ADVANCED PROBABILITY

J. R. NORRIS

Contents

Schedule

This course aims to cover the advanced topics at the core of research in probability. There is an emphasis on techniques needed for the rigorous analysis of stochastic processes such as Brownian motion. The course finishes with two key structural results – Donsker's invariance principle and the Lévy–Khinchin theorem.

It will be assumed that students have some familiarity with the measure-theoretic formulation of probability – at the level of the Part II(B) course Probability and Measure, or Part A of Williams' book.

Review of the basics of measure and integration theory, as covered for example in the Part II(B) course Probability and Measure.

Conditional expectation: discrete case, Gaussian case, conditional density functions; existence and uniqueness; basic properties.

Discrete parameter martingales, submartingales and supermartingales; optional stopping; Doob's inequalities, upcrossings, convergence theorems, backwards martingales.

Applications of martingales: sums of independent random variables, strong law of large numbers, Wald's identity; non-negative martingales and change of measure, Radon–Nikodym theorem, Kakutani's product martingale theorem, consistency of likelihood-ratio tests; Markov chains; stochastic optimal control.

Continuous-time random processes: Kolmogorov's criterion, path regularization theorem for martingales, continuous-time martingales.

Weak convergence in $\mathbb{R}^n$: convergence of distribution functions, convergence with respect to continuous bounded functions, Skorokhod embedding, Helly's theorem. Characteristic functions, Lévy's continuity theorem.

Brownian motion: Wiener's theorem, Scaling and symmetry properties. Martingales associated to Brownian motion, strong Markov property, reflection principle, hitting times. Sample path properties, recurrence and transience. Brownian motion and the Dirichlet problem. Donsker's invariance principle.

Lévy processes: construction of pure jump Lévy processes by integrals with respect to a Poisson random measure. Infinitely divisible laws, Lévy–Khinchin theorem.

Appropriate books

R. Durrett, *Probability: Theory and Examples.* Wadsworth 1991

O. Kallenberg, *Foundations of Morern Probability.* Springer 1997

L.C.G. Rogers and D. Williams, *Diffusions, Markov processes, and Martingales Vol. I (2nd edition).* Chapters I & II. Wiley 1994

D.W. Stroock, *Probability Theory – An analytic view.* Chapters I–V. Cambridge University Press 1993

D. Williams, *Probability with Martingales.* Cambridge University Press 1991

11. Conditional expectation

11.1. **Discrete case.** Let $(G_i : i \in I)$ denote a countable family of disjoint events, whose union is the whole probability space. Set $\mathcal{G} = \sigma(G_i : i \in I)$. For any integrable random variable X, we can define

$$Y = \sum_i \mathbb{E}(X|G_i) 1_{G_i}$$

where we set $\mathbb{E}(X|G_i) = \mathbb{E}(X 1_{G_i})/\mathbb{P}(G_i)$ when $\mathbb{P}(G_i) > 0$ and define $\mathbb{E}(X|G_i)$ in some arbitrary way when $\mathbb{P}(G_i) = 0$. Then it is easy to see 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 $A \in \mathcal{G}$.

11.2. **Gaussian case.** Let (W, X) be a Gaussian random variable in $\mathbb{R}^2$. Set $\mathcal{G} = \sigma(W)$ and $Y = aW + b$, 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 $\mathbb{E}(X 1_A) = \mathbb{E}(Y 1_A)$ for all $A \in \mathcal{G}$.

11.3. **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 a 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 we agree, say, that $0/0 = 0$. 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 $\mathbb{E}(X 1_A) = \mathbb{E}(Y 1_A)$ 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,

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

11.4. **Existence and uniqueness.**

Theorem 11.4.1. *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 $Y = Y'$ a.s..*

We call Y (*a version of*) *the conditional expectation of X given* $\mathcal{G}$ and write $Y = \mathbb{E}(X|\mathcal{G})$ a.s.. In the case $\mathcal{G} = \sigma(G)$ for some random variable G, we also write $Y = \mathbb{E}(X|G)$ a.s.. 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 $X \leq X'$ a.s.. Consider the non-negative random variable $Z = (Y - Y')1_A$, where $A = \{Y \geq Y'\} \in \mathcal{G}$. Then

$$\mathbb{E}(Z) = \mathbb{E}(Y1_A) - \mathbb{E}(Y'1_A) = \mathbb{E}(X1_A) - \mathbb{E}(X'1_A) \leq 0$$

so $Z = 0$ a.s., which implies $Y \leq Y'$ a.s.. In the case $X = X'$, we deduce that $Y = Y'$ a.s..

(*Existence.*) Assume to begin that $X \in L^2(\mathcal{F})$. Since $V = L^2(\mathcal{G})$ is a closed subspace of $L^2(\mathcal{F})$, we have $X = Y + W$ for some $Y \in V$ and $W \in V^\perp$. Then, for any $A \in \mathcal{G}$, we have $1_A \in V$, so

$$\mathbb{E}(X1_A) - \mathbb{E}(Y1_A) = \mathbb{E}(W1_A) = 0.$$

Hence Y satisfies (a) and (b).

Assume now that X is any non-negative random variable. 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 $0 \leq Y_n \leq Y_{n+1}$ a.s.. Set $Y = \lim_{n\to\infty} Y_n$, then Y is $\mathcal{G}$-measurable and, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathbb{E}(X1_A) = \mathbb{E}(Y1_A).$$

In particular, if $\mathbb{E}(X)$ is finite then so is $\mathbb{E}(Y)$.

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). $\square$

11.5. **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 11.4.1:

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

In the proof of Theorem 11.4.1, we showed also

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

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

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

To see this, one checks that the right hand side has the defining properties(a) and (b) of the left hand side.

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

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

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

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

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

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 countably many affine functions:

$$c(x) = \sup_i (a_i x + b_i), \quad x \in \mathbb{R}.$$

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

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

So we obtain

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

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* $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H})$ *a.s.*.

We can always *take out what is known*:

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

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 $\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$ a.s.. The result extends to simple $\mathcal{G}$-measurable random variables Y by linearity, then to the case $X \geq 0$ and any 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* $\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G})$ *a.s.*.

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 Proposition 3.1.4.

Lemma 11.5.1. *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 Lemma 6.2.1, given $\varepsilon > 0$, we can find $\delta > 0$ so that $\mathbb{E}(|X|1_A) \leq \varepsilon$ whenever $\mathbb{P}(A) \leq \delta$. Then choose $\lambda < \infty$ so that $\mathbb{E}(|X|) \leq \lambda\delta$. Suppose $Y = \mathbb{E}(X|\mathcal{G})$, then $|Y| \leq \mathbb{E}(|X||\mathcal{G})$. In particular, $\mathbb{E}(|Y|) \leq \mathbb{E}(|X|)$ so

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

Then

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

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

12. Martingales – theory

12.1. **Definitions.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(E, \mathcal{E})$ be a measurable space and let I be a countable subset of $\mathbb{R}$. A *process in* E is a family $X = (X_t)_{t\in I}$ of random variables in E. A *filtration* $(\mathcal{F}_t)_{t\in I}$ is an increasing family of sub-σ-algebras of $\mathcal{F}$: thus $\mathcal{F}_s \subseteq \mathcal{F}_t$ whenever $s \leq t$. We set $\mathcal{F}_{-\infty} = \cap_{t\in I}\mathcal{F}_t$ and $\mathcal{F}_\infty = \sigma(\mathcal{F}_t : t \in I)$. Every process has a *natural filtration* $(\mathcal{F}_t^X)_{t\in I}$, given by

$$\mathcal{F}_t^X = \sigma(X_s : s \leq t).$$

We will always assume some filtration $(\mathcal{F}_t)_{t\in I}$ to be given. The σ-algebra $\mathcal{F}_t$ is interpreted as modelling the state of our knowledge at time t. In particular, $\mathcal{F}_t^X$ contains all the events which depend (measurably) only on $X_s, s \leq t$, that is, everything we know about the process X by time t. We say that X is *adapted* (*to* $(\mathcal{F}_t)_{t\in I}$) if X_t is $\mathcal{F}_t$-measurable for all t. Of course every process is adapted to its natural filtration. Unless otherwise indicated, it is to be understood from now on that $E = \mathbb{R}$. We say that X is *integrable* if X_t is integrable for all t. A *martingale* X is an adapted integrable process such that, for all $s, t \in I$ with $s \leq t$,

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

On replacing the equality in this condition by $\leq$ or $\geq$, we get the notions of *supermartingale* and *submartingale,* respectively. Note that every process which is a martingale with respect to the given filtration is also a martingale with respect to its natural filtration.

12.2. **Optional stopping.** We say that a random variable $T : \Omega \to I \cup \{\infty\}$ is a *stopping time* if $\{T \leq t\} \in \mathcal{F}_t$ for all t. For a stopping time T, we set

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

It is easy to check that, if $T \equiv t$, then T is a stopping time and $\mathcal{F}_T = \mathcal{F}_t$. Given a process X, we set $X_T(\omega) = X_{T(\omega)}(\omega)$ whenever $T(\omega) < \infty$. We also define the *stopped process* X^T by $X_t^T = X_{T\wedge t}$.

We assume in the following two results that $I = \{0, 1, 2, \dots\}$. In this context, we will write n, m or k for elements of I, rather than t or s.

Proposition 12.2.1. *Let S and T be stopping times and let $X = (X_n)_{n\geq 0}$ be an adapted process. Then*

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

Theorem 12.2.2 (Optional stopping theorem)**.** *Let $X = (X_n)_{n\geq 0}$ 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,*

$$\mathbb{E}(X_T|\mathcal{F}_S) \leq X_{S\wedge T} \quad a.s.,$$

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

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

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

$$X_T = X_{S\wedge T} + \sum_{S\leq k<T}(X_{k+1}-X_k) = X_{S\wedge T} + \sum_{k=0}^{n}(X_{k+1}-X_k)1_{S\leq k<T}. \tag{12.1}$$

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

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

Hence, on multiplying (12.1)by 1_A and taking expectations, we obtain

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

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

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

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

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

It follows that (d) implies (a). $\square$

12.3. **Doob's inequalities.** Let X be a process and let $a, b \in \mathbb{R}$ with $a < b$. For $J \subseteq I$, set

$$U([a,b],J) = \sup\{n : X_{s_1} < a, X_{t_1} > b, \dots, X_{s_n} < a, X_{t_n} > b \text{ for some } s_1 < t_1 < \dots < s_n < t_n \text{ in } J\}.$$

Then $U[a,b] = U([a,b],I)$ is the number of *upcrossings* of $[a,b]$ by X.

Theorem 12.3.1 (Doob's upcrossing inequality)**.** *Let X be a supermartingale. Then*

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

Proof. Since $U([a,b],I) = \lim_{J\uparrow I, J \text{ finite}} U([a,b],J)$,it suffices, by monotone convergence, to consider the case where I is finite. Let us assume then that $I = \{0,1,\dots,n\}$.

Write $U = U[a,b]$ and note that $U \leq n$. Set $T_0 = 0$ and define inductively for $k \geq 0$:

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

As usual $\inf \emptyset = \infty$. Then $U = \max\{k : T_k < \infty\}$. For $k \leq U$, set $G_k = X_{T_k} - X_{S_k}$ and note that $G_k \geq b - a$. Note that $T_U \leq n$ and $T_{U+1} = \infty$. Set

$$R = \begin{cases} X_n - X_{S_{U+1}} & \text{if } S_{U+1} < \infty, \\ 0 & \text{if } S_{U+1} = \infty \end{cases}$$

and note that $R \geq -(X_n - a)^-$.

Then we have

$$\sum_{k=1}^{n} (X_{T_k \wedge n} - X_{S_k \wedge n}) = \sum_{k=1}^{U} G_k + R \geq (b-a)U - (X_n - a)^-. \tag{12.2}$$

Now X is a supermartingale and $S_k \wedge n$ and $T_k \wedge n$ are bounded stopping times, with $S_k \wedge n \leq T_k \wedge n$. Hence, by optional stopping, $\mathbb{E}(X_{T_k \wedge n}) \leq \mathbb{E}(X_{S_k \wedge n})$ and the desired inequality results on taking expectations in (12.2). □

For any process X, for $J \subseteq I$, we set

$$X^*(J) = \sup_{t \in J} |X_t|, \quad X^* = X^*(I).$$

Theorem 12.3.2 (Doob's maximal inequality)**.** *Let X be a martingale or a non-negative submartingale. Then, for all $\lambda \geq 0$,*

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

Proof. Note that

$$\lambda \mathbb{P}(X^* \geq \lambda) = \lim_{\nu \uparrow \lambda} \nu \mathbb{P}(X^* > \nu) \leq \lim_{\nu \uparrow \lambda} \lim_{J \uparrow I, J \text{ finite}} \nu \mathbb{P}(X^*(J) \geq \nu).$$

It therefore suffices to consider the case where I is finite. Let us assume then that $I = \{0, 1, \ldots, n\}$. 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 $T = \inf\{m \geq 0 : X_m \geq \lambda\} \wedge n$. Then T is a stopping time and $T \leq n$ so, by optional stopping,

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

Hence

$$\lambda \mathbb{P}(X^* \geq \lambda) \leq \mathbb{E}(X_n 1_{X^* \geq \lambda}) \leq \mathbb{E}(X_n). \tag{12.3}$$

□

Theorem 12.3.3 (Doob's L^p-inequality)**.** *Let X be a martingale or non-negative submartingale. Then, for all $p > 1$ and $q = p/(p-1)$,*

$$\|X^*\|_p \leq q \sup_{t \in I} \|X_t\|_p.$$

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. Since $X^* = \lim_{J\uparrow I, J \text{ finite}} X^*(J)$, it suffices, by monotone convergence, to consider the case where I is finite. Let us assume then that $I = \{0, 1, \dots, n\}$.

Fix $k < \infty$. By Fubini's theorem, equation (12.3) and Hölder's inequality,

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

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

12.4. **Convergence theorems.** Recall that, for $p \geq 1$, a process X is said to be *bounded in* L^p if $\sup_{t\in I} \|X_t\|_p < \infty$. Also X is *uniformly integrable* if

$$\sup_{t\in I} \mathbb{E}(|X_t| 1_{|X_t|>k}) \to 0 \quad \text{as} \quad k \to \infty.$$

Recall from §6 that, if X is bounded in L^p for some $p > 1$, then X is uniformly integrable. Also if X is uniformly integrable then X is bounded in L^1.
The next two results are stated for the case $\sup I = \infty$.

Theorem 12.4.1 (Almost sure martingale convergence theorem)**.** *Let X be a supermartingale which is bounded in L^1. Then $X_t \to X_\infty$ a.s. as $t \to \infty$, for some $X_\infty \in L^1(\mathcal{F}_\infty)$.*

Note that, if $\inf I \in I$, then a non-negative supermartingale is automatically bounded in L^1.

Proof. By Doob's upcrossing inequality, for all $a < b$,

$$\mathbb{E}(U[a,b]) \leq (b-a)^{-1} \sup_{t\in I} \mathbb{E}(|X_t| + |a|) < \infty.$$

Consider for $a < b$ the sets

$$\Omega_{a,b} = \{\liminf_{t\to\infty} X_t < a < b < \limsup_{t\to\infty} X_t\},$$

$$\Omega_0 = \{X_t \text{ converges in } [-\infty, \infty] \text{ as } t \to \infty\}.$$

Since $U[a,b] = \infty$ on $\Omega_{a,b}$, we must have $\mathbb{P}(\Omega_{a,b}) = 0$. But

$$\Omega_0 \cup (\cup_{a,b\in\mathbb{Q}, a<b} \Omega_{a,b}) = \Omega$$

so we deduce $\mathbb{P}(\Omega_0) = 1$. Define

$$X_\infty = \begin{cases} \lim_{t\to\infty} X_t & \text{on } \Omega_0, \\ 0 & \text{on } \Omega \setminus \Omega_0 \end{cases}$$

Then X_∞ is $\mathcal{F}_\infty$-measurable and, by Fatou's lemma,

$$\mathbb{E}(|X_\infty|) \le \liminf_{t\to\infty} \mathbb{E}(|X_t|) < \infty.$$

so $X_\infty \in L^1$ as required. $\square$

Let us denote by $\mathcal{M}^1$ the set of uniformly integrable martingales and, for $p > 1$, by $\mathcal{M}^p$ the set of martingales bounded in L^p.

