \mathbf{P}) = \Big\{X : (\Omega, \mathcal{A}) \to \mathbf{R}, \qquad \mathcal{A}\text{-measurable with } \mathbf{E}|X|^2 < \infty\Big\}$$

of square integrable functions on the probability space with two functions X and Y identified if $\mathbf{P}\{X = Y\} = 1$. The scalar product on this space is

$$\langle X, Y\rangle = \mathbf{E}\Big\{XY\Big\}\,.$$

Hence the norm is $\|X\|^2 = \mathbf{E}|X|^2$. If you do not know that this defines a Hilbert space, then see the EXERCISES, where the completeness of this space is proved, by checking that every Cauchy sequence converges:

Lemma (Completeness): If $\{X_n\}$ is a sequence of random variables on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ such that $\mathbf{E}|X_n|^2 < \infty$ for all n and $\lim_{k\to\infty} \sup_{n,m\geq k} \mathbf{E}|X_n - X_m|^2 = 0$, then there is a random variable X on $(\Omega, \mathcal{A}, \mathbf{P})$ such that $\lim_{n\to\infty} \mathbf{E}|X_n - X|^2 = 0$.

Hilbert spaces are the mathematical structure that allows to make sense of the notion of projections. For our purpose we need the following.

Lemma 4.4 *Suppose that $V \subset \mathcal{L}^2(\Omega, \mathcal{A}, \mathbf{P})$ is a complete linear subspace and $X \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbf{P})$. Then there exists a $Y \in V$, called the* projection of X on V *such that*

- $\|X - Y\| = I := \inf\{\|X - W\| : W \in V\}$,
- $\langle X - Y, Z\rangle = 0$ *for all* $Z \in V$.

Remark: Y is called the projection of X onto Y and any two projections are equal almost surely.

Proof: Recall that *completeness* of V means that every Cauchy-sequence of random vectors Y_n in V converges in norm to a random variable $Y \in V$. We first choose a sequence $\{Y_n\}$ in V such that

$$\|X - Y_n\| \to I\,.$$

Using the scalar product one can see that

$$2\|\frac{1}{2}(Y_n - Y_m)\|^2 = \|X - Y_n\|^2 + \|X - Y_m\|^2 - 2\|X - \frac{1}{2}(Y_n + Y_m)\|^2\,.$$

Because $\frac{1}{2}(Y_n - Y_m) \in V$, the subtracted term is at least I^2 and hence $\{Y_n\}$ is a Cauchy sequence and converges to some Random variable $Y \in V$. By the triangle inequality

$$I \leq \|X - Y\| \leq \|X - Y_n\| + \|Y - Y_n\| \to I\,,$$

and thus $\|X - Y\| = I$. For every $Z \in V$ we have $Y + tZ \in$ for each $t \in \mathbf{R}$ and hence

$$\|X - Y - tZ\|^2 \geq \|X - Y\|^2\,,$$

which implies $-2t\langle Z, X - Y\rangle + t^2\|Z\|^2 \geq 0$. As a function of t this has a minimum in $t = 0$ and hence the derivative at 0 must vanish, which is the statement. ■

We now show the existence of conditional expectation first for a random variable $X \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbf{P})$. We project X onto the complete subspace

$$V := \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) \subset \mathcal{L}^2(\Omega, \mathcal{A}, \mathbf{P})\,,$$

and call the result Y. As $Y \in V$ it is $\mathcal{F}$-measurable. Moreover,

$$\langle X - Y, Z\rangle = 0 \text{ for all } Z \in V.$$

If $F \in \mathcal{F}$ we may choose $Z = 1_F \in V$ there and get

$$\int_F X\,d\mathbf{P} = \langle X, 1_F\rangle = \langle Y, 1_F\rangle = \int_F Y\,d\mathbf{P}\,.$$

Hence Y is a conditional expectation of X given $\mathcal{F}$. It remains to show the same if X just fulfills $\mathbf{E}|X| < \infty$. It suffices to consider the case $X \geq 0$. Choose $X_n = X \wedge n$ and observe that $X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbf{P})$ and there are conditional expectations $\mathbf{E}\{X_n|\mathcal{F}\}$. By Theorem 4.2 (Positivity) they are positive and increasing. Let

$$Y = \lim_{n\to\infty} \mathbf{E}\{X_n|\mathcal{F}\}\,.$$

Clearly, Y is $\mathcal{F}$-measurable and, as X_n increases to X, we can use monotone convergence to see that, for every $F \in \mathcal{F}$,

$$\int_F Y\,d\mathbf{P}(\omega) = \lim_{n\to\infty} \int_F \mathbf{E}\{X_n|\mathcal{F}\}\,d\mathbf{P}(\omega) = \lim_{n\to\infty} \int_F X_n\,d\mathbf{P}(\omega) = \int_F X\,d\mathbf{P}(\omega)\,.$$

This also implies $\mathbf{E}Y = \mathbf{E}X < \infty$ and hence Y is a conditional expectation of X given $\mathcal{F}$. ■

As an EXERCISE we investigate the situation when two real valued random variables X and Z have a joint density $f(x, z)$. Then, clearly, the density of X is $f_X : x \mapsto \int f(x, z)\,dz$ and the density of Z is $f_Z : z \mapsto \int f(x, z)\,dx$. Assume that

$$\mathbf{E}|X| = \int |x| f_X(x)\,dx = \int |x| \int f(x, z)\,dz\,dx < \infty\,.$$

One can define a function

$$f_{X|Z}(x|z) = \begin{cases} f(x, z)/f_Z(z) & \text{if } f_Z(z) \neq 0\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

With this function define a random variable

$$Y = \int x f_{X|Z}(x|Z)\,dx\,.$$

Then Y is a conditional expectation of X given Z.

4.2 The weak Markov property and Blumenthal's 01-Law

Suppose that $\{X(t) : t \geq 0\}$ is a stochastic process. Intuitively, the *Markov property* of a process says that if we know the process $\{X(t)\}$ on the interval $[0, s]$, this is as useful as just knowing the endpoint $X(s)$. We will see that Brownian motion is a process having this property.

We now work on the space $\Omega = \mathcal{C}([0, \infty))$ of continuous functions with the Borel σ-field $\mathcal{A}$ generated by all increments and the probability measures $\mathbf{P}_x$, which are the distributions of Brownian motion started in x. The corresponding expectations and conditional expectations are denoted $\mathbf{E}_x$ and $\mathbf{E}_x\{\cdot|\cdot\}$. On Ω we have for every $s > 0$ *shift transformations* $\theta_s, \Theta_s : \Omega \to \Omega$ defined by

$$\theta_s B(t) = B(s + t) \qquad \text{and} \qquad \Theta_s B(t) = B(s + t) - B(s)\,.$$

Note that one can consider θ_s and Θ_s as Ω-valued random variables and thus as stochastic processes by

$$\theta_s(t) = \theta_s(t)[B] = \theta_s B(t) = B(t + s)\,,$$

and analogously for Θ_s. Define the σ-field $\mathcal{F}^0(t)$ to be the σ-field generated by the increments $B(t_n) - B(t_{n-1}), \ldots, B(t_2) - B(t_1)$ for all $0 \leq t_1 \leq \cdots \leq t_n \leq t$. Note that

$$\mathcal{F}^0(t) \subset \mathcal{F}^0(s) \text{ for all } t \leq s\,.$$

A family $\{\mathcal{F}(t) : t \geq 0\}$ of σ-fields such that $\mathcal{F}(t) \subset \mathcal{F}(s)$ for all $t \leq s$ is called a *filtration.*

Theorem 4.5 (Weak Markov property) *For every $s \geq 0$ and every bounded random variable $Y : (\Omega, \mathcal{A}) \to \mathbf{R}$ we have*

$$\mathbf{E}_x\{Y \circ \theta_s \,|\, \mathcal{F}^0(s)\} = \mathbf{E}_x\{Y \circ \theta_s \,|\, B(s)\} = \mathbf{E}_{B(s)}\{Y\} \text{ almost surely,}$$

where the right hand side is the function $\varphi(x) = \mathbf{E}_x\{Y\}$ evaluated at $x = B(s)$. In particular, the process $\{B(s + t) - B(s) : t \geq 0\} = \{\Theta_s(t) : ,t \geq 0\}$ is a standard Brownian motion independent of $\mathcal{F}^0(s)$.

In order to prove the result in this generality we need one more measure theoretic tool, the *monotone class theorem.* This is an improvement of our standard measure theory machinery.

Lemma 4.6 (Monotone class theorem) *Let $\mathcal{B}$ be a $\cap$-stable system that contains Ω and generates the σ-field $\mathcal{A}$. Let $\mathcal{H}$ be a collection of real valued functions containg all indicators $1_A, A \in \mathcal{B}$ such that*

- *if $f, g \in \mathcal{H}$ then $f + g \in \mathcal{H}$ and $cf \in \mathcal{H}$ for all real c,*
- *if $f_n \in \mathcal{H}$ is nonnegative and increasing to a bounded function f, then $f \in \mathcal{H}$.*

Then $\mathcal{H}$ contains all bounded $\mathcal{A}$-measurable functions.

The **proof** of this can be found in the book of Durrett, Chapter 5, Theorem (1.5) or in most texts on measure theory.

Proof: The particular statement at the end of the theorem is evident from the independence of the increments of Brownian motion, because $\mathcal{F}^0(s)$ is generated by the increments up to time s and independence of $\mathcal{F}^0(s)$ from the given process starting in 0 is just independence from its increments. In the formal language of conditional expectations we get, for each Y of the special form $Y(B) = X(B(0))Z(\Theta_0 B)$, using $\Theta_0\theta_s = \Theta_s$,

$$\mathbf{E}_x\{Y \circ \theta_s \,|\, \mathcal{F}^0(s)\} = \mathbf{E}_x\{X(B(s))Z(\Theta_s B) \,|\, \mathcal{F}^0(s)\} = X(B(s))\mathbf{E}_x\{Z(\Theta_s B) \,|\, \mathcal{F}^0(s)\},$$

as we may take out what is known. By the particular statement, we have $\mathbf{E}_x\{Z(\Theta_s B) \,|\, \mathcal{F}^0(s)\} = \mathbf{E}_x\{Z(\Theta_s B)\} = \mathbf{E}_{B(s)}\{Z(\Theta_0 B)\}$ and hence, we may proceed the last chain with,

$$= X(B(s))\mathbf{E}_{B(s)}\{Z(\Theta_0 B)\} = \mathbf{E}_{B(s)}\{Y\}.$$

The same argument can be carried out replacing $\mathcal{F}^0(s)$ by the σ-field generated by (the preimages of) $B(s)$. The argument can be extended from these special Y to general bounded Y by applying the monotone class theorem to the collection $\mathcal{H}$ of all functions Y for which the theorem holds and using the sets of the form $\{B(0) \in A_1\} \cap \{\Theta_0 B \in A_2\}$, $A_1 \subset \mathbf{R}$ Borel, $A_2 \in \mathcal{A}$, as $\cap$-stable collection $\mathcal{B}$. ∎

We now discuss a formal generalization of the weak Markov property, which has surprising applications. For this we make each σ-field a bit larger by allowing an infinitesimal glance into the future. Define

$$\mathcal{F}^+(s) = \bigcap_{t>s} \mathcal{F}^0(t).$$

Then $\{\mathcal{F}^+(s)\}$ is a slightly larger filtration, for which the weak Markov property still holds. Recall that we are dealing with Brownian motion defined on $\mathcal{C}([0,\infty))$.

Theorem 4.7 *For every $s \geq 0$ and every bounded random variable $Y : (\Omega, \mathcal{A}) \to \mathbf{R}$ we have*

$$\mathbf{E}_x\{Y \circ \theta_s \,|\, \mathcal{F}^+(s)\} = \mathbf{E}_{B(s)}\{Y\} \text{ almost surely.}$$

Proof: We observe that the process $\{B(t+s) - B(s) \,:\, t \geq 0\}$ is also independent of $\mathcal{F}^+(s)$. This holds because, by continuity,

$$B(t+s) - B(s) = \lim_{n\to\infty} B(s_n + t) - B(s_n)$$

for a strictly decreasing sequence $\{s_n\}$ converging to s, and each increment $B(s_n + t) - B(s_n)$ is independent of $\mathcal{F}^+(s)$. The remainder of the proof is as before. ∎

We now look at the *germ field* $\mathcal{F}^+(0)$, which heuristically comprises all events defined in terms of Brownian motion on an infinitesimal small interval to the right of the origin.

Theorem 4.8 (Blumenthal's 01-law.) *Let $x \in \mathbf{R}$ and $A \in \mathcal{F}^+(0)$. Then $\mathbf{P}_x\{A\} \in \{0,1\}$.*

Proof: Let $A \in \mathcal{F}^+(0)$. Observe that $\theta_0 =$id. By the previous two theorems

$$\mathbf{E}_x\{1_A \,|\, \mathcal{F}^+(0)\} = \mathbf{E}_x\{1_A \,|\, \mathcal{F}^0(0)\}.$$

Now $\mathcal{F}^0(0) = \{\emptyset, \Omega\}$ and thus 1_A is independent of $\mathcal{F}^0(0)$. Hence on the left hand side $\mathbf{E}_x\{1_A \,|\, \mathcal{F}^0(0)\} = \mathbf{E}_x\{1_A\} = \mathbf{P}_x(A)$. On the other hand, 1_A is $\mathcal{F}^+(0)$-measurable and hence the right hand side equals $\mathbf{E}_x\{1_A \,|\, \mathcal{F}^+(0)\} = 1_A$ almost surely. ∎

As a first application we show that Brownian motion has positive and negative values and zeroes in every small interval to the right of the origin. This is quite remarkable!

Theorem 4.9 *Define* $\tau = \inf\{t > 0 : B(t) > 0\}$ *and* $\sigma = \inf\{t > 0 : B(t) = 0\}$. *Then*

$$\mathbf{P}_0\{\tau = 0\} = \mathbf{P}_0\{\sigma = 0\} = 1\,.$$

Proof: The event

$$\{\tau = 0\} = \bigcap_{n=1}^{\infty} \Big\{ \text{ there is } 0 < \varepsilon < 1/n \text{ such that } B(\varepsilon) > 0 \Big\}$$

is clearly in $\mathcal{F}^+(0)$. Hence we just have to show that this event has positive probability. This follows, as, for $t > 0$,

$$\mathbf{P}_0\{\tau \le t\} \ge \mathbf{P}_0\{B(t) > 0\} = 1/2\,.$$

Hence $\mathbf{P}_0\{\tau = 0\} \ge 1/2$ and we have shown the first part. The same argument works replacing $B(t) > 0$ by $B(t) < 0$ and from these two facts $\mathbf{P}_0\{\sigma = 0\} = 1$ follows, using the intermediate value property of continuous functions. ∎

As an EXERCISE we prove another 01-law. Define $\mathcal{G}(t)$ to be the σ-field defined by all increments $B(t_1) - B(t_0)$ for $t \le t_0 \le t_1$. $\mathcal{G}(t)$ describes the future at time t. Let $\mathcal{T} = \bigcap_{t \ge 0} \mathcal{G}(t)$ be the σ-field of all *tail events*.

Theorem 4.10 *Let* $x \in \mathbf{R}$ *and* $A \in \mathcal{T}$. *Then* $\mathbf{P}_x\{A\} \in \{0, 1\}$.

4.3 Stopping times and the strong Markov property

Heuristically, the weak Markov property states that Brownian motion is *started anew* at each deterministic time instance. It is a crucial property of Brownian motion that this holds also for an important class of random times. These random times are called *stopping times*, they are of vital importance.

The basic idea is that a random time T is a stopping time if we can decide whether $\{T < t\}$ by just knowing the path of the stochastic process up to time t. Think of the situation that T is the moment where some random event related to the process happens.

Definition
A random variable T with values in $\mathbf{R} \cup \{\infty\}$ is called a *stopping time* with respect to the filtration $\{\mathcal{F}(t) : t \ge 0\}$ if, for every $t \ge 0$, $\{T < t\} \in \mathcal{F}(t)$. It is called a *strict stopping time* if, for every $t \ge 0$, $\{T \le t\} \in \mathcal{F}(t)$. Every strict stopping time is also a stopping time, because

$$\{T < t\} = \bigcup_{n=1}^{\infty} \{T \le t - 1/n\} \in \mathcal{F}(t)\,.$$

For certain nice filtrations strict stoppping times and stopping times agree. In order to come into this situation we are going to work with the filtration $\{\mathcal{F}^+(t)\}$ in the case of Brownian motion and refer the notions of stopping time, etc. always to this filtration. As this filtration is larger than $\{\mathcal{F}^0(t)\}$, there are *more stopping times.* The crucial property which distinguishes $\{\mathcal{F}^+(t)\}$ from $\{\mathcal{F}^0(t)\}$ is *right-continuity*, which means that

$$\bigcap_{\varepsilon>0} \mathcal{F}^+(t+\varepsilon) = \mathcal{F}^+(t)\,.$$

To see this note that

$$\bigcap_{\varepsilon>0} \mathcal{F}^+(t+\varepsilon) = \bigcap_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \mathcal{F}^0(t+1/n+1/k) = \mathcal{F}^+(t)\,.$$

Theorem 4.11 *Every stopping time T with respect to the filtration $\{\mathcal{F}^+(t)\}$ is automatically a strict stopping time.*

Proof: Suppose that T is a stopping time. Then

$$\{T \le t\} = \bigcap_{k=1}^{\infty} \{T < t+1/k\} \in \bigcap_{n=1}^{\infty} \mathcal{F}^+(t+1/n) = \mathcal{F}^+(t)\,.$$

■

We give some EXAMPLES.

- Of course, every deterministic time $t \ge 0$ is also a stopping time.
- Suppose G is an open set. Then $T = \inf\{t \ge 0 \,:\, B(t) \in G\}$ is a stopping time.

  Proof: Let Q be the rationals in $(0,t)$. Then, by continuity of B,

  $$\{T < t\} = \bigcup_{s\in Q} \{B(s) \in G\} \in \mathcal{F}^+(t)\,.$$

- If $T_n \uparrow T$ is an increasing sequence of stopping times, then T is also a stopping time.

  Proof:

  $$\{T \le t\} = \bigcap_{n=1}^{\infty} \{T_n \le t\} \in \mathcal{F}^+(t)\,.$$

- Suppose H is a closed set, for example a singleton. Then $T = \inf\{t \ge 0 \,:\, B(t) \in H\}$ is a stopping time.

  Proof: Let $G(n)$ be an open neighbourhood of K with $K = \bigcap G(n)$. Then $T_n = \inf\{t \ge 0 \,:\, B(t) \in G(n)\}$ are stopping times, which are increasing to T.

- Let T be a stopping time. Define stopping times

$$T_n = (m+1)2^{-n} \text{ if } m2^{-n} \le T < (m+1)2^{-n}.$$

In other words, we stop at the first time of the form $k2^{-n}$ after T. It is easy to see that T_n is a stopping time. We use it later as a discrete approximation to T.

