ADVANCED PROBABILITY

JAMES NORRIS

Contents

Date: November 19, 2016.

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.* We will usually 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 $k = 1, \dots, n$, 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 $\mu(f)\le\mu(g)$ whenever $f\le g$.

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*

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

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

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 $|f_n|\le g$ 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 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 $\mathbb{E}(X|G_n) = \mathbb{E}(X 1_{G_n})/\mathbb{P}(G_n)$ 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 $\mathbb{E}(X 1_A) = \mathbb{E}(Y 1_A)$ 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 $A \in \mathcal{G}$,

$$\mathbb{E}(X_n1_A) = \mathbb{E}(Y_n1_A)$$

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. □

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* $(\mathcal{F}^X_n)_{n\geq 0}$, given by

$$\mathcal{F}^X_n = \sigma(X_0, \ldots, X_n).$$

Then $\mathcal{F}^X_n$ models what we know about X by time n. We say that $(X_n)_{n\geq 0}$ is *adapted* if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. It is equivalent to require that $\mathcal{F}^X_n \subseteq \mathcal{F}_n$ 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 $\mathbb{E}(X_T) \leq \mathbb{E}(X_S)$.*

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. For a direct proof of the optional stopping theorem, you can write out the implication from (a) to (b) below in the case where $S \leq T$ and $A = \Omega$.

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 $S \le T$,*

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

Proof. For $S \ge 0$ and $T \le n$, we have

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

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

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

Hence, on multiplying (2.1) by 1_A and taking expectations, we obtain

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

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

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

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

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

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

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

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

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

and, on this event,

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

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

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

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

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

Hence, on taking expectations, we obtain

$$(b-a)\mathbb{E}(U_n[a,b]) \le \mathbb{E}\left((X_n - a)^-\right) \tag{2.2}$$

and the desired estimate follows by monotone convergence. □

2.4. **Doob's maximal inequalities.**

Define, for a random process X,

$$X_n^* = \sup_{k\le n}|X_k|, \quad X^* = \sup_{n\ge 0}|X_n|.$$

In the next two theorems, we see that the martingale (or submartingale) property allows us effectively to move the supremum outside the probability or expectation.

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

$$\lambda\mathbb{P}(X^* \ge \lambda) \le \sup_{n\ge 0}\mathbb{E}(|X_n|).$$

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

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

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

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

Hence

$$\lambda\mathbb{P}(X_n^* \ge \lambda) \le \mathbb{E}(X_n 1_{\{X_n^*\ge\lambda\}}) \le \mathbb{E}(X_n). \tag{2.3}$$

On letting $n \to \infty$, we have $X_n^* \uparrow X^*$, so $\mathbb{P}(X_n^* > \lambda) \to \mathbb{P}(X^* > \lambda)$. Hence, from (2.3) we obtain

$$\lambda\mathbb{P}(X^* > \lambda) \le \sup_{n\ge 0}\mathbb{E}(X_n).$$

Finally, for $\lambda > 0$, we apply this to $\lambda' \in [0, \lambda)$ and let $\lambda' \to \lambda$ for the desired inequality. □

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

$$\|X^*\|_p \le q \sup_{n\ge 0}\|X_n\|_p.$$

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, equation (2.3) and Hölder's inequality,

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

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

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. Recall that, for a sequence of real numbers $(x_n)_{n \geq 0}$, as $n \to \infty$, either x_n converges, or $|x_n| \to \infty$, or $\liminf x_n < \limsup x_n$. In the last 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. $\square$

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, we may obtain all $L^1(\mathcal{F}_\infty)$ random variables in this way.*

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

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

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, we may obtain all $L^p(\mathcal{F}_\infty)$ random variables in this way.*

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. □

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 (2.2), 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. □

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. In particular, we have $X_{T \wedge n} \to X_T$ almost surely. Since $\mathcal{F}_{T \wedge n} \subseteq \mathcal{F}_n$, by Theorem 2.2.3 and the tower property,

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

By Lemma 1.5.1, the random process $(X_{T\wedge n})_{n\geq 0}$ is then uniformly integrable. Hence $X_{T\wedge n} \to X_T$ 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. Now, the optional stopping theorem and Theorem 2.2.3 apply at the bounded stopping time $T\wedge n$ to show

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

and the claimed identities follow on letting $n \to \infty$. □

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 7.10.3

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 $\varepsilon > 0$.*

The main point of the next result is that, if a sum of independent random variables converges in L^2, then it also converges almost surely, without passing to a subsequence.

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

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

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. □

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 $A_1, \dots, A_m$ such that $\mathcal{F}_n = \sigma(A_1, \dots, A_m)$. 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. $\square$

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 7.4.4.

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 $x \in D$,

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

□

In Theorem 7.9.3 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 $C < \infty$,*

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

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

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

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

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

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 we may obtain all $L^1(\mathcal{F}_\infty)$ random variables in this way.*

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

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

Then the estimates from Theorems 4.3.1 and 4.3.3 apply to show 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. $\square$

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 we may obtain all $L^p(\mathcal{F}_\infty)$ random variables in this way.*

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*

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

(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. $\square$

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) $\limsup_n \mu_n(C) \le \mu(C)$ *for all closed sets* C,
(c) $\liminf_n \mu_n(G) \ge \mu(G)$ *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 S is said to be *tight* if, for all $\varepsilon > 0$, there exists a compact set K such that $\mu_n(S \setminus K) \le \varepsilon$ 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* S. *Then there exists a subsequence* (n_k) *and a probability measure* μ *on* S *such that* $\mu_{n_k} \to \mu$ *weakly on* S.