Theorem 12.4.2 (L^p martingale convergence theorem)**.** *Let $p \in [1, \infty)$.*

(a) *Suppose $X \in \mathcal{M}^p$. Then $X_t \to X_\infty$ as $t \to \infty$, a.s. and in L^p, for some $X_\infty \in L^p(\mathcal{F}_\infty)$. Moreover, $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$ a.s. for all t.*

(b) *Suppose $Y \in L^p(\mathcal{F}_\infty)$ and set $X_t = \mathbb{E}(Y|\mathcal{F}_t)$. Then $X = (X_t)_{t\in I} \in \mathcal{M}^p$ and $X_t \to Y$ as $t \to \infty$, a.s. and in L^p.*

Thus the map $X \mapsto X_\infty$ is a one-to-one correspondence between $\mathcal{M}^p$ and $L^p(\mathcal{F}_\infty)$.

Proof for $p = 1$. Let X be a uniformly integrable martingale. Then $X_t \to X_\infty$ a.s. by the almost sure martingale convergence theorem. Since X is UI, it follows that $X_t \to X_\infty$ in L^1, by Theorem 6.2.3. Next, for $s \ge t$,

$$\|X_t - \mathbb{E}(X_\infty|\mathcal{F}_t)\|_1 = \|\mathbb{E}(X_s - X_\infty|\mathcal{F}_t)\|_1 \le \|X_s - X_\infty\|_1.$$

Let $s \to \infty$ to deduce $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$ a.s..

Suppose now that $Y \in L^1(\mathcal{F}_\infty)$ and set $X_t = \mathbb{E}(Y|\mathcal{F}_t)$. Then $X = (X_t)_{t\in I}$ is a martingale by the tower property and is uniformly integrable by Lemma 11.5.1. Hence X_t converges a.s. and in L^1, with limit X_∞, say. For all t and all $A \in \mathcal{F}_t$ we have

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

Now $X_\infty, Y \in L^1(\mathcal{F}_\infty)$ and $\cup_t \mathcal{F}_t$ is a π-system generating $\mathcal{F}_\infty$. Hence, by Proposition 3.1.4, $X_\infty = Y$ a.s.. $\square$

Proof for $p > 1$. Let X be a martingale bounded in L^p for some $p > 1$. Then $X_t \to X_\infty$ a.s. by the almost sure martingale convergence theorem. By Doob's L^p-inequality,

$$\|X^*\|_p \le q \sup_{t\in I} \|X_t\|_p < \infty.$$

Since $|X_t - X_\infty|^p \le (2X^*)^p$ for all t, we can use dominated convergence to deduce that $X_t \to X_\infty$ in L^p. It follows that $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$ a.s., as in the case $p = 1$.

Suppose now that $Y \in L^p(\mathcal{F}_\infty)$ and set $X_t = \mathbb{E}(Y|\mathcal{F}_t)$. Then $X = (X_t)_{t\in I}$ is a martingale by the tower property and

$$\|X_t\|_p = \|\mathbb{E}(Y|\mathcal{F}_t)\|_p \le \|Y\|_p$$

for all t, so X is bounded in L^p. Hence X_t converges a.s. and in L^p, with limit X_∞, say, and we can show that $X_\infty = Y$ a.s., as in the case $p = 1$. $\square$

In the next result we assume $\inf I = -\infty$.

Theorem 12.4.3 (Backward martingale convergence theorem)**.** *Let $p \in [1, \infty)$ and let $Y \in L^p$. Set $X_t = \mathbb{E}(Y|\mathcal{F}_t)$. Then $X_t \to \mathbb{E}(Y|\mathcal{F}_{-\infty})$ as $t \to -\infty$, a.s. and in L^p.*

Proof. The argument is a minor modification of that used in Theorems 12.3.1, 12.4.1, 12.4.2. The process X is automatically UI, by Proposition 11.5.1, and is bounded in L^p because $\|X_t\|_p = \|\mathbb{E}(Y|\mathcal{F}_t)\|_p \le \|Y\|_p$ for all t. We leave the details to the reader. □

In the following result we take $I = \{0, 1, 2, \dots\}$.

Theorem 12.4.4 (Optional stopping theorem (continued))**.** *Let X be a UI martingale and let S and T be stopping times. Then*

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

Proof. We have already proved the result when T is bounded. If T is unbounded, then $T \wedge n$ is a bounded stopping time, so

$$\mathbb{E}(X^T_n|\mathcal{F}_S) = \mathbb{E}(X_{T\wedge n}|\mathcal{F}_S) = X_{S\wedge T\wedge n} = X^T_{S\wedge n} \quad \text{a.s..} \tag{12.4}$$

Now

$$\|\mathbb{E}(X^T_n|\mathcal{F}_S) - \mathbb{E}(X_T|\mathcal{F}_S)\|_1 \le \|X^T_n - X^T_\infty\|_1. \tag{12.5}$$

We have $X_n \to X_\infty$ in L^1. So, in the case $T \equiv \infty$, we can pass to the limit in (12.4) to obtain

$$\mathbb{E}(X_\infty|\mathcal{F}_S) = X_S \quad \text{a.s..}$$

Then, returning to (12.5), for general T, we have

$$\|X^T_n - X^T_\infty\|_1 = \|\mathbb{E}(X_n - X_\infty|\mathcal{F}_T)\|_1 \le \|X_n - X_\infty\|_1$$

and the result follows on passing to the limit in (12.4). □

13. Martingales – applications

13.1. **Sums of independent random variables.** Throughout this section $(X_n : n \in \mathbb{N})$ will denote a sequence of independent random variables. We shall use martingale arguments to analyse the behaviour of the sums

$$S_0 = 0, \quad S_n = X_1 + \dots + X_n, \quad n \in \mathbb{N}.$$

Theorem 13.1.1 (Strong law of large numbers)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent and identically distributed random variables in L^1 and set $\mu = \mathbb{E}(X_1)$. Then $S_n/n \to \mu$ a.s. and in L^1.*

Proof. Define for $n \geq 1$

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

Then $\mathcal{F}_{-n} = \sigma(S_n, \mathcal{T}_n)$. Since X_1 is independent of $\mathcal{T}_n$, we have $\mathbb{E}(X_1|\mathcal{F}_{-n}) = \mathbb{E}(X_1|S_n)$ for all n. Now, for all $A \in \mathcal{B}$ and $k = 1, \ldots, n$, by symmetry, $\mathbb{E}(X_k 1_{S_n \in A})$ does not depend on k. Hence $\mathbb{E}(X_k|S_n)$ does not depend on k. But $\mathbb{E}(X_1|S_n) + \cdots + \mathbb{E}(X_n|S_n) = \mathbb{E}(S_n|S_n) = S_n$,so we must have $\mathbb{E}(X_1|S_n) = S_n/n$ a.s..

Set $M_{-n} = S_n/n$. We have shown that $(M_n)_{n\leq 0}$ is an $(\mathcal{F}_n)_{n\leq 0}$-martingale. So, by the backward martingale convergence theorem, S_n/n converges a.s. and in L^1. Finally, by Kolmogorov's zero-one law, the limit, Y say, is a.s. constant. So $Y = \mathbb{E}(Y) = \lim_n \mathbb{E}(S_n/n) = \mu$ a.s.. □

Proposition 13.1.2. *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables in L^2 and set*

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

Suppose that the series $\sum_n \mu_n$ and $\sum_n \sigma_n^2$ both converge in $\mathbb{R}$. Then S_n converges a.s. and in L^2.

Proposition 13.1.3 (Wald's identity)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed random variables with $\mathbb{P}(X_1 = 0) < 1$. Let $a, b \in \mathbb{R}$ with $a < 0 < b$ and set*

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

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

Set $M(\lambda) = \mathbb{E}(\exp(\lambda X_1))$. Then, for any $\lambda \in \mathbb{R}$ such that $M(\lambda) < \infty$ and $\mathbb{E}(M(\lambda)^{-T}) < \infty$, we have

$$\mathbb{E}(M(\lambda)^{-T} \exp(\lambda S_T)) = 1.$$

13.2. **Non-negative martingales and change of measure.**

Proposition 13.2.1. *Let $(X_n)_{n\geq 0}$ be a non-negative adapted process, with $\mathbb{E}(X_n) = 1$ for all n.*

(a) *We can define for each n a probability measure $\tilde{\mathbb{P}}_n$ on $\mathcal{F}_n$ by*

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

These measures 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.

(b) *Assume that $(X_n)_{n\geq 0}$ is a martingale. Then there exists a probability measure $\tilde{\mathbb{P}}$ on $\mathcal{F}_\infty$ 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.*

(c) *Assume that $\mathbb{E}(X_T) = 1$ for all finite stopping times T. Then there exists an $\mathcal{F}_\infty$-measurable random variable X such that $\tilde{\mathbb{P}}(A) = \mathbb{E}(X 1_A)$ for all $A \in \mathcal{F}_\infty$ if and only if $(X_n)_{n\geq 0}$ is uniformly integrable.*

Proof of (b). Since $(X_n)_{n\geq 0}$ is a martingale, by (a), we can define 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, $\tilde{\mathbb{P}}$ 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 adapted partitions $(A_n : n \geq 0)$ of Ω. Hence it suffices to note that adapted partitions are in one-to-one correspondence with finite stopping times T, by $\{T = n\} = A_n$, and then

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

□

Theorem 13.2.2 (Radon–Nikodym theorem)**.** *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$. Assume that $\mathcal{F}$ is* countably generated*, that is, for some sequence of sets $(F_n : n \in \mathbb{N})$,*

$$\mathcal{F} = \sigma(F_n : n \in \mathbb{N}).$$

Then the following are equivalent:

(a) *$\mathbb{P}(A) = 0$ implies $\tilde{\mathbb{P}}(A) = 0$ for all $A \in \mathcal{F}$,*

(b) *there exists a random variable $X \geq 0$ such that*

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

The random variable X, which is unique $\mathbb{P}$-a.s., is called (a version of) the *Radon-Nikodym derivative* of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$. We write $X = d\tilde{\mathbb{P}}/d\mathbb{P}$ a.s. The theorem extends immediately to finite measures by scaling, then to σ-finite measures by breaking Ω into pieces where the measures are finite. The assumption that $\mathcal{F}$ is countably generated can also be removed but we do not give the details here.

Proof. It is obvious that (b) implies (a). Assume then that (a) holds. Set $\mathcal{F}_n = \sigma(F_k : k \leq n)$. For each n, we can define an $\mathcal{F}_n$-measurable random variable X_n such that $\tilde{\mathbb{P}}(A) = \mathbb{E}(X_n 1_A)$ for all $A \in \mathcal{F}_n$. For, we can find disjoint sets $A_1, \dots, A_m$ such that $\mathcal{F}_n = \sigma(A_1, \dots, A_m)$ and then

$$X_n = \sum_{j=1}^{m} \frac{\tilde{\mathbb{P}}(A_j)}{\mathbb{P}(A_j)} 1_{A_j}$$

has the required property. We agree here to set $0/0 = 0$.

The process $(X_n)_{n\geq 0}$ is a martingale, which we will show 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 $\mathbb{Q}(A) = \mathbb{E}(X 1_A)$ for

$A \in \mathcal{F}$. Then $\mathbb{Q}$ is a probability measure and $\mathbb{Q} = \tilde{\mathbb{P}}$ on $\cup_n \mathcal{F}_n$, which is a π-system generating $\mathcal{F}$. Hence $\mathbb{Q} = \tilde{\mathbb{P}}$ on $\mathcal{F}$, 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 $\tilde{\mathbb{P}}(B) < \varepsilon$ whenever $\mathbb{P}(B) < \delta, B \in \mathcal{F}$. 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 $\mathbb{P}(B_n \text{ i.o.}) = 0$ and $\tilde{\mathbb{P}}(B_n \text{ i.o.}) \geq \varepsilon$, contradicting (a). Set $\lambda = 1/\delta$, then, for all n, we have $\mathbb{P}(X_n > \lambda) \leq \mathbb{E}(X_n)/\lambda = 1/\lambda = \delta$, 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. □

Theorem 13.2.3 (Kakutani's product martingale theorem)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent non-negative random variables of mean* 1*. Set*

$$M_0 = 1, \quad M_n = X_1 X_2 \dots X_n, \quad n \in \mathbb{N}.$$

Then $(M_n)_{n\geq 0}$ is a non-negative martingale and $M_n \to M_\infty$ a.s. for some random variable M_∞. Set $a_n = \mathbb{E}(\sqrt{X_n})$, then $a_n \in (0, 1]$. Moreover,

(a) *if $\prod_n a_n > 0$, then $M_n \to M_\infty$ in L^1 and $\mathbb{E}(M_\infty) = 1$,*
(b) *if $\prod_n a_n = 0$, then $M_\infty = 0$ a.s..*

Proof. We have, for all n and a.s.,

$$\mathbb{E}(M_{n+1}|\mathcal{F}_n) = \mathbb{E}(M_n X_{n+1}|\mathcal{F}_n) = M_n \mathbb{E}(X_{n+1}|\mathcal{F}_n) = M_n \mathbb{E}(X_{n+1}) = M_n.$$

So $(M_n)_{n\geq 0}$ is a martingale. Since $M_n \geq 0$, $(M_n)_{n\geq 0}$ is bounded in L^1, so converges a.s. by the a.s. martingale convergence theorem.

Set $Y_n = \sqrt{X_n}/a_n$ and $N_n = Y_1 Y_2 \dots Y_n$, then $(N_n)_{n\geq 0}$ is a martingale just as $(M_n)_{n\geq 0}$ is. Note that $M_n \leq N_n^2$ for all n.

Suppose that $\prod_n a_n > 0$ then

$$\mathbb{E}(N_n^2) = (a_1 a_2 \dots a_n)^{-2} \leq (\prod_n a_n)^{-2} < \infty.$$

so $(N_n)_{n\geq 0}$ is bounded in L^2. Hence by Doob's L^2-inequality,

$$\mathbb{E}\left(\sup_n M_n\right) \leq \mathbb{E}\left(\sup_n N_n^2\right) \leq 4 \sup_n \mathbb{E}(N_n^2) < \infty.$$

Hence $M_n \to M_\infty$ in L^1, by dominated convergence.

