# Part III — Advanced Probability
## Theorems with proof

Based on lectures by M. Lis
Notes taken by Dexter Chua

Michaelmas 2017



The aim of the course is to introduce students to advanced topics in modern probability theory. The emphasis is on tools required in the rigorous analysis of stochastic processes, such as Brownian motion, and in applications where probability theory plays an important role.

**Review of measure and integration:** sigma-algebras, measures and filtrations; integrals and expectation; convergence theorems; product measures, independence and Fubini's theorem.
**Conditional expectation:** Discrete case, Gaussian case, conditional density functions; existence and uniqueness; basic properties.
**Martingales:** Martingales and submartingales in discrete time; optional stopping; Doob's inequalities, upcrossings, martingale convergence theorems; applications of martingale techniques.
**Stochastic processes in continuous time:** Kolmogorov's criterion, regularization of paths; martingales in continuous time.
**Weak convergence:** Definitions and characterizations; convergence in distribution, tightness, Prokhorov's theorem; characteristic functions, Lévy's continuity theorem.
**Sums of independent random variables:** Strong laws of large numbers; central limit theorem; Cramér's theory of large deviations.
**Brownian motion:** Wiener's existence theorem, scaling and symmetry properties; martingales associated with Brownian motion, the strong Markov property, hitting times; properties of sample paths, recurrence and transience; Brownian motion and the Dirichlet problem; Donsker's invariance principle.
**Poisson random measures:** Construction and properties; integrals.
**Lévy processes:** Lévy-Khinchin theorem.

### Pre-requisites

A basic familiarity with measure theory and the measure-theoretic formulation of probability theory is very helpful. These foundational topics will be reviewed at the beginning of the course, but students unfamiliar with them are expected to consult the literature (for instance, Williams' book) to strengthen their understanding.

# Contents

**0 Introduction** **3**

**1 Some measure theory** **4**
1.1 Review of measure theory . . . . . . . . . . . . . . . . . . . . . . . . 4
1.2 Conditional expectation . . . . . . . . . . . . . . . . . . . . . . . . . . 5

**2 Martingales in discrete time** **9**
2.1 Filtrations and martingales . . . . . . . . . . . . . . . . . . . . . . . . 9
2.2 Stopping time and optimal stopping . . . . . . . . . . . . . . . . . . 9
2.3 Martingale convergence theorems . . . . . . . . . . . . . . . . . . . . 10
2.4 Applications of martingales . . . . . . . . . . . . . . . . . . . . . . . . 15

**3 Continuous time stochastic processes** **19**

**4 Weak convergence of measures** **24**

**5 Brownian motion** **28**
5.1 Basic properties of Brownian motion . . . . . . . . . . . . . . . . . . 28
5.2 Harmonic functions and Brownian motion . . . . . . . . . . . . . . . 33
5.3 Transience and recurrence . . . . . . . . . . . . . . . . . . . . . . . . 35
5.4 Donsker's invariance principle . . . . . . . . . . . . . . . . . . . . . . 37

**6 Large deviations** **40**

# 0 Introduction

# 1 Some measure theory

## 1.1 Review of measure theory

**Theorem.** Let $(E, \mathcal{E}, \mu)$ be a measure space. Then there exists a unique function $\tilde{\mu} : m\mathcal{E}^+ \to [0, \infty]$ satisfying

- $\tilde{\mu}(\mathbf{1}_A) = \mu(A)$, where $\mathbf{1}_A$ is the indicator function of $A$.
- Linearity: $\tilde{\mu}(\alpha f + \beta g) = \alpha\tilde{\mu}(f) + \beta\tilde{\mu}(g)$ if $\alpha, \beta \in \mathbb{R}_{\geq 0}$ and $f, g \in m\mathcal{E}^+$.
- Monotone convergence: iff $f_1, f_2, \ldots \in m\mathcal{E}^+$ are such that $f_n \nearrow f \in m\mathcal{E}^+$ pointwise a.e. as $n \to \infty$, then

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

We call $\tilde{\mu}$ the *integral* with respect to $\mu$, and we will write it as $\mu$ from now on.

**Lemma** (Fatou's lemma)**.** Let $f_i \in m\mathcal{E}^+$. Then

$$\mu\left(\liminf_n f_n\right) \leq \liminf_n \mu(f_n).$$

*Proof.* Apply monotone convergence to the sequence $\inf_{m\geq n} f_m$ □

**Theorem** (Dominated convergence theorem)**.** If $f_i \in m\mathcal{E}$ and $f_i \to f$ a.e., such that there exists $g \in L^1$ such that $|f_i| \leq g$ a.e. Then

$$\mu(f) = \lim \mu(f_n).$$

*Proof.* Apply Fatou's lemma to $g - f_n$ and $g + f_n$. □

**Theorem.** If $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ are $\sigma$-finite measure spaces, then there exists a unique measure $\mu$ on $\mathcal{E}_1 \otimes \mathcal{E}_2)$ satisfying

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

for all $A_i \in \mathcal{E}_i$.

This is called the *product measure*.

**Theorem** (Fubini's/Tonelli's theorem)**.** If $f = f(x_1, x_2) \in m\mathcal{E}^+$ with $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2$, then the functions

$$\begin{aligned} x_1 &\mapsto \int f(x_1, x_2)\mathrm{d}\mu_2(x_2) \in m\mathcal{E}_1^+ \\ x_2 &\mapsto \int f(x_1, x_2)\mathrm{d}\mu_1(x_1) \in m\mathcal{E}_2^+ \end{aligned}$$

and

$$\begin{aligned} \int_E f \;\mathrm{d}u &= \int_{E_1}\left(\int_{E_2} f(x_1, x_2)\;\mathrm{d}\mu_2(x_2)\right)\mathrm{d}\mu_1(x_1) \\ &= \int_{E_2}\left(\int_{E_1} f(x_1, x_2)\;\mathrm{d}\mu_1(x_1)\right)\mathrm{d}\mu_2(x_2) \end{aligned}$$

## 1.2 Conditional expectation

**Lemma.** The conditional expectation $Y = \mathbb{E}(X \mid \mathcal{G})$ satisfies the following properties:

- $Y$ is $\mathcal{G}$-measurable
- We have $Y \in L^1$, and
$$\mathbb{E}Y\mathbf{1}_A = \mathbb{E}X\mathbf{1}_A$$
for all $A \in \mathcal{G}$.

*Proof.* It is clear that $Y$ is $\mathcal{G}$-measurable. To show it is $L^1$, we compute

$$\begin{aligned}
\mathbb{E}[|Y|] &= \mathbb{E}\left|\sum_{n=1}^{\infty} \mathbb{E}(X \mid G_n)\mathbf{1}_{G_n}\right| \\
&\leq \mathbb{E}\sum_{n=1}^{\infty} \mathbb{E}(|X| \mid G_n)\mathbf{1}_{G_n} \\
&= \sum \mathbb{E}\left(\mathbb{E}(|X| \mid G_n)\mathbf{1}_{G_n}\right) \\
&= \sum \mathbb{E}|X|\mathbf{1}_{G_n} \\
&= \mathbb{E}\sum |X|\mathbf{1}_{G_n} \\
&= \mathbb{E}|X| \\
&< \infty,
\end{aligned}$$

where we used monotone convergence twice to swap the expectation and the sum.

The final part is also clear, since we can explicitly enumerate the elements in $\mathcal{G}$ and see that they all satisfy the last property. □

**Theorem** (Existence and uniqueness of conditional expectation)**.** Let $X \in L^1$, and $\mathcal{G} \subseteq \mathcal{F}$. Then there exists a random variable $Y$ such that

- $Y$ is $\mathcal{G}$-measurable
- $Y \in L^1$, and $\mathbb{E}X\mathbf{1}_A = \mathbb{E}Y\mathbf{1}_A$ for all $A \in \mathcal{G}$.

Moreover, if $Y'$ is another random variable satisfying these conditions, then $Y' = Y$ almost surely.

We call $Y$ a (version of) the conditional expectation given $\mathcal{G}$.

*Proof.* We first consider the case where $X \in L^2(\Omega, \mathcal{F}, \mu)$. Then we know from functional analysis that for any $\mathcal{G} \subseteq \mathcal{F}$, the space $L^2(\mathcal{G})$ is a Hilbert space with inner product
$$\langle X, Y\rangle = \mu(XY).$$
In particular, $L^2(\mathcal{G})$ is a closed subspace of $L^2(\mathcal{F})$. We can then define $Y$ to be the orthogonal projection of $X$ onto $L^2(\mathcal{G})$. It is immediate that $Y$ is $\mathcal{G}$-measurable. For the second part, we use that $X - Y$ is orthogonal to $L^2(\mathcal{G})$, since that's what orthogonal projection is supposed to be. So $\mathbb{E}(X - Y)Z = 0$ for all $Z \in L^2(\mathcal{G})$. In particular, since the measure space is finite, the indicator function of any measurable subset is $L^2$. So we are done.

