Advanced Probability
Michaelmas 2006
Dr. G. Miermont*

Chris Almost†

Contents

*gregory@statslab.cam.ac.uk
†cda23@cam.ac.uk

0 Basics of Measure Theory

0.1 Convergence theorems

Let $(E, \mathscr{A}, \mu)$ be a measure space. If $f : E \to \mathbb{R}_+$ is a measurable function then we let $\mu(f)$ denote $\int_E f \, d\mu$.

0.1.1 Theorem (Monotone Convergence Theorem).
Let $(f_n, n \geq 0)$ be non-negative measurable functions such that $f_n \leq f_{n+1}$ a.s. for all n. If $f = \lim_{n\to\infty} f_n$ then $\mu(f) = \lim_{n\to\infty} \mu(f_n)$.

0.1.2 Theorem (Fatou's Lemma).
Let $(f_n, n \geq 0)$ be non-negative measurable functions. Then

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

0.1.3 Theorem (Dominanted Convergence Theorem).
Let $(f_n, n \geq 0)$ be measurable functions. Assume that $|f_n| \leq g$ for all n, for some measurable function g such that $\mu(g) < \infty$, and $f_n \to f$ a.s. Then $\mu(f_n) \to \mu(f)$ and in fact $\mu(|f_n - f|) \to 0$.

0.2 Uniqueness of measure

0.2.1 Definition. A *π-system* is a collection Π of sets such that $A \cap B \in \Pi$ for all $A, B \in \Pi$. A *d-system* is a collection $\mathscr{D}$ of sets such that

(i) $\Omega \in \mathscr{D}$;

(ii) if $A \subseteq B$ then $B \setminus A \in \mathscr{D}$ for all $A, B \in \mathscr{D}$;

(iii) if $(A_n, n \in \mathbb{N})$ is an increasing sequence in $\mathscr{D}$ then $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{D}$.

0.2.2 Theorem (Dynkin's Lemma).
Let Π be a π-system and $\mathscr{D}$ be a d-system containing Π. Then $\sigma(\Pi) \subseteq \mathscr{D}$.

0.2.3 Theorem. *Let $(E, \mathscr{A})$ be a measurable space and μ_1, μ_2 be two measures on $(E, \mathscr{A})$ such that $\mu_1(E) = \mu_2(E) < \infty$. Let Π be a π-system such that $\sigma(\Pi) = \mathscr{A}$. If $\mu_1(A) = \mu_2(A)$ for all $A \in \Pi$ then $\mu_1 = \mu_2$ on $\mathscr{A}$.*

0.2.4 Example. Take $(E, \mathscr{A}) = (\mathbb{R}, \mathscr{B}_\mathbb{R})$. Then $\Pi = \{(-\infty, a], a \in \mathbb{Q}\}$ is a π-system that generates $\mathscr{B}_\mathbb{R}$. Thus, a finite measure on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ is entirely determined by its values on $\{(-\infty, a], a \in \mathbb{Q}\}$.

0.3 Product measures

Let $(E, \mathscr{A}, \mu)$ and $(F, \mathscr{B}, \nu)$ be finite measure spaces.

0.3.1 Definition. The *product σ-algebra* on $E \times F$ is

$$\mathscr{A} \otimes \mathscr{B} = \sigma(\{A \times B, A \in \mathscr{A}, B \in \mathscr{B}\}).$$

0.3.2 Theorem. *There is a unique measure (the product measure) $\rho = \mu \otimes \nu$ on $(E \times F, \mathscr{A} \otimes \mathscr{B})$ such that $\rho(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathscr{A}$ and $B \in \mathscr{B}$.*

0.3.3 Theorem (Fubini).

(i) *Let $f : E \times F \to \mathbb{R}_+$ be measurable with respect to the product σ-algebra. Then for all y, $x \mapsto f(x,y)$ and for all x, $y \mapsto f(x,y)$ are measurable. Let $\varphi_E^f = \int f(x,y)\mu(dx)$ and $\varphi_F^f = \int f(x,y)\nu(dy)$. These functions are measurable and*

$$\int f d(\mu \otimes \nu) = \int \varphi_E^f \nu(dy) = \int \varphi_F^f \mu(dx)$$

(ii) *If $f : E \times F \to \mathbb{R}$ is integrable with respect to $\mu \otimes \nu$ then the above results hold and φ_E^f and φ_F^f are integrable.*

0.4 L^p spaces

Let $(E, \mathscr{A}, \mu)$ be a measure space and $p \in [1, \infty)$. Define

$$\mathscr{L}^p(E, \mathscr{A}, \mu) = \{f \text{ measurable}, \int_E |f|^p d\mu < \infty\}$$

and

$$\mathscr{L}^\infty(E, \mathscr{A}, \mu) = \{f \text{ measurable}, \exists M \geq 0, \mu(|f| > M) = 0\}.$$

These are vector spaces and we let

$$\|f\|_p = \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}}$$

and

$$\|f\|_\infty = \text{esssup}|f| := \inf\{M, \mu(|f| > M) = 0\}.$$

These are not norms on the $\mathscr{L}^p$ spaces, since $\|f\|_p = 0$ only implies that f is zero a.e. We let $f \equiv g$ if $f = g$ a.e. Let L^p be $\mathscr{L}^p$ modulo these equivalence classes.

0.4.1 Theorem. *For all $p \in [1, \infty]$, L^p is a Banach space. Further, L^2 is a Hilbert space with inner product*

$$\langle f, g \rangle = \int_E f g d\mu$$

One idea that we will need soon is that if $H \subseteq L^2$ is a closed vector subspace of then we can project orthogonally onto it.

0.4.2 Theorem. *If $f \in L^2$ then there exists a unique $g \in H$ such that $\langle h, f - g \rangle = 0$ for all $h \in H$. In fact g is characterized as the unique $g \in H$ such that $\|f - g\|_2 = \inf_{h \in H} \|f - h\|_2$.*

0.5 Further results from elementary probability

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

0.5.1 Theorem.

(i) *Markov's inequality: If $X \in L^1$ and $a > 0$ then*

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a} \mathbb{E}[|X|].$$

(ii) Chebyshev's inequality: If $X \in L^2$ *and* $a > 0$ *then*

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a^2}\,\mathbb{E}[X^2].$$

0.5.2 Theorem (First Borel-Cantelli Lemma).
Let $(A_n, n \geq 1)$ *be a sequence of events such that* $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. *Then* $\mathbb{P}(\limsup_n A_n) = \mathbb{P}(A_n, \text{i.o.}) = 0$.

0.5.3 Theorem (Second Borel-Cantelli Lemma).
Let $(A_n, n \geq 1)$ *be a sequence of independent events such that* $\sum_{n\geq 1} \mathbb{P}(A_n)$ *diverges to infinity. Then* $\mathbb{P}(\limsup_n A_n) = \mathbb{P}(A_n, \text{i.o.}) = 1$.

1 Conditional Expectation

1.1 The 'discrete' case

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

1.1.1 Definition. Let $A, B \in \mathscr{F}$ and suppose that $\mathbb{P}(B) > 0$. The *conditional probability* of A with respect to B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

The conditional probability is interpreted as the probability of event A happening given that event B has happened.

1.1.2 Definition. More generally, if $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ then the *conditional expectation* of X given B is

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}(X\mathbf{1}_B)}{\mathbb{P}(B)}.$$

1.1.3 Definition. Let $(B_i, i \geq 1)$ be a partition of Ω such that $B_i \in \mathscr{F}$ for all i. Let $\mathscr{G} = \sigma(B_i, i \geq 1)$ and for $X \in L^1$ define the *conditional expectation* of X with respect to $\mathscr{G}$ to be

$$\mathbb{E}[X \mid \mathscr{G}] = \sum_{i\geq 0} \mathbb{E}[X \mid B_i]\mathbf{1}_{B_i},$$

with the convention that $\mathbb{E}[X \mid B_i] = 0$ if $\mathbb{P}(B_i) = 0$.

1.1.4 Lemma. *The conditional expectation with respect to* $\mathscr{G}$ *satisfies*

(i) $\mathbb{E}[X \mid \mathscr{G}]$ *is* $\mathscr{G}$*-measurable;*

(ii) $\mathbb{E}[X \mid \mathscr{G}] \in L^1(\Omega, \mathscr{G}, \mathbb{P})$*; and*

(iii) $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]\mathbf{1}_A]$ *for every* $A \in \mathscr{G}$.

PROOF: Exercise. □

1.1.5 Example. Let $X \in L^1$ and let Y be a r.v. taking values in some countable set E. Then $\Omega = \bigcup_{y \in E}\{Y = y\}$ is a partition of Ω which generates $\sigma(Y)$, and

$$\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)] = \sum_{y \in E} \mathbb{E}[X \mid \{Y = y\}]\mathbf{1}_{\{Y=y\}}.$$

Remember that r.v.'s and conditional expectations are only defined up to a set of measure zero.

1.2 Conditioning with respect to a σ-algebra

We now define the *conditional expectation* with respect to a sub-σ-algebra so that it satisfies the same properties as conditional expectation with respect to $\mathscr{G} = \sigma(B_i, i \geq 0)$. The following definition and theorem are due to Kolmogorov.

1.2.1 Theorem. *Let $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ be a sub-σ-algebra. Then there exists an integrable r.v. X' such that*

(i) X' is $\mathscr{G}$-measurable;

(ii) $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X'\mathbf{1}_A]$ for all $A \in \mathscr{G}$.

Moreover, if X'' is another such r.v. then $X' = X''$ a.s.

We denote the unique element of $L^1(\Omega, \mathscr{G}, \mathbb{P})$ satisfying the properties (i) and (ii) by $\mathbb{E}[X \mid \mathscr{G}]$. Further, we may replace (ii) with the following slightly more general property.

(ii') $\mathbb{E}[ZX] = \mathbb{E}[ZX']$ for every bounded $\mathscr{G}$-measurable r.v. Z.

To prove (ii') from (i) and (ii), first prove it for simple functions Z and then approximate general Z.

PROOF: *Uniqueness:* Suppose X' and X'' both satisfy (i) and (ii). Let $A = \{X' > X''\} \in \mathscr{G}$. Then

$$\mathbb{E}[X'\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X''\mathbf{1}_A]$$

so $\mathbb{E}[(X' - X'')\mathbf{1}_{\{X'>X''\}}] = 0$, which implies that $X' \leq X''$ a.s. The same argument with the reverse inequality proves that $X' = X''$ a.s. □

PROOF: *Existence:* First assume that $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$. Now $L^2(\Omega, \mathscr{G}, \mathbb{P})$ is a closed vector subspace of $L^2(\Omega, \mathscr{F}, \mathbb{P})$, so there is a unique r.v. X' such that $\mathbb{E}[Z(X - X')] = 0$ for all $Z \in L^2(\Omega, \mathscr{G}, \mathbb{P})$, namely the orthogonal projection.

It follows that

$$\mathbb{E}[\cdot \mid \mathscr{G}] : L^2(\Omega, \mathscr{F}, \mathbb{P}) \to L^2(\Omega, \mathscr{G}, \mathbb{P}) : X \mapsto \mathbb{E}[X \mid \mathscr{G}]$$

is simply orthogonal projection, and $\mathbb{E}[X \mid \mathscr{G}]$ is the $\mathscr{G}$-measurable r.v. that best approximates X for the L^2-norm (i.e. $\mathbb{E}[|X - \mathbb{E}[X \mid \mathscr{G}]|^2]$ is minimal over $\{\mathbb{E}[|X - Z|^2] \mid Z \in L^2(\Omega, \mathscr{G}, \mathbb{P})\}$).

Note also that $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$ (since $\mathbf{1} \in L^2$), and if $X \geq 0$ then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$. Indeed,

$$0 \geq \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbf{1}_{\mathbb{E}[x|\mathcal{G}]<0}] = \mathbb{E}[X\mathbf{1}_{\mathbb{E}[x|\mathcal{G}]<0}] \geq 0$$

so $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ a.s.

Now assume that X is a non-negative r.v. (and not necessarily integrable). For all $n \geq 1$, $X \wedge n \in L^2$, and $X \wedge n \nearrow_{n\to\infty} X$ pointwise. Note that $(\mathbb{E}[X \wedge n \mid \mathcal{G}], n \geq 1)$ is an increasing sequence since

$$X \wedge n - X \wedge (n-1) \geq 0$$

and $\mathbb{E}[\cdot \mid \mathcal{G}]$ is linear on L^2. Therefore let $X' = \lim_{n\to\infty} \mathbb{E}[X \wedge n \mid \mathcal{G}]$, which is $\mathcal{G}$-measurable. For any $A \in \mathcal{G}$, as $n \to \infty$, by MCT,

$$\begin{array}{ccc} \mathbb{E}[\mathbf{1}_A X \wedge n] & \longrightarrow & \mathbb{E}[\mathbf{1}_A X] \\ \| & & \| \\ \mathbb{E}[\mathbf{1}_A \mathbb{E}[X \wedge n \mid \mathcal{G}]] & \longrightarrow & \mathbb{E}[\mathbf{1}_A X'] \end{array}$$

Therefore $\mathbb{E}[X \mid \mathcal{G}]$ is X' by uniqueness. Taking $A = \Omega$ shows that if X is integrable then so is $\mathbb{E}[X \mid \mathcal{G}]$.

For any $X \in L^1$, we may write $X = X^+ - X^-$, where $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$. Let

$$X' = \mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}]$$

which is integrable and $\mathcal{G}$-measurable. Finally, for all $A \in \mathcal{G}$,

$$\mathbb{E}[\mathbf{1}_A X'] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X^+ \mid \mathcal{G}]] - \mathbb{E}[\mathbf{1}_A \mathbb{E}[X^- \mid \mathcal{G}]] = \mathbb{E}[\mathbf{1}_A(X^+ - X^-)] = \mathbb{E}[\mathbf{1}_A X],$$

so $\mathbb{E}[X \mid \mathcal{G}]$ is X' by uniqueness. □

We also proved that for all non-negative r.v.'s X there is a non-negative, $\mathcal{G}$-measurable r.v. X' such that $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X'\mathbf{1}_A]$ for all $A \in \mathcal{G}$.

1.2.2 Theorem. *Let X be a non-negative r.v. and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Then there exists a non-negative r.v. X' such that*

(i) X' is $\mathcal{G}$-measurable;

(ii) $\mathbb{E}[X'Z] = \mathbb{E}[XZ]$ for every non-negative $\mathcal{G}$-measurable r.v. Z.

Moreover, if X'' is another such r.v. then $X' = X''$ a.s.

We denote by $\mathbb{E}[X \mid \mathcal{G}]$ the class of X' up to equality a.s.

PROOF: Exercise. □

1.2.3 Proposition.

(i) If $X \geq 0$ then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ a.s.

(ii) If $X, Y \in L^1$ and $a, b \in \mathbb{R}$ then $\mathbb{E}[aX + bY \mid \mathcal{G}] = a\,\mathbb{E}[X \mid \mathcal{G}] + b\,\mathbb{E}[Y \mid \mathcal{G}]$.

(iii) If $X \in L^1$ then $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$.

(iv) If X is $\mathscr{G}$-measurable then $\mathbb{E}[X \mid \mathscr{G}] = X$.

(v) If $X \in L^1$ then $|\mathbb{E}[X \mid \mathscr{G}]| \le \mathbb{E}[|X| \mid \mathscr{G}]$, which implies that $\mathbb{E}[|\mathbb{E}[X \mid \mathscr{G}]|] \le \mathbb{E}[|X|]$.

(vi) If X is independent of $\mathscr{G}$ then $\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}[X]$.

PROOF: Exercise □

1.3 Conditional convergence theorems

Let $\mathscr{G} \subseteq \mathscr{F}$ be a sub-σ-algebra.

1.3.1 Theorem (Conditional MCT).
Let $X_n \ge 0$ $(n \ge 0)$ be such that $X_n \le X_{n+1}$ for all n and let $X = \lim_{n\to\infty} X_n$ a.s. Then

$$\mathbb{E}[X_n \mid \mathscr{G}] \nearrow_{n\to\infty} \mathbb{E}[X \mid \mathscr{G}] \quad \text{a.s.}$$

PROOF: Note that if $0 \le X \le Y$ then $0 \le \mathbb{E}[X \mid \mathscr{G}] \le \mathbb{E}[Y \mid \mathscr{G}]$ (exercise). Let $X_n' = \mathbb{E}[X_n \mid \mathscr{G}])$. Then X_n' increases with n, so let $X' = \lim_{n\to\infty} X_n'$.

Let $A \in \mathscr{G}$. Then

$$\begin{array}{ccc} \mathbb{E}[X_n \mathbf{1}_A] & \xrightarrow{n\to\infty} & \mathbb{E}[X\mathbf{1}_A] \\ \| & & \\ \mathbb{E}[X_n' \mathbf{1}_A] & \xrightarrow{n\to\infty} & \mathbb{E}[X'\mathbf{1}_A] \end{array}$$

by the MCT, so $\mathbb{E}[X'\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ for all $A \in \mathscr{G}$ and $X' = \mathbb{E}[X \mid \mathscr{G}]$ by uniqueness. □

1.3.2 Theorem (Conditional Fatou Lemma).
Let $X_n \ge 0$ $(n \ge 0)$. Then

$$\mathbb{E}[\liminf_{n\to\infty} X_n \mid \mathscr{G}] \le \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{G}] \quad \text{a.s.}$$

PROOF: Exercise (recall the proof of the usual Fatou's Lemma). □

1.3.3 Theorem (Conditional DCT).
Let X_n $(n \ge 1)$ $X = \lim_{n\to\infty} X_n$ a.s., and assume that $\sup_n |X_n| \le Y$ for some integral r.v. Y. Then $\mathbb{E}[X \mid \mathscr{G}]$ is integrable and

$$\mathbb{E}[X_n \mid \mathscr{G}] \xrightarrow{n\to\infty} \mathbb{E}[X \mid \mathscr{G}] \quad \text{a.s.}$$

PROOF: Consider $Y - X_n, Y + X_n \ge 0$. By (ii),

$$\mathbb{E}[Y - X \mid \mathscr{G}] \le \liminf_{n\to\infty} (\mathbb{E}[Y \mid \mathscr{G}] - \mathbb{E}[X_n \mid \mathscr{G}])$$

and

$$\mathbb{E}[Y + X \mid \mathscr{G}] \le \liminf_{n\to\infty} (\mathbb{E}[Y \mid \mathscr{G}] + \mathbb{E}[X_n \mid \mathscr{G}]).$$

Subtracting these we get

$$\liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{G}] \ge \mathbb{E}[X \mid \mathscr{G}] \ge \limsup_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{G}].$$

□

1.3.4 Proposition (Conditional Jensen inequality).
Let $c : \mathbb{R} \to \mathbb{R}$ be a convex function, $X \in L^1$, and assume that $c(X) \in L^1$ or $c \geq 0$. Then $c(\mathbb{E}[X \mid \mathscr{G}]) \leq \mathbb{E}[c(X) \mid \mathscr{G}]$.

PROOF: Let $E = \{(a, b) \mid ax + b \leq c(x) \text{ for all } x \in \mathbb{R}\}$. Then for all x, $c(x) = \sup_{(a,b)\in E; a,b\in\mathbb{Q}} ax + b$, so

$$\begin{aligned} c(\mathbb{E}[X \mid \mathscr{G}]) &= \sup_{(a,b)\in E;a,b\in\mathbb{Q}} a\,\mathbb{E}[X \mid \mathscr{G}] + b \\ &= \sup_{(a,b)\in E;a,b\in\mathbb{Q}} \mathbb{E}[aX + b \mid \mathscr{G}] \\ &\leq \mathbb{E}\left[\sup_{(a,b)\in E;a,b\in\mathbb{Q}} (aX + b) \mid \mathscr{G} \right] \\ &= \mathbb{E}[c(X) \mid \mathscr{G}] \end{aligned}$$

(Show $\sup_{n\geq 1} \mathbb{E}[X_n] \leq \mathbb{E}[\sup_{n\geq 1} X_n]$ to complete the proof.) □

1.3.5 Proposition. *$X \mapsto \mathbb{E}[X \mid \mathscr{G}]$ is a continuous linear operator on L^p of (operator) norm at most one.*

PROOF: $x \mapsto |x|^p$ is non-negative and convex, so

$$\|\mathbb{E}[X \mid \mathscr{G}]\|_p^p = \mathbb{E}[|\mathbb{E}[X \mid \mathscr{G}]|^p] \leq \mathbb{E}[\mathbb{E}[|X|^p \mid \mathscr{G}]] = \mathbb{E}[|X|^p] = \|X\|_p^p. \qquad \square$$

1.4 Specific properties of conditional expectation

1.4.1 Proposition. *Let $\mathscr{G} \subseteq \mathscr{F}$ be a sub-σ-algebra and let X and Y be r.v.'s, and assume that Y is $\mathscr{G}$-measurable, and either $X, Y, XY \in L^1$ or $X, Y \geq 0$. Then $\mathbb{E}[YX \mid \mathscr{G}] = Y\,\mathbb{E}[X \mid \mathscr{G}]$*

This proposition is known as "taking out what is known" for obvious reasons.

PROOF: Assume that $X, Y \geq 0$, and take $A \in \mathscr{G}$. Then

$$\mathbb{E}[\mathbf{1}_A \mathbb{E}[YX \mid \mathscr{G}]] = \mathbb{E}[\mathbf{1}_A YX] = \mathbb{E}[\mathbf{1}_A Y\,\mathbb{E}[X \mid \mathscr{G}]],$$

so $\mathbb{E}[YX \mid \mathscr{G}] = Y\,\mathbb{E}[X \mid \mathscr{G}]$ by uniqueness. The L^1 case follows analogously. □

1.4.2 Proposition (Tower property).
Let $\mathscr{H} \subseteq \mathscr{G} \subseteq \mathscr{F}$ be sub-σ-algebras and $X \in L^1$. Then $\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}] = \mathbb{E}[X \mid \mathscr{H}]$.

PROOF: Let $A \in \mathscr{H}$. Then

$$\mathbb{E}[\mathbf{1}_A \mathbb{E}[X \mid \mathscr{H}]] = \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X \mid \mathscr{G}]] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}],$$

so $\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}] = \mathbb{E}[X \mid \mathscr{H}]$ by uniqueness. □

1.4.3 Proposition. *Let $\mathscr{G}, \mathscr{H} \subseteq \mathscr{F}$ be sub-σ-algebras and $X \in L^1$. Assume that $\mathscr{H}$ is independent of $\sigma(X) \vee \mathscr{G}$. Then $\mathbb{E}[X \mid \mathscr{G} \vee \mathscr{H}] = \mathbb{E}[X \mid \mathscr{G}]$.*

PROOF: Let $A \in \mathscr{G}$ and $B \in \mathscr{H}$. Then

$$\begin{aligned}\mathbb{E}[\mathbf{1}_A \mathbf{1}_B \, \mathbb{E}[X \mid \mathscr{G} \vee \mathscr{H}]] &= \mathbb{E}[\mathbf{1}_A \mathbf{1}_B X] = \mathbb{E}[\mathbf{1}_B \, \mathbb{E}[\mathbf{1}_A X \mid \mathscr{H}]] \\ &= \mathbb{P}(B)\,\mathbb{E}[\mathbf{1}_A X] = \mathbb{P}(B)\,\mathbb{E}[\mathbf{1}_A X \mid \mathscr{G}] = \mathbb{E}[\mathbf{1}_A \mathbf{1}_B \, \mathbb{E}[X \mid \mathscr{G}]]\end{aligned}$$

Independence was used for the third and last equalities. We are done by the monotone class theorem, since $\{A \cap B \mid A \in \mathscr{G}, B \in \mathscr{H}\}$ is a π-system that generates $\mathscr{G} \vee \mathscr{H}$ (see Williams). □

1.4.4 Proposition. *Let X and Y be r.v.'s and g be a non-negative measurable function. Let $\mathscr{G} \subseteq \mathscr{F}$ be a sub-σ-algebra, and assume that Y is $\mathscr{G}$-measurable and $\sigma(X)$ is independent of $\mathscr{G}$. Then*

$$\mathbb{E}[g(X,Y) \mid \mathscr{G}] = \int P_X(dx) g(x,Y).$$

Here $P_X = \mathscr{L}(X)$ is the *law of* X, i.e. $P_X(A) := \mathbb{P}(X \in A)$.