Proof for the case $S = \mathbb{R}$. 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 $F_n(-R) \le \varepsilon$ and $F_n(R) \ge 1 - \varepsilon$ 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. □

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*

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

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

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

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

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,

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

□

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 measre 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*

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

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

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

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

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

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

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. □

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 $(B_t)_{t\geq 0}$ be a continuous random process in $\mathbb{R}^d$. We say that $(B_t)_{t\geq 0}$ is a *Brownian motion in* $\mathbb{R}^d$ if, for all $s,t\geq 0$ with $s<t$,

(i) $B_t - B_s \sim N(0,(t-s)I)$,
(ii) $B_t - B_s$ is independent of $\sigma(B_u : u\leq s)$.

We recall that, for $x\in\mathbb{R}^d$ and $t>0$, we write $X\sim N(x,tI)$ to mean that X is a random variable in $\mathbb{R}^d$ having Gaussian distribution of mean x and covariance matrix tI. Thus, for any bounded measurable function f on $\mathbb{R}^d$,

$$\mathbb{E}(f(X)) = P_t f(x) = \int_{\mathbb{R}^d} p(t,x,y)f(y)dy$$

where $p(t,x,y) = (2\pi t)^{-d/2}e^{-|x-y|^2/(2t)}$. In standard usage, where a process is introduced as a Brownian motion without mentioning the state-space, it is often assumed that this is $\mathbb{R}$. Similarly, where a process is introduced as a Brownian motion without mentioning the initial state, it is often assumed that this is 0.

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 $\mathcal{W}_d = \sigma(X_t : t\geq 0)$. When $d=1$ we write simply W and $\mathcal{W}$. The measure μ identified in the next theorem is called *Wiener measure*.

Theorem 7.2.1 (Wiener's theorem)**.** *There exists a unique probability measure μ on $(W,\mathcal{W})$ such that $(X_t)_{t\geq 0}$ is a Brownian motion starting from 0.*

Proof. Conditions (i) and (ii) determine the finite dimensional distributions of any such measure μ, so there can be at most one. To prove existence it will suffice to construct a Brownian motion B on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. Then $B:\Omega\to W$ is measurable and $\mu = B^{-1}\circ\mathbb{P}$ has the required property.

For $n\geq 0$ denote 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. Then $\mathbb{D}$ is countable so, by a standard argument [PM, Section 2.4], there exists, a probability space $(\Omega,\mathcal{F},\mathbb{P})$, on which there is defined a family of independent $N(0,1)$ random variables $(Y_t : t\in\mathbb{D})$. For $t\in\mathbb{D}_0=\mathbb{Z}^+$, set $\beta_t = Y_1+\cdots+Y_t$. Define recursively, for $n\geq 0$ and $t\in\mathbb{D}_{n+1}\setminus\mathbb{D}_n$,

$$\beta_t = \tfrac{1}{2}(\beta_r+\beta_s)+Z_t$$

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

Suppose inductively for $n\geq 0$ that $(\beta_{t+2^{-n}}-\beta_t : t\in\mathbb{D}_n)$ is a sequence of independent $N(0,2^{-n})$ random variables. Consider the sequence $(\beta_{t+2^{-n-1}}-\beta_t : t\in\mathbb{D}_{n+1})$. Fix $t\in\mathbb{D}_{n+1}\setminus\mathbb{D}_n$ and note that

$$\beta_t-\beta_r = \tfrac{1}{2}(\beta_s-\beta_r)+Z_t,\quad \beta_s-\beta_t = \tfrac{1}{2}(\beta_s-\beta_r)-Z_t.$$

Now

$$\mathrm{var}\left(\tfrac{1}{2}(\beta_s-\beta_r)\right) = 2^{-n-2} = \mathrm{var}(Z_t)$$

so

$$\mathrm{var}(\beta_t - \beta_r) = \mathrm{var}(\beta_s - \beta_t) = 2^{-n-1}, \quad \mathrm{cov}(\beta_t - \beta_r, \beta_s - \beta_t) = 0.$$

On the other hand, we also have

$$\mathrm{cov}(\beta_t - \beta_r, \beta_v - \beta_u) = \mathrm{cov}(\beta_s - \beta_t, \beta_v - \beta_u) = 0$$

for any $u, v \in \mathbb{D}_{n+1}$ with $(u, v] \cap (r, s] = 0$. Hence $\beta_t - \beta_r$ and $\beta_s - \beta_t$ are independent $N(0, 2^{-n-1})$ random variables, which are independent also of $\beta_v - \beta_u$ for all such u, v. The induction proceeds.

We have shown that $(\beta_t)_{t\in\mathbb{D}}$ has independent increments and that $\beta_t - \beta_s$ has $N(0, t-s)$ distribution for all $s, t \in \mathbb{D}$ with $s < t$. Choose $p > 2$ and set $C_p = \mathbb{E}(|\beta_1|^p)$. Then $C_p < \infty$ and

$$\mathbb{E}(|\beta_t - \beta_s|^p) \leq C_p(t-s)^{p/2}$$

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

Let $s, t \geq 0$ with $s < t$ and let $A \in \sigma(B_u : u \leq s)$. There exist sequences $(s_n : n \in \mathbb{N})$ and $(t_n : n \in \mathbb{N})$ such that $s_n, t_n \in \mathbb{D}_n$ and $s \leq s_n < t_n$ for all n and $s_n \to s$, $t_n \to t$. Also, there exists $A_0 \in \sigma(\beta_u : u \leq s, u \in \mathbb{D})$ such that $1_A = 1_{A_0}$ almost surely. Then, for any continous bounded function f on $\mathbb{R}^d$,

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

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

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

Hence $(B_t)_{t\geq 0}$ is a Brownian motion. □

7.3. **Transformations of Brownian motion.** The first two statements concern the case $d = 1$.