We now want to prove that Brownian motion starts anew at each stopping time. For this purpose we define, for every stopping time T, the σ-field

$$\mathcal{F}(T) = \{A \in \mathcal{A} : A \cap \{T < t\} \in \mathcal{F}^+(t) \text{ for all } t \ge 0\}.$$

This means that the part of A that lies in $\{T < t\}$ should be measurable with respect to the information available at time t. Heuristically, this is the collection of events that happened before the stopping time T. As in the proof of the last theorem we can infer that $\{T \le t\}$ may replace $\{T < t\}$ without changing the definition.

We need three lemmas, the third of which is proved as an EXERCISE.

Lemma 4.12 *If $S \le T$ are stopping times, then $\mathcal{F}(S) \subset \mathcal{F}(T)$.*

Proof: If $A \in \mathcal{F}(S)$, then $A \cap \{T \le t\} = (A \cap \{S \le t\}) \cap \{T \le t\} \in \mathcal{F}^+(t)$. ■

Lemma 4.13 *If $T_n \downarrow T$ are stopping times, then $\mathcal{F}(T) = \bigcap_{n=1}^{\infty} \mathcal{F}(T_n)$.*

Proof: By the last lemma, $\mathcal{F}(T_n) \supset \mathcal{F}(T)$ for all n, which proves $\subset$. On the other hand, if $A \in \bigcap_{n=1}^{\infty} \mathcal{F}(T_n)$, then for all $t \ge 0$,

$$A \cap \{T < t\} = \bigcup_{n=1}^{\infty} A \cap \{T_n < t\} \in \mathcal{F}^+(t).$$

Hence $A \in \mathcal{F}(T)$. ■

This result can be used in the proof of the following lemma. Use the discrete approximation of T by a sequence $T_n \downarrow T$, see the last example.

Lemma 4.14 *If T is a stopping time, then the random variable $B(T)$ is $\mathcal{F}(T)$-measurable.*

Theorem 4.15 (Strong Markov property) *For every almost surely finite stopping time $T \ge 0$ and every bounded random variable $Y : \Omega \times \mathbf{R} \to \mathbf{R}$ we have*

$$\mathbf{E}_x\{Y(\theta_T, T) \mid \mathcal{F}(T)\} = \mathbf{E}_{B(T)}\{Y(\cdot, T)\} \text{ almost surely.}$$

In particular, the process

$$\{B(T+t) - B(T) : t \ge 0\}$$

is a standard Brownian motion independent of $\mathcal{F}(T)$.

Proof: We show our statement first for stopping times S such that only countably many values $s_1 < s_2 < s_3 < \ldots$ are taken with positive probability and we can use this to condition with respect to the value of S. Let $A \in \mathcal{F}(S)$, then

$$\int_A Y(\theta_S B, S)\, d\mathbf{P}_x(B) = \sum_{n=1}^{\infty} \int_{A\cap\{S=s_n\}} Y(\theta_S, S)\, d\mathbf{P}_x = \sum_{n=1}^{\infty} \int_{A\cap\{S=s_n\}} Y(\theta_{s_n}, s_n)\, d\mathbf{P}_x\,.$$

Now,

$$A \cap \{S = s_n\} = (A \cap \{S \le s_n\}) \setminus (A \cap \{S \le s_{n-1}\}) \in \mathcal{F}^+(s_n)\,,$$

so by the definition of conditional expectation and the weak Markov property, the sum is equal to

$$\sum_{n=1}^{\infty} \int_{A\cap\{S=s_n\}} \mathbf{E}_{B(s_n)}\{Y(\theta_{s_n}, s_n)|\mathcal{F}^+(s_n)\}\, d\mathbf{P}_x = \sum_{n=1}^{\infty} \int_{A\cap\{S=s_n\}} \mathbf{E}_{B(s_n)}\{Y(\cdot, s_n)\}\, d\mathbf{P}_x$$

$$= \sum_{n=1}^{\infty} \int_{A\cap\{S=s_n\}} \mathbf{E}_{B(S)}\{Y(\cdot, S)\}\, d\mathbf{P}_x = \int_A \mathbf{E}_{B(S)}\{Y(\cdot, S)\}\, d\mathbf{P}_x\,.$$

By definition of conditional expectation and Lemma 4.14 we infer that

$$\mathbf{E}_x\{Y(\theta_S, S)|\mathcal{F}(S)\} = \mathbf{E}_{B(S)}\{Y(\cdot, S)\}\,.$$

It remains to generalize this to general stopping times T. For this purpose we look at the stopping times T_n defined by

$$T_n = (m+1)2^{-n} \text{ if } m2^{-n} \le T < (m+1)2^{-n}\,.$$

For them the strong Markov property holds and we have $T_n \downarrow T$ and hence also $\mathcal{F}(T_n) \supset \mathcal{F}(T)$. We choose Y of the form

$$Y(f, T) = F(f(t_n) - f(t_{n-1}), \ldots, f(t_2) - f(t_1), T)\,,$$

for a continuous bounded function $F : \mathbb{R}^n \to \mathbb{R}$ and $0 \le t_1 \le \ldots \le t_n$. Then, by continuity of B and defintion of T_n,

$$Y(\theta_T B, T) = \lim_{n\to\infty} Y(\theta_{T_n} B, T_n)\,.$$

Let $A \in \mathcal{F}(T)$. Since $T \le T_n$, we infer $A \in \mathcal{F}(T_n)$ and hence, by definition of conditional expectation,

$$\int_A Y(\theta_{T_n}, T_n)\, d\mathbf{P}_x = \int_A \mathbf{E}_{B(T_n)}\{Y(\cdot, T_n)\}\, d\mathbf{P}_x\,.$$

By bounded convergence we infer

$$\int_A Y(\theta_T, T)\, d\mathbf{P}_x = \int_A \mathbf{E}_{B(T)}\{Y(\cdot, T)\}\, d\mathbf{P}_x\,.$$

As the integrand on the right hand side is $\mathcal{F}(T)$-measurable,

$$\mathbf{E}_x\{Y(\theta_T, T)|\mathcal{F}(T)\} = \mathbf{E}_{B(T)}\{Y(\cdot, T)\} \text{ almost surely.}$$

Now, in order to pass to more general Y, one can first pass to an indicator of the form

$$Y(f,T) = 1_A(f(t_n) - f(t_{n-1}), \ldots, f(t_2) - f(t_1), T),$$

for $A \subset \mathbf{R}^n$ open, by applying our result to a monotonously increasing sequence of continuous functions Y_k converging to Y and using the theorem of monotone convergence. Then we can use the monotone class theorem to the collection $\mathcal{H}$ of all functions Y for which the theorem holds, using the inverse images of the open sets $A \subset \mathbf{R}^n$ under

$$(f,T) \mapsto (f(t_n) - f(t_{n-1}), \ldots, f(t_2) - f(t_1), T) \qquad 0 \le t_1 \le \ldots \le t_n .$$

as $\cap$-stable collection $\mathcal{B}$. Finally, note that the particular statement at the end follows immediately from the formula just proven. ■

Remark: Let $T = \inf\{t \ge 0 : B(t) = \max_{0 \le s \le 1} B(s)\}$. It is intuitively clear that T is *not* a stopping time. To prove it, observe that almost surely $T < 1$. $B(t+T) - B(T)$ does not take negative values in a small neighbourhood to the right of 0, which contradicts the strong Markov property and Theorem 4.9.

We will see many applications of the strong Markov property later, the next subsection names an interesting one.

4.4 The reflection principle

The reflection principle states that Brownian motion reflected at some stopping time T is still a Brownian motion. More formally:

Theorem 4.16 (Reflection principle) *If T is a stopping time and $\{B(t)\}$ is a standard Brownian motion, then Brownian motion reflected at T defined by*

$$B^*(t) = B(t)1_{\{t \le T\}} + (2B(T) - B(t))1_{\{t > T\}}$$

is also a standard Brownian motion.

Proof: By the strong Markov property both

$$\{B(t+T) - B(T) : t \ge 0\} \text{ and } \{-(B(t+T) - B(T)) : t \ge 0\}$$

are Brownian motions and independent of the beginning $\{B(t) : t \in [0,T]\}$. Hence the concatenation (glueing together) of the beginning with the first part and the concatenation with the second part have the same distribution. The first is just $\{B(t)\}$, the second is the object $\{B^*(t)\}$ introduced in the statement. ■

Let $M(t) = \max_{0 \le s \le t} B(s)$. A priori it is not at all clear what the distribution of this random variable is, but we can determine it as a consequence of the reflection principle. Recall that $B(t)$ has distribution $\mathcal{N}(0,t)$.

Theorem 4.17 *If $a > 0$ then $\mathbf{P}_0\{M(t) > a\} = 2\mathbf{P}_0\{B(t) > a\} = \mathbf{P}_0\{|B(t)| > a\}$.*

Proof: Let $T = \inf\{t \geq 0 : B(t) = a\}$ and let $\{B^*(t)\}$ be Brownian motion reflected at T. Then $\{M(t) > a\}$ is the disjoint union of the events $\{B(t) > a\}$ and $\{M(t) > a,\ B(t) \leq a\}$ and since the latter is exactly $\{B^*(t) \geq a\}$ the statement follows from the reflection principle. ■

Chapter 5

Martingales

In this chapter we get to know the processes which correspond to fair games, the martingales. We first study the theory mainly for discrete time and later (in *Probability II*) extend the theory to continuous time. Symmetric random walk and the critical Galton-Watson process (whose offspring distribution has expected value one) turn out to be important examples of discrete time martingales. The word *martingales* originally denotes a special gambling strategy, indicating the connection to fair games.

5.1 Martingales: Definition and Examples

We now look at discrete time processes $\{X_n\}$ and a *filtration*, i.e. an increasing sequence

$$\mathcal{F}(0) \subset \mathcal{F}(1) \subset \mathcal{F}(2) \subset \dots$$

of σ-fields. We always assume that the process $\{X_n\}$ is $\{\mathcal{F}(n)\}$-*adapted*, which means that

$$X_n \text{ is } \mathcal{F}(n)\text{-measurable,}$$

but in most cases we even consider the filtration $\mathcal{F}(n)$ generated by the (preimages of) random variables $X_0, \dots, X_n$. This filtration is called the *natural filtration for* $\{X_n\}$.

Definition
A discrete time process $\{X_n : n \geq 0\}$ is called a *martingale* relative to the filtration $\{\mathcal{F}(n)\}$ if

- $\{X_n\}$ is $\{\mathcal{F}(n)\}$-adapted,
- $\mathsf{E}\{|X_n|\} < \infty$ for all n, and
- $\mathsf{E}\{X_n \,|\, \mathcal{F}(n-1)\} = X_{n-1}$ almost surely, for all $n \geq 1$.

If we have just $\leq$ in the last condition, then $\{X_n\}$ is called a *supermartingale*.

Let us convince ourselves, before starting the discussion of martingales, that there are plenty of interesting EXAMPLES.

1) Suppose that $X_1, X_2, \dots$ are independent random variables with $\mathbf{E}|X_n| < \infty$ and $\mathbf{E}X_n = 0$. Let $S_n = \sum_{k=1}^n X_k$ be the partial sums and $\mathcal{F}(n)$ be the natural filtration of the $\{S_n\}$. Observe that this is also the natural filtration for the $\{X_n\}$. Then

$$\mathbf{E}\{S_n \,|\, \mathcal{F}(n-1)\} = \mathbf{E}\Big\{ \sum_{k=1}^{n-1} X_k + X_n \,\Big|\, \mathcal{F}(n-1) \Big\} = \sum_{k=1}^{n-1} X_k + \mathbf{E}X_n = S_{n-1}\,.$$

Hence the *random walk* $\{S_n\}$ is a martingale.

2) Suppose that $X_1, X_2, \dots$ are independent nonnegative random variables with $\mathbf{E}X_n = 1$. Let $M_n = \prod_{k=1}^n X_k$. Let $\mathcal{F}(n)$ be the natural filtration. Then,

$$\mathbf{E}\{M_n \,|\, \mathcal{F}(n-1)\} = M_{n-1}\mathbf{E}\Big\{X_n \Big| \mathcal{F}(n-1)\Big\} = M_{n-1}\mathbf{E}X_n = M_{n-1}\,.$$

Hence $\{M_n\}$ is a martingale.

3) Let the filtration $\{\mathcal{F}(n)\}$ be arbitrary and X an integrable random variable. Define $X_n = \mathbf{E}\{X|\mathcal{F}(n)\}$, one should interpret X_n as the data accumulated about X at time n. The martingale property of $\{X_n\}$ follows from the tower property of conditional expectation

$$\mathbf{E}\{X_n|\mathcal{F}(n-1)\} = \mathbf{E}\{\mathbf{E}\{X|\mathcal{F}(n)\}|\mathcal{F}(n-1)\} = \mathbf{E}\{X|\mathcal{F}(n-1)\} = X_{n-1}\,.$$

4) Recall the definition of the *Galton Watson process* $\{X_n\}$ with offspring distribution $(p_0, p_1, p_2, \dots)$. There are independent random variables Y_k, $k \in A$, such that

$$X_n = \sum_{k \in T, |k| = n-1} Y_k\,.$$

The Y_k are independent with the distribution given by the sequence $(p_0, p_1, \dots)$ and

$$X_{n-1} = \#\{k \in T, |k| = n-1\}\,.$$

Let $\mathcal{F}(n)$ be the σ-field on Ω^* generated by the Y_k with $|k| \le n-1$. Note that X_n is $\mathcal{F}(n)$-measurable. We see that

$$\mathbf{E}\{X_n|\mathcal{F}(n-1)\} = \sum_{k \in A, |k| = n-1} \mathbf{E}\Big\{1_{\{k\in T\}} Y_k \Big| \mathcal{F}(n-1)\Big\} = \sum_{k \in A, |k| = n-1} 1_{\{k \in T\}} \mathbf{E}\Big\{Y_k \Big| \mathcal{F}(n-1)\Big\}\,,$$

where we have taken out what is known. As Y_k is independent of $\mathcal{F}(n-1)$ if $|k| = n-1$ we may continue with

$$\sum_{k \in A, |k| = n-1} 1_{\{k \in T\}} \mathbf{E}\Big\{Y_k \Big| \mathcal{F}(n-1)\Big\} = \sum_{k \in A, |k| = n-1} 1_{\{k \in T\}} \mathbf{E}\{Y_k\} = X_{n-1}\mathbf{E}\{Y_k\}\,.$$

Hence $\{X_n\}$ is a martingale if and only if the expected number of offspring is

$$\mathbf{E}\{Y_k\} = \sum_{n=0}^{\infty} n p_n = 1$$

and a supermartingale if this is ≤ 1.

Let us discuss some consequences of the definition. First note that, for every martingale $\{X_n\}$, from the tower property of conditional expectation, for all $m < n$,

$$\mathbf{E}\{X_n|\mathcal{F}(m)\} = \mathbf{E}\{\mathbf{E}\{X_n|\mathcal{F}(n-1)\}\,|\,\mathcal{F}(m)\} = \mathbf{E}\{X_{n-1}|\mathcal{F}(m)\} = \ldots = X_m\,.$$

Taking expected values gives,

$$\mathbf{E}\{X_n\} = \mathbf{E}\{X_0\} \text{ for all } n.$$

It is immediate from the definition that, for every martingale $\{X_n\}$,

$$\mathbf{E}\{X_n - X_{n-1}|\mathcal{F}(n-1)\} = 0\,. \tag{5.1}$$

Considering $\{X_n\}$ to be the capital of a gambler at time n, this can be interpreted as saying that the game is *fair*, the expected profit in each step is 0. In the supermartingale case this is ≤ 0 and the game is unfavourable. Let us explore this interpretation a bit more.

Suppose that $\{C_n : n \geq 1\}$ is your stake on game n. You have to base your decision on C_n on the history of the game up to time $n-1$. Formally, C_n has to be $\mathcal{F}(n-1)$-measurable. We use this to **define:**

A process $\{C_n : n \geq 1\}$ is called *previsible* if C_n is $\mathcal{F}(n-1)$-measurable.

Your winnings in game n are then $C_n(X_n - X_{n-1})$ and the *total winnings up to time* n are given by

$$Y_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) =: (C \bullet X)_n\,.$$

The process $(C \bullet X)$ is called the *martingale transform* of X by C, it is the discrete analogue of the stochastic integrals. The big question is now: can you choose $\{C_n\}$ such that your expected total winnings are positive? A positive answer to this question would be the most useful result of this lecture, however, we can prove:

Theorem 5.1 (You can't beat the system) *Let C be a bounded, previsible process (i.e. such that $|C_n(\omega)| \leq C$ for all $n \geq 1$ and $\omega \in \Omega$). Then, if $\{X_n\}$ is a martingale, so is $\{(C \bullet X)_n\}$. Moreover, $(C \bullet X)_0 = 0$ and hence $\mathbf{E}\{(C \bullet X)\} = 0$*

Proof: It is clear that $(C \bullet X)_n$ is integrable (as C_n is bounded) and by definition the process $\{Y_n\} = \{(C \bullet X)_n\}$ is adapted to the filtration $\{\mathcal{F}(n)\}$ and starts in 0. We calculate

$$\mathbf{E}\{Y_n|\mathcal{F}(n-1)\} = \mathbf{E}\Big\{\sum_{k=1}^{n} C_k(X_k - X_{k-1})\Big|\mathcal{F}(n-1)\Big\}$$

$$= \sum_{k=1}^{n-1} C_k(X_k - X_{k-1}) + C_n\mathbf{E}\{X_n - X_{n-1}|\mathcal{F}(n-1)\} = Y_{n-1}\,.$$

This proves it all. ■

Remark: The proof also shows, if also $C_n \geq 0$ and $\{X_n\}$ is a supermartingale, so is $\{(C \bullet X)_n\}$.

In the next two sections we study the two most important theorems in martingale theory: Doob's Optional Stopping Theorem and Doob's Martingale Convergence Theorem.

5.2 Doob's Optional Stopping Theorem

We now study stopping times for martingales. The intuition is very similar to the stopping times we have discussed for Brownian motions, but some technical points are easier.