On the other hand, we know that N_n converges a.s. by the a.s. martingale convergence theorem. So if $\prod_n a_n = 0$ we must have also $M_\infty = 0$ a.s.. □

Corollary 13.2.4. *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$. Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables. Assume that, under $\mathbb{P}$ (respectively $\tilde{\mathbb{P}}$), the random variables X_n are independent and X_n has law μ_n (respectively $\tilde{\mu}_n$) for all n. Suppose that $\tilde{\mu}_n = f_n \mu_n$ for all n. Define the* likelihood ratio

$$L_n = \prod_{i=1}^{n} f_i(X_i).$$

Then, under $\mathbb{P}$,

(a) *if $\prod_n \int_{\mathbb{R}} \sqrt{f_n} d\mu_n > 0$, then L_n converges a.s. and in L^1,*
(b) *if $\prod_n \int_{\mathbb{R}} \sqrt{f_n} d\mu_n = 0$, then $L_n \to 0$ a.s..*

In particular, if $\mu_n = \mu$ and $\tilde{\mu}_n = \tilde{\mu}$ for all n, with $\mu \neq \tilde{\mu}$, then

$$\mathbb{P}(L_n \to 0) = 1, \quad \tilde{\mathbb{P}}(L_n \to \infty) = 1.$$

13.3. **Markov chains.** Let E be a countable set. We identify each probability measure λ on E with the row vector $(\lambda_i : i \in E)$, where $\lambda_i = \lambda(\{i\})$. We identify each function f on E with the column vector $(f_i : i \in E)$, where $f_i = f(i)$. A matrix $P = (p_{ij} : i, j \in E)$ is called *stochastic* if each row $(p_{ij} : j \in E)$ is a probability measure.

We suppose given a filtration $(\mathcal{F}_n)_{n \geq 0}$. Let $(X_n)_{n \geq 0}$ be an adapted process 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 $i, j \in E$ and all $A \in \mathcal{F}_n$ with $A \subseteq \{X_n = i\}$,

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

Our notion of Markov chain depends on a choice of filtration. When it is necessary to make this explicit, we shall refer to an $(\mathcal{F}_n)_{n \geq 0}$-Markov chain. The following result shows that our definition agrees with the usual one for the most obvious choice of filtration.

Proposition 13.3.1. *Let $(X_n)_{n \geq 0}$ be a 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 λ and transition matrix P,*
(b) *for all n and all $i_0, i_1, \dots, i_n \in E$,*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}.$$

We introduce some notation. 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 $\mathcal{E}^* = \sigma(X_k : k \geq 0)$.

Proposition 13.3.2. *Let P be a stochastic matrix. Then, for each $i \in E$, there is a unique probability measure $\mathbb{P}^i$ on $(E^*, \mathcal{E}^*)$ such that $(X_n)_{n \geq 0}$ is a Markov chain with transition matrix P and starting from i.*

Proposition 13.3.3. *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*

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

Proposition 13.3.4 (Strong Markov property)**.** *Let $(X_n)_{n\geq 0}$ be an $(\mathcal{F}_n)_{n\geq 0}$-Markov chain with transition matrix P and let T be a bounded stopping time. Set $\tilde{X}_n = X_{T+n}$ and $\tilde{\mathcal{F}}_n = \mathcal{F}_{T+n}$. Then $(\tilde{X}_n)_{n\geq 0}$ is a $(\tilde{\mathcal{F}}_n)_{n\geq 0}$-Markov chain with transition matrix P.*

Suitably reformulated, a version of the strong Markov property holds for all stopping times T, finite or infinite. Let us partition E into two disjoint sets D and ∂D. Set $T = \inf\{n \geq 0 : X_n \in \partial D\}$. Suppose we are given non-negative functions g on D and f on ∂D and define ϕ on E by

$$\phi_i = \mathbb{E}^i\left(\sum_{0\leq n<T} g(X_n) + f(X_T)1_{T<\infty}\right)$$

One can interpret ϕ as the expected cost incurred by $(X_n)_{n\geq 0}$, where cost g_i is incurred on each visit to i before T and cost f_i is incurred on arrival at $i \in \partial D$. In particular, if $f \equiv 0$ and $g = 1_A$ with $A \subseteq D$, then ϕ_i is the expected time spent in A, starting from i, before hitting ∂D. On the other hand, if $g \equiv 0$ and $f = 1_B$ with $B \subseteq \partial D$, then ϕ_i is the probability, starting from i, of entering ∂D through B.

Proposition 13.3.5. *We have*

(a)

$$\begin{cases} \phi = P\phi + g & \text{in } D \\ \phi = f & \text{in } \partial D, \end{cases} \tag{13.1}$$

(b) *if $\psi = (\psi_i : i \in E)$ satisfies*

$$\begin{cases} \psi \geq P\psi + g & \text{in } D \\ \psi \geq f & \text{in } \partial D \end{cases} \tag{13.2}$$

and $\psi_i \geq 0$ for all i, then $\psi_i \geq \phi_i$ for all i,

(c) *if $\mathbb{P}^i(T < \infty) = 1$ for all i, then* (13.1) *has at most one bounded solution.*

13.4. **Stochastic optimal control.** We consider here in a simple context an idea of much wider applicability. Let E be a countable set and let $B \subseteq E$. Suppose we are given an adapted process $(X_n)_{n\geq 0}$ in E, a *pay-off function* $f : B \to [0,\infty)$ and a family of probability measures $(\mathbb{P}_u : u \in U)$. Each $u \in U$ is called a *control.* Set $T = \inf\{n \geq 0 : X_n \in B\}$. Assume that $\mathbb{P}_u(T < \infty) = 1$ for all $u \in U$ and that the distribution of X_0 is the same for all $u \in U$. Consider the following optimization problem:

$$\text{maximize} \quad \mathbb{E}_u(f(X_T)).$$

Proposition 13.4.1 (Bellman's optimality principle)**.** *Suppose we can find a bounded function $V : E \to [0, \infty)$ and a control u^* such that*

(i) *$V = f$ on B,*
(ii) *$M_n = V(X_n^T)$ is a $\mathbb{P}_{u^*}$-martingale,*
(iii) *M_n is a $\mathbb{P}_u$-supermartingale for all $u \in U$.*

Then u^* is optimal and $\mathbb{E}_{u^*}(f(X_T)) = \mathbb{E}(V(X_0))$.

An important case of the set-up we have just considered arises when we are given a family of stochastic matrices $(P(a) : a \in A)$. Let $U = \{u : E \to A\}$ and define $\tilde{P}(u)$ by $\tilde{p}_{ij}(u) = p_{ij}(u(i))$. By Proposition 13.3.2, we can construct on the canonical space $(E^*, \mathcal{E}^*)$, for each $i \in E$ and each $u \in U$, a probability measure $\mathbb{P}_u^i$ making $(X_n)_{n\geq 0}$ a Markov chain with transition matrix $\tilde{P}(u)$ and starting from i. In this case, in order to check conditions (ii) and (iii) of Bellman's optimality principle, it suffices to show that

$$V_i \geq \sum_{j \in E} p_{ij}(u(i))V_j, \quad i \in E \setminus B$$

for all $u \in U$, with equality when $u = u^*$.

14. Continuous-time random processes

14.1. **Definitions.** We may apply to an *any* subset I of $\mathbb{R}$ all of the definitions made in §12.1. However, when I is uncountable, a process $(X_t)_{t\in I}$ can be rather a flaky object, unless we impose some additional regularity condition on the dependence of X_t on t. For statements which depend on the values of X_t for uncountably many t are not in general measurable – for example the statement 'X does not enter the set A.' In the following definitions we take $I = [0, T]$, for some $T > 0$, or $I = [0, \infty)$, and take E to be a topological space. We say that a process X in E is *continuous* (respectively *right-continuous*) if $t \mapsto X_t(\omega) : I \to E$ is continuous (respectively right-continuous) for all ω. We say that X has *left limits* if $\lim_{s\uparrow t, s\in I} X_s(\omega)$ exists in E, for all $t \in I$, for all ω. A right continuous process with left limits is called *cadlag* (continu à droite, limité à gauche). For cadlag processes, the whole process can be determined by its restriction to a countable dense set of times, so the measurability problems raised above go away. Except in the next section, all the continuous-time processes we consider will be at least cadlag.

A continuous process X can be considered as a single random variable

$$\omega \mapsto (X_t(\omega))_{t\in I} : \Omega \to C(I, E),$$

where $C(I,E)$ is the space of continuous functions $x : I \to E$, with the σ-algebra generated by its coordinate functions $x_t : C(I,E) \to E, t \in I$, where $x_t(x) = x(t)$. The same remark applies to any cadlag process, provided we replace $C(I,E)$ by $D(I,E)$, the space of cadlag functions $x : I \to E$, with the corresponding σ-algebra. Thus, each continuous (respectively cadlag) process X has a law which is a probability measure μ_X on $C(I,E)$ (respectively $D(I,E)$).

Given a probability measure μ on $D(I,E)$, to each finite set $J \subseteq I$, there corresponds a probability measure μ^J on E^J, which is the law of $(x_t : t \in J)$ under μ. The probability measures μ^J are called the *finite-dimensional distributions of* μ. When $\mu = \mu_X$, they are called the *finite-dimensional distributions of* X. By a π-system uniqueness argument, μ is uniquely determined by its finite-dimensional distributions. So, when we want to specify the law of a cadlag process, it suffices to describe its finite-dimensional distributions. Of course we have no *a priori* reason to believe there exists a cadlag process whose finite-dimensional distributions coincide with a given family of measures $(\mu^J : J \subseteq I, J \text{ finite})$.

Let X be a process in $\mathbb{R}^n$. We say that X is *Gaussian* if each of its finite-dimensional distributions is Gaussian. Since any Gaussian distribution is determined by its mean and covariance, it follows that the law of a continuous Gaussian process is determined once we specify $\mathbb{E}(X_t)$ and $\mathrm{cov}(X_s, X_t)$ for all $s, t \in I$.

14.2. **Path regularization.** Given two processes X and $\tilde{X}$, we say that $\tilde{X}$ is a *version* of X if $\tilde{X}_t = X_t$ a.s., for all $t \in I$. In this section we present two results which provide criteria for a process X to possess a version $\tilde{X}$ which is continuous or cadlag. Recall that $\mathbb{D}$ denotes the set of dyadic rationals.

Theorem 14.2.1 (Kolmogorov's criterion)**.** *Let $p \geq 1$ and $\beta > 1/p$. Let $I = \mathbb{D} \cap [0,1]$. Suppose $X = (X_t)_{t \in I}$ is a process such that*

$$\|X_s - X_t\|_p \leq C|s-t|^\beta, \quad \textit{for all } s, t \in I$$

for some constant $C < \infty$. Then, for all $\alpha \in [0, \beta - (1/p))$, there exists a random variable $K_\alpha \in L^p$ such that

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

Proof. Let D_n denote the set of integer multiples of 2^{-n} in $[0,1)$. Set

$$K_n = \sup_{t \in D_n} |X_{t+2^{-n}} - X_t|.$$

Then

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

For $s, t \in I$ 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

$$|X_t - X_s| \leq 2 \sum_{n \geq m+1} K_n$$

and so

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

where $K_\alpha = 2\sum_{n\geq 0} 2^{n\alpha} K_n$. But

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

□

Theorem 14.2.2 (Path regularization). *Let $X = (X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-martingale. Set $\tilde{\mathcal{F}}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N})$, where $\mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$ and $\mathcal{N} = \{A \in \mathcal{F} : \mathbb{P}(A) = 0\}$. Then there exists a cadlag $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-martingale $\tilde{X}$ such that*

$$\mathbb{E}(\tilde{X}_t|\mathcal{F}_t) = X_t \quad a.s.$$

In particular, if $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, then $\tilde{X}$ is a version of X.

Proof. Since completion of filtrations preserves the martingale property, we may assume that $\mathcal{N} \subseteq \mathcal{F}_0$ from the outset. Set $I_N = \mathbb{Q} \cap [0, N]$ and let $a < b$. By Doob's upcrossing and maximal inequalities, $U([a, b], I_N)$ and $X^*(I_N)$ are a.s. finite for all $N \in \mathbb{N}$. Hence $\mathbb{P}(\Omega_0) = 1$, where

$$\Omega_0 = \cap_{N\in\mathbb{N}} \cap_{a,b\in\mathbb{Q},a<b} \{U([a, b], I_N) < \infty\} \cap \{X^*(I_N) < \infty\}.$$

For $\omega \in \Omega_0$ the following limits exist in $\mathbb{R}$:

$$X_{t+}(\omega) = \lim_{s\downarrow t, s\in\mathbb{Q}} X_s(\omega), \quad t \geq 0$$
$$X_{t-}(\omega) = \lim_{s\uparrow t, s\in\mathbb{Q}} X_s(\omega), \quad t > 0.$$

Define, for $t \geq 0$,

$$\tilde{X}_t = \begin{cases} X_{t+} & \text{on } \Omega_0, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\tilde{X}$ is cadlag and $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-adapted. By the backward martingale convergence theorem, for any $s > t$,

$$\tilde{X}_t = \mathbb{E}(X_s|\tilde{\mathcal{F}}_t), \quad \text{a.s.}$$

The remaining conclusions follow easily. □

The σ-algebra $(\tilde{\mathcal{F}}_t)_{t\geq 0}$ satisfies the *usual conditions* of *right continuity* and *completeness*:

$$\tilde{\mathcal{F}}_{t+} = \tilde{\mathcal{F}}_t, \quad \mathcal{N} \subseteq \tilde{\mathcal{F}}_t, \quad \text{for all } t.$$

The path regularization theorem shows that, when $I = [0, \infty)$, we do not lose much in restricting our attention to cadlag martingales and filtrations satisfying the usual conditions.

14.3. **Martingales in continuous time.** The following four results for a cadlag process $(X_t)_{t\geq 0}$ are immediate consequences of the corresponding results for the process $(X_t)_{t\in I}$ obtained by restricting $(X_t)_{t\geq 0}$ to the countable index set $I = \mathbb{Q}\cap[0,\infty)$.

Theorem 14.3.1 (Doob's maximal inequality). *Let $(X_t)_{t\geq 0}$ 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 14.3.2 (Doob's L^p-inequality). *Let $(X_t)_{t\geq 0}$ be a cadlag martingale or non-negative submartingale. Then, for all $p > 1$ and $q = p/(p-1)$,*

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

Theorem 14.3.3 (Almost sure martingale convergence theorem). *Let $(X_t)_{t\geq 0}$ be a cadlag martingale which is bounded in L^1. Then $X_t \to X_\infty$ a.s. for some $X_\infty \in L^1(\mathcal{F}_\infty)$.*

Denote by $\mathcal{M}^1[0,\infty)$ the set of uniformly integrable cadlag martingales $(X_t)_{t\geq 0}$ and, for $p > 1$, by $\mathcal{M}^p[0,\infty)$ the set of cadlag martingales which are bounded in L^p.

Theorem 14.3.4 (L^p martingale convergence theorem). *Let $p \in [1, \infty)$.*

(a) *Suppose $(X_t)_{t\geq 0} \in \mathcal{M}^p[0,\infty)$. Then $X_t \to X_\infty$ as $t \to \infty$, a.s. and in L^p, for some $X_\infty \in L^p(\mathcal{F}_\infty)$. Moreover, $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$ a.s. for all t.*

(b) *Assume that $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions. Suppose $Y \in L^p(\mathcal{F}_\infty)$. Then there exists $(X_t)_{t\geq 0} \in \mathcal{M}^p[0,\infty)$ such that $X_t = \mathbb{E}(Y|\mathcal{F}_t)$. Moreover $X_t \to Y$ as $t \to \infty$ a.s. and in L^p.*

Thus, when $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions, the map $(X_t)_{t\geq 0} \to X_\infty$ is a one-to-one correspondence between $\mathcal{M}^p[0,\infty)$ and $L^p(\mathcal{F}_\infty)$.

We recall the following definitions from §12.2. 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 define

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

For a cadlag process X, we set $X_T(\omega) = X_{T(\omega)}(\omega)$ whenever $T(\omega) < \infty$. We also define the *stopped process* X^T by $X^T_t = X_{T\wedge t}$.