We next focus on the case where $X \in m\mathcal{E}^+$. We define

$$X_n = X \wedge n$$

We want to use monotone convergence to obtain our result. To do so, we need the following result:

**Claim.** If $(X, Y)$ and $(X', Y')$ satisfy the conditions of the theorem, and $X' \geq X$ a.s., then $Y' \geq Y$ a.s.

*Proof.* Define the event $A = \{Y' \leq Y\} \in \mathcal{G}$. Consider the event $Z = (Y - Y')\mathbf{1}_A$. Then $Z \geq 0$. We then have

$$\mathbb{E}Y'\mathbf{1}_A = \mathbb{E}X'\mathbf{1}_A \geq \mathbb{E}X\mathbf{1}_A = \mathbb{E}Y\mathbf{1}_A.$$

So it follows that we also have $\mathbb{E}(Y - Y')\mathbf{1}_A \leq 0$. So in fact $\mathbb{E}Z = 0$. So $Y' \geq Y$ a.s. □

We can now define $Y_n = \mathbb{E}(X_n \mid \mathcal{G})$, picking them so that $\{Y_n\}$ is increasing. We then take $Y_\infty = \lim Y_n$. Then $Y_\infty$ is certainly $\mathcal{G}$-measurable, and by monotone convergence, if $A \in \mathcal{G}$, then

$$\mathbb{E}X\mathbf{1}_A = \lim \mathbb{E}X_n\mathbf{1}_A = \lim \mathbb{E}Y_n\mathbf{1}_A = \mathbb{E}Y_\infty\mathbf{1}_A.$$

Now if $\mathbb{E}X < \infty$, then $\mathbb{E}Y_\infty = \mathbb{E}X < \infty$. So we know $Y_\infty$ is finite a.s., and we can define $Y = Y_\infty\mathbf{1}_{Y_\infty < \infty}$.

Finally, we work with arbitrary $X \in L^1$. We can write $X = X^+ - X^-$, and then define $Y^\pm = \mathbb{E}(X^\pm \mid \mathcal{G})$, and take $Y = Y^+ - Y^-$.

Uniqueness is then clear. □

**Lemma.** If $Y$ is $\sigma(Z)$-measurable, then there exists $h : \mathbb{R} \to \mathbb{R}$ Borel-measurable such that $Y = h(Z)$. In particular,

$$\mathbb{E}(X \mid Z) = h(Z) \text{ a.s.}$$

for some $h : \mathbb{R} \to \mathbb{R}$.

**Proposition.**

(i) $\mathbb{E}(X \mid \mathcal{G}) = X$ iff $X$ is $\mathcal{G}$-measurable.

(ii) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}X$

(iii) If $X \geq 0$ a.s., then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$

(iv) If $X$ and $\mathcal{G}$ are independent, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}[X]$

(v) If $\alpha, \beta \in \mathbb{R}$ and $X_1, X_2 \in L^1$, then

$$\mathbb{E}(\alpha X_1 + \beta X_2 \mid \mathcal{G}) = \alpha\mathbb{E}(X_1 \mid \mathcal{G}) + \beta\mathbb{E}(X_2 \mid \mathcal{G}).$$

(vi) Suppose $X_n \nearrow X$. Then

$$\mathbb{E}(X_n \mid \mathcal{G}) \nearrow \mathbb{E}(X \mid \mathcal{G}).$$

(vii) *Fatou's lemma*: If $X_n$ are non-negative measurable, then

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \mid \mathcal{G}\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n \mid \mathcal{G}).$$

(viii) *Dominated convergence theorem*: If $X_n \to X$ and $Y \in L^1$ such that $Y \geq |X_n|$ for all $n$, then

$$\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G}).$$

(ix) *Jensen's inequality*: If $c : \mathbb{R} \to \mathbb{R}$ is convex, then

$$\mathbb{E}(c(X) \mid \mathcal{G}) \geq c(\mathbb{E}(X) \mid \mathcal{G}).$$

(x) *Tower property*: If $\mathcal{H} \subseteq \mathcal{G}$, then

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}).$$

(xi) For $p \geq 1$,

$$\|\mathbb{E}(X \mid \mathcal{G})\|_p \leq \|X\|_p.$$

(xii) If $Z$ is bounded and $\mathcal{G}$-measurable, then

$$\mathbb{E}(ZX \mid \mathcal{G}) = Z\mathbb{E}(X \mid \mathcal{G}).$$

(xiii) Let $X \in L^1$ and $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$. Assume that $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$. Then

$$\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})).$$

*Proof.*

(i) Clear.

(ii) Take $A = \omega$.

(iii) Shown in the proof.

(iv) Clear by property of expected value of independent variables.

(v) Clear, since the RHS satisfies the unique characterizing property of the LHS.

(vi) Clear from construction.

(vii) Same as the unconditional proof, using the previous property.

(viii) Same as the unconditional proof, using the previous property.

(ix) Same as the unconditional proof.

(x) The LHS satisfies the characterizing property of the RHS

(xi) Using the convexity of $|x|^p$, Jensen's inequality tells us

$$\begin{aligned}\|E(X \mid \mathcal{G})\|_p^p &= \mathbb{E}|\mathbb{E}(X \mid \mathcal{G})|^p\\ &\leq \mathbb{E}(\mathbb{E}(|X|^p \mid \mathcal{G}))\\ &= \mathbb{E}|X|^p\\ &= \|X\|_p^p\end{aligned}$$

(xii) If $Z = \mathbf{1}_B$, and let $b \in \mathcal{G}$. Then

$$\mathbb{E}(Z\mathbb{E}(X \mid \mathcal{G})\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \cdot \mathbf{1}_{A\cap B}) = \mathbb{E}(X\mathbf{1}_{A\cap B}) = \mathbb{E}(ZX\mathbf{1}_A).$$

So the lemma holds. Linearity then implies the result for $Z$ simple, then apply our favorite convergence theorems.

(xiii) Take $B \in \mathcal{H}$ and $A \in \mathcal{G}$. Then

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

If instead of $A \cap B$, we had any $\sigma(\mathcal{G}, \mathcal{H})$-measurable set, then we would be done. But we are fine, since the set of subsets of the form $A \cap B$ with $A \in \mathcal{G}$, $B \in \mathcal{H}$ is a generating $\pi$-system for $\sigma(\mathcal{H}, \mathcal{G})$. $\square$

**Lemma.** If $X \in L^1$, then the family of random variables $Y_\mathcal{G} = \mathbb{E}(X \mid \mathcal{G})$ for all $\mathcal{G} \subseteq \mathcal{F}$ is uniformly integrable.

In other words, for all $\varepsilon > 0$, there exists $\lambda > 0$ such that

$$\mathbb{E}(Y_\mathcal{G}\mathbf{1}_{|Y_\mathcal{G} > \lambda|}) < \varepsilon$$

for all $\mathcal{G}$.

*Proof.* Fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that $\mathbb{E}|X|\mathbf{1}_A < \varepsilon$ for any $A$ with $\mathbb{P}(A) < \delta$.

Take $Y = \mathbb{E}(X \mid \mathcal{G})$. Then by Jensen, we know

$$|Y| \leq \mathbb{E}(|X| \mid \mathcal{G})$$

In particular, we have

$$\mathbb{E}|Y| \leq \mathbb{E}|X|.$$

By Markov's inequality, we have

$$\mathbb{P}(|Y| \geq \lambda) \leq \frac{\mathbb{E}|Y|}{\lambda} \leq \frac{\mathbb{E}|X|}{\lambda}.$$

So take $\lambda$ such that $\frac{\mathbb{E}|X|}{\lambda} < \delta$. So we have

$$\mathbb{E}(|Y|\mathbf{1}_{|Y|\geq\lambda}) \leq \mathbb{E}(\mathbb{E}(|X| \mid \mathcal{G})\mathbf{1}_{|Y|\geq\lambda}) = \mathbb{E}(|X|\mathbf{1}_{|Y|\geq\lambda}) < \varepsilon$$

using that $\mathbf{1}_{|Y|\geq\lambda}$ is a $\mathcal{G}$-measurable function. $\square$

# 2 Martingales in discrete time

## 2.1 Filtrations and martingales

## 2.2 Stopping time and optimal stopping

**Proposition.**

(i) If $T, S, (T_n)_{n \geq 0}$ are all stopping times, then
$$T \vee S, T \wedge S, \sup_n T_n, \inf T_n, \limsup T_n, \liminf T_n$$
are all stopping times.

(ii) $\mathcal{F}_T$ is a $\sigma$-algebra

(iii) If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

(iv) $X_T \mathbf{1}_{T<\infty}$ is $\mathcal{F}_T$-measurable.

(v) If $(X_n)$ is an adapted process, then so is $(X_n^T)_{n \geq 0}$ for any stopping time $T$.

(vi) If $(X_n)$ is an integrable process, then so is $(X_n^T)_{n \geq 0}$ for any stopping time $T$. □

**Theorem** (Optional stopping theorem)**.** Let $(X_n)_{n \geq 0}$ be a super-martingale and $S \leq T$ *bounded* stopping times. Then
$$\mathbb{E}X_T \leq \mathbb{E}X_S.$$

*Proof.* Follows from the next theorem. □

**Theorem.** The following are equivalent:

(i) $(X_n)_{n \geq 0}$ is a super-martingale.

(ii) For any bounded stopping times $T$ and any stopping time $S$,
$$\mathbb{E}(X_T \mid \mathcal{F}_S) \leq X_{S \wedge T}.$$

(iii) $(X_n^T)$ is a super-martingale for any stopping time $T$.

(iv) For bounded stopping times $S, T$ such that $S \leq T$, we have
$$\mathbb{E}X_T \leq \mathbb{E}X_S.$$

*Proof.*

– (ii) $\Rightarrow$ (iii): Consider $(X_n^{T'})_{\geq 0}$ for a stopping time $T'$. To check if this is a super-martingale, we need to prove that whenever $m \leq n$,
$$\mathbb{E}(X_{n \wedge T'} \mid \mathcal{F}_m) \leq X_{m \wedge T'}.$$
But this follows from (ii) above by taking $S = m$ and $T = T' \wedge n$.

– (ii) $\Rightarrow$ (iv): Clear by the tower law.