PROOF: Let Z be $\mathscr{G}$-measurable. Then

$$\mathbb{E}[Zg(X,Y)] = \int P_{(X,Y,Z)}(dx,dy,dz) z g(x,y)$$

But X is independent of (Z,Y), so

$$P_{(X,Y,Z)}(dx,dy,dz) = P_X(dx) \otimes P_{(Y,Z)}(dy,dz).$$

By Fubini's Theorem,

$$\begin{aligned}\int P_{(X,Y,Z)}(dx,dy,dz) z g(x,y) &= \int P_X(dx) \int P_{(Y,Z)}(dy,dz) z g(x,y) \\ &= \int P_X(dx) \mathbb{E}[Zg(x,Y)] \\ &= \mathbb{E}\left[Z \int P_X(dx) g(x,Y) \right]\end{aligned}$$

Since $\int P_X(dx) g(x,Y)$ is $\mathscr{G}$-measurable, it is $\mathbb{E}[g(X,Y) \mid \mathscr{G}]$ by uniqueness. □

1.5 Explicit computations of conditional expectation

1.5.1 Example (Conditional density functions). Let (X,Y) be a random vector in $\mathbb{R}^2$ such that

$$P_{(X,Y)}(dx,dy) = f_{(X,Y)}(x,y)dxdy,$$

where $dxdy$ is is Lebesgue measure on $\mathbb{R}^2$. Let $h \geq 0$. We want to compute $\mathbb{E}[h(X) \mid Y]$. To do this we must compute

$$\mathbb{E}[h(X)g(Y)] = \int h(x)g(y)f_{(X,Y)}(x,y)dxdy$$

for $g \geq 0$. Let

$$f_Y(y) := \int_{x\in\mathbb{R}} f_{(X,Y)}(x,y)dx$$

be the density of Y at y. Then

$$\begin{aligned}\mathbb{E}[h(X)g(Y)] &= \int g(y)f_Y(y)dy \int h(x)\frac{f_{(X,Y)}(x,y)}{f_Y(y)}dx\mathbf{1}_{\{f_Y(y)>0\}} \\ &= \mathbb{E}\left[g(Y)\int h(x)f_{X|Y}(x \mid Y)dx\right]\end{aligned}$$

where

$$f_{X|Y}(x,y) := \frac{f_{(X,Y)}(x,y)}{f_Y(y)}\mathbf{1}_{\{f_Y(y)>0\}},$$

so $\mathbb{E}[h(X) \mid Y] = \int h(x)f_{X|Y}(x \mid Y)dx$ since the latter is $\sigma(Y)$-measurable.

Let $\nu(Y,dx)$ be the measure with density $f_{X|Y}(x \mid Y)$ with respect to Lebesgue measure dx. Then $\mathbb{E}[h(X) \mid Y] = \int_{x\in\mathbb{R}} h(x)\nu(Y,dx)$. When we have such a formula, we say that $\nu(Y,dx)$ is the *conditional distribution* of X given Y. Sometimes $\nu(y,dx)$ is called the conditional distribution of X given $Y = y$. These are defined only up a set of zero measure for y. We also say that $f_{X|Y}(x \mid y)$ is the *conditional density function* of X given $Y = y$.

1.5.2 Example (Gaussian case). Let (X,Y) be a Gaussian random vector in $\mathbb{R}^2$ (so $\lambda X + \mu Y$ is Gaussian for all $\lambda,\mu \in \mathbb{R}$). What is $\mathbb{E}[X \mid Y]$? Let $X' = aY + b$, where a and b are chosen so that

(i) $\mathrm{Cov}(X - X', Y) = 0$; and

(ii) $\mathbb{E}[X'] = \mathbb{E}[X]$.

In this case $X - X'$ and Y are independent (this is a property of Gaussian random vectors). Then

$$\mathbb{E}[(X - X')\mathbf{1}_A(Y)] = \mathbb{E}[X - X']\,\mathbb{P}(Y \in A) = 0,$$

so $X' = \mathbb{E}[X \mid Y]$. Notice that $a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}$, so

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathbb{E}[Y]).$$

1.5.3 Exercise. What is the conditional distribution of X given Y? (Hint:

$$\mathbb{E}[h(X) \mid Y] = \mathbb{E}[h(X - \mathbb{E}[X \mid Y] + \mathbb{E}[X \mid Y]) \mid Y],$$

and $X - \mathbb{E}[X \mid Y]$ and $\mathbb{E}[X \mid Y]$ are independent.)

2 Discrete-time Martingales

2.1 General notions on random processes

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $I \subseteq \mathbb{R}$ be the set of times on which we will define a process.

2.1.1 Definition. A *stochastic process* (or *random process*) X indexed by I is a family $(X_t, t \in I)$ of r.v.'s. If X_t is $\mathbb{R}$-valued then X is *integrable* if $X_t \in L^1$ for every $t \in I$.

2.1.2 Definition. A *filtration* is an increasing family of sub-σ-algebras of $\mathscr{F}$ indexed by I, $(\mathscr{F}_t, t \in I)$, such that if $s \le t$ then $\mathscr{F}_s \subseteq \mathscr{F}_t$.

We think of $\mathscr{F}_t$ as "the information available at and before time t." If $(\mathscr{F}_t, t \in I)$ is a filtration, we say that $(\Omega, \mathscr{F}, (\mathscr{F}_t, t \in I), \mathbb{P})$ is a *filtered probability space* (or *f.p.s.*).

2.1.3 Definition. A random process $(X_t, t \in I)$ is *adapted* to the filtration $(\mathscr{F}_t, t \in I)$ if X_t is $\mathscr{F}_t$-measurable for each $t \in I$.

2.1.4 Example. Let (X_t) be a process, and let $\mathscr{F}_t^X = \sigma\{X_s \mid s \le t\}$. Then $(\mathscr{F}_t^X)$ is the *natural filtration* of X. It is the smallest filtration with respect to which X is adapted.

For the remainder of these notes let $(\Omega, \mathscr{F}, (\mathscr{F}_t, t \in I), \mathbb{P})$ be an f.p.s.

2.1.5 Definition. Let X be an adpated real-valued integrable process. Then X is a

(i) *martingale* if $\mathbb{E}[X_t \mid \mathscr{F}_s] = X_s$ for all $s \le t \in I$;

(ii) *super-martingale* if $\mathbb{E}[X_t \mid \mathscr{F}_s] \le X_s$ for all $s \le t \in I$;

(iii) *sub-martingale* if $\mathbb{E}[X_t \mid \mathscr{F}_s] \ge X_s$ for all $s \le t \in I$.

In particular, if X is a

(i) martingale then $\mathbb{E}[X_t] = \mathbb{E}[X_s]$ for all $s, t, \in I$;

(ii) super-martingale then $\mathbb{E}[X_t] \le \mathbb{E}[X_s]$ for all $s \le t \in I$;

(iii) sub-martingale then $\mathbb{E}[X_t] \ge \mathbb{E}[X_s]$ for all $s \le t \in I$.

2.1.6 Example. Let $I = \mathbb{Z}_+$ and $X_1, X_2, \dots$ be i.i.d. integrable r.v.'s. Take $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\mathscr{F}_n = \sigma(X_m, m \le n)$. Let $S_n = X_1 + \dots + X_n$. Then S is a martingale if and only if $\mathbb{E}[X_1] = 0$, since for all n,

$$\mathbb{E}[S_{n+1} \mid \mathscr{F}_n] = S_n + \mathbb{E}[X_n] = S_n + \mathbb{E}[X_1].$$

Similarily, it is a super-martingale if $\mathbb{E}[X_1] \le 0$ or a sub-martingale if $\mathbb{E}[X_1] \ge 0$.

2.1.7 Definition (Doob). Let $T : \Omega \to I \cup \{+\infty\}$ be a random variable. T is a *stopping time* with respect to the filtration $(\mathscr{F}_t, t \in I)$ if $\{T \le t\} \in \mathscr{F}_t$ for all $t \in I$.

Note that constant r.v.'s $T \equiv t$ for some $t \in I$ are stopping times.

2.1.8 Example. Let $I = \mathbb{Z}_+$, $(X_n, n \ge 0)$ be a random process, and $(\mathscr{F}_n) = (\mathscr{F}_n^X)$. Let A be a Borel subset of $\mathbb{R}$ and $T_A(\omega) = \inf\{n \ge 0 \mid X_n(\omega) \in A\}$. Then T_A is a stopping time since

$$\{T_A \le n\} = \bigcup_{m \le n} \{X_m \in A\}.$$

T_A is known as the *first entrance time* into A and is an important example of a stopping time. However, $L_A(\omega) = \sup\{N \ge n \ge 0 \mid X_n(\omega) \in A\}$ is *not* a stopping time in general.

Remark. If I is countable (for example, if $I = \mathbb{Z}_+$) then T is a stopping time with respect to $(\mathscr{F}_t)$ if and only if $\{T = t\} \in \mathscr{F}_t$ for all $t \in I$. *This is not true in general.*

2.1.9 Proposition. *Let S, T, and $(T_n, n \geq 0)$ all be stopping times. Then $S \wedge T$, $S \vee T$, $\inf_{n\geq 0} T_n$, $\sup_{n\geq 0} T_n$, $\liminf_n T_n$, and $\limsup_n T_n$ are all stopping times as well.*

PROOF: Exercise. □

2.1.10 Definition. Let T be stopping time with respect to $(\mathscr{F}_t, t \in I)$, and

$$\mathscr{F}_T = \{A \in \mathscr{F} \mid A \cap \{T \leq t\} \in \mathscr{F}_t \text{ for all } t \in I\}.$$

This defines a σ-algebra, called the σ-algebra of *measurable events before T*.

2.1.11 Exercises.

(i) If S and T are stopping times such that $S \leq T$ then $\mathscr{F}_S \subseteq \mathscr{F}_T$.

(ii) Y is an $\mathscr{F}_T$-measurable r.v. if and only if $Y\mathbf{1}_{\{T\leq t\}}$ is $\mathscr{F}_t$-measurable for all $t \in I$. (Hint: begin with $Y = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$, $A_i \in \mathscr{F}_T$.)

2.1.12 Proposition. *Let $I \subseteq \mathbb{R}$ be countable. Let $(X_t, t \in I)$ be adapted, and let T be a stopping time. Define X_T by $X_T(\omega) = X_{T(\omega)}(\omega)$ on the event $\{T < \infty\}$. Then*

(i) $X_T\mathbf{1}_{\{T<\infty\}}$ is $\mathscr{F}_T$-measurable; and

(ii) the stopped process at time T, $X^T := (X_{t\wedge T}, t \in I)$ is adapted.

PROOF: We have

$$X_T\mathbf{1}_{\{T<\infty\}}\mathbf{1}_{\{T\leq t\}} = \sum_{s\leq t, s\in I} X_s\mathbf{1}_{\{T=s\}},$$

which is $\mathscr{F}_t$-measurable since $X_s\mathbf{1}_{\{T=s\}}$ is $\mathscr{F}_s$-measurable and the sum is countable.

X^T is adapted because

$$X_t^T = X_{T\wedge t} = X_t X_s \mathbf{1}_{\{T>t\}} + \sum_{s\leq t, s\in I} X_{s\wedge t} X_s \mathbf{1}_{\{T=s\}}$$

for every $t \in I$. □

For the time being we consider only the case $I = \mathbb{Z}_+$ and the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n, n \geq 0), \mathbb{P})$.

2.1.13 Proposition. *Let $X = (X_n, n \geq 0)$ be adapted and T a stopping time. If X is integrable then the stopped process X^T is also integrable.*

PROOF: We have

$$X_n^T = X_n\mathbf{1}_{\{T>n\}} + \sum_{m\leq n} X_m\mathbf{1}_{\{T=m\}},$$

which is a finite sum, so

$$\mathbb{E}[|X_n^T|] \leq \mathbb{E}[|X_n|] + \sum_{m\leq n} \mathbb{E}[|X_m|] < \infty.$$

□

Check that an adapted integrable process $(X_n, n \geq 0)$, is a $(\mathscr{F}_n)$-martingale if $\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] = X_n$ for all n. (Hint: use the tower property.) Similarly, we have a super- or sub-martingale when this relation holds with "$\leq$" or "$\geq$" respectively.

2.2 Optional stopping, part I

2.2.1 Definition. Let $(C_n, n \geq 1)$ and $(X_n, n \geq 0)$ be processes taking values in $\mathbb{R}$. Define

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

for $n \geq 1$, and $(C \cdot X)_0 = 0$. Then $C \cdot X$ is sometimes called a *discrete stochastic integral*. We say that $C = (C_n, n \geq 1)$ is *previsible* if C_n is $\mathscr{F}_{n-1}$-measurable for all $n \geq 1$.

2.2.2 Proposition. *Let $(X_n, n \geq 0)$ be a martingale (resp. super-martingale) and let $(C_n, n \geq 1)$ be a bounded, previsible process (resp. bounded, previsible, non-negative). Then the process $((C \cdot X)_n, n \geq 0)$ is a martingale (resp. super-martingale).*

PROOF: $C \cdot X$ is adapted since $\sum_{k=1}^{n} C_k (X_k - X_{k-1})$ is $\mathscr{F}_n$-measurable for every n. Therefore it is integrable since the C_k's are uniformly bounded and the X_k's are integrable.

$$\begin{aligned}\mathbb{E}[(C \cdot X)_{n+1} \mid \mathscr{F}_n] &= \mathbb{E}[(C \cdot X)_n + C_{n+1}(X_{n+1} - X_n) \mid \mathscr{F}_n] \\ &= (C \cdot X)_n + C_{n+1}\, \mathbb{E}[X_{n+1} - X_n \mid \mathscr{F}_n] \\ &= (C \cdot X)_n\end{aligned}$$

by the martingale property. Check the result for super-martingales. □

The above result shows that there is no way to make money out of super-martingales.

2.2.3 Theorem (Optional Stopping, discrete-time).
Let T be a stopping time and let $(X_n, n \geq 0)$ be a martingale (resp. super-martingale). Then X^T is also a martingale (resp. super-martingale).

PROOF: Let $C_n = \mathbf{1}_{\{n \leq T\}}$ for $n \geq 0$. Then $(C_n, n \geq 1)$ is previsible since $\{n \leq T\} = \{T \leq n-1\}^C \in \mathscr{F}_{n-1}$. It is also bounded, integrable, and non-negative, so by the previous proposition $C \cdot X$ is a martingale (resp. super-martingale).

$$(C \cdot X)_n = \sum_{k=1}^{n} \mathbf{1}_{\{k \leq T\}} (X_k - X_{k-1}) = \sum_{k=1}^{n \wedge T} (X_k - X_{k-1}) = X_{n \wedge T} - X_0 = X_n^T - X_0,$$

so X^T is a martingale (resp. super-martingale). □

2.2.4 Proposition (Optional Stopping, bounded times).
Let $(X_n, n \geq 0)$ be a martingale and S and T be bounded stopping times such that $S \leq T \leq K$, where K is a fixed constant. Then $\mathbb{E}[X_T \mid \mathscr{F}_S] = X_S$. In particular, $\mathbb{E}[X_T] = \mathbb{E}[X_S]$.

Warning: This is not true in general (for unbounded stopping times). Consider i.i.d. r.v.'s $X_1, \ldots, X_n$ such that

$$\mathbb{P}(X_n = 1) = \frac{1}{2} = \mathbb{P}(X_n = -1),$$

and let $S_n = X_1 + \cdots + X_n$, which we have seen to be a martingale. Let $T = \inf\{n \geq 1 \mid S_n = 1\}$, a stopping time. It holds that $\mathbb{P}(T < \infty) = 1$, but $\mathbb{E}[S_T] = 1 > 0 = \mathbb{E}[S_0]$.

PROOF: Let $A \in \mathscr{F}_S$, and consider $C_n = \mathbf{1}_A \mathbf{1}_{\{S<n\leq T\}}$. C is previsible since

$$C_n = \mathbf{1}_A \mathbf{1}_{\{S\leq n-1\}} \mathbf{1}_{\{n\leq T\}} = \mathbf{1}_{A\cap\{S\leq n-1\}} \mathbf{1}_{\{n\leq T\}}$$

is $\mathscr{F}_{n-1}$-measurable. Then $((C \cdot X)_n, n \geq 0)$ is a martingale, and we have

$$(C \cdot X)_K = \sum_{k=S+1}^{T} \mathbf{1}_A (X_k - X_{k-1}) = \mathbf{1}_A (X_T - X_S).$$

Since $(C \cdot X)$ is a martingale, $\mathbb{E}[(C \cdot X)_K] = \mathbb{E}[(C \cdot X)_0] = 0$. This says that $\mathbb{E}[\mathbf{1}_A X_T] = \mathbb{E}[\mathbf{1}_A X_S]$ for all $A \in \mathscr{F}_S$, which is the very definition of $\mathbb{E}[X_T \mid \mathscr{F}_S] = \mathbb{E}[X_S]$. □

2.2.5 Exercise. Check that for a super-martingale X and bounded stopping times $S \leq T$, we have $\mathbb{E}[X_T \mid \mathscr{F}_S] \leq X_S$ and $\mathbb{E}[X_T] \leq \mathbb{E}[X_S]$.

2.3 Martingale Convergence Theorem

2.3.1 Theorem (Martingale convergence theorem).
Let $(X_n, n \geq 0)$ be a super-martingale such that $\sup_n \mathbb{E}[|X_n|] < \infty$ (i.e. X is bounded in L^1). Then X_n converges a.s. to a finite limit X_∞ as $n \to \infty$.

2.3.2 Corollary. *If $(X_n, n \geq 0)$ is a non-negative super-martingale then X_n converges a.s. to a finite limit X_∞ as $n \to \infty$.*

PROOF: $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = \mathbb{E}[X_0] < \infty$ for all n, so $(X_n, n \geq 0)$ is bounded in L^1. □

2.3.3 Definition. Let $(x_n, n \geq 0)$ be a real sequence and $a < b \in \mathbb{R}$. Recursively define

(i) $S_1(x) := \inf\{n \geq 0 \mid x_n < a\} \in \mathbb{Z}_+ \cup \{\infty\}$;

(ii) $T_k(x) := \inf\{n \geq S_k(x) \mid x_n > b\}$ for $k \geq 1$;

(iii) $S_{k+1}(x) := \inf\{n \geq T_k(x) \mid x_n < a\}$ for $k \geq 1$;

(iv) $N_n(x, [a, b]) := \sup\{k \geq 1 \mid T_k(x) \leq n\}$;

(v) $N(x, [a, b]) := \sup\{k \geq 1 \mid T_k(x) \leq \infty\}$.

Notice that $N(x, [a, b]) = \nearrow_{n\to\infty} N_n(x, [a, b])$. A little explanation is in order. S_1 is the first time that the sequence drops below a, and T_1 is the first time after S_1 that the sequence exceeds b. T_1 is the time of the first *up-crossing*. Similarly, T_k is the time of the k^{th} up-crossing, and N_n is the number of up-crossings that have occurred before time n.

2.3.4 Lemma. *A real sequence $(x_n, n \geq 0)$ converges in $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ if and only if $N(x, [a, b]) < \infty$ for every $a < b \in \mathbb{Q}$.*

PROOF: Exercise. □

Observe that if $(X_n, n \geq 0)$ is an adapted process then $S_k(X)$ and $T_k(X)$ $(k \geq 1)$ are all stopping times.

2.3.5 Proposition (Doob's Up-crossing Lemma).
Let $(X_n, n \geq 0)$ *be a super-martingale. Then for all* $a < b$ *and all* $n \geq 0$*,*

$$(b-a)\mathbb{E}[N_n(X,[a,b])] \leq \mathbb{E}[(X_n - a)^-].$$

PROOF: For this proof we write S_k, T_k, and N_n for the r.v.'s $S_k(X)$, $T_k(X)$, and $N_n(X,[a,b])$, respectively. Let $C_n = \sum_{k\geq 1} \mathbf{1}_{\{S_k < n \leq T_k\}}$. Since $S_1 < T_1 < S_2 < T_2 < \cdots$, C_n takes values in $\{0,1\}$. Note that the process $(C_n, n \geq 1)$ is previsible (we have already seen that $\mathbf{1}_{\{S<n\leq T\}}$ is previsible when $S \leq T$ are stopping times). Therefore $((C\cdot X)_n, n \geq 0)$ is a super-martingale. Now

$$\begin{aligned}
(C\cdot X)_n &= \sum_{r=1}^{n} C_r(X_r - X_{r-1}) \\
&= \sum_{r=1}^{n}\sum_{k\geq 1} \mathbf{1}_{\{S_k<r\leq T_k\}}(X_r - X_{r-1}) \\
&= \sum_{r=1}^{n}\sum_{k=1}^{N_n} \mathbf{1}_{\{S_k<r\leq T_k\}}(X_r - X_{r-1}) \\
&= \sum_{k=1}^{N_n}\sum_{r=1}^{n} \mathbf{1}_{\{S_k<r\leq T_k\}}(X_r - X_{r-1}) \\
&= \mathbf{1}_{\{S_{N_n+1}\leq n\}}(X_n - X_{S_{N_n+1}}) + \sum_{k=1}^{N_n}(X_{T_k} - X_{S_k}) \\
&\geq \mathbf{1}_{\{S_{N_n+1}\leq n\}}(X_n - a) + (b-a)N_n
\end{aligned}$$

Now either $X_n \geq a$, in which case $\mathbf{1}_{\{S_{N_n+1}\leq n\}}(X_n - a) \geq 0$, or $X_n < a$, in which case $S_{N_n+1} \leq n$, so $\mathbf{1}_{\{S_{N_n+1}\leq n\}}(X_n - a) = (X_n - a)\mathbf{1}_{\{X_n<a\}} = -(X_n - a)^-$. Hence

$$(C\cdot X)_n \geq (b-a)N_n - (X_n - a)^-$$

but

$$0 = \mathbb{E}[(C\cdot X)_0] \geq \mathbb{E}[(C\cdot X)_n] \geq (b-a)\mathbb{E}[N_n] - \mathbb{E}[(X_n - a)^-] \qquad \square$$

PROOF (OF MARTINGALE CONVERGENCE THEOREM): Let $a < b \in \mathbb{Q}$. By the Up-crossing Lemma

$$\mathbb{E}[N_n(X,[a,b])] \leq \mathbb{E}[(X_n - a)^-] \leq \mathbb{E}[|X_n|] + a \leq M < \infty$$

for some constant M, for all n. By the Monotone Convergence Theorem

$$\mathbb{E}[N(X,[a,b])] \leq M,$$

hence for every $a < b \in \mathbb{Q}$, $N(X,[a,b]) < \infty$ a.s. Therefore, by 2.3.4, $(X_n, n \geq 0)$ converges in $\overline{\mathbb{R}}$.

It remains to show that $X_\infty = \lim_{n\to\infty} X_n$ is finite a.s. But $\mathbb{E}[|X_n|] \leq M < \infty$, so by Fatou's Lemma

$$\mathbb{E}[|X_\infty|] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|] \leq M < \infty$$

Thus X_∞ is integrable and hence finite a.s. $\square$

2.3.6 Corollary. *Let $(X_n, n \geq 0)$ be a non-negative super-martingale and let T be a stopping time. Then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ (where $X_T = X_\infty$ on the event $\{T = \infty\}$).*

Warning: We can't turn the "$\leq$" into an "$=$" even if X is a martingale. See the example following 2.2.4.

PROOF: $T \wedge n \nearrow T$ and $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_0]$ since $T \wedge n$ is bounded (by n). Apply Fatou's Lemma and MCT to get

$$\mathbb{E}[X_T] = \mathbb{E}[\liminf_{n\to\infty} X_{T\wedge n}] \leq \liminf_{n\to\infty} \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]. \qquad \square$$

2.4 Convergence in L^p for $p \in (1,\infty)$

We have seen that $X_n \xrightarrow{\text{a.s.}} X_\infty$ if X is bounded in L^1. When can we upgrade this to convergence in L^p?

