ADVANCED PROBABILITY

JAMES NORRIS

Contents

Date: October 3, 2021.

0. Review of measure and integration

This review covers briefly some notions which are discussed in detail in my notes on Probability and Measure (from now on [PM]), Sections 1 to 3.

0.1. **Measurable spaces.** Let E be a set. A set $\mathcal{E}$ of subsets of E is called a *σ-algebra on E* if it contains the empty set $\emptyset$ and, for all $A \in \mathcal{E}$ and every sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{E}$,

$$E \setminus A \in \mathcal{E}, \quad \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}.$$

Let $\mathcal{E}$ be a σ-algebra on E. A pair such as $(E, \mathcal{E})$ is called a *measurable space*. The elements of $\mathcal{E}$ are called *measurable sets*. A function $\mu : \mathcal{E} \to [0, \infty]$ is called a *measure on* $(E, \mathcal{E})$ if $\mu(\emptyset) = 0$ and, for every sequence $(A_n : n \in \mathbb{N})$ of disjoint sets in $\mathcal{E}$,

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

A triple such as $(E, \mathcal{E}, \mu)$ is called a *measure space*.

Given a set E which is equipped with a topology, the *Borel σ-algebra on E* is the smallest σ-algebra containing all the open sets. We denote this σ-algebra by $\mathcal{B}(E)$ and call its elements *Borel sets*. We use this construction most often in the cases where E is the real line $\mathbb{R}$ or the extended half-line $[0, \infty]$. We write $\mathcal{B}$ for $\mathcal{B}(\mathbb{R})$.

0.2. **Integration of measurable functions.** Given measurable spaces $(E, \mathcal{E})$ and $(E', \mathcal{E}')$ and a function $f : E \to E'$, we say that f is *measurable* if $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{E}'$. If we refer to a measurable function f on $(E, \mathcal{E})$ without specifying its range then, by default, we take $E' = \mathbb{R}$ and $\mathcal{E}' = \mathcal{B}$. By a *non-negative measurable function on E* we mean any function $f : E \to [0, \infty]$ which is measurable when we use the Borel σ-algebra on $[0, \infty]$. Note that we allow the value ∞ for non-negative measurable functions but not for real-valued measurable functions. We denote the set of real-valued measurable functions by $m\mathcal{E}$ and the set of non-negative measurable functions by $m\mathcal{E}^+$.

Theorem 0.2.1. *Let $(E, \mathcal{E}, \mu)$ be a measure space. There exists a unique map $\tilde{\mu} : m\mathcal{E}^+ \to [0, \infty]$ with the following properties*

(a) $\tilde{\mu}(1_A) = \mu(A)$ *for all* $A \in \mathcal{E}$,
(b) $\tilde{\mu}(\alpha f + \beta g) = \alpha\tilde{\mu}(f) + \beta\tilde{\mu}(g)$ *for all* $f, g \in m\mathcal{E}^+$ *and all* $\alpha, \beta \in [0, \infty)$,
(c) $\tilde{\mu}(f_n) \to \tilde{\mu}(f)$ *as* $n \to \infty$ *whenever* $(f_n : n \in \mathbb{N})$ *is a non-decreasing sequence in* $m\mathcal{E}^+$ *with pointwise limit* f.

The map $\tilde{\mu}$ is called the *integral with respect to μ*. From now on, we simply write μ instead of $\tilde{\mu}$. We say that f is a *simple function* if it is a finite linear combination of indicator functions of measurable sets, with positive coefficients. Thus f is a simple function if there exist $n \geq 0$, and $\alpha_k \in (0, \infty)$ and $A_k \in \mathcal{E}$ for $k = 1, \ldots, n$, such that

$$f = \sum_{k=1}^{n} \alpha_k 1_{A_k}.$$

Note that properties (a) and (b) force the integral of such a simple function f to be

$$\mu(f) = \sum_{k=1}^{n} \alpha_k \mu(A_k).$$

Note also that property (b) implies that $\mu(f) \leq \mu(g)$ whenever $f \leq g$.

Property (c) is called *monotone convergence*. Given $f \in m\mathcal{E}^+$, we can define a non-decreasing sequence of simple functions $(f_n : n \in \mathbb{N})$ by

$$f_n(x) = (2^{-n}\lfloor 2^n f(x) \rfloor) \wedge n, \quad x \in E.$$

Then $f_n(x) \to f(x)$ as $n \to \infty$ for all $x \in E$. So, by monotone convergence, we have

$$\mu(f) = \lim_{n\to\infty} \mu(f_n).$$

We have proved the uniqueness statement in Theorem 0.2.1.

For measurable functions f and g, we say that $f = g$ *almost everywhere* if

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0.$$

It is straightforward to see that, for $f \in m\mathcal{E}^+$, we have $\mu(f) = 0$ if and only if $f = 0$ almost everywhere.

Lemma 0.2.2 (Fatou's lemma). *Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions. Then*

$$\mu\left(\liminf_{n\to\infty} f_n\right) \leq \liminf_{n\to\infty} \mu(f_n).$$

The proof is by applying monotone convergence to the non-decreasing sequence of functions $(\inf_{m\geq n} f_m : n \in \mathbb{N})$.

Given a (real-valued) measurable function f, we say that f is *integrable with respect to* μ if $\mu(|f|) < \infty$. We write $L^1(E, \mathcal{E}, \mu)$ for the set of such integrable functions, or simply L^1 when the choice of measure space is clear. The integral is extended to L^1 by setting

$$\mu(f) = \mu(f^+) - \mu(f^-)$$

where $f^{\pm} = (\pm f) \vee 0$. Then L^1 is a vector space and the map $\mu : L^1 \to \mathbb{R}$ is linear.

Theorem 0.2.3 (Dominated convergence). *Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions. Suppose that $f_n(x)$ converges as $n \to \infty$, with limit $f(x)$, for all $x \in E$. Suppose further that there exists an integrable function g such that $|f_n| \leq g$ for all n. Then f_n is integrable for all n, and so is f, and $\mu(f_n) \to \mu(f)$ as $n \to \infty$.*

The proof is by applying Fatou's lemma to the two sequences of non-negative measurable functions $(g \pm f_n : n \in \mathbb{N})$.

0.3. **Product measure and Fubini's theorem.** Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be finite (or σ-finite) measure spaces. The *product σ-algebra* $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2$ is the σ-algebra on $E_1 \times E_2$ generated by subsets of the form $A_1 \times A_2$ for $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$.

Theorem 0.3.1. *There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $\mathcal{E}$ such that, for all $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$,*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

Theorem 0.3.2 (Fubini's theorem)**.** *Let f be a non-negative $\mathcal{E}$-measurable function on E. For $x_1 \in E_1$, define a function f_{x_1} on E_2 by $f_{x_1}(x_2) = f(x_1, x_2)$. Then f_{x_1} is $\mathcal{E}_2$-measurable for all $x_1 \in E_1$. Hence, we can define a function f_1 on E_1 by $f_1(x_1) = \mu_2(f_{x_1})$. Then f_1 is $\mathcal{E}_1$-measurable and $\mu_1(f_1) = \mu(f)$.*

By some routine arguments, it is not hard to see that $\mu(f) = \hat{\mu}(\hat{f})$, where $\hat{\mu} = \mu_2 \otimes \mu_1$ and $\hat{f}$ is the function on $E_2 \times E_1$ given by $\hat{f}(x_2, x_1) = f(x_1, x_2)$. Hence, with obvious notation, it follows from Fubini's theorem that, for any non-negative $\mathcal{E}$-measurable function f, we have $\mu_1(f_1) = \mu_2(f_2)$. This is more usually written as

$$\int_{E_1} \left(\int_{E_2} f(x_1, x_2)\mu_2(dx_2) \right) \mu_1(dx_1) = \int_{E_2} \left(\int_{E_1} f(x_1, x_2)\mu_1(dx_1) \right) \mu_2(dx_2).$$

We refer to [PM, Section 3.6] for more discussion, in particular for the case where the assumption of non-negativity is replaced by one of integrability.

1. Conditional expectation

We say that $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space* if it is a measure space with the property that $\mathbb{P}(\Omega) = 1$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The elements of $\mathcal{F}$ are called *events* and $\mathbb{P}$ is called a *probability measure*. A measurable function X on $(\Omega, \mathcal{F})$ is called a *random variable*. The integral of a random variable X with respect to $\mathbb{P}$ is written $\mathbb{E}(X)$ and is called the *expectation of* X. We use *almost surely* to mean almost everywhere in this context.

A probability space gives us a mathematical framework in which to model probabilities of events subject to randomness and average values of random quantities. It is often natural also to take a partial average, which may be thought of as integrating out some variables and not others. This is made precise in greatest generality in the notion of conditional expectation. We first give three motivating examples, then establish the notion in general, and finally discuss some of its properties.

1.1. **Discrete case.** Let $(G_n : n \in \mathbb{N})$ be sequence of disjoint events, whose union is Ω. Set

$$\mathcal{G} = \sigma(G_n : n \in \mathbb{N}) = \{\cup_{n \in I} G_n : I \subseteq \mathbb{N}\}.$$

For any integrable random variable X, we can define

$$Y = \sum_{n \in \mathbb{N}} \mathbb{E}(X|G_n) 1_{G_n}$$

where we set $\mathbb{E}(X|G_n) = \mathbb{E}(X1_{G_n})/\mathbb{P}(G_n)$ when $\mathbb{P}(G_n) > 0$ and set $\mathbb{E}(X|G_n) = 0$ when $\mathbb{P}(G_n) = 0$. It is easy to check that Y has the following two properties

(a) Y is $\mathcal{G}$-measurable,
(b) Y is integrable and $\mathbb{E}(X1_A) = \mathbb{E}(Y1_A)$ for all $A \in \mathcal{G}$.

1.2. **Gaussian case.** Let (W, X) be a Gaussian random variable in $\mathbb{R}^2$. Set

$$\mathcal{G} = \sigma(W) = \{\{W \in B\} : B \in \mathcal{B}\}.$$

Write $Y = aW + b$, where $a, b \in \mathbb{R}$ are chosen to satisfy

$$a\mathbb{E}(W) + b = \mathbb{E}(X), \quad a \operatorname{var} W = \operatorname{cov}(W, X).$$

Then $\mathbb{E}(X - Y) = 0$ and

$$\operatorname{cov}(W, X - Y) = \operatorname{cov}(W, X) - \operatorname{cov}(W, Y) = 0$$

so W and $X - Y$ are independent. Hence Y satisfies

(a) Y is $\mathcal{G}$-measurable,
(b) Y is integrable and $\mathbb{E}(X1_A) = \mathbb{E}(Y1_A)$ for all $A \in \mathcal{G}$.

1.3. **Conditional density functions.** Suppose that U and V are random variables having a joint density function $f_{U,V}(u, v)$ in $\mathbb{R}^2$. Then U has density function f_U given by

$$f_U(u) = \int_{\mathbb{R}} f_{U,V}(u, v)\, dv.$$

The *conditional density function* $f_{V|U}(v|u)$ of V given U is defined by

$$f_{V|U}(v|u) = f_{U,V}(u, v)/f_U(u)$$

where interpret $0/0$ as 0 if necessary. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function and suppose that $X = h(V)$ is integrable. Let

$$g(u) = \int_{\mathbb{R}} h(v) f_{V|U}(v|u)\, dv.$$

Set $\mathcal{G} = \sigma(U)$ and $Y = g(U)$. Then Y satisfies

(a) Y is $\mathcal{G}$-measurable,
(b) Y is integrable and $\mathbb{E}(X1_A) = \mathbb{E}(Y1_A)$ for all $A \in \mathcal{G}$.

To see (b), note that every $A \in \mathcal{G}$ takes the form $A = \{U \in B\}$, for some Borel set B. Then, by Fubini's theorem,

$$\begin{aligned}\mathbb{E}(X1_A) &= \int_{\mathbb{R}^2} h(v) 1_B(u) f_{U,V}(u, v)\, du dv \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(v) f_{V|U}(v|u)\, dv \right) f_U(u) 1_B(u)\, du = \mathbb{E}(Y1_A).\end{aligned}$$

1.4. **Existence and uniqueness.** We will use in this subsection the Hilbert space structure of the set L^2 of square integrable random variables. See [PM, Section 5] for details.

Theorem 1.4.1. *Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Then there exists a random variable Y such that*

(a) *Y is $\mathcal{G}$-measurable,*
(b) *Y is integrable and $\mathbb{E}(X1_A) = \mathbb{E}(Y1_A)$ for all $A \in \mathcal{G}$.*

Moreover, if Y' also satisfies (a) *and* (b)*, then $Y = Y'$ almost surely.*

The same statement holds with 'integrable' replaced by 'non-negative' throughout. We leave this extension as an exercise. We call Y (*a version of*) *the conditional expectation of* X *given* $\mathcal{G}$ and write

$$Y = \mathbb{E}(X|\mathcal{G}) \quad \text{almost surely.}$$

In the case where $\mathcal{G} = \sigma(G)$ for some random variable G, we also write $Y = \mathbb{E}(X|G)$ almost surely. In the case where $X = 1_A$ for some event A, we write $Y = \mathbb{P}(A|\mathcal{G})$ almost surely. The preceding three examples show how to construct explicit versions of the conditional expectation in certain simple cases. In general, we have to live with the indirect approach provided by the theorem.

Proof. (*Uniqueness.*) Suppose that Y satisfies (a) and (b) and that Y' satisfies (a) and (b) for another integrable random variable X', with $X \leq X'$ almost surely. Consider the non-negative random variable $Z = (Y - Y')1_A$, where $A = \{Y \geq Y'\} \in \mathcal{G}$. Then

$$\mathbb{E}(Y1_A) = \mathbb{E}(X1_A) \leq \mathbb{E}(X'1_A) = \mathbb{E}(Y'1_A) < \infty$$

so $\mathbb{E}(Z) \leq 0$ and so $Z = 0$ almost surely, which implies that $Y \leq Y'$ almost surely. In the case $X = X'$, we deduce that $Y = Y'$ almost surely.

(*Existence.*) Assume for now that $X \in L^2(\mathcal{F})$. Since $L^2(\mathcal{G})$ is complete, it is a closed subspace of $L^2(\mathcal{F})$, so X has an orthogonal projection Y on $L^2(\mathcal{G})$, that is, there exists $Y \in L^2(\mathcal{G})$ such that $\mathbb{E}((X - Y)Z) = 0$ for all $Z \in L^2(\mathcal{G})$. In particular, for any $A \in \mathcal{G}$, we can take $Z = 1_A$ to see that $\mathbb{E}(X1_A) = \mathbb{E}(Y1_A)$. Thus Y satisfies (a) and (b).

Assume now that $X \geq 0$. Then $X_n = X \wedge n \in L^2(\mathcal{F})$ and $0 \leq X_n \uparrow X$ as $n \to \infty$. We have shown, for each n, that there exists $Y_n \in L^2(\mathcal{G})$ such that, for all $A \in \mathcal{G}$,

$$\mathbb{E}(X_n 1_A) = \mathbb{E}(Y_n 1_A)$$

and moreover that $0 \leq Y_n \leq Y_{n+1}$ almost surely. Define

$$\Omega_0 = \{\omega \in \Omega : 0 \leq Y_n(\omega) \leq Y_{n+1}(\omega) \text{ for all } n\}$$

and set $Y_\infty = \lim_{n\to\infty} Y_n 1_{\Omega_0}$. Then Y_∞ is a non-negative $\mathcal{G}$-measurable random variable and, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathbb{E}(X1_A) = \mathbb{E}(Y_\infty 1_A).$$

In particular, since X is integrable, we have $\mathbb{E}(Y_\infty) = \mathbb{E}(X) < \infty$ so $Y_\infty < \infty$ almost surely. Set $Y = Y_\infty 1_{\{Y_\infty < \infty\}}$. Then Y is a random variable satisfying (a) and (b).

Finally, for a general integrable random variable X, we can apply the preceding construction to X^- and X^+ to obtain Y^- and Y^+. Then $Y = Y^+ - Y^-$ satisfies (a) and (b). $\square$

1.5. **Properties of conditional expectation.** Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. The following properties follow directly from Theorem 1.4.1

(i) $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$,
(ii) *if* X *is* $\mathcal{G}$*-measurable, then* $\mathbb{E}(X|\mathcal{G}) = X$ *almost surely,*
(iii) *if* X *is independent of* $\mathcal{G}$*, then* $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ *almost surely.*

In the proof of Theorem 1.4.1, we showed also

(iv) *if* $X \geq 0$ *almost surely, then* $\mathbb{E}(X|\mathcal{G}) \geq 0$ *almost surely.*

Next, for $\alpha, \beta \in \mathbb{R}$ and any integrable random variable Y, we have

(v) $\mathbb{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha\mathbb{E}(X|\mathcal{G}) + \beta\mathbb{E}(Y|\mathcal{G})$ *almost surely.*

To see this, one checks that the right hand side satisfies the properties (a) and (b) from Theorem 1.4.1 which characterize the left hand side.

The basic convergence theorems for expectation have counterparts for conditional expectation. Consider a sequence of random variables X_n in the limit $n \to \infty$. If $0 \leq X_n \uparrow X$ almost surely, then $\mathbb{E}(X_n|\mathcal{G}) \uparrow Y$ almost surely, for some $\mathcal{G}$-measurable random variable Y; so, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathbb{E}(X1_A) = \lim \mathbb{E}(X_n 1_A) = \lim \mathbb{E}(\mathbb{E}(X_n|\mathcal{G})1_A) = \mathbb{E}(Y1_A),$$

which implies that $Y = \mathbb{E}(X|\mathcal{G})$ almost surely. We have proved the conditional monotone convergence theorem:

(vi) *if* $0 \leq X_n \uparrow X$ *almost surely, then* $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$ *almost surely.*

Next, by essentially the same arguments used for the original results, we can deduce conditional forms of Fatou's lemma and the dominated convergence theorem

(vii) *if* $X_n \geq 0$ *for all* n*, then* $\mathbb{E}(\liminf X_n|\mathcal{G}) \leq \liminf \mathbb{E}(X_n|\mathcal{G})$ *almost surely,*
(viii) *if* $X_n \to X$ *and* $|X_n| \leq Y$ *for all* n*, almost surely, for some integrable random variable* Y*, then* $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ *almost surely.*

There is a conditional form of Jensen's inequality. Let $c : \mathbb{R} \to (-\infty, \infty]$ be a convex function. Then c is the supremum of a sequence of affine functions

$$c(x) = \sup_{n \in \mathbb{N}} (a_n x + b_n), \quad x \in \mathbb{R}.$$

Hence, $\mathbb{E}(c(X)|\mathcal{G})$ is well defined and, almost surely, for all n,

$$\mathbb{E}(c(X)|\mathcal{G}) \geq a_n \mathbb{E}(X|\mathcal{G}) + b_n.$$

On taking the supremum over $n \in \mathbb{N}$ in this inequality, we obtain

(ix) *if* $c : \mathbb{R} \to (-\infty, \infty]$ *is convex, then* $\mathbb{E}(c(X)|\mathcal{G}) \geq c(\mathbb{E}(X|\mathcal{G}))$ *almost surely.*

In particular, for $1 \leq p < \infty$,

$$\|\mathbb{E}(X|\mathcal{G})\|_p^p = \mathbb{E}(|\mathbb{E}(X|\mathcal{G})|^p) \leq \mathbb{E}(\mathbb{E}(|X|^p|\mathcal{G})) = \mathbb{E}(|X|^p) = \|X\|_p^p.$$

So we have

(x) $\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$ *for all* $1 \leq p < \infty$.

For any σ-algebra $\mathcal{H} \subseteq \mathcal{G}$, the random variable $Y = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})$ is $\mathcal{H}$-measurable and satisfies, for all $A \in \mathcal{H}$

$$\mathbb{E}(Y1_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_A) = \mathbb{E}(X1_A)$$

so we have the *tower property*

(xi) *if* $\mathcal{H} \subseteq \mathcal{G}$*, then* $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H})$ *almost surely.*

We can always *take out what is known*

(xii) *if* Y *is bounded and* $\mathcal{G}$*-measurable, then* $\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$ *almost surely.*

To see this, consider first the case where $Y = 1_B$ for some $B \in \mathcal{G}$. Then, for $A \in \mathcal{G}$,

$$\mathbb{E}(Y\mathbb{E}(X|\mathcal{G})1_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_{A\cap B}) = \mathbb{E}(X1_{A\cap B}) = \mathbb{E}(YX1_A),$$

which implies that $\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$ almost surely. The result extends to simple $\mathcal{G}$-measurable random variables Y by linearity, then to the case $X \geq 0$ and any bounded non-negative $\mathcal{G}$-measurable random variable Y by monotone convergence. The general case follows by writing $X = X^+ - X^-$ and $Y = Y^+ - Y^-$.

Finally,

(xiii) *if $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$, then $\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G})$ almost surely.*

For, suppose $A \in \mathcal{G}$ and $B \in \mathcal{H}$, then

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X|\sigma(\mathcal{G},\mathcal{H}))1_{A\cap B}) &= \mathbb{E}(X1_{A\cap B}) \\ &= \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_A)\mathbb{P}(B) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_{A\cap B}).\end{aligned}$$

The set of such intersections $A \cap B$ is a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$, so the desired formula follows from [PM, Proposition 3.1.4].

Lemma 1.5.1. *Let $X \in L^1$. Then the set of random variables Y of the form $Y = \mathbb{E}(X|\mathcal{G})$, where $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra, is uniformly integrable.*

Proof. By [PM, Lemma 6.2.1], given $\varepsilon > 0$, we can find $\delta > 0$ so that $\mathbb{E}(|X|1_A) \leq \varepsilon$ whenever $\mathbb{P}(A) \leq \delta$. Then choose $\lambda < \infty$ so that $\mathbb{E}(|X|) \leq \lambda\delta$. Suppose $Y = \mathbb{E}(X|\mathcal{G})$, then $|Y| \leq \mathbb{E}(|X||\mathcal{G})$. In particular, $\mathbb{E}(|Y|) \leq \mathbb{E}(|X|)$ so

$$\mathbb{P}(|Y| \geq \lambda) \leq \lambda^{-1}\mathbb{E}(|Y|) \leq \delta.$$

Then

$$\mathbb{E}(|Y|1_{|Y|\geq\lambda}) \leq \mathbb{E}(|X|1_{|Y|\geq\lambda}) \leq \varepsilon.$$

Since λ was chosen independently of $\mathcal{G}$, we are done. $\square$

2. Martingales in discrete time

2.1. **Definitions.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a *filtration*, that is to say, a sequence $(\mathcal{F}_n)_{n\geq 0}$ of σ-algebras such that, for all $n \geq 0$,

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}.$$

Set

$$\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0).$$

Then $\mathcal{F}_\infty \subseteq \mathcal{F}$. We allow the possibility that $\mathcal{F}_\infty \neq \mathcal{F}$. We interpret the parameter n as time, and the σ-algebra $\mathcal{F}_n$ as the extent of our knowledge at time n.