Suppose that $\{X_n\}$ is a martingale for the filtration $\{\mathcal{F}(n)\}$ on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. To avoid technicalities assume that the union of all $\mathcal{F}(n)$ generates $\mathcal{A}$. A map $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is called a *stopping time* if

$$\{T \le n\} = \{\omega : T(\omega) \le n\} \in \mathcal{F}(n) \text{ for all } n < \infty .$$

This corresponds to strict stopping times in the continuous time setting, replacing $\{T \le n\}$ by $\{T < n\}$ would change the definition, but as in discrete time $\{T < n\} = \{T \le n-1\}$ a stopping time in this weaker sense is a stopping time for the shifted filtration $\mathcal{F}'(n) := \mathcal{F}(n+1)$ so that no new theory would evolve from such a change. However one can require

$$\{T = n\} = \{\omega : T(\omega) = n\} \in \mathcal{F}(n) \text{ for all } n < \infty ,$$

and this definition *is* equivalent to our stopping time definition. This is easy to see, as the first definition implies

$$\{T = n\} = \{T \le n\} \setminus \{T \le n-1\} \in \mathcal{F}(n) ,$$

and the second definition implies

$$\{T \le n\} = \bigcup_{k=0}^{n} \{T = k\} \in \mathcal{F}(n) .$$

Interpreting $\{X_n\}$ as a game we interpret the stopping times as those instances when we can quit playing (and obtain our winnings or pay our losses). If we follow that strategy and play unit stakes up to time T and then quit playing, the stake process C is

$$C_n = 1_{\{n \le T\}} = 1 - 1_{\{T \le n-1\}},$$

which is $\mathcal{F}(n-1)$-measurable and hence $\{C_n\}$ is previsible. The winnings process is

$$(C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}) = X_{T\wedge n} - X_0 ,$$

We define X^T, the *process X stopped at T*, as

$$X_n^T(\omega) = X_{T(\omega)\wedge n}(\omega) .$$

Recall that $X^T - X_0$ is the martingale transform of X by the (bounded) stake process C defined above. Theorem 5.1 can be applied and yields:

Theorem 5.2 (Elementary stopping theorem) *If X is a martingale and T a stopping time, then X^T is a martingale. In particular,*

$$\mathbf{E}\{X_{T\wedge n}\} = \mathbf{E}\{X_0\} \text{ for all } n .$$

If X is a supermartingale, then so is X^T and we still have

$$\mathbf{E}\{X_{T\wedge n}\} \le \mathbf{E}\{X_0\} \text{ for all } n .$$

Example: We look at a *simple random walk*. Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables with distribution

$$\mathbf{P}\{Y_n = 1\} = \mathbf{P}\{Y_n = -1\} = \frac{1}{2},$$

and let

$$X_n = \sum_{k=1}^{n} Y_k,$$

the simple random walk. We have seen that this process is a martingale (with respect to the natural filtration) and we have studied this process a little in the lecture Stochastische Methoden. Let

$$T = \inf\{n \geq 0 \,:\, X_n = 1\}.$$

Clearly, T is a stopping time. It is not straightforward to see that

$$\mathbf{P}\{T < \infty\} = 1,$$

but this follows from the facts proved in Stochastische Methoden and should be accepted for the moment. Our theorem then states that

$$\mathbf{E}\{X_{T \wedge n}\} = \mathbf{E}\{X_0\} \text{ for all } n.$$

However, we have $X_T = 1$ almost surely and hence

$$1 = \mathbf{E}X_T \neq \mathbf{E}X_0 = 0.$$

In some sense this means that you *can* beat the system if you have an infinite amount of time and credit.

After this example one would very much like to see conditions which make sure that for nice martingales and stopping times we have $\mathbf{E}X_T = \mathbf{E}X_0$. This is the content of Doob's optional stopping theorem.

Theorem 5.3 (Doob's optional stopping theorem) *Let T be a stopping time and X a martingale. Then X_T is integrable and*

$$\mathbf{E}\{X_T\} = \mathbf{E}\{X_0\},$$

if one of the following conditions hold:

(1) *T is bounded (i.e. there is N such that $T(\omega) < N$ for all ω),*

(2) *T is almost surely finite and X is bounded, (i.e. there is a real K such that $|X_n(\omega)| < K$ for all n and ω),*

(3) *$\mathbf{E}\{T\} < \infty$ and there is $K > 0$ such that, for all n and ω, $|X_n(\omega) - X_{n-1}(\omega)| \leq K$.*

If $\{X_n\}$ is a super-martingale and either one of the three previous conditions or

(4) *X is nonnegative and T almost surely finite.*

holds, then X_T is integrable and $\mathbf{E}\{X_T\} \leq \mathbf{E}\{X_0\}$.

Proof: We assume that $\{X_n\}$ is a supermartingale. Then $\{X_{T\wedge n}\}$ is a supermartingale and, in particular, integrable, and

$$\mathbb{E}\{X_{T\wedge n} - X_0\} \le 0 .$$

For (1) the result follows by choosing $n = N$. For (2) let $n \to \infty$ and use dominated convergence. For (3) we observe that

$$|X_{T\wedge n} - X_0| = \Big| \sum_{k=1}^{T\wedge n} (X_k - X_{k-1}) \Big| \le KT .$$

By assumption KT is an integrable function and we can use dominated convergence again. For (4) we use Fatou's lemma to see

$$\mathbf{E}\{X_T\} = \mathbf{E}\{\liminf_{n\to\infty} X_{T\wedge n}\} \le \liminf_{n\to\infty} \mathbf{E}\{X_{T\wedge n}\} = \mathbf{E}\{X_0\} .$$

The statement for martingales follows by applying the previous to the supermartingales $\{X_n\}$ and $\{-X_n\}$ separately. ■

In the situation of our example these conditions must fail. A glance at the third condition gives a striking corollary:

Corollary 5.4 *For a simple random walk the expected first hitting time of level* 1 *is infinite.*

5.3 Application to Brownian Motion: Exit from an interval

Suppose (a, b) is an interval containing the origin. At which end does Brownian motion leave the interval? In this section we use martingales to calculate the probability that Brownian motion leaves such an interval at the upper or lower end. We use the strong Markov property to embed a martingale into Brownian motion. Let us first state the main result of this section.

Theorem 5.5 *Let $\{B(t) : t \ge 0\}$ be standard Brownian motion and suppose that $a < 0 < b$ and $T = \inf\{t \ge 0 : B(t) \notin (a, b)\}$. Then T is an almost surely finite stopping time and*

$$\mathbf{P}\Big\{B(T) = a\Big\} = \frac{b}{b-a} \text{ and } \mathbf{P}\Big\{B(T) = b\Big\} = \frac{-a}{b-a} .$$

Proof: The fact that T is almost surely finite follows from the fact that, almost surely,

$$\limsup_{t\to\infty} |B(t)/\sqrt{t}| = \infty ,$$

which was proved as an exercise.

We first suppose that a and b are rationals. Then there are integers $n > 0$ and p, q such that $a = p/n$ and $b = q/n$. We define a sequence $T_0, T_1, T_2, \ldots$ of stopping times by $T_0 = 0$,

$$T_{k+1} = \inf\{t > T_k : |B(t) - B(T_k)| = 1/n\} \text{ for all } k \ge 0 .$$

Then we study the discrete time process $X_k = B(T_k)$. This process takes values in the set

$$\mathcal{D} = \{l/n \,:\, l \text{ integer }\}.$$

Denote by $Y_k = X_k - X_{k-1}$ the increments of the process. Then, by the strong Markov property, the Y_k are independent and identically distributed. The individual distributions are, by symmetry, given by

$$\mathbf{P}\{Y_k = 1/n\} = \frac{1}{2} = \mathbf{P}\{Y_k = -1/n\}\,.$$

In particular, as the expected value is zero, the process

$$X_k = \sum_{i=1}^{k} Y_i$$

is a martingale, see example 1. Let $S = \inf\{k \geq 0 \,:\, X_k = p/n \text{ or } X_k = q/n\}$. This is a stopping time for the martingale. The relationship between this stopping time S and the stopping time T is given by

$$B(T) = X_S\,.$$

Hence, we have to show the middle step in the equation

$$\mathbf{P}\{B(T) = a\} = \mathbf{P}\{X_S = p/n\} = \frac{q}{q-p} = \frac{b}{b-a}\,.$$

We look at the expected value of X_S. By definition this is

$$\mathbf{E}\{X_S\} = \frac{p}{n}\mathbf{P}\{X_S = \frac{p}{n}\} + \frac{q}{n}\mathbf{P}\{X_S = \frac{q}{n}\}\,.$$

Now look at the stopped martingale $\{X_n^S\}$ defined by $X_n^S = X_{S\wedge n}$ and apply Doob's optional stopping theorem. Condition (2) is satisfied because X^S is bounded from above by q/n and from below by p/n and S is almost surely finite. Hence,

$$\mathbf{E}\{X_S\} = \mathbf{E}\{X_S^S\} = \mathbf{E}\{X_0^S\} = \mathbf{E}\{X_0\} = 0\,.$$

Abbreviating $P = \mathbf{P}\{X_S = p/n\}$ and $Q = \mathbf{P}\{X_S = q/n\}$ we thus get

$$0 = \frac{p}{n}P + \frac{q}{n}Q \text{ and, clearly, } 1 = P + Q\,.$$

Solving this system of equations gives the required result for X and hence also for Brownian motion.

It remains to look at irrational values $a < 0 < b$. This can be dealt with by approximation. Let $a_n \uparrow a$ be a sequence of rationals increasing to a and $b_n \uparrow b$ be a sequence of rationals increasing to b and T_n the exit time from the interval (a_n, b_n). Then

$$\mathbf{P}\{B(T) = a\} \geq \mathbf{P}\{B(T_n) = a_n\} = \frac{b_n}{b_n - a_n} \longrightarrow \frac{b}{b-a}\,.$$

The other inequality can be obtained by approximating a and b from above by rationals in the same manner. This finishes the proof. ■

A similar theorem can be proved for simple random walk $\{X_n\}$. We choose integers $a < 0 < b$. What is the probability that $\{X_n\}$ hits a before b? Look at the stopping times

$$S^a = \inf\Big\{n \geq 0 \,:\, X_n = a\Big\} \text{ and } S^b = \inf\Big\{n \geq 0 \,:\, X_n = b\Big\}.$$

Then show as an exercise

$$\mathbb{P}\{S^a < S^b\} = \frac{b}{b-a}.$$

We now look at an amusing EXAMPLE of Keeler and Spencer:

Optimal doubling strategy for backgammon: In our idealization we need a continuous time function $B(t) : [0, T] \to [0, 1]$, which models your current chances of winning. For backgammon it is best to choose Brownian motion $\{B(t) \,:\, t \geq 0\}$ started at $1/2$ and stopped upon leaving the interval $[0, 1]$. At the beginning the stake is, say, one. You and the opponent both have the right to announce a doubling of the stake. If player A does this, player B has the right to either accept this — then the game goes on with doubled stakes and only B has the right to announce the next doubling — or player B gives up and loses the current stake.

A doubling strategy consists of two numbers $0 < b < 1/2 < a < 1$ and you announce a double if $B(t) \geq a$ and give up if the opponent announces a double and $B(t) < b$. An optimal strategy a^*, b^* must satisfy:

- when $B(t) = b^*$ accepting and giving up must have the same expected payoff.

Denoting by v your expected winnings from unit stake when you announce a doubling at $B(t) = a^*$, this means, by the previous theorem,

$$-1 = \frac{b^*}{a^*} \cdot 2v + \frac{a^* - b^*}{a^*} \cdot (-2).$$

Now, clearly, $1/2 < a^* \leq 1 - b^*$. The expected winnings per unit stake are $v = 1$ if you announce a doubling at $1 - b^*$, at the instant when the opponent starts giving up. If $a^* < 1 - b^*$ your expected winnings are lower (otherwise he would give up), so that necessarily $a^* = 1 - b^*$ and $v = 1$. Hence,

$$-1 = \frac{b^*}{1 - b^*} \cdot 2 + \frac{1 - 2b^*}{1 - b^*} \cdot (-2).$$

Solving the system we obtain $b^* = 1/5$ and $a^* = 4/5$.

Result: If your opponent announces a double you should give up if you feel your chances of winning are below 20%. You should announce a double if your chances are above 80%.

5.4 Doob's Martingale Convergence Theorem

Doob's famous forward convergence theorem gives a sufficient condition for the almost sure convergence of martingales $\{X_n\}$ to a limiting random variable.

Theorem 5.6 (Martingale Convergence Theorem) *Let $\{X_n\}$ be a supermartingale, which is bounded in L^1, i.e. there is $K > 0$ such that $\mathbf{E}|X_n| \leq K$ for all n. Then there exists a real-valued random variable X on the same probability space such that*

$$\lim_{n\to\infty} X_n = X \text{ almost surely.}$$

Important remark: Note that if X_n is nonnegative, we have

$$\mathbf{E}|X_n| = \mathbf{E}\{X_n\} = \mathbf{E}\{X_0\} := K$$

and thus X_n is automatically bounded in L^1 and $\lim_{n\to\infty} X_n = X$ exists.

The **proof** of the lemma is based on the idea of counting the number of upcrossings of intervals. Basically, if $\{X_n\}$ does not converge it oscillates and hence crosses some interval $[a, b]$ infinitely often.

Formally, pick two numbers $a < b$ and let $I = [a, b]$. The number $U_N[a, b]$ of upcrossings of $[a, b]$ made by $\{X_n\}$ up to time N is defined as the largest integer k such that there are integers

$$0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le N$$

with $X_{s_i} < a$ and $X_{t_i} > b$ for all $1 \le i \le k$. Clearly, $U_n[a, b]$ is a random variable. We shall show:

Lemma 5.7 (Doob's upcrossing lemma) *Let $\{X_n\}$ be a supermartingale and $a < b$. Then, for all N,*

$$(b-a)\mathbf{E}\{U_N[a,b]\} \le \mathbf{E}\{(X_N - a)^-\}\,.$$

Proof: We will discuss this in the language of fair games. Suppose that we bet on the game $\{X_n\}$ with the following anti-cyclic strategy: Initially our stake C_0 is zero. We leave it like this until $X_n < a$ when we choose unit stake $C_{n+1} = 1$. We play unit stakes until $X_n > b$ when we stop and choose $C_{n+1} = 0$, keep it until $X_n < a$, and so fourth. Formally,

$$C_1 = 1_{\{X_0 < a\}} \text{ and } C_n = 1_{\{C_{n-1}=1\}} 1_{\{X_{n-1} \le b\}} + 1_{\{C_{n-1}=0\}} 1_{\{X_{n-1} < a\}}\,.$$

C_n is clearly previsible and we study the martingale transform

$$Y_n = (C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1})\,.$$

Recalling our strategy, we observe that

$$Y_N \ge (b-a) U_N[a,b] - (X_N - a)^-\,.$$

Each upcrossing increases the value of Y_n by at least $b - a$ and the last term is responsible for a possible unfinished upcrossing at the end. As C is previsible, nonnegative and bounded, Y is a supermartingale and

$$0 = \mathbf{E}\{Y_0\} \ge \mathbf{E}\{Y_N\} \ge (b-a)\mathbf{E}\{U_N[a,b]\} - \mathbf{E}\{(X_N - a)^-\}\,.$$

This is the necessary inequality. ∎

From this we can observe that the expected number of upcrossings is bounded if $\mathbf{E}\{(X_N - a)^-\}$ is bounded.

Lemma 5.8 *Let $\{X_n\}$ be a supermartingale, which is bounded in L^1. Then, for the increasing limit*

$$U[a,b] := \lim_{N\to\infty} U_N[a,b]$$

we have

$$(b-a)\mathbf{E}\{U[a,b]\} \le |a| + \sup_n \mathbf{E}|X_n| < \infty\,.$$

In particular, $\mathbf{P}\{U[a,b]=\infty\}=0\,.$

Proof: Recall that, by the upcrossing lemma,

$$(b-a)\mathbf{E}\{U_N[a,b]\} \le \mathbf{E}\{(X_N-a)^-\} \le |a| + \mathbf{E}|X_N| \le |a| + \sup_n \mathbf{E}|X_n|\,.$$

Now let $N\to\infty$ using monotone convergence. ∎

We are almost done, look at the event

$$M[a,b] := \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\}$$

and observe that if X_n does neither converges nor diverges to $\pm\infty$, there are rationals $a<b$ such that event $M[a,b]$ takes place. But if $M[a,b]$ takes place, then $U[a,b]=\infty$, which has probability zero. Taking the union over the countable collection of rationals $a<b$ we obtain that almost surely $\{X_n\}$ converges to a possibly infinite random variable X. But, by Fatou's lemma,

$$\mathbf{E}|X| = \mathbf{E}\{\liminf_{n\to\infty} |X_n|\} \le \liminf_{n\to\infty} \mathbf{E}|X_n| \le K < \infty\,,$$

hence $|X|<\infty$ almost surely. This finishes the proof of Doob's martingale convergence theorem.

Examples. **a)** Let $\{X_n\}$ be a simple random walk. Then $\{X_n\}$ is a martingale, which does not converge. Namely, by the recurrence of simple random walks, both level 0 and level 1 are reached infinitely often, contradicting convergence. Clearly, $\{X_n\}$ is not L^1-bounded.

b) If $\mathcal{F}(0)\subset\mathcal{F}(1)\subset\ldots$ is a filtration of a measurable space whose σ-field is generated by the union of the $\mathcal{F}(n)$ and X is a nonnegative random variable, then we infer from the martingale convergence theorem that

$$\lim_{n\to\infty} \mathbf{E}\{X|\mathcal{F}(n)\}$$

exists almost surely. It is a natural suspicion (and we will show in the next section) that this limit is almost surely equal to X.

c) We now look at the Galton-Watson process $\{X_n\}$ offspring distribution given by $(p_0,p_1,\ldots)$. Assume that the mean offspring number is

$$\sum_{n=0}^{\infty} np_n = 1\,.$$

As $\{X_n\}$ is a nonnegative martingale it converges almost surely to a random variable $X\ge 0$. Because $\{X_n\}$ is integer-valued we must have $X_n=X$ for sufficiently large n almost surely.

Now assume $p_1 < 1$ (in the case $p_1 = 1$ trivially $X_n = 1$ for all n). For every $k > 0$ we then have, for any K,

$$\mathbb{P}\{X_n = k \text{ for all } n \geq K\} \leq \lim_{M\to\infty} \prod_{n=K+1}^{M} \mathbf{P}\{X_n = k \,|\, X_{n-1} = k\} = 0\,,$$

hence $X = 0$ almost surely. In other words, *the critical Galton-Watson process becomes extinct in finite time.* However, recall that $\mathbf{E}\{X_n\} = 1$ for all n. The convergence in the martingale convergence theorem does not hold for the expected values.

5.5 Uniformly integrable martingales

The key to the difference between the two examples is the question when the almost sure convergence in the martingale convergence theorem can be replaced by L^1-convergence. We first study a general criterion for L^1-convergence of an arbitrary sequence of random variables. Recall that a sequence $\{X_n\}$ of random variables converges in L^1 to X iff $\mathbf{E}|X_n - X| \to 0$. This also implies that $\mathbf{E}\{X_n\} \to \mathbf{E}\{X\}$.

Definition: A sequence $\{X_n\}$ of random variables is called *uniformly integrable* if, for every $\varepsilon > 0$ there is $K \geq 0$ such that

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbf{P} < \varepsilon \text{ for all } n\,.$$

Theorem 5.9 *(a) Every bounded sequence of random variables is uniformly integrable.*

(b) If a sequence of random variables is dominated by an integrable, nonnegative random variable Y, i.e. if $|X_n| \leq Y$ for all n, then the sequence is uniformly integrable.

(c) Let $p \geq 1$. A sequence is L^p-bounded if $\sup_n \mathbf{E}\{|X_n|^p\} < \infty$. Every sequence, which is L^p-bounded for some $p > 1$ is uniformly integrable.

(d) Every uniformly integrable sequence is L^1-bounded.