2.4.1 Proposition (Doob's Maximal Inequality).
Let $(X_n, n \geq 0)$ be a sub-martingale, and define $\tilde{X}_n = \max_{0\leq k\leq n} X_k$. Then for $a > 0$,

$$a\,\mathbb{P}(\tilde{X}_n > a) \leq \mathbb{E}[X_n \mathbf{1}_{\{\tilde{X}_n > a\}}].$$

PROOF: Let $T = \inf\{n \geq 0 \mid X_n > a\}$, a stopping time. Notice that $\{T \leq n\} = \{\tilde{X}_n > a\}$, and the stopped process $X^T = (X_{T\wedge n}, n \geq 0)$ is a submartingale. Thus

$$\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T \mathbf{1}_{\{T\leq n\}}] + \mathbb{E}[X_n \mathbf{1}_{\{T>n\}}].$$

Moreover, since $T \wedge n$ is bounded (by n) the OST implies that $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n]$. Thus

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_T \mathbf{1}_{\{T\leq n\}}] + \mathbb{E}[X_n \mathbf{1}_{\{T>n\}}].$$

Rearranging, $\mathbb{E}[X_n \mathbf{1}_{\{T\leq n\}}] \geq a\,\mathbb{P}(T \leq n)$. But $\{T \leq n\} = \{\tilde{X}_n > a\}$, so we are done. $\square$

2.4.2 Proposition (Doob's L^p-inequality).
Let $p \in (1,\infty)$ and $(X_n, n \geq 0)$ be a martingale. Define $X_n^* = \max_{0\leq k\leq n} |X_k|$. Then (recalling $\|X\|_p = (\mathbb{E}[|X|^p])^{\frac{1}{p}}$)

$$\|X_n^*\|_p \leq \frac{p}{p-1}\|X_n\|_p.$$

PROOF: For any random variable

$$\mathbb{E}[Y^p] = p\int_0^\infty x^{p-1}\,\mathbb{P}(Y > x)dx.$$

$(|X_n|, n \geq 0)$ is a sub-martingale by the conditional Jensen inequality, so by Doob's Maximal Inequality,

$$\mathbb{P}(X_n^* > x) \leq \frac{1}{x}\mathbb{E}[|X_n| \mathbf{1}_{\{X_n^* > x\}}].$$

Combining these and Hölder's inequality, where $\frac{1}{p}+\frac{1}{q}=1$,

$$\begin{aligned}
\mathbb{E}[(X_n^*)^p] &= p\int_0^\infty x^{p-1}\,\mathbb{P}[X_n^*>x]dx \\
&\le \frac{p}{p-1}\int_0^\infty (p-1)x^{p-2}\,\mathbb{E}[|X_n|\mathbf{1}_{\{X_n^*>x\}}]dx \\
&\le \frac{p}{p-1}\,\mathbb{E}\left[|X_n|\int_0^\infty (p-1)x^{p-2}\mathbf{1}_{\{x<X_n^*\}}dx\right] \\
&= \frac{p}{p-1}\,\mathbb{E}[|X_n|(X_n^*)^{p-1}] \\
&\le \frac{p}{p-1}\,\mathbb{E}[|X_n|^p]^{\frac{1}{p}}\,\mathbb{E}[(X_n^*)^{(p-1)q}]^{\frac{1}{q}}.
\end{aligned}$$

But $p+q=pq$, so $\|X_n^*\|_p \le \frac{p}{p-1}\|X_n\|_p$. □

2.4.3 Definition. When $X_n=\mathbb{E}[Z \mid \mathscr{F}_n]$ for all $n\ge 0$ for some Z then $(X_n, n\ge 0)$ is a martingale, and if $Z\in L^p(\omega,\mathscr{F},\mathbb{P})$ then we say that X is *closed in* L^p.

2.4.4 Theorem. *Let $(X_n, n\ge 0)$ be a martingale. Then for $p>1$ the following are equivalent.*

(i) (X_n) is bounded in L^p (i.e. $\sup_n \|X_n\|_p<\infty$).

(ii) $X_n\to X_\infty$ a.s. and in L^p.

(iii) (X_n) is closed in L^p.

Proof: (i) implies (ii): Bounded in L^p implies bounded in L^1, so by the martingale convergence theorem X_n converges a.s. to a finite limit X_∞. By Doob's inequality,

$$\|X_n^*\|_p \le \frac{p}{p-1}\sup_{m\ge 1}\|X_m\|_p < \infty,$$

so $X_n^* \nearrow_{n\to\infty} X_\infty^* = \sup_{m\ge1}|X_m|$. By the MCT, $\|X_\infty^*\|_p<\infty$, so $|X_n|\le X_\infty^*$ for all n, which implies that $|X_\infty|\le X_\infty^*$, so $X_\infty\in L^p$. Moreover, $|X_n-X_\infty|^p\le (2X_\infty)^p\in L^1$, so by the DCT, $\mathbb{E}[|X_n-X_\infty|^p]\to 0$ as $n\to\infty$, or equivalently $X_n\to X_\infty$ in L^p.

(ii) implies (iii): Suppose $X_n\to X_\infty$ in L^p. Since $X_n=\mathbb{E}[X_{n+k}\mid\mathscr{F}_n]\to\mathbb{E}[X_\infty\mid\mathscr{F}_n]$ as $k\to\infty$, because $\mathbb{E}[\cdot\mid\mathscr{G}]$ is continuous, $X_n=\mathbb{E}[X_\infty\mid\mathscr{F}_n]$, so take $Z=X_\infty$.

(iii) implies (i): Suppose $X_n=\mathbb{E}[Z\mid\mathscr{F}_n]$ for some $Z\in L^p$. By Jensen, $|X_n|^p\le\mathbb{E}[|Z|^p\mid\mathscr{F}_n]$, so $\sup_{n\ge0}\mathbb{E}[|X_n|^p]\le\mathbb{E}[|Z|^p]<\infty$. □

Suppose that we have that $X_n=\mathbb{E}[Z\mid\mathscr{F}_n]$. Then $\mathbb{E}[Z\mid\mathscr{F}_\infty]=X_\infty$, where

$$\mathscr{F}_\infty=\bigvee_{n\ge0}\mathscr{F}_n=\sigma\Big(\bigcup_{n\ge0}\mathscr{F}_n\Big).$$

Indeed, let $A\in\bigcup_{n\ge0}\mathscr{F}_n$, say $A\in\mathscr{F}_m$. Then $\mathbb{E}[Z\mathbf{1}_A]=\mathbb{E}[X_m\mathbf{1}_A]=\mathbb{E}[X_n\mathbf{1}_A]$ for any $n\ge m$. Therefore as $n\to\infty$, by L^p convergence, $\mathbb{E}[Z\mathbf{1}_A]=\mathbb{E}[X_\infty\mathbf{1}_A]$. To conclude that this is true for every $A\in\bigvee_{n\ge0}\mathscr{F}_n$, note that $\bigcup_{n\ge0}\mathscr{F}_n$ is a π-system that spans $\mathscr{F}_\infty$ and apply the monotone class theorem. Finally, note that $X_\infty=\limsup_{n\to\infty}X_n$

is an $\mathscr{F}_\infty$-measurable r.v. In particular, if $\mathscr{F} = \mathscr{F}_\infty$ then X_∞ is the only possible Z for which $X_n = \mathbb{E}[Z \mid \mathscr{F}_n]$ for all n.

Yet otherwise said,

$$L^p(\Omega, \mathscr{F}_\infty, \mathbb{P}) \to \{L^p\text{-bounded martingales}\} : Z \mapsto (\mathbb{E}[Z \mid \mathscr{F}_n], n \geq 0)$$

is a bijection.

2.5 Convergence in L^1

2.5.1 Definition. A sequence $(X_n, n \geq 0)$ is *uniformly integrable* (or *u.i.*) if $\sup_{n\geq 0} \mathbb{E}[|X_n| \mathbf{1}_{|X_n|>a}] \to 0$ as $a \to \infty$.

2.5.2 Lemma.

(i) If $X_n \to X_\infty$ a.s. then $X_n \to X_\infty$ in L^1 if and only if (X_n) is u.i.

(ii) If $(\mathscr{F}_n, n \geq 0)$ is a filtration and $Z \in L^1$ then $(\mathbb{E}[Z \mid \mathscr{F}_n], n \geq 0)$ is u.i.

PROOF: Exercise (see example sheet). □

2.5.3 Theorem. *Let $(X_n, n \geq 0)$ be a martingale. The following are equivalent.*

(i) (X_n) is uniformly integrable.

(ii) $X_n \to X_\infty$ a.s. and in L^1.

(iii) (X_n) is closed in L^1.

PROOF: (i) implies (ii): From example sheet 0, u.i. implies bounded in L^1, so $X_n \to X_\infty$ a.s. by the martingale convergence theorem. By 2.5.2, this implies that $X_n \to X_\infty$ in L^1 since (X_n) is u.i.

(ii) implies (iii): As the proof of 2.4.4, we have $X_n = \mathbb{E}[X_{n+k} \mid \mathscr{F}_n]$, and let $k \to \infty$.

(iii) implies (i): The 2.5.2 implies that $(\mathbb{E}[Z \mid \mathscr{F}_n], n \geq 0)$ is u.i. for any $Z \in L^1$. □

As before, $Z = X_\infty$ is the only $\mathscr{F}_\infty$-measurable r.v. for which $X_n = \mathbb{E}[Z \mid \mathscr{F}_n]$ for all n. Similarly,

$$L^1(\Omega, \mathscr{F}_\infty, \mathbb{P}) \to \{\text{u.i. martingales}\} : Z \to (\mathbb{E}[Z \mid \mathscr{F}_n], n \geq 0)$$

is a bijection.

2.6 Optional stopping, part II

Recall that $\mathbb{E}[X_T \mid \mathscr{F}_S] = X_S$ when $S \leq T$ are bounded stopping times. It turns out that we may eliminate the boundedness of the stopping times when the martingale is u.i.

2.6.1 Theorem (Optional Stopping, u.i. martingale).
Let $(X_n, n \geq 0)$ be a uniformly integrable martingale and let S and T be stopping times with $S \leq T$. Then $\mathbb{E}[X_T \mid \mathscr{F}_S] = X_S$, where we take $X_T = X_T \mathbf{1}_{T<\infty} + X_\infty \mathbf{1}_{T=\infty}$. In particular, in this case $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

PROOF: X_T is integrable since

$$X_T = X_\infty \mathbf{1}_{T=\infty} + \sum_{n\geq 0} X_n \mathbf{1}_{T=n} = X_\infty \mathbf{1}_{T=\infty} + \sum_{n\geq 0} \mathbb{E}[X_\infty \mid \mathscr{F}_n] \mathbf{1}_{T=n}$$

by uniform integrability. Therefore

$$|X_T| \leq |X_\infty| \mathbf{1}_{T=\infty} + \sum_{n\geq 0} \mathbb{E}[|X_\infty| \mathbf{1}_{T=n} \mid \mathscr{F}_n],$$

so

$$\mathbb{E}[|X_T|] \leq \sum_{n=0}^{n=\infty} \mathbb{E}[|X_\infty| \mathbf{1}_{T=n}] = \mathbb{E}[|X_\infty|].$$

Suppose that $T = \infty$. Let $A \in \mathscr{F}_S$. Then

$$\begin{aligned}
\mathbb{E}[X_\infty \mathbf{1}_A] &= \sum_{n=0}^{n=\infty} \mathbb{E}[X_\infty \mathbf{1}_A \mathbf{1}_{S=n}] \\
&= \mathbb{E}[X_\infty \mathbf{1}_A \mathbf{1}_{S=\infty}] + \sum_{n\geq 0} \mathbb{E}[\mathbb{E}[X_\infty \mid \mathscr{F}_n] \mathbf{1}_A \mathbf{1}_{S=n}] \\
&= \sum_{n=0}^{n=\infty} \mathbb{E}[X_S \mathbf{1}_A \mathbf{1}_{S=n}] \\
&= \mathbb{E}[X_S \mathbf{1}_A]
\end{aligned}$$

Thus $\mathbb{E}[X_\infty \mid \mathscr{F}_S] = X_S$. For general T, $\mathbb{E}[X_T \mid \mathscr{F}_S] = \mathbb{E}[\mathbb{E}[X_\infty \mid \mathscr{F}_T] \mid \mathscr{F}_S] = X_S$. □

2.6.2 Exercise. Let $(S_n)_{n\geq 0}$ be a simple random walk, $S_n = X_1 + \cdots + X_n$, where $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$. Let $T_x = \inf\{n \geq 0 \mid S_n = x\}$, and $T = T_a \wedge T_{-b}$ for $a, b \in N$. Compute $\mathbb{P}(T_a < T_{-b}) = \mathbb{P}(S_T = a)$.

SOLUTION: By the u.i. OST (though how do we show that (S_n) is u.i.?) we have

$$a\mathbf{1}_{T_a < T_{-b}} - b(1 - \mathbf{1}_{T_a < T_{-b}}) = \mathbb{E}[S_T] = \mathbb{E}[S_0] = 0.$$

Whence $\mathbb{P}(T_a < T_{-b}) = \frac{b}{a+b}$. ✠

2.7 Backward martingales

For this section we let $I = \mathbb{Z}_- = \{\dots, -2, -1, 0\}$ and let $(\mathscr{F}_n, n \leq 0)$ be a filtration.

2.7.1 Definition. A *backward martingale* is a martingale $(X_n, n \leq 0)$ with respect to the filtration $(\mathscr{F}_n, n \leq 0)$.

Note that for all $n \leq 0$, $\mathbb{E}[X_0 \mid \mathscr{F}_n] = X_n$, so a backwards martingale is always closed.

2.7.2 Theorem. *Let $(X_n, n \leq 0)$ be a backwards martingale. Then X_n converges a.s. and in L^1 to a limit $X_{-\infty}$ as $n \to -\infty$ and $X_{-\infty} = \mathbb{E}[X_0 \mid \mathscr{F}_{-\infty}]$, where $\mathscr{F}_{-\infty} = \bigcap_{n\leq 0} \mathscr{F}_n$.*

PROOF: Let $a < b$, and let $N_n(x,[a,b])$ be the number of upcrossings of X between times $-n$ and 0 from a to b. Consider $(X_{-n+k}, 0 \le k \le n)$, a martingale with respect to the filtration $(\mathscr{F}_{-n+k}, 0 \le k \le n)$. Doob's Up-crossing Lemma gives that

$$(b-a)\mathbb{E}[N_n(X,[a,b])] \le \mathbb{E}[(X_0-a)^-].$$

Letting $n \to \infty$, we obtain

$$(b-a)\mathbb{E}[N(X,[a,b])] \le \mathbb{E}[|X_0|] + a < \infty.$$

Therefore, for all $a < b \in \mathbb{Q}$, $N(X,[a,b]) < \infty$ a.s., so $X_n \to X_{-\infty} \in \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ a.s. as $n \to -\infty$. Again by Fatou's Lemma, $X_{-\infty} \in \mathbb{R}$ a.s. Since $X_n = \mathbb{E}[X_0 \mid \mathscr{F}_n]$ for all n, the family $(X_n, n \le 0)$ is u.i. by 2.5.2. Therefore $X_n \to X_{-\infty}$ in L^1. Finally, let $A \in \mathscr{F}_{-\infty}$. Then

$$\mathbb{E}[\mathbf{1}_A X_0] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X_0 \mid \mathscr{F}_n]] = \mathbb{E}[\mathbf{1}_A X_n] \to \mathbb{E}[\mathbf{1}_A X_{-\infty}]$$

as $n \to -\infty$ since $X_n \to X_{-\infty}$ in L^1. Therefore $X_{-\infty} = \mathbb{E}[X_0 \mid \mathscr{F}_{-\infty}]$ since it is $\mathscr{F}_{-\infty}$-measurable. □

Remark. Sometimes backwards martingales are defined as a forwards process $(Y_n, n \ge 0)$ with respect to a backwards filtration $\mathscr{G}_0 \supseteq \mathscr{G}_1 \supseteq \mathscr{G}_2 \supseteq \cdots$ such that Y_n is adapted and in L^1, and $\mathbb{E}[Y_n \mid \mathscr{G}_{n+1}] = Y_{n+1}$. This is equivalent to our definition by taking $Y_n = X_{-n}$ and $\mathscr{G}_n = F_{-n}$ for all $n \ge 0$.

3 Applications of Discrete Time Martingales

3.1 Strong law of large numbers

We begin with a classical result of Kolmogorov.

3.1.1 Theorem (Kolmogorov's 0-1 law). *Let $X_0, X_1, X_2, \dots$ be independent r.v.'s. Define*

$$\mathscr{G}_n = \sigma(X_n, X_{n+1}, \dots) \qquad \textit{and} \qquad \mathscr{G}_\infty = \bigcap_{n \ge 0} \mathscr{G}_n.$$

Then $\mathbb{P}(A) \in \{0,1\}$ for all $A \in \mathscr{G}_\infty$.

$\mathscr{G}_\infty$ is the *tail σ-algebra* of $(X_0, X_1, X_2, \dots)$.

PROOF: Let $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$ and $A \in \mathscr{G}_\infty$. By definition, $\mathscr{G}_\infty$ is independent of $\mathscr{F}_n$ since $\mathscr{G}_\infty \subseteq \mathscr{G}_{n+1}$ and the X_i's are independent. Thus $\mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_n] = \mathbb{P}(A)$. On the other hand, $(\mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_n], n \ge 0)$ is a martingale with respect to the filtration $(\mathscr{F}_n, n \ge 0)$. Thus $\mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_n] \to \mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_\infty]$ a.s. as $n \to \infty$, where $\mathscr{F}_\infty = \bigvee_{n \ge 0} \mathscr{F}_n$. But $\mathscr{F}_\infty \supseteq \mathscr{G}_\infty$ since $\mathscr{F}_\infty = \sigma(X_0, X_1, \dots) \supseteq \mathscr{G}_n$ for all n. Therefore $\mathbf{1}_A = \mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_\infty] = \mathbb{P}(A)$ a.s. □

3.1.2 Theorem (Strong Law of Large Numbers).
Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with $\mathbb{E}[|X_1|] < \infty$, and let $S_n = X_1 + \dots + X_n$. Then $\frac{S_n}{n} \to \mathbb{E}[X_1]$ a.s. as $n \to \infty$.

PROOF: Let $\mathscr{F}_n = \sigma(S_1, \dots, S_n) = \mathscr{F}^S_n$, and

$$\mathscr{G}_n = \sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, X_{n+1}, X_{n+2}, \dots).$$

We claim that $(\frac{S_{-n}}{-n}, n \le 0)$ is a backwards martingale with respect to $(\mathscr{G}_{-n}, n \le 0)$. Indeed,

$$\mathbb{E}\left[\frac{S_n}{n} \mid \mathscr{G}_{n+1}\right] = \mathbb{E}\left[\frac{S_{n+1} - X_{n+1}}{n} \mid \mathscr{G}_{n+1}\right] = \frac{S_{n+1}}{n} - \frac{1}{n}\mathbb{E}[X_{n+1} \mid \mathscr{G}_{n+1}].$$

But

$$\mathbb{E}[X_{n+1} \mid \mathscr{G}_{n+1}] = \mathbb{E}[X_{n+1} \mid S_{n+1}, X_{n+2}, X_{n+3}, \dots] = \mathbb{E}[X_{n+1} \mid S_{n+1}]$$

by a property of conditional expectation, since the r.v.'s $X_{n+2}, X_{n+3}, \dots$ are independent of X_{n+1} and S_{n+1}. Note that $\mathbb{E}[X_{n+1} \mid S_{n+1}] = \mathbb{E}[X_k \mid \mathscr{F}_{n+1}]$ for all $k \in \{0, 1, \dots, n+1\}$ since $\mathbb{E}[X_{n+1} f(S_{n+1})] = \mathbb{E}[X_k f(S_{n+1})]$ since S_{n+1} is a symmetric function of $X_1, \dots, X_{n+1}$. Therefore

$$\mathbb{E}[\mathbb{E}[X_{n+1} \mid S_{n+1}] f(S_{n+1})] = \mathbb{E}[\mathbb{E}[X_k \mid S_{n+1}] f(S_{n+1})]$$

so

$$\mathbb{E}[X_{n+1} \mid S_{n+1}] = \frac{\sum_{k=1}^{n+1} \mathbb{E}[X_k \mid S_{n+1}]}{n+1} = \frac{\mathbb{E}[\sum_{k=1}^{n+1} X_k \mid S_{n+1}]}{n+1} = \frac{S_{n+1}}{n+1}$$

Finally,

$$\mathbb{E}\left[\frac{S_n}{n} \mid \mathscr{G}_{n+1}\right] = \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n}(1 - \frac{1}{n+1}) = \frac{S_{n+1}}{n+1},$$

and the claim is proved.

Therefore $\frac{S_n}{n}$ converges to some finite limit L a.s. and in L^1 as $n \to \infty$. We must finally check that $L = \mathbb{E}[X_1]$. We have $\mathbb{E}[L] = \mathbb{E}[\frac{S_n}{n}] = \mathbb{E}[\frac{S_1}{1}] = \mathbb{E}[X_1]$. Note that L is measurable with respect to the tail σ-algebra of $X_1, X_2, \dots$ (since $\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{S_n - S_k}{n}$ for all k, so L is $\sigma(X_k, X_{k+1}, \dots)$-measurable for all k.) By Kolmogorov's 0-1 law, $\mathbb{P}(L = a) \in \{0, 1\}$ for all a, implying that L is constant a.s. Therefore $L = \mathbb{E}[X_1]$. □

3.2 Radon-Nikodym theorem

Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let Q be a non-negative, finite measure on $(\Omega, \mathscr{F})$. We would like to find conditions under which there exists some non-negative r.v. X such that $Q = X \cdot P$, in the sense that $Q(A) = \mathbb{E}^P[\mathbf{1}_A X]$ for all $A \in \mathscr{F}$. In this case X is called a *density* or *Radon-Nikodym derivative* of Q with respect to P, and we write $X = \frac{dQ}{dP}$.

Let $(\mathscr{F}_n, n \ge 0)$ be a filtration on $(\Omega, \mathscr{F})$, with $\mathscr{F}_\infty = \mathscr{F}$. Assume that $Q|_{\mathscr{F}_n} = X_n \cdot P|_{\mathscr{F}_n}$ for some non-negative $\mathscr{F}_n$-measurable r.v. X_n, for every n. Write $P_n = P|_{\mathscr{F}_n}$ and $Q_n = Q|_{\mathscr{F}_n}$.

3.2.1 Lemma. *With the scenario as described above, $(X_n, n \ge 0)$ is a $(\mathscr{F}_n)$-martingale.*

PROOF: Indeed, let $A \in \mathscr{F}_n$. Then

$$\mathbb{E}^{P_n}[X_{n+1}\mathbf{1}_A] = \mathbb{E}^{P_{n+1}}[X_{n+1}\mathbf{1}_A] = Q_{n+1}(A) = Q_n(A) = \mathbb{E}^{P_n}[X_{n+1} \mid \mathscr{F}_n]$$

since $A \in \mathscr{F}_n$. Therefore $X_n = \mathbb{E}[X_{n+1} \mid \mathscr{F}_n]$. □

By the martingale convergence theorem, $X_n \to X_\infty \geq 0$ a.s. as $n \to \infty$. Is it true that $Q = X_\infty \cdot P$?

3.2.2 Proposition. *Q admits a density X with respect to P if and only if the martingale $(X_n, n \geq 0)$, defined above, is u.i. In this case $X = X_\infty$.*

PROOF: Assume that $(X_n, n \geq 0)$ is u.i. Then $X_n \to X_\infty$ a.s. and in L^1. Hence if $A \in F_m$ then for $n \geq m$,

$$Q(A) = \mathbb{E}[X_m \mathbf{1}_A] = \mathbb{E}[X_n \mathbf{1}_A] \xrightarrow{n\to\infty} \mathbb{E}[X_\infty \mathbf{1}_A]$$

by L^1 convergence. Therefore, for all $A \in \bigcup_{n\geq 0} F_n$, $\mathbb{E}[X_\infty \mathbf{1}_A] = Q(A)$, and by the monotone class theorem this also holds for every $A \in F_\infty = \sigma(\bigcup_{n\geq 0} F_n)$, so X_∞ is a density of Q with respect to P.