By a *random process* (*in discrete time*) we mean a sequence of random variables $(X_n)_{n\geq 0}$. Each random process $X = (X_n)_{n\geq 0}$ has a *natural filtration* $(\mathcal{F}_n^X)_{n\geq 0}$, given by

$$\mathcal{F}_n^X = \sigma(X_0, \dots, X_n).$$

Then $\mathcal{F}_n^X$ models what we know about X by time n. We say that $(X_n)_{n\geq 0}$ is *adapted* (*to* $(\mathcal{F}_n)_{n\geq 0}$) if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. It is equivalent to require that $\mathcal{F}_n^X \subseteq \mathcal{F}_n$ for all n. In this section we consider only real-valued or non-negative random processes. We say that $(X_n)_{n\geq 0}$ is *integrable* if X_n is an integrable random variable for all $n \geq 0$.

A *martingale* is an adapted integrable random process $(X_n)_{n\geq 0}$ such that, for all $n \geq 0$,

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \quad \text{almost surely.}$$

If equality is replaced in this condition by $\leq$, then we call X a *supermartingale*. On the other hand, if equality is replaced by $\geq$, then we call X a *submartingale*. Note that every process which is a martingale with respect to the given filtration $(\mathcal{F}_n)_{n\geq 0}$ is also a martingale with respect to its natural filtration.

2.2. **Optional stopping.** We say that a random variable

$$T : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$$

is a *stopping time* if $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$. For a stopping time T, we set

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n \geq 0\}.$$

It is easy to check that, if $T(\omega) = n$ for all ω, then T is a stopping time and $\mathcal{F}_T = \mathcal{F}_n$. Given a process X, we define

$$X_T(\omega) = X_{T(\omega)}(\omega) \quad \text{whenever} \quad T(\omega) < \infty$$

and we define the *stopped process* X^T by

$$X_n^T(\omega) = X_{T(\omega)\wedge n}(\omega), \quad n \geq 0.$$

Proposition 2.2.1. *Let S and T be stopping times and let X be an adapted process. Then*

(a) *$S \wedge T$ is a stopping time,*
(b) *$\mathcal{F}_T$ is a σ-algebra,*
(c) *if $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$,*
(d) *$X_T 1_{T<\infty}$ is an $\mathcal{F}_T$-measurable random variable,*
(e) *X^T is adapted,*
(f) *if X is integrable, then X^T is integrable.*

Throughout these notes, a 'Proposition' indicates a straightforward result whose proof is left as an exercise.

Theorem 2.2.2 (Optional stopping theorem)**.** *Let X be a supermartingale and let S and T be bounded stopping times with $S \leq T$. Then $\mathbb{E}(X_T) \leq \mathbb{E}(X_S)$.*

Proof. Fix $n \geq 0$ such that $T \leq n$. Then

$$X_T = X_S + \sum_{S\leq k<T} (X_{k+1} - X_k) = X_S + \sum_{k=0}^{n} (X_{k+1} - X_k) 1_{S\leq k<T}. \tag{2.1}$$

Now $\{S \leq k\} \in \mathcal{F}_k$ and $\{T > k\} \in \mathcal{F}_k$, so

$$\mathbb{E}((X_{k+1} - X_k) 1_{S\leq k<T}) \leq 0.$$

Hence, on taking expectations in (2.1), we obtain

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_S).$$

□

Note that X is a submartingale if and only if $-X$ is a supermartingale, and X is a martingale if and only both X and $-X$ are supermartingales. So the optional stopping theorem immediately implies a submartingale version with $\mathbb{E}(X_T) \geq \mathbb{E}(X_S)$ and a martingale version with $\mathbb{E}(X_T) = \mathbb{E}(X_0) = \mathbb{E}(X_S)$. We will prove a more comprehensive result on the relationship between supermartingales and stopping times.

Theorem 2.2.3. *Let X be an adapted integrable process. Then the following are equivalent*

(a) *X is a supermartingale,*
(b) *for all bounded stopping times T and all stopping times S,*

$$\mathbb{E}(X_T|\mathcal{F}_S) \leq X_{S\wedge T} \quad \textit{almost surely,}$$

(c) *for all stopping times T, the stopped process X^T is a supermartingale,*
(d) *for all bounded stopping times T and all stopping times $S \leq T$,*

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_S).$$

Proof. For $S \geq 0$ and $T \leq n$, we have

$$X_T = X_{S\wedge T} + \sum_{S\leq k<T}(X_{k+1} - X_k) = X_{S\wedge T} + \sum_{k=0}^{n}(X_{k+1} - X_k)1_{S\leq k<T}. \tag{2.2}$$

Suppose that X is a supermartingale and that S and T are stopping times, with $T \leq n$. Let $A \in \mathcal{F}_S$. Then $A \cap \{S \leq k\} \in \mathcal{F}_k$ and $\{T > k\} \in \mathcal{F}_k$, so

$$\mathbb{E}((X_{k+1} - X_k)1_{S\leq k<T}1_A) \leq 0.$$

Hence, on multiplying (2.2) by 1_A and taking expectations, we obtain

$$\mathbb{E}(X_T 1_A) \leq \mathbb{E}(X_{S\wedge T}1_A).$$

We have shown that (a) implies (b).

It is obvious that (b) implies (c) and (d) and that (c) implies (a).

Let $m \leq n$ and $A \in \mathcal{F}_m$. Set $T = m1_A + n1_{A^c}$. Then T is a stopping time and $T \leq n$. We note that

$$\mathbb{E}(X_n 1_A) - \mathbb{E}(X_m 1_A) = \mathbb{E}(X_n) - \mathbb{E}(X_T).$$

It follows that (d) implies (a). □

2.3. **Doob's upcrossing inequality.** Let X be a random process and let $a, b \in \mathbb{R}$ with $a < b$. Fix $\omega \in \Omega$. By an *upcrossing* of $[a, b]$ by $X(\omega)$, we mean an interval of times $\{j, j+1, \dots, k\}$ such that $X_j(\omega) < a$ and $X_k(\omega) > b$. Write $U_n[a, b](\omega)$ for the number of disjoint upcrossings contained in $\{0, 1, \dots, n\}$ and write $U[a, b](\omega)$ for the total number of disjoint upcrossings. Then, as $n \to \infty$, we have

$$U_n[a, b] \uparrow U[a, b].$$

Theorem 2.3.1 (Doob's upcrossing inequality)**.** *Let X be a supermartingale. Then*

$$(b - a)\mathbb{E}(U[a, b]) \leq \sup_{n\geq 0} \mathbb{E}((X_n - a)^-).$$

Proof. Set $T_0 = 0$ and define recursively for $k \geq 0$

$$S_{k+1} = \inf\{m \geq T_k : X_m < a\}, \quad T_{k+1} = \inf\{m \geq S_{k+1} : X_m > b\}.$$

Note that, if $T_k < \infty$, then $\{S_k, S_k+1, \dots, T_k\}$ is an upcrossing of $[a,b]$ by X, and indeed T_k is the time of completion of the kth disjoint upcrossing. Note that $U_n[a,b] \leq n$. For $m \leq n$, we have

$$\{U_n[a,b] = m\} = \{T_m \leq n < T_{m+1}\}$$

and, on this event,

$$X_{T_k \wedge n} - X_{S_k \wedge n} = \begin{cases} X_{T_k} - X_{S_k} \geq b - a, & \text{if } k \leq m, \\ X_n - X_{S_k} \geq X_n - a, & \text{if } k = m+1 \text{ and } S_{m+1} \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, on summing over $k \leq n$, we obtain

$$\sum_{k=1}^{n} (X_{T_k \wedge n} - X_{S_k \wedge n}) \geq (b-a)U_n[a,b] - (X_n - a)^-.$$

Since X is a supermartingale and $T_k \wedge n$ and $S_k \wedge n$ are bounded stopping times with $S_k \leq T_k$, by optional stopping,

$$\mathbb{E}(X_{T_k \wedge n}) \leq \mathbb{E}(X_{S_k \wedge n}).$$

Hence, on taking expectations, we obtain

$$(b-a)\mathbb{E}(U_n[a,b]) \leq \mathbb{E}\left((X_n - a)^-\right) \tag{2.3}$$

and the desired estimate follows by monotone convergence. □

2.4. **Doob's maximal inequalities.** Define, for a random process X,

$$X_n^* = \sup_{k \leq n} |X_k|.$$

In the next two theorems, we see that the martingale (or submartingale) property allows us to obtain estimates on this supremum in terms of expectations for X_n itself.

Theorem 2.4.1 (Doob's maximal inequality)**.** *Let X be a martingale or non-negative submartingale. Then, for all $\lambda \geq 0$,*

$$\lambda \mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E}(|X_n| 1_{\{X_n^* \geq \lambda\}}) \leq \mathbb{E}(|X_n|).$$

Proof. If X is a martingale, then $|X|$ is a non-negative submartingale. It therefore suffices to consider the case where X is non-negative. Set

$$T = \inf\{k \geq 0 : X_k \geq \lambda\} \wedge n.$$

Then T is a stopping time and $T \leq n$ so, by optional stopping,

$$\mathbb{E}(X_n) \geq \mathbb{E}(X_T) = \mathbb{E}(X_T 1_{\{X_n^* \geq \lambda\}}) + \mathbb{E}(X_T 1_{\{X_n^* < \lambda\}}) \geq \lambda \mathbb{P}(X_n^* \geq \lambda) + \mathbb{E}(X_n 1_{\{X_n^* < \lambda\}}).$$

Hence

$$\lambda \mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E}(X_n 1_{\{X_n^* \geq \lambda\}}) \leq \mathbb{E}(X_n).$$

□

Theorem 2.4.2 (Doob's L^p-inequality)**.** *Let X be a martingale or non-negative submartingale. Then, for all $p > 1$ and $q = p/(p-1)$,*

$$\|X_n^*\|_p \leq q \|X_n\|_p.$$

Proof. If X is a martingale, then $|X|$ is a non-negative submartingale. So it suffices to consider the case where X is non-negative. Fix $k < \infty$. By Fubini's theorem, Doob's maximal inequality, and Hölder's inequality,

$$\mathbb{E}[(X_n^* \wedge k)^p] = \mathbb{E}\int_0^k p\lambda^{p-1} 1_{\{X_n^* \geq \lambda\}}\, d\lambda = \int_0^k p\lambda^{p-1}\mathbb{P}(X_n^* \geq \lambda)\, d\lambda$$

$$\leq \int_0^k p\lambda^{p-2}\mathbb{E}(X_n 1_{\{X_n^* \geq \lambda\}})\, d\lambda = q\mathbb{E}(X_n(X_n^* \wedge k)^{p-1}) \leq q\|X_n\|_p \|X_n^* \wedge k\|_p^{p-1}.$$

Hence $\|X_n^* \wedge k\|_p \leq q\|X_n\|_p$ and the result follows by monotone convergence on letting $k \to \infty$. $\square$

Doob's maximal and L^p inequalities have versions which apply, under the same hypotheses, to the full supremum

$$X^* = \sup_{n \geq 0} |X_n|.$$

Since $X_n^* \uparrow X^*$, on letting $n \to \infty$, we obtain, for all $\lambda \geq 0$,

$$\lambda\mathbb{P}(X^* > \lambda) = \lim_{n\to\infty} \lambda\mathbb{P}(X_n^* > \lambda) \leq \sup_{n\geq 0} \mathbb{E}(|X_n|).$$

We can then replace $\lambda\mathbb{P}(X^* > \lambda)$ by $\lambda\mathbb{P}(X^* \geq \lambda)$ by taking limits from the right in λ. Similarly, for $p \in (1, \infty)$, by monotone convergence,

$$\|X^*\|_p \leq q \sup_{n\geq 0} \|X_n\|_p.$$

2.5. **Doob's martingale convergence theorems.** We say that a random process X is *L^p-bounded* if

$$\sup_{n\geq 0} \|X_n\|_p < \infty.$$

We say that X is *uniformly integrable* if

$$\sup_{n\geq 0} \mathbb{E}\left(|X_n| 1_{\{|X_n|>\lambda\}}\right) \to 0 \quad \text{as} \quad \lambda \to \infty.$$

By Hölder's inequality, if X is L^p-bounded for some $p > 1$, then X is uniformly integrable. On the other hand, if X is uniformly integrable, then X is L^1-bounded.

Theorem 2.5.1 (Almost sure martingale convergence theorem)**.** *Let X be an L^1-bounded supermartingale. Then there exists an integrable $\mathcal{F}_\infty$-measurable random variable X_∞ such that $X_n \to X_\infty$ almost surely as $n \to \infty$.*

Proof. For a sequence of real numbers $(x_n)_{n\geq 0}$, as $n \to \infty$, either x_n converges (in $\mathbb{R}$), or $x_n \to \pm\infty$, or $\liminf x_n < \limsup x_n$. In the second case, we have $\liminf |x_n| = \infty$. In the third case, since the rationals are dense, there exist $a, b \in \mathbb{Q}$ such that $\liminf x_n < a < b < \limsup x_n$. Set

$$\Omega_0 = \Omega_\infty \cap \left(\bigcap_{a,b\in\mathbb{Q},\, a<b} \Omega_{a,b} \right)$$

where

$$\Omega_\infty = \{\liminf |X_n| < \infty\}, \quad \Omega_{a,b} = \{U[a,b] < \infty\}.$$

Then $X_n(\omega)$ converges for all $\omega \in \Omega_0$. By Fatou's lemma and Doob's upcrossing inequality, for all $a < b$,

$$\mathbb{E}(\liminf |X_n|) \leq \liminf \mathbb{E}|X_n|, \quad (b-a)\mathbb{E}(U[a,b]) \leq |a| + \sup_{n \geq 0} \mathbb{E}|X_n|.$$

So, since $(X_n)_{n\geq 0}$ is L^1-bounded, we have $\mathbb{P}(\Omega_0) = 1$. Define

$$X_\infty = \lim_{n\to\infty} X_n 1_{\Omega_0}.$$

Then $X_n \to X_\infty$ almost surely, X_∞ is $\mathcal{F}_\infty$-measurable and $|X_\infty| \leq \liminf |X_n|$ so X_∞ is integrable. □

Note, in particular, that every non-negative supermartingale is L^1-bounded and hence, by the theorem, converges almost surely.

Theorem 2.5.2 (L^1 martingale convergence theorem)**.** *Let $(X_n)_{n\geq 0}$ be a uniformly integrable martingale. Then there exists a random variable $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ as $n \to \infty$ almost surely and in L^1. Moreover, $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ almost surely for all $n \geq 0$. Moreover, for all $Y \in L^1(\mathcal{F}_\infty)$, on choosing a version X_n of $\mathbb{E}(Y|\mathcal{F}_n)$ for all n, we obtain a uniformly integrable martingale $(X_n)_{n\geq 0}$ such that $X_n \to Y$ almost surely and in L^1.*

Proof. Let $(X_n)_{n\geq 0}$ be a uniformly integrable martingale. By the almost sure martingale convergence theorem, there exists $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ almost surely. Since X is uniformly integrable, it follows that $X_n \to X_\infty$ in L^1, by [PM, Theorems 2.5.1 and 6.2.3]. Next, for $m \geq n$,

$$\|X_n - \mathbb{E}(X_\infty|\mathcal{F}_n)\|_1 = \|\mathbb{E}(X_m - X_\infty|\mathcal{F}_n)\|_1 \leq \|X_m - X_\infty\|_1.$$

Let $m \to \infty$ to deduce $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ almost surely.

Suppose now that $Y \in L^1(\mathcal{F}_\infty)$ and let X_n be a version of $\mathbb{E}(Y|\mathcal{F}_n)$ for all n. Then $(X_n)_{n\geq 0}$ is a martingale by the tower property and is uniformly integrable by Lemma 1.5.1. Hence there exists $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ almost surely and in L^1. For all $n \geq 0$ and all $A \in \mathcal{F}_n$ we have

$$\mathbb{E}(X_\infty 1_A) = \lim_{m\to\infty} \mathbb{E}(X_m 1_A) = \mathbb{E}(Y 1_A).$$

Now $X_\infty, Y \in L^1(\mathcal{F}_\infty)$ and $\cup_n \mathcal{F}_n$ is a π-system generating $\mathcal{F}_\infty$. Hence, by [PM, Proposition 3.1.4], $X_\infty = Y$ almost surely. □

This theorem can be seen as setting up a bijection between the set of uniformly integrable martingales and $L^1(\mathcal{F}_\infty)$, given by $X \mapsto X_\infty$, provided that we identify martingales and random variables which agree almost surely.

Theorem 2.5.3 (L^p martingale convergence theorem)**.** *Let $p \in (1,\infty)$. Let $(X_n)_{n\geq 0}$ be an L^p-bounded martingale. Then there exists a random variable $X_\infty \in L^p(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ as $n \to \infty$ almost surely and in L^p. Moreover, $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ almost surely for all $n \geq 0$. Moreover, for all $Y \in L^p(\mathcal{F}_\infty)$, on choosing a version X_n of $\mathbb{E}(Y|\mathcal{F}_n)$ for all n, we obtain an L^p-bounded martingale $(X_n)_{n\geq 0}$ such that $X_n \to Y$ almost surely and in L^p.*

Proof. Let $(X_n)_{n\geq 0}$ be an L^p-bounded martingale. By the almost sure martingale convergence theorem, there exists $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ almost surely. By Doob's L^p-inequality,

$$\|X^*\|_p \leq q \sup_{n\geq 0} \|X_n\|_p < \infty.$$

Since $|X_n - X_\infty|^p \leq (2X^*)^p$ for all n, it follows by dominated convergence that $X_n \to X_\infty$ in L^p. Then $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ almost surely for all $n \geq 0$, as in the L^1 case.

Suppose now that $Y \in L^p(\mathcal{F}_\infty)$ and let X_n be a version of $\mathbb{E}(Y|\mathcal{F}_n)$ for all n. Then $(X_n)_{n\geq 0}$ is a martingale by the tower property and

$$\|X_n\|_p = \|\mathbb{E}(Y|\mathcal{F}_n)\|_p \leq \|Y\|_p$$

for all n, so $(X_n)_{n\geq 0}$ is L^p-bounded. Hence there exists $X_\infty \in L^p(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ almost surely and in L^p. Finally, we must have $X_\infty = Y$ almost surely, as in the L^1 case. $\square$

In the next result, we dispense with the filtration $(\mathcal{F}_n)_{n\geq 0}$ and suppose given instead a *backward filtration* $(\hat{\mathcal{F}}_n)_{n\geq 0}$, that is to say, a sequence of σ-algebras $\hat{\mathcal{F}}_n$ such that, for all $n \geq 0$,

$$\mathcal{F} \supseteq \hat{\mathcal{F}}_n \supseteq \hat{\mathcal{F}}_{n+1}.$$

We write $\hat{\mathcal{F}}_\infty$ for the σ-algebra given by

$$\hat{\mathcal{F}}_\infty = \bigcap_{n\geq 0} \hat{\mathcal{F}}_n.$$

Theorem 2.5.4 (Backward martingale convergence theorem)**.** *For all $Y \in L^1(\mathcal{F})$, we have $\mathbb{E}(Y|\hat{\mathcal{F}}_n) \to \mathbb{E}(Y|\hat{\mathcal{F}}_\infty)$ as $n \to \infty$, almost surely and in L^1.*

Proof. Write $X_n = \mathbb{E}(Y|\hat{\mathcal{F}}_n)$ for all $n \geq 0$. Fix $n \geq 0$. By the tower property, $(X_{n-k})_{0\leq k\leq n}$ is a martingale for the filtration $(\hat{\mathcal{F}}_{n-k})_{0\leq k\leq n}$. For $a < b$, the number $U_n[a,b]$ of upcrossings of $[a,b]$ by $(X_k)_{0\leq k\leq n}$ equals the number of upcrossings of $[-b,-a]$ by $(-X_{n-k})_{0\leq k\leq n}$. Hence, from (2.3), we obtain

$$(b-a)\mathbb{E}(U_n[a,b]) \leq \mathbb{E}((X_0 - b)^+)$$

and so, by monotone convergence,

$$(b-a)\mathbb{E}(U[a,b]) \leq \mathbb{E}((X_0 - b)^+) \leq \mathbb{E}|Y| + |b| < \infty.$$

Also, we have

$$\mathbb{E}(\liminf |X_n|) \leq \liminf \mathbb{E}|X_n| \leq \mathbb{E}|Y| < \infty.$$

Hence the argument used in the proof of the almost sure martingale convergence theorem applies to show that $\mathbb{P}(\hat{\Omega}_0) = 1$, where

$$\hat{\Omega}_0 = \{X_n \text{ converges as } n \to \infty\}.$$

Set

$$X_\infty = 1_{\hat{\Omega}_0} \lim_{n\to\infty} X_n.$$

Then $X_\infty \in L^1(\hat{\mathcal{F}}_\infty)$ and $X_n \to X_\infty$ almost surely. Now $(X_n)_{n\geq 0}$ is uniformly integrable by Lemma 1.5.1, so $X_n \to X_\infty$ also in L^1. Finally, for all $A \in \hat{\mathcal{F}}_\infty$, we have

$$\mathbb{E}((X_\infty - \mathbb{E}(Y|\hat{\mathcal{F}}_\infty))1_A) = \lim_{n\to\infty} \mathbb{E}((X_n - Y)1_A) = 0$$

and this implies that $X_\infty = \mathbb{E}(Y|\hat{\mathcal{F}}_\infty)$ almost surely. $\square$

Recall that, for a stopping time T and a random process X, X_T has been defined only on the event $\{T < \infty\}$. Given an almost sure limit X_∞ for X, we define $X_T = X_\infty$ on $\{T = \infty\}$. Then the optional stopping theorem extends to all stopping times for uniformly integrable martingales.

Theorem 2.5.5. *Let X be a uniformly integrable martingale and let T be any stopping time. Then $\mathbb{E}(X_T) = \mathbb{E}(X_0)$. Moreover, for all stopping times S and T, we have*

$$\mathbb{E}(X_T|\mathcal{F}_S) = X_{S\wedge T} \quad \textit{almost surely.}$$

Proof. By the L^1 martingale convergence theorem, there exists $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $X_n \to X_\infty$ as $n \to \infty$, almost surely and in L^1, and $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ almost surely, for all n. Hence X_T is well defined and $X_{T\wedge n} \to X_T$ almost surely as $n \to \infty$.