(e) There are L^1-bounded sequences, which are not uniformly integrable.

Proof: Clearly, (a) follows from (b), which we now prove. Observe that,

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbf{P} \leq \int_{\{|Y|>K\}} |Y| \, d\mathbf{P}\,.$$

Now, using monotone convergence,

$$\mathbf{E}|Y| = \liminf_{K\to\infty} \int 1_{\{|Y|\leq K\}} |Y| \, d\mathbf{P}\,.$$

Hence $\lim_{K\to\infty} \int_{\{|Y|>K\}} |Y| \, d\mathbf{P} = 0$ and this implies uniform integrability of our family.
We come to (c) and suppose that $\sup_n \mathbf{E}\{|X_n|^p\} < C$ for some $p > 1$. Observe that, for every K,

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbf{P} \leq K^{1-p} \int_{\{|X_n|>K\}} |X_n|^p \, d\mathbf{P} \leq K^{1-p} C\,.$$

Choosing K large gives uniform integrability. (d) is easy, since

$$\mathbf{E}|X_n| \le \int_{\{|X_n|>K\}} |X_n|\, d\mathbf{P} + \int_{\{|X_n|\le K\}} |X_n|\, d\mathbf{P} \le 1 + K$$

for suitably chosen K. Finally, suppose that U is uniformly distributed on $(0,1)$ and let

$$X_n = n\mathbb{1}_{\{U\le 1/n\}} \ge 0\,.$$

Then $\mathbf{E}|X_n| = 1$ and hence the family is L^1-bounded, but

$$\int_{\{|X_n|>K\}} |X_n|\, d\mathbf{P} = 1 \text{ for all } n > K\,,$$

and hence the family cannot be uniformly integrable. ■

Recall that every almost surely convergent sequence also converges *in probability*, which means that

$$\lim_{n\to\infty} \mathbf{P}\{|X_n - X| > \varepsilon\} = 0 \text{ for all } \varepsilon > 0\,.$$

We can now state the key theorem of this section.

Theorem 5.10 (Uniform Integrability Theorem) *Suppose that $\{X_n\}$ is a uniformly integrable sequence of random variables, which converges in probability to a random variable X. Then the sequence converges also in L^1.*

Proof: For every $K \ge 0$ we define the cutoff-function

$$\varphi_K(x) = \begin{cases} K & \text{if } x > K \\ x & \text{if } |x| \le K \\ -K & \text{if } x < -K. \end{cases}$$

By uniform integrability one can choose K such that, for all n,

$$\mathbf{E}\{|\varphi_K(X_n) - X_n|\} < \frac{\varepsilon}{3} \text{ and } \mathbf{E}\{|\varphi_K(X) - X|\} < \frac{\varepsilon}{3}\,.$$

Since

$$|\varphi_K(X) - \varphi_K(X_n)| \le |X - X_n|$$

we infer that $\{\varphi_K(X_n)\}$ converges in probability to $\{\varphi_K(X)\}$. Hence there is N such that

$$\mathbf{P}\Big\{|\varphi_K(X_n) - \varphi_K(X)| > \frac{\varepsilon}{6}\Big\} < \frac{\varepsilon}{12K} \text{ for all } n \ge N\,.$$

Then, for $n \ge N$,

$$\begin{aligned} \mathbf{E}\{|\varphi_K(X_n) - \varphi_K(X)|\} &\le \int_{\{|\varphi_K(X_n)-\varphi_K(X)|>\varepsilon/6\}} |\varphi_K(X_n) - \varphi_K(X)|\, d\mathbf{P} \\ &\quad + \int_{\{|\varphi_K(X_n)-\varphi_K(X)|\le\varepsilon/6\}} |\varphi_K(X_n) - \varphi_K(X)|\, d\mathbf{P} \\ &\le 2K\mathbf{P}\Big\{|\varphi_K(X_n) - \varphi_K(X)| > \frac{\varepsilon}{6}\Big\} + \frac{\varepsilon}{6} \le \varepsilon/3\,. \end{aligned}$$

We thus get, from the triangular inequality, for all $n \geq N$,

$$\mathbf{E}\{|X_n - X|\} \leq \mathbf{E}\{|X_n - \varphi_K(X_n)|\} + \mathbf{E}\{|\varphi_K(X_n) - \varphi_K(X)|\} + \mathbf{E}\{|X - \varphi_K(X)|\} \leq \varepsilon\,.$$

This completes the proof. ∎

Note that, by Theorem 5.9(b), the uniform integrability theorem is a generalization of the dominated convergence theorem.

We go back to the study of martingales. Let $\{\mathcal{F}(n)\}$ be a filtration and $\{X_n\}$ a martingale with respect to this filtration. The next theorem shows that every uniformly integrable martingale is of the form that data is accumulated about a random variable X.

Theorem 5.11 (Convergence for uniformly integrable martingales) *Suppose that the martingale $\{X_n\}$ is uniformly integrable. Then there is an almost surely finite random variable X such that*

$$\lim_{n\to\infty} X_n = X \text{ almost surely and in } L^1.$$

Moreover, for every n, $X_n = \mathbf{E}\{X|\mathcal{F}(n)\}$.

Proof: Because $\{X_n\}$ is uniformly integrable, it is in particular L^1-bounded and thus, by the martingale convergence theorem, almost surely convergent to a real-valued random variable X. By the previous theorem this convergence holds also in the L^1-sense. To check the last assertion, we verify the two properties of conditional expectation. $\mathcal{F}(n)$-measurability of X_n is clear by definition, so let $F \in \mathcal{F}(n)$. For all $m \geq n$ we have, by the martingale property,

$$\int_F X_m \, d\mathbf{P} = \int_F X_n \, d\mathbf{P}\,.$$

We let $m \to \infty$. Then

$$\Big| \int_F X_m \, d\mathbf{P} - \int_F X \, d\mathbf{P} \Big| \leq \int |X_m - X| \, d\mathbf{P} \to 0\,,$$

hence we obtain, as required,

$$\int_F X \, d\mathbf{P} = \int_F X_n d\mathbf{P}\,.$$

∎

The previous theorem shows that every uniformly integrable martingale is of the type that data about some (hidden) random variable is accumulated (see example 3). Conversely, every martingale of this type is uniformly integrable and convergent to the hidden variable. This is the content of Lévy's upward theorem.

Theorem 5.12 (Lévy's upward theorem) *Let X be an integrable random variable on $(\Omega, \mathcal{A}, \mathcal{P})$ and $\{\mathcal{F}(n)\}$ be a filtration such that the union of the $\mathcal{F}(n)$ generates $\mathcal{A}$. Define $X_n = \mathbf{E}\{X|\mathcal{F}(n)\}$. Then $\{X_n\}$ is a uniformly integrable martingale and*

$$\lim_{n\to\infty} X_n = X \text{ almost surely and in } L^1.$$

The key to the proof of the theorem is the following lemma.

Lemma 5.13 *Let X be an integrable random variable and $\{\mathcal{F}(n)\}$ be a sequence of sigma-fields. If $X_n = \mathbf{E}\{X|\mathcal{F}(n)\}$, the sequence $\{X_n\}$ is uniformly integrable.*

Proof: Let $\varepsilon > 0$ be given. Choose $\delta > 0$ such that, for $F \in \mathcal{A}$,

$$\mathbf{P}\{F\} < \delta \text{ implies } \int_F |X|\, d\mathbf{P} < \varepsilon\,.$$

This is possible, since otherwise we could find a sequence $F(n) \in \mathcal{A}$ of events with $\mathbf{P}(F(n)) < 2^{-n}$ and $\int_{F(n)} |X| d\mathbf{P} \geq \varepsilon$. Then look at the event H that infinitely many of the events $F(n)$ happen. By Borel-Cantelli $\mathbf{P}(H) = 0$ but at the same time

$$\int_H |X|\, d\mathbf{P} = \int \limsup_{n\to\infty} 1_{F(n)} |X|\, d\mathbf{P} \geq \limsup_{n\to\infty} \int 1_{F(n)} |X|\, d\mathbf{P} \geq \varepsilon\,,$$

using Fatou's lemma. This is a contradiction.

Having δ at our disposal, we choose K larger than $\mathbf{E}|X|/\delta$. By considering positive and negative part of X separately, we obtain, almost surely,

$$\Big|\mathbf{E}\{X|\mathcal{F}(n)\}\Big| \leq \mathbf{E}\Big\{|X| \,\Big|\, \mathcal{F}(n)\Big\}\,.$$

Hence, using Markov's inequality,

$$K\mathbf{P}\{|\mathbf{E}\{X|\mathcal{F}(n)\}| > K\} \leq \mathbf{E}\Big|\mathbf{E}\{X|\mathcal{F}(n)\}\Big| \leq \mathbf{E}\Big\{\mathbf{E}\{|X|\,|\mathcal{F}(n)\}\Big\} = \mathbf{E}|X|\,,$$

which implies

$$\mathbf{P}\{|\mathbf{E}\{X|\mathcal{F}(n)\}| > K\} \leq \frac{\mathbf{E}|X|}{K} < \delta\,.$$

Note that this event is in $\mathcal{F}(n)$. We obtain, from the definition of conditional expectation, for all n,

$$\int_{\{|\mathbf{E}\{X|\mathcal{F}(n)\}|>K\}} |\mathbf{E}\{X|\mathcal{F}(n)\}|\, d\mathbf{P} \leq \int_{\{|\mathbf{E}\{X|\mathcal{F}(n)\}|>K\}} \mathbf{E}\{|X|\,|\mathcal{F}(n)\}\, d\mathbf{P}$$

$$= \int_{\{|\mathbf{E}\{X|\mathcal{F}(n)\}|>K\}} |X|\, d\mathbf{P} < \varepsilon\,.$$

This finishes our proof. ∎

Proof of the Upward Theorem. We already know that $\{X_n\}$ is a martingale (see example section) and that it is uniformly integrable. Hence there is a random variable Y such that

$$\lim_{n\to\infty} X_n = Y \text{ almost surely and in } L^1.$$

We have to show that $X = Y$ almost surely. Now the uniqueness theorem enters again. We may assume that $X \geq 0$, which also implies $X_n \geq 0$ and hence $Y \geq 0$ almost surely. Observe that

$$\mathbf{E}\{X\} = \mathbf{E}\{X_n\} \to \mathbf{E}\{Y\}\,.$$

Define probability measures P and Q on $\mathcal{A}$ by

$$P(A) = \frac{1}{\mathbf{E}\{X\}} \int_A X \, d\mathbf{P} \text{ and } Q(A) = \frac{1}{\mathbf{E}\{X\}} \int_A Y \, d\mathbf{P} \, .$$

Now $\mathcal{F} := \bigcup \mathcal{F}(n)$ is a $\cap$-stable system which generates $\mathcal{A}$. If $A \in \mathcal{F}(n)$, then, for all $m \geq n$,

$$P(A) = \frac{1}{\mathbf{E}\{X\}} \int_A X \, d\mathbf{P} = \frac{1}{\mathbf{E}\{X\}} \int_A X_m \, d\mathbf{P} \, .$$

Also,

$$\lim_{m\to\infty} \frac{1}{\mathbf{E}\{X\}} \int_A X_m \, d\mathbf{P} = \frac{1}{\mathbf{E}\{X\}} \int_A \lim_{m\to\infty} X_m \, d\mathbf{P} = Q(A) \, ,$$

where the first equality follows, as before, from L^1-convergence. The uniqueness theorem now yields $P(A) = Q(A)$ for all $A \in \mathcal{A}$. Let $A = \{X > Y\} \in \mathcal{A}$. Then $\mathbf{P}(A) > 0$ would imply, by definition, $P(A) > Q(A)$, which is a contradiction. Hence, $\mathbf{P}(A) = 0$, which means $X \leq Y$ almost surely. In the same manner one can show $X \geq Y$ almost surely, and this finishes the proof of the Upward Theorem.

Theorem 5.14 (Lévy's downward theorem) *Suppose that $\{\mathcal{G}(-n) : n \geq 0\}$ is a collection of σ-fields such that*

$$\mathcal{G}(-\infty) := \bigcap_{k=0}^{\infty} \mathcal{G}(-k) \subset \cdots \subset \mathcal{G}(-n) \subset \cdots \subset \mathcal{G}(-2) \subset \mathcal{G}(-1) \, .$$

Let X be an integrable random variable and define

$$X_{-n} = \mathbf{E}\{X | \mathcal{G}(-n)\} \, .$$

Then

$$\lim_{n\to\infty} X_{-n} = \mathbf{E}\{X | \mathcal{G}(-\infty)\} \text{ almost surely and in } L^1.$$

Proof: Fix a positive integer N. We look at the filtration $\{\mathcal{F}(n)\}$ given by

$$\mathcal{F}(n) = \mathcal{G}((n-N) \wedge (-1))$$

and the adapted process $Y_n = X((n-N) \wedge (-1))$. Because, by the tower property,

$$\mathbf{E}\{Y_n \,|\, \mathcal{F}(n-1)\} = \mathbf{E}\Big\{\mathbf{E}\{X \,|\, \mathcal{G}((n-N) \wedge (-1))\} \,\Big|\, \mathcal{G}((n-1-N) \wedge (-1))\Big\}$$

$$= \mathbf{E}\Big\{X \,\Big|\, \mathcal{G}((n-1-N) \wedge (-1))\Big\} = Y_{n-1} \, ,$$

this is indeed a martingale. We obtain, from the upcrossing lemma,

$$(b-a)\mathbf{E}\{U_N[a,b]\} \leq \mathbf{E}\{(X_{-1} - a)^-\} \, .$$

Letting $N \to \infty$ shows that, almost surely, the total number of downcrossings of $[a,b]$ by the process $\{X_{-n}\}$ is finite. Now one has this simultaneously for all rationals $a < b$ and we can argue for $\{X_{-n}\}$ as in the martingale convergence theorem. Hence, $\lim X_{-n} = X_{-\infty}$ exists almost surely. By Lemma 5.13 the sequence is even uniformly integrable and hence convergence holds in L^1. To see that $X_{-\infty} = \mathbf{E}\{X|\mathcal{G}(-\infty)\}$, first observe that, for all m and $G \in \mathcal{G}(-\infty) \subset \mathcal{G}(-m)$,

$$\int_G X \, d\mathbf{P} = \int_G X_{-m} \, d\mathbf{P} \,,$$

and then let $m \to \infty$ to see that $X_{-\infty}$ satisfies the conditions of a conditional probability of X given $\mathcal{F}(-\infty)$. ■

Example: **Exponential increase of a rabbit population.**

Suppose that $\{X_n\}$ is a Galton-Watson process. Recall that in the *critical case* $\mu = 1$ we have seen that the process *dies* almost surely in finite time. Now we shall give conditions that the process *grows exponentially* with positive probability. Assume that the offspring distribution (given by the sequence $\{p_n\}$) has

- mean $\mu = \sum_{n=0}^{\infty} np_n > 1$ (*supercritical case*),
- positive and finite variance $\sigma^2 = \sum_{n=0}^{\infty} n^2 p_n - \mu^2$.

We first show that $M_n = X_n/\mu^n$ defines a martingale. To see this recall that there are independent random variables Y_k, $k \in A$, such that

$$X_n = \sum_{k \in A, |k| = n-1} 1_{\{k \in T\}} Y_k \,.$$

The Y_k are independent with the distribution given by the sequence $(p_0, p_1, \ldots)$ and

$$X_{n-1} = \#\{k \in T, |k| = n-1\} \,.$$

Let $\mathcal{F}(n)$ be the σ-field on Ω^* generated by the Y_k with $|k| \le n-1$. Note that X_n is $\mathcal{F}(n)$-measurable. We see that

$$\begin{aligned}
\mathbf{E}\Big\{\frac{X_n}{\mu^n}\Big|\mathcal{F}(n-1)\Big\} &= \mu^{-n} \sum_{k \in A, |k| = n-1} \mathbf{E}\Big\{1_{\{k \in T\}} Y_k \Big| \mathcal{F}(n-1)\Big\} \\
&= \mu^{-n} \sum_{k \in A, |k| = n-1} 1_{\{k \in T\}} \, \mathbf{E}\Big\{Y_k \Big| \mathcal{F}(n-1)\Big\} \\
&= \mu^{-n} X_{n-1} \mathbf{E}\{Y_k\} = \frac{X_{n-1}}{\mu^{n-1}} \,.
\end{aligned}$$

Hence $\{M_k\}$ is a martingale. As $M_k \ge 0$ there exists a random variable $M \ge 0$ with

$$\lim_{n \to \infty} M_n = M \text{ almost surely.}$$

We want to show that $\mathbf{P}\{M > 0\} > 0$. For this purpose we show that $\{M_n\}$ is uniformly integrable. If this holds we have

$$\mathbf{E}M = \lim_{n \to \infty} \mathbf{E}M_n = 1 \,,$$

hence $M \geq 0$ cannot be identically zero so that $\mathbf{P}\{M > 0\} > 0$. To check uniform integrability it suffices to check L^2-boundedness (see Theorem 5.9c). We have

$$\begin{aligned}
\mathbf{E}\{M_n^2|\mathcal{F}(n-1)\} &= \mathbf{E}\{M_{n-1}^2|\mathcal{F}(n-1)\} + \mathbf{E}\{2M_{n-1}(M_n - M_{n-1})|\mathcal{F}(n-1)\} \\
&\quad + \mathbf{E}\{(M_n - M_{n-1})^2|\mathcal{F}(n-1)\} \\
&= M_{n-1}^2 + \mathbf{E}\{(M_n - M_{n-1})^2|\mathcal{F}(n-1)\}\,.
\end{aligned}$$

To compute the second term observe

$$\mathbf{E}\{(M_n - M_{n-1})^2|\mathcal{F}(n-1)\} = \mu^{-2n}\mathbf{E}\{(X_n - \mu X_{n-1})^2|\mathcal{F}(n-1)\}$$

and that, on $\{X_{n-1} = N\}$,

$$\mathbf{E}\Big\{(X_n - \mu X_{n-1})^2|\mathcal{F}(n-1)\Big\} = \mathbf{E}\Big\{\Big(\sum_{k\in T, |k|=n-1} Y_k - \mu N\Big)^2\Big|\mathcal{F}(n-1)\Big\} = N\sigma^2 = X_{n-1}\sigma^2\,.$$

Combining this yields

$$\mathbf{E}\{M_n^2\} = \mathbf{E}\{M_{n-1}^2\} + (\sigma^2/\mu^{2n})\mathbf{E}\{X_{n-1}\}\,.$$

Now, $\mathbf{E}\{X_{n-1}\} = \mu^{n-1}\mathbf{E}\{M_{n-1}\} = \mu^{n-1}$. Hence,

$$\mathbf{E}\{M_n^2\} = \mathbf{E}\{M_{n-1}^2\} + \frac{\sigma^2}{\mu^{n+1}}\,,$$

which implies by induction,

$$\mathbf{E}\{M_n^2\} \leq 1 + \sigma^2 \sum_{n=1}^{\infty} \frac{1}{\mu^{k+1}} < \infty\,,$$

and we infer that the martingale $\{M_n\}$ is L^2-bounded. Altogether we have shown that, almost surely, for all large n,

$$X_n = M_n\mu^n \geq \frac{M}{2}\mu^n\,,$$

so that if $M > 0$ the Galton Watson process increases exponentially and the event $\{M > 0\}$ has positive probability.