Conversely, assume that $Q = X \cdot P$ for some $X \geq 0$. Then $Q(A) = \mathbb{E}[X \mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X \mid F_n] \mathbf{1}_A]$ for all $A \in F_n$. Thus $Q_n = \mathbb{E}[X \mid F_n] \cdot P_n$. By uniqueness of the density, $X_n := \frac{dQ_n}{dP_n} = \mathbb{E}[X \mid F_n]$. Now, $\infty > Q(\omega) = \mathbb{E}[X]$, so $X \in L^1$. Since $(X_n, n \geq 0)$ is closed in L^1, it is u.i. □

3.2.3 Theorem (Radon-Nikodym). *Assume that $\mathscr{F}$ is separable, i.e. that there are events $F_1, F_2, \ldots$ such that $\mathscr{F} = \sigma(F_1, F_2, \ldots)$. Let Q be a non-negative finite measure on $(\Omega, \mathscr{F})$ and P be a probability on $(\Omega, \mathscr{F})$. Then the following are equivalent.*

(i) $P(A) = 0$ implies $Q(A) = 0$ for all $A \in \mathscr{F}$ (in this case we say that Q is absolutely continuous with respect to P, and write $Q \ll P$).

(ii) For all $\varepsilon > 0$ there is $\delta > 0$ such that for all $A \in \mathscr{F}$, $P(A) < \delta$ implies $Q(A) < \varepsilon$.

(iii) There exists a r.v. X such that $Q = X \cdot P$ (i.e. Q admits a density with respect to P).

PROOF: (iii) implies (i): Trivial, since if $Q(A) = \mathbb{E}^P[X \mathbf{1}_A]$ for some X then $P(A) = 0$ implies $\mathbf{1}_A = 0$ a.s. under P, so $Q(A) = 0$.

(i) implies (ii): Assume that (ii) does not hold. Then there is $\varepsilon > 0$ such that for every $\delta > 0$ there is $B \in \mathscr{F}$ such that $P(B) < \delta$ and $Q(B) \geq \varepsilon$. By taking $\delta = \frac{1}{2^n}$, we can find $(B_n, n \geq 0)$ such that $P(B_n) < \frac{1}{2^n}$, but $Q(B_n) \geq \varepsilon > 0$. Therefore $\sum_n P(B_n) < \infty$, so by the Borel-Cantelli Lemma, $P(B_n \text{ happens infinitely often}) = 0$. On the other hand,

$$Q\left(\bigcap_{k\geq n} B_k\right) \geq Q(B_n) \geq \varepsilon.$$

The lefthand side converges to $Q(\limsup_n B_n)$ as $n \to \infty$, so

$$Q(B_n \text{ happens infinitely often}) > 0$$

and (i) does not hold.

(ii) implies (iii): Let $\mathscr{F}_n = \sigma(F_1, \ldots, F_n)$. Then $\mathscr{F}_n = \sigma(A^\varepsilon, \varepsilon \in \{0,1\}^n)$, where $A^\varepsilon = \cap_{i=1}^n F_i^{\varepsilon_i}$, $F_i^0 = F_i$, and $F_i^1 = \Omega \setminus F_i$. The A^ε partition Ω and generate $\mathscr{F}_n$. Consider $Q_n = Q|_{\mathscr{F}_n}$ and $P_n = P|_{\mathscr{F}_n}$, and let

$$X_n = \sum_{\varepsilon \in \{0,1\}^n} \frac{Q(A^\varepsilon)}{P(A^\varepsilon)} \mathbf{1}_{A^\varepsilon}$$

with the convention that $\frac{0}{0} = 0$.

Claim. $X_n = \frac{dQ_n}{dP_n}$.

Indeed, if $A = A^\varepsilon$ for some $\varepsilon \in \{0,1\}^n$ then

$$Q_n(A) = \frac{Q_n(A)}{P_n(A)} P_n(A) = \mathbb{E}^{P_n}\left[\frac{Q(A)}{P(A)}\mathbf{1}_A\right] = \mathbb{E}^{P_n}\left[\mathbf{1}_A \sum_{\varepsilon\in\{0,1\}^n} \frac{Q(A^\varepsilon)}{P(A^\varepsilon)}\mathbf{1}_{A^\varepsilon}\right].$$

Therefore $Q_n(A) = \mathbb{E}[X_n \mathbf{1}_A]$. By linearity this holds for all $A \in \mathscr{F}_n$.

By the previous proposition, it remains to show that (X_n) is u.i. to obtain that $Q = X_\infty \cdot P$. We must check that $\sup_{n\geq 0} \mathbb{E}[X_n \mathbf{1}_{X_n>\lambda}] \to 0$ as $\lambda \to \infty$. But $\mathbb{E}[X_n \mathbf{1}_{X_n>\lambda}] = Q(X_n > \lambda)$, and note that

$$P(X_n > \lambda) \leq \frac{1}{\lambda}\mathbb{E}[X_n] = \frac{1}{\lambda}Q(\Omega).$$

Need to show for all $\varepsilon > 0$ there is $\delta > 0$ such that $\sup_n P(X_n \geq \lambda) \leq \delta$ (take $\lambda > \frac{Q(\Omega)}{\lambda}$). By (ii), fixing $\varepsilon > 0$, there is $\delta > 0$ such that $P(A) \leq \delta$ implies $Q(A) < \varepsilon$. By choosing λ as above, $P(X_n > \lambda) \leq \delta$ implies $Q(X_n > \lambda) \leq \varepsilon$.

Therefore for all $\varepsilon > 0$ there is $\delta > 0$ such that $\sup_n \mathbb{E}[X_n \mathbf{1}_{X_n>\lambda}] \leq \varepsilon$. Hence $(X_n, \geq 0)$ is u.i., and $Q = X_\infty \cdot P$. □

3.3 Kakutani's theorem for product martingales

3.3.1 Definition. A *product martingale* is a martingale of the form $X_n = \prod_{i=1}^n Y_i$, where the Y_i are independent, non-negative r.v.'s such that $\mathbb{E}[Y_i] = 1$ for all $i = 1, \dots, n$.

A product martingale is indeed a martingale, since $\mathbb{E}[X_n] = \prod_{i=1}^n \mathbb{E}[Y_i] = 1$ by independence, and

$$\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] = \mathbb{E}[Y_{n+1}X_n \mid \mathscr{F}_n] = X_n \mathbb{E}[Y_{n+1} \mid \mathscr{F}_n] = X_n \mathbb{E}[Y_{n+1}] = X_n,$$

since Y_{n+1} is independent of $\mathscr{F}_n$, where

$$\mathscr{F}_n := \mathscr{F}_n^X = \sigma(X_1, \dots, X_n) = \sigma(Y_1, \dots, Y_n).$$

Now $X_n \geq 0$, so $X_n \to X_\infty \geq 0$ finite a.s. as $n \to \infty$.

3.3.2 Theorem (Kakutani). *With the notation as above, let $a_n = \mathbb{E}[\sqrt{Y_n}]$. The following are equivalent.*

(i) $(X_n, n \geq 0)$ *is a u.i. martingale;*

(ii) $\mathbb{E}[X_\infty] = 1$*;*

(iii) $P(X_\infty > 0) > 0$*;*

(iv) $\prod_{n\geq 1} a_n > 0$.

By the Cauchy-Schwartz inequality, $\mathbb{E}[\sqrt{Y_n}]^2 \leq \mathbb{E}[Y_n] = 1$, so $a_n \leq 1$ for all n. Recall that $\prod_{n\geq 1} a_n \in [0,1]$, and $\prod_{n\geq 1} a_n > 0$ if and only if $\sum_{n\geq 1}(1 - a_n) < \infty$ (this is seen by taking logarithms).

PROOF: (iii) implies (iv): Suppose that $\prod_{n\geq 0} a_n = 0$. Let $M_n = \prod_{i=1}^n \frac{\sqrt{Y_i}}{a_i}$, so that M_n is also a (non-negative) product martingale. By the martingale convergence theorem, $M_n \to M_\infty \geq 0$ a.s. We have

$$M_n = \frac{1}{\prod_{i=1}^n a_i} \sqrt{X_n}$$

so $X_n = M_n^2 \prod_{i=1}^n a_i^2 \to 0$ as $n \to \infty$ since M_∞ is finite valued and $\prod_{n\geq 0} a_n = 0$. Thus $X_\infty \equiv 0$ and the contrapositive is proved.

(iv) implies (i): Assume that $\prod_{n\geq 0} > 0$. Then

$$\mathbb{E}[M_n^2] = \mathbb{E}\left[\frac{X_n}{\prod_{i=1}^n a_i^2}\right] = \frac{1}{\prod_{i=1}^n a_i^2} \leq \frac{1}{\prod_{i\geq 0} a_i^2} < \infty.$$

Therefore M_n is bounded in L^2. Doob's inequality gives that

$$\mathbb{E}[(M_n^*)^2] \leq \left(\frac{2}{2-1}\right)^2 \mathbb{E}[M_n^2] \leq \frac{4}{\prod_{i\geq 1} a_i^2},$$

where $M_n^* = \max_{1\leq i\leq n} M_i$. Therefore $\sup_{n\geq 0} \mathbb{E}[(M_n^*)^2] < \infty$. Then $M_\infty^* = \sup_{n\geq 0} \frac{1}{\prod_{i=1}^n a_i} \sqrt{X_n}$, which implies that

$$(M_\infty^*)^2 = \sup_{n\geq 0} \frac{X_n}{\prod_{i=1}^n a_i^2} \qquad \text{so} \qquad \sup_{n\geq 0} X_n \leq (M_\infty^*)^2.$$

X_n is dominated by an integrable r.v., so $(X_n, n \geq 0)$ is u.i. (Clean this up using the notes).

(i) implies (ii): If X is u.i. then $X_n \to X_\infty$ a.s. and in L^1 as $n \to \infty$. We have $1 = \mathbb{E}[X_n] \to \mathbb{E}[X_\infty]$.

(ii) implies (iii): If $\mathbb{E}[X_\infty] > 0$ then $\mathbb{P}(X_\infty > 0) > 0$ since $X_\infty \geq 0$. □

Remark. In the case that every Y_n is positive a.s. for every n, we are assured that $X_n > 0$ for all n as well. Therefore $\{X_\infty = 0\}$ does not depend on the first few Y_i's, so it is a tail event. By Kolmogorov's 0-1 law, $\mathbb{P}(X_\infty = 0) \in \{0, 1\}$. In this case, $\mathbb{P}(X_\infty > 0) > 0$ if and only if $\mathbb{P}(X_\infty > 0) = 1$.

3.4 Likelihood ratio

A very general statement of a basic problem in applied statistics is as follows. Let $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}(\mathbb{R})^{\otimes\mathbb{N}})$ be the measurable space of real sequences and

$$X_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R} : \omega = (\omega_i, i \geq 0) \mapsto \omega_n$$

be the canonical projections. Let $(f_i, i \geq 0)$ and $(g_i, i \geq 0)$ be sequences of strictly positive p.d.f.'s. Let P and Q be the probability distributions under which X_i is a r.v. with distribution $f_i(x)dx$ and $g_i(x)dx$, respectively. Further suppose that the X_i's are independent under P and Q. When is $Q \ll P$?

3.4.1 Definition. The *likelihood ratio* is

$$L_n = \prod_{i=1}^n \frac{g_i(X_i)}{f_i(X_i)},$$

for $n \geq 0$ (taking $L_0 = 1$).

Let $\mathscr{F}_n = \mathscr{F}_n^X = \sigma(X_1, \dots, X_n)$.

3.4.2 Proposition. *$(L_n, n \geq 0)$ is a $(\mathscr{F}_n)$-martingale in the probability space $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}(\mathbb{R})^{\otimes\mathbb{N}}, P)$, and $Q|_{\mathscr{F}_n} = L_n \cdot P|_{\mathscr{F}_n}$.*

PROOF: Let $A_1, \dots, A_n$ be measurable Borel sets in $\mathbb{R}$. Compute

$$\begin{aligned} Q|_{F_n}(X_1 \in A_1, \dots, X_n \in A_n) &= \int_{A_1 \times \cdots \times A_n} g_1(x_1)dx_1 \cdots g_n(x_n)dx_n \\ &= \int_{A_1 \times \cdots \times A_n} \frac{g_1(x_1) \cdots g_n(x_n)}{f_1(x_1) \cdots f_n(x_n)} f_1(x_1) \cdots f_n(x_n) dx_1 \cdots dx_n \\ &= \mathbb{E}^{P|_{F_n}}(L_n \mathbf{1}_{\{X_1 \in A_1, \dots, X_n \in A_n\}}) \qquad \square \end{aligned}$$

Thus $Q = L_\infty \cdot P$ (or $L_\infty = \frac{dQ}{dP}$) if and only if $(L_n, n \geq 0)$ is a u.i. martingale. But $(L_n, n \geq 0)$ is a product martingale (under P) with $Y_n = \frac{g_n(X_n)}{f_n(X_n)}$ for all n. These are independent under P and have $\mathbb{E}[Y_n] = 1$ for all n, so L is u.i. if and only if

$$\prod_{n \geq 1} \mathbb{E}^P \left[\sqrt{\frac{g_n(X_n)}{f_n(X_n)}} \right] > 0$$

by Kakutani's theorem. But

$$\int_{\mathbb{R}} dx_n f_n(x_n) \sqrt{\frac{g_n(x_n)}{f_n(x_n)}} = \int_{\mathbb{R}} \sqrt{f_n g_n}.$$

Using the fact that $(\sqrt{f} - \sqrt{g})^2 = f + g - 2\sqrt{fg}$, it is easy to check that $\prod_{n \geq 1} \int \sqrt{f_n g_n} > 0$ if and only if $\sum_{n \geq 1} \int (\sqrt{f_n} - \sqrt{g_n})^2 < \infty$ (exercise).

3.4.3 Example. Assume that $f_n = f$ and $g_n = g$ for all n. Thus under P and Q, the X_n's are i.i.d. r.v.'s, namely $f(x)dx$ and $g(x)dx$, respectively. Is it true that $Q \ll P$? It is true if and only if $\sum_{n \geq 1} \int (\sqrt{f} - \sqrt{g})^2 < \infty$, so $Q \ll P$ if and only if $\int (\sqrt{f} - \sqrt{g})^2 = 0$, or equivalently when $f = g$ a.s.

This has applications to statistical experiments. Suppose that $X_1, \dots, X_n$ are outcomes of an experiment, known to be i.i.d., either distributed according to $f(x)dx$ or $g(x)dx$. We will test hypotheses

$$H_0 : \text{Law}(X_1) = f(x)dx \qquad \text{against} \qquad H_1 : \text{Law}(X_1) = g(x)dx.$$

The test is as following. If $L_n = \prod_{i=1}^n \frac{g_i(X_i)}{f_i(X_i)} < 1$ then H_0 is validated, otherwise reject H_0. This test is consistent because if H_0 holds we showed that $L_n \to 0$ a.s. as $n \to \infty$, while if H_1 holds then $L_n \to \infty$ a.s.

4 Continuous-time Processes

For this chapter we typically take $I \subseteq \mathbb{R}$ to be an interval, and most of the time we take I to be the non-negative reals, $I = \mathbb{R}_+$.

4.1 Technical issues in dealing with continuous-time

Recall that a filtration is an increasing family $(\mathscr{F}_t, t \in I)$ of sub-σ-algebras of $\mathscr{F}$, and a process $(X_t, t \in I)$ is a family of r.v.'s, and is said to be adapted if X_t is $\mathscr{F}_t$-measurable for all $t \in I$. A stopping time is a r.v. $T : \Omega \to I \cup \{\infty\}$ such that $\{T \le t\} \in \mathscr{F}_t$ for all $t \in I$.

Now $\omega \mapsto X_t(\omega)$ is measurable (with respect to $(\Omega, \mathscr{F})$) for all $t \in I$, but what about the *sample path* $t \mapsto X_t(\omega)$ given by $\omega \in \Omega$ (with respect to $(I, \mathscr{B}(I))$ as a sub-σ-algebra of $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$)? Also, if T is a stopping time then $X_T \mathbf{1}_{T<\infty} = \sum_{s \in I} X_s \mathbf{1}_{T=s}$ is not a r.v. in general. Worse, there are "few" stopping times. Typically, for a Borel subset A of $\mathbb{R}$, $T_A = \inf\{t \ge 0 \mid X_t \in A\}$ has no reason to be a stopping time, since $\{T_A \le t\} = \bigcup_{s \le t, s \in I} \{T_A = s\}$ is typically an uncountable union.

Let $(X_t, t \in I)$ be a process with values in some metric space (E, d) (usually we will have $E = \mathbb{R}^n$ for some n). $C(I, E)$ denotes the set of continuous functions $I \to E$ and $D(I, E)$ denotes the set of *càdlàg* functions, those that are right-continuous and admit left-limits at every point (from the French). If f is càdlàg then, for all $t \in I$, $f(t^+) = \lim_{s \to t^+} f(s) = f(t)$ and $f(t^-) = \lim_{s \to t^-} f(s)$ exists.

4.1.1 Proposition.

(i) Let $(X_t, t \ge 0)$ be a continuous process (so that $t \mapsto X_t(\omega) \in C(I, E)$ for all ω), and let $A \subseteq E$ be closed. Then $T_A = \inf\{t \ge 0 \mid X_t \in A\}$ is a stopping time with respect to $\mathscr{F}^X$.

(ii) If X is adapted to $(\mathscr{F}_t, t \ge 0)$ and T is a stopping time then $X_T \mathbf{1}_{T<\infty}$ is an $\mathscr{F}_T$-measurable r.v. In particular, the process $X^T = (X_{T \wedge t}, t \ge 0)$ is adapted if X is adapted.

PROOF: (i) If X is continuous with and $X_s \in A$ for some $s \in [0, t]$ then we can find a nieghbourhood V of s in $[0, t]$ on which $d(X_s, X_q) \le \varepsilon$ for some fixed $\varepsilon > 0$ and all $q \in V$. Therefore $\{T_A \le t\} = \inf_{q \in [0,t], q \in \mathbb{Q}} \{d(X_q, A) = 0\}$ ($\subseteq$ is clear, and $\supseteq$ comes from compactness (prove this)).

(ii) See notes. □

If A is open then T_A is not an $\mathscr{F}^X$-stopping time in general. However, define $\mathscr{F}_{t^+} = \bigcap_{\varepsilon \ge 0} \mathscr{F}_{t+\varepsilon}$. If $(X_t, t \ge 0)$ is $(\mathscr{F}_t, t \ge 0)$-adapted then it is $(\mathscr{F}_{t^+}, t \ge 0)$-adapted and T_A is a stopping time with respect to $(\mathscr{F}_{t^+}, t \ge 0)$ for $A \subseteq E$ open.

4.2 Finite dimensional marginal distributions, Versions

Let $(X_t, t \in I)$ be a process. We may consider X to be a *random function* in the following sense. Endow E^I with the product σ-algebra, i.e. the smallest σ-algebra such that $\pi_t : f \mapsto f(t)$ is a measurable function for all $t \in I$. By definition, X_t is a random element of E^I with the given σ-algebra.

4.2.1 Definition. For a finite subset J of I, let μ_J be the law of the finite seqence $(X_j, j \in J)$. Elements of the family $(\mu_J, J \subseteq I$ finite) are called the *finite marginal distributions* of the process X.

Note that if X and X' are two processes with the same finite marginal distributions then they define the same r.v. in E^I (i.e. they have the same distribution).

Indeed, the product σ-algebra on E^I is generated by finite rectangles of the form $\{\pi_{t_1}^{-1}(A_1)\cap\cdots\cap\pi_{t_k}^{-1}(A_k)\}$, which is a π-system. Yet otherwise said, the probability of events $\mu_J(A_1\times\cdots\times A_k)=\mathbb{P}(X_{t_1}\in A_1,\ldots,X_{t_k}\in A_k)$ suffice to determine the law of X.

4.2.2 Example. Consider $X_t=0$ for all $t\in[0,1]$ and $X'_t=\mathbf{1}_{\{U\}}(t)$ for $t\in[0,1]$, where U is a uniform r.v. in $[0,1]$. Then the finite marginal distribution of X is $\mu_J=\delta_{(0,\ldots,0)}$ for all $J\subseteq[0,1]$ finite. But the finite marginal distribution of X' is the same since the probability that $U\in\{t_1,\ldots,t_k\}$ is zero.

4.2.3 Definition. Let X and X' be two processes. We say that X' is a *version* of X if $\mathbb{P}(X_t=X'_t)=1$ for every t.

Warning: In general this condition is weaker than $\mathbb{P}(X_t=X'_t \text{ for all } t)=1$.

If X,X' are continuous and X' is a version of X then $X=X'$. (Indeed, $X_t=X'_t$ for all t rational).

4.3 Martingales in continuous-time

Fortunately, we can assume that martingales in continuous time are càdlàg, due to the following theorem.

4.3.1 Definition. Let $(\mathscr{F}_t,t\in I)$ be a filtration. If $(\mathscr{F}_t)$ satisfies the two conditions below then it is said to satisfy the *usual conditions*.

(i) $\mathscr{F}_t=\mathscr{F}_{t^+}=\bigcap_{\varepsilon>0}\mathscr{F}_{t+\varepsilon}$ for all $t\in I$; and

(ii) $N\in\mathscr{F}_t$ for all $N\in\mathscr{F}$ such that $\mathbb{P}(N)=0$, for all t.

4.3.2 Theorem. *Let $(\mathscr{F}_t,t\geq 0)$ satisfy the usual conditions, and let $(X_t,t\geq 0)$ be an adapted martingale. Then there exists a version of X which is a càdlàg $(\mathscr{F}_t)$-martingale.*

4.4 Convergence theorems for continuous-time

From now on, all martingales $(X_t,t\geq 0)$ are assumed to be càdlàg, so $X_t=X_{t^+}$ and X_{t^-} exists for all t.

4.4.1 Theorem (Martingale convergence theorem).
Assume that $(X_t,t\geq 0)$ is a càdlàg martingale which is bounded in L^1. Then $X_t\to X_\infty$ a.s. as $n\to\infty$ for some finite X_∞.

PROOF: Let $N(X,I,[a,b])=\sup\{n\mid \exists s_1<t_1<s_2<t_2<\cdots<s_n<t_n \text{all in } I, \text{ such that} X_{s_i}<a<b<X_{t_i}\forall i\}=\sup_{J\subseteq I finite}N(X,J,[a,b])$. Since X is right-continuous, $N(X,\mathbb{R}_+,[a,b])=N(X,\mathbb{Q}_+,[a,b])$. Indeed, if $s_1<t_1<s_2<t_2<\cdots<s_n<t_n\in\mathbb{R}_+$ such as in the definition of $N(X,\mathbb{R}_+,[a,b])$ then we can find $s_1<s'_1<t_1<t'_1<\cdots<s_n<s'_n<t_n<t'_n$ such that $s'_i,t'_i\in\mathbb{Q}_+$ and $X_{s'_i}<a<b<X_{t'_i}$, which implies "$\leq$" (the reverse inequality is trivial).

Let $J=\{t_1,\ldots t_k\}\subseteq\mathbb{Q}_+$. Then $(X_{t_i},1\leq i\leq k)$ is martingale with respect to $(\mathscr{F}_{t_i},1\leq i\leq k)$. Therefore by Doob's Up-crossings Lemma

$$(b-a)\mathbb{E}[N(X,J,[a,b]]\leq\mathbb{E}[(X_{t_k}-a)^-]\leq\mathbb{E}[|X_{t_k}|]+a\leq M<\infty$$

for every J, by boundedness in L^1. Taking the supremum over J,

$$(b-a)\mathbb{E}[N(X,\mathbb{Q}_+,[a,b]] \le M < \infty$$

so for every $a < b$ rational, $N(X,\mathbb{R}_+,[a,b]) < \infty$ a.s. Since $\mathbb{Q}$ is countable, a.s. $N(X,\mathbb{R}_+,[a,b]) < \infty$ for every $a < b$ rational. Thus $X_t \to X_\infty \in \overline{\mathbb{R}}$. By the usual arugment with Fatou's Lemma, X_∞ is finite a.s. □

4.4.2 Theorem (Doob's Inequalities).
Let $(X_t, t \ge 0)$ be a càdlàg martingale and $X_t^* := \sup_{0\le s\le t} |X_s|$. For all $a \ge 0$,

$$a\,\mathbb{P}(X_t^* \ge a) \le \mathbb{E}[|X_t|]$$

and for $p > 1$,

$$\|X_t^*\|_p \le \frac{p}{p-1}\|X_t\|_p$$

for all t.