- (iii) $\Rightarrow$ (i): Take $T = \infty$.
- (i) $\Rightarrow$ (ii): Assume $T \leq n$. Then

$$\begin{aligned} X_T &= X_{S \wedge T} + \sum_{S \leq k < T} (X_{k+1} - X_k) \\ &= X_{S \wedge T} + \sum_{k=0}^{n} (X_{k+1} - X_k) \mathbf{1}_{S \leq k < T} \end{aligned} \tag{$*$}$$

  Now note that $\{S \leq k < T\} = \{S \leq k\} \cap \{T \leq k\}^c \in \mathcal{F}_k$. Let $A \in \mathcal{F}_S$. Then $A \cap \{S \leq k\} \in \mathcal{F}_k$ by definition of $\mathcal{F}_S$. So $A \cap \{S \leq k < T\} \in \mathcal{F}_k$.

  Apply $\mathbb{E}$ to to $(*) \times \mathbf{1}_A$. Then we have

$$\mathbb{E}(X_T \mathbf{1}_A) = \mathbb{E}(X_{S \wedge T} \mathbf{1}_A) + \sum_{k=0}^{n} \mathbb{E}(X_{k+1} - X_k) \mathbf{1}_{A \cap \{S \leq k < T\}}.$$

  But for all $k$, we know

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

  since $X$ is a super-martingale. So it follows that for all $A \in \mathcal{F}_S$, we have

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

  But since $X_{S \wedge T}$ is $\mathcal{F}_{S \wedge T}$ measurable, it is in particular $\mathcal{F}_S$ measurable. So it follows that for all $A \in \mathcal{F}_S$, we have

$$\mathbb{E}(\mathbb{E}(X_T \mid \mathcal{F}_S) \mathbf{1}_A) \leq \mathbb{E}(X_{S \wedge T} \mathbf{1}_A).$$

  So the result follows.
- (iv) $\Rightarrow$ (i): Fix $m \leq n$ and $A \in \mathcal{F}_m$. Take

$$T = m \mathbf{1}_A + \mathbf{1}_{A^C}.$$

  One then manually checks that this is a stopping time. Now note that

$$X_T = X_m \mathbf{1}_A + X_n \mathbf{1}_{A^C}.$$

  So we have

$$\begin{aligned} 0 &\geq \mathbb{E}(X_n) - \mathbb{E}(X_T) \\ &= \mathbb{E}(X_n) - \mathbb{E}(X_n \mathbf{1}_{A^c}) - \mathbb{E}(X_m \mathbf{1}_A) \\ &= \mathbb{E}(X_n \mathbf{1}_A) - \mathbb{E}(X_m \mathbf{1}_A). \end{aligned}$$

  Then the same argument as before gives the result. □

## 2.3 Martingale convergence theorems

**Theorem** (Almost sure martingale convergence theorem)**.** Suppose $(X_n)_{n \geq 0}$ is a super-martingale that is bounded in $L^1$, i.e. $\sup_n \mathbb{E}|X_n| < \infty$. Then there exists an $\mathcal{F}_\infty$-measurable $X_\infty \in L^1$ such that

$$X_n \to X_\infty \text{ a.s. as } n \to \infty.$$

**Lemma.** Let $(x_n)_{n\geq 0}$ be a sequence of numbers. Then $x_n$ converges in $\mathbb{R}$ if and only if

(i) $\liminf |x_n| < \infty$.

(ii) For all $a, b \in \mathbb{Q}$ with $a < b$, we have $U[a, b, (x_n)] < \infty$.

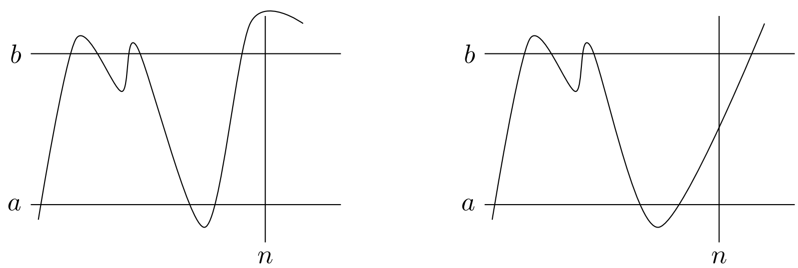
**Lemma** (Doob's upcrossing lemma)**.** If $X_n$ is a super-martingale, then

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

*Proof.* Assume that $X$ is a positive super-martingale. We define stopping times $S_k, T_k$ as follows:

- $T_0 = 0$
- $S_{k+1} = \inf\{n : X_n \leq a, n \geq T_n\}$
- $T_{k+1} = \inf\{n : X_n \geq b, n \geq S_{k+1}\}$.

Given an $n$, we want to count the number of upcrossings before $n$. There are two cases we might want to distinguish:




Now consider the sum

$$\sum_{k=1}^{n} X_{T_k \wedge n} - X_{S_k \wedge n}.$$

In the first case, this is equal to

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

In the second case, it is equal to

$$\sum_{k=1}^{U_n} X_{T_k} - X_{S_k} + (X_n - X_{S_{U_n+1}}) + \sum_{k=U_n+2}^{n} X_n - X_n \geq (b-a)U_n + (X_n - X_{S_{U_n+1}}).$$

Thus, in general, we have

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

By definition, $S_k < T_k \leq n$. So the expectation of the LHS is always non-negative by super-martingale convergence, and thus

$$0 \geq (b-a)\mathbb{E}U_n + \mathbb{E}(X_n - X_{S_{U_n+1} \wedge n}).$$

Then observe that

$$X_n - X_{S_{U_n+1}} \geq -(X_n - a)^-. \qquad \square$$

**Lemma** (Maximal inequality)**.** Let $(X_n)$ be a sub-martingale that is non-negative, or a martingale. Define

$$X_n^* = \sup_{k \leq n} |X_k|, \quad X^* = \lim_{n \to \infty} X_n^*.$$

If $\lambda \geq 0$, then

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

In particular, we have

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

*Proof.* If $X_n$ is a martingale, then $|X_n|$ is a sub-martingale. So it suffices to consider the case of a non-negative sub-martingale. We define the stopping time

$$T = \inf\{n : X_n \geq \lambda\}.$$

By optional stopping,

$$\begin{aligned}
\mathbb{E}X_n &\geq \mathbb{E}X_{T \wedge n} \\
&= \mathbb{E}X_T \mathbf{1}_{T \leq n} + \mathbb{E}X_n \mathbf{1}_{T > n} \\
&\geq \lambda \mathbb{P}(T \leq n) + \mathbb{E}X_n \mathbf{1}_{T > n} \\
&= \lambda \mathbb{P}(X_n^* \geq \lambda) + \mathbb{E}X_n \mathbf{1}_{T > n}.
\end{aligned}$$

**Lemma** (Doob's $L^p$ inequality)**.** For $p > 1$, we have

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

for all $n$.

*Proof.* Let $k > 0$, and consider

$$\|X_n^* \wedge k\|_p^p = \mathbb{E}|X_n^* \wedge k|^p.$$

We use the fact that

$$x^p = \int_0^x p s^{p-1}\;\mathrm{d}s.$$

So we have

$$\begin{aligned}
\|X_n^* \wedge k\|_p^p &= \mathbb{E}|X_n^* \wedge k|^p \\
&= \mathbb{E}\int_0^{X_n^* \wedge k} px^{p-1}\;\mathrm{d}x \\
&= \mathbb{E}\int_0^k px^{p-1}\mathbf{1}_{X_n^* \geq x}\;\mathrm{d}x \\
&= \int_0^k px^{p-1}\mathbb{P}(X_n^* \geq x)\;\mathrm{d}x && \text{(Fubini)} \\
&\leq \int_0^k px^{p-2}\mathbb{E}X_n\mathbf{1}_{X_n^* \geq x}\;\mathrm{d}x && \text{(maximal inequality)} \\
&= \mathbb{E}X_n\int_0^k px^{p-2}\mathbf{1}_{X_n^* \geq x}\;\mathrm{d}x && \text{(Fubini)} \\
&= \frac{p}{p-1}\mathbb{E}X_n(X_n^* \wedge k)^{p-1} \\
&\leq \frac{p}{p-1}\|X_n\|_p\left(\mathbb{E}(X_n^* \wedge k)^p\right)^{\frac{p-1}{p}} && \text{(H older)} \\
&= \frac{p}{p-1}\|X_n\|_p\|X_n^* \wedge k\|_p^{p-1}
\end{aligned}$$

Now take the limit $k \to \infty$ and divide by $\|X_n^*\|_p^{p-1}$. □

**Theorem** ($L^p$ martingale convergence theorem)**.** Let $(X_n)_{n\geq 0}$ be a martingale, and $p > 1$. Then the following are equivalent:

(i) $(X_n)_{n\geq 0}$ is bounded in $L^p$, i.e. $\sup_n \mathbb{E}|X_i|^p < \infty$.

(ii) $(X_n)_{n\geq 0}$ converges as $n \to \infty$ to a random variable $X_\infty \in L^p$ almost surely and in $L^p$.

(iii) There exists a random variable $Z \in L^p$ such that

$$X_n = \mathbb{E}(Z \mid \mathcal{F}_n)$$

Moreover, in (iii), we always have $X_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$.

*Proof.*

- (i) ⇒ (ii): If $(X_n)_{n\geq 0}$ is bounded in $L^p$, then it is bounded in $L^1$. So by the martingale convergence theorem, we know $(X_n)_{n\geq 0}$ converges almost surely to $X_\infty$. By Fatou's lemma, we have $X_\infty \in L^p$.