5.6 Kolmogorov's strong law of large numbers

In this section we give a very short proof of Kolmogorov's strong law of large numbers under minimal moment conditions. Martingale theory will serve as the major tool.

Theorem 5.15 *Suppose that $\{X_n\}$ is a sequence of independent and identically distributed integrable random variables. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mu \textit{ almost surely and in } L^1\,,$$

where μ is the common expected value of the X_n.

Proof: Abbreviate $S_n = \sum_{k=1}^n X_k$. Let $\mathcal{G}(-n)$ be the σ-field generated by the random variables $S_n, S_{n+1}, S_{n+2}, \dots$ and $\mathcal{G}(-\infty)$ the intersection of all these fields. From the exercises we know that

$$\mathbf{E}\{X_1|\mathcal{G}(-n)\} = \frac{S_n}{n} \text{ almost surely.}$$

Hence, by Lévy's downward theorem,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k = \lim_{n\to\infty} \mathbf{E}\{X_1|\mathcal{G}(-n)\} = \mathbf{E}\{X_1|\mathcal{G}(-\infty)\} \text{ almost surely and in } L^1 .$$

Now the limit is measurable with respect to the tail σ-field $\mathcal{T}$ of the sequence $\{X_n\}$ and, as this has only events of probability zero or one (by Kolmogorov's 0-1-law), we have $\mathbf{E}\{X_1|\mathcal{G}(-\infty)\} = \mu$ almost surely for some constant value μ. Because we have L^1-convergence we also have

$$\mu = \mathbf{E}\Big\{\lim_{k\to\infty} \frac{S_k}{k}\Big\} = \lim_{k\to\infty} \frac{\mathbf{E}\{S_k\}}{k} = \mathbf{E}X_n \text{ for all } n,$$

and this finishes the proof. ∎

Chapter 6

The Donsker Invariance Principle

Suppose we have a high precision machine which produces units of given length. Many very small effects change the adjustment of the machine slightly, so that after some time the small perturbations have summed up and the machine needs readjustment. Indeed each of these small perturbations is again the sum of many small effects, so that the total perturbation after time t is

$$P(t) = \sum_{k=1}^{[nt]} X_k$$

where X_k are the independent perturbations occurring in the time interval $[k/n, (k+1)/n)$ satisfying $\mathbf{E}\{X_k\} = 0$ and variance of order $1/n$. If we consider increasingly small effects, we increase the n and the question is whether this leads to convergence of the process $\{P(t)\}$ to a limit process, which would then be a good model for the perturbing process. This is the problem of a *functional central limit theorem*, which we address in this chapter. The main result is that all random walks whose increments have finite variance can be rescaled so that they converge in distribution to Brownian motion.

We start with a proper definition of the *convergence of distributions* on function (and more general metric) spaces.

6.1 Convergence of distributions on metric spaces

The concepts of convergence of random variables we have studied so far,

- almost sure convergence,
- convergence in probability,
- L^1-convergence (and L^p-convergence),

refer to the sequences of random variables $\{X_n\}$ converging to a random variable X all on the same probability space. The values of the approximating sequences lead to conclusions about the values of the limit random variable. This is entirely different for *convergence in distribution*, which we now study. Intuitively if $\{X_n\}$ converges in distribution to X, this just means that the shape of the distributions of X_n for large n is like the shape of the distribution of X. Sample values from X_n allow no inference towards sample values from X and, indeed, there is no need to define X_n and X on the same probability space. In fact, convergence in distribution is only related to the convergence of the distributions of the random variables and not to the random variables themselves.

We start by giving a definition of convergence in distributions for random variables in metric spaces, explore some of its properties and then show that the concept of convergence in distribution for real-valued random variables known from the previous course is consistent with our definition.

Definition: Suppose (M, d) is a metric space and $\mathcal{A}$ the Borel-σ-field on M. Suppose that X_n and X are M-valued random variables. Then we say that X_n *converges in distribution* to X, if, for every bounded continuous $g : M \to \mathbf{R}$,

$$\lim_{n\to\infty} \mathbf{E}\{g(X_n)\} = \mathbf{E}\{g(X)\}.$$

We write $X_n \Rightarrow X$ for convergence in distribution.

Warning: If $M = \mathbf{R}$ and $X_n \Rightarrow X$ this does not imply that $\mathbf{E}\{X_n\}$ converges to $\mathbf{E}\{X\}$. Note that $g(x) = x$ is not a bounded function on $\mathbf{R}$.

Here is an alternative approach, which shows that convergence in distribution is in fact a convergence *of* the distributions. The statement of the following proposition is trivial.

Proposition 6.1 *Let Prob(M) be the set of probability measures on $(M, \mathcal{A})$. A sequence $\{P_n\} \subset$ Prob(M) converges weakly to a limit $P \in$ Prob if, for every continuous, bounded function $g : M \to \mathbf{R}$,*

$$\lim_{n\to\infty} \int g\, dP_n = \int g\, dP\,.$$

Then the limit of a convergent sequence is uniquely determined. Suppose that X_n and X are M-valued random variables. Then X_n converges in distribution to X, if and only if the distributions of X_n converge weakly to the distribution of X.

Proof: Only the uniqueness of the limit needs proof. If P and Q are two limits of the same sequence, then $\int f\, dP = \int f\, dQ$ for all bounded continuous $f : M \to \mathbf{R}$. For every open set $G \subset M$ we may choose an increasing sequence $f_n(x) = nd(x, G^c) \wedge 1$ of continuous functions converging to 1_G and infer from monotone convergence that $P(G) = Q(G)$. Now $P = Q$ follows from the Uniqueness Theorem. ■

In complete, separable metric spaces M weak convergence stems from a suitably defined metric on the space Prob(M). The case $M = \mathbf{R}$ will be discussed as an exercise.

EXAMPLES:

- Suppose $M = \{1, \ldots, m\}$ is finite and $d(x, y) = 1 - 1_{\{x=y\}}$. Then $X_n \Rightarrow X$ if and only if $\lim_{n\to\infty} \mathbf{P}\{X_n = k\} = \mathbf{P}\{X = k\}$ for all $k \in M$.

- Let $M = [0, 1]$ and $X_n = 1/n$ almost surely. Then $X_n \Rightarrow X$, where $X = 0$ almost surely. However, note that $\lim_{n\to\infty} \mathbf{P}\{X_n = 0\} = 0 \neq \mathbf{P}\{X = 0\} = 1$.

Theorem 6.2 *Suppose a sequence $\{X_n\}$ of random variables converges almost surely to a random variable X (of course, all on the same probability space). Then X_n converges in distribution to X.*

Proof: Suppose g is bounded and continuous. The $g(X_n)$ converges almost surely to $g(X)$. As the sequence is bounded it is also uniformly integrable, hence convergence holds also in the L^1-sense and this implies convergence of the expectations, i.e. $\mathbf{E}\{g(X_n)\} \to \mathbf{E}\{g(X)\}$. ■

Theorem 6.3 (Portmanteau Theorem) *The following statements are equivalent*

(i) $X_n \Rightarrow X$.

(ii) For all closed sets $K \subset M$, $\limsup_{n\to\infty} \mathbf{P}\{X_n \in K\} \le \mathbf{P}\{X \in K\}$.

(iii) For all open sets $G \subset M$, $\liminf_{n\to\infty} \mathbf{P}\{X_n \in G\} \ge \mathbf{P}\{X \in G\}$.

(iv) For all Borel sets $A \subset M$ *with* $\mathbf{P}\{X \in \partial A\} = 0$, $\lim_{n\to\infty} \mathbf{P}\{X_n \in A\} = \mathbf{P}\{X \in A\}$.

(v) For all bounded measurable functions $g : M \to \mathbf{R}$ *with* $\mathbf{P}\{g \text{ is discontinuous at } X\} = 0$ *we have* $\mathbf{E}\{g(X_n)\} \to \mathbf{E}\{g(X)\}$.

Proof:

(i)⇒(ii) Let $g_n(x) = 1 - (nd(x,K) \wedge 1)$, which is continuous and bounded, is 1 on K and converges pointwise to 1_K. Then, for every n,

$$\limsup_{k\to\infty} \mathbf{P}\{X_k \in K\} \le \limsup_{k\to\infty} \mathbf{E}\{g_n(X_k)\} = \mathbf{E}\{g_n(X)\}\,.$$

Let $n \to \infty$. The integrand on the right hand side is bounded by 1 and converges pointwise and hence in the L^1-sense to $1_K(X)$.

(ii)⇒(iii) Follows from $1_G = 1 - 1_K$ for the closed set $K = G^c$.

(iii)⇒(iv) Let G be the interior and K the closure of A. Then, by assumption, $\mathbf{P}\{X \in G\} = \mathbf{P}\{X \in K\} = \mathbf{P}\{X \in A\}$ and we may use (iii) and (ii) (which follows immediately from (iii)) to get

$$\limsup_{n\to\infty} \mathbf{P}\{X_n \in A\} \le \limsup_{n\to\infty} \mathbf{P}\{X_n \in K\} \le \mathbf{P}\{X \in K\} = \mathbf{P}\{X \in A\},$$

$$\liminf_{n\to\infty} \mathbf{P}\{X_n \in A\} \ge \liminf_{n\to\infty} \mathbf{P}\{X_n \in G\} \ge \mathbf{P}\{X \in G\} = \mathbf{P}\{X \in A\}.$$

(iv)⇒(v) From (iv) we infer that the convergence holds for g of the form $g(x) = \sum_{n=1}^{N} a_n 1_{A_n}$ where A_n satisfies $\mathbf{P}\{X \in A_n\} = 0$. Let us call such functions elementary. Given g as in (v) we observe that for every $a < b$ with possibly a countable set of exceptions

$$\mathbf{P}\Big\{X \in \partial\{x \,:\, g(x) \in (a,b]\}\Big\} = 0\,.$$

Indeed, if $X \in \partial\{x : g(x) \in (a,b]\}$ then either g is discontinuous in X or $g(X) = a$ or $g(X) = b$. The first event has probability zero and so have the last two except possibly for a countable set of values of a, b. By decomposing the real axis in small suitable intervals we thus obtain an increasing sequence g_n and a decreasing sequence h_n of elementary functions both converging pointwise to g. Now, for all k,

$$\limsup_{n\to\infty} \mathbf{E}\{g(X_n)\} \le \limsup_{n\to\infty} \mathbf{E}\{h_k(X_n)\} = \mathbf{E}\{h_k(X)\},$$

and

$$\liminf_{n\to\infty} \mathbf{E}\{g(X_n)\} \ge \liminf_{n\to\infty} \mathbf{E}\{g_k(X_n)\} = \mathbf{E}\{g_k(X)\}.$$

and the right sides converge, as $k \to \infty$, by bounded convergence, to $\mathbf{E}\{g(X)\}$.

(v)⇒(i) This is trivial.

■

To remember the directions of the inequalities in the Portmanteau Theorem it is useful to recall the last example $X_n = 1/n \to 0$ and choose $G = (0,1)$ and $K = \{0\}$ to obtain cases where the opposite inequalities fail.

The notion of convergence in distribution for real valued random variables is already known from the previous course. We can now see, that the concepts agree for the case of real random variables.

Theorem 6.4 (Helly-Bray Theorem) *Let X_n and X be real valued random variables and define the associated distribution functions $F_n(x) = \mathbf{P}\{X_n \le x\}$ and $F(x) = \mathbf{P}\{X \le x\}$. Then the following assertions are equivalent.*

(a) X_n converges in distribution to X,

(b) $\lim_{n\to\infty} F_n(x) = F(x)$ for all x such that F is continuous in x.

Proof:

(a)⇒(b) Use property (iv) for the set $A = (-\infty, x]$.

(b)⇒(a) We choose a dense sequence $\{x_n\}$ with $\mathbf{P}\{X = x_n\} = 0$ and note that every open set $G \subset \mathbf{R}$ can be written as the countable union of disjoint intervals $I_k = (a_k, b_k]$ with a_k, b_k chosen from the sequence. We have

$$\lim_{n\to\infty} \mathbf{P}\{X_n \in I_k\} = \lim_{n\to\infty} F_n(b_k) - F_n(a_k) = F(b_k) - F(a_k) = \mathbf{P}\{X \in I_k\}.$$

Hence, for all N,

$$\liminf_{n\to\infty} \mathbf{P}\{X_n \in G\} \ge \sum_{k=1}^{N} \liminf_{n\to\infty} \mathbf{P}\{X_n \in I_k\} = \sum_{k=1}^{N} \mathbf{P}\{X \in I_k\},$$

and as $N \to \infty$ the last term converges to $\mathbf{P}\{X \in G\}$.

■

We finish this section with an easy observation, which follows directly from the definition:

Lemma 6.5 *If $X_n \Rightarrow X$ and $g : M \to \mathbf{R}$ is continuous, then $g(X_n) \Rightarrow g(X)$.*

6.2 The Donsker Invariance Principle: Statement and Applications

Let $\{X_n\}$ be a sequence of independent and identically distributed random variables and assume that they are normalized, so that $\mathbf{E}\{X_n\} = 0$ and $\mathrm{Var}(X_n) = 1$. This assumption is no loss of generality, because if X_n has finite variance we can always consider the normalization

$$\frac{X_n - \mathbf{E}\{X_n\}}{\sqrt{\mathrm{Var}(X_n)}}.$$

We look at the *random walk* generated by the sequence

$$S_n = \sum_{k=1}^{n} X_k \,,$$

and interpolate linearly between the integer points, i.e.

$$S(t) = S_{[t]} + (t - [t])(S_{[t]+1} - S_{[t]}) \,.$$

This defines a random function $S \in C[0, \infty)$. We now define a sequence $\{S^n\}$ of random functions in $C[0,1]$ by scaling S with a factor n in the time and a factor $\sqrt{n}$ in the space axis. More precisely

$$S^n(t) = \frac{S(nt)}{\sqrt{n}} \text{ for all } t \in [0,1].$$

Theorem 6.6 (Donsker's Invariance Principle) *On the space $C[0,1]$ of continuous functions on the unit interval with the metric induced by the sup-norm, the sequence $\{S^n\}$ converges in distribution to a standard Brownian motion $\{B(t) : t \in [0,1]\}$.*

The theorem is also called *functional central limit theorem.* Before discussing the proof, we shall discuss consequences and applications of the theorem. Using the normalization indicated before one obtains that $\{X_n\}$ is any sequence of independent and identically distributed random variables with expectation μ and finite, positive variance σ^2, then $\{S^n\}$ converges in distribution to a Brownian motion with drift parameter μ and diffusion parameter σ^2.

As a first application we prove a central limit theorem with minimal moment conditions.

Theorem 6.7 (Central Limit Theorem) *Suppose that $\{X_k\}$ is a sequence of independent, identically distributed random variables with $\mathbf{E}\{X_k\} = 0$ and $\mathrm{Var}(X_k) = 1$. Then*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \Rightarrow X \,,$$

where X is a standard normally distributed random variable.

Proof: Consider the continuous function $g : C[0,1] \to \mathbb{R}$ defined by $g(f) = f(1)$. Then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k = g(S^n)$$

and $g(B)$ is standard normally distributed. Hence the statement follows from Donsker's theorem by means of Lemma 6.5. ∎

The next two theorems give examples why the name *invariance principle* is justified. The limits we obtain are invariant under (i.e. do not depend on) the choice of the exact distribution of the random variables X_n. A special case of interest is $\mathbb{P}\{X_n = 1\} = 1/2 = \mathbb{P}\{X_n = -1\}$ in which case the associated random walk is the symmetric random walk.

Theorem 6.8 *Suppose that $\{X_k\}$ is a sequence of independent, identically distributed random variables with $\mathbf{E}\{X_k\} = 0$ and $\mathrm{Var}(X_k) = 1$. Let $\{S_n\}$ be the associated random walk and*

$$M_n = \max\{S_k \,:\, 0 \le k \le n\}$$

its maximal value up to time n. Then, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{P}\{M_n \ge x\sqrt{n}\} = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\, dy\,.$$

Proof: Suppose that $g : \mathbb{R} \to \mathbb{R}$ is a continuous bounded function. Define a continuous bounded function $G : C[0,1] \to \mathbb{R}$ by

$$G(f) = g\Big(\max_{x\in[0,1]} f(x)\Big)\,.$$

Then, by definition,

$$\mathbf{E}\Big\{G(S^n)\Big\} = \mathbf{E}\Big\{g\Big(\max_{0\le t\le 1} \frac{S(tn)}{\sqrt{n}}\Big)\Big\} = \mathbf{E}\Big\{g\Big(\frac{\max_{0\le k\le n} S_k}{\sqrt{n}}\Big)\Big\}\,,$$

and

$$\mathbf{E}\Big\{G(B)\Big\} = \mathbf{E}\Big\{g\Big(\max_{0\le t\le 1} B(t)\Big)\Big\}.$$

Hence, by Donsker's Theorem,

$$\lim_{n\to\infty} \mathbf{E}\Big\{g\Big(\frac{M_n}{\sqrt{n}}\Big)\Big\} = \mathbf{E}\Big\{g\Big(\max_{0\le t\le 1} B(t)\Big)\Big\}.$$

From the Portmanteau Theorem and the reflection principle we infer

$$\lim_{n\to\infty} \mathbf{P}\{M_n \ge x\sqrt{n}\} = \mathbf{P}\{\max_{0\le t\le 1} B(t) \ge x\} = 2\mathbf{P}\{B(1) \ge x\}\,,$$

and the latter probability is the given integral. ∎

Theorem 6.9 (Arc-sine law for the last sign-change) *Suppose that $\{X_k\}$ is a sequence of independent, identically distributed random variables with $\mathsf{E}\{X_k\} = 0$ and $\mathrm{Var}(X_k) = 1$. Let $\{S_n\}$ be the associated random walk and*

$$N_n = \max\{1 \le k \le n \,:\, S_k S_{k-1} \le 0\}$$

the last time the random walk changes its sign before time n. Then N_n/n converges in distribution to a random variable with density

$$\frac{1}{\pi\sqrt{x(1-x)}} \text{ for } x \in (0,1).$$

In particular, for all $x \in (0,1)$,

$$\lim_{n\to\infty} \mathsf{P}\{N_n \le xn\} = \frac{2}{\pi}\arcsin(\sqrt{x})\,.$$

Remark: Note that this is surprising: the probability is high that N_n/n is near 0 or 1.