Consider, for each $n \geq 0$, the bounded stopping time $T \wedge n$. By the optional stopping theorem and Theorem 2.2.3

$$\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0), \quad \mathbb{E}(X_{T\wedge n}|\mathcal{F}_S) = X_{S\wedge T\wedge n} \quad \text{almost surely.} \tag{2.4}$$

Since $\mathcal{F}_{T\wedge n} \subseteq \mathcal{F}_n$, by Theorem 2.2.3 and the tower property,

$$X_{T\wedge n} = \mathbb{E}(X_n|\mathcal{F}_{T\wedge n}) = \mathbb{E}(X_\infty|\mathcal{F}_{T\wedge n}).$$

Then, by Lemma 1.5.1, the random process $(X_{T\wedge n})_{n\geq 0}$ is uniformly integrable. So $X_{T\wedge n} \to X_T$ in L^1 and so also $\mathbb{E}(X_{T\wedge n}|\mathcal{F}_S) \to \mathbb{E}(X_T|\mathcal{F}_S)$ in L^1. Hence we can let $n \to \infty$ in (2.4) to obtain the claimed identities. $\square$

3. Applications of martingale theory

3.1. **Sums of independent random variables.** We use martingale arguments to analyse some aspects of the behaviour of the partial sums

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

of a sequence $(X_n)_{n\geq 1}$ of independent random variables. We will have more to say about such sums in Theorem 6.1.1 and Theorem 7.11.2

Theorem 3.1.1 (Strong law of large numbers)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of independent, identically distributed, integrable random variables. Set $\mu = \mathbb{E}(X_1)$. Then $S_n/n \to \mu$ as $n \to \infty$ almost surely and in L^1.*

Proof. Define for $n \geq 1$

$$\hat{\mathcal{F}}_n = \sigma(S_m : m \geq n), \quad \mathcal{T}_n = \sigma(X_m : m \geq n+1), \quad \mathcal{T} = \cap_{n\geq 1}\mathcal{T}_n.$$

Then $\hat{\mathcal{F}}_n = \sigma(S_n, \mathcal{T}_n)$ and $(\hat{\mathcal{F}}_n)_{n\geq 1}$ is a backward filtration. Since $\sigma(X_1, S_n)$ is independent of $\mathcal{T}_n$, we have $\mathbb{E}(X_1|\hat{\mathcal{F}}_n) = \mathbb{E}(X_1|S_n)$ almost surely for all n. For $k \leq n$ and all Borel sets B, we have $\mathbb{E}(X_k 1_{\{S_n\in B\}}) = \mathbb{E}(X_1 1_{\{S_n\in B\}})$ by symmetry, so $\mathbb{E}(X_k|S_n) = \mathbb{E}(X_1|S_n)$ almost surely. But

$$\mathbb{E}(X_1|S_n) + \cdots + \mathbb{E}(X_n|S_n) = \mathbb{E}(S_n|S_n) = S_n \quad \text{almost surely}$$

so we must have

$$\mathbb{E}(X_1|\hat{\mathcal{F}}_n) = \mathbb{E}(X_1|S_n) = S_n/n \quad \text{almost surely.}$$

Then, by the backward martingale convergence theorem,

$$S_n/n \to Y \quad \text{almost surely and in } L^1$$

for some random variable Y. Then Y is $\mathcal{T}$-measurable so, by Kolmogorov's zero-one law [PM, Theorem 2.6.1], Y is constant almost surely. Hence

$$Y = \mathbb{E}(Y) = \lim_{n\to\infty} \mathbb{E}(S_n/n) = \mu \quad \text{almost surely.}$$

□

Since almost sure convergence implies convergence in probability [PM, Theorem 2.5.1], the following is an immediate corollary.

Corollary 3.1.2 (Weak law of large numbers)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of independent, identically distributed, integrable random variables. Set $\mu = \mathbb{E}(X_1)$. Then $\mathbb{P}(|S_n/n - \mu| > \varepsilon) \to 0$ as $n \to \infty$ for all $\varepsilon > 0$.*

Proposition 3.1.3. *Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables in L^2. Set $S_n = X_1 + \cdots + X_n$ and write*

$$\mu_n = \mathbb{E}(S_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n), \quad \sigma_n^2 = \mathrm{var}(S_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n).$$

Then the following are equivalent:

(a) *the sequences $(\mu_n)_{n\geq 1}$ and $(\sigma_n^2)_{n\geq 1}$ converge in $\mathbb{R}$,*
(b) *there exists a random variable S such that $S_n \to S$ almost surely and in L^2.*

The following identities allow estimation of exit probabilities and the mean exit time for a random walk in an interval. They are of some historical interest, having been developed by Wald in the 1940's to compute the efficiency of the sequential probability ratio test.

Proposition 3.1.4 (Wald's identities)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of independent, identically distributed random variables, having mean μ and variance $\sigma^2 \in (0, \infty)$. Fix $a, b \in \mathbb{R}$ with $a < 0 < b$ and set*

$$T = \inf\{n \geq 0 : S_n \leq a \text{ or } S_n \geq b\}.$$

Then $\mathbb{E}(T) < \infty$ and

$$\mathbb{E}(S_T) = \mu\mathbb{E}(T).$$

Moreover, in the case $\mu = 0$, we have

$$\mathbb{E}(S_T^2) = \sigma^2\mathbb{E}(T)$$

while, in the case $\mu \neq 0$, if we can find $\lambda^* \neq 0$ such that $\mathbb{E}(e^{\lambda^* X_1}) = 1$, then

$$\mathbb{E}(e^{\lambda^* S_T}) = 1.$$

3.2. **Non-negative martingales and change of measure.** Given a random variable X, with $X \geq 0$ and $\mathbb{E}(X) = 1$, we can define a new probability measure $\tilde{\mathbb{P}}$ on $\mathcal{F}$ by

$$\tilde{\mathbb{P}}(A) = \mathbb{E}(X1_A), \quad A \in \mathcal{F}.$$

Moreover, by [PM, Proposition 3.1.4], given $\tilde{\mathbb{P}}$, this equation determines X uniquely, up to almost sure modification. We say that $\tilde{\mathbb{P}}$ *has a density with respect to* $\mathbb{P}$ and X *is a version of the density.*

Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration in $\mathcal{F}$ and assume for simplicity that $\mathcal{F} = \mathcal{F}_\infty$. Let $(X_n)_{n\geq 0}$ be an adapted random process, with $X_n \geq 0$ and $\mathbb{E}(X_n) = 1$ for all n. We can define for each n a probability measure $\tilde{\mathbb{P}}_n$ on $\mathcal{F}_n$ by

$$\tilde{\mathbb{P}}_n(A) = \mathbb{E}(X_n 1_A), \quad A \in \mathcal{F}_n.$$

Since we require X_n to be $\mathcal{F}_n$-measurable, this equation determines X_n uniquely, up to almost sure modification.

Proposition 3.2.1. *The measures $\tilde{\mathbb{P}}_n$ are consistent, that is $\tilde{\mathbb{P}}_{n+1}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n$ for all n, if and only if $(X_n)_{n\geq 0}$ is a martingale. Moreover, there is a measure $\tilde{\mathbb{P}}$ on $\mathcal{F}$, which has a density with respect to $\mathbb{P}$, such that $\tilde{\mathbb{P}}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n$ for all n, if and only if $(X_n)_{n\geq 0}$ is a uniformly integrable martingale.*

This construction can also give rise to new probability measures which do not have a density with respect to $\mathbb{P}$ on $\mathcal{F}$, as the following result suggests.

Theorem 3.2.2. *There exists a measure $\tilde{\mathbb{P}}$ on $\mathcal{F}$ such that $\tilde{\mathbb{P}}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n$ for all n if and only if $\mathbb{E}(X_T) = 1$ for all finite stopping times T.*

Proof. Suppose that $\mathbb{E}(X_T) = 1$ for all finite stopping times T. Then, since bounded stopping times are finite, $(X_n)_{n\geq 0}$ is a martingale, by optional stopping. Hence we can define consistently a set function $\tilde{\mathbb{P}}$ on $\cup_n \mathcal{F}_n$ such that $\tilde{\mathbb{P}}|_{\mathcal{F}_n} = \tilde{\mathbb{P}}_n$ for all n. Note that $\cup_n \mathcal{F}_n$ is a ring. By Carathéodory's extension theorem [PM, Theorem 1.6.1], $\tilde{\mathbb{P}}$ extends to a measure on $\mathcal{F}_\infty$ if and only if $\tilde{\mathbb{P}}$ is countably additive on $\cup_n \mathcal{F}_n$. Since each $\tilde{\mathbb{P}}_n$ is countably additive, it is not hard to see that this condition holds if and only if

$$\sum_{n=1}^{\infty} \tilde{\mathbb{P}}(A_n) = 1$$

for all partitions $(A_n : n \geq 0)$ of Ω such that $A_n \in \mathcal{F}_n$ for all n. But such partitions are in one-to-one correspondence with finite stopping times T, by $\{T = n\} = A_n$, and then

$$\mathbb{E}(X_T) = \sum_{n=1}^{\infty} \tilde{\mathbb{P}}(A_n).$$

Hence $\tilde{\mathbb{P}}$ extends to a measure on $\mathcal{F}$ with the claimed property. Conversely, given such a measure, the last equation shows that $\mathbb{E}(X_T) = 1$ for all finite stopping times T. $\square$

Theorem 3.2.3 (Radon–Nikodym theorem)**.** *Let μ and ν be σ-finite measures on a measurable space $(E, \mathcal{E})$. Then the following are equivalent*

(a) $\nu(A) = 0$ *for all* $A \in \mathcal{E}$ *such that* $\mu(A) = 0$,

(b) *there exists a measurable function f on E such that $f \geq 0$ and*

$$\nu(A) = \mu(f 1_A), \quad A \in \mathcal{E}.$$

The function f, which is unique up to modification μ-almost everywhere, is called (*a version of*) *the Radon-Nikodym derivative of ν with respect to μ*. We write

$$f = \frac{d\nu}{d\mu} \quad \text{almost everywhere.}$$

We will give a proof for the case where $\mathcal{E}$ is countably generated. Thus, *we assume further that there is a sequence $(G_n : n \in \mathbb{N})$ of subsets of E which generates $\mathcal{E}$*. This holds, for example, whenever $\mathcal{E}$ is the Borel σ-algebra of a topology with countable basis. A further martingale argument, which we omit, allows to deduce the general case.

Proof. It is obvious that (b) implies (a). Assume then that (a) holds. There is a countable partition of E by measurable sets on which both μ and ν are finite. It will suffice to show that (b) holds on each of these sets, so we reduce without loss to the case where μ and ν are finite.

The case where $\nu(E) = 0$ is clear. Assume then that $\nu(E) > 0$. Then also $\mu(E) > 0$, by (a). Write $\Omega = E$ and $\mathcal{F} = \mathcal{E}$ and consider the probability measures $\mathbb{P} = \mu/\mu(E)$ and $\tilde{\mathbb{P}} = \nu/\nu(E)$ on $(\Omega, \mathcal{F})$. It will suffice to show that there is a random variable $X \geq 0$ such that $\tilde{\mathbb{P}}(A) = \mathbb{E}(X 1_A)$ for all $A \in \mathcal{F}$.

Set $\mathcal{F}_n = \sigma(G_k : k \leq n)$. There exist $m \in \mathbb{N}$ and a partition of Ω by events $A_1, \ldots, A_m$ such that $\mathcal{F}_n = \sigma(A_1, \ldots, A_m)$. Set

$$X_n = \sum_{j=1}^{m} a_j 1_{A_j}$$

where $a_j = \tilde{\mathbb{P}}(A_j)/\mathbb{P}(A_j)$ if $\mathbb{P}(A_j) > 0$ and $a_j = 0$ otherwise. Then $X_n \geq 0$, X_n is $\mathcal{F}_n$-measurable and, using (a), we have $\tilde{\mathbb{P}}(A) = \mathbb{E}(X_n 1_A)$ for all $A \in \mathcal{F}_n$. Observe that $(\mathcal{F}_n)_{n \geq 0}$ is a filtration and $(X_n)_{n \geq 0}$ is a non-negative martingale. *We will show that $(X_n)_{n \geq 0}$ is uniformly integrable.* Then, by the L^1 martingale convergence theorem, there exists a random variable $X \geq 0$ such that $\mathbb{E}(X 1_A) = \mathbb{E}(X_n 1_A)$ for all $A \in \mathcal{F}_n$. Define a probability measure $\mathbb{Q}$ on $\mathcal{F}$ by $\mathbb{Q}(A) = \mathbb{E}(X 1_A)$. Then $\mathbb{Q} = \tilde{\mathbb{P}}$ on $\cup_n \mathcal{F}_n$, which is a π-system generating $\mathcal{F}$. Hence $\mathbb{Q} = \tilde{\mathbb{P}}$ on $\mathcal{F}$, by uniqueness of extension [PM, Theorem 1.7.1], which implies (b).

It remains to show that $(X_n)_{n \geq 0}$ is uniformly integrable. Given $\varepsilon > 0$ we can find $\delta > 0$ such that $\tilde{\mathbb{P}}(B) < \varepsilon$ for all $B \in \mathcal{F}$ with $\mathbb{P}(B) < \delta$. For, if not, there would be a sequence of sets $B_n \in \mathcal{F}$ with $\mathbb{P}(B_n) < 2^{-n}$ and $\tilde{\mathbb{P}}(B_n) \geq \varepsilon$ for all n. Then

$$\mathbb{P}(\cap_n \cup_{m \geq n} B_m) = 0, \quad \tilde{\mathbb{P}}(\cap_n \cup_{m \geq n} B_m) \geq \varepsilon$$

which contradicts (a). Set $\lambda = 1/\delta$, then $\mathbb{P}(X_n > \lambda) \leq \mathbb{E}(X_n)/\lambda = 1/\lambda = \delta$ for all n, so

$$\mathbb{E}(X_n 1_{X_n > \lambda}) = \tilde{\mathbb{P}}(X_n > \lambda) < \varepsilon.$$

Hence $(X_n)_{n \geq 0}$ is uniformly integrable. □

3.3. **Markov chains.** Let E be a countable set. We identify each measure μ on E with its mass function $(\mu_x : x \in E)$, where $\mu_x = \mu(\{x\})$. Then, for each function f on E, the integral is conveniently written as the matrix product

$$\mu(f) = \mu f = \sum_{x \in E} \mu_x f_x$$

where we consider μ as a row vector and identify f with the column vector $(f_x : x \in E)$ given by $f_x = f(x)$. A *transition matrix* on E is a matrix $P = (p_{xy} : x, y \in E)$ such that each row $(p_{xy} : y \in E)$ is a probability measure.

Let a filtration $(\mathcal{F}_n)_{n \geq 0}$ be given and let $(X_n)_{n \geq 0}$ be an adapted process with values in E. We say that $(X_n)_{n \geq 0}$ is a *Markov chain with transition matrix* P if, for all $n \geq 0$, all $x, y \in E$ and all $A \in \mathcal{F}_n$ with $A \subseteq \{X_n = x\}$ and $\mathbb{P}(A) > 0$,

$$\mathbb{P}(X_{n+1} = y | A) = p_{xy}.$$

Our notion of Markov chain depends on the choice of $(\mathcal{F}_n)_{n \geq 0}$. The following result shows that our definition agrees with the usual one for the most obvious such choice.

Proposition 3.3.1. *Let $(X_n)_{n \geq 0}$ be a random process in E and take*

$$\mathcal{F}_n = \sigma(X_k : k \leq n).$$

The following are equivalent

(a) *$(X_n)_{n \geq 0}$ is a Markov chain with initial distribution μ and transition matrix P,*
(b) *for all n and all $x_0, x_1, \dots, x_n \in E$,*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu_{x_0} p_{x_0 x_1} \dots p_{x_{n-1} x_n}.$$

Proposition 3.3.2. *Let E^* denote the set of sequences $x = (x_n : n \geq 0)$ in E and define $X_n : E^* \to E$ by $X_n(x) = x_n$. Set $\mathcal{E}^* = \sigma(X_k : k \geq 0)$. Let P be a transition matrix on E. Then, for each $x \in E$, there is a unique probability measure $\mathbb{P}_x$ on $(E^*, \mathcal{E}^*)$ such that $(X_n)_{n \geq 0}$ is a Markov chain with transition matrix P and starting from x.*

A example of a Markov chain in $\mathbb{Z}^d$ is the simple symmetric random walk, whose transition matrix is given by

$$p_{xy} = \begin{cases} 1/(2d), & \text{if } |x - y| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

The following result shows a simple instance of a general relationship between Markov processes and martingales. We will see a second instance of this for Brownian motion in Theorem 7.5.1.

Proposition 3.3.3. *Let $(X_n)_{n \geq 0}$ be an adapted process in E. Then the following are equivalent*

(a) *$(X_n)_{n \geq 0}$ is a Markov chain with transition matrix P,*
(b) *for all bounded functions f on E the following process is a martingale*

$$M_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (P - I) f(X_k).$$

A bounded function f on E is said to be *harmonic* if $Pf = f$, that is to say, if

$$\sum_{y \in E} p_{xy} f_y = f_x, \quad x \in E.$$

Note that, if f is a bounded harmonic function, then $(f(X_n))_{n \geq 0}$ is a bounded martingale. Then, by Doob's convergence theorems, $f(X_n)$ converges almost surely and in L^p for all $p < \infty$.

More generally, for $D \subseteq E$, a bounded function f on E is *harmonic in* D if

$$\sum_{y \in E} p_{xy} f_y = f_x, \quad x \in D.$$

Suppose we set $\partial D = E \setminus D$ fix a bounded function f on ∂D. Set

$$T = \inf\{n \geq 0 : X_n \in \partial D\}$$

and define a function u on E by

$$u(x) = \mathbb{E}_x(f(X_T) 1_{\{T < \infty\}}).$$

Theorem 3.3.4. *The function u is bounded, harmonic in D, and $u = f$ on ∂D. Moreover, if $\mathbb{P}_x(T < \infty) = 1$ for all $x \in D$, then u is the unique bounded extension of f which is harmonic in D.*

Proof. It is clear that u is bounded and $u = f$ on ∂D. For all $x, y \in E$ with $p_{xy} > 0$, under $\mathbb{P}_x$, conditional on $\{X_1 = y\}$, $(X_{n+1})_{n \geq 0}$ has distribution $\mathbb{P}_y$. So, for $x \in D$,

$$u(x) = \sum_{y \in E} p_{xy} u(y)$$

showing that u is harmonic in D.

On the other hand, suppose that g is a bounded function, harmonic in D and such that $g = f$ on ∂D. Then $M = M^g$ is a martingale and T is a stopping time, so M^T is also a martingale by optional stopping. But $M_{T \wedge n} = g(X_{T \wedge n})$. So, if $\mathbb{P}_x(T < \infty) = 1$ for all $x \in D$, then

$$M_{T \wedge n} \to f(X_T) \quad \text{almost surely}$$

so, by bounded convergence, for all $x \in D$,

$$g(x) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_{T \wedge n}) \to \mathbb{E}_x(f(X_T)) = u(x).$$

$\square$

In Theorem 7.9.2 we will prove an analogous result for Brownian motion

4. Random processes in continuous time

4.1. **Definitions.** A *continuous random process* is a family of random variables $(X_t)_{t\geq 0}$ such that, for all $\omega \in \Omega$, the *path* $t \mapsto X_t(\omega) : [0,\infty) \to \mathbb{R}$ is continuous.

A function $x : [0,\infty) \to \mathbb{R}$ is said to be *cadlag* if it is right-continuous with left limits, that is to say, for all $t \geq 0$

$$x_s \to x_t \quad \text{as } s \to t \text{ with } s > t$$

and, for all $t > 0$, there exists $x_{t-} \in \mathbb{R}$ such that

$$x_s \to x_{t-} \quad \text{as } s \to t \text{ with } s < t.$$

The term is a French acronym for *continu à droite, limité à gauche*. A *cadlag random process* is a family of random variables $(X_t)_{t\geq 0}$ such that, for all $\omega \in \Omega$, the path $t \mapsto X_t(\omega) : [0,\infty) \to \mathbb{R}$ is cadlag.

The spaces of continuous and cadlag functions on $[0,\infty)$ are denoted $C([0,\infty),\mathbb{R})$ and $D([0,\infty),\mathbb{R})$ respectively. We equip both these spaces with the σ-algebra generated by the coordinate functions $\sigma(x \mapsto x_t : t \geq 0)$. A continuous random process $(X_t)_{t\geq 0}$ can then be considered as a random variable X in $C([0,\infty),\mathbb{R})$ given by

$$X(\omega) = (t \mapsto X_t(\omega) : t \geq 0).$$

A cadlag random process can be thought of as a random variable in $D([0,\infty),\mathbb{R})$. The *finite-dimensional distributions* of a continuous or cadlag process X are the laws $\mu_{t_1,\dots,t_n}$ on $\mathbb{R}^n$ given by

$$\mu_{t_1,\dots,t_n}(A) = \mathbb{P}((X_{t_1},\dots,X_{t_n}) \in A), \quad A \in \mathcal{B}(\mathbb{R}^n)$$

where $n \in \mathbb{N}$ and $t_1,\dots,t_n \in [0,\infty)$ with $t_1 < \cdots < t_n$. Since the *cylinder sets* $\{(X_{t_1},\dots,X_{t_n}) \in A\}$ form a generating π-system, they determine uniquely the law of X. We make analogous definitions when $\mathbb{R}$ is replaced by a general topological space.

4.2. **Kolmogorov's criterion.** This result allows us to prove pathwise Hölder continuity for a random process starting from L^p-Hölder continuity, by giving up $\frac{1}{p}$ in the exponent. In particular, it is a means to construct continuous random processes.

Theorem 4.2.1 (Kolmogorov's criterion)**.** *Let $p \in (1,\infty)$ and $\beta \in (\frac{1}{p}, 1]$. Let I be a dense subset of $[0,1]$ and let $(\xi_t)_{t\in I}$ be a family of random variables such that, for some constant $C < \infty$,*

$$\|\xi_s - \xi_t\|_p \leq C|s-t|^\beta, \quad \textit{for all } s,t \in I. \tag{4.1}$$

Then there exists a continuous random process $(X_t)_{t\in[0,1]}$ such that

$$X_t = \xi_t \quad \textit{almost surely,} \quad \textit{for all } t \in I.$$

Moreover $(X_t)_{t\in[0,1]}$ may be chosen so that, for all $\alpha \in [0, \beta - \frac{1}{p})$, there exists $K_\alpha \in L^p$ such that

$$|X_s - X_t| \leq K_\alpha |s-t|^\alpha, \quad \textit{for all } s,t \in [0,1].$$

Proof. For $n \geq 0$, write

$$\mathbb{D}_n = \{k2^{-n} : k \in \mathbb{Z}^+\}, \quad \mathbb{D} = \cup_{n\geq 0}\mathbb{D}_n, \quad D_n = \mathbb{D}_n \cap [0,1), \quad D = \mathbb{D} \cap [0,1].$$

By taking limits in L^p, we can extend $(\xi_t)_{t\in I}$ to all parameter values $t \in D$ and so that (4.1) holds for all $s, t \in D \cup I$. For $n \geq 0$ and $\alpha \in [0, \beta - \frac{1}{p})$, define non-negative random variables by

$$K_n = \sup_{t\in D_n} |\xi_{t+2^{-n}} - \xi_t|, \quad K_\alpha = 2\sum_{n\geq 0} 2^{n\alpha} K_n.$$

Then

$$\mathbb{E}(K_n^p) \leq \mathbb{E}\sum_{t\in D_n} |\xi_{t+2^{-n}} - \xi_t|^p \leq 2^n C^p (2^{-n})^{\beta p}$$

so

$$\|K_\alpha\|_p \leq 2\sum_{n\geq 0} 2^{n\alpha}\|K_n\|_p \leq 2C\sum_{n\geq 0} 2^{-(\beta-\alpha-1/p)n} < \infty.$$

For $s, t \in D$ with $s < t$, choose $m \geq 0$ so that $2^{-m-1} < t - s \leq 2^{-m}$. The interval $[s, t)$ can be expressed as the finite disjoint union of intervals of the form $[r, r + 2^{-n})$, where $r \in D_n$ and $n \geq m + 1$ and where no three intervals have the same length. Hence

$$|\xi_t - \xi_s| \leq 2\sum_{n\geq m+1} K_n$$

and so

$$|\xi_t - \xi_s|/(t-s)^\alpha \leq 2\sum_{n\geq m+1} K_n 2^{(m+1)\alpha} \leq K_\alpha.$$

Now define

$$X_t(\omega) = \begin{cases} \lim_{s\to t,\, s\in D} \xi_s(\omega) & \text{if } K_\alpha(\omega) < \infty \text{ for all } \alpha \in [0, \beta - \frac{1}{p}), \\ 0 & \text{otherwise.} \end{cases}$$

Then $(X_t)_{t\in[0,1]}$ is a continuous random process with the claimed properties. $\square$

4.3. **Martingales in continuous time.** We assume in this section that our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a *continuous-time filtration*, that is, a family of σ-algebras $(\mathcal{F}_t)_{t\geq 0}$ such that

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, \quad s \leq t.$$

Define for $t \geq 0$

$$\mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s, \quad \mathcal{F}_\infty = \sigma(\mathcal{F}_t : t \geq 0), \quad \mathcal{N} = \{A \in \mathcal{F}_\infty : \mathbb{P}(A) = 0\}.$$

The filtration $(\mathcal{F}_t)_{t\geq 0}$ is said to satisfy the *usual conditions* if $\mathcal{N} \subseteq \mathcal{F}_0$ and $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t. A continuous adapted integrable random process $(X_t)_{t\geq 0}$ is said to be a *continuous martingale* if, for all $s, t \geq 0$ with $s \leq t$,

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s \quad \text{almost surely.}$$

We define analogously the notion of a *cadlag martingale*. If equality is replaced in this condition by $\leq$ or $\geq$, we obtain notions of *supermartingale* and *submartingale* respectively.