Proposition 7.3.1. *Let $(B_t)_{t\geq 0}$ be a continuous random process starting from* 0*. The following are equivalent*

(a) *$(B_t)_{t\geq 0}$ is a Brownian motion,*
(b) *$(B_t)_{t\geq 0}$ is a zero-mean Gaussian process with $\mathbb{E}(B_sB_t) = s \wedge t$ for all $s, t \geq 0$.*

Proposition 7.3.2. *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from* 0*. Then, for $s \geq 0$ and $c > 0$, the following processes are also Brownian motions starting from* 0

(a) *$(-B_t : t \geq 0)$,*
(b) *$(B_{s+t} - B_s : t \geq 0)$,*
(c) *$(cB_{c^{-2}t} : t \geq 0)$,*
(d) *$(tB_{1/t} : t \geq 0)$*

where in (d) *the process is defined to take the value* 0 *when $t = 0$, and for all $t > 0$ if this is necessary to make it continuous.*

In the last proposition, (c) is called the *scaling property* and (d) is called the *time inversion property.*

Proposition 7.3.3. *Let $(B_t)_{t\geq 0} = (B_t^1, \ldots, B_t^d)_{t\geq 0}$ be a random process in $\mathbb{R}^d$, starting from 0. Let $x \in \mathbb{R}^d$. The following are equivalent*

(a) *$(B_t)_{t\geq 0}$ is a Brownian motion,*
(b) *$(x + B_t)_{t\geq 0}$ is a Brownian motion starting from x,*
(c) *$(B_t^1)_{t\geq 0}, \ldots, (B_t^d)_{t\geq 0}$ are independent Brownian motions in $\mathbb{R}$.*

The last proposition makes clear that, for all $x \in \mathbb{R}^d$, there exists a Brownian motion $(B_t)_{t\geq 0}$ starting from x, whose law on $(W_d, \mathcal{W}_d)$ is unique. We denote this law by μ_x and call it *Wiener measure starting from* x.

Proposition 7.3.4. *The map $(x, A) \mapsto \mu_x(A) : \mathbb{R}^d \times \mathcal{W}_d \to [0, 1]$ is a measurable probability kernel.*

Proposition 7.3.5. *Let $(B_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ and let U be an orthogonal $d \times d$ matrix. Then $(UB_t)_{t\geq 0}$ is also a Brownian motion in $\mathbb{R}^d$.*

7.4. **Martingales.** In this section, we assume that our probability space is equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$. Let $\Omega_0 \in \mathcal{F}_0$. Let $(B_t)_{t\geq 0}$ be a continuous adapted random process in $\mathbb{R}^d$ defined on Ω_0. We say that $(B_t)_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$*-Brownian motion* if, for all $s, t \geq 0$ with $s < t$ and all bounded measurable functions f on $\mathbb{R}^d$, we have

$$\mathbb{E}(f(B_t)|\mathcal{F}_s) = P_{t-s}f(B_s) \quad \text{almost surely.}$$

By a monotone class argument [PM, Theorem 2.1.2], it is equivalent to require this condition only for bounded continuous functions f. We default to the case $\Omega_0 = \Omega$ unless otherwise indicated.

Proposition 7.4.1. *Let $(B_t)_{t\geq 0}$ be a continuous random process in $\mathbb{R}^d$. The following are equivalent*

(a) *$(B_t)_{t\geq 0}$ is a Brownian motion in the sense of Section 7.1*
(b) *$(B_t)_{t\geq 0}$ is an $(\mathcal{F}_t^B)_{t\geq 0}$-Brownian motion.*

Proposition 7.4.2. *Let $(B_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion defined on $\Omega_0 \in \mathcal{F}_0$ and let F be a bounded measurable function on W_d. Then we can define a bounded measurable function f on $\mathbb{R}^d$ by*

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

and we have

$$\mathbb{E}(F(B)1_{\Omega_0}) = \mathbb{E}(f(B_0)1_{\Omega_0}).$$

Proposition 7.4.3. *Let $(B_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion in $\mathbb{R}$ starting from 0. Fix $\lambda \in \mathbb{R}$ and define*

$$Q_t = B_t^2 - t, \quad Z_t = \exp\{\lambda B_t - \lambda^2 t/2\}, \quad t \geq 0.$$

Then $(B_t)_{t\geq 0}$, $(Q_t)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ are continuous $(\mathcal{F}_t)_{t\geq 0}$-martingales.

Theorem 7.4.4. *Let $f \in C_b^{1,2}([0, \infty) \times \mathbb{R}^d)$ and let $(B_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion. Define $(M_t)_{t\geq 0}$ by*

$$M_t = f(t, B_t) - f(0, B_0) - \int_0^t \left(\tfrac{\partial}{\partial s} + \tfrac{1}{2}\Delta\right) f(s, B_s)ds.$$

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

Proof. It is straightforward to see that $(M_t)_{t\geq 0}$ is continuous, adapted and integrable. It remains to show, for $s, t \geq 0$, that

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

Fix $s \geq 0$ and set

$$\tilde{f}(t,x) = f(s+t,x), \quad \tilde{B}_t = B_{s+t}, \quad \tilde{\mathcal{F}}_t = \mathcal{F}_{s+t}.$$