PROOF: Observe that $X_t^* = \max(X_t, \sup_{s\in[0,1],s\in\mathbb{Q}} |X_s|)$ by the càdlàg property. Therefore $X_t^* = \sup_{J\subseteq\{t\}\cup([0,t]\cap\mathbb{Q}) finite} \max_{s\in J} |X_s|$. Apply Doob's inequality to $(X_s, s \in J) = (X_{t_i} \mid 1 \le i \le k)$, where $J = \{t_1,\dots,t_k\}$. Hence

$$a\,\mathbb{P}(\max_{s\in J} |X_s| \ge a) \le \mathbb{E}[|X_{t_k}|] \le \mathbb{E}[|X_t|]$$

whenever $\sup J \le t$ because $(|X_t|, t \ge 0)$ is a sub-martingale. Then pass to the supremum over J.

The L^p-inequality is proved in the same way (exercise). □

4.4.3 Theorem (L^p convergence criteria).

(i) *For $p > 1$, X_t converges a.s. and in L^p if and only if X is bounded in L^p, and this if and only if $X_t = \mathbb{E}[Z \mid \mathscr{F}_t]$ for some $Z \in L^p$ (which may be taken to be X_∞).*

(ii) *For $p = 1$, X_t converges a.s. and in L^1 if and only if X is u.i., and this if and only if $X_t = \mathbb{E}[Z \mid \mathscr{F}_t]$ for all t for some $Z \in L^1$ (which may be taken to be X_∞).*

PROOF: As in the discrete case. □

4.4.4 Theorem (Optional stopping, càdlàg u.i.).
Let (X_t) be a u.i. càdlàg martingale and $S \le T$ be two stopping times. Then $\mathbb{E}[X_T \mid \mathscr{F}_S] = X_S$.

PROOF: Let $n \ge 1$ and let $T_n = 2^{-n}\lceil 2^n T\rceil$. Then T_n is a stopping time such that $T_n \searrow_{n\to\infty} T$. Moreover T_n takes its values in the set $D_n = \{k2^{-n} \mid k \in \mathbb{Z}_+ \cup \{\infty\}\}$. We want to compute

$$\mathbb{E}[X_\infty \mid \mathscr{F}_{T_n}] = \sum_{d\in D_n} \mathbf{1}_{T_n=d}\,\mathbb{E}[X_\infty \mid \mathscr{F}_{T_n}] = \sum_{d\in D_n} \mathbf{1}_{T_n=d}\,\mathbb{E}[X_\infty \mid \mathscr{F}_d]$$

(Indeed, we used that for T a stopping time and Z a r.v., $\mathbb{E}[Z \mid \mathscr{F}_T]\mathbf{1}_{T=t} = \mathbb{E}[Z \mid \mathscr{F}_t]\mathbf{1}_{T=t}$. But if $A \in \mathscr{F}_T$, $\mathbb{E}[\mathbf{1}_A\mathbf{1}_{T=t}Z] = \mathbb{E}[\mathbf{1}_A\mathbf{1}_{T=t}\,\mathbb{E}[Z \mid \mathscr{F}_t]] = \mathbb{E}[\mathbf{1}_A\mathbf{1}_{T=t}\,\mathbb{E}[Z \mid$

$\mathscr{F}_T]] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[\mathbf{1}_{T=t} Z \mid \mathscr{F}_T]]$ since $\mathbf{1}_{T=t} \in \mathscr{F}_T \cap \mathscr{F}_t$.) Finally, since $X_t \to X_\infty$ a.s. and in L^1 since X is u.i.,

$$\mathbb{E}[X_\infty \mid \mathscr{F}_d] = \lim_{n\to\infty} \mathbb{E}[X_t \mid \mathscr{F}_d]$$

implying that

$$\mathbb{E}[X_\infty \mid \mathscr{F}_{T_n}] = \sum_{d\in D_n} \mathbf{1}_{T_n=d} X_d = X_{T_n} \to X_T$$

as $n \to \infty$. Note that $(\mathbb{E}[X_\infty \mid \mathscr{F}_{T_{-n}}], n \le 0)$ is a backwards martingale with respect to $(\mathscr{F}_{T_{-n}}, n \ge 0)$. As such, $\mathbb{E}[X_\infty \mid \mathscr{F}_{T_{-n}}] \to \mathbb{E}[X_\infty \mid \bigcap_{n\ge 1} \mathscr{F}_{T_n}]$ a.s. as $n \to \infty$. Finally, $X_T = \mathbb{E}[X_\infty \mid \bigcap_{n\ge 1} \mathscr{F}_{T_n}]$, so condition on $\mathscr{F}_T$ and use the tower property to prove $X_T = \mathbb{E}[X_\infty \mid \mathscr{F}_T]$.

For general S, T, use $\mathbb{E}[X_T \mid \mathscr{F}_S] = \mathbb{E}[\mathbb{E}[X_\infty \mid \mathscr{F}_T] \mid \mathscr{F}_S] = \mathbb{E}[X_\infty \mid \mathscr{F}_S] = X_S$. □

When martingales are càdlàg they behave much like discrete parameter martingales.

4.5 Kolmorgorov's continuity criterion

4.5.1 Theorem. *Let $(X_t, 0 \le t \le 1)$ be a stochastic process with values in $\mathbb{R}^d$. Assume that there exists $C, p, \varepsilon > 0$ such that for every $s, t \in [0,1]$,*

$$\mathbb{E}[|X_t - X_s|^p] \le C|t-s|^{1+\varepsilon}.$$

Then there exists a version of X which is continuous, and further, α-Hölder-continuous with index $\alpha \in (0, \frac{\varepsilon}{p})$.

Recall that $f : I \to \mathbb{R}^d$ is *α-Hölder-continuous* if there is $C > 0$ such that $|f(t) - f(t)| \le C|t-s|^\alpha$ for all $s, t \in I$.

PROOF: See the supplemental notes. □

Remark. Suppose we are given a process $(X_t, t \in D)$ and satisfying the hypothesis of Kolmogorov's continuity criterion (only with $s, t \in D$ instead of $[0,1]$). Then by the very same proof one may extend X to a continuous process $(X_t, t \in [0,1])$.

5 Weak Convergence

5.1 Weak convergence for probability measures

We have a variety of notions of convergence for r.v.'s, including almost sure convergence, convergence in L^p, and convergence in probability. All these notions depend on the r.v.'s themselves. Weak convergence is a notion of convergence for (probability) *measures*. If $(\mu_n, n \ge 0)$ is a sequence of measures on $(\Omega, \mathscr{F})$, we could say that $\mu_n \to \mu$ if $\mu_n(A) \to \mu(A)$ for all $A \in \mathscr{F}$, but this form of convergence (i.e. pointwise convergence) is very rigid. Instead we use the following definition.

5.1.1 Definition. Let (M,d) be a metric space and endow it with its Borel σ-algebra. Let $(\mu_n, n \geq 1)$ be a sequence of probability measures defined on $(M, \mathscr{B}(M))$, and let μ be another (probability) measure. We say that (μ_n) *converges weakly* to μ if for every continuous bounded function $f : M \to \mathbb{R}$, $\mu_n(f) \to \mu(f)$ as $n \to \infty$, and write $\mu_n \xrightarrow{(w)} \mu$.

Notation. $C_b(M)$ is the set of continuous bounded functions $M \to \mathbb{R}$.

5.1.2 Examples.

(i) $x_n \to x$ in $\mathbb{R}$ if and only if $\delta_{x_n} \xrightarrow{(w)} \delta_x$. Indeed, the latter is equivalent to requiring that $f(x_n) \to f(x)$ for every continuous bounded function f.

(ii) Let $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{\frac{k}{n}}$ on $[0,1]$. Then $\mu_n \xrightarrow{(w)} dx$, Lebesgue measure on $[0,1]$. Indeed, $\mu_n(f)$ is a Riemann sum of f.

5.1.3 Theorem. *Let $(\mu_n, n \geq 1)$ and μ be probability measures on M. The following are equivalent.*

(i) $\mu_n \to \mu$ weakly.

(ii) For every $G \subseteq M$ open, $\liminf_n \mu_n(G) \geq \mu(G)$.

(iii) For every $F \subseteq M$ closed, $\limsup_n \mu_n(F) \leq \mu(F)$.

(iv) For every $A \subseteq M$ Borel such that $\mu(\partial A) = 0$, $\lim_n \mu_n(A) = \mu(A)$.

The idea is the mass can be "won or lost" only through the boundary.

PROOF: (i) implies (ii): Let G be an open set and introduce the sequence of functions $f_k(x) := 1 \wedge kd(x, G^c)$ (for $x \in M$). For each $k \geq 1$, f_k is continuous and bounded (in fact, uniformly continuous), so $\mu_n(f_k) \to \mu(f_k)$. As G is open, $x \in G$ if and only if $d(x, G^c) > 0$. Thus $f_k \nearrow \mathbf{1}_G$ pointwise as $k \to \infty$. For every n, $\mu_n(G) = \mu_n(\mathbf{1}_G) \geq \mu_n(f_k)$, so $\liminf_n \mu_n(G) \geq \mu(f_k)$ for all k. By the MCT, taking $k \to \infty$ we have $\liminf_n \mu_n(G) \geq \mu(G)$.

(ii) if and only if (iii): Let F be a closed set. Then

$$\mu(F) = 1 - \mu(F^c) \geq 1 - \liminf_n \mu_n(F^c) = \liminf_n \mu_n(F)$$

(ii) & (iii) imply (iv): Let A be a Borel set such that $\mu(\partial A) = 0$. Then

$$\begin{aligned}\limsup_n \mu_n(A) &\leq \limsup_n \mu_n(\overline{A}) \\ &\leq \mu(\overline{A}) = \mu(A^\circ \cup \partial A) = \mu(A^\circ) \\ &\leq \liminf_n \mu_n(A^\circ) \leq \liminf_n \mu_n(A).\end{aligned}$$

It follows that the limit exists is equals $\mu(A)$.

(iv) implies (i): Let $f \in C_b^+(M)$. We must show that $\mu_n(f) \to \mu(f)$ as $n \to \infty$. Write $K = \|f\|_\infty$, so

$$\mu_n(f) = \int_0^\infty \mu_n(f \geq t)dt = \int_0^K \mu_n(f \geq t)dt.$$

Note that the set $\{f \geq t\} = f^{-1}([t,\infty))$ is closed, while the set $\{f > t\}$ is open, so $\partial\{f \geq t\} \subseteq \{f = t\}$. The set $A := \{t \mid \mu(f = t) > 0\}$ is countable since it is equal to $\bigcup_{n\geq 1}\{t \mid \mu(f = t) \geq n^{-1}\}$, and each set in the union is finite since μ is a probability measure. Therefore

$$\mu_n(f) = \int_{[0,K]\setminus A} \mu_n(f \geq t)dt \to \int_{[0,K]\setminus A} \mu(f \geq t)dt = \int_0^K dt\mu_n(f \geq t) = \mu(f)$$

by DCT, since by (iv), $\mu_n(f \geq t) \to \mu(f \geq t)$ for all $t \notin A$. Finally, to check the definition of weak convergence, decompose elements of $C_b(M)$ into positive and negative parts. □

For a probability measure μ on $\mathbb{R}$, let $F_\mu(x) := \mu((-\infty, x])$ denote the *distribution function* of μ. Note that F_μ is increasing, càdlàg and $\lim_{x\to-\infty} F_\mu(x) = 0$ and $\lim_{x\to+\infty} F_\mu(x) = 1$.

5.1.4 Corollary. *Let μ_n, μ be probability measures on $\mathbb{R}$. The following are equivalent.*

(i) $\mu_n \to \mu$ weakly.

(ii) $F_{\mu_n}(x) \to F_\mu(x)$ for every $x \in \mathbb{R}$ that is a continuity point of F_μ.

PROOF: (i) implies (ii) is (iv) in the previous theorem. Indeed, $\partial(-\infty, x] = \{x\}$, so if F_μ continuous at x then $\mu(\{x\}) = 0$.

(ii) implies (i): Let G be open in $\mathbb{R}$. We may write $G = \bigcup_{k\geq 0}(a_k, b_k)$ for pairwise disjoint intervals (a_k, b_k). Then $\mu_n(G) = \sum_{k\geq 0}\mu_n(a_k, b_k)$. Now,

$$\mu_n(a_k, b_k) = F_{\mu_n}(b_k-) - F_{\mu_n}(a_k) \geq F_{\mu_n}(b) - F_{\mu_n}(a)$$

for every $a_k < a < b < b_k$. Choosing a and b to be continuity points of F_μ (such points are dense in $\mathbb{R}$), we obtain $\liminf \mu_n(a_k, b_k) \geq F_\mu(b) - F_\mu(a)$. Letting $a \searrow a_k$ and $b \nearrow b_k$ along continuity points of F_μ, we obtain

$$\liminf \mu_n(a_k, b_k) \geq F_\mu(b_k-) - F_\mu(a_k) = \mu(a_k, b_k).$$

Finally,

$$\liminf_n \mu_n(G) \geq \sum_{k\geq 0} \liminf_n \mu_n(a_k, b_k) \geq \sum_{k\geq 0} \mu(a_k, b_k) = \mu(G).$$

□

5.2 Convergence in distribution for random variables

5.2.1 Definition. Let $(X_n, n \geq 0)$, X be r.v.'s taking values (M, d). We say that X_n *converges in distribution* to X as $n \to \infty$ if $\mathscr{L}(X_n) \xrightarrow{(w)} \mathscr{L}(X)$, and we write $X_n \xrightarrow{(d)} X$.

By definition of weak convergence $X_n \xrightarrow{(d)} X$ if and only if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in C_b(M)$.

5.2.2 Proposition. *Let X_n, X be r.v.'s. If $X_n \to X$ in probability, i.e.*

$$\lim_{n\to\infty} \mathbb{P}(d(X_n,X) > \varepsilon) \to 0 \quad \textit{for all } \varepsilon > 0$$

then $X_n \to X$ in distribution. If $X_n \to c$ in distribution for some constant c then $X_n \to c$ in probability.

PROOF: Exercise (see example sheet 3). □

5.2.3 Examples.

(i) If $X_n = x_n$ is constant and $x_n \to x$ then $X_n \xrightarrow{(d)} X$ since $\delta_{x_n} \xrightarrow{(w)} \delta_x$.

(ii) Let U be a uniform r.v. in $[0,1]$, and write $X_n = n^{-1}\lfloor nU \rfloor$. Then $X_n \to U$ a.s., so $X_n \xrightarrow{(d)} U$. Indeed, $\mathscr{L}(X_n) = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{\frac{k}{n}} \xrightarrow{(w)} dx = \mathscr{L}(U)$.

(iii) Central Limit Theorem. Let $X_1, X_2, \dots$ be i.i.d. r.v.'s in $L^2(\mathbb{R})$. If $\mu := \mathbb{E}[X]$ and $\sigma^2 := \mathrm{Var}(X) = \mathbb{E}[X^2]$ then for all $a < b$,

$$\mathbb{P}\left(a < \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} < b\right) \to \int_a^b \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

which is saying that $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \xrightarrow{(d)} N$ in distribution, where N is a Gaussian r.v. with mean 0 and variance 1.

5.3 Tightness

5.3.1 Definition. Let $(\mu_i)_{i\in I}$ be a family of probability measures on (M,d). We say that $(\mu_i)_{i\in I}$ is *tight* if for all $\varepsilon > 0$ there is $K_\varepsilon \subseteq M$ compact such that $\sup_{i\in I} \mu_i(M \setminus K_\varepsilon) < \varepsilon$.

5.3.2 Theorem (Prokhorov). *Let $(\mu_n, n \ge 0)$ be probability measures on (M,d) such that $(\mu_n, n \ge 0)$ is tight. Then there is a subsequence $(\mu_{n_k}, k \ge 0)$) and a probability measure μ such that $\mu_{n_k} \xrightarrow{(w)} \mu$ as $k \to \infty$.*

To motivate this theorem, notice that for (δ_{x_n}), tightness is exactly the requirement that the x_n are contained in a compact set, so Prokhorov's Theorem may (loosely) be seen as a generalization of the Bolzano-Weierstrass Theorem.

PROOF: We prove the case $M = \mathbb{R}$. Consider $(F_{\mu_n}(q), n \ge 0) \subseteq [0,1]$ for q a rational number. There exists a subsequence n_k^q such that $F_{\mu_{n_k^q}}(q) \to F(q)$ (i.e. the subsequence has a limit in $[0,1]$). Since $\mathbb{Q}$ is countable, by a diagonal procedure we can find $(n_k, k \ge 0)$ such that $F_{n_k(q)} \to F(q)$ for all $q \in \mathbb{Q}$. Then F is non-decreasing on $\mathbb{Q}$, so extend it to $\mathbb{R}$ via $F(t) = \lim_{q\searrow t, q\in\mathbb{Q}} F(q)$. Further, F is càdlàg.

Check that for every continuity point t of F, $F_{\mu_{n_k}}(t) \to F(t)$ as $k \to \infty$. Let $\varepsilon > 0$, and pick $\eta > 0$ such that $F(t+\eta) \le F(t) + \varepsilon$ and $t + \eta \in \mathbb{Q}$. Then $F_{\mu_{n_k}}(t+\eta) \to F(t+\eta) \le F(t) + \varepsilon$. Thus

$$\limsup_k F_{\mu_{n_k}}(t) \le \limsup_k F_{\mu_{n_k}}(t+\eta) = F(t+\eta) \le F(t) + \varepsilon,$$

so $\limsup F_{\mu_{n_k}}(t) \le F(t)$. Similarly we get that $F(t)$ is less than the limit infimum.

We want to show that $F = F_\mu$ for some probability measure μ. It is sufficient to check that $F(t) \to 0, 1$ as $t \to -\infty, +\infty$. By tightness we can find $K = [A, B]$ such that $\mu_n((B, \infty)) < \varepsilon + \mu_n((-\infty, A))$ for all $n \ge 0$. But $\varepsilon > \mu_{n_k}((B, \infty)) = 1 - F_{n_k}(B) \to 1 - F(B)$ and $\varepsilon > \mu_{n_k}((-\infty, A)) = F_{n_k}(A) \to F(A)$. Therefore $F = F_\mu$ for some probability μ and $F_{\mu_{n_k}} \to F_\mu$ at every continuity point of F_μ. Therefore $\mu_{n_k} \to \mu$ weakly. □

5.4 Levy's convergence theorem

5.4.1 Definition. Let X be a r.v. taking values in $\mathbb{R}^d$. The *characteristic function* of X is $\phi_X : \mathbb{R}^d \to \mathbb{C} : \xi \mapsto \mathbb{E}[\exp(i\langle \xi, X\rangle)]$.

5.4.2 Theorem (Levy).

(i) Let $(X_n, n \ge 0)$, X be r.v.'s such that $X_n \xrightarrow{(d)} X$. Then $\phi_{X_n}(\xi) \to \phi_X(\xi)$ for all $\xi \in \mathbb{R}^d$.

(ii) Let $(X_n, n \ge 0)$ be r.v.'s such that $\phi_{X_n}(\xi) \to \phi(\xi)$ for all $\xi \in \mathbb{R}^d$ for some function $\phi : \mathbb{R}^d \to \mathbb{C}$ that is continuous at 0. There is a r.v. X such that $\phi = \phi_X$ and $X_n \xrightarrow{(d)} X$.

PROOF:

(i) This is a consequence of the definition of convergence in distribution, since $x \to \exp(i\langle \xi, X\rangle)$ is continuous and bounded for every ξ. (Recall that $X_n \to X$ in distribution if and only if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in C_b(\mathbb{R}^d)$.)

(ii) We shall prove the case $d = 1$ (see the notes for the general proof). We must control quantities of the form $\sup_n \mathbb{P}(|X_n| \ge K)$.

Claim. This is a universal constant $C > 0$ such that for every r.v. X taking values in $\mathbb{R}$,

$$\mathbb{P}(|X| \ge K) \le CK \int_0^{\frac{1}{K}} (1 - \Re\phi_X(u))du.$$

Indeed, $\Re\phi_X(u) = \Re\,\mathbb{E}[\exp(iuX)] = \mathbb{E}[\cos(uX)]$. By Fubini's Theorem,

$$K \int_0^{\frac{1}{K}} (1 - \Re\phi_X(u))du = \mathbb{E}\left[1 - K \int_0^{\frac{1}{K}} \cos(uX)du\right] = \mathbb{E}\left[1 - \frac{\sin(\frac{X}{K})}{\frac{X}{K}}\right]$$

Consider $x \mapsto 1 - \frac{\sin x}{x}$. There is C such that $\mathbf{1}_{|x|\ge 1} \le C(1 - \frac{\sin x}{x})$. We obtain

$$CK \int_0^{\frac{1}{K}} (1 - \Re\phi_X(u))du = C\,\mathbb{E}\left[1 - \frac{\sin(\frac{X}{K})}{\frac{X}{K}}\right] \ge \mathbb{E}[\mathbf{1}_{|\frac{X}{K}|\ge 1}] = \mathbb{P}(|X| \ge K).$$

Now,

$$\mathbb{P}(|X_n| \ge K) \le CK \int_0^{\frac{1}{K}} (1 - \Re\phi_{X_n}(u))du \to CK \int_0^{\frac{1}{K}} (1 - \Re\phi(u))du$$

as $n \to \infty$, by the DCT, since $\phi_{X_n}(u) \to \phi(u)$. Since ϕ is continuous at zero and $\phi(0) = 1$, for any $\varepsilon > 0$ there is K large enough such that $|1 - \Re\phi(u)| \le \frac{\varepsilon}{2C}$ for any $u \in [0, \frac{1}{K}]$. Thus, for such K, $\limsup_n \mathbb{P}(|X_n| \ge K) \le \frac{\varepsilon}{2}$. Thus there is N_ε such that for all $n > N_\varepsilon$, $\mathbb{P}(|X_n| \ge K) \le \varepsilon$. Up to taking K even larger, we may assume that $\max_{1 \le n \le N_\varepsilon} \mathbb{P}(|X_n| \ge K) \le \varepsilon$. Now for such K, $\sup_n \mathbb{P}(|X_n \ge K) \le \varepsilon$ if and only if $\sup_n \mu_n([-K,K]^C) \le \varepsilon$, where $\mu_n = \mathscr{L}(X_n)$. Therefore $(\mathscr{L}(X_n))_{n \ge 1}$ is a tight family of measures. By Prokhorov's Theorem, there is a subsequence such that $\mathscr{L}(X_{n_k}) \xrightarrow{(w)} \mu$, for some probability measure μ on $\mathbb{R}$. Hence $X_{n_k} \xrightarrow{(d)} X$ for some r.v. X. But by part (i), $\phi_{X_{n_k}} \to \phi_X$ pointwise, so $\phi = \phi_X$ is the characteristic function of a r.v. Let us assume that X_n does not converge in distribution to X. Then we could find $f \in C_b(\mathbb{R})$ such that $\mathbb{E}[f(X_n)]$ does not converge to $\mathbb{E}[f(X)]$. Thus we could find an extraction n_k and $\varepsilon > 0$ such that $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \ge \varepsilon$. But $(\mathscr{L}(X_{n_k}), k \ge 1)$ is tight (since it is contained in a tight family), so we can further extract a subsequence such that $X_{n_{k_r}} \to X'$ in distribution for some r.v. X'. Similarly, $\phi_X = \phi_{X'}$, so X and X' have the same distribution. This is a contradiction since $\mathbb{E}[f(X_{n_{k_r}})] \to \mathbb{E}[f(X)]$. □

6 Brownian Motion

6.1 Historical notes

Around 1827 Brown observed the erratic motion of pollen particles in water. Later, Langevin and Einstein realized that Brownian motion is an isotropic Gaussian process. The first mathematical construction of Brownian motion is due to Wiener in 1923, using Fourier Analysis. Lévy studied the sample path properties of Brownian motion, and Kakutani and Doob made the link with potential theory. Itô's calculus was developed in 1950.