Recall that we write, for $n \geq 0$,

$$\mathbb{D}_n = \{k2^{-n} : k \in \mathbb{Z}^+\}, \quad \mathbb{D} = \cup_{n\geq 0}\mathbb{D}_n.$$

Define, for a cadlag random process X,

$$X^* = \sup_{t\geq 0} |X_t|, \quad X^{(n)*} = \sup_{t\in\mathbb{D}_n} |X_t|.$$

The cadlag property implies that

$$X^{(n)*} \to X^* \quad \text{as } n \to \infty$$

while, if $(X_t)_{t\geq 0}$ is a cadlag martingale, then $(X_t)_{t\in\mathbb{D}_n}$ is a discrete-time martingale, for the filtration $(\mathcal{F}_t)_{t\in\mathbb{D}_n}$, and similarly for supermartingales and submartingales. Thus, on applying Doob's inequalities to $(X_t)_{t\in\mathbb{D}_n}$ and passing to the limit we obtain the following results.

Theorem 4.3.1 (Doob's maximal inequality)**.** *Let X be a cadlag martingale or non-negative submartingale. Then, for all $\lambda \geq 0$,*

$$\lambda\mathbb{P}(X^* \geq \lambda) \leq \sup_{t\geq 0} \mathbb{E}(|X_t|).$$

Theorem 4.3.2 (Doob's L^p-inequality)**.** *Let X be a cadlag martingale or non-negative submartingale. Then, for all $p > 1$ and $q = p/(p-1)$,*

$$\|X^*\|_p \leq q \sup_{t\geq 0} \|X_t\|_p.$$

Similarly, the cadlag property implies that every upcrossing of a non-trivial interval by $(X_t)_{t\geq 0}$ corresponds, eventually as $n \to \infty$, to an upcrossing by $(X_t)_{t\in\mathbb{D}_n}$. This leads to the following estimate.

Theorem 4.3.3 (Doob's upcrossing inequality)**.** *Let X be a cadlag supermartingale and let $a, b \in \mathbb{R}$ with $a < b$. Then*

$$(b-a)\mathbb{E}(U[a,b]) \leq \sup_{t\geq 0} \mathbb{E}((X_t - a)^-)$$

where $U[a,b]$ is the total number of disjoint upcrossings of $[a,b]$ by X.

Then, arguing as in the discrete-time case, we obtain continuous-time versions of each martingale convergence theorem, where the notions of L^p-bounded and uniformly integrable are adapted in the obvious way.

Theorem 4.3.4 (Almost sure martingale convergence theorem)**.** *Let X be an L^1-bounded cadlag supermartingale. Then there exists an integrable $\mathcal{F}_\infty$-measurable random variable X_∞ such that $X_t \to X_\infty$ almost surely as $t \to \infty$.*

The following result shows, in particular, that, under the usual conditions on $(\mathcal{F}_t)_{t\geq 0}$, martingales are naturally cadlag.

Theorem 4.3.5 (L^1 martingale convergence theorem)**.** *Let $(X_t)_{t\geq 0}$ be a uniformly integrable cadlag martingale. Then there exists a random variable $X_\infty \in L^1(\mathcal{F}_\infty)$ such that $X_t \to X_\infty$ as $t \to \infty$ almost surely and in L^1. Moreover, $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$ almost surely for all $t \geq 0$. Moreover, if $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions, then, for all $Y \in L^1(\mathcal{F}_\infty)$, there exists a uniformly integrable cadlag martingale $(X_t)_{t\geq 0}$ such that $X_t = \mathbb{E}(Y|\mathcal{F}_t)$ almost surely for all t, and $X_t \to Y$ almost surely and in L^1.*

Proof. The proofs of the first two assertions are straightforward adaptations of the corresponding discrete-time proofs. We give details only for the final assertion. Suppose that $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions and that $Y \in L^1(\mathcal{F}_\infty)$. Choose a version ξ_t of $\mathbb{E}(Y|\mathcal{F}_t)$ for all $t \in \mathbb{D}$. Then $(\xi_t)_{t\in\mathbb{D}}$ is uniformly integrable and $(\xi_t)_{t\in\mathbb{D}_n}$ is a discrete-time martingale for all $n \geq 0$. Set $\xi^* = \sup_{t\in\mathbb{D}} |\xi_t|$ and write $u[a,b]$ for the total number of disjoint upcrossings of $[a,b]$ by $(\xi_t)_{t\in\mathbb{D}}$. Set

$$\Omega_0 = \Omega^* \cap \bigcap_{a,b\in\mathbb{Q},\, a<b} \Omega_{a,b}$$

where

$$\Omega^* = \{\xi^* < \infty\}, \quad \Omega_{a,b} = \{u[a,b] < \infty\}.$$

By the arguments leading to Theorems 4.3.1 and 4.3.3, we obtain the estimates

$$\lambda\mathbb{P}(\xi^* \geq \lambda) \leq \mathbb{E}|Y|, \quad (b-a)\mathbb{E}(u[a,b]) \leq \mathbb{E}|Y| + |a|$$

which then imply that $\mathbb{P}(\Omega_0) = 1$. Define for $t \geq 0$

$$X_t = \lim_{s\to t,\, s>t,\, s\in\mathbb{D}} \xi_s 1_{\Omega_0}.$$

The usual conditions ensure that $(X_t)_{t\geq 0}$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$. It is straightforward to check that $(X_t)_{t\geq 0}$ is cadlag and $X_t = \mathbb{E}(Y|\mathcal{F}_t)$ almost surely for all $t \geq 0$, so $(X_t)_{t\geq 0}$ is a uniformly integrable cadlag martingale. Moreover, X_t converges, with limit X_∞ say, as $t \to \infty$, and then $X_\infty = Y$ almost surely by the same argument used for the discrete-time case. □

Theorem 4.3.6 (L^p martingale convergence theorem)**.** *Let $p \in (1,\infty)$. Let $(X_t)_{t\geq 0}$ be an L^p-bounded cadlag martingale. Then there exists a random variable $X_\infty \in L^p(\mathcal{F}_\infty)$ such that $X_t \to X_\infty$ as $t \to \infty$ almost surely and in L^p. Moreover, $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$ almost surely for all $t \geq 0$. Moreover, if $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions, then, for all $Y \in L^p(\mathcal{F}_\infty)$, there exists an L^p-bounded cadlag martingale $(X_t)_{t\geq 0}$ such that $X_t = \mathbb{E}(Y|\mathcal{F}_t)$ almost surely for all t, and $X_t \to Y$ almost surely and in L^p.*

We say that a random variable

$$T : \Omega \to [0,\infty]$$

is a *stopping time* if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. For a stopping time T, we set

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

Given a cadlag random process X, we define X_T and the stopped process X^T by

$$X_T(\omega) = X_{T(\omega)}(\omega), \quad X_t^T(\omega) = X_{T(\omega)\wedge t}(\omega)$$

where we leave $X_T(\omega)$ undefined if $T(\omega) = \infty$ and $X_t(\omega)$ fails to converge as $t \to \infty$.

Proposition 4.3.7. *Let S and T be stopping times and let X be a cadlag adapted process. Then*

(a) *$S \wedge T$ is a stopping time,*
(b) *$\mathcal{F}_T$ is a σ-algebra,*
(c) *if $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$,*
(d) *$X_T 1_{T<\infty}$ is an $\mathcal{F}_T$-measurable random variable,*
(e) *X^T is adapted.*

Theorem 4.3.8. *Let X be a cadlag adapted integrable process. Then the following are equivalent*

(a) X *is a martingale,*
(b) *for all bounded stopping times* T *and all stopping times* S, X_T *is integrable and*

$$\mathbb{E}(X_T|\mathcal{F}_S) = X_{S\wedge T} \quad \textit{almost surely,}$$

(c) *for all stopping times* T, *the stopped process* X^T *is a martingale,*
(d) *for all bounded stopping times* T, X_T *is integrable and*

$$\mathbb{E}(X_T) = \mathbb{E}(X_0).$$

Moreover, if X *is uniformly integrable, then* (b) *and* (d) *hold for all stopping times* T.

Proof. Suppose (a) holds. Let S and T be stopping times, with T bounded, $T \leq t$ say. Let $A \in \mathcal{F}_S$. For $n \geq 0$, set

$$S_n = 2^{-n}\lceil 2^n S\rceil, \quad T_n = 2^{-n}\lceil 2^n T\rceil.$$

Then S_n and T_n are stopping times and $S_n \downarrow S$ and $T_n \downarrow T$ as $n \to \infty$. Since $(X_t)_{t\geq 0}$ is right continuous, $X_{T_n} \to X_T$ almost surely as $n \to \infty$. By Theorem 2.2.3, $X_{T_n} = \mathbb{E}(X_{t+1}|\mathcal{F}_{T_n})$ so $(X_{T_n} : n \geq 0)$ is uniformly integrable and so $X_{T_n} \to X_T$ in L^1. In particular, X_T is integrable. Similarly $X_{S_n\wedge T_n} \to X_{S\wedge T}$ in L^1. By Theorem 2.2.3 again,

$$\mathbb{E}(X_{T_n} 1_A) = \mathbb{E}(X_{S_n\wedge T_n} 1_A).$$

On letting $n \to \infty$, we deduce that (b) holds. For the rest of the proof we argue as in the discrete-time case. □

5. Weak convergence

5.1. **Definitions and characterizations.** Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on a metric space E, and let μ be another probability measure on E. We say that μ_n *converges to* μ *weakly on* E and write $\mu_n \to \mu$ *weakly on* E if $\mu_n(f) \to \mu(f)$ for all bounded continuous functions f on E. Here is a general result, which we will not prove, on characterizations of weak convergence.

Theorem 5.1.1. *The following are equivalent*

(a) $\mu_n \to \mu$ *weakly on* E,
(b) $\limsup_n \mu_n(C) \leq \mu(C)$ *for all closed sets* C,
(c) $\liminf_n \mu_n(G) \geq \mu(G)$ *for all open sets* G,
(d) $\lim_n \mu_n(A) = \mu(A)$ *for all Borel sets* A *with* $\mu(\partial A) = 0$.

Here is a result of similar type for the case $E = \mathbb{R}$. A proof that (b) implies (c) is given in [PM, Theorem 2.5.2].

Proposition 5.1.2. *Let* μ_n *and* μ *be probability measures on* $\mathbb{R}$. *Denote by* F_n *and* F *the corresponding distribution functions. The following are equivalent*

(a) $\mu_n \to \mu$ *weakly on* $\mathbb{R}$,
(b) $F_n(x) \to F(x)$ *for all* $x \in \mathbb{R}$ *such that* $F(x-) = F(x)$,
(c) *on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *for all* n, *there exist random variables* X *and* X_n, *with laws* μ *and* μ_n *respectively, such that* $X_n \to X$ *almost surely.*

5.2. **Prohorov's theorem.** A sequence of probability measures $(\mu_n : n \in \mathbb{N})$ on a metric space E is said to be *tight* if, for all $\varepsilon > 0$, there exists a compact set K such that $\mu_n(E\backslash K) \leq \varepsilon$ for all n.

Theorem 5.2.1 (Prohorov's theorem)**.** *Let $(\mu_n : n \in \mathbb{N})$ be a tight sequence of probability measures on E. Then there exists a subsequence (n_k) and a probability measure μ on E such that $\mu_{n_k} \to \mu$ weakly on E.*

Proof for the case $E = \mathbb{R}$. Write F_n for the distribution function of μ_n. By a diagonal argument and by passing to a subsequence, it suffices to consider the case where $F_n(x)$ converges, with limit $g(x)$ say, for all rationals x. Then g is non-decreasing on the rationals, so has a non-decreasing extension G to $\mathbb{R}$, and G has at most countably many discontinuities. It is easy to check that, if G is continuous at $x \in \mathbb{R}$, then $F_n(x) \to G(x)$. Set $F(x) = G(x+)$. Then F is non-decreasing and right-continuous and $F_n(x) \to F(x)$ at every point of continuity x of F. By tightness, for every $\varepsilon > 0$, there exists $R < \infty$ such that $F_n(-R) \leq \varepsilon$ and $F_n(R) \geq 1 - \varepsilon$ for all n. It follows that $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$, so F is a distribution function. The result now follows from Proposition 5.1.2. $\square$

5.3. **Weak convergence and characteristic functions.** For a probability measure μ on $\mathbb{R}^d$, we define the *characteristic function* ϕ by

$$\phi(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle}\mu(dx), \quad u \in \mathbb{R}^d.$$

Lemma 5.3.1. *Let μ be a probability measure on $\mathbb{R}$ with characteristic function ϕ. Then*

$$\mu(|y| \geq \lambda) \leq C\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\phi(u))du$$

for all $\lambda \in (0, \infty)$, where $C = (1 - \sin 1)^{-1} < \infty$.

Proof. It is elementary to check that, for all $t \geq 1$,

$$Ct^{-1} \int_0^t (1 - \cos v)dv \geq 1.$$

By a substitution, we deduce that, for all $y \in \mathbb{R}$,

$$1_{|y|\geq\lambda} \leq C\lambda \int_0^{1/\lambda} (1 - \cos uy)du.$$

Then, by Fubini's theorem,

$$\mu(|y| \geq \lambda) \leq C\lambda \int_{\mathbb{R}} \int_0^{1/\lambda} (1 - \cos uy)du\mu(dy) = C\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\phi(u))du.$$

$\square$

Theorem 5.3.2. *Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on $\mathbb{R}^d$ and let μ be another probability measure on $\mathbb{R}^d$. Write ϕ_n and ϕ for the characteristic functions of μ_n and μ respectively. Then the following are equivalent*

(a) *$\mu_n \to \mu$ weakly on $\mathbb{R}^d$,*
(b) *$\phi_n(u) \to \phi(u)$, for all $u \in \mathbb{R}^d$.*

Proof for $d = 1$. It is clear that (a) implies (b). Suppose then that (b) holds. Since ϕ is a characteristic function, it is continuous at 0, with $\phi(0) = 1$. So, given $\varepsilon > 0$, we can find $\lambda < \infty$ such that

$$C\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\phi(u))du \leq \varepsilon/2.$$

By bounded convergence we have

$$\int_0^{1/\lambda} (1 - \operatorname{Re}\phi_n(u))du \to \int_0^{1/\lambda} (1 - \operatorname{Re}\phi(u))du$$

as $n \to \infty$. So, for n sufficiently large,

$$\mu_n(|y| \geq \lambda) \leq \varepsilon.$$

Hence the sequence $(\mu_n : n \in \mathbb{N})$ is tight. By Prohorov's theorem, there is at least one weak limit point ν.

Fix a bounded continuous function f on $\mathbb{R}$ and suppose for a contradiction that $\mu_n(f) \not\to \mu(f)$. Then there is a subsequence (n_k) such that $|\mu_{n_k}(f) - \mu(f)| \geq \varepsilon$ for all k, for some $\varepsilon > 0$. But then, by the argument just given, we may choose (n_k) so that moreover μ_{n_k} converges weakly on $\mathbb{R}$, with limit ν say. Then $\phi_{n_k}(u) \to \psi(u)$ for all u, where ψ is the characteristic function of ν. But then $\psi = \phi$ so $\nu = \mu$, by uniqueness of characteristic functions [PM, Theorem 7.7.1], so $\mu_{n_k}(f) \to \mu(f)$, which is impossible. It follows that $\mu_n \to \mu$ weakly on $\mathbb{R}$. □

The argument just given in fact establishes the following stronger result (in the case $d = 1$).

Theorem 5.3.3 (Lévy's continuity theorem)**.** *Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on $\mathbb{R}^d$. Let μ_n have characteristic function ϕ_n and suppose that $\phi_n(u) \to \phi(u)$ for all $u \in \mathbb{R}^d$, for some function ϕ which is continuous at 0. Then ϕ is the characteristic function of a probability measure μ and $\mu_n \to \mu$ weakly on $\mathbb{R}^d$.*

6. Large deviations

In some probability models, one is concerned not with typical behaviour but with rare events, say of a catastrophic nature. The study of probabilities of rare events, in certain structured asymptotic contexts, is known as the study of *large deviations*. We will illustrate how this may be done in a simple case.

6.1. Cramér's theorem.

Theorem 6.1.1 (Cramér's theorem)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed, integrable random variables. Set*

$$m = \mathbb{E}(X_1), \quad S_n = X_1 + \cdots + X_n.$$

Then, for all $a \geq m$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -\psi^*(a)$$

where ψ^* is the Legendre transform *of the* cumulant generating function *ψ, given by*

$$\psi(\lambda) = \log \mathbb{E}(e^{\lambda X_1}), \quad \psi^*(x) = \sup_{\lambda \geq 0}\{\lambda x - \psi(\lambda)\}.$$

Before giving the proof, we discuss two simple examples. Consider first the case where X_1 has $N(0,1)$ distribution. Then $\psi(\lambda) = \lambda^2/2$ and so $\psi^*(x) = x^2/2$. Thus we find that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -\frac{a^2}{2}.$$

Since S_n has $N(0,n)$ distribution, it is straightforward to check this directly.

Consider now a second example, where X_1 has exponential distribution of parameter 1. Then

$$\mathbb{E}(e^{\lambda X_1}) = \int_0^\infty e^{\lambda x} e^{-x} dx = \begin{cases} 1/(1-\lambda), & \text{if } \lambda < 1, \\ \infty, & \text{otherwise} \end{cases}$$

so

$$\psi^*(x) = x - 1 - \log x.$$

In this example, for $a \geq 1$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -(a - 1 - \log a).$$

According to the central limit theorem, $(S_n - n)/\sqrt{n}$ converges in distribution to $N(0,1)$. Thus, for all $a \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}(S_n \geq n + a\sqrt{n}) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

However that the large deviations for S_n do not show the same behaviour as $N(0,1)$.

The proof of Cramér's theorem relies on certain properties of the functions ψ and ψ^* which we collect in the next two results. Write μ for the distribution of X_1 on $\mathbb{R}$. We exclude the trivial case $\mu = \delta_m$, for which the theorem may be checked directly. For $\lambda \geq 0$ with $\psi(\lambda) < \infty$, define the *tilted distribution* μ_λ by

$$\mu_\lambda(dx) \propto e^{\lambda x} \mu(dx).$$

For $K \geq m$, define the conditioned distribution $\mu(.|x \leq K)$ by

$$\mu(dx|x \leq K) \propto 1_{\{x \leq K\}} \mu(dx).$$

The associated cumulant generating function ψ_K and Legendre transform ψ_K^* are then given, for $\lambda \geq 0$ and $x \geq m$, by

$$\psi_K(\lambda) = \log \mathbb{E}(e^{\lambda X_1} | X_1 \leq K), \quad \psi_K^*(x) = \sup_{\lambda \geq 0} \{\lambda x - \psi_K(\lambda)\}.$$

Note that $m_K \uparrow m$ as $K \to \infty$, where $m_K = \mathbb{E}(X_1 | X_1 \leq K)$.

Proposition 6.1.2. *Assume that X_1 is integrable and not almost surely constant. For all $K \geq m$ and all $\lambda \geq 0$, we have $\psi_K(\lambda) < \infty$ and*

$$\psi_K(\lambda) \uparrow \psi(\lambda) \quad \text{as } K \to \infty.$$

Moreover, in the case where $\psi(\lambda) < \infty$ for all $\lambda \geq 0$, the function ψ has a continuous derivative on $[0,\infty)$ and is twice differentiable on $(0,\infty)$, with

$$\psi'(\lambda) = \int_{\mathbb{R}} x \mu_\lambda(dx), \quad \psi''(\lambda) = \text{var}(\mu_\lambda)$$

and ψ' maps $[0,\infty)$ homeomorphically to $[m, \sup(\text{supp}(\mu)))$.

Lemma 6.1.3. *Let $a \geq m$ be such that $\mathbb{P}(X_1 > a) > 0$. Then*

$$\psi_K^*(a) \downarrow \psi^*(a) \quad \text{as } K \to \infty.$$

Moreover, in the case where $\psi(\lambda) < \infty$ for all $\lambda \geq 0$, the function ψ^* is continuous at a, with

$$\psi^*(a) = \lambda^* a - \psi(\lambda^*)$$

where $\lambda^* \geq 0$ is determined uniquely by $\psi'(\lambda^*) = a$.

Proof. Suppose for now that $\psi(\lambda) < \infty$ for all $\lambda \geq 0$. Then the map $\lambda \mapsto \lambda a - \psi(\lambda)$ is strictly concave on $[0, \infty)$ with unique stationary point λ^* determined by $\psi'(\lambda^*) = a$. Hence

$$\psi^*(a) = \sup_{\lambda \geq 0}\{\lambda x - \psi(\lambda)\} = \lambda^* a - \psi(\lambda^*)$$

and ψ^* is continuous at a because ψ' is a homeomorphism.