Then $\tilde{B}$ is an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion and $M_{s+t} - M_s = \tilde{M}_t$, where

$$\begin{aligned}
\tilde{M}_t &= f(s+t, B_{s+t}) - f(s, B_s) - \int_s^{s+t} \left(\tfrac{\partial}{\partial r} + \tfrac{1}{2}\Delta\right) f(r, B_r) dr \\
&= \tilde{f}(t, \tilde{B}_t) - \tilde{f}(0, \tilde{B}_0) - \int_0^t \left(\tfrac{\partial}{\partial r} + \tfrac{1}{2}\Delta\right) \tilde{f}(r, \tilde{B}_r) dr
\end{aligned}$$

We have to show $\mathbb{E}(\tilde{M}_t | \tilde{\mathcal{F}}_0) = 0$ almost surely. Since this is the same problem for all $s \geq 0$, it will suffice to show that $\mathbb{E}(M_t | \mathcal{F}_0) = 0$ almost surely. Now $\mathbb{E}(M_t | \mathcal{F}_0) = m(B_0)$ almost surely, where $m(x) = \mathbb{E}_x(M_t)$ and the subscript x specifies the case $B_0 = x$. So it will suffice to show that $\mathbb{E}_x(M_t) = 0$ for all $x \in \mathbb{R}^d$.

Now $\mathbb{E}_x(M_s) \to 0$ as $s \to 0$, so it will suffice to show that $\mathbb{E}_x(M_t - M_s) = 0$ for all $x \in \mathbb{R}^d$ and all $0 < s < t$. We compute

$$\begin{aligned}
\mathbb{E}_x(M_t - M_s) &= \mathbb{E}_x\left(f(t, B_t) - f(s, B_s) - \int_s^t \left(\tfrac{\partial}{\partial r} + \tfrac{1}{2}\Delta\right) f(r, B_r) dr\right) \\
&= \mathbb{E}_x f(t, B_t) - \mathbb{E}_x f(s, B_s) - \int_s^t \mathbb{E}_x \left(\tfrac{\partial}{\partial r} + \tfrac{1}{2}\Delta\right) f(r, B_r) dr \\
&= \int_{\mathbb{R}^d} p(t,x,y) f(t,y) dy - \int_{\mathbb{R}^d} p(s,x,y) f(s,y) dy \\
&\qquad - \int_s^t \int_{\mathbb{R}^d} p(r,x,y) \left(\tfrac{\partial}{\partial r} + \tfrac{1}{2}\Delta\right) f(r,y) dy dr.
\end{aligned}$$

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

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

Hence $\mathbb{E}_x(M_t - M_s) = 0$ as required. $\square$

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.

7.5. **Strong Markov property.**

Theorem 7.5.1 (Strong Markov property). *Let $(B_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion and let T be a stopping time. Then $(B_{T+t})_{t\geq 0}$ is an $(\mathcal{F}_{T+t})_{t\geq 0}$-Brownian motion defined on $\{T < \infty\}$.*

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

$$\mathbb{E}(f(B_{T_n+t})1_{A_k})=\mathbb{E}(f(B_{t_k+t})1_{A_k})=\mathbb{E}(P_{t-s}f(B_{t_k+s})1_{A_k})=\mathbb{E}(P_{t-s}f(B_{T_n+s})1_{A_k})$$

On summing over k, we obtain

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

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

$$\mathbb{E}(f(B_{T+t})1_A)=\mathbb{E}(P_{t-s}f(B_{T+s})1_A)$$

and hence, since m and A were arbitrary, we have shown

$$\mathbb{E}(f(B_{T+t})|\mathcal{F}_{T+s})=P_{t-s}f(B_{T+s})\quad\text{almost surely on } \{T<\infty\}$$

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

We specialize to the case $d=1$.

Corollary 7.5.2 (Reflection principle). *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from 0 and let $a>0$. Set $T=\inf\{t\geq 0: B_t=a\}$ and define*

$$X_t=\begin{cases}2a-B_t, & \text{if } T\leq t\\ B_t, & \text{otherwise.}\end{cases}$$

Then $(X_t)_{t\geq 0}$ is also a Brownian motion starting from 0.

Proof. Note that T is a stopping time and $B_T=a$ on $\{T<\infty\}$. On the event $\{T<\infty\}$, set

$$\tilde{B}_t=B_{T+t}-B_T,\quad t\geq 0.$$

By the strong Markov property, conditional on $\{T<\infty\}$, $(\tilde{B}_t)_{t\geq 0}$ is a Brownian motion starting from 0 and independent of $\mathcal{F}_T$. Hence the same is true for $(-\tilde{B}_t)_{t\geq 0}$. But

$$B_t=B_{T\wedge t}+\tilde{B}_{(t-T)^+}1_{\{T<\infty\}},\quad X_t=B_{T\wedge t}-\tilde{B}_{(t-T)^+}1_{\{T<\infty\}},\quad t\geq 0.$$

Hence $(B_t)_{t\geq 0}$ and $(X_t)_{t\geq 0}$ have the same distribution. □

7.6. **Hitting times.** Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from 0. For $a\in\mathbb{R}$ we define the *hitting time*

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

Proposition 7.6.1. *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, given by

$$f(t)=(a/\sqrt{2\pi t^3})e^{-a^2/2t},\quad t\geq 0.$$

7.7. **Sample path properties.**

Proposition 7.7.1. *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from 0. Then, almost surely,*

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

Theorem 7.7.2. *Let B be a Brownian motion. Then, almost surely,*

(a) *for all* $\alpha < 1/2$, *B is locally Hölder continuous of exponent* α,
(b) *for all* $\alpha > 1/2$, *B 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

$$\|B_s - B_t\|_p \leq C|s-t|^{1/2}$$

where $C = \|B_1\|_p < \infty$. Then, by Kolmogorov's criterion, there exists $K \in L^p$ such that