Proposition 14.3.5. *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) *if $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$,*
(c) *$X_T 1_{T<\infty}$ is an $\mathcal{F}_T$-measurable random variable,*

(d) X^T *is adapted.*

Theorem 14.3.6 (Optional stopping theorem). *Let X be a cadlag adapted 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*
$$\mathbb{E}(X_T|\mathcal{F}_S) = X_{S\wedge T} \quad a.s.,$$
(c) *for all stopping times T, 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 UI, then (b) *and* (d) *hold for* all *stopping times T.*

Proof. Suppose (a) holds. Let S and T be stopping times, with T bounded, $T \le t$ say. Let $A \in \mathcal{F}_S$. For $n \ge 0$, 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 $(X_t)_{t\ge 0}$ is right continuous, $X_{T_n} \to X_T$ a.s. as $n \to \infty$. By the discrete-time optional stopping theorem, $X_{T_n} = \mathbb{E}(X_{t+1}|\mathcal{F}_{T_n})$ so $(X_{T_n} : n \ge 0)$ is UI 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 the discrete-time optional stopping theorem 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. □

15. Weak convergence

15.1. **Definitions.** Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on a metric space S. We say that μ_n converges *weakly* to μ and write $\mu_n \Rightarrow \mu$ if $\mu_n(f) \to \mu(f)$ for all bounded continuous functions f on S.

There are a number of equivalent characterizations of weak convergence:

Theorem 15.1.1. *The following are equivalent:*

(a) $\mu_n \Rightarrow \mu$,
(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 the same type for the case $S = \mathbb{R}$.

Theorem 15.1.2. *Let $\mu_n, n \in \mathbb{N}$, and μ be probability measures on $\mathbb{R}$. Denote by F_n and F the corresponding distribution functions. The following are equivalent:*

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

Proof. Suppose $\mu_n \Rightarrow \mu$. Fix $x \in \mathbb{R}$ with $F(x-) = F(x)$. Given $\varepsilon > 0$, choose $\delta > 0$ so that $F(x-\delta) \geq F(x) - \varepsilon$ and $F(x+\delta) \leq F(x) + \varepsilon$. For some continuous functions f and g,

$$1_{(-\infty, x-\delta]} \leq f \leq 1_{(-\infty,x]} \leq g \leq 1_{(-\infty,x+\delta]}.$$

Then $\mu_n(f) \leq F_n(x) \leq \mu_n(g)$ for all n. Also $\mu(f) \geq F(x) - \varepsilon$ and $\mu(g) \leq F(x) + \varepsilon$. Hence $\liminf_n F_n(x) \geq F(x) - \varepsilon$ and $\limsup_n F_n(x) \leq F(x) + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this proves (b).

Suppose now that (b) holds. We use the construction of random variables discussed in §2.3. (It is this which makes the case $S = \mathbb{R}$ relatively straightforward.) Take $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1], \mathcal{B}((0,1]), dx)$ and set

$$X_n(\omega) = \inf\{x : \omega \leq F_n(x)\}, \quad X(\omega) = \inf\{x : \omega \leq F(x)\}.$$

Then X_n has law μ_n and X has law μ. For any a with $F(a-) = F(a)$ and any ω such that $X(\omega) > a$, we have $\omega > F(a)$ so $\omega > F_n(a)$ eventually and so $X_n(\omega) > a$ eventually. Since F has at most countably many points of discontinuity, the set of such a is dense and so $\liminf_n X_n(\omega) \geq X(\omega)$ for all ω. Now let $\hat{F}$ denote the distribution function of $-X$ and set

$$\hat{X}(\omega) = \inf\{x : 1 - \omega \leq \hat{F}(x)\} = -\sup\{y : \omega \geq F(y-)\}.$$

Define similarly $\hat{X}_n$. Then $\liminf_n \hat{X}_n(\omega) \geq \hat{X}(\omega)$ for all ω. Note that $F(y-) \leq \omega \leq F(x)$ for some ω implies $y \leq x$. Hence $-\hat{X}(\omega) \leq X(\omega)$ for all ω. But X and $-\hat{X}$ have the same distribution, so we must have $-\hat{X} = X$ a.s. and similarly $-\hat{X}_n = X_n$ a.s., for all n. Hence $\limsup_n X_n(\omega) \leq X(\omega)$ a.s.. We have shown that (b) implies (c).

Finally, if (c) holds, then (a) follows by bounded convergence. □

15.2. **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 $\mu_n(K^c) \leq \varepsilon$ for all n.

Theorem 15.2.1 (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* $\mu_{n_k} \Rightarrow \mu$.

Proof for the case $S = \mathbb{R}$. By a diagonal argument and by passing to a subsequence, it suffices to consider the case where the corresponding distribution functions F_n converge pointwise on $\mathbb{Q}$, with limit G, say. Then $G : \mathbb{Q} \to [0,1]$ must be increasing, so must have an increasing extension G to $\mathbb{R}$, with 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 increasing 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 N such that

$$F_n(-N) \leq \varepsilon, \quad F_n(N) \geq 1 - \varepsilon, \quad \text{for all } n.$$

It follows that

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$$

so F is a distribution function. The result now follows from Theorem 15.1.2. □

15.3. **Weak convergence and characteristic functions.** Recall that, 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 15.3.1. *Let μ be a probability measure on $\mathbb{R}$ with characteristic function ϕ. Then*

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

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

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

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

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

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

On integrating this inequality with respect to μ, we obtain our result. □

Theorem 15.3.2. *Let $\mu_n, n \in \mathbb{N}$, and μ be probability measures on $\mathbb{R}^d$, having characteristic functions ϕ_n and ϕ respectively. Then the following are equivalent:*

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

Proof for $d = 1$. It is trivial that (a) implies (b). Assume 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. But if $\mu_{n_k}\Rightarrow\nu$ then $\phi_{n_k}(u)\to\psi(u)$ for all u, where ψ is the characteristic function of ν. Hence $\psi=\phi$ and so $\nu=\mu$, by uniqueness of characteristic functions. It follows that $\mu_n\Rightarrow\mu$. □

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

Theorem 15.3.3 (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 $\mu_n\Rightarrow\mu$ as $n\to\infty$.*

16. Brownian motion

16.1. **Wiener's theorem.** Let $B=(B_t)_{t\geq0}$ be a continuous process in $\mathbb{R}^n$. We say that B is a *Brownian motion in* $\mathbb{R}^n$ if

(i) $B_t-B_s\sim N(0,(t-s)I)$, for all $s<t$,
(ii) B has independent increments, independent of B_0.

In the case $n=1$, or if is already established that B takes values in $\mathbb{R}^n$ for some $n\geq2$, we say simply that B is a *Brownian motion.* It is easy to check that B is a Brownian motion in $\mathbb{R}^n$ if and only if the components of $(B_t-B_0)_{t\geq0}$ are independent Brownian motions, starting from 0 and independent of B_0.

Let $W=C([0,\infty),\mathbb{R})$ and define for $t\geq0$ the coordinate function $X_t : W\to\mathbb{R}$ by $X_t(x)=x(t)$. Set $\mathcal{W}=\sigma(X_t : t\geq0)$.

Theorem 16.1.1 (Wiener's theorem)**.** *There exists a unique probability measure μ on $(W,\mathcal{W})$ such that $(X_t)_{t\geq0}$ is a Brownian motion starting from* 0.

The measure μ is called *Wiener measure.*

Proof. Conditions (i) and (ii) determine the finite dimensional distributions of any such measure μ, so there can be at most one. To show there is exactly one it will suffice to construct a Brownian motion on some probability space.

For $n\geq0$ denote by D_n the set of integer multiples of 2^{-n} in $[0,\infty)$ and denote by D the union of these sets. Then D is countable so, by an argument given in §2.4, we know there exists, on some probability space, a family of independent $N(0,1)$ random variables $(Y_t : t\in D)$. Let us say that a process $(B_t)_{t\in D_n}$ is a Brownian motion if conditions (i) and (ii) hold on D_n. For $t \in D_0 = \mathbb{Z}^+$, set $B_t = Y_1 + \cdots + Y_t$. Then $(B_t)_{t \in D_0}$ is a Brownian motion.

Suppose, inductively for $n \geq 1$, that we have constructed a Brownian motion $(B_t)_{t \in D_{n-1}}$. For $t \in D_n \setminus D_{n-1}$, set $r = t - 2^{-n}$ and $s = t + 2^{-n}$ so that $r, s \in D_{n-1}$ and define

$$Z_t = 2^{-(n+1)/2} Y_t, \quad B_t = \tfrac{1}{2}(B_r + B_s) + Z_t.$$

We then have two new increments:

$$B_t - B_r = \tfrac{1}{2}(B_s - B_r) + Z_t, \quad B_s - B_t = \tfrac{1}{2}(B_s - B_r) - Z_t.$$

We compute

$$\mathbb{E}[(B_t - B_r)^2] = \mathbb{E}[(B_s - B_t)^2] = \tfrac{1}{4} 2^{-(n-1)} + 2^{-(n+1)} = 2^{-n},$$

$$\mathbb{E}[(B_t - B_r)(B_s - B_t)] = \tfrac{1}{4} 2^{-(n-1)} - 2^{-(n+1)} = 0.$$

The two new increments, being Gaussian, are therefore independent and have the required variance. Moreover, being constructed from $B_s - B_r$ and Y_t, they are independent of increments over intervals disjoint from (r, s). Hence $(B_t)_{t \in D_n}$ is a Brownian motion. By induction, we obtain a process $(B_t)_{t \in D}$, having independent increments and such that, for $s < t$, we have $B_t - B_s \sim N(0, t - s)$. In particular, for $p \in [1, \infty)$,

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

where $C_p = \mathbb{E}(|B_1|^p) < \infty$. Hence, by Kolmogorov's criterion, there is a continuous process $(\tilde{B}_t)_{t \geq 0}$ such that $\tilde{B}_t = B_t$ for all $t \in D$ a.s.. (Moreover, since p can be chosen arbitrarily large, we can choose $(\tilde{B}_t)_{t \geq 0}$ so that $t \mapsto \tilde{B}_t$ is Hölder continuous of exponent α, for every $\alpha < 1/2$.)

It remains to show that $(\tilde{B}_t)_{t \geq 0}$ is a Brownian motion. Write $p(t, .)$ for the density function of a Gaussian of mean 0 and variance t. Given $0 < t_1 < \cdots < t_n$, we can find sequences $(t_k^m)_{m \in \mathbb{N}}$ in D such that $0 < t_1^m < \cdots < t_n^m$ for all m and $t_k^m \to t_k$ for all k. Set $t_0 = t_0^m = 0$. Then, for all continuous bounded functions f, by continuity of $(\tilde{B}_t)_{t \geq 0}$ and bounded convergence,

$$\begin{aligned}
\mathbb{E}(f(\tilde{B}_{t_1} - \tilde{B}_{t_0}, \ldots, \tilde{B}_{t_n} - \tilde{B}_{t_{n-1}})) &= \lim_{m \to \infty} \mathbb{E}(f(B_{t_1^m} - B_{t_0^m}, \ldots, B_{t_n^m} - B_{t_{n-1}^m})) \\
= \lim_{m \to \infty} \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) \prod_{k=1}^n p(t_k^m - t_{k-1}^m, x_k) dx_k \\
= \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) \prod_{k=1}^n p(t_k - t_{k-1}, x_k) dx_k
\end{aligned}$$

This shows that $(\tilde{B}_t)_{t \geq 0}$ has the required finite-dimensional distributions and so is a Brownian motion. $\square$

16.2. **Invariance properties.**

Proposition 16.2.1. *Let B be a continuous process. Then the following are equivalent:*

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

Proposition 16.2.2. *Let B be a Brownian motion starting from 0. Then so are the following processes:*

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

where in (d) *the process is defined to take the value* 0 *when* $t = 0$.

Part (c) is called the *scaling* property of Brownian motion. Part (a) generalizes to the following *rotational invariance* property of Brownian motion in $\mathbb{R}^n$.

Proposition 16.2.3. *Let $U \in O(n)$. If $(B_t)_{t \geq 0}$ is a Brownian motion in $\mathbb{R}^n$, then so is $(UB_t)_{t \geq 0}$.*

16.3. **Martingales.** There are many martingales associated with Brownian motion and these provide a useful tool for its study. For example, if B is a Brownian motion starting from 0, then you can easily check that both $(B_t)_{t \geq 0}$ and $(B_t^2 - t)_{t \geq 0}$ are martingales starting from 0. This fact is useful for the proof of Proposition 16.5.1. We begin this section with a discussion of the relationship between filtrations and Brownian motion. Then we will give a characterization of Brownian motion by means of exponential martingales, which will lead to the strong Markov property. Finally we shall give a general theorem for constructing martingales from Brownian motion, which will be used in our discussion of the relationship between Brownian motion and the Dirichlet problem.

For any process X, we set

$$\mathcal{F}_t^X = \sigma(X_s : s \leq t), \quad \mathcal{F}_{t\infty}^X = \sigma(X_s - X_t : s > t).$$

Let B be a Brownian motion in $\mathbb{R}^n$ and let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration. We say that B and $(\mathcal{F}_t)_{t \geq 0}$ are *compatible* or that *B is an $(\mathcal{F}_t)_{t \geq 0}$-Brownian motion* if

(i) $\mathcal{F}_t^B \subseteq \mathcal{F}_t$ for all t (B is adapted),
(ii) $\mathcal{F}_{t\infty}^B$ and $\mathcal{F}_t$ are independent for all t.

Obviously, these two properties are satisfied if $\mathcal{F}_t = \mathcal{F}_t^B$ for all t. More generally, if B is a process in $\mathbb{R}^n$ defined on $\Omega_0 \in \mathcal{F}_0$, then we say that B is an $(\mathcal{F}_t)_{t \geq 0}$-*Brownian motion defined on* Ω_0 if B is an $(\tilde{\mathcal{F}}_t)_{t \geq 0}$-Brownian motion under $\tilde{\mathbb{P}}$, where

$$\tilde{\mathcal{F}} = \{A \in \mathcal{F} : A \subseteq \Omega_0\}, \quad \tilde{\mathcal{F}}_t = \{A \in \mathcal{F}_t : A \subseteq \Omega_0\}$$

and
$$\tilde{\mathbb{P}}(A) = \mathbb{P}(A)/\mathbb{P}(\Omega_0), \quad A \in \tilde{\mathcal{F}}.$$

Proposition 16.3.1. *Let B be a continuous process in $\mathbb{R}^n$. Define for $u \in \mathbb{R}^n$*
$$Z_t^u = \exp(i\langle u, B_t\rangle + |u|^2 t/2).$$
Then the following are equivalent:

(a) *B is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion,*
(b) *Z^u is an $(\mathcal{F}_t)_{t\geq 0}$-martingale for all $u \in \mathbb{R}^n$.*

Proposition 16.3.2. *Let $\Omega_0 = \cup_{n\geq 1}\Omega_n$ with $\Omega_n \in \mathcal{F}_0$ for all n and let B be a process in $\mathbb{R}^n$, defined on Ω_0. Then B is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion defined on Ω_0 if and only if $B|_{\Omega_n}$ is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion defined on Ω_n for all n.*

Proof. Note that B is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion if and only if
$$\mathbb{E}(e^{iu(B_t - B_s)}1_A) = e^{-u^2(t-s)/2}\mathbb{P}(A)$$
for all $A \in \mathcal{F}_s$, for all $s \leq t$. Similarly, B is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion defined on Ω_0 if and only if
$$\mathbb{E}(e^{iu(B_t - B_s)}1_A) = e^{-u^2(t-s)/2}\mathbb{P}(A)$$
for all $A \in \mathcal{F}_s$ with $A \subseteq \Omega_0$, for all $s \leq t$. So, if B is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion defined on Ω_n for all n, we have
$$\mathbb{E}(e^{iu(B_t - B_s)}1_A) = e^{-u^2(t-s)/2}\mathbb{P}(A)$$
for all $A \in \mathcal{F}_s$ with $A \subseteq \Omega_n$ for some n, for all $s \leq t$. If $\cup_n\Omega_n = \Omega_0$, a simple dominated convergence argument extends the identity to all $A \subseteq \Omega_0$, as required. $\square$

Theorem 16.3.3. *Let B be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion in $\mathbb{R}^n$ and let $f \in C^{1,2}([0,\infty)\times\mathbb{R}^n)$ with all derivatives having no more than exponential growth on $\mathbb{R}^n$, uniformly on compacts in $[0,\infty)$. Set*
$$M_t^f = f(t, B_t) - f(0, B_0) - \int_0^t \left(\frac{\partial}{\partial s} + \frac{1}{2}\Delta\right) f(s, B_s)ds.$$
Then M^f is an $(\mathcal{F}_t)_{t\geq 0}$-martingale.