We return to the general case and note first that $\psi_K^*(a)$ is non-increasing in K, with $\psi_K^*(a) \geq \psi^*(a)$ for all K. For K sufficiently large, we have

$$\mathbb{P}(X_1 > a | X_1 \leq K) > 0$$

and $a \geq m \geq m_K$, and $\psi_K(\lambda) < \infty$ for all $\lambda \geq 0$, so we may apply the preceding argument to μ_K to see that

$$\psi_K^*(a) = \lambda_K^* a - \psi_K(\lambda_K^*)$$

where $\lambda_K^* \geq 0$ is determined by $\psi_K'(\lambda_K^*) = a$. Now $\psi_K'(\lambda)$ is non-decreasing in K and λ, so $\lambda_K^* \downarrow \lambda^*$ for some $\lambda^* \geq 0$. Also $\psi_K'(\lambda) \geq m_K$ for all $\lambda \geq 0$, so

$$\psi_K(\lambda_K^*) \geq \psi_K(\lambda^*) + m_K(\lambda_K^* - \lambda^*).$$

Then

$$\psi_K^*(a) = \lambda_K^* a - \psi_K(\lambda_K^*) \leq \lambda_K^* a - \psi_K(\lambda^*) - m_K(\lambda_K^* - \lambda^*) \to \lambda^* a - \psi(\lambda^*) \leq \psi^*(a).$$

So $\psi_K^*(a) \downarrow \psi^*(a)$ as $K \to \infty$ as claimed. $\square$

Proof of Theorem 6.1.1. First we prove an upper bound. Fix $a \geq m$ and note that, for all $n \geq 1$ and all $\lambda \geq 0$,

$$\mathbb{P}(S_n \geq an) \leq \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda an}) \leq e^{-\lambda an}\mathbb{E}(e^{\lambda S_n}) = e^{-(\lambda a - \psi(\lambda))n}$$

so

$$\log \mathbb{P}(S_n \geq an) \leq -(\lambda a - \psi(\lambda))n$$

and so, on optimizing over $\lambda \geq 0$, we obtain

$$\log \mathbb{P}(S_n \geq an) \leq -\psi^*(a)n.$$

The proof will be completed by proving a complementary lower bound. Consider first the case where $\mathbb{P}(X_1 \leq a) = 1$. Set $p = \mathbb{P}(X_1 = a)$. Then $\mathbb{E}(e^{\lambda(X_1 - a)}) \to p$ as $\lambda \to \infty$ by bounded convergence, so

$$\lambda a - \psi(\lambda) = -\log \mathbb{E}(e^{\lambda(X_1 - a)}) \to -\log p$$

and hence $\psi^*(a) \geq -\log p$. Now, for all $n \geq 1$, we have $\mathbb{P}(S_n \geq an) = p^n$, so

$$\log \mathbb{P}(S_n \geq an) \geq -\psi^*(a)n.$$

When combined with the upper bound, this proves the claimed limit.

Consider next the case where $\mathbb{P}(X_1 > a) > 0$ and $\psi(\lambda) < \infty$ for all $\lambda \geq 0$. Fix $\varepsilon > 0$ and set $b = a + \varepsilon$ and $c = a + 2\varepsilon$. We choose ε small enough so that $\mathbb{P}(X_1 > b) > 0$. Then there exists $\lambda > 0$ such that $\psi'(\lambda) = b$. Fix $n \geq 1$ and define a new probability measure $\mathbb{P}_\lambda$ by

$$d\mathbb{P}_\lambda = e^{\lambda S_n - \psi(\lambda)n} d\mathbb{P}.$$

Under $\mathbb{P}_\lambda$, the random variables $X_1, \dots, X_n$ are independent, with distribution μ_λ, so $\mathbb{E}_\lambda(X_1) = \psi'(\lambda) = b$. Consider the event

$$A_n = \{|S_n/n - b| \leq \varepsilon\} = \{an \leq S_n \leq cn\}.$$

Then $\mathbb{P}_\lambda(A_n) \to 1$ as $n \to \infty$ by the weak law of large numbers. Now

$$\mathbb{P}(S_n \geq an) \geq \mathbb{P}(A_n) = \mathbb{E}_\lambda(e^{-\lambda S_n + \psi(\lambda)n} 1_{A_n}) \geq e^{-\lambda cn + \psi(\lambda)n} \mathbb{P}_\lambda(A_n)$$

so

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) \geq -\lambda c + \psi(\lambda) \geq -\psi^*(c).$$

On letting $\varepsilon \to 0$, we have $c \to a$, so $\psi^*(c) \to \psi^*(a)$. Hence

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) \geq -\psi^*(a)$$

which, when combined with the upper bound, gives the claimed limit.

It remains to deal with the case where $\mathbb{P}(X_1 > a) > 0$ without the restriction that $\psi(\lambda) < \infty$ for all $\lambda \geq 0$. Fix $n \geq 1$ and $K \in (a, \infty)$ and define a new probability measure $\mathbb{P}_K$ by

$$d\mathbb{P}_K \propto 1_{\{X_1 \leq K, \dots, X_n \leq K\}} d\mathbb{P}.$$

Under $\mathbb{P}_K$, the random variables $X_1, \dots, X_n$ are independent, with common distribution $\mu(.|x \leq K)$. We have $a \geq m \geq \mathbb{E}(X_1 | X_1 \leq K)$ and $\psi_K(\lambda) < \infty$ for all $\lambda \geq 0$, so

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_K(S_n \geq an) \geq -\psi_K^*(a).$$

But

$$\mathbb{P}(S_n \geq an) \geq \mathbb{P}_K(S_n \geq an)$$

and $\psi_K^*(a) \downarrow \psi^*(a)$ as $K \to \infty$. Hence

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) \geq -\psi^*(a)$$

which is the desired lower bound. $\square$

7. Brownian motion

7.1. **Definition.** Let $(X_t)_{t\geq 0}$ be a random process with state-space $\mathbb{R}^d$. We say that $(X_t)_{t\geq 0}$ is a *Brownian motion* if

(i) for all $s,t\geq 0$, the random variable $X_{s+t}-X_s$ is Gaussian, of mean 0 and variance tI, and is independent of $\mathcal{F}_s^X=\sigma(X_u:u\leq s)$,
(ii) for all $\omega\in\Omega$, the map $t\mapsto X_t(\omega):[0,\infty)\to\mathbb{R}^d$ is continuous.

Condition (i) states that, for all $s\geq 0$, all $t>0$, any Borel set $B\subseteq\mathbb{R}^d$ and any $A\in\mathcal{F}_s^X$,

$$\mathbb{P}(\{X_{s+t}-X_s\in B\}\cap A)=\mathbb{P}(A)\int_B(2\pi t)^{-d/2}e^{-|y|^2/(2t)}dy.$$

By a monotone class argument, it is equivalent to the following statement expressed in terms of conditional expectation: for all $s,t\geq 0$, and for all bounded measurable functions f on $\mathbb{R}^d$, almost surely,

$$\mathbb{E}(f(X_{s+t})|\mathcal{F}_s^X)=P_tf(X_s)$$

where $(P_t)_{t\geq 0}$ is the *heat semigroup*, given by $P_0f=f$ and, for $t>0$,

$$P_tf(x)=\int_{\mathbb{R}^d}p(t,x,y)f(y)dy. \tag{7.1}$$

Here $p(t,x,.)$ is the Gaussian probability density function on $\mathbb{R}^d$ of mean x and variance tI, given by

$$p(t,x,y)=(2\pi t)^{-d/2}e^{-|x-y|^2/(2t)}.$$

In the case where $X_0=x$ for some $x\in\mathbb{R}^d$, we call $(X_t)_{t\geq 0}$ a *Brownian motion starting from* x. Then condition (i) may also be expressed in terms of the finite-dimensional distributions of $(X_t)_{t\geq 0}$ as follows: for all $n\in\mathbb{N}$ and all $t_1,\dots,t_n\geq 0$ with $t_1<\dots<t_n$, for any Borel set $B\subseteq(\mathbb{R}^d)^n$,

$$\mathbb{P}((X_{t_1},\dots,X_{t_n})\in B)=\int_B\prod_{i=1}^n p(s_i,x_{i-1},x_i)dx_i$$

where $x_0=x$, $t_0=0$ and $s_i=t_i-t_{i-1}$. It follows easily from the definition that, if $(X_t^1)_{t\geq 0},\dots,(X_t^d)_{t\geq 0}$ are independent one-dimensional Brownian motions starting from 0, and if $x\in\mathbb{R}^d$, then the process $(X_t)_{t\geq 0}$, given by

$$X_t=x+(X_t^1,\dots,X_t^d)$$

is a Brownian motion in $\mathbb{R}^d$ starting from x, and we can obtain every Brownian motion in $\mathbb{R}^d$ starting from x in this way.

For a Brownian motion $(X_t)_{t\geq 0}$ in $\mathbb{R}^d$ starting from x, for all $s,t\geq 0$ and all $i,j=1,\dots,d$,

$$\mathbb{E}(X_t)=x,\quad \operatorname{cov}(X_s^i,X_t^j)=\mathbb{E}(X_s^iX_t^j)=(s\wedge t)\delta_{ij}. \tag{7.2}$$

Recall that Gaussian distributions are determined by their means and covariances. Hence, given that $(X_t)_{t\geq 0}$ is a continuous Gaussian process, the simple properties (7.2) determine the finite-dimensional distributions $(X_t)_{t\geq 0}$ and hence characterize $(X_t)_{t\geq 0}$ as a Brownian motion. This provides a convenient way to identify certain linear transformations of Brownian motion as Brownian motions themselves.

7.2. **Wiener's theorem.** Write W_d for the set of continuous paths $C([0,\infty),\mathbb{R}^d)$. For $t \geq 0$, define the *coordinate function* $x_t : W_d \to \mathbb{R}^d$ by $x_t(w) = w(t)$. We equip W_d with the σ-algebra $\mathcal{W}_d = \sigma(x_t : t \geq 0)$. Given any continuous random process $(X_t)_{t\geq 0}$ with state-space $\mathbb{R}^d$, we can define a measurable map $X : \Omega \to W_d$ and a probability measure μ on $(W_d, \mathcal{W}_d)$ by

$$X(\omega)(t) = X_t(\omega), \quad \mu(A) = \mathbb{P}(X \in A).$$

Then μ is called the *law* of $(X_t)_{t\geq 0}$ on W_d. The measure μ_x identified in the next theorem is called *Wiener measure starting from* x.

Theorem 7.2.1 (Wiener's theorem)**.** *For all $d \geq 1$ and all $x \in \mathbb{R}^d$, there exists a unique probability measure μ_x on $(W_d, \mathcal{W}_d)$ such that the coordinate process $(x_t)_{t\geq 0}$ is a Brownian motion starting from x.*

Proof. Conditions (i) and (ii) determine the finite-dimensional distributions of any such measure μ_x, so there can be at most one. To prove existence, it will suffice to construct, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a one-dimensional Brownian motion $(X_t)_{t\geq 0}$ starting from 0. Then, for $d = 1$ and $x \in \mathbb{R}$, the law μ_x of $(x + X_t)_{t\geq 0}$ on $(W_1, \mathcal{W}_1)$ is a measure with the required property, and, for $d \geq 2$ and $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, the product measure $\mu_x = \mu_{x_1} \otimes \cdots \otimes \mu_{x_d}$ on $(W_d, \mathcal{W}_d)$ has the required property.

For $N \geq 0$ denote (as above) by $\mathbb{D}_N$ the set of integer multiples of 2^{-N} in $[0,\infty)$ and denote by $\mathbb{D}$ the union of these sets. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is defined a family of independent standard Gaussian random variables $(Y_t : t \in \mathbb{D})$. For $t \in \mathbb{D}_0 = \mathbb{Z}^+$, set $\xi_t = Y_1 + \cdots + Y_t$. Define recursively, for $N \geq 0$ and $t \in \mathbb{D}_{N+1} \setminus \mathbb{D}_N$,

$$\xi_t = \frac{\xi_r + \xi_s}{2} + Z_t$$

where $r = t - 2^{-N-1}$, $s = t + 2^{-N-1}$ and $Z_t = 2^{-(N+2)/2} Y_t$. Note that the random variables $(\xi_t : t \in \mathbb{D})$ are jointly Gaussian and have mean 0, and that $(\xi_{t+1} - \xi_t : t \in \mathbb{D}_0)$ is a sequence of independent zero-mean Gaussian random variables of variance 1.

Suppose inductively for $N \geq 0$ that $(\xi_{t+2^{-N}} - \xi_t : t \in \mathbb{D}_N)$ is a sequence of independent zero-mean Gaussian random variables of variance 2^{-N}. Consider the sequence $(\xi_{t+2^{-N-1}} - \xi_t : t \in \mathbb{D}_{N+1})$. Fix $t \in \mathbb{D}_{N+1} \setminus \mathbb{D}_N$ and note that

$$\xi_t - \xi_r = \frac{\xi_s - \xi_r}{2} + Z_t, \quad \xi_s - \xi_t = \frac{\xi_s - \xi_r}{2} - Z_t.$$

Now

$$\operatorname{var}\left(\frac{\xi_s - \xi_r}{2}\right) = 2^{-N-2} = \operatorname{var}(Z_t)$$

so

$$\operatorname{var}(\xi_t - \xi_r) = \operatorname{var}(\xi_s - \xi_t) = 2^{-N-1}, \quad \operatorname{cov}(\xi_t - \xi_r, \xi_s - \xi_t) = 0.$$

On the other hand,

$$\operatorname{cov}(\xi_t - \xi_r, \xi_v - \xi_u) = \operatorname{cov}(\xi_s - \xi_t, \xi_v - \xi_u) = 0$$

for any $u, v \in \mathbb{D}_{N+1}$ with $(u, v] \cap (r, s] = 0$. Hence $\xi_t - \xi_r$ and $\xi_s - \xi_t$ are independent zero-mean Gaussian random variables of variance 2^{-N-1}, which are independent also of $\xi_v - \xi_u$ for all such u, v. The induction proceeds.

It follows that $(\xi_t)_{t\in\mathbb{D}}$ has independent increments and, for all $s,t\in\mathbb{D}$ with $s<t$, $\xi_t-\xi_s$ is a zero-mean Gaussian random variable of variance $t-s$. Choose $p>2$ and set $C_p=\mathbb{E}(|\xi_1|^p)$. Then $C_p<\infty$ and

$$\mathbb{E}(|\xi_t-\xi_s|^p)\le C_p(t-s)^{p/2}.$$

Hence, by Kolmogorov's criterion, there is a continuous process $(X_t)_{t\ge 0}$ starting from 0 such that $X_t=\xi_t$ for all $t\in\mathbb{D}$ almost surely.

Let $s\ge 0$ and $t>0$, and let $A\in\mathcal{F}_s^X$. There exist sequences $(s_n:n\in\mathbb{N})$ and $(t_n:n\in\mathbb{N})$ in $\mathbb{D}$ such that $s_n\ge s$ and $t_n>0$ for all n, and such that $s_n\to s$ and $t_n\to t$. Also, there exists $A_0\in\sigma(\xi_u:u\le s,u\in\mathbb{D})$ such that $1_A=1_{A_0}$ almost surely. Then, for any continuous bounded function f on $\mathbb{R}^d$,

$$\mathbb{E}(f(X_{s_n+t_n}-X_{s_n})1_A)=\mathbb{E}(f(\xi_{s_n+t_n}-\xi_{s_n})1_{A_0})=\mathbb{P}(A_0)\int_{\mathbb{R}^d}p(t_n,0,y)f(y)dy$$

so, on letting $n\to\infty$, by bounded convergence,

$$\mathbb{E}(f(X_{s+t}-X_s)1_A)=\mathbb{P}(A)\int_{\mathbb{R}^d}p(t,0,y)f(y)dy.$$

Hence $(X_t)_{t\ge 0}$ is a Brownian motion. $\square$

7.3. **Symmetries of Brownian motion.** In the study of Brownian motion, we can take advantage of the following symmetries.

Proposition 7.3.1. *Let $(X_t)_{t\ge 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0. Let $\sigma\in(0,\infty)$ and let U be an orthogonal $d\times d$-matrix. Then*

(a) *$(\sigma X_{\sigma^{-2}t})_{t\ge 0}$ is a Brownian motion in $\mathbb{R}^d$ starting from 0,*
(b) *$(UX_t)_{t\ge 0}$ is a Brownian motion in $\mathbb{R}^d$ starting from 0.*

We call (a) the *scaling property* and we call (b) *rotation invariance.*

7.4. **Brownian motion in a given filtration.** Suppose given a filtration $(\mathcal{F}_t)_{t\ge 0}$ on $(\Omega,\mathcal{F},\mathbb{P})$. Let $(X_t)_{t\ge 0}$ be a random process with state-space $\mathbb{R}^d$. We say that $(X_t)_{t\ge 0}$ is an $(\mathcal{F}_t)_{t\ge 0}$-Brownian motion if

(i) $(X_t)_{t\ge 0}$ is $(\mathcal{F}_t)_{t\ge 0}$-adapted,
(ii) for all $s,t\ge 0$, the random variable $X_{s+t}-X_s$ is Gaussian, of mean 0 and variance tI, and is independent of $\mathcal{F}_s$,
(iii) for all $\omega\in\Omega$, the map $t\mapsto X_t(\omega):[0,\infty)\to\mathbb{R}^d$ is continuous.

Then every $(\mathcal{F}_t)_{t\ge 0}$-Brownian motion is a Brownian motion, in the sense used above, and every Brownian motion is an $(\mathcal{F}_t^X)_{t\ge 0}$-Brownian motion.

The following result allows us to compute conditional expectations for Brownian motion in terms of Wiener measure.

Proposition 7.4.1. *Let $(X_t)_{t\ge 0}$ be an $(\mathcal{F}_t)_{t\ge 0}$-Brownian motion and let F be a bounded measurable function on W_d. Define a function f on $\mathbb{R}^d$ by*

$$f(x)=\int_{W_d}F(w)\mu_x(dw).$$

Then f is measurable and, almost surely,

$$\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0).$$

7.5. **Martingales of Brownian motion.** We now identify a useful class of martingales associated to Brownian motion, in terms of the *Laplacian* Δ on $\mathbb{R}^d$, which is the second-order differential operator given by

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}.$$

Theorem 7.5.1. *Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion and let $f \in C_b^2(\mathbb{R}^d)$. Define $(M_t)_{t\geq 0}$ by*

$$M_t = f(X_t) - f(X_0) - \int_0^t \tfrac{1}{2}\Delta f(X_s)ds.$$

Then $(M_t)_{t\geq 0}$ is a continuous $(\mathcal{F}_t)_{t\geq 0}$-martingale.

Proof. It is clear that $(M_t)_{t\geq 0}$ is continuous, adapted and integrable. For simplicity of writing, we do the case $d = 1$. The argument for general d is the same. Fix $T \geq 0$ and set

$$\delta_n = \sup\{|X_s - X_t| : s, t \leq T, |s-t| \leq 2^{-n}\},$$
$$\varepsilon_n = \sup\{|f''(y) - f''(X_t)| : t \leq T, |y - X_t| \leq \delta_n\}.$$

Then $\varepsilon_n \leq 2\|f''\|_\infty$ for all n and $\delta_n \to 0$ and hence $\varepsilon_n \to 0$ almost surely since f'' is continuous. Hence $\|\varepsilon_n\|_2 \to 0$ by bounded convergence.

We will show that, for $r \leq t \leq T$ with $t - r \leq 2^{-n}$,

$$\|\mathbb{E}(M_t - M_r|\mathcal{F}_r)\|_1 \leq 3(t-r)\|\varepsilon_n\|_2.$$

Then, by the tower property, for $s \leq r \leq t \leq T$ with $t - r \leq 2^{-n}$,

$$\|\mathbb{E}(M_t - M_r|\mathcal{F}_s)\|_1 \leq 3(t-r)\|\varepsilon_n\|_2.$$

By the triangle inequality, the condition $t - r \leq 2^{-n}$ may be dropped. Then on taking $r = s$ and letting $n \to \infty$ we obtain

$$\|\mathbb{E}(M_t|\mathcal{F}_s) - M_s\|_1 = 0$$

so $(M_t)_{t\geq 0}$ has the martingale property, as claimed.

Fix $s < t \leq T$ with $t - s \leq 2^{-n}$. By Taylor's theorem

$$f(X_t) = f(X_s) + (X_t - X_s)f'(X_s) + \tfrac{1}{2}(X_t - X_s)^2 f''(X_s) + (X_t - X_s)^2 E_1(s,t)$$

where

$$E_1(s,t) = \int_0^1 (1-u)(f''(uX_t + (1-u)X_s) - f''(X_s))du.$$

Also

$$\int_s^t \tfrac{1}{2} f''(X_r)dr = \tfrac{1}{2}(t-s)f''(X_s) + (t-s)E_2(s,t)$$

where

$$(t-s)E_2(s,t) = \int_s^t \tfrac{1}{2}(f''(X_r) - f''(X_s))dr.$$

Note that $|E_1(s,t)| \leq \varepsilon_n$ and $|E_2(s,t)| \leq \varepsilon_n$. Now

$$\begin{aligned} M_t - M_s &= f(X_t) - f(X_s) - \int_s^t \tfrac{1}{2} f''(X_r)dr \\ &= (X_t - X_s)f'(X_s) + ((X_t - X_s)^2 - (t-s))f''(X_s) \\ &\qquad + (X_t - X_s)^2 E_1(s,t) - (t-s)E_2(s,t). \end{aligned}$$

and

$$\mathbb{E}((X_t - X_s)f'(X_s)|\mathcal{F}_s) = f'(X_s)\mathbb{E}(X_t - X_s|\mathcal{F}_s) = 0,$$
$$\mathbb{E}(((X_t - X_s)^2 - (t-s))f''(X_s)|\mathcal{F}_s) = f''(X_s)\mathbb{E}((X_t - X_s)^2 - (t-s)|\mathcal{F}_s) = 0.$$

Hence

$$\mathbb{E}(M_t - M_s|\mathcal{F}_s) = \mathbb{E}((X_t - X_s)^2 E_1(s,t)|\mathcal{F}_s) - (t-s)\mathbb{E}(E_2(s,t)|\mathcal{F}_s)$$

and so

$$\begin{aligned} \|\mathbb{E}(M_t - M_s|\mathcal{F}_s)\|_1 &\leq \|(X_t - X_s)^2 E_1(s,t)\|_1 + (t-s)\|E_2(s,t)\|_1 \\ &\leq \|(X_t - X_s)^2\|_2 \|E_1(s,t)\|_2 + (t-s)\|E_2(s,t)\|_2 \leq 3\|\varepsilon_n\|_2 \end{aligned}$$

where we used Cauchy–Schwarz and the fact that

$$\mathbb{E}((X_t - X_s)^4) = (t-s)^2 \mathbb{E}(X_1^4) = 3(t-s)^2.$$

□

Proof. It is clear that $(M_t)_{t \geq 0}$ is continuous, adapted and integrable. Fix $t > 0$. Consider for now the case where $X_0 = x$ for some $x \in \mathbb{R}^d$ and set

$$m(x) = \mathbb{E}(M_t) = \int_{W_d} \left(f(w(t)) - f(w(0)) - \int_0^t \tfrac{1}{2}\Delta f(w(s))ds \right) \mu_x(dw).$$

Then, for all $s \in (0,t]$, by Fubini,

$$\begin{aligned} \mathbb{E}(M_t - M_s) &= \mathbb{E}\left(f(X_t) - f(X_s) - \int_s^t \tfrac{1}{2}\Delta f(X_r)dr \right) \\ &= \mathbb{E}\left(f(X_t)\right) - \mathbb{E}\left(f(X_s)\right) - \int_s^t \mathbb{E}\left(\tfrac{1}{2}\Delta f(X_r)\right) dr \\ &= \int_{\mathbb{R}^d} p(t,x,y)f(y)dy - \int_{\mathbb{R}^d} p(s,x,y)f(y)dy - \int_s^t \int_{\mathbb{R}^d} p(r,x,y)\tfrac{1}{2}\Delta f(y)dydr. \end{aligned}$$

Now p satisfies the heat equation $\dot{p} = \frac{1}{2}\Delta p$ so, on integrating by parts twice in $\mathbb{R}^d$, we obtain

$$\begin{aligned} \int_s^t \int_{\mathbb{R}^d} p(r,x,y)\tfrac{1}{2}\Delta f(y)dydr &= \int_s^t \int_{\mathbb{R}^d} \dot{p}(r,x,y)f(y)dydr \\ &= \int_{\mathbb{R}^d} p(t,x,y)f(y)dy - \int_{\mathbb{R}^d} p(s,x,y)f(y)dy. \end{aligned}$$

Hence $\mathbb{E}(M_t) = \mathbb{E}(M_s)$. On letting $s \to 0$, we have $M_s \to 0$, so $\mathbb{E}(M_s) \to 0$ by bounded convergence. Hence $m(x) = \mathbb{E}(M_t) = 0$.