Proof: Step 1: Define a bounded function g on $C[0,1]$ by

$$g(f) = \max\{t \le 1 \,:\, f(t) = 0\}.$$

It is clear that $g(S^n)$ differs from N_n/n by a term, which is bounded by $1/n$ and hence vanishes asymptotically. Hence Donsker would imply convergence of N_n/n in distribution to

$$g(B) = \sup\{t \le 1 \,:\, B(t) = 0\}$$

if g was continuous. g is not continuous, but we shall see that g is continuous on the set $\mathcal{C}$ of all $f \in C[0,1]$ such that f takes positive and negative values in every neighbourhood of every zero and $f(1) \neq 0$. We shall also see that Brownian motion is almost surely in $\mathcal{C}$. From property (v) in the Portmanteau Theorem we can infer that, for every continuous bounded $h : \mathbf{R} \to \mathbf{R}$,

$$\lim_{n\to\infty} \mathsf{E}\Big\{h\Big(\frac{N_n}{n}\Big)\Big\} = \lim_{n\to\infty} \mathsf{E}\Big\{h \circ g(S^n)\Big\} = \mathsf{E}\Big\{h \circ g(B)\Big\} = \mathsf{E}\Big\{h(\sup\{t \le 1 \,:\, B(t) = 0\})\Big\}.$$

Step 2: g is continuous on $\mathcal{C}$.

Let $\varepsilon > 0$ is given and $f \in \mathcal{C}$. Let

$$\delta_0 = \min_{t \in [g(f)+\varepsilon, 1]} |f(t)|\,,$$

and choose δ_1 such that

$$(-\delta_1, \delta_1) \subset f(g(f) - \varepsilon, g(f) + \varepsilon)\,.$$

Let $0 < \delta < \delta_0 \wedge \delta_1$. If now $\|h - f\|_\infty < \delta$, then h has no zero in $(g(f) + \varepsilon, 1]$, but has a zero in $(g(f) - \varepsilon, g(f) + \varepsilon)$, because there are $s, t \in (g(f) - \varepsilon, g(f) + \varepsilon)$ with $h(t) < 0$ and $h(s) > 0$. Thus $|g(h) - g(f)| < \varepsilon$.

Step 3: Almost surely, every neighbourhood of every zero of Brownian motion $\{B(t) \,:\, 0 \le t \le 1\}$ contains positive and negative values and $B(1) \neq 0$.

Obviously, $B(1) \neq 0$ almost surely. For each rational $q \in [0,1)$ let t_q be the first zero of Brownian motion after q. As t_q is a stopping time we know from the strong Markov property that $\{B(t+t_q) : t \geq 0\}$ is a Brownian motion and by Theorem 4.9, almost surely, there exist positive and negative values in every small interval to the right of t_q. Taking the union over all rationals gives that, almost surely, all the zeroes of the form t_q have positive and negative values in every neighbourhood. But if t is a zero which is not equal to t_q for any rational q there exists, for every rational $q < t$ a zero t_q in the interval $[q, t)$, hence t has positive and negative values in every small interval to its left.

Step 4: Calculate the distribution of the random variable $L = \sup\{t \leq 1 : B(t) = 0\}$.

This will be done with the help of the Markov property and the reflection principle. Write

$$T_a = \inf\{t > 0 : B(t) = a\},$$

which is a stopping time. We get the distribution of T_a from the reflection principle, let $a \geq 0$,

$$\mathbf{P}\{T_a \leq t\} = \mathbf{P}\{\sup_{0 \leq s \leq t} B(s) \geq a\} = 2\mathbf{P}\{B(t) \geq a\} = 2\int_a^\infty \frac{1}{\sqrt{2\pi t}} \exp(-x^2/2t)\, dx,$$

change variables $x = (\sqrt{t/s})a$, $dx/ds = -a\sqrt{t}s^{-3/2}/2$,

$$= 2\int_t^0 \frac{1}{\sqrt{2\pi t}} \exp(-a^2/2s)(-\sqrt{t}a/2s^{3/2})\, ds = \int_0^t \frac{1}{\sqrt{2\pi s^3}} a \exp(-a^2/2s)\, ds.$$

The latter integrand is hence the density of T_a. Recall now that B under $\mathbf{P}_x$ is a Brownian motion with start in x. We now use the Markov property,

$$\begin{aligned}
\mathbf{P}\{L \leq s\} &= \mathbf{E}\Big\{\mathbf{E}\{1_{\{T_0(\theta_s) > 1-s\}} | \mathcal{F}^0(s)\}\Big\} \\
&= \mathbf{E}\Big\{\mathbf{P}_{B(s)}\{T_0 > 1-s\}\Big\} \\
&= 2\int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp(-x^2/2s) \int_{1-s}^\infty \frac{1}{\sqrt{2\pi r^3}} x \exp(-x^2/2r)\, dr\, dx \\
&= \frac{1}{\pi}\int_{1-s}^\infty \frac{1}{\sqrt{sr^3}} \int_0^\infty x \exp(-x^2(r+s)/2rs)\, dx\, dr \\
&= \frac{1}{\pi}\int_{1-s}^\infty \frac{1}{\sqrt{sr^3}} \frac{rs}{r+s}\, dr \\
&= \frac{1}{\pi}\int_{1-s}^\infty \Big(\frac{(r+s)^2}{rs}\Big)^{1/2} \frac{s}{(r+s)^2}\, dr,
\end{aligned}$$

and substitute $t = s/(r+s)$ with $dt = -s/(r+s)^2 dr$ to see that this equals

$$\frac{1}{\pi}\int_0^s \frac{1}{\sqrt{t(1-t)}}\, dt = \frac{2}{\pi}\arcsin(\sqrt{x}).$$

■

For the proof of the Donsker invariance principle we have essentially two possibilities:

- Suppose that a subsequence of $\{S^n\}$ converges in distribution to a limit X. This limit is a continuous random function, which is easily seen to have stationary, independent increments. Hence it is a Brownian motion and one can check that it must have drift 0 and variance 1. So Brownian motion is the only possible limit point of the sequence $\{S^n\}$. The difficult part of this proof is to show that every subsequence of $\{S^n\}$ has a convergent subsubsequence (the *tightness property*).

- We will follow the idea behind Theorem 6.2 and construct the random variables $X_1, X_2, X_3, \ldots$ on the same probability space as the Brownian motion in such a way that $\{S^n\}$ is with high probability close to a Brownian motion. This is called embedding the random walk in to Brownian motion and can be done with the famous embedding theorem of Skorokhod (1965).

As the first approach is more measure theoretic, we will prefer the second approach, which represents a typical technique in probability theory. The proof uses almost all the techniques we have developed in the lecture so far.

6.3 The Skorokhod Embedding Theorem

Theorem 6.10 (Skorokhod Embedding Theorem) *Suppose that X is a real valued random variable with $\mathbf{E}\{X\} = 0$ and $\mathbf{E}\{X^2\} < \infty$. Then X and a Brownian motion $\{B(t) : t \geq 0\}$ can be defined on a joint probability space such that there exists a stopping time T such that $\mathbf{E}\{T\} = \mathbf{E}\{X^2\} < \infty$ and, almost surely, $B(T) = X$.*

Examples:

- Define the product space $\Omega_1 \otimes C[0, \infty)$, where Ω_1 is a probability space on which X can be defined and the second factor is Wiener space. Simply let
$$T = \inf\{t \geq 0 : B(t) = X\}.$$
Then $B(T) = X$, as required, but this simple recipe gives no guarantee that $\mathbf{E}\{T\} < \infty$. In fact, the next example shows that this cannot always be the case.

- To see that *some* conditions have to be imposed on the X for the theorem to hold true consider $X = 1$ constant. In order to have T with $B(T) = X = 1$ we have to choose $T \geq \inf\{t \geq 0 : B(t) = 1\}$. We have seen (exercise) that this implies $\mathbf{E}\{T\} = \infty$, so that the theorem cannot hold in this situation.

- Now assume that X may take two values $a < b$. In order that $\mathbf{E}\{X\} = 0$ we must have $a < 0 < b$ and
$$\mathbf{P}\{X = a\} = \frac{b}{b-a} \text{ and } \mathbf{P}\{X = b\} = \frac{-a}{b-a}.$$
Choosing
$$T = \inf\{t : B(t) \notin (a, b)\},$$
we have seen in Theorem 5.5 that $B(T) = X$ has the given distribution. Moreover, by a recent exercise, $\mathbf{E}\{T\} = -ab$ is finite.

Proof: We will use the last example as a building block for the general case. The idea is to choose the boundaries of an interval (U, V) at random so that the Brownian motion at the first exit time from this interval has the right distribution.

Write P for the distribution of X. Then
$$c = E\{X^-\} = \int_{-\infty}^{0} (-u)\, dP(u) = E\{X^+\} = \int_0^\infty v\, dP(v).$$

If φ is bounded, measurable with $\varphi(0)=0$, then

$$\begin{aligned} c\int \varphi(x)\,dP(x) &= \Big(\int_0^\infty \varphi(v)\,dP(v)\Big)\int_{-\infty}^0 (-u)\,dP(u) \\ &\quad + \Big(\int_{-\infty}^0 \varphi(u)\,dP(u)\Big)\int_0^\infty v\,dP(v) \\ &= \int_0^\infty dP(v)\int_{-\infty}^0 dP(u)\,(v\varphi(u)-u\varphi(v)). \end{aligned}$$

We thus obtain the key formula

$$\int \varphi(x)\,dP(x) = \frac{1}{c}\int_0^\infty dP(v)\int_{-\infty}^0 dP(u)(v-u)\Big\{\frac{v}{v-u}\varphi(u)+\frac{-u}{v-u}\varphi(v)\Big\}.$$

Note that the expression in the curly brackets is $\mathbf{E}\{\varphi(B(T))\}$, where T is the first exit time of the interval (u,v). We now let $(\Omega_1,\mathcal{A}_1,\mathbf{P}_1)$ be a probability space on which a pair (U,V) of random variables can be defined with $U\le 0$ and $V\ge 0$ and distribution given by

$$\mathbf{P}_1\{(U,V)=(0,0)\} = P\{X=0\}$$

and, for Borel sets $A\subset(-\infty,0)\times(0,\infty)$,

$$\mathbf{P}_1\{(U,V)\in A\} = \frac{1}{c}\iint_{(u,v)\in A} dP(u)\,dP(v)\,(v-u)\,.$$

Note that this properly defines the distribution of the random vector (U,V). We formally define probability measures $\mu_{u,v}$ on $\{u,v\}$ by $\mu_{0,0}\{0\}=1$ and

$$\mu_{u,v}\{u\} = \frac{v}{v-u} \text{ and } \mu_{u,v}\{v\} = \frac{-u}{v-u} \text{ for } u<0<v.$$

With this definition the key formula can be written as follows. For all bounded measurable functions φ,

$$\int \varphi(x)\,dP(x) = \mathbf{E}_1\Big\{\int \varphi(x)\mu_{U,V}(dx)\Big\}.$$

Now define the product space

$$(\Omega,\mathcal{A},\mathbb{P}) = (\Omega_1,\mathcal{A}_1,\mathbf{P}_1)\otimes(C[0,\infty),\mathcal{A}_0,\mathbf{P}_0)$$

of Ω_1 and Wiener space. Observe that (U,V) and $\{B(t)\,:\,t\ge 0\}$ are independent random variables on this space. Define

$$T = \inf\{t\ge 0\,:\,B(t)\notin(U,V)\}.$$

This can be seen either as a stopping time with respect to $\mathcal{F}^+(t)\subset\mathcal{A}$ conditional on $U=u$ and $V=v$ or, on the whole space, as a stopping time with respect to the filtration $\mathcal{F}(t)=\mathcal{A}\otimes\mathcal{F}^+(t)$. Our key formula gives (denoting expectations with respect to $\mathbf{P}_0$ by $\mathbf{E}_0$ and expectations with respect to $\mathbb{P}_1$ by $\mathbf{E}_1$), by Fubini's Theorem,

$$\mathbf{E}\{\varphi(B(T))\} = \mathbf{E}_1\{\mathbf{E}_0\{\varphi(B(T))\}\} = \mathbf{E}_1\Big\{\int\varphi(x)\mu_{U,V}(dx)\Big\} = \int\varphi(x)\,dP(x) = E\{\varphi(X)\}.$$

Hence $B(T)$ and X have the same distribution. It remains to show that $\mathbf{E}\{T\} < \infty$. First recall that

$$\mathbf{E}\{T\} = \mathbf{E}_1\{\mathbf{E}_0\{T\}\} = \mathbf{E}_1\{-UV\}.$$

Now we use the distribution of (U, V), using the alternative expressions for $c = \int_0^\infty v\, dP(v) = \int_{-\infty}^0 (-u)\, dP(u)$,

$$\begin{aligned}
\mathbf{E}_1\{-UV\} &= \frac{1}{c}\int_{-\infty}^0 dP(u)(-u)\int_0^\infty dP(v)\, v(v-u) \\
&= \int_{-\infty}^0 dP(u)(-u)\Big\{ -u + \int_0^\infty dP(v)\,\frac{v^2}{c}\Big\} \\
&= \int_{-\infty}^0 (-u)^2 dP(u) + \int_0^\infty v^2\, dP(v) \\
&= E\{(X^-)^2\} + E\{(X^+)^2\} \\
&= E\{X^2\}.
\end{aligned}$$

■

Note that we had to enlarge Wiener space in order to define T. This can be avoided (*Dubin's stopping rule*). For the purpose of our proof the Skorokhod representation is satisfactory, it allows, by means of the Markov property, to embed the whole random walk in the Brownian motion.

Corollary 6.11 *Suppose that $X_1, X_2, X_3, \ldots$ are independent, identically distributed real valued random variables with mean zero and variance 1. Then there exists a sequence of stopping times $0 = T_0 \le T_1 \le T_2 \le T_3 \le \ldots$ such that the increments $T_n - T_{n-1}$ are independent, identically distributed, $\mathbf{E}\{T_n\} = n < \infty$ and, almost surely, the sequence $\{B(T_n) : n \ge 1\}$ has the distribution of the random walk $\{S_n\}$ associated with $\{X_n\}$.*

Proof: We define the space $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1)$ such that a sequence of independent variables (U_n, V_n) can be defined for X_n as in the proof before. Then we can define

$$T_1 = \inf\{t \ge 0 : B(t) \notin (U_1, V_1)\}$$

and obtain $B(T_1) = X_1$ in distribution and $\mathbf{E}\{T_1\} = 1$. By the strong Markov property

$$\{B_2(t) : t \ge 0\} = \{B(T_1 + t) - B(T_1) : t \ge 0\}$$

is again a Brownian motion and independent of $\mathcal{F}(T)$ and, in particular, of $(T_1, B(T_1))$. Hence we can define a stopping time

$$T_2 = T_1 + \inf\{t \ge 0 : B_2(t) \notin (U_2, V_2)\}$$

and observe that $T_2 - T_1$ is independent of T_1 with the same distribution and $B(T_2) - B(T_1) = B_2(T_2 - T_1)$ has the same distribution at X_2 and is independent of X_1. Furthermore $\mathbf{E}\{T_2\} = 2$. We can proceed inductively to get the corollary. ■

We have thus embedded the random walk $\{S_n\}$ into Brownian motion. In the next section we use this to prove Donsker's Theorem.

6.4 The Donsker Invariance Principle: Proof

We can now work with the sequence

$$0 = T_0 \le T_1 \le T_2 \le T_3 < \dots$$

defined by the corollary to the Skorokhod embedding theorem, such that $S_n = B(T_n)$ is the embedded random walk. We define $W^n \in C[0,1]$ by

$$W^n(t) = \frac{B(nt)}{\sqrt{n}} \text{ for } t \in [0,1] .$$

From the scaling property of Brownian motion we can see that all the random functions W^n are standard Brownian motions. We show that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\Big\{\|W^n - S^n\|_{\sup} > \varepsilon\} = 0 . \tag{6.1}$$

Let us first see why (6.1) implies the theorem. Suppose that $K \subset C[0,1]$ is closed and define

$$K[\varepsilon] = \{f \in C[0,1] \,:\, \|f-g\|_{\sup} \le \varepsilon \text{ for some } g \in K\}.$$

Then

$$\mathbf{P}\{S^n \in K\} \le \mathbf{P}\{W^n \in K[\varepsilon]\} + \mathbf{P}\{\|S^n - W^n\|_{\sup} > \varepsilon\Big\} .$$

By (6.1) the second term goes to 0 and the first term is equal to $\mathbf{P}\{B \in K[\varepsilon]\}$ for a Brownian motion B, independently of n. As K is closed we have

$$\lim_{\varepsilon\to 0} \mathbf{P}\{B \in K[\varepsilon]\} = \mathbf{P}\Big\{B \in \bigcap_{\varepsilon>0} K[\varepsilon]\Big\} = \mathbf{P}\{B \in K\}.$$

Altogether,

$$\limsup_{n\to\infty} \mathbf{P}\{S^n \in K\} \le \mathbf{P}\{B \in K\} ,$$

which is condition (iii) in the Portmanteau Theorem. It just remains to prove (6.1).

Let A_n be the event that there exists $t \in [0,1]$ such that $|S^n(t) - W^n(t)| > \varepsilon$. We have to show $\mathbf{P}(A_n) \to 0$. Let $k = k(t)$ be the unique integer with $(k-1)/n \le t < k/n$. Because S^n is linear on such an interval we have

$$A_n \subset \Big\{\exists t \,:\, \Big|\frac{S_k}{\sqrt{n}} - W^n(t)\Big| > \varepsilon\Big\} \cup \Big\{\exists t \,:\, \Big|\frac{S_{k-1}}{\sqrt{n}} - W^n(t)\Big| > \varepsilon\Big\}.$$

As $S_k = B(T_k) = \sqrt{n} W^n(T_k/n)$, we obtain

$$A_n \subset \Big\{\exists t \,:\, \Big|W^n\Big(\frac{T_k}{n}\Big) - W^n(t)\Big| > \varepsilon\Big\} \cup \Big\{\exists t \,:\, \Big|W^n\Big(\frac{T_{k-1}}{n}\Big) - W^n(t)\Big| > \varepsilon\Big\}.$$

For given $0 < \delta < 1$ the event on the right implies that either

$$\Big\{\exists s,t \in [0,2] \,:\, |s-t| < \delta \,,\, |W^n(s) - W^n(t)| > \varepsilon\Big\}$$

or

$$\Big\{\exists t \in [0,1] \,:\, |T_k/n - t| \vee |T_{k-1}/n - t| \ge \delta\Big\} .$$

Note that the probability of the first event does not depend on n. Choosing $\delta > 0$ small, we can make the first probability as small as we wish, since Brownian motion is uniformly continuous on $[0,2]$. It remains to show that for arbitrary, fixed $\delta > 0$,

$$\lim_{n\to\infty} \mathbf{P}\Big\{\exists t \, : \, |T_k/n - t| \vee |T_{k-1}/n - t| \geq \delta\Big\} = 0 \,. \tag{6.2}$$

To prove this we use that

$$\lim_{n\to\infty} \frac{T_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (T_k - T_{k-1}) = 1 \text{ almost surely.}$$

This is Kolmogorov's law of large numbers for the sequence $\{T_k - T_{k-1}\}$ of independent identically distributed random variables with mean 1. Observe that for every sequence $\{a_n\}$ of reals one has

$$\lim_{n\to\infty} \frac{a_n}{n} = 1 \Rightarrow \lim_{n\to\infty} \sup_{0\leq k\leq n} |a_k - k|/n = 0 \,.$$

This is a matter of plain (deterministic) arithmetic. Hence we have,

$$\lim_{n\to\infty} \mathbf{P}\Big\{ \sup_{0\leq k\leq n} \frac{|T_k - k|}{n} \geq \delta \Big\} = 0 \,. \tag{6.3}$$

Now recall that $t \in [(k-1)/n, k/n]$ and let $n > 2/\delta$. Then

$$\begin{aligned} &\mathbf{P}\Big\{\exists t \, : \, |T_k/n - t| \vee |T_{k-1}/n - t| \geq \delta\Big\} \\ &\leq \mathbf{P}\Big\{ \sup_{0\leq k\leq n} \frac{(T_k - (k-1)) \vee (k - T_{k-1})}{n} \geq \delta \Big\} \\ &\leq \mathbf{P}\Big\{ \sup_{0\leq k\leq n} \frac{T_k - k}{n} \geq \delta/2 \Big\} + \mathbf{P}\Big\{ \sup_{1\leq k\leq n} \frac{(k-1) - T_{k-1}}{n} \geq \delta/2 \Big\} \,, \end{aligned}$$

which by (6.3) converges to 0. Hence (6.2) and Donsker's invariance principle are proved.