Proof. Write $M = M^f$. Let $s, t \geq 0$. Our assumptions on f allow us to show that $M_{s+t} - M_s$ is integrable, with $\mathbb{E}|m_{s+t} - M_s| \to 0$ as $t \downarrow 0$. We have to show that $\mathbb{E}(M_{s+t} - M_s|\mathcal{F}_s) = 0$ a.s.. Now
$$\begin{aligned} M_{s+t} - M_s &= f(s+t, B_{s+t}) - f(s, B_s) - \int_s^{s+t} \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) f(r, B_r)dr \\ &= \tilde{f}(t, \tilde{B}_t) - \tilde{f}(0, \tilde{B}_0) - \int_0^t \left(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\right) \tilde{f}(r, \tilde{B}_r)dr \end{aligned}$$

where $\tilde{f}(t,x) = f(s+t,x)$ and $\tilde{B}_t = B_{s+t}$. Note that $\tilde{B}$ is an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion starting from B_s, where $\tilde{\mathcal{F}}_t = \mathcal{F}_{s+t}$. Hence it will suffice to show that $\mathbb{E}(M_t|\mathcal{F}_0) = 0$, a.s.. Now $\mathbb{E}(M_t|\mathcal{F}_0) = m(B_0)$ a.s., where $m(x) = \mathbb{E}^x(M_t)$ and the superscript x specifies the case $B_0 = x$. So we just have to show that $m(x) = 0$ for all $x \in \mathbb{R}^n$.

As we noted above, $\mathbb{E}(M_s) \to 0$ as $s \downarrow 0$. Hence it will suffice to show that $\mathbb{E}^x(M_t - M_s) = 0$ for all $x \in \mathbb{R}^n$ and all $0 < s < t$. We compute

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

Now p satifies the heat equation

$$\left(\frac{\partial}{\partial t} - \frac{1}{2}\Delta_y\right) p(t,x,y) = 0.$$

We integrate by parts twice in $\mathbb{R}^n$ to obtain

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

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

16.4. **Strong Markov property.**

Theorem 16.4.1 (Strong Markov property). *Let B be an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion in $\mathbb{R}^n$ and let T be a stopping time. Set $\tilde{\mathcal{F}}_t = \mathcal{F}_{T+t}$ and define $\tilde{B}_t = B_{T+t}$ on $\{T < \infty\}$. Then $\tilde{B}$ is an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion defined on $\{T < \infty\}$.*

Proof. By Proposition 16.3.2, it suffices to show that $\tilde{B}$ is an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion defined on $\{T \leq n\}$ for all $n \in \mathbb{N}$. For each n, this property of B is unaltered if we replace T by $T \wedge n$. So we may assume without loss that T is bounded.

Define for $u \in \mathbb{R}^n$

$$\tilde{Z}_t^u = \exp(i\langle u, \tilde{B}_t\rangle + |u|^2 t/2).$$

Then $\tilde{Z}^u$ is integrable, $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-adapted and $\tilde{Z}_t^u = Z_{T+t}^u \exp(-|u|^2 T/2)$, where Z^u is the exponential martingale from Proposition 16.3.1. Hence, for $A \in \tilde{\mathcal{F}}_s$ and $s < t$, by optional stopping,

$$\mathbb{E}(\tilde{Z}^u_t 1_A) = \mathbb{E}(Z^u_{T+t}\exp(-|u|^2T/2)1_A) = \mathbb{E}(Z^u_{T+s}\exp(-|u|^2T/2)1_A) = \mathbb{E}(\tilde{Z}^u_s 1_A).$$

Hence $\tilde{Z}^u$ is a martingale for all u, so $\tilde{B}$ is an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion by Proposition 16.3.1. □

Corollary 16.4.2 (Reflection principle)**.** *Let B be a Brownian motion starting from 0 and let $a>0$. Set $T=\inf\{t : B_t > a\}$ and define*

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

Then X is also a Brownian motion starting from 0.

Proof. Note that T is a stopping time and that $X_T = a$ on $\{T<\infty\}$. We will show more generally that, for any stopping time T, Y is a Brownian motion, where

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

It suffices to check that $(Y_t)_{t\leq n}$ is a Brownian motion for each $n\in\mathbb{N}$, so we may replace T by the bounded stopping time $T\wedge n$ as this leaves $(Y_t)_{t\leq n}$ unchanged. Assume then that T is bounded. By the strong Markov property, $(B_{T+t}-B_T)_{t\geq 0}$ is a Brownian motion starting from 0 and independent of $\mathcal{F}_T$. Hence so is $(-(B_{T+t}-B_T))_{t\geq 0}$. It follows that Y has the same distribution as B. □

16.5. **Hitting times.** Let B be a Brownian motion starting from 0. For $a\in\mathbb{R}$ we define the *hitting time*

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

Then H_a is a stopping time.

Proposition 16.5.1. *For $a,b>0$, we have*

$$\mathbb{P}(H_{-a} < H_b) = b/(a+b), \quad \mathbb{E}(H_{-a}\wedge H_b) = ab.$$

Proposition 16.5.2. *The hitting time H_a has a density function, given by*

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

16.6. **Sample path properties.**

Proposition 16.6.1. *Let B be a Brownian motion starting form 0. Then, almost surely,*

(a) $B_t/t \to 0$ *as* $t\to\infty$,
(b) $\sup_t B_t = -\inf_t B_t = \infty$,
(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 16.6.2. *Let B be a Brownian motion. Then, almost surely,*

(a) *$t \mapsto B_t$ is Hölder continuous of exponent α for all $\alpha < 1/2$,*
(b) *there is no interval (r, s) on which $t \mapsto B_t$ is Hölder continuous of exponent α for any $\alpha > 1/2$.*

Proof. For (a) we refer to the proof of Wiener's theorem 16.1.1. We turn to (b). We use the notation D and D_n from the proof of Wiener's theorem. Let $r, s \in D_N$ with $r < s$ and let $n \geq N$. Then, for $n \geq N$,

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{t\in D_n, r\leq t<s} (B_{t+2^{-n}} - B_t)^2 - (s-r)\right)^2\right] \\
&\qquad = \operatorname{var}\left(\sum_{t\in D_n, r\leq t<s} (B_{t+2^{-n}} - B_t)^2\right) \\
&\qquad = \sum_{t\in D_n, r\leq t<s} \operatorname{var}\left((B_{t+2^{-n}} - B_t)^2\right) = 2^n(s-r)2^{-2n}\operatorname{var}(B_1^2).
\end{aligned}$$

Now $\operatorname{var}(B_1^2) = 2 < \infty$ so

$$\sum_{t\in D_n, r\leq t<s} (B_{t+2^{-n}} - B_t)^2 \to (s-r)$$

in L^2 as $n \to \infty$.

On the other hand, if B is Hölder continuous of exponent α and constant K on $[r, s]$, then

$$\sum_{t\in D_n, r\leq t<s} (B_{t+2^{-n}} - B_t)^2 \leq \sup_{t\in D_n, r\leq t<s} |B_{t+2^{-n}} - B_t|^{2-1/\alpha} \sum_{t\in D_n, r\leq t<s} K^{1/\alpha}2^{-n} \to 0$$

almost surely, since $\sum_{t\in D_n, r\leq t<s} 2^{-n} = (s-r)$ and B is uniformly continuous on $[r, s]$. Hence, almost surely there is no such interval $[r, s]$. □

The preceding result shows in particular that almost surely there is no interval on which B is differentiable. In fact an even stronger result holds.

Theorem 16.6.3. *Almost all Brownian paths are nowhere differentiable.*

Proof. Let B be a Brownian motion. For $1 \leq k \leq n+2$, set $\Delta_{k,n} = |B_{(k-1)/n} - B_{k/n}|$ and consider, for $K > 0$, the event

$$A_n = A_n^K = \{\max\{\Delta_{k,n}, \Delta_{k+1,n}, \Delta_{k+2,n}\} \leq K/n \text{ for some } k = 1, \dots, n\}.$$

The density of $B_{1/n}$ is bounded by $\sqrt{n/2\pi}$ so

$$\mathbb{P}(\Delta_{k,n} \leq K/n) \leq C(K)/\sqrt{n}.$$

Hence, by independence of increments,

$$\mathbb{P}(A_n) \leq n\mathbb{P}(\Delta_{k,n} \leq K/n)^3 \leq C(K)/\sqrt{n}.$$

Consider now the event

$$G_N^K = \{\text{for some } s \in [0,1], |B_s - B_t| \le K|s-t| \\ \text{for all } t \in [0, 1+\frac{1}{N}] \text{ with } |s-t| \le \frac{1}{N}\}.$$

It is an elementary exercise to show that $G_N^K \subseteq A_n^{5K}$ for all $n \ge 3N$. Hence $\mathbb{P}(G_N^K) = 0$ for all N and K. But

$$\{\text{for some } t \in [0,1), s \mapsto B_s \text{ is differentiable at } t\} \subseteq \cup_{N\in\mathbb{N},K\in\mathbb{N}} G_N^K.$$

$\square$

Proposition 16.6.4 (Blumenthal's zero-one law). *Let B be a Brownian motion in $\mathbb{R}^n$ starting from 0. If $A \in \mathcal{F}_{0+}^B$ then $\mathbb{P}(A) \in \{0,1\}$.*

Proposition 16.6.5. *Let A be a non-empty open subset of the unit sphere in $\mathbb{R}^n$ and let $\varepsilon > 0$. Consider the cone*

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

Let B be a Brownian motion in $\mathbb{R}^n$ starting from 0 and let

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

Then $T_C = 0$ a.s..

16.7. **Recurrence and transience.**

Theorem 16.7.1. *Let B be a Brownian motion in $\mathbb{R}^n$.*

(a) *If $n = 1$, then*

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

(b) *If $n = 2$, then*

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

but, for any $\varepsilon > 0$,

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

(c) *If $n \ge 3$, 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 $\mathbb{R}^n, n \ge 3$, is *transient*.

Proof. For (a) we refer to Proposition 16.6.1(c). To prove (b) we fix $0 < a < 1 < b$ and consider the process $X_t = f(B_t)$, where $f \in C_b^2(\mathbb{R}^2)$ is such that

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

Note that $\Delta f(x) = 0$ for $a \le x \le b$. Consider the stopping time

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

By Theorem 16.3.3, M^f is a martingale. Hence, by optional stopping, $\mathbb{E}(M^f_T) = \mathbb{E}(M^f_0) = 0$. Assume for now that $|B_0| = 1$. Then $M^f_T = \log |B_T|$, so $p = p(a, b) = \mathbb{P}(|B_T| = a)$ satisfies

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

Consider first the limit $a \downarrow 0$ 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 $x = 0$ we have for all $\varepsilon > 0$, by the Markov property,

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

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

Consider now the limit $b \uparrow \infty$ 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, $\mathbb{P}(|B_t| < \varepsilon \text{ for some } t > n) = 1$ and hence that $\mathbb{P}(\{t \ge 0 : |B_t| < \varepsilon\} \text{ is unbounded}) = 1$.

We turn to the proof of (c). Since the first three components of a Brownian motion in $\mathbb{R}^n, n \ge 3$, form a Brownian motion in $\mathbb{R}^3$, it suffices to consider the case $n = 3$. 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 $(T_k : k \ge 0)$ by setting $S_0 = 0$ and, for $k \ge 0$,

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

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, $\mathbb{P}(T_1 < \infty) \le \mathbb{P}^N(T_1 < \infty) = p$ and for $k \ge 2$, $\mathbb{P}(T_k < \infty) = \mathbb{P}(T_1 < \infty)\mathbb{P}^N(T_{k-1} < \infty)$. Hence $\mathbb{P}(T_k < \infty) \le p^k$ and

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

as required. □

16.8. **Brownian motion and the Dirichlet problem.** Let D be a connected open set in $\mathbb{R}^n$ with smooth boundary ∂D and let $f : \partial D \to [0, \infty)$ and $g : D \to [0, \infty)$ be measurable functions. By a *solution to the Dirichlet problem* (*in D with data f and g*), we shall 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 $\ge$ (twice) in this definition we say that ψ is a *supersolution*.

We shall 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 16.8.1. *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 $\mathbb{R}^n$. 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 16.8.2. *We have*
(a)

$$\|P_t g\|_\infty \leq (1 \wedge (2\pi t)^{n/2}\,\mathrm{vol}(\mathrm{supp}\, g))\|g\|_\infty,$$

(b) *for $n \geq 3$,*

$$\|Gg\|_\infty \leq (1 + \mathrm{vol}(\mathrm{supp}\, g))\|g\|_\infty,$$

(c) *for $n \geq 3$ and for $g \in C^2(\mathbb{R}^n)$ of compact support, $Gg \in C_b^2(\mathbb{R}^n)$ and*

$$-\tfrac{1}{2}\Delta Gg = g.$$

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 $Gg \in C_b^2(\mathbb{R}^n)$.

To show that $-\frac{1}{2}\Delta Gg = g$, we fix $0 < s < t$ and write

$$Gg(x) = \mathbb{E}^0 \int_0^s g(x + B_r)dr + \int_s^t \int_{\mathbb{R}^n} p(r,x,y)g(y)dydr + \mathbb{E}^0 \int_t^\infty g(x + B_r)dr.$$

By differentiating under the integral we obtain

$$\tfrac{1}{2}\Delta Gg(x) = \tfrac{1}{2}\int_0^s \mathbb{E}^0 \Delta g(x+B_r)dr + \tfrac{1}{2}\int_s^t \int_{\mathbb{R}^n} \Delta_x p(r,x,y)g(y)dydr + \tfrac{1}{2}\int_t^\infty \mathbb{E}^0 \Delta g(x+B_r)dr.$$

By the estimate in (a), the first and third terms on the right tend to 0 as $s \downarrow 0$ and $t \uparrow \infty$. Since $\partial p/\partial t = \frac{1}{2}\Delta p$, the second term equals

$$\begin{aligned}\int_{\mathbb{R}^n} \int_s^t (\partial/\partial r)p(r,x,y)g(y)drdy &= \int_{\mathbb{R}^n} p(t,x,y)g(y)dy - \int_{\mathbb{R}^n} p(s,x,y)g(y)dy \\ &= P_t g(x) - \mathbb{E}^x g(B_s) \to -g(x)\end{aligned}$$

as $s \downarrow 0$ and $t \uparrow \infty$. □

Theorem 16.8.3. *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 $T = \inf\{t \geq 0 : B_t \in \partial D\}$.