As a first approximation, consider $S_n = X_1 + \cdots + X_n$, where the X_j are $\mathbb{R}^d$-valued i.i.d. r.v.'s and $X_1 = \pm e_i$ with probability $\frac{1}{2d}$, where $e_i = (0, \dots, 0, 1, 0 \dots, 0)$. By the CLT,

$$\frac{S_n}{\sqrt{n}} \xrightarrow{(d)} \mathscr{N}(0, I_d) \overset{(d)}{=} \vec{N},$$

where N_i are i.i.d. $\mathscr{N}(0,1)$. Consider $B_t^{(n)} = \frac{S_{\lfloor tn \rfloor}}{\sqrt{n}}$ for $t \ge 0$, so in particular $\mathscr{L}(B_1^{(n)}) \xrightarrow{(w)} \mathscr{N}(0, I_d)$. Brownian motion is a continuous process B such that "$B^{(n)} \to B$ in distribution."

The finite dimensional marginal distributions $\mathscr{L}(B_{t_1}, \dots, B_{t_k})$ for every k should be given by

$$\lim_{n \to \infty} \mathscr{L}\left(B_{t_1}^{(n)}, \dots, B_{t_k}^{(n)}\right).$$

6.1.1 Proposition. *For fixed* $0 = t_0 < t_1 < \cdots < t_k$, $(B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)}, 1 \le i \le n)$ *converges in distribution to* $(N_i, 1 \le i \le k)$*, where the* N_i*'s are independent and* $N_i \sim \mathscr{N}(0, (t_i - t_{i-1}) I_d)$*. In particular,* $\mathscr{L}(B_t^{(n)}) \to \mathscr{N}(0, t I_d)$.

PROOF: (For $d=1$.) Using characteristic functions. Let $\xi \in \mathbb{R}^k$.

$$\begin{aligned}
\mathbb{E}\left[\exp\left(i\sum_{j=1}^{k}\xi_j(B^{(n)}_{t_i}-B^{(n)}_{t_{i-1}})\right)\right] &= \prod_{j=1}^{k}\mathbb{E}\left[\exp\left(i\xi_j(B^{(n)}_{t_i}-B^{(n)}_{t_{i-1}})\right)\right] \\
&= \prod_{j=1}^{k}\mathbb{E}\left[\exp\left(i\xi_j\sum_{r=1}^{\lfloor nt_j\rfloor-\lfloor nt_{j-1}\rfloor}X_r\right)\right] \\
&\xrightarrow{CLT} \prod_{j=1}^{k}\mathbb{E}\left[\exp(i\xi_j N_j)\right] \\
&= \mathbb{E}\left[\exp\left(i\sum_{j=1}^{k}\xi_j N_j\right)\right]
\end{aligned}$$

since $B^{(n)}_{t_i}-B^{(n)}_{t_{i-1}} = \frac{1}{\sqrt{n}}\sum_{r=\lfloor nt_{j-1}\rfloor+1}^{\lfloor nt_j\rfloor}X_r$, so the $\{B^{(n)}_{t_i}-B^{(n)}_{t_{i-1}}\}$ are k independent r.v.'s. And further

$$B^{(n)}_{t_i}-B^{(n)}_{t_{i-1}} = \frac{1}{\sqrt{n}}\sum_{r=1}^{\lfloor nt_j\rfloor-\lfloor nt_{j-1}\rfloor}X_r.$$

Conclude with Lévy's Theorem. □

6.1.2 Definition. Let $(B_t)_{t\geq 0}$ be a stochastic process in $\mathbb{R}^d$. We say that (B_t) is a *standard Brownian motion* if

(i) $B_0 = 0$;

(ii) For $0 = t_0 < t_1 < \cdots < t_k$, $(B_{t_i}-B_{t_{i-1}}, 1 \leq i \leq k)$ are independent;

(iii) $B_t - B_s$ has distribution $\mathcal{N}(0,(t-s)I_d)$ for all $s < t$.

(iv) For all $\omega \in \Omega$, $t \mapsto B_t(\omega)$ is continuous.

Here "standard" refers to the fact that $B_0 = 0$ and $\operatorname{Cov} B_1 = I_d$. The finite dimensional marginal distribution for $t_1 < \cdots < t_k$ is the law of $(B_{t_1},\dots,B_{t_k})$. Let

$$p_t(x) = \frac{1}{(2\pi t)^{\frac{d}{2}}}e^{-\frac{|x|^2}{2t}}$$

be the probability density function of $\mathcal{N}(0,tI_d)$. Then

$$\mathbb{E}[F(B_{t_1},\dots,B_{t_k})] = \int_{(\mathbb{R}^d)^k}F(x_1,\dots,x_k)\prod_{i=1}^{k}p_{t_i-t_{i-1}}(x_i-x_{i-1})dx_1\cdots dx_k$$

With a appropriate change of variables, we may write

$$\mathbb{E}[G(B_{t_i}-B_{t_{i-1}}, 1\leq i\leq k)] = \int_{(\mathbb{R}^d)^k}G(x_1,\dots,x_k)\prod_{i=1}^{k}p_{t_i-t_{i-1}}(x_i)dx_1\cdots dx_k$$

6.1.3 Theorem (Weiner, 1923). *There exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ on which a standard Brownian motion is defined.*

PROOF: Let $D_n = \{k2^{-n} \mid 0 \le k < 2^n\}$ for $n \ge 1$, and $D_0 = \{0,1\}$. Let us construct $(B_d)_{d\in D}$, where $D = \bigcup_n D_n$ are the dyadic rationals, that satisfy (i), (ii), and (iii) of the definition. (We are doing the case $d = 1$ on the time interval $[0,1]$.) Let $(Z_d)_{d\in D}$ be a family of independent r.v.'s with distribution $\mathcal{N}(0,1)$. We'll construct B_d such that for all d, $B_d \in \text{Vect}(Z_{d'}, d' \in D) =: G$. For $X_1,\dots,X_k \in G$, the (X_i)'s are independent if and only if they are pairwise orthogonal, i.e. $\mathbb{E}[X_iX_j] = 0$ for all $i \ne j$.

The construction is by induction. For $n = 0$ take $B_0 = 0$ and $B_1 = Z_0$. Suppose now that $(B_d)_{d\in D_{n-1}}$ satisfies properties (i), (ii), and (iii) of the definition, and that $(B_d)_{d\in D_{n-1}}$ is independent of $(Z_d)_{d\in D\setminus D_{n-1}}$. Let $d \in D_n \setminus D_{n-1}$, and defined $d_- = d - 2^{-n}$ and $d_+ = d + 2^{-n}$, so $d_-, d_+ \in D_{n-1}$. Let $B_d = \frac{1}{2}(B_{d_+} + B_{d_-}) + \frac{1}{2^{\frac{n+1}{2}}} Z_d$. Now $B_d - B_{d_-}$, $B_d - B_{d_+}$ are independent $\mathcal{N}(0, \frac{1}{2^n})$ (since (check that) if N_1, N_2 are $\mathcal{N}(0,\sigma^2)$ and independent then $N_1 + N_2$ and $N_1 - N_2$ are independent $\mathcal{N}(0, 2\sigma^2)$. Indeed, $\text{Cov}(N_1 + N_2), N_1 - N_2) = \text{Cov}(N_1, N_1) - \text{Cov}(N_2, N_2) = 0$.)

Further, the same method (check this) proves that these increments are independent of all other $B_{d'} - B_{d'_-}$) for $d' \in D \setminus D_{n-1}$, so the induction step is proved. By induction we get $(B_d)_{d\in D}$ satisfying (i), (ii), (iii) of the definition.

For $s,t \in D$, $s < t$, $\mathbb{E}[|B_s - B_t|^p] = |t-s|^{\frac{p}{2}}\, \mathbb{E}[N^p]$ for $p > 2$, where $N \sim \mathcal{N}(0,1)$, since $B_s - B_t \sim \mathcal{N}(0, t-s) \sim \sqrt{t-s}\mathcal{N}(0,1)$. But this latter is $C_p|t-s|^{1+(\frac{p}{2}-1)}$ for some $C_p > 0$. By the Kolmogorov Criterion, there exists a.s. a continuous function $(B_t)_{t\in[0,1]}$ that extends $(B_d)_{d\in D}$. Let $\Omega_0 = \{\omega \mid (B_d(\omega))_{d\in D}\}$ does not have a continuous extension to $[0,1]$. On Ω_0, let $B_t(w) = 0$ for all $t \in [0,1]$. (Notice that $\mathbb{P}(\Omega_0) = 0$.)

Let us check (i), (ii), (iii) for $(B_t)_{t\in[0,1]}$. (i) is trivial. Let $0 = t_0 < t_1 < \cdots < t_k$, and consider $0 \le t_1^n < \cdots < t_k^n$ such that $t_i^n \in D_n$ for all n and $t_i^n \to t_i$ as $n \to \infty$. We know that $(B_{t_i^n} - B_{t_{i-1}^n}, 1 \le i \le k)$ are independent $\mathcal{N}(0, t_i^n - t_{i-1}^n)$ for all n.

Claim. If $(N_1^n,\dots,N_k^n)$ are independent Gaussian $N_i \sim \mathcal{N}(0, (\sigma_i^n)^2)$, then if $\sigma_i^n \to \sigma_i$ then $(N_1^n,\dots,N_k^n) \to (N_1,\dots,N_k)$ in distribution, where N_i are independent $N_i \sim \mathcal{N}(0, \sigma_i^2)$.

Indeed, use Lévy's convergence theorem.

By continuity of B, $B_{t_i^n} - B_{t_{i-1}^n} \to B_{t_i} - B_{t_{i-1}}$ for every ω, so by the lemma we do have $(B_{t_i} - B_{t_{i-1}}, 1 \le i \le k)$ are independent $\mathcal{N}(0, t_i - t_{i-1})$, so $(B_t)_{t\in[0,1]}$ is a Brownian motion.

To get a Brownian motion $(B_t)_{t\ge0}$, take Brownian motions $(B_t^n)_{t\in[0,1]}$ independent on $[0,1]$ for $n \ge 1$ and define $B_t = \sum_{n=0}^{\lfloor t\rfloor-1} B_1^n + B_{t-\lfloor t\rfloor}^{\lfloor t\rfloor}$ for $t \ge 0$. In higher dimensions $d > 1$ take $B^{(1)},\dots,B^{(d)}$ independent Brownian motion in $\mathbb{R}$ and set $B_t = (B_t^{(1)},\dots,B_t^{(d)})$, $t \ge 0$, which is a Brownian motion in $\mathbb{R}^d$ (check). □

6.1.4 Definition. Let $\Omega_W = C(\mathbb{R}_+, \mathbb{R}^d)$, the *Weiner space*. The *Weiner measure* $W_0(dw)$ is the law of the standard Brownian motion, the unique measure on Ω_W such that for every $A_1,\dots,A_k \in \mathscr{B}(\mathbb{R}^d)$,

$$W_0(\{w \in \Omega_W \mid w_{t_1} \in A_1,\dots,w_{t_k} \in A_k\}) = \int_{A_1\times\cdots\times A_k} \prod_{i=1}^k p_{t_i - t_{i-1}}(x_i - x_{i-1})\,dx_1\cdots dx_k.$$

Here Ω_W is endowed with the product σ-algebra, i.e. the smallest σ-algebra such that $X_t : w \mapsto w_t$ is measureable for all $t \in \mathbb{R}_+$. Under the probability space $(\Omega_W, P, W_0(dw))$, the process $(X_t, t \geq 0)$ (where $X_t(w) = w_t$) is a standard Brownian motion, called the *canonical Brownian motion*.

6.2 First properties

6.2.1 Proposition.

(i) *Let $(B_t, t \geq 0)$ be a standard Brownian motion in $\mathbb{R}^d$, and let $U \in O(d)$ be an orthogonal matrix, then $(UB_t, t \geq 0)$ is a standard Brownian motion as well. In particular, $(-B_t, t \geq 0)$ is a standard Brownian motion.*

(ii) *(Scaling) Let $\lambda > 0$, then $(\frac{1}{\sqrt{\lambda}} B_{\lambda t}, t \geq 0)$ is a standard Brownian motion.*

(iii) *(Simple Markov property) Let $B^{(t)} := (B_{t+s} - B_t, s \geq 0)$, then $B^{(t)}$ is a standard Brownian motion, independent of $\mathscr{F}_t^B$, where $\mathscr{F}_t^B = \sigma(B_s, 0 \leq s \leq t)$.*

PROOF:

(i) Use the fact that $N \sim \mathcal{N}(0, \sigma^2 I_d)$ then $UN \sim \mathcal{N}(0, \sigma^2 I_d)$.

(ii) Use $\frac{1}{\sqrt{\lambda}} B_{\lambda t} \sim \frac{1}{\sqrt{\lambda}} \mathcal{N}(0, \lambda t) \sim \mathcal{N}(0, t)$ and apply this fact to the increments of $(\frac{1}{\sqrt{\lambda}} B_{\lambda t}, t \geq 0)$.

(iii) Let $t_1, \ldots, t_k \leq t \leq s_1, \ldots, s_{k'}$, and let $F, G \geq 0$ be measurable functions. Then $\mathbb{E}[F(B_{t_1}, \ldots, B_{t_k}) G(B_{s_1}^{(t)}, \ldots, B_{s_{k'}}^{(t)})]$ is going to split because of the independent increment property of Brownian motion (since $B_{s_i}^{(t)} - B_{s_{i-1}}^{(t)} = B_{s_i} - B_{s_{i-1}}$). □

6.2.2 Proposition (Blumenthal's 0-1 law). *Let $(B_t, t \geq 0)$ be a standard Brownian motion and let $(\mathscr{F}_t^B)$ be its natural filtration, and $\mathscr{F}_{0^+}^B = \bigcap_{t>0} \mathscr{F}_t^B$. Then for all $A \in \mathscr{F}_{0^+}^B$, $\mathbb{P}(A) \in \{0, 1\}$.*

PROOF: Let $A \in \mathscr{F}_{0^+}^B$, so for all $\varepsilon > 0$, $A \in \mathscr{F}_\varepsilon^B$. Let F be a continuous bounded function $(\mathbb{R}^d)^k \to \mathbb{R}$ and $t_1, \ldots, t_k > 0$. Then

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_A F(B_{t_1}, \ldots, B_{t_k})] &= \lim_{\varepsilon \to 0} \mathbb{E}[\mathbf{1}_A F(B_{t_1-\varepsilon}^{(\varepsilon)}, \ldots, B_{t_k-\varepsilon}^{(\varepsilon)})] && \text{by DCT} \\
&= \lim_{\varepsilon \to 0} \mathbb{P}(A) \mathbb{E}[F(B_{t_1-\varepsilon}^{(\varepsilon)}, \ldots, B_{t_k-\varepsilon}^{(\varepsilon)})] && \text{SM} \\
&= \mathbb{P}(A) \mathbb{E}[F(B_{t_1}, \ldots, B_{t_k})]
\end{aligned}$$

This implies that A is independent of $\sigma(B_{t_i}, 1 \leq i \leq k)$, for all $t_1, \ldots, t_k > 0$. Therefore A is independent of $\sigma(B_s, s \geq 0) = \mathscr{F}_\infty^B \supseteq \mathscr{F}_{0^+}^B$, so $\mathscr{F}_{0^+}^B$ is independent of itself. Thus for any $A \in \mathscr{F}_{0^+}^B$ $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, so $\mathbb{P}(A) \in \{0, 1\}$. □

6.2.3 Proposition. *Let $(B_t, t \geq 0)$ be standard Brownian motion in dimension $d = 1$. Let $S_t = \sup_{0 \leq s \leq t} B_s$ and $I_t = \inf_{0 \leq s \leq t} B_s$. Then*

(i) *a.s. for all $t > 0$, $S_t > 0$ and $I_t < 0$;*

(ii) *a.s. $S_\infty = \sup_{t \geq 0} B_t = +\infty$ and $I_\infty = \inf_{t \geq 0} B_t = -\infty$;*

(iii) *In dimensions $d \geq 2$, let C be an open cone with origin at 0 and non-empty interior. Let $H_C = \inf\{t > 0 \mid B_t \in C\}$. Then a.s. $H_C = 0$.*

PROOF:

(i) Let ε_k be a positive decreasing sequence with $\varepsilon_k \to 0$ as $k \to \infty$. Let $A_k = \{B_{\varepsilon_k} > 0\}$, so $\limsup_{k\to\infty} A_k = \{B_{\varepsilon_k} > 0 \text{ i.o.}\} \in \mathscr{F}^B_\varepsilon$ for any $\varepsilon > 0$. Therefore it is in $\mathscr{F}^B_{0^+}$. By the reverse Fatou Lemma,

$$\mathbb{P}(\limsup_k A_k) \geq \limsup_k \mathbb{P}(A_k) = \limsup_k \mathbb{P}(B_{\varepsilon_k} > 0) = \frac{1}{2}.$$

On the other hand, by Blumenthal's Law, $\mathbb{P}(\limsup_k A_k) \in \{0, 1\}$, so it must be one. But $\limsup_k A_k \subseteq \{S_t > 0 \mid \forall\ t > 0\}$ we have the result. $(-B_t, t \geq 0)$ is a Brownian motion, so the result for I_t follows.

(ii) $S_\infty = \sup_{t\geq 0} B_t = \sup_{t\geq 0} B_{\lambda t}$ for any $\lambda > 0$. By the scaling property, $(B_{\lambda t}, t \geq 0) \stackrel{(d)}{=} (\sqrt{\lambda} B_t, t \geq 0)$. Therefore $S_\infty \stackrel{(d)}{=} \sqrt{\lambda} \sup_{t\geq 0} B_t = \sqrt{\lambda} S_\infty$ for all $\lambda > 0$, so $\mathbb{P}(S_\infty \in (\varepsilon, \varepsilon^{-1})) = \mathbb{P}(\lambda S_\infty \in (\varepsilon, \varepsilon^{-1})) \to 0$ as $\lambda \to \infty$. Therefore $S_\infty \in \{0, \infty\}$. But $S_\infty \geq S_1 > 0$ a.s. by (i). Same for I_t.

(iii) Copy the proof of (i), $A_k = \{B_{\varepsilon_k} \in C\}$, and note that since $\lambda C = C$ for all $\lambda > 0$, $\mathbb{P}(A_k) = \mathbb{P}(\mathscr{N}(0, I_d) \in C) > 0$ since C has non-empty interior (and the density is everywhere positive?). □

Since $B_t - B_s \sim \mathscr{N}(0, (t-s)I_d)$ we have, for $p > 2$,

$$\mathbb{E}[|B_t - B_s|^p] \leq C_p |t-s|^{1+(\frac{p}{2}-1)}.$$

By Kolmogorov's continuity criterion, $(B_t, t \geq 0)$ is Hölder continuous with index $\alpha \in (0, \frac{1}{2} - \frac{1}{p})$ over any compact interval. Letting $p \to \infty$ we see that, a.s., $(B_t, t \geq 0)$ is α-Hölder continuous for $\alpha \in (0, \frac{1}{2})$ over any compact set.

6.2.4 Proposition. *Almost surely, $(B_t, t \geq 0)$ is not $\frac{1}{2}$-Hölder continuous over any interval with non-empty interior.*

6.2.5 Proposition. *Let $(B_t, t \geq 0)$ be a standard Brownian motion in dimension $d = 1$.*

(i) $\limsup_{t\to 0} \frac{B_t}{\sqrt{t}} = +\infty$ *a.s., and* $\liminf_{t\to 0} \frac{B_t}{\sqrt{t}} = -\infty$ *a.s.*

(ii) *For all $t_0 \geq 0$, (B_t) is not differentiable at t_0, a.s.*

(iii) *(B_t) is not differentiable at $t = t_0$ for any t_0, a.s.*

6.3 Strong Markov property, Reflection principle

We would like to consider starting a Brownian motion from a random initial value. A process $(B_t, t \geq 0)$ is a *Brownian motion* (starting at a random point B_0) if $(B_t - B_0, t \geq 0)$ is a standard Brownian motion that is independent of B_0.

6.3.1 Example. If $B_0 = x \in \mathbb{R}^d$ a.s. then a Brownian motion started at x is just $(x + B_t, t \geq 0)$, where $(B_t, t \geq 0)$ is a standard Brownian motion. Its distribution is the measure $W_x(dw)$ on Ω_W (Weiner space) which is the image measure of $W_0(dw)$ (Weiner measure) by the map

$$\Omega_W \to \Omega_W : w \mapsto (x + w_t, t \geq 0) = x + w.$$

The law of a Brownian motion started at a r.v. B_0 is determined by

$$\begin{aligned}\mathbb{E}[F(B)] &= \mathbb{E}[F((B_t - B_0, t \geq 0) + B_0)] \\ &= \int P(B_0 \in dx) \int W_0(dw) F(w + x) \\ &= \int P(B_0 \in dx) W_x(F) \\ &=: \mathbb{E}[W_{B_0}(F)].\end{aligned}$$

Equivalently, the law of B_t given B_0 is W_{B_0}.

6.3.2 Definition. Let $(\mathscr{F}_t, t \geq 0)$ be a filtration and $(B_t, t \geq 0)$ be a Brownian motion. We say that $(\mathscr{F}_t)$ is a *Brownian filtration* (or (B_t) is an $(\mathscr{F}_t)$-Brownian motion) if $\mathscr{F}_t^B \subseteq \mathscr{F}_t$ for all t (so in particular B is adapted) and for all t, $B^{(t)}$ is a standard Brownian motion independent of $\mathscr{F}_t$ (this is the *simple Markov property*).

6.3.3 Theorem (Strong Markov Property).
Let (B_t) be an $(\mathscr{F}_t)$-Brownian motion and let T be an $(\mathscr{F}_t)$-stopping time. Then, conditionally on $\{T < \infty\}$, the process

$$B^{(T)} = \begin{cases} (B_{T+t} - B_T, t \geq 0) & \text{on } \{T <, \infty\} \\ 0 & \text{otherwise} \end{cases}$$

is a standard Brownian motion which is independent of $\mathscr{F}_T$.

Equivalently, given $\mathscr{F}_T$, the process $(B_{t+T}, t \geq 0)$ is an $(\mathscr{F}_{T+t}, t \geq 0)$-Brownian motion.