We return to the case of general initial state X_0. Then, by Proposition 7.4.1, almost surely,

$$\mathbb{E}(M_t|\mathcal{F}_0) = m(X_0) = 0.$$

Finally, for all $s \geq 0$, since

$$M_{s+t} - M_s = f(X_{s+t}) - f(X_s) - \int_0^t \tfrac{1}{2}\Delta f(X_{s+r})dr$$

we may apply the preceding formula to $(X_{s+t})_{t\geq 0}$, which is an $(\mathcal{F}_{s+t})_{t\geq 0}$-Brownian motion, to obtain, almost surely,

$$\mathbb{E}(M_{s+t} - M_s|\mathcal{F}_s) = 0$$

showing that $(M_t)_{t\geq 0}$ is a martingale. □

In Theorem 7.5.1, the conditions of boundedness on f and its derivative can be relaxed, while taking care that $(M_t)_{t\geq 0}$ remains integrable and the integrations by parts remain valid. There is a natural alternative proof via Itô's formula once one has access to stochastic calculus.

7.6. **Strong Markov property.**

Theorem 7.6.1 (Strong Markov property)**.** *Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion and let T be a stopping time. Then, conditional on $\{T < \infty\}$, the process $(X_{T+t})_{t\geq 0}$ is an $(\mathcal{F}_{T+t})_{t\geq 0}$-Brownian motion.*

Proof. It is clear that $(X_{T+t})_{t\geq 0}$ is continuous on $\{T < \infty\}$. Also X_{T+t} is $\mathcal{F}_{T+t}$-measurable on $\{T < \infty\}$ for all $t \geq 0$, so $(X_{T+t})_{t\geq 0}$ is $(\mathcal{F}_{T+t})_{t\geq 0}$-adapted on $\{T < \infty\}$. Let f be a bounded continuous function on $\mathbb{R}^d$. Let $s \geq 0$ and $t > 0$ and let $m \in \mathbb{N}$ and $A \in \mathcal{F}_{T+s}$ with $A \subseteq \{T \leq m\}$. Fix $n \geq 1$ and set $T_n = 2^{-n}\lceil 2^n T\rceil$. For $k \in \{0, 1, \ldots, m2^n\}$, set $t_k = k2^{-n}$ and consider the event

$$A_k = A \cap \{T \in (t_k - 2^{-n}, t_k]\}.$$

Then $A_k \in \mathcal{F}_{t_k+s}$ and $T_n = t_k$ on A_k, so

$$\mathbb{E}(f(X_{T_n+s+t})1_{A_k}) = \mathbb{E}(f(X_{t_k+s+t})1_{A_k}) = \mathbb{E}(P_t f(X_{t_k+s})1_{A_k}) = \mathbb{E}(P_t f(X_{T_n+s})1_{A_k})$$

On summing over k, we obtain

$$\mathbb{E}(f(X_{T_n+t})1_A) = \mathbb{E}(P_{t-s} f(X_{T_n+s})1_A).$$

Then, by bounded convergence, on letting $n \to \infty$, we deduce that

$$\mathbb{E}(f(X_{T+s+t})1_A) = \mathbb{E}(P_t f(X_{T+s})1_A).$$

Since m and A were arbitrary, this implies that, almost surely on $\{T < \infty\}$,

$$\mathbb{E}(f(X_{T+s+t})|\mathcal{F}_{T+s}) = P_t f(X_{T+s})$$

so, conditional on $\{T < \infty\}$, $(X_{T+t})_{t\geq 0}$ is an $(\mathcal{F}_{T+t})_{t\geq 0}$-Brownian motion. □

7.7. **Properties of one-dimensional Brownian motion.**

Proposition 7.7.1. *Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}$ starting from 0. For $a \in \mathbb{R}$, set*

$$T_a = \inf\{t \geq 0 : X_t = a\}.$$

Then, for $a, b > 0$, we have

$$\mathbb{P}(T_a < \infty) = 1, \quad \mathbb{P}(T_{-a} < T_b) = b/(a+b), \quad \mathbb{E}(T_{-a} \wedge T_b) = ab.$$

Moreover, T_a has a density function f_a on $[0, \infty)$, given by

$$f_a(t) = (a/\sqrt{2\pi t^3})e^{-a^2/2t}.$$

Moreover, the following properties hold almost surely

(a) *$X_t/t \to 0$ as $t \to \infty$,*
(b) *$\inf_{t\geq 0} X_t = -\infty$ and $\sup_{t\geq 0} X_t = \infty$,*
(c) *for all $s \geq 0$, there exist $t, u \geq s$ with $X_t < 0 < X_u$,*
(d) *for all $s > 0$, there exist $t, u \leq s$ with $X_t < 0 < X_u$.*

Theorem 7.7.2. *Let X be a Brownian motion in $\mathbb{R}$. Then, almost surely,*

(a) *for all $\alpha < 1/2$, X is locally Hölder continuous of exponent α,*
(b) *for all $\alpha > 1/2$, X is not Hölder continuous of exponent α on any non-trivial interval.*

Proof. Fix $\alpha < 1/2$ and choose $p < \infty$ so that $\alpha < 1/2 - 1/p$. By scaling, we have

$$\|X_s - X_t\|_p \leq C|s-t|^{1/2}$$

where $C = \|X_1\|_p < \infty$. Then, by Kolmogorov's criterion, there exists $K \in L^p$ such that

$$|X_s - X_t| \leq K|s-t|^\alpha, \quad s, t \in [0, 1].$$

Hence, by scaling, X is locally Hölder continuous of exponent α, almost surely. Then (a) follows by considering a sequence $\alpha_n > 1/2$ with $\alpha_n \to 1/2$.

Define for $m, n \geq 0$ with $m \geq n$ and for $s, t \in \mathbb{D}_n$ with $s < t$.

$$[X]^m_{s,t} = \sum_\tau (X_{\tau+2^{-m}} - X_\tau)^2$$

where the sum is taken over all $\tau \in \mathbb{D}_m$ such that $s \leq \tau < t$. The random variables $(X_{\tau+2^{-m}} - X_\tau)^2$ are then independent, of mean 2^{-m} and variance 2^{-2m+1}. For the variance, we used scaling and the fact that $\mathrm{var}(X_1^2) = 2$. Hence

$$\mathbb{E}([X]^m_{s,t}) = t - s, \quad \mathrm{var}([X]^m_{s,t}) = 2^{-m+1}(t-s)$$

so $[X]^m_{s,t} \to t - s > 0$ almost surely as $m \to \infty$. On the other hand, if X is Hölder continuous of some exponent $\alpha > 1/2$ and constant K on $[s, t]$, then we have

$$(X_{\tau+2^{-m}} - X_\tau)^2 \leq K^2 2^{-2m\alpha}$$

so

$$[X]^m_{s,t} \leq K^2 2^{-2m\alpha+m}(t-s) \to 0.$$

Hence, almost surely, X is not Hölder continuous of any exponent $\alpha > 1/2$ on $[s, t]$. Now, for any non-trivial interval I, there exist $n \geq 0$ and $s, t \in \mathbb{D}_n$ with $s < t$ such that $[s, t] \subseteq I$.

Hence, almost surely, for all $\alpha > 1/2$, there is no non-trivial interval on which $(X_t)_{t\geq 0}$ is Hölder continuous of exponent α. $\square$

7.8. **Recurrence and transience of Brownian motion.** The statements in the following theorem are sometimes expressed by saying that Brownian motion in $\mathbb{R}$ is *point recurrent*, that Brownian motion in $\mathbb{R}^2$ is *neighbourhood recurrent* but does not *hit points* and that Brownian motion in $\mathbb{R}^d$ is *transient* for all $d \geq 3$.

Theorem 7.8.1. *Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$.*

(a) *In the case $d = 1$,*

$$\mathbb{P}(\{t \geq 0 : X_t = 0\} \text{ is unbounded}) = 1.$$

(b) *In the case $d = 2$,*

$$\mathbb{P}(X_t = 0 \text{ for some } t > 0) = 0$$

but, for any $\varepsilon > 0$,

$$\mathbb{P}(\{t \geq 0 : |X_t| < \varepsilon\} \text{ is unbounded}) = 1.$$

(c) *In the case $d \geq 3$,*

$$\mathbb{P}(|X_t| \to \infty \text{ as } t \to \infty) = 1.$$

Proof. By Proposition 7.4.1, it suffices to consider the case where $X_0 = x$ for some $x \in \mathbb{R}^d$. Then (a) follows easily from Proposition 7.7.1. We turn to (b). Set

$$p_0(x) = \mathbb{P}(X_t = 0 \text{ for some } t > 0) = \mu_x(w(t) = 0 \text{ for some } t > 0)$$

and, for $\varepsilon > 0$, set

$$p_\varepsilon(x) = \mathbb{P}(|X_t| < \varepsilon \text{ for some } t > 0) = \mu_x(|w(t)| < \varepsilon \text{ for some } t > 0).$$

Fix $a \in (0, 1)$ and $b \in (1, \infty)$. There exists a function $f \in C_b^2(\mathbb{R}^2)$ such that

$$f(x) = \log |x|, \quad \text{for } a \leq |x| \leq b.$$

Then, since $\log |x|$ is harmonic in $\mathbb{R}^2 \setminus \{0\}$, we have $\Delta f(x) = 0$ for $a \leq |x| \leq b$. Consider the process

$$M_t = f(X_t) - f(X_0) - \int_0^t \tfrac{1}{2}\Delta f(X_s)ds$$

and the stopping time

$$T = \inf\{t \geq 0 : |X_t| = a \text{ or } |X_t| = b\}.$$

Then $(M_t)_{t\geq 0}$ is a martingale, by Theorem 7.5.1, and $\mathbb{P}(T < \infty) = 1$ by (a). Hence, by optional stopping, since $(M_t)_{t\geq 0}$ is bounded up to T, we have

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) = 0.$$

Assume for now that $|x| = 1$. Then $M_T = \log |X_T|$. Set $p = p(a, b) = \mathbb{P}(|X_T| = a)$. Then

$$0 = \mathbb{E}(M_T) = p \log a + (1 - p) \log b.$$

Consider first the limit $a \to 0$ with b fixed. Then $\log a \to -\infty$ so $p(a, b) \to 0$. Hence $p_0(x) = 0$ whenever $|x| = 1$. A scaling argument extends this to all $x \neq 0$. In the remaining case, when $x = 0$, for all $n \geq 1$, by the Markov property,

$$\mathbb{P}(X_t = 0 \text{ for some } t > 1/n) = \int_{\mathbb{R}^2} p(1/n, 0, y)p_0(y)dy = 0.$$

Since $n \geq 1$ is arbitrary, we deduce that $p_0(0) = 0$.

Consider now the limit $b \to \infty$ with $a = \varepsilon > 0$ fixed. Then $\log b \to \infty$, so $p(a, b) \to 1$. Hence $p_\varepsilon(x) = 1$ whenever $|x| = 1$. A scaling argument extends this to all $x \neq 0$ and it is obvious by continuity for $x = 0$. Then, by the Markov property, for all x and all $n \geq 1$,

$$\mathbb{P}(|X_t| < \varepsilon \text{ for some } t > n) = \int_{\mathbb{R}^2} p(n, x, y) p_\varepsilon(y) dy = 1.$$

Since $n \geq 1$ is arbitrary, it follows that $\mathbb{P}(\{t \geq 0 : |X_t| < \varepsilon\} \text{ is unbounded}) = 1$.

We turn to the proof of (c). It will suffice to show, for all $N \geq 1$, that

$$\mathbb{P}(\{t \geq 0 : |X_t| \leq N\} \text{ is unbounded}) = 0.$$

Since the first three components of a Brownian motion in $\mathbb{R}^d$ form a Brownian motion in $\mathbb{R}^3$, it suffices to consider the case $d = 3$. The function $1/|x|$ is harmonic in $\mathbb{R}^3 \setminus \{0\}$. We adapt the argument for (b), replacing $\log |x|$ by $1/|x|$ to see that, in the case $X_0 = x$ with $|x| = 1$, we have

$$\frac{p}{a} + \frac{1-p}{b} = 1$$

so, on letting $b \to \infty$ we obtain, for $|x| = 1$,

$$\mu_x(|w(t)| = a \text{ for some } t \geq 0) = a.$$

Hence, by scaling, for all $N \geq 1$ and for $|x| = N + 1$,

$$\mu_x(|w(t)| = N \text{ for some } t \geq 0) = \frac{N}{N+1}.$$

Set $T_0 = 0$ and define sequences of stopping times $(S_k : k \geq 1)$ and $(T_k : k \geq 1)$ by

$$S_k = \inf\{t \geq T_{k-1} : |X_t| = N + 1\}, \quad T_k = \inf\{t \geq S_k : |X_t| = N\}.$$

For $k \geq 1$, we can apply the strong Markov property at S_k to see that

$$\mathbb{P}(T_k < \infty) \leq \mathbb{P}(T_{k-1} < \infty) \frac{N}{N+1} \leq \left(\frac{N}{N+1} \right)^k.$$

Set $K = \sup\{k \geq 0 : T_k < \infty\}$. Then $K < \infty$ almost surely. Now $S_{k+1} < \infty$ almost surely on $\{T_k < \infty\}$ for all $k \geq 0$, so $S_{K+1} < \infty$ almost surely. But $|X_t| > N$ for all $t \geq S_{K+1}$, so we have shown that

$$\mathbb{P}(\{t \geq 0 : |X_t| \leq N\} \text{ is unbounded }) = 0.$$

□

7.9. **Brownian motion and the Dirichlet problem.** Let D be a connected open set in $\mathbb{R}^d$ with boundary ∂D and let $c : D \to [0, \infty)$ and $f : \partial D \to [0, \infty)$ be measurable functions. We shall be interested in the *potential* or *expected total cost* function ϕ, defined on the closure $\bar{D}$ of D by

$$\phi(x) = \mathbb{E}\left(\int_0^T c(X_t) dt + f(X_T) 1_{\{T < \infty\}} \right) \tag{7.3}$$

where $(X_t)_{t \geq 0}$ is a Brownian motion in $\mathbb{R}^d$ starting from x, and T is its exit time from D, given by

$$T = \inf\{t \geq 0 : X_t \notin D\}.$$

We call any function $\psi \in C(\bar{D}) \cap C^2(D)$ satisfying

$$\begin{aligned} -\tfrac{1}{2}\Delta\psi &= c \quad \text{in } D, \\ \psi &= f \quad \text{in } \partial D \end{aligned}$$

a *solution of the Dirichlet problem* (*in D with data c and f*). More generally, a function $\psi \in C(\bar{D}) \cap C^2(D)$ is called a *supersolution of the Dirichlet problem* if

$$\begin{aligned} -\tfrac{1}{2}\Delta\psi &\geq c \quad \text{in } D, \\ \psi &\geq f \quad \text{in } \partial D. \end{aligned}$$

Theorem 7.9.1. *Suppose that D is bounded and has a C^1 boundary ∂D. Suppose further that c has a C^2 extension to $\mathbb{R}^d$ and that f is continuous on ∂D. Then the expected total cost function ϕ given by* (7.3) *is the unique solution of the Dirichlet problem in D with data c and f.*

It may be helpful to consider the following quick argument, while noting that it contains a number of gaps. Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from x and let T be its exit time from D. Suppose we knew that T was finite and that there was a function $\psi \in C_b^2(\mathbb{R}^d)$ whose restriction to $\bar{D}$ was a solution to the Dirichlet problem. Then we could set

$$M_t = \psi(X_t) - \psi(X_0) - \int_0^t \tfrac{1}{2}\Delta\psi(X_s)ds$$

and $(M_t)_{t\geq 0}$ would be a martingale. Suppose we could justify applying optional stopping to $(M_t)_{t\geq 0}$ at time T. Then we would have $\mathbb{E}(M_T) = \mathbb{E}(M_0)$ so

$$\begin{aligned} \psi(x) = \psi(X_0) &= \mathbb{E}\left(\psi(X_T) - \int_0^T \tfrac{1}{2}\Delta\psi(X_t)dt\right) \\ &= \mathbb{E}\left(f(X_T) + \int_0^T c(X_t)dt\right) = \phi(x). \end{aligned}$$

For a rigorous argument, we will have to work around the obvious gaps.

The proof of Theorem 7.9.1 is given in a series of steps below. These in fact lead to the following stronger result.

Theorem 7.9.2. *Let ϕ be the expected total cost function given for $x \in \bar{D}$ by* (7.3)*. Thus*

$$\phi(x) = \mathbb{E}\left(\int_0^T c(X_t)dt + f(X_T)1_{\{T<\infty\}}\right)$$

where $(X_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^d$ starting from x, and T is its exit time from D.

(a) *For any non-negative supersolution ψ of the Dirichlet problem in D with data c and f, we have $\phi \leq \psi$.*

(b) *For any bounded solution ψ of the Dirichlet problem in D with data c and f, such that*

$$\mathbb{E}\left(\psi(X_t)1_{\{t<T\}}\right) \to 0 \quad \text{as } t \to \infty \tag{7.4}$$

for all starting points $x \in D$, we have $\phi = \psi$.

(c) *Assume that c extends to $\mathbb{R}^d$ as a C^2 function and f is continuous on ∂D. Assume further that D satisfies the exterior cone condition* (7.9) *and that ϕ is locally bounded. Then ϕ is a solution of the Dirichlet problem in D with data c and f.*

Since we impose that ψ is bounded, condition (7.4) holds whenever T is almost surely finite. So, by Theorem 7.8.1, condition (7.4) holds in each of the following cases

(a) $d = 1$ and ∂D is non-empty,
(b) $d = 2$ and $\mathbb{R}^2 \setminus D$ contains an open ball,
(c) $d \geq 3$ and D is bounded.

An examination of the proof of Theorem 7.9.2(b) shows that ϕ always satisfies (7.4) if it is a bounded solution of the Dirichlet problem.

Under the hypotheses of Theorem 7.9.1, $\mathbb{E}(T)$ is bounded, uniformly in the starting point x, by Proposition 7.7.1. Since D is bounded, so are the functions c, f and ϕ, and so is any solution ψ of the Dirichlet problem. Since ∂D is C^1, D satisfies the exterior cone condition. So ϕ is a solution of the Dirichlet problem by Theorem 7.9.2(c). Moreover, (7.4) holds for ψ, so $\phi = \psi$ by Theorem 7.9.2(b), and so ϕ is the unique solution. Hence Theorem 7.9.1 follows from Theorem 7.9.2.

Proof of Theorem 7.9.2(a). Let ψ be a supersolution of the Dirichlet problem. It is clear that $\phi \leq \psi$ on ∂D. Fix $x \in D$ and let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from x. Fix $N \geq 1$ and set

$$D_N = \{x \in D : |x| < N \text{ and } |x - \partial D| > 1/N\}.$$

There exists $g \in C_b^2(\mathbb{R}^d)$ with $g = \psi$ on D_N. Set

$$M_t = g(X_t) - g(X_0) - \int_0^t \tfrac{1}{2}\Delta g(X_s)ds.$$

Then $(M_t)_{t\geq 0}$ is a martingale, by Theorem 7.5.1. Denote by T_N the exit time from D_N. Then, by optional stopping, for N sufficiently large and for all $t \geq 0$,

$$\psi(x) = \mathbb{E}\left(\psi(X_{T_N\wedge t})\right) + \mathbb{E}\int_0^{T_N\wedge t}(-\tfrac{1}{2}\Delta)\psi(X_s)ds. \tag{7.5}$$

We now let $t \to \infty$ and $N \to \infty$. By monotone convergence,

$$\mathbb{E}\int_0^{T_N\wedge t}(-\tfrac{1}{2}\Delta)\psi(X_t)dt \geq \mathbb{E}\int_0^{T_N\wedge t} c(X_t)dt \to \mathbb{E}\int_0^T c(X_t)dt.$$

On the other hand, on the event $\{T < \infty\}$, we have

$$\psi(X_{T_N\wedge t}) \to \psi(X_T) \geq f(X_T).$$

Since $\psi \geq 0$, this implies that

$$\liminf \psi(X_{T_N\wedge t}) \geq f(X_T)1_{T<\infty}$$

so, by Fatou's lemma,

$$\liminf \mathbb{E}\left(\psi(X_{T_N\wedge t})\right) \geq \mathbb{E}(f(X_T)1_{T<\infty}).$$

Hence, on taking the lim inf in (7.5), we obtain $\psi(x) \geq \phi(x)$. $\square$

Next we show that, under suitable conditions, we can replace inequalities by equalities in the preceding argument.

Proof of Theorem 7.9.2(b). Let ψ be a bounded solution of the Dirichlet problem. It is clear that $\phi = \psi$ on ∂D. Fix $x \in D$. Let $(X_t)_{t \geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from x, and let T be its exit time from D. Consider the limit $N \to \infty$ and then $t \to \infty$ in (7.5). By monotone convergence,

$$\mathbb{E} \int_0^{T_N \wedge t} (-\tfrac{1}{2}\Delta)\psi(X_t)dt = \mathbb{E} \int_0^{T_N \wedge t} c(X_t)dt \to \mathbb{E} \int_0^{T} c(X_t)dt.$$

On the other hand

$$\mathbb{E}\left(\psi(X_{T_N \wedge t})\right) = \mathbb{E}\left(\psi(X_{T_N})1_{\{T_N \leq t\}}\right) + \mathbb{E}\left(\psi(X_t)1_{\{t < T_N\}}\right).$$

For the first term on the right, almost surely,

$$\psi(X_{T_N})1_{\{T_N \leq t\}} \to \psi(X_T)1_{\{T<\infty\}} = f(X_T)1_{\{T<\infty\}}$$

so, by bounded convergence,

$$\mathbb{E}\left(\psi(X_{T_N})1_{\{T_N \leq t\}}\right) \to \mathbb{E}\left(f(X_T)1_{\{T<\infty\}}\right)$$

while, for the second term on the right, if ψ satisfies (7.4) then, as $t \to \infty$,

$$\lim_{N\to\infty} \mathbb{E}\left(\psi(X_t)1_{\{t<T_N\}}\right) = \mathbb{E}\left(\psi(X_t)1_{\{t<T\}}\right) \to 0.$$

Now, take the limit $N \to \infty$ and then $t \to \infty$ in (7.5) to obtain $\psi(x) = \phi(x)$. $\square$

It remains to find conditions under which we can show that ϕ is a solution of the Dirichlet problem.