Chapter 7

An outlook to stochastic integration

Looking back through the material of this course one can see that the emphasis was on *continuous time* processes, and in particular Brownian motion, rather than *discrete time* processes. The only notable exception was the chapter on martingales, where we were dealing exclusively with discrete time martingales (and, in fact, some extra effort was needed to apply the results to Brownian motion). It is quite easy to define martingales also in continuous time.

Definition: Let $(\Omega, \mathcal{A}, \mathsf{P})$ be a probability space and $\{\mathcal{F}(t) : t \geq 0\}$ a collection of sub-σ-fields with $\mathcal{F}(s) \subset \mathcal{F}(t)$ for all $s \leq t$, in other words a *filtration.* A process $\{X(t) : t \geq 0\}$ is called a *martingale* if

$$\mathsf{E}\Big\{X(t)\Big|\mathcal{F}(s)\Big\} = X(s) \text{ almost surely, for all } s \leq t.$$

Theorem 7.1 *Standard Brownian motion is a martingale.*

Proof: Choose the filtration $\mathcal{F}(t) = \mathcal{F}^+(t)$ and use the weak Markov property, to see

$$\mathsf{E}\Big\{X(t)\Big|\mathcal{F}(s)\Big\} = \mathsf{E}_{X(s)}\{B(t-s)\} = X(s)\,,$$

for all $s \leq t$. ■

The first major step in the discrete time martingale theory and the key step to all later theorems was the You-can't-beat-the-system-Theorem. Finding a continuous time analogue of this is a major problem for us. Recall that we were using a bounded, previsible process $\{C_n\}$, to define the martingale transform

$$(C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

C_n, the stake at the game $X_n - X_{n-1}$, had to be measurable with respect to $\mathcal{F}(n-1)$. All major theorems about martingales were derived from the fact that the martingale transform again defines a martingale. Recall that $(C \bullet X)_k$ can be interpreted as our total profit at time k if C_n is the number of items of a good we possess at time n and $\{X_n\}$ is the price process of one item of the good. It is clear that such a process would be of great practical and theoretical interest in a continuous time setting, however it is unclear how it could be defined and also what the appropriate condition for the stake process $\{C(t)\}$ could be.

Instead of solving this problem now, we first try a pathwise approach for integration with respect to a standard Brownian motion $\{B(t) : t \in [0,1]\}$. The most natural idea would be to define a Stieltjes integral

$$\int_0^t f(s)\, dB(s).$$

This would mean we let, for $n \geq 2$,

$$\Delta_n = \{0 = t_0^n < t_1^n < t_2^n < \cdots < t_n^n = 1\}$$

be a collection of partitions with $\Delta_n \subset \Delta_{n+1}$ such that the mesh of the partition

$$\delta_n = \max_{i=1}^{n} (t_i^n - t_{i-1}^n)$$

converges to 0. One would then hope that

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(t_{k-1}^n)(B(t_k^n) - B(t_{k-1}^n))$$

converges almost surely to a limit, which would be a reasonable generalization of the martingale transform. However, we shall see that such a limit may fail to exist.

Theorem 7.2 *Suppose $\{\Delta_n\}$ is a sequence of partitions as above with mesh $\delta_n \to 0$. Then, almost surely, there exists a measurable, continuous function $f : [0,1] \to [-1,1]$ such that*

$$\limsup_{n\to\infty} \sum_{k=1}^{n} f(t_{k-1}^n)(B(t_k^n) - B(t_{k-1}^n)) = \infty.$$

To prove this, we first prove a positive result of independent interest. We show that Brownian motion has *finite quadratic variation.*

Theorem 7.3 (Quadratic variation) *Suppose $\{\Delta_n\}$ is a sequence of partitions*

$$\Delta_n = \{s = t_0^n < t_1^n < t_2^n < \cdots < t_n^n = t\}$$

with $\Delta_n \subset \Delta_{n+1}$ and mesh $\delta_n \to 0$. Then,

$$\lim_{n\to\infty} \sum_{k=1}^{n} (B(t_k^n) - B(t_{k-1}^n))^2 = t - s \textit{ in the } L^2\textit{-sense.}$$

In particular, a subsequence converges almost surely.

Proof: We have to show that

$$\lim_{n\to\infty} \mathbf{E}\Big\{\Big(\sum_{k=1}^{n} (B(t_k^n) - B(t_{k-1}^n))^2 - (t-s)\Big)^2\Big\} = 0\,.$$

Using the independence of the increments of Brownian motion, one can get that,

$$\begin{aligned}
&\mathbf{E}\Big\{\Big(\sum_{k=1}^{n}(B(t_k^n)-B(t_{k-1}^n))^2-(t-s)\Big)^2\Big\} \\
&= \mathbf{E}\Big\{\Big(\sum_{k=1}^{n}(B(t_k^n)-B(t_{k-1}^n))^2\Big)^2\Big\} - 2(t-s)\mathbf{E}\Big\{\sum_{k=1}^{n}(B(t_k^n)-B(t_{k-1}^n))^2\Big\} + (t-s)^2 \\
&= \sum_{k=1}^{n}\sum_{l=1}^{n}\mathbf{E}\Big\{\Big((B(t_k^n)-B(t_{k-1}^n))^2(B(t_l^n)-B(t_{l-1}^n))^2\Big\} - (t-s)^2 \\
&= \sum_{k=1}^{n}\mathbf{E}\Big\{(B(t_k^n)-B(t_{k-1}^n))^4\Big\} + \sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}(t_k^n-t_{k-1}^n)(t_l^n-t_{l-1}^n) - (t-s)^2 \\
&\leq \sum_{k=1}^{n}\mathbf{E}\Big\{(B(t_k^n)-B(t_{k-1}^n))^4\Big\}.
\end{aligned}$$

For every normally distributed random variable X with $\mu = 0$ we have, by partial integration,

$$\begin{aligned}
\mathbf{E}\{X^4\} &= \frac{2}{\sqrt{2\pi\sigma^2}}\int_0^\infty x^4\exp(-x^2/2\sigma^2)\,dx \\
&= \frac{2}{\sqrt{2\pi\sigma^2}}\int_0^\infty 3x^2\sigma^2\exp(-x^2/2\sigma^2)\,dx \\
&= 3\sigma^2\mathbf{E}\{X^2\} = 3\mathbf{E}\{X^2\}^2.
\end{aligned}$$

We infer that

$$\sum_{k=1}^{n}\mathbf{E}\Big\{(B(t_k^n)-B(t_{k-1}^n))^4\Big\} = 3\sum_{k=1}^{n}(t_k^n-t_{k-1}^n)^2 \leq 3\delta_n(t-s)\to 0\,,$$

which proves the L^2-convergence. It is clear that this implies the existence of an almost surely convergent subsequence. ■

Corollary 7.4 (Unbounded variation) *Suppose $\{\Delta_n\}$ is a sequence of partitions*

$$\Delta_n = \{s = t_0^n < t_1^n < t_2^n < \cdots < t_n^n = t\}$$

of a nondegenerate interval $[s,t]$ with $\Delta_n \subset \Delta_{n+1}$ and mesh $\delta_n \to 0$. Then, almost surely,

$$\limsup_{n\to\infty}\sum_{k=1}^{n}|B(t_k^n)-B(t_{k-1}^n)| = \infty.$$

Proof: By the Hölder property we can find, for any $\alpha \in (0,1/2)$, an n such that $|B(a)-B(b)| \leq |a-b|^\alpha$ for all $a,b\in[s,t]$ with $|a-b|\leq\delta_n$. Using the quadratic variation of Brownian motion in the penultimate step, almost surely,

$$\limsup_{n\to\infty}\sum_{k=1}^{n}|B(t_k^n)-B(t_{k-1}^n)| \geq \limsup_{n\to\infty}\frac{1}{\delta_n^\alpha}\sum_{k=1}^{n}(B(t_k^n)-B(t_{k-1}^n))^2 \geq \lim_{n\to\infty}\frac{(t-s)}{\delta_n^\alpha} = \infty.$$ ■

Proof of Theorem 7.2: Given $\{\Delta_n\}$ we observe that the set of partition points $\bigcup_{n=1}^\infty \Delta_n$ is countable and let $\Omega_1 \subset \Omega$ be a set with $\mathsf{P}(\Omega_1) = 1$ such that the statement of the corollary holds simultaneously for the partitions induced by Δ_n on all intervals bounded by a fixed pair of partition points. We have to construct the function f so that

$$\limsup_{n\to\infty} \sum_{l=1}^{n} f(t_{l-1}^n)(B(t_l^n) - B(t_{l-1}^n)) = \infty\,.$$

By the previous corollary we first find a large $n = n(0)$ such that

$$\sum_{l=1}^{n-1} |B(t_l^n) - B(t_{l-1}^n)| > 3/2,$$

and $|B(t_n^n) - B(t_{n-1}^n)| < 1/2$. Then define f to be constant 1 or -1 on $[t_{l-1}^n, t_l^n)$, according to the sign of the increment $B(t_l^n) - B(t_{l-1}^n)$. Then

$$\sum_{l=1}^{n} f(t_{l-1}^n)(B(t_l^n) - B(t_{l-1}^n)) = \sum_{l=1}^{n-1} |B(t_l^n) - B(t_{l-1}^n)| + f(t_{n-1}^n)(B(t_n^n) - B(t_{n-1}^n)) > 1,$$

if $|f(t_{n-1}^n)| \leq 1$. Now choose $n = n(1)$ so large that, in the remaining interval $I = [t_{n(0)-1}^{n(0)}, t_{n(0)}^{n(0)}]$,

$$\sum_{l=m}^{n} |B(t_l^n) - B(t_{l-1}^n)| > 2 + 1/2,$$

where m is the index with $t_{m-1}^n = t_{n(0)-1}^{n(0)}$, and $|B(t_n^n) - B(t_{n-1}^n)| < 1/2$. On each interval $[t_{l-1}^n, t_l^n)$, for $m \leq l < n$, choose f constant equal to $1/2$ or $-1/2$ according to the sign of the increment. Proceed like this inductively, always refining until the variation in the remaining interval exceeds $2^k + 1/2$ and choosing the value of f on the partition sets from $\pm 2^{-k}$. This defines a measurable function f, bounded by 1, because after k steps the values of f are all determined up to distance 2^{-k} in the sup-norm. Also, for every k the sum

$$\sum_{l=1}^{n} f(t_{l-1}^n)(B(t_l^n) - B(t_{l-1}^n))\,,$$

with $n = n(k)$ is at least k. This proves Theorem 7.2. ■

Altogether, we have seen that this pathwise approach does not lead to a satisfactory notion of an integral with respect to Brownian motion. We have carried out the explicit construction of the bad function f in order to demonstrate one point: f was chosen dependently on the Brownian motion and, moreover, the construction of f on an interval $(0, t)$ required knowledge about the Brownian motion at times $s > t$. As we are interested in integrands which, in a suitable sense, do *not* look in the future, there might be other ways of defining an integral. Starting point of such an approach would be to define exactly the class of integrands we will consider.

The concept of *stochastic integration* is a solution to this problem and permits a powerful extension of martingale theory to a continuous setting. It will turn out that the fact that Brownian motion is of unbounded variation destroys many typical properties of the integral — most notably the change of variables formula. But they are replaced by new ones, so that the *stochastic integrals provides a new type of calculus, the stochastic calculus.* In the next course we shall present the cornerstones of this theory. Stochastic calculus will enable us to obtain deeper insight in the properties of Brownian motion and also define interesting new processes, the *diffusions.* Stochastic calculus allows elegant proofs of nontrivial probabilistic theorems and opens the door to some nontrivial applications, for example in the mathematics of financial markets or stochastic differential equations.

Chapter 8

Exercises

1st Question: An electrical circuit consists of a series of n identical elements. The lifetimes of these elements are independent and with parameter α exponentially distributed, i.e. for the lifetime of the nth element we have a distribution function F_n with $F_n(t) = 1 - e^{-\alpha t}$. Suppose that we have one reserve element of the same type which can instantly replace each element in the series in case of a breakdown.

(a) Show that the distribution function F of the total lifetime of this system satisfies

$$F(t) = \begin{cases} 1 - (n\alpha t + 1)e^{-n\alpha t} & \text{for } t > 0. \\ 0 & \text{for } t \leq 0. \end{cases}$$

(b) What is the expected lifetime of the system?

2nd Question: We choose a random chord on a circle of radius r in the plane and let X be its length. Find an appropriate probability space for the situation and calculate the probability that X is larger than the sidelength of the unilateral triangle inscribed in the circle if

(a) the two endpoints of the chord are chosen independently and uniformly on the circumference of the circle,

(b) the distance of the chord to the origin is uniformly chosen from $(0, r)$,

(c) the centre of the chord is uniformly chosen from the disk?

3rd Question:

(a) Give an example of a measurable space $(\Omega, \mathcal{A})$ and two different probability measures P_1, P_2, which agree on a generator of $\mathcal{A}$.

(b) Suppose that $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ are measurable spaces and $X : \Omega \to \Omega'$ a mapping. Show that if the σ-field $\mathcal{A}'$ is generated by a collection $\mathcal{B} \subset \mathcal{A}'$ and $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, then X is measurable.

4th Question: Suppose $\Omega \subset \mathbb{R}^2$ is a Borel set with $\ell^2(\Omega) = 1$ and consider on the probability space $(\Omega, \mathcal{B}(\Omega), \ell^2|_\Omega)$ the random variables

$$X : (x, y) \mapsto x, \quad Y : (x, y) \mapsto y.$$

Characterize

(a) those sets Ω, for which X and Y are independent,

(b) those sets Ω, for which X and Y are independent and identically distributed.

5th Question: Suppose that $X : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\Omega', \mathcal{A}')$ is measurable. Then, for all $\omega_1 \in \Omega_1$, the mappings

$$X_{\omega_1} : (\Omega_2, \mathcal{A}_2) \to (\Omega', \mathcal{A}'), \qquad X_{\omega_1}(\omega_2) = X(\omega_1, \omega_2)\,,$$

and, for all $\omega_2 \in \Omega_2$, the mappings

$$X^{\omega_2} : (\Omega_1, \mathcal{A}_1) \to (\Omega', \mathcal{A}'), \qquad X^{\omega_2}(\omega_1) = X(\omega_1, \omega_2)\,,$$

are measurable.

Hint: First reduce the problem to the case that X is the indicator function of $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$.

6th Question: Suppose that X is a random point uniformly chosen from the sphere

$$S^2(r) = \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 + z^2 = r^2\}$$

and $P \in \mathbb{R}^3 \setminus S^2(r)$ a fixed point. Define a random variable Z as the inverse of the Euclidean distance of P and X. Find $\mathbb{E}Z$ and $\lim_{P\to\infty} \|P\|\mathbb{E}Z$ and interpret the results.

Hint: By rotation one can assume that $P = (0, 0, p)$ for $p \neq r$. Recall from higher dimensional calculus that the contribution of the upper hemisphere $S^2_+(r)$ parametrized by some $f : A \to S^2_+(r)$ is given by

$$\int_{f(A)} \frac{1}{\|P - x\|}\, d\mathbf{P}_X(x) = \frac{1}{4\pi r^2} \int_A \sqrt{\det Df(y)^T Df(y)}\, \frac{d\ell(y)}{\|P - f(y)\|}\,.$$

7th Question: Let $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ be bounded increasing functions. Prove that, for all random variables X,

$$\mathbf{E}(f(X)g(X)) \geq \mathbf{E}(f(X))\ \mathbf{E}(g(X)).$$

In other words, $f(X)$ and $g(X)$ are *positively correlated*. Interpret this result.

8th Question: Suppose $X_1, X_2, \ldots$ is a sequence of independent random variables with real values, defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Define

$$\mathcal{F}_n = \sigma\{X_i^{-1}(B_i) : B_i \in \mathcal{B},\ 1 \leq i \leq n\}$$

and

$$\mathcal{G}_n = \sigma\{X_i^{-1}(B_i) : B_i \in \mathcal{B},\ i \geq n\},$$

and let $\mathcal{G} = \bigcap_{n=1}^{\infty} \mathcal{G}_n$. $\mathcal{G}$ is the σ-field of tail events.

(a) Give (nontrivial) examples of tail events.

(b) Show that the σ-fields $\mathcal{F}_n$ and $\mathcal{G}_{n+1}$ are independent.
Hint: Define suitable $\cap$-stable generators $\mathcal{E}$ of $\mathcal{F}_n$ and $\mathcal{E}'$ of $\mathcal{G}_{n+1}$ and first show that for $A \in \mathcal{E}$ the measure P_A defined by $P_A(B) = \frac{P(A\cap B)}{P(A)}$ agrees with P on $\mathcal{E}'$ and hence on $\mathcal{G}_{n+1}$.

(c) Use (b) to prove Kolmogorov's Zero-One-Law:
Every $A \in \mathcal{G}$ satisfies $P(A) = 0$ or $P(A) = 1$.

9th Question: A sequence $X_1, X_2, \dots$ of random variables $X_j : (\Omega, \mathcal{A}, \mathbf{P}) \to (\Omega_j, \mathcal{A}_j)$ is called *independent* if for all $A_1, A_2, \dots$ with $A_j \in \mathcal{A}_j$ and $n \in \mathbb{N}$,

$$\mathbf{P}\Big\{X_1 \in A_1, \dots, X_n \in A_n\Big\} = \prod_{j=1}^{n} \mathbf{P}\Big\{X_j \in A_j\Big\}.$$

(a) Show that on a product space $\Omega = \prod_{i=1}^{\infty} \Omega_i$ equipped with a product measure $\mathbf{P} = \bigotimes_{i=1}^{\infty} P_i$ the sequence of projections $X_j(\omega_1, \omega_2, \dots) = \omega_j$ is independent.