  Now by monotone convergence, we have

  $$\|X^*\|_p = \lim_n \|X_n^*\|_p \leq \frac{p}{p-1}\sup_n \|X_n\|_p < \infty.$$

  By the triangle inequality, we have

  $$|X_n - X_\infty| \leq 2X^* \text{ a.s.}$$

  So by dominated convergence, we know that $X_n \to X_\infty$ in $L^p$.

– (ii) ⇒ (iii): Take $Z = X_\infty$. We want to prove that

$$X_m = \mathbb{E}(X_\infty \mid \mathcal{F}_m).$$

To do so, we show that $\|X_m - \mathbb{E}(X_\infty \mid \mathcal{F}_m)\|_p = 0$. For $n \geq m$, we know this is equal to

$$\|E(X_n \mid \mathcal{F}_m) - \mathbb{E}(X_\infty \mid \mathcal{F}_m)\|_p = \|\mathbb{E}(X_n - X_\infty \mid \mathcal{F}_m)\|_p \leq \|X_n - X_\infty\|_p \to 0$$

as $n \to \infty$, where the last step uses Jensen's. But it is also a constant. So we are done.

– (iii) ⇒ (i): Since expectation decreases $L^p$ norms, we already know that $(X_n)_{n\geq 0}$ is $L^p$-bounded.

To show the "moreover" part, note that $\bigcup_{n\geq 0} \mathcal{F}_n$ is a $\pi$-system that generates $\mathcal{F}_\infty$. So it is enough to prove that

$$\mathbb{E}X_\infty \mathbf{1}_A = \mathbb{E}(\mathbb{E}(Z \mid \mathcal{F}_\infty)\mathbf{1}_A).$$

But if $A \in \mathcal{F}_N$, then

$$\begin{aligned}\mathbb{E}X_\infty \mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}X_n \mathbf{1}_A\\ &= \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Z \mid \mathcal{F}_n)\mathbf{1}_A)\\ &= \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Z \mid \mathcal{F}_\infty)\mathbf{1}_A),\end{aligned}$$

where the last step relies on the fact that $\mathbf{1}_A$ is $\mathcal{F}_n$-measurable. □

**Theorem** (Convergence in $L^1$)**.** Let $(X_n)_{n\geq 0}$ be a martingale. Then the following are equivalent:

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

(ii) $(X_n)_{n\geq 0}$ converges almost surely and in $L^1$.

(iii) There exists $Z \in L^1$ such that $X_n = \mathbb{E}(Z \mid \mathcal{F}_n)$ almost surely.

Moreover, $X_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$.

*Proof.*

– (i) ⇒ (ii): Let $(X_n)_{n\geq 0}$ be uniformly integrable. Then $(X_n)_{n\geq 0}$ is bounded in $L^1$. So the $(X_n)_{n\geq 0}$ converges to $X_\infty$ almost surely. Then by measure theory, uniform integrability implies that in fact $X_n \to L^1$.

– (ii) ⇒ (iii): Same as the $L^p$ case.

– (iii) ⇒ (i): For any $Z \in L^1$, the collection $\mathbb{E}(Z \mid \mathcal{G})$ ranging over all $\sigma$-subalgebras $\mathcal{G}$ is uniformly integrable. □

**Theorem.** If $(X_n)_{n\geq 0}$ is a uniformly integrable martingale, and $S, T$ are arbitrary stopping times, then $\mathbb{E}(X_T \mid \mathcal{F}_S) = X_{S\wedge T}$. In particular $\mathbb{E}X_T = X_0$.

*Proof.* By optional stopping, for every $n$, we know that

$$\mathbb{E}(X_{T\wedge n} \mid \mathcal{F}_S) = X_{S\wedge T\wedge n}.$$

We want to be able to take the limit as $n \to \infty$. To do so, we need to show that things are uniformly integrable. First, we apply optional stopping to write $X_{T\wedge n}$ as

$$\begin{aligned} X_{T\wedge n} &= \mathbb{E}(X_n \mid \mathcal{F}_{T\wedge n}) \\ &= \mathbb{E}(\mathbb{E}(X_\infty \mid \mathcal{F}_n) \mid \mathcal{F}_{T\wedge n}) \\ &= \mathbb{E}(X_\infty \mid \mathcal{F}_{T\wedge n}). \end{aligned}$$

So we know $(X_n^T)_{n\geq 0}$ is uniformly integrable, and hence $X_{n\wedge T} \to X_T$ almost surely and in $L^1$.

To understand $\mathbb{E}(X_{T\wedge n} \mid \mathcal{F}_S)$, we note that

$$\|\mathbb{E}(X_{n\wedge T} - X_T \mid \mathcal{F}_S)\|_1 \leq \|X_{n\wedge T} - X_T\|_1 \to 0 \text{ as } n \to \infty.$$

So it follows that $\mathbb{E}(X_{n\wedge T} \mid \mathcal{F}_S) \to \mathbb{E}(X_T \mid \mathcal{F}_S)$ as $n \to \infty$. □

## 2.4 Applications of martingales

**Theorem.** Let $Y \in L^1$, and let $\hat{\mathcal{F}}_n$ be a backwards filtration. Then

$$\mathbb{E}(Y \mid \hat{\mathcal{F}}_n) \to \mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)$$

almost surely and in $L^1$.

*Proof.* We first show that $\mathbb{E}(Y \mid \hat{F}_n)$ converges. We then show that what it converges to is indeed $\mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)$.

We write

$$X_n = \mathbb{E}(Y \mid \hat{\mathcal{F}}_n).$$

Observe that for all $n \geq 0$, the process $(X_{n-k})_{0\leq k\leq n}$ is a martingale by the tower property, and so is $(-X_{n-k})_{0\leq k\leq n}$. Now notice that for all $a < b$, the number of upcrossings of $[a, b]$ by $(X_k)_{0\leq k\leq n}$ is equal to the number of upcrossings of $[-b, -a]$ by $(-X_{n-k})_{0\leq k\leq n}$.

Using the same arguments as for martingales, we conclude that $X_n \to X_\infty$ almost surely and in $L^1$ for some $X_\infty$.

To see that $X_\infty = \mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)$, we notice that $X_\infty$ is $\hat{F}_\infty$ measurable. So it is enough to prove that

$$\mathbb{E}X_\infty \mathbf{1}_A = \mathbb{E}(\mathbb{E}(Y \mid \hat{\mathcal{F}}_\infty)\mathbf{1}_A)$$

for all $A \in \hat{\mathcal{F}}_\infty$. Indeed, we have

$$\begin{aligned} \mathbb{E}X_\infty \mathbf{1}_A &= \lim_{n\to\infty} \mathbb{E}X_n \mathbf{1}_A \\ &= \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Y \mid \hat{\mathcal{F}}_n)\mathbf{1}_A) \\ &= \lim_{n\to\infty} \mathbb{E}(Y \mid \mathbf{1}_A) \\ &= \mathbb{E}(Y \mid \mathbf{1}_A) \\ &= \mathbb{E}(\mathbb{E}(Y \mid \hat{\mathcal{F}}_n)\mathbf{1}_A). \end{aligned}$$ □

**Theorem** (Kolmogorov 0-1 law)**.** Let $(X_n)_{n\geq 0}$ be independent random variables. Then, let
$$\hat{\mathcal{F}}_n = \sigma(X_{n+1}, X_{n+2}, \ldots).$$
Then the *tail $\sigma$-algebra* $\hat{\mathcal{F}}_\infty$ is trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \hat{\mathcal{F}}_\infty$.

*Proof.* Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n$ and $\hat{\mathcal{F}}_n$ are independent. Then for all $A \in \hat{\mathcal{F}}_\infty$, we have
$$\mathbb{E}(\mathbf{1}_A \mid \mathcal{F}_n) = \mathbb{P}(A).$$
But the LHS is a martingale. So it converges almost surely and in $L^1$ to $\mathbb{E}(\mathbf{1}_A \mid \mathcal{F}_\infty)$. But $\mathbf{1}_A$ is $\mathcal{F}_\infty$-measurable, since $\hat{\mathcal{F}}_\infty \subseteq \mathcal{F}_\infty$. So this is just $\mathbf{1}_A$. So $\mathbf{1}_A = \mathbb{P}(A)$ almost surely, and we are done. □

**Theorem** (Strong law of large numbers)**.** Let $(X_n)_{n\geq 1}$ be iid random variables in $L^1$, with $\mathbb{E}X_1 = \mu$. Define
$$S_n = \sum_{i=1}^n X_i.$$
Then
$$\frac{S_n}{n} \to \mu \text{ as } n \to \infty$$
almost surely and in $L^1$.

*Proof.* We have
$$S_n = \mathbb{E}(S_n \mid S_n) = \sum_{i=1}^n \mathbb{E}(X_i \mid S_n) = n\mathbb{E}(X_1 \mid S_n).$$
So the problem is equivalent to showing that $\mathbb{E}(X_1 \mid S_n) \to \mu$ as $n \to \infty$. This seems like something we can tackle with our existing technology, except that the $S_n$ do not form a filtration.