(a) *Let ψ be a supersolution of the Dirichlet problem. If $\psi \geq 0$ then $\psi \geq \phi$.*
(b) *Let ψ be a solution of the Dirichlet problem. If ψ is bounded and $\mathbb{P}^x(T < \infty) = 1$ for all $x \in D$, then $\psi = \phi$.*
(c) *Assume that $f \in C(\partial D)$ and $g \in C^2(\mathbb{R}^n)$. If ϕ is locally bounded then it is a solution of the Dirichlet problem.*

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

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

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

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

is a martingale, by Theorem 16.3.3. Denote by T_N the hitting time of ∂D_N. Then, by optional stopping, for $x \in D_N$,

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

We now let $N \to \infty$. Since ψ is a supersolution,

$$\mathbb{E}^x \int_0^{T_N} (-\tfrac{1}{2}\Delta)\psi(B_t)dt \geq \mathbb{E}^x \int_0^{T_N} g(B_t)dt \uparrow \mathbb{E}^x \int_0^T g(B_t)dt$$

and $\psi(B_{T_N}) \to \psi(B_T) \geq f(B_T)$ on $\{T < \infty\}$. Hence, if $\psi \geq 0$,

$$\liminf_N \psi(B_{T_N}) \geq f(B_T)1_{T<\infty}$$

and so, by Fatou's lemma,

$$\liminf_N \mathbb{E}^x \psi(B_{T_N}) \geq \mathbb{E}^x(f(B_T)1_{T<\infty}).$$

Hence $\psi(x) \geq \phi(x)$. □

Proof of (b). In the case where ψ is a bounded solution of the Dirichlet problem and $\mathbb{P}^x(T < \infty) = 1$ for all $x \in D$, we have

$$\mathbb{E}^x \int_0^{T_N} (-\tfrac{1}{2}\Delta)\psi(B_t)dt \uparrow \mathbb{E}^x \int_0^T g(B_t)dt$$

and $\psi(B_{T_N}) \to f(B_T)$ a.s.. So, by bounded convergence,

$$\lim_N \mathbb{E}^x \psi(B_{T_N}) = \mathbb{E}^x(f(B_T)).$$

Hence $\psi(x) = \phi(x)$. $\square$

Proof of (c). Let D_0 be a bounded open subset of D and set $T_0 = \inf\{t \geq 0 : B_t \notin D_0\}$. Then T_0 is a stopping time and $T_0 < \infty$ a.s.. Set $\tilde{B}_t = B_{T_0+t}$, $\tilde{\mathfrak{F}}_t = \mathfrak{F}_{T_0+t}$ and $\tilde{T} = \inf\{t \geq 0 : \tilde{B}_t \notin D\}$. Note that $\tilde{T} < \infty$ if and only if $T < \infty$ and then $B_T = \tilde{B}_{\tilde{T}}$. By the strong Markov property, $\tilde{B}$ is an $(\tilde{\mathfrak{F}}_t)_{t\geq 0}$-Brownian motion, so

$$\begin{aligned}
\phi(x) &= \mathbb{E}^x\left(\int_0^{T_0} g(B_t)dt + \int_0^{\tilde{T}} g(\tilde{B}_t)dt + f(\tilde{B}_{\tilde{T}})1_{\tilde{T}<\infty}\right) \\
&= \mathbb{E}^x\left(\int_0^{T_0} g(B_t)dt\right) + \mathbb{E}^x\left(\mathbb{E}\left(f(\tilde{B}_{\tilde{T}})1_{\tilde{T}<\infty} + \int_0^{\tilde{T}} g(\tilde{B}_t)dt \,\middle|\, \tilde{\mathfrak{F}}_0\right)\right) \qquad (16.1)\\
&= \mathbb{E}^x\left(\int_0^{T_0} g(B_t)dt + \phi(B_{T_0})\right).
\end{aligned}$$

It is trivial that $\phi = f$ on ∂D. We can now prove that, for $y \in \partial D$ we have $\phi(x) \to f(y)$ as $x \to y, x \in D$. Choose $D_0 = U \cap D$, where U is a bounded open set in $\mathbb{R}^n$ 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

$$\phi(x) = \mathbb{E}^0\left(\int_0^{T_0(x)} g(x + B_t)dt + \phi(x + B_{T_0(x)})\right).$$

Since ∂D is smooth, there is an open cone C such that $y + C \subseteq D^c$. By Proposition 16.6.5, $\mathbb{P}^0(T_C = 0) = 1$, where T_C is the hitting time of C. Note that $T_C = 0$ implies that $x + B_{T_0(x)} \in \partial D$ for x sufficiently close to y and $x + B_{T_0(x)} \to y$ as $x \to y$. Since f is continuous on ∂D, this further implies that $\phi(x + B_{T_0(x)}) \to f(y)$ 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 required.

Consider now the case where $g \equiv 0$. Fix $x \in D$ and take $D_0 = B(x, \rho)$ with $B(x, \rho) \subseteq D$. 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 16.8.1 that ϕ is harmonic in D.

By linearity, it now suffices to treat the case where $f \equiv 0$. Moreover, it also suffices to treat the case where $n \geq 3$. For, if $n < 3$ we can simply apply the result for $n = 3$ to cylindrical regions D and to functions g which depend only on the first and second coordinates. Assume then that $f \equiv 0$ and $n \geq 3$. Assume also, 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 16.8.2, $\phi_0 \in C_b^2(\mathbb{R}^n)$ with $-\frac{1}{2}\Delta\phi_0 = \tilde{g}$. On taking $\phi = \phi_0$ and $D = \mathbb{R}^n, D_0 = D$ in (16.1), we obtain

$$\phi_0(x) = \phi(x) + \phi_1(x)$$

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 (16.1) 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. □

16.9. **Donsker's invariance principle.** In this section we shall 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 16.9.1 (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

$$\mu_\pm(A) = \mu(\pm A), \quad A \in \mathcal{B}([0, \infty)).$$

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 inductively for $n \geq 0$

$$T_{n+1} = \inf\{t \geq T_n : B_{T_n+t} - B_{T_n} = -X_{n+1} \text{ or } Y_{n+1}\}.$$

Then T_n is a stopping time for all n. Note that, since μ has mean 0, we must have

$$C\int_{-\infty}^0 (-x)\mu(dx) = C\int_0^\infty y\mu(dy) = 1.$$

Write $T = T_1$,$X = X_1$ and $Y = Y_1$.

By Proposition 16.5.1, conditional on $X = x$ and $Y = y$, we have $T < \infty$ a.s. and

$$\mathbb{P}(B_T = y | X = x \text{ and } Y = y) = x/(x+y),$$
$$\mathbb{E}(T | X = x \text{ and } Y = y) = xy.$$

So, for $A \in \mathcal{B}([0,\infty))$,

$$\mathbb{P}(B_T \in A) = \int_A \int_0^\infty \frac{x}{x+y} C(x+y)\mu_-(dx)\mu_+(dy)$$

so $\mathbb{P}(B_T \in A) = \mu(A)$. A similar argument shows this identity holds also for $A \in \mathcal{B}((-\infty,0])$. Next

$$\begin{aligned}\mathbb{E}(T) &= \int_0^\infty \int_0^\infty xyC(x+y)\mu_-(dx)\mu_-(dy) \\ &= \int_{-\infty}^0 (-x)^2\mu(dx) + \int_0^\infty y^2\mu(dy) = \sigma^2.\end{aligned}$$

Now by the strong Markov property for each $n \geq 0$ the process $(B_{T_n+t} - B_{T_n})_{t\geq 0}$ is a Brownian motion, independent of $\mathcal{F}_{T_n}$. So by the above argument $B_{T_{n+1}} - B_{T_n}$ has law μ, $T_{n+1} - T_n$ has mean σ^2, and both are independent of $\mathcal{F}_{T_n}$. The result follows. □

For $x \in C([0,1],\mathbb{R})$ we set $\|x\| = \sup_t |x_t|$. This *uniform norm* makes $C([0,1],\mathbb{R})$ into a metric space so we can consider weak convergence of probability measures. The associated Borel σ-algebra coincides with the σ-algebra generated by the coordinate functions.

Theorem 16.9.2 (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*

$$S_{n+t} = (1-t)S_n + tS_{n+1}, \quad t \in [0,1]$$

and set $S_t^{[N]} = N^{-1/2}S_{Nt}$. Then the law of $(S_t^{[N]})_{0\leq t\leq 1}$ converges weakly to Wiener measure on $C([0,1],\mathbb{R})$.

Proof. Take $(B_t)_{t\geq 0}$ and $((X_n, Y_n) : n \in \mathbb{N})$ as in the proof of Theorem 16.9.1. For each $N \geq 1$, set $B_t^{(N)} = N^{1/2}B_{N^{-1}t}$. Then $(B_t^{(N)})_{t\geq 0}$ is a Brownian motion. Perform the Skorohod embedding construction, with $(B_t)_{t\geq 0}$ replaced by $(B_t^{(N)})_{t\geq 0}$, to obtain stopping times $T_n^{(N)}$. Then set $S_n^{(N)} = B^{(N)}(T_n^{(N)})$ and interpolate linearly to form $(S_t^{(N)})_{t\geq 0}$. For all N, we have

$$\left(\left(T_n^{(N)}\right)_{n\geq 0}, \left(S_t^{(N)}\right)_{t\geq 0}\right) \sim \left((T_n)_{n\geq 0}, (S_t)_{t\geq 0}\right).$$

Next set $\tilde{T}_n^{(N)} = N^{-1}T_n^{(N)}$ and $\tilde{S}_t^{(N)} = N^{-1/2}S_{Nt}^{(N)}$. Then

$$\left(\tilde{S}_t^{(N)}\right)_{t\geq 0} \sim \left(S_t^{[N]}\right)_{t\geq 0}$$

and $\tilde{S}^{(N)}_{n/N} = B_{\tilde{T}^{(N)}_n}$ for all n. We have to show, for all bounded continuous functions $F : C([0,1],\mathbb{R}) \to \mathbb{R}$, that, as $N \to \infty$,

$$\mathbb{E}(F(S^{[N]})) \to \mathbb{E}(F(B)).$$

In fact we shall show, for all $\varepsilon > 0$,

$$\mathbb{P}\big(\sup_{0\le t\le 1} |\tilde{S}^{(N)}_t - B_t| > \varepsilon\big) \to 0.$$

Since F is continuous, this implies that $F(\tilde{S}^{(N)}) \to F(B)$ in probability, which by bounded convergence is enough.

By the strong law of large numbers $T_n/n \to 1$ a.s. as $n \to \infty$. So, as $N \to \infty$,

$$N^{-1} \sup_{n\le N} |T_n - n| \to 0 \quad \text{a.s..}$$

Hence, for all $\delta > 0$,

$$\mathbb{P}\big(\sup_{n\le N} |\tilde{T}^{(N)}_n - n/N| > \delta\big) \to 0.$$

By the intermediate value theorem, for $n/N \le t \le (n+1)/N$ we have $\tilde{S}^{(N)}_t = B_u$ for some $\tilde{T}^{(N)}_n \le u \le \tilde{T}^{(N)}_{n+1}$. Hence

$$\begin{aligned}
&\big\{|\tilde{S}^{(N)}_t - B_t| > \varepsilon \text{ for some } t \in [0,1]\big\} \\
&\quad\subseteq \big\{|\tilde{T}^{(N)}_n - n/N| > \delta \text{ for some } n \le N\big\} \\
&\qquad\qquad \cup\{|B_u - B_t| > \varepsilon \text{ for some } t \in [0,1] \text{ and } |u - t| \le \delta + 1/N\} \\
&\qquad\qquad = A_1 \cup A_2.
\end{aligned}$$

The paths of $(B_t)_{t\ge 0}$ are uniformly continuous on $[0,1]$. So given $\varepsilon > 0$ we can find $\delta > 0$ so that $\mathbb{P}(A_2) \le \varepsilon/2$ whenever $N \ge 1/\delta$. Then, by choosing N even larger if necessary, we can ensure $\mathbb{P}(A_1) \le \varepsilon/2$ also. Hence $\tilde{S}^{(N)} \to B$, uniformly on $[0,1]$ in probability, as required. □

17. Poisson random measures

17.1. **Construction and basic properties.** For $\lambda \in (0,\infty)$ we say that a random variable X in $\mathbb{Z}^+$ 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 17.1.1 (Addition property). *Let $N_k, k \in \mathbb{N}$, be independent random variables, with $N_k \sim P(\lambda_k)$ for all k. Then*

$$\sum_k N_k \sim P(\sum_k \lambda_k).$$

Proposition 17.1.2 (Splitting property)**.** *Let $N, Y_n, n \in \mathbb{N}$, be independent random variables, with $N \sim P(\lambda), \lambda < \infty$ and $\mathbb{P}(Y_n = j) = p_j$, for $j = 1, \dots, k$ and all n. Set*

$$N_j = \sum_{n=1}^{N} 1_{Y_n = j}.$$

Then $N_1, \dots, N_k$ are independent random variables with $N_j \sim P(\lambda p_j)$ for all j.

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) $M(A_k), k \in \mathbb{N}$, are 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 17.1.3. *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 independent random variables N and $Y_n, n \in \mathbb{N}$, with $N \sim P(\lambda)$ and $Y_n \sim \mu/\lambda$ for all n. Set

$$M(A) = \sum_{n=1}^{N} 1_{Y_n \in A}, \quad A \in \mathcal{E}. \tag{17.1}$$

It is easy to check, by 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 there exist disjoint sets $E_k \in \mathcal{E}, k \in \mathbb{N}$, such that $\cup_k E_k = E$ and $\mu(E_k) < \infty$ for all k. We can construct, on some probability space, independent Poisson random measures $M_k, k \in \mathbb{N}$, with M_k having intensity $\mu|_{E_k}$. Set

$$M(A) = \sum_{k \in \mathbb{N}} M_k(A \cap E_k), \quad A \in \mathcal{E}.$$

It is easy to check, by 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. $\square$

17.2. **Integrals with respect to a Poisson random measure.**

Theorem 17.2.1. *Let M be a Poisson random measure on E with intensity μ and let g be a measurable function on E. If $\mu(E)$ is finite or g is integrable, then*

$$X = \int_E g(y) M(dy)$$

is a well-defined random variable with

$$\mathbb{E}(e^{iuX}) = \exp\{\int_E (e^{iug(y)} - 1)\mu(dy)\}.$$

Moreover, if g is integrable, then so is X and

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

Proof. Assume for now that $\mu(E) < \infty$. Then $M(E)$ is finite a.s. so X is well defined. If $g = 1_A$ for some $A \in \mathcal{E}$, then $X = M(A)$, so X is a random variable. This extends by linearity and by taking limits to all measurable functions g.

Since the value of $\mathbb{E}(e^{iuX})$ depends only on the law μ^* of M on E^*, we can assume that M is given as in (17.1). Then

$$\mathbb{E}(e^{iuX} | N = n) = \mathbb{E}(e^{iug(Y_1)})^n = \left(\int_E e^{iug(y)} \frac{\mu(dy)}{\lambda} \right)^n$$

so

$$\begin{aligned} \mathbb{E}(e^{iuX}) &= \sum_{n=0}^{\infty} \mathbb{E}(e^{iuX} | N = n) \mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \left(\int_E e^{iug(y)} \frac{\mu(dy)}{\lambda} \right)^n e^{-\lambda} \lambda^n / n! = \exp\{\int_E (e^{iug(y)} - 1)\mu(dy)\}. \end{aligned}$$

If g is integrable, then formulae for $\mathbb{E}(X)$ and $\mathrm{var}(X)$ may be obtained by a similar argument.