(b) For all disjoint sets $A \subset \mathbb{N}$ and $B \subset \mathbb{N}$ define the random variables $X_A = (X_a : a \in A)$ and $X_B = (X_b : b \in B)$ with values in the product spaces $(\Omega_A, \mathcal{A}_A)$ and $(\Omega_B, \mathcal{A}_B)$, where $\Omega_A = \prod_{a\in A} \Omega_a$ and $\mathcal{A}_A = \bigotimes_{a\in A} \mathcal{A}_a$. Show that X_A and X_B are independent.

(c) Give an example of a sequence $X_1, X_2, \dots$ such that for each $i \neq j$ the random variables X_i and X_j are independent, but the sequence is not independent.

10th Question: Suppose $\{X_n\}$ is a Galton-Watson process with offspring distribution given by the sequence $(p_0, p_1, \dots)$. Define a function

$$G : [0,1] \to [0,1], \qquad G(x) = \sum_{n=0}^{\infty} p_n x^n .$$

a) Show that the extinction probability

$$\pi := \mathbf{P}\Big\{ \text{ there is } n \in \mathbb{N} \text{ such that } X_n = 0\Big\}$$

is the smallest fixed point of the mapping G.

Hint: Show that $\mathbf{P}\{X_n = 0\} = G(\mathbf{P}\{X_{n-1} = 0\})$.

b) Show that the extinction probability $\pi = 1$ if $\sum_{n=0}^{\infty} np_n \leq 1$ and $\pi < 1$ otherwise.

11th Question: Suppose U is uniformly distributed on $[0,1]$ and define stochastic processes $\{X_t\}$ by

$$X_t(\omega) = \begin{cases} 1 & \text{if } t = U(\omega)\,, \\ 0 & \text{otherwise,} \end{cases}$$

and Y_t by $Y_t(\omega) = 0$ for all $\omega \in \Omega$. Show that the asociated random functions X and Y on the space $F = \{f : [0,1] \to \mathbb{R}\}$ with the σ-field $\mathcal{F}$ generated by the sets $\{f \in F : f(t) \in A\}$ for $A \in \mathbb{R}$ Borel and $t \in [0,1]$ have the same distribution. Infer from this that the set $\{f \in F : f \text{ is continuous }\}$ is not in the σ-field $\mathcal{F}$.

12th Question: Give an example of a sequence of random variables $X_1, X_2, \dots$ such that

$$\lim_{n\to\infty} X_n = 0 \text{ almost surely, but } \lim_{n\to\infty} \mathbf{E}X_n = 1 .$$

13th Question: Suppose that $\{X(t) : t \geq 0\}$ is a stochastic process modelling the number of customers in a shop. $\{X(t)\}$ should satisfy the following assumptions:

- The number of customers arriving during one time interval does not affect the number of customers arriving in another disjoint time interval. Mathematically, this means that the process has independent increments.
- The rate at which customers arrive should be constant, more precisely, there is some $\lambda \geq 0$ such that $\mathsf{E}[X(t)] = \lambda t$.
- Customers arrive one at a time. To make this precise we assume that $X(t)$ takes values in $\mathbb{N}$, is increasing, and we have

$$\begin{aligned} \mathsf{P}\{X(t+h) = X(t)+1\} &= \lambda h + o(h), \\ \mathsf{P}\{X(t+h) \geq X(t)+2\} &= o(h). \end{aligned}$$

Show that for every $t > s$ the increments of the process $X(t)-X(s)$ are *Poisson distributed with parameter* $\lambda(t-s)$, i.e.

$$\mathsf{P}\{X(t) - X(s) = k\} = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!} .$$

Infer that the process is uniquely determined up to equivalence.

Hint: One could use the Poisson approximation of the binomial distribution.

14th Question: Show that a process $\{X(t) : t \geq 0\}$ with the properties of Question 13 exists. Proceed as follows:

Let S be a Poisson distributed random variable with parameter λ and $Y_1, Y_2, Y_3, \ldots$ independent random variables with uniform distribution on $[0,1]$. For $0 \leq t \leq 1$ let

$$X(t) = \#\{Y_i : Y_i \leq t \text{ and } i \leq S\} .$$

Show that X satisfies the assumptions of Question 13 on the interval $[0,1)$ and extend X to $[0,\infty)$ by glueing together independent copies of X.

15th Question: Let X be Poisson distributed and Y the number of successes in X Bernoulli trials with success probability $p \in (0,1)$. Show that Y and $Z = X - Y$ are independent and Poisson distributed with parameters λp resp. $(1-p)\lambda$.

16th Question: Let $\{X(t) : t \geq 0\}$ be a Poisson process with intensity λ modelling the number of cars arriving at a petrol station up to time t. Each car independently requires Diesel with probability p and petrol otherwise. Let $Y(t)$ be the number of cars requiring Diesel and $Z(t)$ the number of cars requiring petrol up to time t.

(a) Construct a probability space on which the processes $\{Y(t) : t \geq 0\}$ and $\{Z(t) : t \geq 0\}$ can be defined.

(b) Show that $\{Y(t) : t \geq 0\}$ and $\{Z(t) : t \geq 0\}$ are independent Poisson processes with intensity $p\lambda$ resp. $(1-p)\lambda$.

17th Question (Waiting time paradox): At a bus stop the waiting times $T_1, T_2, \ldots$ between two consecutive busses are independent and exponentially distributed with expectation $1/\alpha$.

(a) Show that the exponential distribution has the *lack of memory property*, i.e. for all $s, t \geq 0$,

$$\mathbb{P}\{T_i \geq s + t \,|\, T_i \geq s\} = \mathbf{P}\{T_i \geq t\}.$$

(b) Compute the density of the distribution of the partial sums $S_n = T_1 + \cdots + T_n$.

(c) Mr. Hickleby arrives at time t. What is the expectation $\mathbf{E}W(t)$ of his personal waiting time $W(t)$ for the next bus? Two contradictory answers stand to reason:

- The lack of memory property implies that the distribution of $W(t)$ should not depend on the time of Mr. Hickleby's arrival. So $\mathbf{E}W(t) = 1/\alpha$.
- The time of Mr. Hickleby's arrival is chosen "at random" in the interval between two consecutive arrivals. For reasons of symmetry the expected time should be half the expected time between two arrivals, that is $\mathbf{E}W(t) = 1/2\alpha$.

Interpret your result.

18th Question: Show that every Brownian motion with drift μ and variance parameter σ^2 is a Gaussian process, i.e. given times $t_1 \leq \ldots \leq t_n$ find a matrix A and a vector b such for a standard Gaussian vector X, we have

$$(B(t_1), \ldots, B(t_n)) = AX + b.$$

19th Question: Suppose that B is a standard Brownian motion. Show that, for every $\varepsilon > 0$, $\mathbf{P}\{B \text{ has a zero in } (0, \varepsilon)\} > 0$.

20th Question: Let $\alpha > 1/2$. Show that, almost surely, for every $t > 0$, there exists $h > 0$ such that $|B(t+h) - B(t)| > h^\alpha$.

21st Question: Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ a sub-σ-field and X a random variable with $\mathbf{E}|X| < \infty$. Then every conditional probability $\mathbf{E}\{X|\mathcal{F}\}$ has the following properties.

Positivity If $X \geq 0$ almost surely, then $\mathbf{E}\{X|\mathcal{F}\} \geq 0$ almost surely.

Monotone Convergence If $0 \leq X_n \uparrow X$, then $\mathbf{E}\{X_n|\mathcal{F}\} \uparrow \mathbf{E}\{X|\mathcal{F}\}$ almost surely.

Fatou If $0 \leq X_n$ and $\mathbf{E}\{X_n|\mathcal{F}\} < \infty$, then

$$\mathbf{E}\{\liminf_{n\to\infty} X_n \,|\, \mathcal{F}\} \leq \liminf_{n\to\infty} \mathbf{E}\{X_n \,|\, \mathcal{F}\} \text{ almost surely.}$$

Dominated Convergence If there is a random variable Z such that $\mathbf{E}Z < \infty$ and $|X_n| \leq Z$ for all n, and if $X_n \to X$ almost surely, then $\mathbf{E}\{X_n|\mathcal{F}\} \to \mathbf{E}\{X|\mathcal{F}\}$ almost surely.

22nd Question: We investigate conditional expectations in the situation when two real valued random variables X and Z have a jont density $f(x, z)$.

a) Show that $f_X : x \mapsto \int f(x,z)\,dz$ is a density of X. Then, clearly, the function $f_Z : z \mapsto \int f(x,z)\,dx$ is a density of Z.

b) Assume that

$$\mathbf{E}|X| = \int |x| f_X(x)\,dx = \int |x| \int f(x,z)\,dz\,dx < \infty\,.$$

One can define a function

$$f_{X|Z}(x|z) = \begin{cases} f(x,z)/f_Z(z) & \text{if } f_Z(z) \neq 0\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

With this function define a random variable

$$Y(\omega) = \int x f_{X|Z}(x|Z(\omega))\,dx\,.$$

Show that Y is a conditional expectation of X given Z.

23rd Question: Suppose that B is a standard Brownian motion. Show that, almost surely,

$$\limsup_{t\to\infty} |B(t)/\sqrt{t}| = \infty\,.$$

24th Question: Let $p \geq 1$. If $\{X_n\}$ is a sequence of random variables on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ such that $\mathbf{E}|X_n|^p < \infty$ for all n and

$$\lim_{k\to\infty} \sup_{n,m\geq k} \mathbf{E}|X_n - X_m|^p = 0\,,$$

then there is a random variable X on $(\Omega, \mathcal{A}, \mathbf{P})$ such that

$$\lim_{n\to\infty} \mathbf{E}|X_n - X|^p = 0\,.$$

Hint: Show that a subsequence of $\{X_n\}$ converges almost surely.

25th Question: Suppose that $X : (\Omega, \mathcal{A}, \mathbf{P}) \to (\Omega', \mathcal{A}')$ is a random variable and $\mathcal{G}, \mathcal{H} \subset \mathcal{A}$ are sub-σ-fields, such that $\mathcal{H}$ and $\mathcal{G} \vee X^{-1}(\mathcal{A}')$ are independent. Then, almost surely,

$$\mathbf{E}\{X|\mathcal{G} \vee \mathcal{H}\} = \mathbf{E}\{X|\mathcal{G}\}\,.$$

26th Question: Suppose that $\{X_i\}$ is a sequence of independent and identically distributed random variables with $\mathbf{E}|X_1| < \infty$ and denote $S_n = X_1 + \cdots + X_n$.

a) Show that, for each n,

$$\mathbf{E}\{X_1|S_n\} = \ldots = \mathbf{E}\{X_n|S_n\} = \frac{S_n}{n}\,.$$

b) Use the answer to the 25th question to show that this implies

$$\mathbf{E}\{X_1|S_n, S_{n+1}, \ldots\} = \frac{S_n}{n}\,.$$

27th Question (Kolmogorov-01-law for Brownian motion): Define $\mathcal{G}(t)$ to be the σ-field defined by the random variables $B(s)$ for $t \le s$. $\mathcal{G}(t)$ describes the future at time t of the Brownian motion. Let $\mathcal{T} = \bigcap_{t\ge 0} \mathcal{G}(t)$ be the σ-field of all *tail events*.

a) Give nontrivial examples of tail events.
b) Show that, for all $x \in \mathbb{R}$ and $A \in \mathcal{T}$, $\mathbb{P}_x\{A\} \in \{0,1\}$.
c) Use this statement to show that Brownian motion is *recurrent*, i.e. that, almost surely, for every large n there is a time $t > n$ such that $B(t) = 0$.

28th Question: Let $\{B(t) : t \ge 0\}$ be a Brownian motion. Show that for a stopping time T, the random variable $B(T)$ is $\mathcal{F}(T)$-measurable.

Hint: Approximate T from above by stopping times T_n taking values in the set $\{k/2^n : k \ge 1\}$.

29th Question: Show that, almost surely, the Brownian sample path $t \mapsto B(t)$ is monotone in no interval.

30th Question: For every $a \ge 0$ we define

$$T(a) = \inf\{t \ge 0 \ : \ B(t) = a\}\,.$$

a) Show that $T(a)$ is an almost surely finite stopping time.
b) Show that the process $\{T(a) \ : \ a \ge 0\}$ has stationary, independent increments.
c) Show that, for all $\lambda > 0$, the process $\{\lambda^2 T(a/\lambda) \ : \ a \ge 0\}$ has the same distribution as the process $\{T(a) \ : \ a \ge 0\}$.

An increasing process $\{T(a) \ : \ a \ge 0\}$ with the properties described in b) and c) is called a *stable subordinator of index* $1/2$.

31st Question: Suppose that $\{X_n\}$ is a simple random walk started at $x \in \{0, \dots, N\}$ and define

$$T := \inf\{n \ge 0 \ : \ X_n = 0 \text{ or } X_n = N\}\,.$$

Show that $\mathbb{E}\{T\} = x(N-x)$.

Hint: Use induction with respect to N and some form of the optional stopping theorem.

32nd Question: Suppose that $\{X_n\}$ is a martingale with $|X_{n+1} - X_n| \le 1$ for all n. Show that, with probability one,

- *either* $\{X_n\}$ converges to a finite limit,
- *or* $\limsup_{n\to\infty} X_n = \infty$ and $\liminf_{n\to\infty} X_n = -\infty$.

33rd Question: (An improved Borel-Cantelli-Lemma) Let $\{\mathcal{F}(n)\}$ be a filtration with $\mathcal{F}(0) = \{\emptyset, \Omega\}$ and $A_n \in \mathcal{F}(n)$ a sequence of events. Show that infinitely many of the events A_n take place if and only if

$$\sum_{n=1}^{\infty} \mathbb{P}\{A_n \,|\, \mathcal{F}(n-1)\} = \infty\,.$$

Hint: Show that $X_n = \sum_{m=1}^{n} (1_{A_m} - \mathbb{P}\{A_m | \mathcal{F}(m-1)\})$ is a martingale satisfying the requirement of Question 32. Then look at the two cases separately.

34th Question: Suppose a box contains initially one black and one white ball. At each step we pick a ball at random from the box and return it together with another ball of the same colour. After n steps the box contains $n+2$ balls and we denote by M_n the proportion of white balls in the box.

a) Show that $\{M_n : n \in \mathbb{N}\}$ is a martingale.

b) Show that M_n converges almost surely and determine the distribution of the limiting random variable.

35th Question (Optional Sampling Theorem): Suppose that $\{X_n\}$ is a uniformly integrable martingale and T an almost surely finite stopping time with $\mathbf{E}|X_T| < \infty$. Show that $\mathbf{E}\{X_T\} = \mathbf{E}\{X_0\}$.

36th Question (Hitting probabilities for biased random walks): Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables with

$$\mathbf{P}\{X_n = 1\} = p \text{ and } \mathbf{P}\{X_n = -1\} = 1-p\,,$$

for some $0 < p < 1/2$. Let N be a positive integer and $a \in \{0, \ldots, N\}$. Define the biased random walk with start in a as

$$S_n = a + \sum_{k=1}^{n} X_k\,.$$

Let T be the first time that $\{S_n\}$ reaches either 0 or N.

(a) Let $M_n = [(1-p)/p]^{S_n}$. Show that $\{M_{n\wedge T}\}$ is a uniformly integrable martingale.

(b) Use the optional sampling theorem to compute the probabilities $\mathbf{P}\{S_T = 0\}$ and $\mathbf{P}\{S_T = N\}$.

37th Question (Wald's equation): Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed, integrable random variables with mean μ. Let $\mathcal{F}(n)$ be the natural filtration for $\{X_n\}$ and T be a stopping time with respect to this filtration with $\mathbf{E}\{T\} < \infty$.

(a) Let $Y = \sum_{n=1}^{T} |X_n|$. Show that $\mathbf{E}|Y| < \infty$.

(b) Let $T_n = T \wedge n$ and define $M_n = X_1 + \cdots + X_{T_n} - \mu T_n$.
Prove that $\{M_n\}$ is a uniformly integrable martingale.

(c) Prove Wald's equation

$$\mathbf{E}\Big\{\sum_{n=1}^{T} X_n\Big\} = \mu \mathbf{E}\{T\}\,.$$

38th Question (Doob decomposition): Let $\{X_n\}$ be an adapted process with $\mathbf{E}|X_n| < \infty$ for all n. Then there is a martingale $\{M_n\}$ and a previsible process $\{C_n\}$ such that $X_n = M_n + C_n$.

39th Question: Let $\{M_n\}$ be a martingale with $\mathbf{E}\{M_n^2\} < \infty$ for all n. Show that if

$$\sum_{n=1}^{\infty} \mathbf{E}\{(M_n - M_{n-1})^2\} < \infty,$$

there is a random variable M such that $\lim_{n\to\infty} M_n = M$ almost surely and in L^2.

40th Question: Suppose that $\{X_k\}$ is a sequence of independent random variables with $\mathsf{E}\{X_k\} = 0$ and finite variance σ_k^2 for all k.

(a) Show that $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$ implies that the random series $\sum_{k=1}^{\infty} X_k$ converges almost surely.

(b) Suppose that there is a $K > 0$ such that $|X_n| \leq K$ for all n.
Show that if $\sum_{k=1}^{\infty} X_k$ converges almost surely, then $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$.
Hint: Show that $N_n = (\sum_{k=1}^{n} X_k)^2 - \sum_{k=1}^{n} \sigma_k^2$ defines a martingale.

41st Question: Suppose that $\{B(t) \,:\, t \geq 0\}$ is a standard Brownian motion.

(a) For $a < 0 < b$ define the stopping time $T = \inf\{t \geq 0 \,:\, B(t) \notin (a, b)\}$ and show that $\mathsf{E}\{T\} = -ab$.

(b) For $a > 0$ define the stopping time $T = \inf\{t \geq 0 \,:\, B(t) > a\}$ and show that $\mathsf{E}\{T\} = \infty$.

42nd Question: For two probability distributions P and Q on $\mathbb{R}$ with the associated distribtuion functions F and G define the *Lévy distance* as

$$d(P, Q) = \inf\{\epsilon > 0 \,:\, F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon \text{ for all } x\}.$$

(a) Show that d defines a metric on the set $\mathrm{Prob}(\mathbb{R})$.

(b) Show that $\lim_{n\to\infty} d(P_n, P) = 0$ if and only if $\{P_n\}$ converges weakly to P.

Recommended Reading:

On the measure theoretic background:

P. HALMOS Measure Theory, 1950.

J. ELSTRODT Maß- und Integrationstheorie, 1996.

On probability theory:

R. DURRETT Probability: Theory and examples, 1996.

G. LAWLER Introduction to stochastic processes, 1995.

D. WILLIAMS Probability with martingales, 1995.

D. FREEDMAN Brownian motion and diffusion, 1995.

Y. PERES Path properties of Brownian motion, 1998.

Other sources:

J. NEVEU Arbres et processus de Galton-Watson. Ann. Inst. H. Poincare, Probab. et Stat. 22 (1986) 199-207.

F.B. KNIGHT Essentials of Brownian motion and diffusion. AMS Surveys, Number 18.