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

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

For any non-trivial interval I, there exist $n \geq 0$ and $s, t \in \mathbb{D}_n$ such that$[s,t] \subseteq I$. Here $\mathbb{D}_n = \{k2^{-n} : k \in \mathbb{Z}^+\}$. Fix $m \in \mathbb{N}$ and let $s, t \in \mathbb{D}_n$ with $s < t$. Define for $m \geq n$

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

where the sum is taken over all $\tau \in \mathbb{D}_m$ such that $s \leq \tau < t$. The random variables $(B_{\tau+2^{-m}} - B_\tau)^2$ are then independent, of mean 2^{-m} and variance 2^{-2m+1}. For the variance, we used scaling and the fact that $\mathrm{var}(B_1^2) = 2$. Hence

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

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

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

so

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

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

Proposition 7.7.3 (Blumenthal's zero-one law)**.** *Let B be a Brownian motion in $\mathbb{R}^d$ starting from 0. Then*

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

Proposition 7.7.4. *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 $(B_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ starting from 0 *and let*

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

Then $T_C = 0$ almost surely.

7.8. **Recurrence and transience.**

Theorem 7.8.1. *Let B be a Brownian motion in $\mathbb{R}^d$.*

(a) *If $d = 1$, then*
$$\mathbb{P}(\{t \geq 0 : B_t = 0\} \text{ is unbounded}) = 1.$$

(b) *If $d = 2$, then*
$$\mathbb{P}(B_t = 0 \text{ for some } t > 0) = 0$$
but, for any $\varepsilon > 0$,
$$\mathbb{P}(\{t \geq 0 : |B_t| < \varepsilon\} \text{ is unbounded}) = 1.$$

(c) *If $d \geq 3$, then*
$$\mathbb{P}(|B_t| \to \infty \text{ as } t \to \infty) = 1.$$

The conclusions of this 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 $d \geq 3$.

Proof. Proposition 7.7.1(c) implies (a). To prove (b), we fix $a \in (0, 1)$ and $b > 1$ and consider the process $X_t = f(B_t)$, where $f \in C_b^2(\mathbb{R}^2)$ is chosen so that

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

Note that $\Delta f(x) = 0$ for $a \leq |x| \leq b$. Consider the process

$$M_t = f(B_t) - f(B_0) - \tfrac{1}{2}\int_0^t \Delta f(B_s)ds$$

and the stopping time

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

By Theorem 7.4.4, $(M_t)_{t\geq 0}$ is a martingale. Then, 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 $|B_0| = 1$. Then $M_T = \log|B_T|$, so $p = p(a, b) = \mathbb{P}(|B_T| = a)$ satisfies

$$p\log a + (1 - p)\log b = 0.$$

Consider first the limit $a \to 0$ with b fixed. Then $\log a \to -\infty$ so $p(a, b) \to 0$. Hence $\mathbb{P}_x(B_t = 0 \text{ for some } t > 0) = 0$ whenever $|x| = 1$. A scaling argument extends this to the case $|x| > 0$. For $x = 0$ and for all $\varepsilon > 0$, by the Markov property,

$$\mathbb{P}_0(B_t = 0 \text{ for some } t > \varepsilon) = \int_{\mathbb{R}^d} p(\varepsilon, 0, y)\mathbb{P}_y(B_t = 0 \text{ for some } t > 0)dy = 0.$$

Since $\varepsilon > 0$ is arbitrary, we deduce that $\mathbb{P}_0(B_t = 0 \text{ for some } t > 0) = 0$.

Consider now the limit $b \to \infty$ with $a = \varepsilon > 0$ fixed. Then $\log b \to \infty$, so $p(a, b) \to 1$. Hence $\mathbb{P}_x(|B_t| < \varepsilon \text{ for some } t > 0) = 1$ whenever $|x| = 1$. A scaling argument extends this to the case $|x| > 0$ and it is obvious by continuity for $x = 0$. It follows by the Markov property that, for all n, $\mathbb{P}(|B_t| < \varepsilon$ for some $t > n) = 1$ and hence that $\mathbb{P}(\{t \geq 0 : |B_t| < \varepsilon\}$ is unbounded$) = 1$.

We turn to the proof of (c). 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 $d = 3$. We have to show that, almost surely, for all $N \in \mathbb{N}$, $|B_t| > N$ for all sufficiently large t. Fix $N \in \mathbb{N}$. Define a sequence of stopping times $(T_k : k \geq 0)$ by setting $S_0 = 0$ and, for $k \geq 0$,

$$T_k = \inf\{t \geq S_k : |B_t| = N\}, \quad S_{k+1} = \inf\{t \geq T_k : |B_t| = N + 1\}.$$

Set $p = \mathbb{P}_x(|B_t| = N \text{ for some } t)$, where $|x| = N + 1$. We can use an argument similar to that used in (b), replacing the function $\log|x|$ by $1/|x|$, to see that $p = N/(N + 1) < 1$. By the strong Markov property,

$$\mathbb{P}(T_1 < \infty) \leq \mathbb{P}_N(T_1 < \infty) = p$$

and, for $k \geq 2$,

$$\mathbb{P}(T_k < \infty) = \mathbb{P}(T_1 < \infty)\mathbb{P}_N(T_{k-1} < \infty).$$