PROOF: Let T be such that $\mathbb{P}(T < \infty) = 1$. Let $A \in \mathscr{F}_T$ and $0 \leq t_1 \leq \cdots \leq t_k$ be fixed times. Let $F : (\mathbb{R}^d)^k \to \mathbb{R}$ be a bounded continuous function. Assume first that T takes values in D, the dyadic rationals. Then

$$\begin{aligned}\mathbb{E}[\mathbf{1}_A F(B_{t_1}^{(T)}, \ldots, B_{t_k}^{(T)})] &= \sum_{d \in D} \mathbb{E}[\mathbf{1}_{A \cap \{T=d\}} F(B_{t_1}^{(T)}, \ldots, B_{t_k}^{(T)})] \\ &= \sum_{d \in D} \mathbb{P}(A, T = d) \mathbb{E}[F(B_{t_1}, \ldots, B_{t_k})] \qquad \text{SM} \\ &= \mathbb{P}(A) \mathbb{E}[F(B_{t_1}, \ldots, B_{t_k})]\end{aligned}$$

In general, let $T_n = 2^{-n}\lceil 2^n T \rceil$, so that $T_n \searrow_{n\to\infty} T$. $\mathscr{F}_T \subseteq \mathscr{F}_{T_n}$ since $T \leq T_n$, so $A \in \mathscr{F}_{T_n}$, and by the previous argument,

$$\mathbb{E}[\mathbf{1}_A F(B_{t_1}^{(T_n)}, \ldots, B_{t_k}^{(T_n)})] = \mathbb{P}(A) \mathbb{E}[F(B_{t_1}, \ldots, B_{t_k})].$$

Since F is continuous and bounded and Brownian motion is continuous, by the DCT the left hand side goes to $\mathbb{E}[\mathbf{1}_A F(B^{(T)}_{t_1}, \dots, B^{(T)}_{t_k})]$ as $n \to \infty$. For $A = \Omega$, we obtain that $B^{(T)}$ is a standard Brownian motion. By letting A vary in $\mathscr{F}_T$, we obtain that $\mathscr{F}_T$ is independent of $\sigma(B^{(T)}_{t_1}, \dots, B^{(T)}_{t_k})$, and is therefore independent of $B^{(T)}$. Finally, if $\mathbb{P}(T < \infty) < 1$ then use the same argument with $A \cap \{T < \infty\}$ instead of A, and divide both sides of the above equation by $\mathbb{P}(T < \infty)$. □

6.3.4 Theorem (Reflection Principle).
Let (B_t) be a standard $(\mathscr{F}_t)$-Brownian motion and let T be an $(\mathscr{F}_t)$-stopping time. Let

$$\tilde{B}_t = B_t \mathbf{1}_{t \le T} + (2B_T - B_t)\mathbf{1}_{t > T}$$

for $t \ge 0$. Then $\tilde{B}$ is a standard $(\mathscr{F}_t)$-Brownian motion as well.

PROOF: We know by the Strong Markov Property that $(B_t, 0 \le t \le T)$ and $B^{(T)}$ are independent. We know also by invariance properties of Brownian motion, $-B^{(T)} \overset{(d)}{=} B^{(T)}$. Therefore $((B_t, 0 \le t \le T), B^{(T)})$ has the same joint distribution as $((B_t, 0 \le t \le T), -B^{(T)})$. Hence for all $F \ge 0$ measurable, $\mathbb{E}[F(B)] = \mathbb{E}[F(\tilde{B})]$. □

6.3.5 Corollary. *Let $d = 1$ and $S_t = \sup_{0 \le s \le t} B_s$ for $t \ge 0$. For every $b > 0$ and for all $a \le b$, $\mathbb{P}(S_t \ge b, B_t \le a) = \mathbb{P}(B_t \ge 2b - a)$.*

PROOF: Introduce, for $x > 0$, $T_x = \inf\{t \ge 0 \mid B_t \ge x\}$, a stopping time since Brownian motion is continuous and $[x, \infty)$ is closed. The event $\{S_t \ge b, B_t \le a\} = \{T_b \le t, 2b - B_t \ge 2b - a\} = \{\tilde{B}_t \ge 2b - a\}$ since $\tilde{B}_t \ge 2b - a \ge b$ implies $T_b \le t$. The result follows by the reflection principle for $T = T_b$. □

6.3.6 Corollary. *For all $t \ge 0$, $S_t \overset{(d)}{=} |B_t|$.*

PROOF: For any $b \ge 0$,

$$\begin{aligned}
\mathbb{P}(S_t \ge b) &= \mathbb{P}(S_t \ge b, B_t \le b) + \mathbb{P}(S_t \ge b, B_t \ge b) \\
&= \mathbb{P}(B_t \ge b) + \mathbb{P}(B_t \ge b) \\
&= 2\,\mathbb{P}(B_t \ge b) = \mathbb{P}(|B_t| \ge b).
\end{aligned}$$

by the symmetry of the Gaussian law. □

Warning: It is not true that S and B have the same distribution as processes, since (S_t) is monotone while $(|B_t|)$ is not.

6.3.7 Corollary. *For all $x \ge 0$, the stopping time $T_x \overset{(d)}{=} (\frac{x}{B_1})^2$.*

PROOF: Indeed,

$$\begin{aligned}
\mathbb{P}(T_x \le t) &= \mathbb{P}(S_t \ge x) \\
&= \mathbb{P}(|B_t| \ge x) \\
&= \mathbb{P}(\sqrt{t}|B_1| \ge x) \\
&= \mathbb{P}((\tfrac{x}{B_1})^2 \le t)
\end{aligned}$$

implying that they have the same probability distribution function. □

6.4 Martingales associated with Brownian motion

6.4.1 Proposition. *Let (B_t) be an $(\mathscr{F}_t)$-Brownian motion.*

(i) In $d=1$, $(B_t, t \geq 0)$ is an $(\mathscr{F}_t)$-martingale when $B_0 \in L^1$.

(ii) In $d=1$, $(B_t^2 - t, t \geq 0)$ is an $(\mathscr{F}_t)$-martingale when $B_0 \in L^2$.

(iii) For $\xi \in \mathbb{R}^d$, $(\exp(i\langle \xi, B_t\rangle + \frac{1}{2}t|\xi|^2), t \geq 0)$ is an $(\mathscr{F}_t)$-martingale.

PROOF:

(i) $\mathbb{E}[B_t - B_s \mid \mathscr{F}_s] = 0$ for all $s < t$.

(ii) $B_t^2 = (B_t - B_s + B_s)^2 = (B_t - B_s)^2 + 2B_s(B_t - B_s) + B_s^2$, so $\mathbb{E}[B_t^2 \mid \mathscr{F}_s] = \mathbb{E}[(B_t - B_s)^2 \mid \mathscr{F}_s] + B_s^2 = t - s + B_s^2$.

(iii) $\mathbb{E}[\exp(i\langle \xi, B_t - B_s\rangle + i\langle \xi, B_s\rangle) \mid \mathscr{F}_s] = \exp(i\langle \xi, B_s\rangle)\exp(-\frac{1}{2}(t-s)|\xi|^2)$. □

6.4.2 Proposition. *Let (B_t) be a standard Brownian motion and $d=1$, and for $x \in \mathbb{R}$, $T_x = \inf\{t \geq 0 \mid B_t = x\}$. Then for $x, y > 0$, $\mathbb{P}(T_x < T_{-y}) = \frac{y}{x+y}$ and $\mathbb{E}[T_x \wedge T_{-y}] = xy$.*

PROOF: We use martingales stopped at $T_x \wedge T_{-y}$.

$$0 = \mathbb{E}[B_{T_x \wedge T_{-y}}] = x\,\mathbb{P}(T_x < T_{-y}) - y(1 - \mathbb{P}(T_x < T_{-y})),$$

so $\mathbb{P}(T_x < T_{-y}) = \frac{y}{x+y}$. Similarly,

$$\mathbb{E}[T_x \wedge T_{-y}] = \mathbb{E}[B^2_{T_x \wedge T_{-y}}] = x^2 \frac{y}{x+y} + y^2 \frac{x}{x+y} = xy.$$

□

6.4.3 Theorem. *Let $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{C}$ be continuously differentiable in the first coordinate (time) and twice continuously differentiable in the second coordinate (space), and be such that all partial derivatives are bounded. If (B_t) is an $(\mathscr{F}_t)$-Brownian motion then*

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

is an $(\mathscr{F}_t)$-martingale, where $\Delta f(t,x) = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(t,x)$ is the Laplaceian.

PROOF: We prove the case where $f(t,x) = f(x)$ does not depend on time. Let $p_t(x) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp(-\frac{1}{2t}|x|^2)$ and let g be bounded and continuous. We claim that if (B_t) is a standard Brownian motion then

$$\int_0^t \int_{\mathbb{R}^d} g(x) \frac{\partial}{\partial s} p_s(x) dx ds = \mathbb{E}[g(B_t)] - g(0).$$

Indeed,

$$\begin{aligned}\int_0^t \int_{\mathbb{R}^d} g(x)\frac{\partial}{\partial s}p_s(x)dxds &= \lim_{\varepsilon\to 0}\int_\varepsilon^t \int_{\mathbb{R}^d} g(x)\frac{\partial}{\partial s}p_s(x)dxds \\ &= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^d} g(x)(p_t(x)-p_\varepsilon(x))dx \\ &= \lim_{\varepsilon\to 0}\mathbb{E}[g(B_t)]-\mathbb{E}[g(B_\varepsilon)] \\ &= \mathbb{E}[g(B_t)]-g(0)\end{aligned}$$

by DCT since $B_\varepsilon \to 0$ and g is bounded.

For $s,t\geq 0$,

$$\begin{aligned}&\mathbb{E}\left[f(B_{s+t})-f(B_t)-\int_t^{t+s}\frac{\Delta}{2}f(B_u)du \mid \mathscr{F}_t\right] \\ &\quad = \underbrace{\int p_s(x)f(x+B_t)dx - f(B_t)}_{(1)} - \underbrace{\mathbb{E}\left[\int_t^{t+s}\frac{\Delta}{2}f(B_u-B_t+B_t)du \mid \mathscr{F}_t\right]}_{(2)}.\end{aligned}$$

But

$$\begin{aligned}(2) &= \mathbb{E}\left[\int_0^s \frac{\Delta}{2}f(B_{t+u}-B_t+B_t)du \mid \mathscr{F}_t\right] \\ &= \int_{\Omega_W} W_0(dw)\int_0^s \frac{\Delta}{2}f(w_u+B_t)du \\ &= \int_0^s du\int_{\Omega_W} W_0(dw)\frac{\Delta}{2}f(w_u+B_t) \\ &= \int_0^s du\int p_u(x)dx\frac{\Delta}{2}f(x+B_t)\end{aligned}$$

By integrating by parts, this is equal to

$$(2) = \int_0^s du\int \frac{\Delta}{2}p_u(x)f(x+B_t)dx.$$

It can be checked that $(\frac{\partial}{\partial t}-\frac{\Delta}{2})p_t(x)=0$, i.e. the density of the Gaussian law is a solution to the heat equation. Therefore

$$(2) = \int_0^s du\int \partial_u p_u(x)f(x+B_t)dx = \int_{\mathbb{R}^d} p_s(x)f(x+B_t)dx - f(B_t) = (1)$$

by the claim. □

6.5 Recurrence and transience properties

Assume given $(\Omega,\mathscr{F},(P_x)_{x\in\mathbb{R}^d})$ probability spaces such that (B_t) is under P_x a Brownian motion started at $x\in\mathbb{R}^d$. (For example, $(\Omega_W, prod, (W_x)_{x\in\mathbb{R}^d})$).

6.5.1 Theorem.

(i) *In $d = 1$, a.s. under P_0, the sets $\{t \mid B_t = x\}$ are unbounded, i.e. Brownian motion in dimension one is "point recurrent."*

(ii) *In $d = 2$, a.s. under P_x, $\{t \mid |B_t| \le \varepsilon\}$ is unbounded for all $\varepsilon > 0$, i.e. Brownian motion in dimension two is "neighbourhood recurrent." However, if $H_x = \inf\{t > 0 \mid B_t = x\}$ then $P_x(H_y < \infty) = 0$, i.e. "points are polar."*

(iii) *In $d \ge 3$, (B_t) is "transient," i.e. $|B_t| \to \infty$ as $t \to \infty$.*

PROOF:

(i) We already know that $\sup_{t\ge 0} B_t = -\inf_{t\ge 0} B_t = \infty$, so we have point recurrence by continuity.

(ii) Let $\varepsilon > 0$, $R > \varepsilon$, and $D_{\varepsilon,R} = \{x \in \mathbb{R}^2 \mid \varepsilon \le |x| \le R\}$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be such that $f : x \mapsto \log|x|$ on $D_{\varepsilon,R}$, f is bounded, is C^∞, and has bounded derivatives. Check that $\Delta \log|x| = 0$ on the interior of $D_{\varepsilon,R}$. Let $T_\varepsilon = \inf\{t \ge 0 \mid |B_t| \le \varepsilon\}$ and $T_R = \inf\{t \ge 0 \mid |B_t| \ge R\}$. By the previous theorem,

$$f(B_t) - f(x) - \int_0^t \frac{\Delta}{2} f(B_s)ds$$

is a martingale. Therefore if $T = T_\varepsilon \wedge T_R$, then

$$f(B_{t\wedge T}) - f(x) - \int_0^{t\wedge T} \frac{\Delta}{2} f(B_s)ds = f(B_{t\wedge T}) - f(x)$$

(since $\Delta f = 0$ on $D_{\varepsilon,R}$) is a martingale, so $(f(B_{t\wedge T}), t \ge 0)$ is a bounded martingale. By optional stopping,

$$\log|x| = \mathbb{E}^{P_x}[f(B_T)] = (\log\varepsilon)P(T_\varepsilon < T_R) + (\log R)P_x(T_R < T_\varepsilon),$$

so

$$P_x(T_\varepsilon < T_R) = \frac{\log|x| - \log R}{\log\varepsilon - \log R}. \qquad (1)$$

Letting $R \to \infty$ in (1), we see that $P_x(T_\varepsilon < \infty) = 1$. But

$$\begin{aligned}
P_x(\exists t \ge n : |B_t| \le \varepsilon) &= P_x(\exists t \ge n : |B_t - B_n + B_n| \le \varepsilon) \\
&= P_x(\exists t \ge n : |B_t^{(n)} + B_n| \le \varepsilon) \\
&= \int P_x(B_n \in dy) P_y(\exists t \ge 0 : |B_t| \le \varepsilon) \qquad \text{SM} \\
&= P_y(T_\varepsilon < \infty) = 1
\end{aligned}$$

so $\{t \mid |B_t| \le \varepsilon\}$ is unbounded. On the other hand, letting $\varepsilon \to 0$ in (1) shows that $P_x(T_0 \le T_R) = 0$, and letting $R \to \infty$ gives $P_x(T_0 < \infty) = 0$ for all $x \ne 0$. For $x = 0$,

$$\begin{aligned}
P_0(\exists t \ge a : B_t = 0) &\xrightarrow{a\to 0} P_0(\exists t > 0 : B_t = 0) \\
&= P_0(\exists t > 0 : B_t^{(a)} + B_a = 0) \\
&= \int_{\mathbb{R}^2} P_0(B_a \in dx) P_x(\exists t \ge 0 : B_t = 0) \qquad \text{SM} \\
&= 0
\end{aligned}$$

because the law of B_a under P_0 is a Gaussian law that does not charge $\{0\}$. Therefore for each $x, y \in \mathbb{R}^2$, $P_x(H_y < \infty) = 0$.

(iii) Since the first three components of Brownian motion in $\mathbb{R}^d$ are a Brownian motion in $\mathbb{R}^3$, it suffices to prove the result for $d = 3$. Let f be a bounded C^∞ functions with all derivatives bounded such that $f(x) = \frac{1}{|x|}$ on $D_{\varepsilon,R}$. Again, $\Delta f = 0$ on $D_{\varepsilon,R}$, so $(f(B_{T\wedge t}), t \geq 0)$ is a bounded martingale. Applying Optional Stopping, we have

$$P_x(T_\varepsilon < T_R) = \frac{\frac{1}{R} - \frac{1}{|x|}}{\frac{1}{R} - \frac{1}{\varepsilon}} \tag{2}$$

for $x \in D_{\varepsilon,R}$. Letting $R \to \infty$ in (2) shows that $P_x(T_\varepsilon < \infty) = \frac{\varepsilon}{|x|}$. Fix $r > 0$ and define $S_1 = \inf\{t \geq 0 \mid |B_t| \leq r\}$, and for all $k \geq 1$ define $T_k = \inf\{t \geq S_k \mid |B_t| \geq 2r\}$ and $S_{k+1} = \inf\{t \geq T_k \mid |B_t| \leq r\}$. Now $\{S_k < \infty\} = \{T_k < \infty\}$ because $S_k < T_k$ (giving one inclusion) and Brownian motion is unbounded (giving the other inclusion).

$$\begin{aligned} P_x(S_{k+1} < \infty \mid S_k < \infty) &= P_x(S_{k+1} < \infty \mid T_k < \infty) \\ &= P_x(\exists t \geq T_k : |B_t| \leq r \mid T_k < \infty) \\ &= P_x(\exists t \geq T_k : |B_t^{(T_k)} + B_{T_k}| \leq r \mid T_k < \infty) \\ &= \int_{\mathbb{R}^3} P_x(B_{T_k} \in dy) P_y(\exists t \geq 0 : |B_t| \leq r) \end{aligned}$$

by the Strong Markov property. But $|B_{T_k}| = 2r$, so a.s. under $P_x(B_{T_k} \in dy)$, $|y| = 2r$, whence $P_y(\exists t : |B_t| \leq r) = P_y(T_r < \infty) = \frac{1}{2}$. Therefore $P_x(S_{k+1} < \infty \mid S_k < \infty) = \frac{1}{2}$. It follows that $P_x(S_k < \infty) = \frac{1}{2^{k-1}} P_x(S_1 < \infty)$ given that $S_1 < \infty$. By the Borel-Cantelli Lemma, $\sup\{k \mid S_k < \infty\} < \infty$ a.s. for all r. Letting $r \to \infty$ along $\mathbb{Z}_+$, this shows that eventually $|B_t| \geq r$ for any r, so $|B_t| \to \infty$ a.s. □

Remark. For 2-dimensional Brownian motion, since $\{t \mid B_t \in B(x, \varepsilon)\}$ is unbounded for every $x \in \mathbb{R}^2$ and $\varepsilon > 0$ and $\mathbb{R}^2$ may be covered by a countable union of balls of a fixed radius, the trajectory of Brownian motion is dense in $\mathbb{R}^2$. On the other hand, a.s. for all $t > 0$, $B_t \notin \mathbb{Q}^2$.

6.6 Dirichlet problem

Let $D \subseteq \mathbb{R}^d$ $(d \geq 2)$ be a connected open set (i.e. a *domain*), and let g be a measureable function on ∂D. $f \in C^2(D) \cap C(\overline{D})$ is said to satisfy the *Dirichlet problem* with boundry condition g if $\Delta f = 0$ on D and $f|_{\partial D} = g$. We will see that in many cases $f(x) := \mathbb{E}^{P_x}[g(B_T)]$ is the unique solution, where T is the first exit time of B from D. Recall that $(\frac{\partial}{\partial t} - \frac{\Delta}{2})p_t(x) = 0$. We restrict our attention to the case where $g \in C(\partial D, \mathbb{R})$ and D is such that $P_x(T < \infty) = 1$ for all $x \in D$, where $T = \inf\{t \geq 0 \mid B_t \notin D\}$.

6.6.1 Proposition. *Let v be a bounded solution of the Dirichlet problem with boundry condition g. Then necessarily $v(x) = \mathbb{E}_x[g(B_T)]$ for all $x \in \overline{D}$.*

PROOF: Let $D_N = B(0,N) \cap \{x \in D \mid d(x,\partial D) > \frac{1}{N}\}$, and let $\tilde{v}_N$ be a bounded C^2 function which agrees with v on D_N and has bounded partials (it is non-trivial that such a $\tilde{v}_N$ exists). Let $T_N = \inf\{t \geq 0 \mid B_t \notin D_N\}$. Since $\Delta\tilde{v}_N = 0$ on D_N,

$$\tilde{v}_N(B_{t\wedge T_N}) - \tilde{v}_N(x) = \tilde{v}_N(B_{t\wedge T_N}) - \tilde{v}_N(x) - \int_0^{t\wedge T} \Delta\tilde{v}_N(B_s)ds$$

is a martingale. Since it is bounded, Optional Stopping implies that

$$v(x) - \mathbb{E}_x[\tilde{v}_N(B_{T_N})] = \mathbb{E}_x[v(B_{T_N})].$$

Letting $N \to \infty$, $B_{T_N} \to B_T$ (this uses the fact that $T < \infty$ a.s.) Since v is bounded on $\overline{D}$, DCT implies that $v(x) = \mathbb{E}_x[v(B_T)] = \mathbb{E}_x[g(B_T)]$. □

6.6.2 Definition. A locally bounded measurable function $h : D \to \mathbb{R}$ is *harmonic* if for all $x \in D$ and $r > 0$ such that $\overline{B(x,r)} \subseteq D$,

$$h(x) = \int_{S_{x,r}} h(y)d\sigma_{x,r}(y),$$

where $S_{x,r}$ is the sphere centered at x of radius r and $\sigma_{x,r}$ is the uniform distribution on $S_{x,r}$.

Remark. $\sigma_{x,r}$ is the unique probability measure on $S_{x,r}$ which is invariant under isometries fixing x.

6.6.3 Proposition. *A harmonic function h is $C^\infty(D)$ and satisfies $\Delta h = 0$.*

PROOF: See lecture notes. □

6.6.4 Proposition. *For $g \in C_b(D)$, the function $u(x) = \mathbb{E}_x[g(B_T)]$ is harmonic on D.*

PROOF: See the lecture notes for a proof that u is bounded and measurable. Fix $x \in D$ and $r > 0$ such that $\overline{B(x,r)} \subseteq D$. Define $S = \inf\{t \geq 0 \mid |B_t - x| \geq r\}$, a stopping time such that $S \leq T < \infty$ a.s. under P_x. Then

$$\begin{aligned} u(x) &= \mathbb{E}_x[g(B_T - B_S + B_S)] \\ &= \mathbb{E}_x[g(B^{(S)}_{T-S} + B_S)] \qquad\qquad T - S = \inf\{t \geq 0 \mid B^{(S)}_t + B_S \notin D\} \\ &= \int P_x(B_S \in dy)\mathbb{E}_y[g(B_T)]. \end{aligned}$$

Now $B_S \in S_{x,r}$ a.s. and by the invariance of standard Brownian motion under the action of $O(d)$, the law of B_S under P_x is invariant under isometries preserving x. Therefore $P_x(B_S \in dy) = \sigma_{x,r}(dy)$, so $u(x) = \int_{S_{x,r}} \sigma_{x,r}(dy)u(y)$ and u is harmonic. □

It is not true in general that $u(x) \to g(y)$ as $x \to y$ in D. For example, if $D = B(0,1) \setminus \{0\}$ and $g = \mathbf{1}_{\{0\}}$ then $u(x) = 0$ for all $x \in D$. The problem can also arise with very complicated boundries.

6.6.5 Theorem. *Let D be a domain with smooth boundry (i.e. for all $y \in \partial D$ there is a neighbourhood V of y such that $\partial D \cap V$ is a smooth surface) and assume that $P_x(T < \infty) = 1$ for all $x \in D$. Then for all $g \in C_b(\partial D)$, $u(x) = \mathbb{E}_x[g(B_T)]$ is the unique bounded solution to the Dirichlet problem.*

6.6.6 Lemma. *Let D be a domain with smooth boundry. Then for all $y \in D$, $P_x(T > \eta) \to 0$ as $x \to y$ in D for all $\eta > 0$.*

PROOF: Add me! □

6.6.7 Corollary. *If D is a bounded domain and $g \in C(\partial D)$ then the unique solution to the Dirichlet problem is $u(x) = \mathbb{E}_x[g(B_T)]$.*

6.7 Donsker's invariance principle

Let $X_1, X_2, \dots$ be i.i.d. real-valued r.v.'s such that $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1^2] = 1$. Consider $S_n = X_1 + \cdots + X_n$, and define a continuous process from this but interpolating linearly between values (so $S_t = (1 - \{t\})S_{\lfloor t \rfloor} + \{t\}S_{\lfloor t \rfloor + 1}$ for all $t \in \mathbb{R}$).

6.7.1 Theorem. *Let $S_t^{\lfloor N \rfloor} = \frac{S_{Nt}}{\sqrt{N}}$ for all $t \in [0,1]$. Then $(S_t^{\lfloor N \rfloor}, 0 \leq t \leq 1) \to (B_t, 0 \leq t \leq 1)$ in distribution, where B is standard Brownian motion.*

By $S^{\lfloor N \rfloor} \to B$ in distribution we mean that it converges in distribution in the space $(C([0,1], \mathbb{R}), \|\cdot\|_\infty)$ with the Borel σ-algebra. Using this theorem one can prove, for example, that $\frac{1}{\sqrt{N}} \sup_{0 \leq n \leq N} S_n \to \sup_{0 \leq t \leq 1} B_t$ in distribution at $N \to \infty$.