Thus, define a backwards filtration
$$\hat{\mathcal{F}}_n = \sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots) = \sigma(S_n, \tau_n),$$
where $\tau_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. We now use the property of conditional expectation that we've never used so far, that adding independent information to a conditional expectation doesn't change the result. Since $\tau_n$ is independent of $\sigma(X_1, S_n)$, we know
$$\frac{S_n}{n} = \mathbb{E}(X_1 \mid S_n) = \mathbb{E}(X_1 \mid \hat{\mathcal{F}}_n).$$
Thus, by backwards martingale convergence, we know
$$\frac{S_n}{n} \to \mathbb{E}(X_1 \mid \hat{\mathcal{F}}_\infty).$$
But by the Kolmogorov 0-1 law, we know $\hat{\mathcal{F}}_\infty$ is trivial. So we know that $\mathbb{E}(X_1 \mid \hat{\mathcal{F}}_\infty)$ is almost constant, which has to be $\mathbb{E}(\mathbb{E}(X_1 \mid \hat{\mathcal{F}}_\infty)) = \mathbb{E}(X_1) = \mu$. □

**Theorem** (Radon–Nikodym)**.** Let $(\Omega, \mathcal{F})$ be a measurable space, and $\mathbb{Q}$ and $\mathbb{P}$ be two probability measures on $(\Omega, \mathcal{F})$. Then the following are equivalent:

(i) $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$, i.e. for any $A \in \mathcal{F}$, if $\mathbb{P}(A) = 0$, then $\mathbb{Q}(A) = 0$.

(ii) For any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$, if $\mathbb{P}(A) \leq \delta$, then $\mathbb{Q}(A) \leq \varepsilon$.

(iii) There exists a random variable $X \geq 0$ such that

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}(X\mathbf{1}_A).$$

In this case, $X$ is called the *Radon–Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$, and we write $X = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$.

*Proof.* We shall only treat the case where $\mathcal{F}$ is *countably generated*, i.e. $\mathcal{F} = \sigma(F_1, F_2, \ldots)$ for some sets $F_i$. For example, any second-countable topological space is countably generated.

– (iii) $\Rightarrow$ (i): Clear.

– (ii) $\Rightarrow$ (iii): Define the filtration

$$\mathcal{F}_n = \sigma(F_1, F_2, \ldots, F_n).$$

Since $\mathcal{F}_n$ is finite, we can write it as

$$\mathcal{F}_n = \sigma(A_{n,1}, \ldots, A_{n,m_n}),$$

where each $A_{n,i}$ is an *atom*, i.e. if $B \subsetneq A_{n,i}$ and $B \in \mathcal{F}_n$, then $B = \emptyset$. We define

$$X_n = \sum_{n=1}^{m_n} \frac{\mathbb{Q}(A_{n,i})}{\mathbb{P}(A_{n,i})} \mathbf{1}_{A_{n,i}},$$

where we skip over the terms where $\mathbb{P}(A_{n,i}) = 0$. Note that this is exactly designed so that for any $A \in \mathcal{F}_n$, we have

$$\mathbb{E}_{\mathbb{P}}(X_n \mathbf{1}_A) = \mathbb{E}_{\mathbb{P}} \sum_{A_{n,i} \subseteq A} \frac{\mathbb{Q}(A_{n,i})}{\mathbb{P}(A_n, i)} \mathbf{1}_{A_{n,i}} = \mathbb{Q}(A).$$

Thus, if $A \in \mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, we have

$$\mathbb{E} X_{n+1} \mathbf{1}_A = \mathbb{Q}(A) = \mathbb{E} X_n \mathbf{1}_A.$$

So we know that

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n.$$

It is also immediate that $(X_n)_{n \geq 0}$ is adapted. So it is a martingale.

We next show that $(X_n)_{n \geq 0}$ is uniformly integrable. By Markov's inequality, we have

$$\mathbb{P}(X_n \geq \lambda) \leq \frac{\mathbb{E} X_n}{\lambda} = \frac{1}{\lambda} \leq \delta$$

for $\lambda$ large enough. Then

$$\mathbb{E}(X_n \mathbf{1}_{X_n \geq \lambda}) = \mathbb{Q}(X_n \geq \lambda) \leq \varepsilon.$$

So we have shown uniform integrability, and so we know $X_n \to X$ almost surely and in $L^1$ for some $X$. Then for all $A \in \bigcup_{n \geq 0} \mathcal{F}_n$, we have

$$\mathbb{Q}(A) = \lim_{n \to \infty} \mathbb{E} X_n \mathbf{1}_A = \mathbb{E} X \mathbf{1}_A.$$

So $\mathbb{Q}(-)$ and $\mathbb{E} X \mathbf{1}_{(-)}$ agree on $\bigcup_{n \geq 0} \mathcal{F}_n$, which is a generating $\pi$-system for $\mathcal{F}$, so they must be the same.

– (i) $\Rightarrow$ (ii): Suppose not. Then there exists some $\varepsilon > 0$ and some $A_1, A_2, \ldots \in \mathcal{F}$ such that

$$\mathbb{Q}(A_n) \geq \varepsilon, \quad \mathbb{P}(A_n) \leq \frac{1}{2^n}.$$

Since $\sum_n \mathbb{P}(A_n)$ is finite, by Borel–Cantelli, we know

$$\mathbb{P} \limsup A_n = 0.$$

On the other hand, by, say, dominated convergence, we have

$$\begin{aligned} \mathbb{Q} \limsup A_n &= \mathbb{Q}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) \\ &= \lim_{k \to \infty} \mathbb{Q}\left(\bigcap_{n=1}^{k} \bigcup_{m=n}^{\infty} A_m\right) \\ &\geq \lim_{k \to \infty} \mathbb{Q}\left(\bigcup_{m=k}^{\infty} A_k\right) \\ &\geq \varepsilon. \end{aligned}$$

This is a contradiction. □

**Proposition.** If $F$ is harmonic and bounded, and $(X_n)_{n \geq 0}$ is Markov, then $(f(X_n))_{n \geq 0}$ is a martingale.

# 3 Continuous time stochastic processes

**Theorem** (Kolmogorov's criterion)**.** Let $(\rho_t)_{t\in I}$ be random variables, where $I \subseteq [0,1]$ is dense. Assume that for some $p > 1$ and $\beta > \frac{1}{p}$, we have

$$\|\rho_t - \rho_s\|_p \leq C|t-s|^\beta \text{ for all } t, s \in I. \tag{$*$}$$

Then there exists a continuous process $(X_t)_{t\in I}$ such that for all $t \in I$,

$$X_t = \rho_t \text{ almost surely,}$$

and moreover for any $\alpha \in [0, \beta - \frac{1}{p})$, there exists a random variable $K_\alpha \in L^p$ such that

$$|X_s - X_t| \leq K_\alpha |s-t|^\alpha$$

for all $s, t \in [0,1]$.

*Proof.* First note that we may assume $D \subseteq I$. Indeed, for $t \in D$, we can define $\rho_t$ by taking the limit of $\rho_s$ in $L^p$ since $L^p$ is complete. The equation $(*)$ is preserved by limits, so we may work on $I \cup D$ instead.

By assumption, $(\rho_t)_{t\in I}$ is Hölder in $L^p$. We claim that it is almost surely pointwise Hölder.

**Claim.** There exists a random variable $K_\alpha \in L^p$ such that

$$|\rho_s - \rho_t| \leq K_\alpha |s-t|^\alpha \text{ for all } s, t \in D.$$

Moreover, $K_\alpha$ is increasing in $\alpha$.

Given the claim, we can simply set

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

Then this is a continuous process, and satisfies the desired properties.

To construct such a $K_\alpha$, observe that given any $s, t \in D$, we can pick $m \geq 0$ such that

$$2^{-(m+1)} < t - s \leq 2^{-m}.$$

Then we can pick $u = \frac{k}{2^{m+1}}$ such that $s < u < t$. Thus, we have

$$u - s < 2^{-m}, \quad t - u < 2^{-m}.$$

Therefore, by binary expansion, we can write

$$u - s = \sum_{i \geq m+1} \frac{x_i}{2^i}, \quad t - u = \sum_{i \geq m+1} \frac{y_i}{2^i},$$

for some $x_i, y_i \in \{0,1\}$. Thus, writing

$$K_n = \sup_{t \in D_n} |S_{t+2^{-n}} - S_t|,$$

we can bound

$$|\rho_s - \rho_t| \leq 2 \sum_{n=m+1}^\infty K_n,$$

and thus

$$\frac{|\rho_s - \rho_t|}{|s-t|^\alpha} \leq 2\sum_{n=m+1}^\infty 2^{(m+1)\alpha} K_n \leq 2\sum_{n=m+1}^\infty 2^{(n+1)\alpha} K_n.$$

Thus, we can define

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

We only have to check that this is in $L^p$, and this is not hard. We first get

$$\mathbb{E} K_n^p \leq \sum_{t \in D_n} \mathbb{E}|\rho_{t+2^{-n}} - \rho_t|^p \leq C^p 2^n \cdot 2^{-n\beta} = C 2^{n(1-p\beta)}.$$

Then we have

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

**Proposition.** Let $(X_t)_{t \geq 0}$ be a cadlag adapted process and $S, T$ stopping times. Then

(i) $S \wedge T$ is a stopping time.

(ii) If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

(iii) $X_T \mathbf{1}_{T < \infty}$ is $\mathcal{F}_T$-measurable.

(iv) $(X_t^T)_{t \geq 0} = (X_{T \wedge t})_{t \geq 0}$ is adapted.

**Lemma.** A random variable $Z$ is $\mathcal{F}_T$-measurable iff $Z\mathbf{1}_{\{T \leq t\}}$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.

*Proof of (iii) of proposition.* We need to prove that $X_T \mathbf{1}_{\{T \leq t\}}$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.

We write

$$X_T \mathbf{1}_{T \leq t} = X_T \mathbf{1}_{T < t} + X_t \mathbf{1}_{T = t}.$$

We know the second term is measurable. So it suffices to show that $X_T \mathbf{1}_{T < t}$ is $\mathcal{F}_t$-measurable.