Proof of Theorem 7.9.2(c). Step I. We restrict for now to the case where $d \geq 3$ and $D = \mathbb{R}^d$, and where c has compact support. Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0. Let g be a continuous function on $\mathbb{R}^d$ of compact support. Then

$$\mathbb{E} \int_0^\infty g(x + X_t)dt = \int_0^\infty P_t g(x)dt.$$

From the explicit formula (7.1) for the heat semigroup, we obtain the following estimates

$$\|P_t g\|_\infty \leq \|g\|_\infty, \quad \|P_t g\|_\infty \leq (2\pi t)^{-d/2} \operatorname{vol}(\operatorname{supp} g)\|g\|_\infty \tag{7.6}$$

and so, by splitting the integral at $t = 1$,

$$\mathbb{E} \int_0^\infty |g(x + X_t)|dt \leq \int_0^\infty \|P_t g\|_\infty dt \leq (1 + \operatorname{vol}(\operatorname{supp} g))\|g\|_\infty.$$

Fix $\varepsilon > 0$ and set $g_\varepsilon(x) = \sup_{|y-x|\leq\varepsilon} |g(y)|$. Then g_ε is also continuous and of compact support, so we see that

$$\mathbb{E} \int_0^\infty \sup_{|y-x|\leq\varepsilon} |g(y + X_t)|dt < \infty.$$

We use this estimate, applied to the first and second derivatives of c, to justify differentiating the formula

$$\phi(x) = \mathbb{E} \int_0^\infty c(x + X_t)dt$$

twice under the integral, to see that $\phi \in C^2(\mathbb{R}^d)$ with

$$\Delta\phi(x) = \mathbb{E}\int_0^\infty \Delta c(x + X_t)dt.$$

Take $s, t \in (0, \infty)$ with $s < t$ and split the integral into three pieces

$$\Delta\phi(x) = \left(\int_0^s + \int_s^t + \int_t^\infty\right) \mathbb{E}(\Delta c(x + X_u))du.$$

Consider the limit where $s \to 0$ and $t \to \infty$. Using the estimates (7.6), we see that the first and third integrals on the right tend to 0. On the other hand, for the second integral, we have

$$\begin{aligned}
\frac{1}{2}\int_s^t \mathbb{E}(\Delta c(x + X_u))du &= \frac{1}{2}\int_s^t \int_{\mathbb{R}^d} p(u, x, y)\Delta c(y)dydu \\
&= \frac{1}{2}\int_s^t \int_{\mathbb{R}^d} \Delta p(u, x, y)c(y)dydu \\
&= \int_s^t \int_{\mathbb{R}^d} \dot{p}(u, x, y)c(y)dydu \\
&= \int_{\mathbb{R}^d} p(t, x, y)c(y)dy - \int_{\mathbb{R}^d} p(s, x, y)c(y)dy \\
&= P_t c(x) - \mathbb{E}(c(x + X_s)) \to -c(x).
\end{aligned}$$

Hence $\frac{1}{2}\Delta\phi = -c$, showing that ϕ is a solution of the Dirichlet problem. □

We will use several times the following identity for the expected total cost function, which is a consequence of the strong Markov property.

Lemma 7.9.3. *Let D_0 be a bounded open subset of D and let $x \in \bar{D}$. Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from x, and write T_0 for its exit time from D_0. Then T_0 is almost surely finite and the expected total cost function ϕ satisfies*

$$\phi(x) = \mathbb{E}\left(\int_0^{T_0} c(X_t)dt + \phi(X_{T_0})\right).$$

Proof. Set $\tilde{\mathcal{F}}_t = \mathcal{F}_{T_0+t}$ and $\tilde{X}_t = X_{T_0+t}$, and write $\tilde{T}$ for the exit time of $(\tilde{X}_t)_{t\geq 0}$ from D. Then $\tilde{T} < \infty$ if and only if $T < \infty$, and if both are finite, then $X_T = \tilde{X}_{\tilde{T}}$. By the strong Markov property, $(\tilde{X}_t)_{t\geq 0}$ is an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion, so

$$\begin{aligned}
\phi(x) &= \mathbb{E}\left(\int_0^{T_0} c(X_t)dt + \int_{T_0}^T c(X_t)dt + f(X_T)1_{\{T<\infty\}}\right) \\
&= \mathbb{E}\left(\int_0^{T_0} c(X_t)dt\right) + \mathbb{E}\left(\mathbb{E}\left(\int_0^{\tilde{T}} c(\tilde{X}_t)dt + f(\tilde{X}_{\tilde{T}})1_{\{\tilde{T}<\infty\}}\,\middle|\, \tilde{\mathcal{F}}_0\right)\right) \\
&= \mathbb{E}\left(\int_0^{T_0} c(X_t)dt + \phi(X_{T_0})\right). \qquad (7.7)
\end{aligned}$$

□

We will use the following characterization of harmonic functions in terms of averages. Denote by $\sigma_{x,\rho}$ the uniform distribution on the sphere $S(x,\rho)$ of radius ρ and centre x.

Lemma 7.9.4. *Let ϕ be a non-negative measurable function on D. Suppose that*

$$\phi(x) = \int_{S(x,\rho)} \phi(y)\sigma_{x,\rho}(dy) \tag{7.8}$$

whenever $\overline{B}(x,\rho) \subseteq D$. Then, either $\phi(x) = \infty$ for all $x \in D$, or $\phi \in C^\infty(D)$ with $\Delta\phi = 0$ in D.

Proof. By taking a suitable average of the equation (7.8) over the possible values of ρ, we can see that ϕ also satisfies the ball-average property

$$\phi(x) = \int_{B(x,\rho)} \phi(y)\beta_{x,\rho}(dy)$$

where $\beta_{x,\rho}$ is the uniform distribution on the ball $B(x,\rho)$. Hence, if $B(x,\rho) \subseteq B(y,\tau) \subseteq D$, then $\phi(x) \leq (\tau/\rho)^d\phi(y)$. Since D is connected, this implies that either $\phi(x) = \infty$ for all $x \in D$, or ϕ is locally bounded in D.

Given $\varepsilon > 0$, there exists a C^∞ probability density function f on $\mathbb{R}^d$, which is rotationally invariant and supported in $B(0,\varepsilon)$. Let Y be a random variable in $\mathbb{R}^d$ having density f. Then, for any $x \in D$ at distance at least ε from ∂D, by taking a suitable average of the equation (7.8), we obtain

$$\phi(x) = \mathbb{E}(\phi(x+Y)) = \int_{\mathbb{R}^d} \phi(x+y)f(y)dy = \int_{\mathbb{R}^d} \phi(z)f(z-x)dz.$$

In the case where ϕ is locally bounded, we can differentiate in the last integral to see that $\phi \in C^\infty(D)$.

Consider then the Taylor expansion

$$\phi(x+ty) = \phi(x) + t\phi'(x)y + t^2\phi''(x)y \otimes y/2 + O(t^3).$$

By rotational invariance

$$\int_{\mathbb{R}^d} yf(y)dy = 0, \quad \int_{\mathbb{R}^d} y^iy^jf(y)dy = \delta_{ij}\mathbb{E}(|Y|^2)/d$$

so, on putting $y = Y$ and taking the expectation, we obtain, for all $t \in (0,1]$,

$$\phi(x) = \mathbb{E}(\phi(x+tY)) = \phi(x) + t^2\Delta\phi(x)\mathbb{E}(|Y|^2)/(2d) + O(t^3)$$

from which it follows that $\Delta\phi(x) = 0$. $\square$

Proof of Theorem 7.9.2(c). Step II. We will show, in the case where $c = 0$, that, provided ϕ is finite-valued, we have $\phi \in C^\infty(D)$ and $\Delta\phi = 0$ in D. Fix $x \in D$ and take $D_0 = B(x,\rho)$ where $\rho > 0$ is chosen so that $\overline{B}(x,\rho) \subseteq D$. Let $(X_t)_{t\geq 0}$ and T_0 be as in Lemma 7.9.3. By rotational invariance, X_{T_0} has the uniform distribution $\sigma_{x,\rho}$ on $S(x,\rho)$. Hence

$$\phi(x) = \mathbb{E}(\phi(X_{T_0})) = \int_{S(x,\rho)} \phi(y)\sigma_{x,\rho}(dy).$$

Since ϕ is finite-valued, it follows by Lemma 7.9.4 that $\phi \in C^\infty(D)$ with $\Delta\phi = 0$ in D. $\square$

We now show, under suitable conditions, that ϕ extends continuously to the boundary. For this we will need to understand the behaviour of Brownian motion just after time 0.

Theorem 7.9.5 (Blumenthal's zero-one law). *Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0. Then*

$$\mathbb{P}(A) \in \{0,1\} \quad \textit{for all} \quad A \in \mathcal{F}^X_{0+} = \bigcap_{t>0} \mathcal{F}^X_t.$$

Proof. Set

$$\mathcal{A} = \bigcup_{s>0} \sigma(X_t - X_s : t \geq s).$$

Then $\mathcal{A}$ is a π-system and $\mathbb{P}(A_0 \cap A) = \mathbb{P}(A_0)\mathbb{P}(A)$ for all $A_0 \in \mathcal{F}^X_{0+}$ and all $A \in \mathcal{A}$. Hence this holds also for all A in the generated σ-algebra $\sigma(\mathcal{A})$. Now $X_t - X_s$ is $\sigma(\mathcal{A})$-measurable for all $s, t > 0$ with $s \leq t$. But $X_s \to 0$ as $s \to 0$, so X_t is $\sigma(\mathcal{A})$-measurable for all $t > 0$ and so $\sigma(\mathcal{A}) = \mathcal{F}^X_\infty$. Hence, if $A \in \mathcal{F}^X_{0+}$, then $A \in \sigma(\mathcal{A})$, so

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$$

and so $\mathbb{P}(A) \in \{0,1\}$. $\square$

Proposition 7.9.6. *Let A be a non-empty open subset of the unit sphere in $\mathbb{R}^d$ and let $\varepsilon > 0$. Consider the cone*

$$C = \{x \in \mathbb{R}^d : x = ty \textit{ for some } 0 < t < \varepsilon, y \in A\}.$$

Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0 and let

$$T_C = \inf\{t \geq 0 : X_t \in C\}.$$

Then $T_C = 0$ almost surely.

We say that D satisfies the *exterior cone condition* if, for all $y \in \partial D$, there exists $\varepsilon > 0$ and a non-empty open subset A of the unit sphere such that

$$\{y + tz : z \in A, t \in (0, \varepsilon)\} \cap D = \emptyset \tag{7.9}$$

Geometrically, this means that, for every point in $y \in \partial D$, there is an open cone in $\mathbb{R}^d \setminus D$ with apex at y. This condition is always satisfied if ∂D is C^1, that is to say if, for all $y \in \partial D$, there is a neighbourhood U of y in $\mathbb{R}^d$ and a C^1 map $F = (F_1, \dots, F_d) : U \to \mathbb{R}^d$ such that $F(y) = 0$, $F'(y)$ is invertible, and $D \cap U = \{x \in U : F_1(x) > 0\}$.

Proof of Theorem 7.9.2(c). Step III. Note that $\phi = f$ on ∂D. Fix $y \in \partial D$. We will show that, for $x \in \bar{D}$, we have $\phi(x) \to f(y)$ as $x \to y$. Choose $D_0 = U \cap D$, where U is a bounded open set in $\mathbb{R}^d$ containing y. Let $(X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0. Consider the stopping time

$$T_0(x) = \inf\{t \geq 0 : x + X_t \notin D_0\}.$$

Then, by Lemma 7.9.3,

$$\phi(x) = \mathbb{E}\left(\int_0^{T_0(x)} c(x + X_t)dt + \phi(x + X_{T_0(x)})\right). \tag{7.10}$$

There exists an open cone C in $\mathbb{R}^d$ of positive height such that $y + C$ is disjoint from D. By Proposition 7.9.6, $T_C = \inf\{t \geq 0 : y + X_t \in C\} = 0$ almost surely. Now, on the event $\{T_C = 0\}$, in the limit $x \to y$, we must have $T_0(x) \to 0$, so $x + X_{T_0(x)} \to y$ and $x + X_{T_0(x)} \in \partial D$ eventually. Since $\phi = f$ and f is continuous on ∂D, this then implies that $\phi(x + X_{T_0(x)}) \to f(y)$ as $x \to y$. Now $\mathbb{E}(\sup_{x \in D_0} T_0(x)) < \infty$ and c and ϕ are locally bounded, so we can use dominated convergence in (7.10) to see that $\phi(x) \to f(y)$ as $x \to y$. $\square$

Proof of Theorem 7.9.2(c). Step IV. In the case where $c(x) = 0$ for all $x \in D$, the proof is already complete, by Steps II and III. In the case where $f(x) = 0$ for all $x \in \partial D$, by Step III, it remains to show that $\phi \in C^2(D)$ and $-\frac{1}{2}\Delta\phi = c$ in D. By linearity, it will suffice to complete this second case. Moreover, it will suffice to treat the case where $d \geq 3$. For the cases $d = 1$ or $d = 2$ then follow by applying the result for $d = 3$ to cylindrical regions D and to functions c which depend only on the first and second coordinates. Assume, for now, that D is bounded. Let $(X_t)_{t \geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0. Set

$$\phi_0(x) = \mathbb{E}\int_0^\infty \tilde{c}(x + X_t)dt$$

where $\tilde{c} \in C^2(\mathbb{R}^d)$ is a compactly supported function agreeing with c on D. By Step I, we have $\phi_0 \in C_b^2(\mathbb{R}^d)$ with $-\frac{1}{2}\Delta\phi_0 = \tilde{c}$. On taking $\phi = \phi_0$ and $D = \mathbb{R}^d$ and $D_0 = D$ in Lemma 7.9.3, we find that $\phi_0(x) = \phi(x) + \phi_1(x)$ for all $x \in D$, where

$$\phi_1(x) = \mathbb{E}(\phi_0(x + X_{T(x)}))$$

and where $T(x)$ is the exit time of $(x + X_t)_{t \geq 0}$ from D. As we showed in Step II, this implies that $\phi_1 \in C^\infty(D)$ with $\Delta\phi_1 = 0$ in D, so $\phi \in C^2(D)$ with $-\frac{1}{2}\Delta\phi = c$ in D. Finally, if D is unbounded, then, by Lemma 7.9.3, in any bounded open set $D_0 \subseteq D$, we have $\phi = \phi_0 + \phi_1$, where

$$\phi_0(x) = \mathbb{E}\int_0^{T_0(x)} c(x + X_t)dt, \quad \phi_1(x) = \mathbb{E}(\phi(x + X_{T_0(x)}))$$

where $T_0(x)$ is the exit time of $(x + X_t)_{t \geq 0}$ from D_0. Then $\phi_0 \in C^2(D_0)$ with $-\frac{1}{2}\Delta\phi_0 = c$ in D_0 by the preceding argument. On the other hand, since ϕ is locally bounded, ϕ_1 is bounded so, by the argument of Step II, we have $\phi_1 \in C^2(D_0)$ with $\Delta\phi_1 = 0$ in D_0. Since D_0 is arbitrary, this shows that $\phi \in C^2(D)$ with $-\frac{1}{2}\Delta\phi = c$ in D. $\square$

7.10. **Skorohod embedding for random walks.**

Theorem 7.10.1 (Skorohod embedding for random walks)**.** *Let μ be a probability measure on $\mathbb{R}$ of mean 0 and variance $\sigma^2 < \infty$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t \geq 0}$, on which is defined a Brownian motion $(B_t)_{t \geq 0}$ and a sequence of stopping times $0 = T_0 \leq T_1 \leq T_2 \leq \ldots$ such that, setting $S_n = B_{T_n}$,*

(i) *$(T_n)_{n \geq 0}$ is a random walk with step mean σ^2,*
(ii) *$(S_n)_{n \geq 0}$ is a random walk with step distribution μ.*

Proof. Define Borel measures $\mu_\pm$ on $[0, \infty)$ by

$$\mu_+(A) = \mu(A \cap [0, \infty)), \quad \mu^-(A) = \hat{\mu}(A \cap (0, \infty))$$

where $\hat{\mu}(A) = \mu(\{x \in \mathbb{R} : -x \in A\})$. There exists a probability space on which are defined a Brownian motion $(B_t)_{t \geq 0}$ and a sequence $((X_n, Y_n) : n \in \mathbb{N})$ of independent random variables in $\mathbb{R}^2$ with law ν given by

$$\nu(dx, dy) = C(x + y)\mu_-(dx)\mu_+(dy)$$

where C is a suitable normalizing constant. Set $\mathcal{F}_0 = \sigma(X_n, Y_n : n \in \mathbb{N})$ and $\mathcal{F}_t = \sigma(\mathcal{F}_0, \mathcal{F}_t^B)$. Set $T_0 = 0$ and define recursively for $n \geq 0$

$$T_{n+1} = \inf\{t \geq T_n : B_t - B_{T_n} \in \{-X_{n+1}, Y_{n+1}\}\}.$$

Then T_n is a stopping time for all n. Note that, since μ has mean 0, we must have

$$C \int_{[0,\infty)} x\mu^-(dx) = C \int_{[0,\infty)} y\mu^+(dy) = 1.$$

Define a non-negative measurable function τ on $W \times [0,\infty)^2$ by

$$\tau(w, x, y) = \inf\{t \geq 0 : w(t) \in \{-x, y\}\}.$$

Then $T_1 = \tau(B, X_1, Y_1)$ so, by Proposition 7.7.1 and Fubini,

$$\begin{aligned}\mathbb{E}(T_1) &= \int_{[0,\infty)^2} \int_W \tau(w, x, y)\mu_0(dw)\nu(dx, dy) \\ &= C \int_{[0,\infty)^2} xy(x+y)\mu^-(dx)\mu^+(dy) = \int_{[0,\infty)} x^2\mu^-(dx) + \int_{[0,\infty)} y^2\mu^+(dy) = \sigma^2.\end{aligned}$$

and, for any Borel set $A \subseteq [0, \infty)$,

$$\begin{aligned}\mathbb{P}(B_{T_1} \in A) &= \int_{[0,\infty)^2} \int_W 1_{\{w(\tau(w,x,y)) \in A\}}\mu_0(dw)\nu(dx, dy) \\ &= C \int_{[0,\infty)^2} \frac{x}{x+y} 1_{\{y \in A\}}(x+y)\mu^-(dx)\mu^+(dy) = C \int_{[0,\infty)} x\mu^-(dx) \int_A \mu^+(dy) = \mu(A).\end{aligned}$$

Similarly, $\mathbb{P}(B_{T_1} \in A) = \mu(A)$ also for $A \subseteq (-\infty, 0)$, so B_{T_1} has distribution μ.

Now, by the strong Markov property, for each $n \geq 1$, the process $(B_{T_n+t} - B_{T_n})_{t \geq 0}$ is a Brownian motion, independent of $\mathcal{F}_{T_n}$. Hence $S_{n+1} - S_n = B_{T_{n+1}} - B_{T_n}$ has law μ, $T_{n+1} - T_n$ has the same distribution as T_1, and both these increments are independent of $(T_1, S_1), \ldots, (T_n, S_n)$. The result follows. □

7.11. **Donsker's invariance principle.** In this section we show that Brownian motion provides a universal scaling limit for random walks having steps of zero mean and finite variance. This can be considered as a generalization to processes of the central limit theorem. We give $C([0,\infty), \mathbb{R})$ the topology of uniform convergence on compact time intervals. The associated Borel σ-algebra then coincides with the σ-algebra generated by the coordinate functions.

Theorem 7.11.1 (Donsker's invariance principle)**.** *Let $(S_n)_{n \geq 0}$ be a random walk with steps of mean* 0 *and variance* 1*. Write $(S_t)_{t \geq 0}$ for the linear interpolation*

$$S_{n+t} = (1-t)S_n + tS_{n+1}, \quad t \in [0, 1].$$

Then the law of $(N^{-1/2}S_{Nt})_{t \geq 0}$ converges weakly to Wiener measure on $C([0,\infty), \mathbb{R})$.

Proof. Let $(B_t)_{t \geq 0}$ be a Brownian motion and let $(X_n, Y_n)_{n \geq 1}$ be a sequence of independent random variables, as in the proof of Theorem 7.10.1. Fix $N \geq 1$ and set $B_t^{(N)} = N^{1/2}B_{N^{-1}t}$. Then $(B_t^{(N)})_{t \geq 0}$ is also a Brownian motion. Define a sequence of stopping times $(T_n^{(N)})_{n \geq 0}$ as in Theorem 7.10.1, but using $(B_t^{(N)})_{t \geq 0}$ in place of $(B_t)_{t \geq 0}$. Set

$$S_n^{(N)} = B^{(N)}(T_n^{(N)})$$

and interpolate linearly to form $(S_t^{(N)})_{t\geq 0}$. Set

$$\tilde{T}_n^{(N)} = N^{-1}T_n^{(N)}, \quad \tilde{S}_t^{(N)} = N^{-1/2}S_{Nt}^{(N)}.$$

Then $(\tilde{S}_t^{(N)})_{t\geq 0}$ has the same law as $(N^{-1/2}S_{Nt})_{t\geq 0}$ on $C([0,\infty),\mathbb{R})$ and, for all $n \geq 0$,

$$\tilde{S}_{n/N}^{(N)} = B_{\tilde{T}_n^{(N)}}.$$

We will show that, for all $\tau < \infty$

$$\sup_{t\in[0,\tau]} |\tilde{S}_t^{(N)} - B_t| \to 0 \quad \text{in probability.}$$

Then, for any bounded continuous function F on $C([0,\infty),\mathbb{R})$, we have

$$F(\tilde{S}^{(N)}) \to F(B) \quad \text{in probability}$$

so, by bounded convergence,

$$\mathbb{E}(F((N^{-1/2}S_{Nt})_{t\geq 0})) = \mathbb{E}(F(\tilde{S}^{(N)})) \to \mathbb{E}(F(B))$$

as required.