Hence $\mathbb{P}(T_k < \infty) \leq p^k$ and

$$\mathbb{P}(\{t \geq 0 : |B_t| = N\} \text{ is unbounded }) = \mathbb{P}(T_k < \infty \text{ for all } k) = 0$$

as required. □

7.9. **Brownian motion and the Dirichlet problem.** Let D be a connected open set in $\mathbb{R}^d$ with boundary ∂D and let $f : \partial D \to [0, \infty)$ and $g : D \to [0, \infty)$ be measurable functions. We assume that ∂D satisfies the following *exterior cone condition*: for all $y \in \partial D$, there exists $\varepsilon > 0$ and a relatively open set A in the unit sphere such that, for all $z \in A$ and all $t \in (0, \varepsilon)$, $y + tz \notin D$. This condition is satisfied if, for all $y \in \partial D$, there is a neighbourhood U of y in $\mathbb{R}^d$ and a C^1 map $F : U \to \mathbb{R}^d$ such that $F(y) = 0$, $F'(y)$ is invertible, and $D \cap U = \{x \in U : F(x) > 0\}$.

By a *solution to the Dirichlet problem* (*in* D *with data* f *and* g), we mean any function $\psi \in C^2(D) \cap C(\bar{D})$ satisfying

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

When $=$ is replaced by $\geq$ in this definition, twice, we say that ψ is a *supersolution.*

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

Proposition 7.9.1. *Let ϕ be a non-negative measurable function on D. Suppose that*

$$\phi(x) = \int_{S(x,\rho)} \phi(y)\mu_{x,\rho}(dy)$$

whenever $S(x, \rho) \subseteq D$. Then, either $\phi \equiv \infty$, or $\phi \in C^\infty(D)$ with $\Delta\phi = 0$.

Let B be a Brownian motion in $\mathbb{R}^d$. For a measurable function g and $t \geq 0$, we define functions $P_t g$ and Gg by

$$P_t g(x) = \mathbb{E}_x(g(B_t)), \quad Gg(x) = \mathbb{E}_x \int_0^\infty g(B_t)dt,$$

whenever the defining integrals exist.

Proposition 7.9.2. *We have*

(a)
$$\|P_t g\|_\infty \le (1 \wedge (2\pi t)^{-d/2} \operatorname{vol}(\operatorname{supp} g))\|g\|_\infty,$$

(b) *for* $d \ge 3$,
$$\|Gg\|_\infty \le (1 + \operatorname{vol}(\operatorname{supp} g))\|g\|_\infty,$$

(c) *for* $d \ge 3$ *and for* $g \in C^2(\mathbb{R}^d)$ *of compact support,* $Gg \in C_b^2(\mathbb{R}^d)$ *and*
$$-\tfrac{1}{2}\Delta G g = g.$$

Proof of (c). Note that
$$Gg(x) = \mathbb{E}_0 \int_0^\infty g(x + B_t)dt.$$
By differentiating this formula under the integral, using the estimate in (b), we see that $Gg \in C_b^2(\mathbb{R}^d)$.

To show that $-\frac{1}{2}\Delta Gg = g$, we fix $0 < s < t$ and write
$$Gg(x) = \mathbb{E}_0 \int_0^s g(x + B_r)dr + \int_s^t \int_{\mathbb{R}^d} p(r, x, y)g(y)dydr + \mathbb{E}_0 \int_t^\infty g(x + B_r)dr.$$
By differentiating under the integral we obtain
$$\begin{aligned} \tfrac{1}{2}\Delta Gg(x) = \tfrac{1}{2} \int_0^s \mathbb{E}_0 \Delta g(x + B_r)dr \\ + \tfrac{1}{2} \int_s^t \int_{\mathbb{R}^d} \Delta_x p(r, x, y)g(y)dydr + \tfrac{1}{2} \int_t^\infty \mathbb{E}_0 \Delta g(x + B_r)dr. \end{aligned}$$
We consider the limit $s \to 0$ and $t \to \infty$. By the estimate in (a), the first and third terms on the right tend to 0. Since $\frac{\partial}{\partial t} p = \frac{1}{2}\Delta p$, the second term is given by
$$\begin{aligned} &\int_{\mathbb{R}^d} \int_s^t \tfrac{\partial}{\partial r} p(r, x, y)g(y)drdy \\ &\qquad = \int_{\mathbb{R}^d} p(t, x, y)g(y)dy - \int_{\mathbb{R}^d} p(s, x, y)g(y)dy = P_t g(x) - \mathbb{E}_x g(B_s). \end{aligned}$$
Now $P_t g(x) \to 0$ and $\mathbb{E}_x g(B_s) \to g(x)$, so we obtain the desired identity. □

Theorem 7.9.3. *For* $x \in \bar{D}$, *set*
$$\phi(x) = \mathbb{E}_x \left(\int_0^T g(B_t)dt + f(B_T)1_{T<\infty} \right)$$
where $T = \inf\{t \ge 0 : B_t \in \partial D\}$. *Then*

(a) $\phi \le \psi$ *for any non-negative supersolution* ψ *of the Dirichlet problem,*
(b) $\phi = \psi$ *for any bounded solution* ψ *of the Dirichlet problem, provided* $T < \infty$, $\mathbb{P}_x$-*almost surely, for all* $x \in D$,
(c) ϕ *is a solution of the Dirichlet problem, provided* $f \in C(\partial D)$, $g \in C^2(\mathbb{R}^d)$ *and* ϕ *is locally bounded.*

Proof of (a). Let ψ be a supersolution of the Dirichlet problem. Fix $N \in \mathbb{N}$ and set