Define $T_n = 2^{-n}\lceil 2^n T\rceil$. This is a stopping time, since we always have $T_n \geq T$. Since $(X_t)_{t \geq 0}$ is cadlag, we know

$$X_T \mathbf{1}_{T < t} = \lim_{n \to \infty} X_{T_n \wedge t} \mathbf{1}_{T < t}.$$

Now $T_n \wedge t$ can take only countably (and in fact only finitely) many values, so we can write

$$X_{T_n \wedge t} = \sum_{q \in D_n, q < t} X_q \mathbf{1}_{T_n = q} + X_t \mathbf{1}_{T < t < T_n},$$

and this is $\mathcal{F}_t$-measurable. So we are done. □

**Proposition.** Let $A \subseteq \mathbb{R}$ be a closed set and $(X_t)_{t \geq 0}$ be continuous. Then $T_A$ is a stopping time.

*Proof.* Observe that $d(X_q, A)$ is a continuous function in $q$. So we have

$$\{T_A \le t\} = \left\{\inf_{q \in \mathbb{Q}, q < t} d(X_q, A) = 0\right\}. \qquad \square$$

**Proposition.** Let $(X_t)_{t \ge 0}$ be an adapted process (to $(\mathcal{F}_t)_{t \ge 0}$) that is cadlag, and let $A$ be an open set. Then $T_A$ is a stopping time with respect to $\mathcal{F}_t^+$.

*Proof.* Since $(X_t)_{t \ge 0}$ is cadlag and $A$ is open. Then

$$\{T_A < t\} = \bigcup_{q < t, q \in \mathbb{Q}} \{X_q \in A\} \in \mathcal{F}_t.$$

Then

$$\{T_A \le t\} = \bigcap_{n \ge 0} \left\{T_A < t + \frac{1}{n}\right\} \in \mathcal{F}_t^+. \qquad \square$$

**Theorem** (Optional stopping theorem)**.** Let $(X_t)_{t \ge 0}$ be an adapted cadlag process in $L^1$. Then the following are equivalent:

(i) For any bounded stopping time $T$ and any stopping time $S$, we have $X_T \in L^1$ and
$$\mathbb{E}(X_T \mid \mathcal{F}_S) = X_{T \wedge S}.$$

(ii) For any stopping time $T$, $(X_t^T)_{t \ge 0} = (X_{T \wedge t})_{t \ge 0}$ is a martingale.

(iii) For any bounded stopping time $T$, $X_T \in L^1$ and $\mathbb{E}X_T = \mathbb{E}X_0$.

*Proof.* We show that (i) $\Rightarrow$ (ii), and the rest follows from the discrete case similarly.

Since $T$ is bounded, assume $T \le t$, and we may wlog assume $t \in \mathbb{N}$. Let

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

We have $T_n \searrow T$ as $n \to \infty$, and so $X_{T_n} \to X_T$ as $n \to \infty$.

Since $T_n \le t + 1$, by restricting our sequence to $D_n$, discrete time optional stopping implies
$$\mathbb{E}(X_{t+1} \mid \mathcal{F}_{T_n}) = X_{T_n}.$$
In particular, $X_{T_n}$ is uniformly integrable. So it converges in $L^1$. This implies $X_T \in L^1$.

To show that $\mathbb{E}(X_t \mid \mathcal{F}_S) = X_{T \wedge S}$, we need to show that for any $A \in \mathcal{F}_S$, we have
$$\mathbb{E}X_t \mathbf{1}_A = \mathbb{E}X_{S \wedge T}\mathbf{1}_A.$$
Since $\mathcal{F}_S \subseteq \mathcal{F}_{S_n}$, we already know that

$$\mathbb{E}X_{T_n}\mathbf{1}_A = \lim_{n \to \infty} \mathbb{E}X_{S_n \wedge T_n}\mathbf{1}_A$$

by discrete time optional stopping, since $\mathbb{E}(X_{T_n} \mid \mathcal{F}_{S_n}) = X_{T_n \wedge S_n}$. So taking the limit $n \to \infty$ gives the desired result. $\square$

**Theorem.** Let $(X_t)_{t \ge 0}$ be a super-martingale bounded in $L^1$. Then it converges almost surely as $t \to \infty$ to a random variable $X_\infty \in L^1$.

*Proof.* Define $U_s[a,b,(x_t)_{t\geq 0}]$ be the number of upcrossings of $[a,b]$ by $(x_t)_{t\geq 0}$ up to time $s$, and

$$U_\infty[a,b,(x_t)_{t\geq 0}] = \lim_{s\to\infty} U_s[a,b,(x_t)_{t\geq 0}].$$

Then for all $s \geq 0$, we have

$$U_s[a,b,(x_t)_{t\geq 0}] = \lim_{n\to\infty} U_s[a,b,(x_t^{(n)})_{t\in D_n}].$$

By monotone convergence and Doob's upcrossing lemma, we have

$$\mathbb{E}U_s[a,b,(X_t)_{t\geq 0}] = \lim_{n\to\infty} \mathbb{E}U_s[a,b,(X_t)_{t\in D_n}] \leq \frac{\mathbb{E}(X_s-a)^-}{b-1} \leq \frac{\mathbb{E}|X_s|+a}{b-a}.$$

We are then done by taking the supremum over $s$. Then finish the argument as in the discrete case.

This shows we have pointwise convergence in $\mathbb{R}\cup\{\pm\infty\}$, and by Fatou's lemma, we know that

$$\mathbb{E}|X_\infty| = \mathbb{E}\liminf_{t_n\to\infty}|X_{t_n}| \leq \liminf_{t_n\to\infty}\mathbb{E}|X_{t_n}| < \infty.$$

So $X_\infty$ is finite almost surely. □

**Lemma** (Maximal inequality)**.** Let $(X_t)_{t\geq 0}$ be a cadlag martingale or a non-negative sub-martingale. Then for all $t\geq 0$, $\lambda \geq 0$, we have

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

**Lemma** (Doob's $L^p$ inequality)**.** Let $(X_t)_{t\geq 0}$ be as above. Then

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

**Theorem** (Regularization of martingales)**.** Let $(X_t)_{t\geq 0}$ be a martingale with respect to $(\mathcal{F}_t)$, and suppose $\mathcal{F}_t$ satisfies the usual conditions. Then there exists a version $(\tilde{X}_t)$ of $(X_t)$ which is cadlag.

*Proof.* For all $M > 0$, define

$$\Omega_0^M = \left\{\sup_{q\in D\cap[0,M]}|X_q| < \infty\right\} \cap \bigcap_{a<b\in\mathbb{Q}}\left\{U_M[a,b,(X_t)_{t\in D\cap[0,M]}] < \infty\right\}$$

Then we see that $\mathbb{P}(\Omega_0^M) = 1$ by Doob's upcrossing lemma. Now define

$$\tilde{X}_t = \lim_{s\geq t, s\to t, s\in D} X_s \mathbf{1}_{\Omega_0^t}.$$

Then this is $\mathcal{F}_t$ measurable because $\mathcal{F}_t$ satisfies the usual conditions.

Take a sequence $t_n \searrow t$. Then $(X_{t_n})$ is a backwards martingale. So it converges almost surely in $L^1$ to $\tilde{X}_t$. But we can write

$$X_t = \mathbb{E}(X_{t_n} \mid \mathcal{F}_t).$$

Since $X_{t_n} \to \tilde{X}_t$ in $L^1$, and $\tilde{X}_t$ is $\mathcal{F}_t$-measurable, we know $X_t = \tilde{X}_t$ almost surely.

The fact that it is cadlag is an exercise. □

**Theorem** ($L^p$ convergence of martingales)**.** Let $(X_t)_{t\geq 0}$ be a cadlag martingale. Then the following are equivalent:

(i) $(X_t)_{t\geq 0}$ is bounded in $L^p$.

(ii) $(X_t)_{t\geq 0}$ converges almost surely and in $L^p$.

(iii) There exists $Z \in L^p$ such that $X_t = \mathbb{E}(Z \mid \mathcal{F}_t)$ almost surely.

**Theorem** ($L^1$ convergence of martingales)**.** Let $(X_t)_{t\geq 0}$ be a cadlag martingale. Then the folloiwng are equivalent:

(i) $(X_t)_{t\geq 0}$ is uniformly integrable.

(ii) $(X_t)_{t\geq 0}$ converges almost surely and in $L^1$ to $X_\infty$.

(iii) There exists $Z \in L^1$ such that $\mathbb{E}(Z \mid \mathcal{F}_t) = X_t$ almost surely.

**Theorem** (Optional stopping theorem)**.** Let $(X_t)_{t\geq 0}$ be a uniformly integrable martingale, and let $S, T$ b e any stopping times. Then

$$\mathbb{E}(X_T \mid \mathcal{F}_s) = X_{S\wedge T}.$$

# 4 Weak convergence of measures

**Proposition.** Let $(\mu_n)_{n\geq 0}$ be as above. Then, the following are equivalent:

(i) $(\mu_n)_{n\geq 0}$ converges weakly to $\mu$.

(ii) For all open $G$, we have

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G).$$

(iii) For all closed $A$, we have

$$\limsup_{n\to\infty} \mu_n(A) \leq \mu(A).$$

(iv) For all $A$ such that $\mu(\partial A) = 0$, we have

$$\lim_{n\to\infty} \mu_n(A) = \mu(A)$$

(v) (when $M = \mathbb{R}$) $F_{\mu_n}(x) \to F_\mu(x)$ for all $x$ at which $F_\mu$ is continuous, where $F_\mu$ is the *distribution function* of $\mu$, defined by $F_\mu(x) = \mu_n((-\infty, t])$.