By the strong law of large numbers $T_n^{(1)}/n \to 1$ almost surely as $n \to \infty$. So, as $N \to \infty$,

$$N^{-1} \sup_{n\leq N\tau} |T_n^{(1)} - n| \to 0 \quad \text{almost surely}$$

and hence

$$N^{-1} \sup_{n\leq N\tau} |T_n^{(N)} - n| \to 0 \quad \text{in probability}$$

Hence, for all $\delta > 0$,

$$\mathbb{P}\left(\sup_{n\leq N\tau} |\tilde{T}_n^{(N)} - n/N| > \delta\right) \to 0.$$

By the intermediate value theorem, for $n/N \leq t \leq (n+1)/N$ we have $\tilde{S}_t^{(N)} = B_u$ for some $\tilde{T}_n^{(N)} \leq u \leq \tilde{T}_{n+1}^{(N)}$. Hence

$$\{|\tilde{S}_t^{(N)} - B_t| > \varepsilon \text{ for some } t \in [0,\tau]\} \subseteq A_1 \cup A_2$$

where

$$A_1 = \{|\tilde{T}_n^{(N)} - n/N| > \delta \text{ for some } n \leq N\tau\}$$

and

$$A_2 = \{|B_u - B_t| > \varepsilon \text{ for some } t \in [0,\tau] \text{ and } |u - t| \leq \delta + 1/N\}.$$

The paths of $(B_t)_{t\geq 0}$ are uniformly continuous on $[0,\tau]$. So given $\varepsilon > 0$ we can find $\delta > 0$ so that $\mathbb{P}(A_2) \leq \varepsilon/2$ whenever $N \geq 1/\delta$. Then, by choosing N even larger if necessary, we can ensure also that $\mathbb{P}(A_1) \leq \varepsilon/2$. Hence $\tilde{S}^{(N)} \to B$, uniformly on $[0,\tau]$ in probability, as required. □

We did not use the central limit theorem in this proof, so we have the following corollary

Corollary 7.11.2 (Central limit theorem)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed random variables, of mean 0 and variance 1. Set $S_n = X_1 + \cdots + X_n$. Then $S_n/\sqrt{n}$ converges weakly to the Gaussian distribution of mean 0 and variance 1.*

Proof. Let f be a continuous bounded function on $\mathbb{R}$ and define $x_1 : C([0,\infty),\mathbb{R}) \to \mathbb{R}$ by $x_1(w) = w_1$. Set $F = f \circ x_1$. Then F is a continuous bounded function on $C([0,\infty),\mathbb{R})$. So

$$\mathbb{E}(f(S_n/\sqrt{n})) = \mathbb{E}(F(S^{(n)})) \to \mathbb{E}(F(B)) = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} e^{-|x|/2} dx.$$

□

8. Poisson random measures

8.1. **Construction and basic properties.** For $\lambda \in (0,\infty)$ we say that a random variable X in $\mathbb{Z}^+ \cup \{\infty\}$ is *Poisson of parameter* λ and write $X \sim P(\lambda)$ if

$$\mathbb{P}(X = n) = e^{-\lambda}\lambda^n/n!$$

We also write $X \sim P(0)$ to mean $X \equiv 0$ and write $X \sim P(\infty)$ to mean $X \equiv \infty$.

Proposition 8.1.1 (Addition property). *Let $(N_k : k \in \mathbb{N})$ be a sequence of independent random variables, with $N_k \sim P(\lambda_k)$ for all k. Then*

$$\sum_k N_k \sim P\left(\sum_k \lambda_k\right).$$

Proposition 8.1.2 (Splitting property). *Let $N \sim P(\lambda)$ and let $(Y_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed random variables in $\mathbb{N}$, independent of N. Set*

$$N_k = \sum_{n=1}^{N} 1_{\{Y_n = k\}}.$$

Then $(N_k : k \in \mathbb{N})$ is a sequence of independent random variables, with $N_k \sim P(\lambda p_k)$ for all k, where $p_k = \mathbb{P}(Y_1 = k)$.

Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space. A *Poisson random measure with intensity* μ is a map

$$M : \Omega \times \mathcal{E} \to \mathbb{Z}^+ \cup \{\infty\}$$

satisfying, for all sequences $(A_k : k \in \mathbb{N})$ of disjoint sets in $\mathcal{E}$,

(i) $M(\cup_k A_k) = \sum_k M(A_k)$,
(ii) $(M(A_k) : k \in \mathbb{N})$ is a sequence of independent random variables,
(iii) $M(A_k) \sim P(\mu(A_k))$ for all k.

Denote by E^* the set of $\mathbb{Z}^+ \cup \{\infty\}$-valued measures on $\mathcal{E}$ and define, for $A \in \mathcal{E}$,

$$X : E^* \times \mathcal{E} \to \mathbb{Z}^+ \cup \{\infty\}, \quad X_A : E^* \to \mathbb{Z}^+ \cup \{\infty\}$$

by

$$X(m, A) = X_A(m) = m(A).$$

Set $\mathcal{E}^* = \sigma(X_A : A \in \mathcal{E})$.

Theorem 8.1.3. *There exists a unique probability measure μ^* on $(E^*, \mathcal{E}^*)$ such that X is a Poisson random measure with intensity μ.*

Proof. (*Uniqueness.*) Consider the subset $\mathcal{A}$ of $\mathcal{E}^*$ consisting of sets of the form

$$A^* = \{m \in E^* : m(A_1) = n_1, \dots, m(A_k) = n_k\}$$

where $k \in \mathbb{N}$, $A_1, \dots, A_k \in \mathcal{E}$ and $n_1, \dots, n_k \in \mathbb{Z}^+$. Note that each such set A^* is a finite union of elements of $\mathcal{A}$ such that the sets $A_1, \dots, A_k$ are disjoint. Also, in this disjoint case, if μ^* makes X into a Poisson random measure with intensity μ, then

$$\mu^*(A^*) = \prod_{j=1}^{k} e^{-\mu(A_j)} \mu(A_j)^{n_j}/n_j!$$

This condition thus determines the values of μ^* on $\mathcal{A}$ and, since $\mathcal{A}$ is a π-system generating $\mathcal{E}^*$, this implies that μ^* is uniquely determined on $\mathcal{E}^*$.

(*Existence.*) Consider first the case where $\lambda = \mu(E) < \infty$. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which are defined a random variable $N \sim P(\lambda)$ and a sequence of independent random variables $(Y_n : n \in \mathbb{N})$, independent of N and all having distribution μ/λ. Set

(8.1) $$M(A) = \sum_{n=1}^{N} 1_{\{Y_n \in A\}}, \quad A \in \mathcal{E}.$$

It is easy to check, using the Poisson splitting property, that M is a Poisson random measure with intensity μ.

More generally, if $(E, \mathcal{E}, \mu)$ is σ-finite, then $E = \cup_k E_k$ for some sequence $(E_k : k \in \mathbb{N})$ of disjoint sets in $\mathcal{E}$ such that $\mu(E_k) < \infty$ for all k. We can construct, on some probability space, a sequence $(M_k : k \in \mathbb{N})$ of independent Poisson random measures, such that M_k has intensity $1_{E_k}\mu$ for all k. Set

$$M(A) = \sum_{k \in \mathbb{N}} M_k(A), \quad A \in \mathcal{E}.$$

It is easy to check, using the Poisson addition property, that M is a Poisson random measure with intensity μ. The law μ^* of M on E^* is then a measure with the required properties. $\square$

8.2. Integrals with respect to a Poisson random measure.

Theorem 8.2.1. *Let M be a Poisson random measure on E with intensity μ. Assume that $\mu(E) < \infty$. Let g be a measurable function on E. Define*

$$M(g) = \begin{cases} \int_E g(y) M(dy), & \text{if } M(E) < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Then $M(g)$ is a well-defined random variable and

$$\mathbb{E}(e^{iuM(g)}) = \exp\left\{ \int_E (e^{iug(y)} - 1) \mu(dy) \right\}.$$

Moreover, if $g \in L^1(\mu)$, then $M(g) \in L^1(\mathbb{P})$ and

$$\mathbb{E}(M(g)) = \int_E g(y) \mu(dy), \quad \operatorname{var}(M(g)) = \int_E g(y)^2 \mu(dy).$$

Proof. Set $E_0^* = \{m \in E^* : m(E) < \infty\}$ and note that $M \in E_0^*$ almost surely. For any $m \in E_0^*$, we have $m(|g| > n) = 0$ for sufficiently large $n \in \mathbb{N}$, so $g \in L^1(m)$. Moreover the map $m \mapsto m(g) : E_0^* \to \mathbb{R}$ is measurable. To see this, we note that in the case $g = 1_A$ for $A \in \mathcal{E}$, this is by definition of $\mathcal{E}^*$. This extends to g simple by linearity, then to g non-negative by monotone convergence, then to all g by linearity again.

Hence $M(g)$ is well defined random variable and

$$\mathbb{E}(e^{iuM(g)}) = \int_{E_0^*} e^{ium(g)} \mu^*(dm).$$

It will suffice then to prove the claimed formulas in the case where M is given as in (8.1). Then

$$\mathbb{E}(e^{iuM(g)}|N = n) = \mathbb{E}(e^{iug(Y_1)})^n = \left(\int_E e^{iug(y)} \mu(dy) \right)^n \lambda^{-n}$$

so

$$\begin{aligned} \mathbb{E}(e^{iuM(g)}) &= \sum_{n=0}^{\infty} \mathbb{E}(e^{iuM(g)}|N = n)\mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \left(\int_E e^{iug(y)} \mu(dy) \right)^n e^{-\lambda}/n! = \exp\left\{ \int_E (e^{iug(y)} - 1)\mu(dy) \right\}. \end{aligned}$$

If $g \in L^1(\mu)$ is integrable, then formulae for $\mathbb{E}(M(g))$ and $\mathrm{var}(M(g))$ may be obtained by a similar argument. □

We now fix a σ-finite measure space $(E, \mathcal{E}, K)$ and denote by μ the product measure on $(0, \infty) \times E$ determined by

$$\mu((0, t] \times A) = tK(A), \quad t \geq 0, A \in \mathcal{E}.$$

Let M be a Poisson random measure with intensity μ and set $\tilde{M} = M - \mu$. We call $\tilde{M}$ a *compensated Poisson random measure with intensity* μ. We use the filtration $(\mathcal{F}_t)_{t \geq 0}$ given by $\mathcal{F}_t = \sigma(\mathcal{F}_t^M, \mathcal{N})$, where

$$\mathcal{F}_t^M = \sigma(M((0, s] \times A) : s \leq t, A \in \mathcal{E}), \quad \mathcal{N} = \{B \in \mathcal{F}_\infty^M : \mathbb{P}(B) = 0\}.$$

Proposition 8.2.2. *Assume that $K(E) < \infty$. Let $g \in L^1(K)$. Set*

$$\tilde{M}_t(g) = \begin{cases} \int_{(0,t] \times E} g(y)\tilde{M}(ds, dy), & \text{if } M((0, t] \times E) < \infty \text{ for all } t \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then $(\tilde{M}_t(g))_{t \geq 0}$ is a cadlag martingale with stationary independent increments. Moreover

$$\mathbb{E}(\tilde{M}_t^2(g)) = t \int_E g(y)^2 K(dy) \tag{8.2}$$

and

$$\mathbb{E}(e^{iu\tilde{M}_t(g)}) = \exp\left\{ t \int_E (e^{iug(y)} - 1 - iug(y))K(dy) \right\}. \tag{8.3}$$

Theorem 8.2.3. *Let $g \in L^2(K)$. Let $(E_n : n \in \mathbb{N})$ be a sequence in $\mathcal{E}$ with $E_n \uparrow E$ and $K(E_n) < \infty$ for all n. Then the restriction $\tilde{M}^n$ of $\tilde{M}$ to $(0, \infty) \times E_n$ is a compensated*

Poisson random measure with intensity $1_{E_n}\mu$*. Set* $X_t^n = \tilde{M}_t^n(g)$*. Then there exists a cadlag martingale* $(X_t)_{t\geq 0}$ *such that, for all* $t \geq 0$*,*

$$\mathbb{E}\left(\sup_{s\leq t}|X_s^n - X_s|^2\right) \to 0.$$

Set $\tilde{M}_t(g) = X_t$*. Then* $(\tilde{M}_t(g))_{t\geq 0}$ *has stationary independent increments and* (8.2) *and* (8.3) *remain valid.*

The process $(\tilde{M}_t(g))_{t\geq 0}$ is (*a version of*) *the stochastic integral of* g *with respect to* $\tilde{M}$. We write

$$(\tilde{M}_t(g))_{t\geq 0} = \int_{(0,t]\times E} g(y)\tilde{M}(ds, dy) \quad \text{almost surely.}$$

Note that there is in general no preferred version and this 'integral' does not converge absolutely.

Proof. Set $g_n = 1_{E_n}g$. Fix $t > 0$. By Doob's L^2-inequality and Proposition 8.2.2,

$$\mathbb{E}\left(\sup_{s\leq t}|X_s^n - X_s^m|^2\right) \leq 4\mathbb{E}((X_t^n - X_t^m)^2) = 4t\int_E (g_n - g_m)^2 dK \to 0$$

as $n, m \to \infty$. Then there is a subsequence (n_k) such that, almost surely as $j, k \to \infty$, for all $t \geq 0$,

$$\sup_{s\leq t}|X_s^{n_k} - X_s^{n_j}| \to 0.$$

The uniform limit of cadlag functions is cadlag, so there is a cadlag process $(X_t)_{t\geq 0}$ such that, almost surely as $k \to \infty$, for all $t \geq 0$,

$$\sup_{s\leq t}|X_s^{n_k} - X_s| \to 0.$$

Then, by Fatou's lemma, as $n \to \infty$,

$$\mathbb{E}\left(\sup_{s\leq t}|X_s^n - X_s|^2\right) \leq 4t\int_E (g_n - g)^2 dK \to 0.$$

In particular $X_t^n \to X_t$ in L^2 for all t, from which it is easy to deduce (8.2) and that $(X_t)_{t\geq 0}$ inherits the martingale property. Moreover, using the inequality

$$|e^{iug} - 1 - iug| \leq u^2 g^2/2,$$

for $s, t \geq 0$ with $s < t$ and $A \in \mathcal{F}_s$, we can pass to the limit in the identity

$$\mathbb{E}(e^{iu(X_t^n - X_s^n)}1_A) = \exp\left\{(t-s)\int_{E_n}(e^{iug(y)} - 1 - iug(y))K(dy)\right\}\mathbb{P}(A)$$

to see that $(X_t)_{t\geq 0}$ has stationary independent increments and (8.3) holds. □

9. Lévy processes

9.1. **Definition and examples.** A *Lévy process* is a cadlag process starting from 0 with stationary independent increments. We call (a, b, K) a *Lévy triple* if $a \in [0, \infty)$, $b \in \mathbb{R}$ and K is a Borel measure on $\mathbb{R}$ with $K(\{0\}) = 0$ and

$$\int_{\mathbb{R}} (1 \wedge |y|^2) K(dy) < \infty.$$

We call a the *diffusivity*, b the *drift* and K the *Lévy measure*. These notions generalize naturally to processes with values in $\mathbb{R}^d$ but we will consider only the case $d = 1$. Let B be a Brownian motion and let M be a Poisson random measure, independent of B, with intensity μ on $(0, \infty) \times \mathbb{R}$, where $\mu(dt, dy) = dtK(dy)$, as in the preceding section. Set

$$X_t = \sqrt{a} B_t + bt + \int_{(0,t] \times \{|y| \leq 1\}} y \tilde{M}(ds, dy) + \int_{(0,t] \times \{|y| > 1\}} y M(ds, dy).$$

We interpret the last integral as 0 on the null set $\{M((0, t] \times \{|y| > 1\}) = \infty$ for some $t \geq 0\}$. Then $(X_t)_{t \geq 0}$ is a Lévy process and, for all $t \geq 0$,

$$\mathbb{E}(e^{iuX_t}) = e^{t\psi(u)}$$

where

$$\psi(u) = \psi_{a,b,K}(u) = ibu - \tfrac{1}{2} a u^2 + \int_{\mathbb{R}} (e^{iuy} - 1 - iuy 1_{|y| \leq 1}) K(dy).$$

Thus, to every Lévy triple there corresponds a Lévy process. Moreover, given $(X_t)_{t \geq 0}$, we can recover M by

$$M((0, t] \times A) = \#\{s \leq t : X_s - X_{s-} \in A\}$$

and so we can also recover b and $\sqrt{a} B$. Hence the law of the Lévy process $(X_t)_{t \geq 0}$ determines the Lévy triple (a, b, K).

9.2. Lévy–Khinchin theorem.

Theorem 9.2.1 (Lévy–Khinchin theorem). *Let X be a Lévy process. Then there exists a unique Lévy triple (a, b, K) such that, for all $t \geq 0$ and all $u \in \mathbb{R}$,*

$$\mathbb{E}(e^{iuX_t}) = e^{t\psi_{a,b,K}(u)}.$$

Proof. For $t \geq 0$ and $u \in \mathbb{R}$, set $\phi_t(u) = \mathbb{E}(e^{iuX_t})$. Then $\phi_t : \mathbb{R} \to \mathbb{C}$ is continuous. Since $(X_t)_{t \geq 0}$ has stationary independent increments and

$$X_{nt} = X_t + (X_{2t} - X_t) + \cdots + (X_{nt} - X_{(n-1)t})$$

we obtain, on taking characteristic functions, for all $n \in \mathbb{N}$,

$$\phi_{nt}(u) = (\phi_t(u))^n.$$

Since $(X_t)_{t \geq 0}$ is cadlag, as $t \to s$ with $t > s$, we have $X_t \to X_s$, so

$$|\phi_t(u) - \phi_s(u)| \leq \mathbb{E}|e^{iu(X_t - X_s)} - 1| \leq \mathbb{E}((u|X_t - X_s|) \wedge 2) \to 0$$

uniformly on compacts in u. In particular, $\phi_t(u) \to 1$ as $t \to 0$, so

$$|\phi_t(u)|^{1/n} = |\phi_{t/n}(u)| \to 1 \quad \text{as } n \to \infty$$

which implies that $\phi_t(u) \neq 0$ for all $t \geq 0$ and all $u \in \mathbb{R}$. Set

$$\psi_t(u) = \int_1^{\phi_t(u)} \frac{dz}{z}$$

where we integrate along a contour homotopic to $(\phi_t(r) : r \in [0, u])$ in $\mathbb{C} \setminus \{0\}$. Then $\psi_t : \mathbb{R} \to \mathbb{C}$ is the unique continuous function such that $\psi_t(0) = 0$ and, for all $u \in \mathbb{R}$,

$$\phi_t(u) = e^{\psi_t(u)}.$$

Moreover, we then have, for all $n \in \mathbb{N}$,

$$\psi_{nt}(u) = n\psi_t(u)$$

and

$$\psi_t(u) \to \psi_s(u) \quad \text{as } t \to s \text{ with } t > s.$$

Hence, by a standard argument, for all $t \geq 0$,

$$\phi_t(u) = e^{t\psi(u)}$$

where $\psi = \psi_1$, and it remains to show that $\psi = \psi_{a,b,K}$ for some Lévy triple (a, b, K).

Write ν_n for the law of $X_{1/n}$. Then, uniformly on compacts in u, as $n \to \infty$,

$$\int_{\mathbb{R}} (e^{iuy} - 1) n\nu_n(dy) = n(\phi_{1/n}(u) - 1) \to \psi(u)$$

so

$$\int_{\mathbb{R}} (1 - \cos uy) n\nu_n(dy) \to -\operatorname{Re}\psi(u).$$

There is a constant $C < \infty$ such that, for all $y \in \mathbb{R}$

$$y^2 1_{\{|y| \leq 1\}} \leq C(1 - \cos y)$$

and, for all $\lambda \in (0, \infty)$,

$$1_{\{|y| \geq \lambda\}} \leq C\lambda \int_0^{1/\lambda} (1 - \cos uy) du.$$

Consider the measure η_n on $\mathbb{R}$, given by

$$\eta_n(dy) = n(1 \wedge |y|^2)\nu_n(dy).$$

Then, as $n \to \infty$,

$$\begin{aligned}
\eta_n([-1, 1]) &= \int_{\mathbb{R}} y^2 1_{\{|y| \leq 1\}} n\nu_n(dy) \\
&\leq C \int_{\mathbb{R}} (1 - \cos y) n\nu_n(dy) \to -C \operatorname{Re}\psi(1)
\end{aligned}$$

and, for $\lambda \geq 1$,

$$\begin{aligned}
\eta_n(\mathbb{R} \setminus (-\lambda, \lambda)) &= \int_{\mathbb{R}} 1_{\{|y| \geq \lambda\}} n\nu_n(dy) \\
&\leq C\lambda \int_0^{1/\lambda} \int_{\mathbb{R}} (1 - \cos uy) n\nu_n(dy) du \\
&\to -C\lambda \int_0^{1/\lambda} \operatorname{Re}\psi(u) du.
\end{aligned}$$

Note that, since $\psi(0) = 0$, the final limit can be made arbitrarily small by choosing λ sufficiently large. Hence the sequence $(\eta_n : n \in \mathbb{N})$ is bounded in total mass and tight. By Prohorov's theorem, there is a subsequence (n_k) and a finite measure η on $\mathbb{R}$ such that $\eta_{n_k} \to \eta$ weakly on $\mathbb{R}$. Fix a continuous function χ on $\mathbb{R}$ with

$$1_{\{|y|\le 1\}} \le \chi(y) \le 1_{\{|y|\le 2\}}.$$

We have

$$\begin{aligned}\int_{\mathbb{R}} (e^{iuy} - 1) n\nu_n(dy) &= \int_{\mathbb{R}\setminus\{0\}} (e^{iuy} - 1)\frac{\eta_n(dy)}{1 \wedge y^2} \\ &= \int_{\mathbb{R}\setminus\{0\}} \frac{(e^{iuy} - 1 - iuy\chi(y))}{1 \wedge y^2}\eta_n(dy) + \int_{\mathbb{R}\setminus\{0\}} \frac{iuy\chi(y)}{1 \wedge y^2}\eta_n(dy) \\ &= \int_{\mathbb{R}} \theta(u, y)\eta_n(dy) + iub_n\end{aligned}$$

where

$$\theta(u, y) = \begin{cases} (e^{iuy} - 1 - iuy\chi(y))/(1 \wedge y^2), & \text{if } y \neq 0, \\ -u^2/2, & \text{if } y = 0. \end{cases}$$

and

$$b_n = \int_{\mathbb{R}} \frac{y\chi(y)}{1 \wedge y^2}\eta_n(dy).$$

Now $\theta(u, .)$ is a bounded continuous function for each $u \in \mathbb{R}$. So, on letting $k \to \infty$,

$$\int_{\mathbb{R}} \theta(u, y)\eta_{n_k}(dy) \to \int_{\mathbb{R}} \theta(u, y)\eta(dy) = \int_{\mathbb{R}} (e^{iuy} - 1 - iuy\chi(y))K(dy) - \tfrac{1}{2}au^2$$

where

$$K(dy) = (1 \wedge y^2)^{-1} 1_{\{y\neq 0\}}\eta(dy), \quad a = \eta(\{0\}).$$

Then b_{n_k} must also converge, to β say, so we obtain

$$\psi(u) = i\beta u - \tfrac{1}{2}au^2 + \int_{\mathbb{R}} (e^{iuy} - 1 - iuy\chi(y))K(dy) = \psi_{a,b,K}(u)$$

where

$$b = \beta - \int_{\mathbb{R}} y(\chi(y) - 1_{\{|y|\le 1\}})K(dy).$$

□