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

We can find $\theta \in C_b^2(\mathbb{R}^d)$ with $\theta = \psi$ on D_N. Set

$$M_t = \theta(B_t) - \theta(B_0) - \int_0^t \tfrac{1}{2}\Delta\theta(B_s)ds.$$

Then $(M_t)_{t\ge0}$ is a martingale, by Theorem 7.4.4. Denote by T_N the hitting time of ∂D_N. Then, by optional stopping, for $x \in D_N$,

$$\psi(x) = \mathbb{E}_x\psi(B_{T_N\wedge t}) + \mathbb{E}_x \int_0^{T_N\wedge t} (-\tfrac{1}{2}\Delta)\psi(B_t)dt.$$

We now let $t \to \infty$ and $N \to \infty$. Since ψ is a supersolution, by monotone convergence,

$$\mathbb{E}_x \int_0^{T_N\wedge t} (-\tfrac{1}{2}\Delta)\psi(B_t)dt \ge \mathbb{E}_x \int_0^{T_N\wedge t} g(B_t)dt \to \mathbb{E}_x \int_0^T g(B_t)dt$$

and on $\{T < \infty\}$ we have $\psi(B_{T_N\wedge t}) \to \psi(B_T) \ge f(B_T)$. Hence, if $\psi \ge 0$, then

$$\liminf \psi(B_{T_N\wedge t}) \ge f(B_T)1_{T<\infty}$$

and so, by Fatou's lemma,

$$\liminf \mathbb{E}_x\psi(B_{T_N\wedge t}) \ge \mathbb{E}_x(f(B_T)1_{T<\infty}).$$

Hence $\psi(x) \ge \phi(x)$. □

Proof of (b). In the case where ψ is a bounded solution of the Dirichlet problem and $T < \infty$, $\mathbb{P}_x$-almost surely, for all $x \in D$, we have, as $t \to \infty$ and $N \to \infty$,

$$\mathbb{E}_x \int_0^{T_N\wedge t} (-\tfrac{1}{2}\Delta)\psi(B_t)dt \to \mathbb{E}_x \int_0^T g(B_t)dt$$

and $\psi(B_{T_N\wedge t}) \to f(B_T)$ almost surely so, by bounded convergence,

$$\lim \mathbb{E}_x\psi(B_{T_N\wedge t}) = \mathbb{E}_x(f(B_T)).$$

Hence $\psi(x) = \phi(x)$. □

Proof of (c). Let D_0 be a bounded open subset of D. Set $T_0 = \inf\{t \ge 0 : B_t \notin D_0\}$. Then T_0 is a stopping time and $T_0 < \infty$ almost surely. Set

$$\tilde{B}_t = B_{T_0+t}, \quad \tilde{\mathcal{F}}_t = \mathcal{F}_{T_0+t}, \quad \tilde{T} = \inf\{t \ge 0 : \tilde{B}_t \notin D\}.$$

Note that $\tilde{T} < \infty$ if and only if $T < \infty$ and then $B_T = \tilde{B}_{\tilde{T}}$. By the strong Markov property, $(\tilde{B}_t)_{t\ge0}$ is an $(\tilde{\mathcal{F}}_t)_{t\ge0}$-Brownian motion, so

$$\begin{aligned}\phi(x) &= \mathbb{E}_x\left(\int_0^{T_0} g(B_t)dt + \int_0^{\tilde{T}} g(\tilde{B}_t)dt + f(\tilde{B}_{\tilde{T}})1_{\tilde{T}<\infty}\right)\\ &= \mathbb{E}_x\left(\int_0^{T_0} g(B_t)dt\right) + \mathbb{E}_x\left(\mathbb{E}\left(\int_0^{\tilde{T}} g(\tilde{B}_t)dt + f(\tilde{B}_{\tilde{T}})1_{\tilde{T}<\infty}\,\middle|\,\tilde{\mathcal{F}}_0\right)\right) \qquad (7.1)\\ &= \mathbb{E}_x\left(\int_0^{T_0} g(B_t)dt + \phi(B_{T_0})\right).\end{aligned}$$

Note that $\phi = f$ on ∂D. We now prove that, for $x \in D$ and $y \in \partial D$, 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. Consider, under $\mathbb{P}_0$, for each $x \in \bar{D}$, the stopping time $T_0(x) = \inf\{t \geq 0 : x + B_t \in \partial D_0\}$. Then

$$\phi(x) = \mathbb{E}_0\left(\int_0^{T_0(x)} g(x + B_t)dt + \phi(x + B_{T_0(x)})\right).$$

There exists an open cone C in $\mathbb{R}^d$ of positive height such that $y + C$ is disjoint from D. By Proposition 7.7.4, $T_C = \inf\{t \geq 0 : y + B_t \in C\} = 0$, $\mathbb{P}_0$-almost surely. Note that $T_C = 0$ implies that $x + B_{T_0(x)} \in \partial D$ for x sufficiently close to y and $T_0(x) \to 0$ and $x + B_{T_0(x)} \to y$ as $x \to y$. Since f is continuous on ∂D, this then implies that $\phi(x + B_{T_0(x)}) = f(x + B_{T_0(x)}) \to f(y)$ as $x \to y$. We have assumed that ϕ is locally bounded. Hence, by bounded convergence, $\phi(x) \to f(y)$ as $x \to y$, as claimed.