It remains to deal with the case where g is integrable and $\mu(E) = \infty$. Assume for now that $g \geq 0$, then X is obviously well defined. We can find $0 \leq g_n \uparrow g$ with $\mu(|g_n| > 0) < \infty$ for all n. The conclusions of the theorem are then valid for the corresponding integrals X_n. Note that $X_n \uparrow X$ and $\mathbb{E}(X_n) \leq \mu(g) < \infty$ for all n. It follows that X is a random variable and, by monotone convergence, $X_n \to X$ in $L^1(\mathbb{P})$. Note the estimate $|e^{iux} - 1| \leq |ux|$. We can then obtain the desired formulae for X by passing to the limit. Finally, for a general integrable function g, we have

$$\mathbb{E} \int_E |g(y)| M(dy) = \int_E |g(y)| \mu(dy)$$

so X is well defined. Also $X = X_+ - X_-$, where

$$X_\pm = \int_{\{\pm g > 0\}} g(y) M(dy)$$

and X_+ and X_- are independent. Hence the formulae for X follow from those for $X_\pm$. □

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 $\tilde{M} = M - \mu$. Then $\tilde{M}$ is a *compensated Poisson random measure with intensity* μ.

Proposition 17.2.2. *Let g be an integrable function on E. Set*

$$X_t = \int_{(0,t] \times E} g(y) \tilde{M}(ds, dy).$$

Then $(X_t)_{t \geq 0}$ is a cadlag martingale with stationary independent increments. Moreover

$$\mathbb{E}(e^{iuX_t}) = \exp\{t \int_E (e^{iug(y)} - 1 - iug(y)) K(dy)\}$$

$$\mathbb{E}(X_t^2) = t \int_E g(y)^2 K(dy).$$

Theorem 17.2.3. *Let $g \in L^2(K)$ and let $(g_n : n \in \mathbb{N})$ be a sequence of integrable functions such that $g_n \to g$ in $L^2(K)$. Set*

$$X_t^n = \int_{(0,t] \times E} g_n(y) \tilde{M}(ds, dy).$$

Then there exists a cadlag martingale $(X_t)_{t \geq 0}$ such that

$$\mathbb{E}\left(\sup_{s \leq t} |X_s^n - X_s|^2 \right) \to 0$$

for all $t \geq 0$*. Moreover* $(X_t)_{t\geq 0}$ *has stationary independent increments and*

$$\mathbb{E}(e^{iuX_t}) = \exp\{t \int_E (e^{iug(y)} - 1 - iug(y))K(dy)\}.$$

The notation $\int_{(0,t]\times E} g(y)\tilde{M}(ds, dy)$ is used for X_t, even when g is not integrable with respect to K. Of course $(X_t)_{t\geq 0}$ does not depend on the choice of approximating sequence (g_n). This is a simple example of a *stochastic integral.*

Proof. Fix $t > 0$. By Doob's L^2-inequality and Proposition 17.2.2,

$$\mathbb{E}\left(\sup_{s\leq t} |X^n_s - X^m_s|^2\right) \leq 4\mathbb{E}((X^n_t - X^m_t)^2) = 4t \int_E (g_n - g_m)^2 dK \to 0$$

as $n, m \to \infty$. Hence X^n_s converges in L^2 for all $s \leq t$. For some subsequence we have

$$\sup_{s\leq t} |X^{n_k}_s - X^{n_j}_s| \to 0 \quad \text{a.s.}$$

as $j, k \to \infty$. The uniform limit of cadlag functions is cadlag, so there is a cadlag process $(X_s)_{s\leq t}$ such that

$$\sup_{s\leq t} |X^{n_k}_s - X_s| \to 0 \quad \text{a.s.}.$$

Since X^n_s converges in L^2 for all $s \leq t$, $(X_s)_{s\leq t}$ is a martingale and so by Doob's L^2-inequality

$$\mathbb{E}\left(\sup_{s\leq t} |X^n_s - X_s|^2\right) \leq 4\mathbb{E}((X^n_t - X_t)^2) \to 0.$$

Note that $|e^{iug} - 1 - iug| \leq u^2g^2/2$. Hence, for $s < t$ we have

$$\begin{aligned}\mathbb{E}(e^{iu(X_t - X_s)}|\mathcal{F}^M_s) &= \lim_n \mathbb{E}(e^{iu(X^n_t - X^n_s)}|\mathcal{F}^M_s)\\ &= \lim_n \exp\{(t-s)\int_E (e^{iug_n(y)} - 1 - iug_n(y))K(dy)\}\\ &= \exp\{(t-s)\int_E (e^{iug(y)} - 1 - iug(y))K(dy)\}\end{aligned}$$

which shows that $(X_t)_{t\geq 0}$ has independent increments with the claimed characteristic function. □

18. Lévy processes

18.1. **Definition and examples.** A *Lévy process* is a cadlag process starting from 0 with stationary independent increments. A *Lévy system* is a triple (a, b, K), where $a = \sigma^2 \in [0, \infty)$ is the *diffusivity*, $b \in \mathbb{R}$ is the *drift* and K, the *Lévy measure*, is a Borel measure on $\mathbb{R}$ with $K(\{0\}) = 0$ and

$$\int_{\mathbb{R}} (1 \wedge |y|^2)K(dy) < \infty.$$

Let B be a Brownian motion and let M be a Poisson random measure with intensity μ on $(0,\infty)\times\mathbb{R}$, where $\mu(dt,dy)=dtK(dy)$, as in the preceding section. Set

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

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

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

Thus, to every Lévy system 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 σB. Hence the law of the Lévy process $(X_t)_{t\ge 0}$ determines the Lévy system (a,b,K).

18.2. **Lévy–Khinchin theorem.**

Theorem 18.2.1 (Lévy–Khinchin theorem)**.** *Let X be a Lévy process. Then there exists a unique Lévy system (a,b,K) such that, for all $t\ge 0$,*

$$\mathbb{E}(e^{iuX_t})=e^{t\psi(u)} \tag{18.1}$$

where

$$\psi(u)=ibu-\tfrac{1}{2}au^2+\int_{\mathbb{R}}(e^{iuy}-1-iuy1_{|y|\le 1})K(dy). \tag{18.2}$$

Proof. First we shall show that there is a continuous function $\psi:\mathbb{R}\to\mathbb{C}$ with $\psi(0)=0$ such that (18.1) holds for all $u\in\mathbb{R}$ and for $t=1/n$ for all $n\in\mathbb{N}$. Let ν_n denote the law, and ϕ_n the characteristic function, of $X_{1/n}$. Note that ϕ_n is continuous and $\phi_n(0)=1$. Let I_n denote the largest open interval containing 0 where $|\phi_n|>0$. There is a unique continuous function $\psi_n: I_n\to\mathbb{C}$ such that $\psi_n(0)=0$ and

$$\phi_n(u)=e^{\psi_n(u)/n},\quad u\in I_n.$$

Since X is a Lévy process, we have $(\phi_n)^n=\phi_1$, so we must have $I_n=I_1$ and $\psi_n=\psi_1$ for all n. Write $I=I_1$ and $\psi=\psi_1$. Then $\phi_n\to 1$ on I as $n\to\infty$ and $\phi_n=0$ on ∂I for all n. By the argument used in Theorem 15.3.2, $(\nu_n:n\in\mathbb{N})$ is then tight, so for some subsequence $\phi_{n_k}\to\phi$ on $\mathbb{R}$, for some characteristic function ϕ. This forces $\partial I=\emptyset$, so $I=\mathbb{R}$.

We have shown that (18.1) holds for all $t\in\mathbb{Q}^+$. Since X is cadlag, this extends to all $t\in\mathbb{R}^+$ using

$$X_t=\lim_{n\to\infty}X_{2^{-n}\lceil 2^n t\rceil}.$$

It remains to show that ψ can be written in the form (18.2). We note that it suffices to find a similar representation where $1_{|y|\leq 1}$ is replaced by $\chi(y)$ for some continuous function χ with

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

We have

$$\int_{\mathbb{R}} (e^{iuy} - 1) n\nu_n(dy) = n(\phi_n(u) - 1) \to \psi(u)$$

as $n \to \infty$, uniformly on compacts in u. Hence

$$\int_{\mathbb{R}} (1 - \cos uy) n\nu_n(dy) \to -\operatorname{Re}\psi(u).$$

Now there is a constant $C < \infty$ such that

$$\begin{aligned} y^2 1_{|y|\leq 1} &\leq C(1-\cos y) \\ 1_{|y|\geq\lambda} &\leq C\lambda \int_0^{1/\lambda} (1 - \cos uy) du, \quad \lambda \in (0,\infty). \end{aligned}$$

Set $\eta_n(dy) = n(1 \wedge y^2)\nu_n(dy)$. Then, as $n \to \infty$,

$$\begin{aligned} \eta_n(|y| \leq 1) &= \int_{\mathbb{R}} y^2 1_{|y|\leq 1} n\nu_n(dy) \\ &\leq C \int_{\mathbb{R}} (1 - \cos y) n\nu_n(dy) \to -C \operatorname{Re}\psi(1) \end{aligned}$$

and, for $\lambda \geq 1$,

$$\begin{aligned} \eta_n(|y| \geq \lambda) &= \int_{\mathbb{R}} 1_{|y|\geq\lambda} n\nu_n(dy) \\ &\leq C\lambda \int_0^{1/\lambda} \int_{\mathbb{R}} (1-\cos uy) n\nu_n(dy) du \\ &\to -C\lambda \int_0^{1/\lambda} \operatorname{Re}\psi(u) du. \end{aligned}$$

We 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 $\eta_{n_k}(\theta) \to \eta(\theta)$ for all bounded continuous functions θ on

$\mathbb{R}$. Now

$$\begin{aligned}\int_{\mathbb{R}}(e^{iuy}-1)n\nu_n(dy) &= \int_{\mathbb{R}}(e^{iuy}-1)\frac{\eta_n(dy)}{1\wedge y^2}\\ &= \int_{\mathbb{R}}\frac{(e^{iuy}-1-iuy\chi(y))}{1\wedge y^2}\eta_n(dy)+\int_{\mathbb{R}}\frac{iuy\chi(y)}{1\wedge y^2}\eta_n(dy)\\ &= \int_{\mathbb{R}}\theta(u,y)\eta_n(dy)+iub_n\end{aligned}$$

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, for each u, $\theta(u,.)$ is a bounded continuous function. So, on letting $k\to\infty$,

$$\begin{aligned}\int_{\mathbb{R}}\theta(u,y)\eta_{n_k}(dy)\to&\int_{\mathbb{R}}\theta(u,y)\eta(dy)\\ &=\int_{\mathbb{R}}(e^{iuy}-1-iuy\chi(y))K(dy)-\tfrac{1}{2}au^2\end{aligned}$$

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, say to b, and we obtain the desired formula

$$\psi(u)=ibu-\tfrac{1}{2}au^2+\int_{\mathbb{R}}(e^{iuy}-1-iuy\chi(y))K(dy).$$

□

Exercises

Students should attempt Exercises 11.1–13.4 for their first supervision, then 13.5–14.3, 15.1–16.8 and 16.9–18.4 for later supervisions.

11.1 Let X and Y be integrable random variables and suppose that

$$\mathbb{E}(X|Y)=Y,\quad \mathbb{E}(Y|X)=X\quad \text{a.s.}$$

Show that $X=Y$ a.s.

11.2 Prove the conditional forms of Fatou's lemma and the dominated convergence theorem, stated in §11.5.

12.1 Let $(X_n : n\in\mathbb{N})$ be a sequence of independent integrable random variables. Set $S_0=0$, $P_0=1$ and $S_n=X_1+\dots+X_n$, $P_n=X_1\dots X_n$, $n\in\mathbb{N}$. Show that

(i) if $\mathbb{E}(X_n)=0$ for all n, then $(S_n)_{n\geq 0}$ is a martingale,
(ii) if $\mathbb{E}(X_n)=1$ for all n, then $(P_n)_{n\geq 0}$ is a martingale.

12.2 Let $X = (X_n)_{n\geq 0}$ be an integrable process, taking values in a countable set $E \subseteq \mathbb{R}$. Show that X is a martingale if and only if, for all n and for all $i_0, \dots, i_n \in E$, we have

$$\mathbb{E}(X_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = i_n.$$

12.3 Let $(X_n)_{n\geq 0}$ be a Markov chain in E with transition matrix P. Let $f : E \to \mathbb{R}$ be a bounded function. Find necessary and sufficient conditions on f for $(f(X_n))_{n\geq 0}$ to be a martingale.

12.4 Find a simple direct argument to show that for any martingale $(X_n)_{n\geq 0}$ and any bounded stopping time T we have $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

12.5 Let S_1 and S_2 be defined as in the proof of Theorem 12.3.1. Show that S_1 and S_2 are stopping times.

12.6 Let $(X_t : t \in I)$ be a countable family of non-negative random variables and suppose that, for all $s, t \in I$, there exists $u \in I$ such that $X_u \geq \max(X_s, X_t)$. Show carefully that

$$\mathbb{E}(\sup_{t\in I} X_t) = \sup_{t\in I} \mathbb{E}(X_t).$$

12.7 Let $X = (X_n)_{n\geq 0}$ be a martingale in L^2. Show that X is bounded in L^2 if and only if

$$\sum_{n=0}^{\infty} \mathbb{E}\big((X_{n+1} - X_n)^2\big) < \infty.$$

12.8 Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration and set $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$. Let $X \in L^2$. Set $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$. Show, by a direct argument, that X_n converges in L^2 and that

$$X_n \to X \text{ in } L^2 \quad \Leftrightarrow \quad X \text{ is } \mathcal{F}_\infty\text{-measurable}.$$

12.9 Write out the details of the proof of the backward martingale convergence theorem, say for $p = 1$.

13.1 Prove Propositions 13.1.2 and 13.1.3.

Examples 13.2–13.7 are taken from Williams, Probability with Martingales.

13.2(a) *Pólya's urn.* At time 0, an urn contains 1 black ball and 1 white ball. At each time $1, 2, 3, \dots$, a ball is chosen at random from the urn and is replaced together with a new ball of the same colour. Just after time n, there are therefore $n + 2$ balls in the urn, of which $B_n + 1$ are black, where B_n is the number of black balls chosen by time n. Let $M_n = (B_n + 1)/(n + 2)$ the proportion of black balls in the urn just after time n. Prove that, relative to a natural filtration which you should specify, M is a martingale.

Prove also that $\mathbb{P}(B_n = k) = (n + 1)^{-1}$ for $0 \leq k \leq n$.

What is the distribution of Θ, where $\Theta := \lim M_n$?

Prove that for $0 < \theta < 1$, $(N_n^\theta)_{n\geq 0}$ is a martingale, where

$$N_n^\theta := \frac{(n+1)!}{B_n!(n-B_n)!}\,\theta^{B_n}(1-\theta)^{n-B_n}.$$

13.2(b) *Bayes' urn.* A random number Θ is chosen uniformly between 0 and 1, and a coin with probability Θ of heads is minted. The coin is tossed repeatedly. Let B_n be the number of heads in n tosses. Prove that (B_n) has exactly the same probabilistic structure as the (B_n) sequence in **13.2(a)**. Prove that N_n^θ is a conditional density function of Θ given $B_1, B_2, \dots, B_n$.