7 Poisson Random Measures

7.1 Motivation

The Poisson distribution $\mathscr{P}(\lambda)$ ($\lambda \geq 0$) is such that if N has distribution $\mathscr{P}(\lambda)$ then $\mathbb{P}(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}$ for $n \geq 0$. It is a "law of rare events," in that $\text{Bin}(n, \frac{\lambda}{n}) \to \mathscr{P}(\lambda)$ weakly as $n \to \infty$.

Let $(E, \mathscr{E})$ be a measurable space. Define E^* to be the set of σ-finite measures on $(E, \mathscr{E})$ of the form $m = \sum_{i \in I} \delta_{x_i}$ for some countable set I and $x_i \in E$ for all $i \in I$. Let $\mathscr{E}^*$ be the smallest σ-algebra for which X_A is measurable for all $A \in \mathscr{E}$, where $X_A : E^* \to \mathbb{R}_+ \cup \{\infty\} : m \mapsto m(A)$.

7.1.1 Definition. Let μ be a σ-finite on $(E, \mathscr{E})$. An E^*-valued r.v. M on $\mathscr{E}$ is a *Poisson random measure* with *intensity* μ if for all $A_1, \dots, A_k \in \mathscr{E}$ pairwise disjoint,

(i) $M(A_1), \dots, M(A_k)$ are independent; and

(ii) for all $j \in \{1, \dots, k\}$, $M(A_j) \sim \mathscr{P}(\mu(A_j))$ if $\mu(A_j) < \infty$.

7.1.2 Lemma. *If M is a Poisson random measure with intensity μ (PRM(μ)) then its law is uniquely determined.*

PROOF: Let $A_1, \dots, A_k \in \mathscr{E}$ be pairwise disjoint and of finite μ-mass, and $i_1 \geq 0, \dots, i_k \geq 0$. Consider $\{m \in E^* \mid m(A_1) = i_1, \dots, m(A_k) = i_k\} = \{X_{A_1} = i_1, \dots, X_{A_k} =$

$i_k\}$. Such events are stable under intersection and generate the σ-algebra $\mathcal{E}^*$. Now if M is $PRM(\mu)$ then

$$\begin{aligned}\mathbb{P}(M(A_1)=i_1,\dots,M(A_k)=i_k) &= \prod_{j=1}^{k} P(M(A_j)=i_j) \\ &= \prod_{j=1}^{k} e^{-\mu(A_j)}\frac{(\mu(A_j))^{i_j}}{i_j!}.\end{aligned}$$

Thus the probabilities $\mathbb{P}(B)$ are determined by the definition of $PRM(\mu)$ for all B in a π-system that generates $\mathcal{E}^*$, so the definition determines the law of M uniquely. □

7.1.3 Lemma.

(i) If $N_1, N_2 \sim \mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)$ then $N_1+N_2 \sim \mathcal{P}(\lambda_1+\lambda_2)$.

(ii) Let $N \sim \mathcal{P}(\lambda)$ and let $Y_1, Y_2, \dots$ be i.i.d. and taking values in $\{1,\dots,k\}$, independent of N, and such that $\mathbb{P}(Y_i=j)=p_j$ for some $p_1,\dots,p_k$ which sum to 1. If $N_i=\sum_{n=1}^{N}\mathbf{1}_{\{Y_n=i\}}$. Then $N_1,\dots,N_k$ are independent and $N_i \sim \mathcal{P}(p_i\lambda)$ for $1 \le i \le k$.

PROOF: Exercise. □

7.1.4 Proposition. *There exists a $PRM(\mu)$.*

PROOF: First assume that $\mu(E)<\infty$. Let N be a $\mathcal{P}(\mu(E))$ r.v. Independently of N, let $X_1, X_2, \dots$ be i.i.d. with law $\frac{\mu(\cdot)}{\mu(E)}$. Set $M=\sum_{i=1}^{N}\delta_{X_i}$. Then M is a $PRM(\mu)$. Indeed, let $A_1,\dots,A_k$ be pairwise disjoint, then $M(A_j)=\sum_{i=1}^{N}\mathbf{1}_{X_i\in A_j}$. Setting $Y_i=j$ if and only if $X_i \in A_j$, the Y_i's are independent and $\mathbb{P}(Y_i=j)=\mathbb{P}(X_i\in A_j)=\frac{\mu(A_j)}{\mu(E)}$. By the lemma, the $M(A_j)$ are independent and follow a $\mathcal{P}(\mu(E)\frac{\mu(A_j)}{\mu(E)})=\mathcal{P}(\mu(A_j))$, as we wanted.

If $\mu(E)=\infty$ then we can find E_n $(n\ge1)$ that partition E and $\mu(E_n)<\infty$ for all n. Let M_n, $n\ge1$ be independent Poisson random measures with intensities $\mu(\cdot\cap E_n)$, the restriction of μ to E_n. Then $M=\sum_{n\ge1}M_n$ is $PRM(\mu)$. Indeed, if $A_1,\dots,A_k$ are disjoint and of finite μ-mass then $M(A_j)=\sum_{n\ge1}M_n(A_j)=\sum_{n\ge1}M_n(A_j\cap E_n)$. By definition, $(M_n(A_j\cap E_n), 1\le j\le k)$ are independent and these vectors are independent over n. Thus $M(A_1),\dots,M(A_k)$ are independent and $M(A_j)\sim\mathcal{P}(\sum_{n\ge1}\mu(A_j\cap E_n))=\mathcal{P}(\mu(A_j))$. □

7.1.5 Corollary. *If M is a $PRM(\mu)$ and if $A_1,\dots,A_k$ are pairwise disjoint with finite μ-mass then $M|_{A_1},\dots,M|_{A_k}$ are independent, and moreover, $M|_{A_j}$, given $M(A_j)=n$, has the same distribution as $\sum_{i=1}^{n}\delta_{X_i}$, where $X_1,\dots,X_n$ are i.i.d. with respect to the measure $\frac{\mu(\cdot\cap A_j)}{\mu(A_j)}$ $(=: \mu(\cdot\mid A_j))$.*

Note that $\mathbb{E}[M(A)]=\mu(A)$. Indeed, $\mathbb{E}[\mathcal{P}(\lambda)]=\lambda$.

7.2 Integrating with respect to a Poisson measure

7.2.1 Proposition. *For $f : E \to \mathbb{R}_+$ measurable, $M(f) = \int_E f(x)M(dx)$ defines a positive r.v. with $\mathbb{E}[M(f)] = \mu(f)$, the first moment formula. Moreover,*

$$\mathbb{E}[e^{-M(f)}] = \exp(\mu(e^{-f} - 1)),$$

the Laplace functional formula.

PROOF: If $f = \mathbf{1}_A$ for some $A \in \mathscr{E}$ with $\mu(A) < \infty$ then $M(f) = M(A)$ is a r.v. By approximating f by simple functions (i.e. functions of form $\sum_{i \in I} \alpha_i \mathbf{1}_{A_i}$), the usual argument shows that $M(f)$ is a r.v. If the Laplace functional formula holds then replacing f by λf $(\lambda \geq 0)$ and differentiaing with respect λ at $\lambda = 0$, we directly obtain $\mathbb{E}[M(f)] = \int_E \mu(dx)f(x) = \mu(f)$.

Now assume that $\mu(E) < \infty$ and $M = \sum_{i=1}^N \delta_{X_i}$, where $N \sim \mathscr{P}(\mu(E))$ and $(X_i, i \geq 1)$ are i.i.d. and independent of N, with $\mathscr{L}(X_i) = \frac{\mu}{\mu(E)}$. Then

$$\begin{aligned}\mathbb{E}[e^{-M(f)}] &= \mathbb{E}\left[e^{-\sum_{i=1}^N f(X_i)}\right] \\ &= \sum_{n \geq 0} e^{-\mu(E)} \frac{\mu(E)^n}{n!} \mathbb{E}\left[e^{-\sum_{i=1}^n f(X_i)}\right] \\ &= \sum_{n \geq 0} e^{-\mu(E)} \frac{\mu(E)^n}{n!} \left(\int_E \frac{\mu(dx)}{\mu(E)} e^{-f(x)}\right)^n\end{aligned}$$

hence, since $\mu(E) = \int_E \mu(dx)$,

$$\mathbb{E}[e^{-M(f)}] = e^{-\mu(E)} e^{\int_E \mu(dx) e^{f(x)}} = \exp(\mu(e^{-f} - 1)).$$

If $\mu(E) = \infty$ then find E_n that partition E with $\mu(E_n) < \infty$ for all n. The measures $M|_{E_n}$ are independent, and $M(f\mathbf{1}_{E_n})$ are independent, so

$$\mathbb{E}[e^{-M(f\mathbf{1}_{E_n})}] = \exp\left(\int_{E_n} (e^{-f(x)} - 1)\mu(dx)\right).$$

Thus

$$\begin{aligned}\mathbb{E}[e^{-M(f)}] &= \prod_{n \geq 0} \mathbb{E}[e^{-M(f\mathbf{1}_{E_n})}] \\ &= \exp\left(\sum_{n \geq 0} \int_{E_n} (e^{-f(x)} - 1)\mu(dx)\right) \\ &= \exp(\mu(e^{-f} - 1)).\end{aligned}$$

□

7.2.2 Corollary. *If $f \in L^1(\mu)$ then $f \in L^1(M)$ a.s.,*

$$\mathbb{E}[e^{iM(f)}] = \exp\left(\int \mu(dx)(e^{if(x)} - 1)\right),$$

and $\mathbb{E}[M(f)] = \mu(f)$.

PROOF: Apply the previous proof to $|f|$ to get $\mathbb{E}[M(|f|)] = \mu(|f|) < \infty$. Therefore $M(f) < \infty$ a.s., so $f \in L^1(M)$. Then redo the previous proof and note that $|e^{if(x)} - 1| \le |f(x)| \in L^1(\mu)$. (Apply DCT?) □

Note that if moreover we have $f \in L^2(\mu)$ then $M(f)$ has finite variance

$$\mathrm{Var}(M(f)) = \int f(x)^2 d\mu(x) = \mu(f^2),$$

the *second moment formula*.

7.2.3 Proposition.

(i) *Let M be a Poisson random measure on $(E,\mathscr{E})$, let $(F,\mathscr{F})$ be a measurable space, let $f : E \to F$ be measurable, and let f_* denote the push-forward of f. Then f_*M (i.e. $\sum \delta_{f(X_i)}$ if $M = \sum \delta_{X_i}$) is a PRM$(f_*\mu)$ on F.*

(ii) *(Marking property) Let M be a PRM(μ) written in the form $M = \sum \delta_{X_i}$ and let $(Y_i, i \ge 1)$ be i.i.d. and independent of M, with $\mathscr{L}(Y_i) = \nu$ on some space $(F,\mathscr{F})$. Then $M^* = \sum_{i\ge1} \delta_{(X_i,Y_i)}$ is a PRM$(\mu \otimes \nu)$ on $E \times F$.*

PROOF: Exercise (apparently it is obvious given the definitions). □

7.3 Poisson point processes

7.3.1 Definition. Let $(E,\mathscr{E},G)$ be a σ-finite measure space. A *Poisson point process* with intensity G (or $PPP(G)$) is a Poisson random measure on $\mathbb{R}_+ \times E$ with intensity $dt \otimes G(dx)$, where dt is Lebesgue measure on $\mathbb{R}_+$.

Such a process M can be written $M(dt\,dx) = \sum_{i\in I} \delta_{(t_i,x_i)}$, where I is a countable index set. Since dt is *diffuse* (i.e. it has no atoms), with probability 1, for all t, there is at most one $i \in I$ such that $t_i = t$. Therefore we can define a process $(e_t, t \ge 0)$ with values in $E \amalg \{*\}$ by

$$e_t = \begin{cases} x_i & \text{if } t = t_i \text{ for some } i \in I \\ * & \text{otherwise} \end{cases}$$

Then $(e_t, t \ge 0)$ is also called a Poisson point process with intensity G.

7.3.2 Definition. Let $(X_t, t \ge 0)$ be a stochastic process with values in $\mathbb{R}$. We say that (X_t) is a *Lévy process* if

(i) For all $s < t$, $X_t - X_s \overset{(d)}{=} X_{t-s}$ (increments are stationary);

(ii) For all $0 = t_0 < t_1 < \cdots < t_k$, $(X_{t_i} - X_{t_{i-1}}, 1 \le i \le k)$ are independent.

7.3.3 Proposition. *Let $M(dt\,dx)$ (resp. $(e_t, t \ge 0)$) be a $PPP(G)$ on E. Let $f \in L^1(G)$ and define*

$$X_t^f := \int_{[0,1]\times E} f(x) M(ds\,dx)$$

(resp. $X_t^f := \sum_{0\le s\le t} f(e_s)$, where $f(*) := 0$) for $t \ge 0$. Then X_t^f is a Lévy process, and moreover

(i) $M_t^f := X_t^f - t\int_E f(x)G(dx)$ *defines a martingale; and*

(ii) if $f \in L^2(G) \cap L^1(G)$ *then* $((M_t^f)^2 - t\int_E f^2(x)G(dx), t \geq 0)$ *is a martingale.*

Proof: Let $M' : \mathbb{R}_+ \times E \to E$ denote the push-forward of M by projection on the second coordinate. Then $\int_{[0,t]\times E} f(x)M(ds\,dt) = M(f) = M'(f)$ and it is easy to see that its intensity is equal to $tG(x)$. Since $G(f) < \infty$, we obtain that $(s,x) \mapsto f(x)$ is in $L^1(ds\mathbf{1}_{[0,1]} \times G(dx))$.

Check Lévy:

(i) Let $0 \leq s < t$. Then

$$X_t^f - X_s^f = \int_{(s,t]\times E} f(x)M(ds\,dt)$$

and notice $M'(f) = M(f\mathbf{1}_{(s,t]})$ and M' is a $PRM((t-s)G)$, whence $X_t^f - X_s^f$ has the same distribution as X_{t-s}^f.

(ii) Let $0 = t_0 < t_1 < \cdots < t_k$. Then

$$X_{t_i} - X_{t_{i-1}} = \int_{(t_i,t_{i-1}]\times E} f(x)M(ds\,dt) = M|_{(t_i,t_{i-1}]\times E}(f)$$

which are independent since the bands $(t_i, t_{i-1}] \times E$ are pairwise disjoint.

Check martingale:

(i) Apply the formula for the first moment and note that the latter half of the right hand side of the first equation below is independent of $\mathscr{F}_t$.

$$\begin{aligned}
\mathbb{E}[X_{t+s}^f \mid \mathscr{F}_t] &= \mathbb{E}[\int_{[0,t]\times E} f(x)M(ds\,dx) + \int_{[t,t+s]\times E} f(x)M(ds\,dx) \mid \mathscr{F}_t] \\
&= X_t^f + \mathbb{E}[\int_{[t,t+s]\times E} f(x)M(ds\,dx)] \\
&= X_t^f + \int_{[t,t+s]\times E} f(x)ds G(dx) \\
&= X_t^f + sG(f) \\
&= M_t^f + (t+s)G(f)
\end{aligned}$$

(ii) Use the fact that $(M_t^f)^2 = (M_{t+s}^f - M_t^f)^2 - (M_t^f)^2 + 2M_t^f M_{t+s}^f$, and use the formula for the variance. □

7.3.4 Example (Poisson process). Taking $E = \{0\}$, $G = \theta\delta_0$, and $f(0) = 1$ gives the Poisson process which we have seen in its alternate form as a sum of i.i.d. exponential r.v.'s with parameter θ. Let $N_t^\theta = \sum_{k=1}^\infty \mathbf{1}_{T_1+\cdots+T_k \leq t}$, the *Poisson process* with intensity θ.

7.3.5 Example (Compound Poisson process). Let ν be a finite measure on $\mathbb{R}$. Let $M(dt\,dx) = \sum \delta_{(t_i,x_i)}$, a $PPP(\nu)$. Then for any t there are only a finite number of t_i in the interval $[0,t]$, so we may label them in increasing order $0(=t_0) < t_1 < t_2 < \cdots$. Then

$$N_t^\nu := \int_{[0,t]\times\mathbb{R}} x M(ds\,dx) = \sum_{i\geq 1} x_i \mathbf{1}_{t_i \leq t},$$

for $t \geq 0$, is the *compound Poisson process*. (By finiteness this makes sense even if $\int_{\mathbb{R}} |x|\nu(dx) = \infty$.) It is easy to see that $\sum_{i\geq 1} \delta_{t_i}$ is a $PRM(\theta dt)$, where $\theta = \nu(\mathbb{R})$, a Poisson point process corresponding to a homogeneous Poisson process with intensity θ. By the Marking property of Poisson random measures, we obtain the following alternative construction of $(N_t^\nu, t \geq 0)$. Take $T_1 < T_2 < \cdots$ such that the increments $(T_i - T_{i-1}, i > 1)$ are independent and exponentially distributed with parameter θ. Take an i.i.d. $(Y_i, i \geq 1)$ independent of (T_i) with $\mathscr{L}(Y_i) = \frac{\nu}{\nu(\mathbb{R})} = \frac{\nu}{\theta}$. Then (N_t^ν) has the same distribution as $(\sum_{i\geq 1} Y_i \mathbf{1}_{T_i \leq t}, t \geq 0)$. The law of N_1^ν is called the *compound Poisson distribution*, $CP(\nu)$, so

$$CP(\nu) = \sum_{n\geq 0} \underbrace{e^{-\theta}\frac{\theta^n}{n!}}_{\substack{\text{prob of } n \\ \text{atoms before} \\ \text{time 1}}} \underbrace{\left(\frac{\nu}{\theta}\right)^{*n}}_{\substack{\text{Law of} \\ Y_1+\cdots+Y_n}} = e^{-\theta}\sum_{n\geq 0}\frac{\nu^{*n}}{n!}.$$

(Recall that $\mu * \nu$ is the convolution of μ and ν, the unique measure such that $\mu * \nu(f) = \int f(x+y)\mu(dx)\nu(dy)$.)

Remark. Notice that

$$\begin{aligned}
\mathbb{E}[e^{i\lambda N_t^\nu}] &= \exp\left(\int_{[0,1]\times\mathbb{R}} ds\nu(dx)(e^{i\lambda x} - 1)\right) \\
&= \exp\left(t\theta \int_{\mathbb{R}} \frac{\nu(dx)}{\theta}(e^{i\lambda x} - 1)\right) \\
&= \exp\Big(t\theta(\phi_{\frac{\nu}{\theta}}(\lambda) - 1)\Big)
\end{aligned}$$

where $\phi_{\frac{\nu}{\theta}}$ is the characteristic function of $\frac{\nu}{\theta}$.

7.4 Lévy processes in $\mathbb{R}$

Recall that $(X_t, t \geq 0)$ is a Lévy process if it has stationary independent increments. Notice that the law of the Lévy process (X_t) is determined by the 1-dimensional marginals $(\mathscr{L}(X_t), t \geq 0)$, since

$$\mathscr{L}(X_{t_1}, \ldots, X_{t_k}) = \mathscr{L}(F((X_{t_i} - X_{t_{i-1}})_{1\leq i\leq k}))$$

for some function F, and each of the $X_{t_i} - X_{t_{i-1}}$ are independent with law $\mathscr{L}(X_{t_i - t_{i-1}})$.

7.4.1 Definition. A triple (a, q, Π) is a *Lévy triple* if

(i) $a \in \mathbb{R}$;

(ii) $q \geq 0$; and

(iii) Π is a σ-finite measure on $\mathbb{R}$ such that $\Pi(\{0\}) = 0$ and

$$\int_{\mathbb{R}} \Pi(dx)(x^2 \wedge 1) < \infty.$$

7.4.2 Theorem. *There is a one-to-one correspondence between càdlàg Lévy processes $(X_t, t \geq 0)$ and Lévy triples (a, q, Π) in such a way that if $(X_t, t \geq 0)$ is the càdlàg Lévy process associated with (a, q, Π) then $\mathbb{E}[e^{i\lambda X_t}] = e^{t\Psi(\lambda)}$ where*

$$\Psi(\lambda) = ia\lambda - q\frac{\lambda^2}{2} + \int_{\mathbb{R}} \Pi(dx)(e^{i\lambda x} - 1 - i\lambda x \mathbf{1}_{|x|\leq 1}).$$

This is the Lévy-Khintchine formula.

7.4.3 Examples.

(i) If $q = 0$ and Π is the zero measure then $\mathbb{E}[e^{i\lambda X_t}] = e^{i\lambda at}$, so $X_t = at$ and X is just a (deterministic) linear function.

(ii) If $a = 0$ and Π is the zero measure then $\mathbb{E}[e^{i\lambda X_t}] = e^{-tq\frac{\lambda^2}{2}}$, so X_t has distribution $\mathcal{N}(0, qt)$ and hence $X_t = \sqrt{q}B_t$, where B is a standard Brownian motion.

(iii) If $q = 0$, $\Pi(dx)$ is a finite measure (i.e. $\int_{\mathbb{R}} \Pi(dx)|x| < \infty$), and if $a = \int_{-1}^{1} x\Pi(dx)$, then $\Psi(\lambda) = \int_{\mathbb{R}} \Pi(dx)(e^{i\lambda x} - 1)$, so X_t is a compound Poisson process with intensity Π.

In general, Lévy processes have jumps which are only square-summable over compact intervals and the term $i\lambda x \mathbf{1}_{|x|\leq 1}$ is a compensation for this.

How can we construct the Lévy process with triple (a, q, Π)? Let $(Y^n, n \geq 1)$ be independent compound Poisson processes with respective intensities equal to $\Pi_n := \Pi|_{(\frac{1}{n+1}, \frac{1}{n}]}$. Then $\Pi_n(|x|) < \infty$ for all n, so $M^n = (Y_t^n - t\int_{\mathbb{R}} x\Pi_n(dx), t \geq 0)$ is a martingale. Let $\overline{M}_t^n = \sum_{k=1}^{n} M_t^k$, and note that

$$\mathbb{E}[e^{i\lambda M_t^n}] = \exp\left(t \int_{\frac{1}{n+1}}^{\frac{1}{n}} \Pi_n(dx)(e^{i\lambda x} - 1 - i\lambda x) \right).$$

7.4.4 Theorem. *For every $t \geq 0$, $\overline{M}_t^n \to M_t^\infty$ in L^2, where $(M_t^\infty, t \geq 0)$ is an L^2-martingale with a càdlàg modification, which we will also denote by M^∞. If we are given*

(i) a standard BM (B_t);

(ii) a CP process Y^0 with intensity $(\Pi|_{(0,\infty)})$; and

(iii) the martingale M^∞;

then $X_t = at + \sqrt{q}B_t + Y_t^0 + M_t^\infty$ is a càdlàg Lévy process with characteristic function given by the Lévy-Khintchine formula.

PROOF: Let $n \geq m \geq 1$ and notice

$$\mathbb{E}\left[\sup_{0\leq s\leq t} |\overline{M}_s^n - \overline{M}_s^m|^2\right] \leq 4\,\mathbb{E}[|\overline{M}_t^n - \overline{M}_t^m|^2]$$

by Doob's L^2-inequality. Now

$$\mathbb{E}\left[|\sum_{k=m+1}^{n} M_t^k|^2\right] = t\int_{x=\frac{1}{n+1}}^{\frac{1}{m+1}} x^2\Pi(dx)$$

by the second moment formula for PRM's since

$$mg \longleftrightarrow PPP(\Pi(dx)|_{]\frac{1}{n+1},\frac{1}{m+1}]})??$$

Thus, for all $\varepsilon > 0$ and for all n, m large, $\mathbb{E}[\sup_{0\leq s\leq t} |\overline{M}_s^n - \overline{M}_s^m|^2] \leq \varepsilon$, so in particular $(\overline{M}_t^n)_{n\geq 1}$ is an L^2-Cauchy sequence. Therefore $\overline{M}_t^n \to M_t^\infty$ in L^2 as $n \to \infty$. By the Borel-Cantelli Lemma and the above we even get that $\overline{M}_t^n \to M^\infty$ uniformly over compact intervals, up to extraction. A uniform limit of càdlàg processes is càdlàg, so M^∞ is a càdlàg martingale. □

Index