Consider for now the case where $g(x) = 0$ for all $x \in D$. Fix $x \in D$ and take $D_0 = B(x, \rho)$ where $\rho > 0$ is chosen so that $B(x, \rho) \subseteq D$. By rotational invariance, under $\mathbb{P}_x$, B_{T_0} has the uniform distribution $\mu_{x,\rho}$ on $S(x, \rho)$. Hence

$$\phi(x) = \mathbb{E}_x(\phi(B_{T_0})) = \int_{S(x,\rho)} \phi(y)\mu_{x,\rho}(dy).$$

Since ϕ is finite, it follows by Proposition 7.9.1 that $\phi \in C^\infty(D)$ with $\Delta\phi = 0$ in D. Then $\phi \in C(\bar{D})$ and ϕ is a solution of the Dirichlet problem.

By linearity, it now suffices to treat the case where $f(x) = 0$ for all $x \in \partial D$. Moreover, it also suffices to treat the case where $d \geq 3$. For, if $d = 1$ or $d = 2$, we can simply apply the result for $d = 3$ to cylindrical regions D and to functions g which depend only on the first and second coordinates. Assume, for now, that D is bounded. Set

$$\phi_0(x) = \mathbb{E}_x \int_0^\infty \tilde{g}(B_t)dt$$

where $\tilde{g}$ is a compactly supported function agreeing with g on D. By Proposition 7.9.2, $\phi_0 \in C_b^2(\mathbb{R}^d)$ with $-\frac{1}{2}\Delta\phi_0 = \tilde{g}$. On taking $\phi = \phi_0$ and $D = \mathbb{R}^d, D_0 = D$ in (7.1), we find that $\phi_0(x) = \phi(x) + \phi_1(x)$ where

$$\phi_1(x) = \mathbb{E}_x(\phi_0(B_T)).$$

We showed above that this implies ϕ_1 is harmonic in D so $-\frac{1}{2}\Delta\phi = g$ in D as required. Finally, if D is unbounded, we can go back to (7.1) to see that $-\frac{1}{2}\Delta\phi = g$ in D_0, for all bounded open sets $D_0 \subseteq D$, and hence in D. □

7.10. **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.

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 $0 = T_0 \leq T_1 \leq T_2 \leq \dots$ 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 7.6.1 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 $(T_1, S_1), \dots, (T_n, S_n)$. The result follows. □

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 7.10.2 (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 $T_n/n \to 1$ almost surely as $n \to \infty$. So, as $N \to \infty$,

$$N^{-1} \sup_{n\leq N\tau} |T_n - n| \to 0 \quad \text{almost surely.}$$

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 7.10.3 (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.*) For disjoint sets $A_1, \dots, A_k \in \mathcal{E}$ and $n_1, \dots, n_k \in \mathbb{Z}^+$, set

$$A^* = \{m \in E^* : m(A_1) = n_1, \dots, m(A_k) = n_k\}.$$

Then, for any measure μ^* making X a Poisson random measure with intensity μ,

$$\mu^*(A^*) = \prod_{j=1}^{k} e^{-\mu(A_j)} \mu(A_j)^{n_j}/n_j!$$

Since the set of such sets 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

$$M(A) = \sum_{n=1}^{N} 1_{\{Y_n \in A\}}, \quad A \in \mathcal{E}. \tag{8.1}$$

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

$$\mathbb{E}(M(g)) = \int_E g(y)\mu(dy), \quad \text{var}(M(g)) = \int_E g(y)^2 \mu(dy).$$

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. $\square$

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 $\mu(E_n) < \infty$ 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

$$\mathbb{E}(e^{iu(X_t^n - X_s^n)} 1_A) = \exp\left\{(t - s)\int_E (e^{iug_n(y)} - 1 - iug_n(y))K(dy)\right\}\mathbb{P}(A)$$

to see that $(X_t)_{t\geq 0}$ has stationary independent increments and (8.3) holds. $\square$

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

$$X_t = \sqrt{a}B_t + bt + \int_{(0,t]\times\{|y|\le 1\}} y\tilde{M}(ds, dy) + \int_{(0,t]\times\{|y|>1\}} yM(ds, dy).$$

We interpret the last integral as 0 on the null set $\{M((0,t] \times \{|y| > 1\}) = \infty \text{ for some } t \ge 0\}$. Then $(X_t)_{t\ge 0}$ is a Lévy process and, for all $t \ge 0$,

$$\mathbb{E}(e^{iuX_t}) = e^{t\psi(u)}$$

where

$$\psi(u) = \psi_{a,b,K}(u) = ibu - \tfrac{1}{2}au^2 + \int_{\mathbb{R}} (e^{iuy} - 1 - iuy1_{|y|\le 1}) K(dy).$$

Thus, to every Lévy triple there corresponds a Lévy process. Moreover, given $(X_t)_{t\ge 0}$, we can recover M by

$$M((0,t] \times A) = \#\{s \le t : X_s - X_{s-} \in A\}$$

and so we can also recover b and $\sqrt{a}B$. Hence the law of the Lévy process $(X_t)_{t\ge 0}$ 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 $t \ge 0$ and all $u \in \mathbb{R}$,*

$$\mathbb{E}(e^{iuX_t}) = e^{t\psi_{a,b,K}(u)}.$$

Proof. For $t \ge 0$ 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 $(X_t)_{t\ge 0}$ 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 $(X_t)_{t\ge 0}$ is cadlag, as $t \to s$ with $t > s$, we have $X_t \to X_s$, so

$$|\phi_t(u) - \phi_s(u)| \le \mathbb{E}|e^{iu(X_t - X_s)} - 1| \le \mathbb{E}((u|X_t - X_s|) \wedge 2) \to 0$$

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

$$1_{\{|y|\leq 1\}} \leq \chi(y) \leq 1_{\{|y|\leq 2\}}.$$

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

$$b = \beta - \int_{\mathbb{R}} y(\chi(y) - 1_{\{|y|\leq 1\}})K(dy).$$

□