*Proof.*

- (i) $\Rightarrow$ (ii): The idea is to approximate the open set by continuous functions. We know $A^c$ is closed. So we can define

  $$f_N(x) = 1 \wedge (N \cdot \operatorname{dist}(x, A^c)).$$

  This has the property that for all $N > 0$, we have

  $$f_N \leq \mathbf{1}_A,$$

  and moreover $f_N \nearrow \mathbf{1}_A$ as $N \to \infty$. Now by definition of weak convergence,

  $$\liminf_{n\to\infty} \mu(A) \geq \liminf_{n\to\infty} \mu_n(f_N) = \mu(F_N) \to \mu(A) \text{ as } N \to \infty.$$

- (ii) $\Leftrightarrow$ (iii): Take complements.
- (iii) and (ii) $\Rightarrow$ (iv): Take $A$ such that $\mu(\partial A) = 0$. Then

  $$\mu(A) = \mu(\mathring{A}) = \mu(\bar{A}).$$

  So we know that

  $$\liminf_{n\to\infty} \mu_n(A) \geq \liminf_{n\to\infty} \mu_n(\mathring{A}) \geq \mu(\mathring{A}) = \mu(A).$$

  Similarly, we find that

  $$\mu(A) \geq \limsup_{n\to\infty} \mu_n(A).$$

  So we are done.

– (iv) ⇒ (i): We have

$$\begin{aligned}\mu(f) &= \int_M f(x)\,\mathrm{d}\mu(x)\\ &= \int_M \int_0^\infty \mathbf{1}_{f(x)\geq t}\,\mathrm{d}t\,\mathrm{d}\mu(x)\\ &= \int_0^\infty \mu(\{f \geq t\})\,\mathrm{d}t.\end{aligned}$$

Since $f$ is continuous, $\partial\{f \leq t\} \subseteq \{f = t\}$. Now there can be only countably many $t$'s such that $\mu(\{f = t\}) > 0$. So replacing $\mu$ by $\lim_{n\to\infty} \mu_n$ only changes the integrand at countably many places, hence doesn't affect the integral. So we conclude using bounded convergence theorem.

– (iv) ⇒ (v): Assume $t$ is a continuity point of $F_\mu$. Then we have

$$\mu(\partial(-\infty, t]) = \mu(\{t\}) = F_\mu(t) - F_\mu(t_-) = 0.$$

So $\mu_n(\partial_n(-\infty, t]) \to \mu((-\infty, t])$, and we are done.

– (v) ⇒ (ii): If $A = (a, b)$, then

$$\mu_n(A) \geq F_{\mu_n}(b') - F_{\mu_n}(a')$$

for any $a \leq a' \leq b' \leq b$ with $a', b'$ continuity points of $F_\mu$. So we know that

$$\liminf_{n\to\infty} \mu_n(A) \geq F_\mu(b') - F_\mu(a') = \mu(a', b').$$

By taking supremum over all such $a', b'$, we find that

$$\liminf_{n\to\infty} \mu_n(A) \geq \mu(A).$$ □

**Theorem** (Prokhorov's theorem)**.** If $(\mu_n)_{n\geq 0}$ is a sequence of tight probability measures, then there is a subsequence $(\mu_{n_k})_{k\geq 0}$ and a measure $\mu$ such that $\mu_{n_k} \Rightarrow \mu$.

*Proof.* Take $\mathbb{Q} \subseteq \mathbb{R}$, which is dense and countable. Let $x_1, x_2, \ldots$ be an enumeration of $\mathbb{Q}$. Define $F_n = F_{\mu_n}$. By Bolzano–Weierstrass, and some fiddling around with sequences, we can find some $F_{n_k}$ such that

$$F_{n_k}(x_i) \to y_i \equiv F(x_i)$$

as $k \to \infty$, for each fixed $x_i$.

Since $F$ is non-decreasing on $\mathbb{Q}$, it has left and right limits everywhere. We extend $F$ to $\mathbb{R}$ by taking right limits. This implies $F$ is cadlag.

Take $x$ a continuity point of $F$. Then for each $\varepsilon > 0$, there exists $s < x < t$ rational such that

$$|F(s) - F(t)| < \frac{\varepsilon}{2}.$$

Take $n$ large enough such that $|F_n(s) - F(s)| < \frac{\varepsilon}{4}$, and same for $t$. Then by monotonicity of $F$ and $F_n$, we have

$$|F_n(x) - F(x)| \leq |F(s) - F(t)| + |F_n(s) - F(s)| + |F_n(t) - F(t)| \leq \varepsilon.$$

It remains to show that $F(x) \to 1$ as $x \to \infty$ and $F(x) \to 0$ as $x \to -\infty$. By tightness, for all $\varepsilon > 0$, there exists $N > 0$ such that

$$\mu_n((-\infty, N]) \leq \varepsilon, \quad \mu_n((N, \infty) \leq \varepsilon.$$

This then implies what we want. $\square$

**Proposition.** If $\varphi_X = \varphi_Y$, then $\mu_X = \mu_Y$.

**Theorem** (Lévy's convergence theroem)**.** Let $(X_n)_{n\geq 0}$, $X$ be random variables taking values in $\mathbb{R}^d$. Then the following are equivalent:

(i) $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$.

(ii) $\varphi_{X_n} \to \varphi_X$ pointwise.

**Theorem** (Lévy)**.** Let $(X_n)_{n\geq 0}$ be as above, and let $\varphi_{X_n}(t) \to \psi(t)$ for all $t$. Suppose $\psi$ is continuous at 0 and $\psi(0) = 1$. Then there exists a random variable $X$ such that $\varphi_X = \psi$ and $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$.

**Lemma.** Let $X$ be a real random variable. Then for all $\lambda > 0$,

$$\mu_X(|x| \geq \lambda) \leq c\lambda \int_0^{1/\lambda} (1 - \operatorname{Re} \varphi_X(t))\,\mathrm{d}t,$$

where $C = (1 - \sin 1)^{-1}$.

*Proof.* For $M \geq 1$, we have

$$\int_0^M (1 - \cos t)\,\mathrm{d}t = M - \sin M \geq M(1 - \sin 1).$$

By setting $M = \frac{|x|}{\lambda}$, we have

$$\mathbf{1}_{|X|\geq \lambda} \leq C\frac{\lambda}{|X|} \int_0^{|X|/\lambda} (1 - \cos t)\,\mathrm{d}t.$$

By a change of variables with $t \mapsto Xt$, we have

$$\mathbf{1}_{|X|\geq\lambda} \leq c\lambda \int_0^1 (1 - \cos Xt)\,\mathrm{d}t.$$

Apply $\mu_X$, and use the fact that $\operatorname{Re}\varphi_X(t) = \mathbb{E}\cos(Xt)$. $\square$

*Proof of theorem.* It is clear that weak convergence implies convergence in characteristic functions.

Now observe that if $\mu_n \Rightarrow \mu$ iff from every subsequence $(n_k)_{k\geq 0}$, we can choose a further subsequence $(n_{k_\ell})$ such that $\mu_{n_{k_\ell}} \Rightarrow \mu$ as $\ell \to \infty$. Indeed, $\Rightarrow$ is clear, and suppose $\mu_n \not\Rightarrow \mu$ but satisfies the subsequence property. Then we can choose a bounded and continuous function $f$ such that

$$\mu_n(f) \not\Rightarrow \mu(f).$$

Then there is a subsequence $(n_k)_{k\geq 0}$ such that $|\mu_{n_k}(f) - \mu(f)| > \varepsilon$. Then there is no further subsequence that converges.

Thus, to show $\Leftarrow$, we need to prove the existence of subsequential limits (uniqueness follows from convergence of characteristic functions). It is enough to prove tightness of the whole sequence.

By the mean value theorem, we can choose $\lambda$ so large that

$$c\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\psi(t))\; \mathrm{d}t < \frac{\varepsilon}{2}.$$

By bounded convergence, we can choose $\lambda$ so large that

$$c\lambda \int_0^{1/\lambda} (1 - \operatorname{Re}\varphi_{X_n}(t))\; \mathrm{d}t \leq \varepsilon$$

for all $n$. Thus, by our previous lemma, we know $(\mu_{X_n})_{n \geq 0}$ is tight. So we are done. $\square$

# 5 Brownian motion

## 5.1 Basic properties of Brownian motion

**Theorem** (Wiener's theorem)**.** There exists a Brownian motion on some probability space.

*Proof.* We first prove existence on $[0,1]$ and in $d = 1$. We wish to apply Kolmogorov's criterion.

Recall that $D_n$ are the dyadic numbers. Let $(Z_d)_{d\in D}$ be iid $N(0,1)$ random variables on some probability space. We will define a process on $D_n$ inductively on $n$ with the required properties. We wlog assume $x = 0$.

In step 0, we put

$$B_0 = 0, \quad B_1 = Z_1.$$

Assume that we have already constructed $(B_d)_{d\in D_{n-1}}$ satisfying the properties. Take $d \in D_n \setminus D_{n-1}$, and set

$$d^\pm = d \pm 2^{-n}.$$

These are the two consecutive numbers in $D_{n-1}$ such that $d_- < d < d_+$. Define

$$B_d = \frac{B_{d_+} + B_{d_-}}{2} + \frac{1}{2^{(n+1)/2}} Z_d.$$

The condition (i) is trivially satisfied. We now have to check the other two conditions.