13.3 Your winnings per unit stake on game n are ε_n, where the ε_n are independent random variables with

$$\mathbb{P}(\varepsilon_n = 1) = p, \quad \mathbb{P}(\varepsilon_n = -1) = q,$$

where $p \in (\frac{1}{2}, 1)$ and $q = 1-p$. Your stake C_n on game n must lie between 0 and Z_{n-1}, where Z_{n-1} is your fortune at time $n-1$. Your object is to maximize the expected 'interest rate' $\mathbb{E}\log(Z_N/Z_0)$, where N is a given integer representing the length of the game, and Z_0, your fortune at time 0, is a given constant. Let $\mathcal{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$. Show that if C is any *previsible* strategy, that is C_n is $\mathcal{F}_{n-1}$-measurable for all n, then $\log Z_n - n\alpha$ is a supermartingale, where α denotes the *entropy*

$$\alpha = p\log p + q\log q + \log 2,$$

so that $\mathbb{E}\log(Z_n/Z_0) \leq N\alpha$, but that, for a certain strategy, $\log Z_n - n\alpha$ is a *martingale*. What is the best strategy?

13.4 *ABRACADABRA.* At each of times $1, 2, 3, \dots$, a monkey types a capital letter at random, the sequence of letters typed forming a sequence of independent random variables, each chosen uniformly from amongst the 26 possible capital letters.

Just before each time $n = 1, 2, \dots$, a new gambler arrives on the scene. He bets \$1 that

the n^{th} letter will be A.

If he loses, he leaves. If he wins, he receives \$26 all of which he bets on the event that

the $(n+1)^{\text{th}}$ letter will be B.

If he loses, he leaves. If he wins, he bets his whole current fortune \$$26^2$ that

the $(n+2)^{\text{th}}$ letter will be R

and so on through the ABRACADABRA sequence. Let T be the first time by which the monkey has produced the consecutive sequence ABRACADABRA. Prove, by a martingale argument, that

$$\mathbb{E}(T) = 26^{11} + 26^4 + 26.$$

13.5 *'What always stands a reasonable chance of happening will (almost surely) happen—sooner rather than later.'* Suppose that T is a stopping time such that for some $N \in \mathbb{N}$ and some $\varepsilon > 0$, we have, for every n:

$$\mathbb{P}(T \leq n + N \mid \mathcal{F}_n) > \varepsilon, \quad \text{a.s.}$$

Prove by induction using $\mathbb{P}(T > kN) = \mathbb{P}(T > kN; T > (k-1)N)$ that for $k = 1, 2, 3, \ldots$

$$\mathbb{P}(T > kN) \leq (1 - \varepsilon)^k.$$

Show that $\mathbb{E}(T) < \infty$.

13.6 *Gambler's Ruin.* Suppose that $X_1, X_2, \ldots$ are independent random variables with

$$\mathbb{P}(X = +1) = p, \quad \mathbb{P}(X = -1) = q,$$

where $p \in (0,1)$, $q = 1 - p$ and $p \neq q$. Suppose that a and b are integers with $0 < a < b$. Define

$$S_n := a + X_1 + \cdots + X_n, \quad T := \inf\{n : S_n = 0 \text{ or } S_n = b\}.$$

Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Explain why T satisfies the conditions in **13.5** Prove that

$$M_n := \left(\frac{q}{p}\right)^{S_n} \quad \text{and} \quad N_n = S_n - n(p - q)$$

define martingales M and N. Deduce the values of $\mathbb{P}(S_T = 0)$ and $\mathbb{E}(T)$.

13.7 *Azuma–Hoeffding Inequality.*
(a) Show that if Y is a random variable with values in $[-c, c]$ and with $\mathbb{E}(Y) = 0$, then, for $\theta \in \mathbb{R}$,

$$\mathbb{E}(e^{\theta Y}) \leq \cosh \theta c \leq \exp\left(\frac{1}{2}\,\theta^2 c^2\right).$$

(b) Prove that if M is a martingale, with $M_0 = 0$ and such that for some sequence $(c_n : n \in \mathbb{N})$ of positive constants, $|M_n - M_{n-1}| \leq c_n$ for all n, then, for $x > 0$,

$$\mathbb{P}\Big(\sup_{k \leq n} M_k \geq x\Big) \leq \exp\Big(-\frac{1}{2}\,x^2 \Big/ \sum_{k=1}^{n} c_k^2\Big).$$

Hint for (a). Let $f(z) := \exp(\theta z)$, $z \in [-c, c]$. Then, since f is convex,

$$f(y) \leq \frac{c - y}{2c}\, f(-c) + \frac{c + y}{2c}\, f(c).$$

Hint for (b). Optimize over θ.

13.8 Let $(\Omega, \mathcal{F})$ denote the set of real sequences $\omega = (\omega_n : n \geq 0)$ such that

$$\limsup_{n \to \infty} \omega_n = -\liminf_{n \to \infty} \omega_n = \infty,$$

with the σ-algebra generated by the coordinate functions $X_n(\omega) = \omega_n$. Show that, for $p = 1/2$ and for no other $p \in (0,1)$, there exists a probability measure $\mathbb{P}_p$ on $(\Omega, \mathcal{F})$ making $(X_n)_{n\geq 0}$ into a simple random walk with

$$\mathbb{P}(X_1 = 1) = p, \quad \mathbb{P}(X_1 = -1) = 1 - p.$$

Let $\mathbb{P}_{p,n}$ denote the unique probability measure on $(\Omega, \mathcal{F}_n)$ making $(X_k)_{0\leq k\leq n}$ into a simple random walk with $\mathbb{P}(X_1 = 1) = p$, where $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$. Fix $p \in (0,1) \setminus \{1/2\}$. Identify the martingale

$$M_n = d\mathbb{P}_{p,n}/d\mathbb{P}_{1/2,n}.$$

Find a finite stopping T such that

$$\mathbb{E}_{1/2}(M_T) < 1.$$

13.9 Let $f : [0,1] \to \mathbb{R}$ be Lipschitz, that is, suppose that, for some $K < \infty$ and all $x, y \in [0,1]$

$$|f(x) - f(y)| \leq K|x - y|.$$

Denote by f_n the simplest piecewise linear function agreeing with f on $\{k2^{-n} : k = 0, 1, \dots, 2^n\}$. Set $M_n = f_n'$. Show that M_n converges a.e. and in L^1 and deduce that f is the indefinite integral of a bounded function.

13.10 Let X be a non-negative random variable with $\mathbb{E}(X) = 1$. Show that

$$\mathbb{E}\big(\sqrt{X}\big) \leq 1$$

with equality only if $X = 1$ a.s.

13.11 In an experiment to determine a parameter θ, it is possible to make a series of independent measurements of declining accuracy, so that the kth measurement $X_k \sim N(\theta, \sigma_k^2)$. Let $\hat{\Theta}_n$ denote the maximum likelihood estimate for θ based on the first n measurements. Determine for which sequences $(\sigma_k)_{k\in\mathbb{N}}$ we have $\hat{\Theta}_n \to \theta$ a.s. as $n \to \infty$. Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Show that, for all θ, θ' and all n, $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta'}$ are mutually absolutely continuous on $\mathcal{F}_n$. Is the same true for $\mathcal{F}_\infty$?

13.12 Prove Propositions 13.3.1 and 13.3.2.

13.13 Let $(X_n)_{n\geq 0}$ be a Markov chain and suppose that

$$\mathbb{P}^i(X_n = i \text{ for some } n \geq 1) = 1.$$

Define inductively

$$T_{k+1} = \inf\{n \geq 1 : X_{T_1+\dots+T_k+n} = i\}.$$

Show that the random variables $T_1, T_2, \dots$ are independent and identically distributed.

14.1 Prove Proposition 14.3.5(c).

14.2 Let $T \sim E(\lambda)$. Define

$$Z_t = \begin{cases} 0 & \text{if } t < T \\ 1 & \text{if } t \geq T \end{cases}, \quad \mathcal{F}_t = \sigma\{Z_s : s \leq t\}, \quad M_t = \begin{cases} 1 - e^{\lambda t} & \text{if } t < T \\ 1 & \text{if } t \geq T. \end{cases}$$

Prove that $\mathbb{E}|M_t| < \infty$, and that $\mathbb{E}(M_t; \{T > r\}) = \mathbb{E}(M_s; \{T > r\})$ for $r \leq s \leq t$, and hence deduce that M_t is a cadlag martingale with respect to the filtration $\{\mathcal{F}_t\}$.

Is M bounded in L^1? Is M uniformly integrable? Is M_{T-} in L^1?

14.3 Let T be a random variable with values in $(0, \infty)$ and with strictly positive continuous density f on $(0, \infty)$ and distribution function $F(t) = \mathbb{P}(T \leq t)$. Define

$$A_t = \int_0^t \frac{f(s)ds}{1 - F(s)}, \qquad 0 \leq t < \infty.$$

By expressing the distribution function of $A_T, G(t) = \mathbb{P}(A_T \leq t)$, in terms of the inverse function A^{-1} of A, or otherwise, deduce that A_T has the exponential distribution of mean 1.

Define Z_t and $\mathcal{F}_t$ as in **14.2** above, and prove that $M_t = Z_t - A_{t \wedge T}$ is a cadlag martingale relative to $\{\mathcal{F}_t\}$. The function A_t is called the *hazard function* for T.

15.1 Assuming Prohorov's theorem, prove that if $(\mu_n : n \in \mathbb{N})$ is a tight sequence of finite measures on $\mathbb{R}$ and if

$$\sup_n \mu_n(\mathbb{R}) < \infty$$

then there is a subsequence (n_k) and a finite measure μ on $\mathbb{R}$ such that $\mu_{n_k} \Rightarrow \mu$.

15.2 Let $(X_n : n \in \mathbb{N})$ be a sequence of independent, identically distributed, integrable random variables. Set $S_n = X_1 + \cdots + X_n$. Use characteristic functions to show that

$$S_n/n \Rightarrow \mathbb{E}(X_1).$$

15.3 Let $(X_n : n \in \mathbb{N})$ be a sequence of random variable and suppose that

$$X_n \Rightarrow X.$$

Show that, if X is a.s. constant, then also X_n converges to X in probability. Is the condition that X is a.s. constant necessary?

16.1 Let $(B_t)_{t \geq 0}$ be a Brownian motion starting from 0. Show that

$$\limsup_{t \downarrow 0} B_t/t = -\liminf_{t \downarrow 0} B_t/t = \infty \quad \text{a.s.}$$

16.2 Let $(B_t)_{t \geq 0}$ be a Brownian motion starting from 0. Set

$$L = \sup\{t > 0 : B_t = at\}.$$

Show that L has the same distribution as H_a^{-1}.

16.3 Let $(B_t)_{t\geq 0}$ be a Brownian motion. Find all polynomials $f(t,x)$, of degree 3 in x, such that $(M_t)_{t\geq 0}$ is a martingale, where

$$M_t = f(t, B_t).$$

16.4 Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from 0. Find the distribution of $(B_t, \max_{s\leq t} B_s)$.

16.5 Let $(B_t)_{t\geq 0}$ be a Brownian motion starting from 0. Show that $(tB_{1/t})_{t>0}$ and $(B_t)_{t>0}$ have the same distribution.

Show also that $tB_{1/t} \to 0$ a.s. as $t \to 0$.

16.6 Let D be a dense subset of $[0,1]$ and suppose that $f : D \to \mathbb{R}$ satisfies, for some $K < \infty$ and $\alpha \in (0,1]$

$$(*) \qquad |f(s) - f(t)| \leq K|t-s|^\alpha$$

for all $s,t \in D$. Show that f has a unique extension $\tilde{f} : [0,1] \to \mathbb{R}$ such that $(*)$ holds for all $s,t \in [0,1]$.

16.7 Prove Propositions 16.2.1, 16.2.3, 16.3.2, 16.5.1, 16.5.2, 16.6.1, 16.6.4 and 16.6.5.

16.8 Let $(B_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^3$. Set $R_t = 1/|B_t|$. Show that

(i) $(R_t : t \geq 1)$ is bounded in L^2,
(ii) $\mathbb{E}(R_t) \to 0$ as $t \to \infty$,
(iii) R_t is a supermartingale.

Deduce that $|B_t| \to \infty$ a.s. as $t \to \infty$.

16.9 Let μ denote Wiener measure on $W = \{x \in C([0,1],\mathbb{R}) : x_0 = 0\}$. For $a \in \mathbb{R}$, define a new probability measure μ_a on W by

$$d\mu_a/d\mu(x) = \exp(ax_1 - a^2/2).$$

Show that under μ_a the coordinate process remains Gaussian, and identify its distribution.

Deduce that $\mu(A) > 0$ for every non-empty open set $A \subseteq W$.

16.10 Let $B = (B_t)_{0\leq t\leq 1}$ be a Brownian motion starting from 0. Denote by μ the law of B on $W = C([0,1],\mathbb{R})$. For each $y \in \mathbb{R}$, set

$$Z_t^y = yt + (B_t - tB_1)$$

and denote by μ^y the law of $Z^y = (Z_t^y)_{0\leq t\leq 1}$ on W. Show that, for any bounded measurable function $F : W \to \mathbb{R}$ and for $f(y) = \mu^y(F)$ we have

$$\mathbb{E}(F(B) \mid B_1) = f(B_1) \quad \text{a.s..}$$

16.11 Let D be a bounded open set in $\mathbb{R}^n$ and let $h : \overline{D} \to \mathbb{R}$ be a bounded continuous function, harmonic in D. Show that, for all $x \in D$,

$$\inf_{y \in \partial D} h(y) \leq h(x) \leq \sup_{y \in \partial D} h(y).$$

16.12 (i) Let $(B_t)_{t \geq 0}$ be a Brownian motion in $\mathbb{R}^2$ starting from (x, y). Compute the distribution of B_T, where

$$T = \inf\{t \geq 0 : B_t \notin H\}$$

and where H is the upper half plane $\{(x, y) : y > 0\}$.

(ii) Show that, for any bounded continuous function $u : \overline{H} \to \mathbb{R}$, harmonic in H, with $u(x, 0) = f(x)$ for all $x \in \mathbb{R}$, we have

$$u(x, y) = \int_{\mathbb{R}} f(s) \frac{1}{\pi} \frac{y}{(x - s)^2 + y^2} \, ds.$$

(iii) Let D be any open set in $\mathbb{R}^2$ for which there exists a continuous homeomorphism $g : \overline{H} \to \overline{D}$, which is conformal in H. Show that, if u is harmonic in D, then $u \circ g$ is harmonic in H.

(iv) Find an explicit integral representation for bounded continuous functions $u : \overline{D} \to \mathbb{R}$, harmonic in D, in terms of their values on the boundary of D.

(v) Determine the exit distribution of Brownian motion from D.

18.1 Let $(X_t)_{t \geq 0}$ be a Lévy process with characteristic exponent ψ. Show that, for all $u \in \mathbb{R}$, the following process is a martingale:

$$M_t^u = \exp\{iuX_t - t\psi(u)\}.$$

18.2 By generalizing the case of Brownian motion, formulate and prove a strong Markov property for Lévy processes.

18.3 Say that a Lévy process $(X_t)_{t \geq 0}$ satisfies the scaling relation with exponent $\alpha \in (0, \infty)$ if

$$(cX_{c^{-\alpha} t})_{t \geq 0} \sim (X_t)_{t \geq 0}, \quad c \in (0, \infty).$$

For example Brownian motion satisfies the scaling relation with exponent 2. Find, for each $\alpha \in (0, 2)$, a Lévy process having a scaling relation with exponent α.

18.4 Let $(X_t)_{t \geq 0}$ be the Lévy process corresponding to the Lévy triple (a, b, K). Show that, if K consists of finitely many atoms, then $(X_t)_{t \geq 0}$ can be written as a linear combination of a Brownian motion, a uniform drift and finitely many Poisson processes.

INDEX