Consider

$$\begin{aligned} B_{d_+} - B_d &= \frac{B_{d_+} - B_{d_-}}{2} - \frac{1}{2^{(n+1)/2}} Z_d \\ B_d - B_{d_-} &= \underbrace{\frac{B_{d_+} - B_{d_-}}{2}}_{N} + \underbrace{\frac{1}{2^{(n+1)/2}} Z_d}_{N'}. \end{aligned}$$

Notice that $N$ and $N'$ are normal with variance $\mathrm{var}(N') = \mathrm{var}(N) = \frac{1}{2^{n+1}}$. In particular, we have

$$\mathrm{cov}(N - N', N + N') = \mathrm{var}(N) - \mathrm{var}(N') = 0.$$

So $B_{d_+} - B_d$ and $B_d - B_{d_-}$ are independent.

Now note that the vector of increments of $(B_d)_{d\in D_n}$ between consecutive numbers in $D_n$ is Gaussian, since after dotting with any vector, we obtain a linear combination of independent Gaussians. Thus, to prove independence, it suffice to prove that pairwise correlation vanishes.

We already proved this for the case of increments between $B_d$ and $B_{d_\pm}$, and this is the only case that is tricky, since they both involve the same $Z_d$. The other cases are straightforward, and are left as an exercise for the reader.

Inductively, we can construct $(B_d)_{d\in D}$, satisfying (i), (ii) and (iii). Note that for all $s, t \in D$, we have

$$\mathbb{E}|B_t - B_s|^p = |t - s|^{p/2} \mathbb{E}|N|^p$$

for $N \sim N(0,1)$. Since $\mathbb{E}|N|^p < \infty$ for all $p$, by Kolmogorov's criterion, we can extend $(B_d)_{d\in D}$ to $(B_t)_{t\in[0,1]}$. In fact, this is $\alpha$-Hölder continuous for all $\alpha < \frac{1}{2}$.

Since this is a continuous process and satisfies the desired properties on a dense set, it remains to show that the properties are preserved by taking continuous limits.

Take $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$, and $0 \leq t_1^n < t_2^n < \cdots < t_m^n \leq 1$ such that $t_i^n \in D_n$ and $t_i^n \to t_i$ as $n \to \infty$ and $i = 1, \ldots m$.

We now apply Lévy's convergence theorem. Recall that if $X$ is a random variable in $\mathbb{R}^d$ and $X \sim N(0, \Sigma)$, then

$$\varphi_X(u) = \exp\left(-\frac{1}{2}u^T \Sigma u\right).$$

Since $(B_t)_{t \in [0,1]}$ is continuous, we have

$$\begin{aligned}\varphi_{(B_{t_2^n} - B_{t_1^n}, \ldots, B_{t_m^n} - B_{t_{m-1}}^n)}(u) &= \exp\left(-\frac{1}{2}u^T \Sigma u\right)\\ &= \exp\left(-\frac{1}{2}\sum_{i=1}^{m-1}(t_{i+1}^n - t_i^n)u_i^2\right).\end{aligned}$$

We know this converges, as $n \to \infty$, to $\exp\left(-\frac{1}{2}\sum_{i=1}^{m-1}(t_{i+1} - t_i)u_i^2\right)$.

By Lévy's convergence theorem, the law of $(B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_n} - B_{t_{m-1}})$ is Gaussian with the right covariance. This implies that (ii) and (iii) hold on $[0, 1]$.

To extend the time to $[0, \infty)$, we define independent Brownian motions $(B_t^i)_{t \in [0,1], i \in \mathbb{N}}$ and define

$$B_t = \sum_{i=0}^{\lfloor t \rfloor - 1} B_1^i + B_{t - \lfloor t \rfloor}^{\lfloor t \rfloor}$$

To extend to $\mathbb{R}^d$, take the product of $d$ many independent one-dimensional Brownian motions. □

**Lemma.** Brownian motion is a Gaussian process, i.e. for any $0 \leq t_1 < t_2 < \cdots < t_m \leq 1$, the vector $(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ is Gaussian with covariance

$$\operatorname{cov}(B_{t_1}, B_{t_2}) = t_1 \wedge t_2.$$

*Proof.* We know $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}})$ is Gaussian. Thus, the sequence $(B_{t_1}, \ldots, B_{t_m})$ is the image under a linear isomorphism, so it is Gaussian. To compute covariance, for $s \leq t$, we have

$$\operatorname{cov}(B_s, B_t) = \mathbb{E}B_s B_t = \mathbb{E}B_s B_T - \mathbb{E}B_s^2 + \mathbb{E}B_s^2 = \mathbb{E}B_s(B_t - B_s) + \mathbb{E}B_s^2 = s. \ \square$$

**Proposition** (Invariance properties)**.** Let $(B_t)_{t \geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$.

(i) If $U$ is an orthogonal matrix, then $(UB_t)_{t \geq 0}$ is a standard Brownian motion.

(ii) *Brownian scaling*: If $a > 0$, then $(a^{-1/2}B_{at})_{t \geq 0}$ is a standard Brownian motion. This is known as a *random fractal property*.

(iii) (*Simple*) *Markov property*: For all $s \geq 0$, the sequence $(B_{t+s} - B_s)_{t \geq 0}$ is a standard Brownian motion, independent of $(\mathcal{F}_s^B)$.

(iv) *Time inversion*: Define a process

$$X_t = \begin{cases} 0 & t = 0 \\ tB_{1/t} & t > 0 \end{cases}.$$

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

*Proof.* Only (iv) requires proof. It is enough to prove that $X_t$ is continuous and has the right finite-dimensional distributions. We haves

$$(X_{t_1}, \ldots, X_{t_m}) = (t_1 B_{1/t_1}, \ldots, t_m B_{1/t_m}).$$

The right-hand side is the image of $(B_{1/t_1}, \ldots, B_{1/t_m})$ under a linear isomorphism. So it is Gaussian. If $s \leq t$, then the covariance is

$$\operatorname{cov}(sB_s, tB_t) = st \operatorname{cov}(B_{1/s}, B_{1/t}) = st\left(\frac{1}{s} \wedge \frac{1}{t}\right) = s = s \wedge t.$$

Continuity is obvious for $t > 0$. To prove continuity at 0, we already proved that $(X_q)_{q>0, q\in\mathbb{Q}}$ has the same law (as a process) as Brownian motion. By continuity of $X_t$ for positive $t$, we have

$$\mathbb{P}\left(\lim_{q\in\mathbb{Q}_+, q\to 0} X_q = 0\right) = \mathbb{P}\left(\lim_{q\in\mathbb{Q}_+, q\to 0} B_q = 0\right) = \mathbf{1}_B$$

by continuity of $B$. □

**Theorem.** For all $s \geq t$, the process $(B_{t+s} - B_s)_{t\geq 0}$ is independent of $\mathcal{F}_s^+$.

*Proof.* Take a sequence $s_n \to s$ such that $s_n > s$ for all $n$. By continuity,

$$B_{t+s} - B_s = \lim_{n\to\infty} B_{t+s_n} - B_{s_n}$$

almost surely. Now each of $B_{t+s_n} - B_{s_n}$ is independent of $\mathcal{F}_s^+$, and hence so is the limit. □

**Theorem** (Blumenthal's 0-1 law)**.** The $\sigma$-algebra $\mathcal{F}_0^+$ is trivial, i.e. if $A \in \mathcal{F}_0^+$, then $\mathbb{P}(A) \in \{0, 1\}$.

*Proof.* Apply our previous theorem. Take $A \in \mathcal{F}_0^+$. Then $A \in \sigma(B_s : s \geq 0)$. So $A$ is independent of itself. □

**Proposition.**

(i) If $d = 1$, then

$$\begin{aligned} 1 &= \mathbb{P}(\inf\{t \geq 0 : B_t > 0\} = 0) \\ &= \mathbb{P}(\inf\{t \geq 0 : B_t < 0\} = 0) \\ &= \mathbb{P}(\inf\{t > 0 : B_t = 0\} = 0) \end{aligned}$$

(ii) For any $d \geq 1$, we have

$$\lim_{t\to\infty} \frac{B_t}{t} = 0$$

almost surely.

(iii) If we define
$$S_t = \sup_{0 \leq s \leq t} B_t, \quad I_t = \inf_{0 \leq s \leq t} B_t,$$
then $S_\infty = \infty$ and $I_\infty = -\infty$ almost surely.

(iv) If $A$ is open an $\mathbb{R}^d$, then the cone of $A$ is $C_A = \{tx : x \in A, t > 0\}$. Then $\inf\{t \geq 0 : B_t \in C_A\} = 0$ almost surely.

*Proof.*

(i) It suffices to prove the first equality. Note that the event $\{\inf\{t \geq 0 : B_k > 0\} = 0\}$ is trivial. Moreover, for any finite $t$, the probability that $B_t > 0$ is $\frac{1}{2}$. Then take a sequence $t_n$ such that $t_n \to 0$, and apply Fatou to conclude that the probability is positive.

(ii) Follows from the previous one since $tB_{1/t}$ is a Brownian motion.

(iii) By scale invariance, because $S_\infty = aS_\infty$ for all $a > 0$.

(iv) Same as (i). □

**Theorem** (Strong Markov property)**.** Let $(B_t)_{t \geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$, and let $T$ be an almost-surely finite stopping time with respect to $(\mathcal{F}_t^+)_{t \geq 0}$. Then
$$\tilde{B}_t = B_{T+t} - B_T$$
is a standard Brownian motion with respect to $(\mathcal{F}_{T+t}^+)_{t \geq 0}$ that is independent of $\mathcal{F}_T^+$.

*Proof.* Let $T_n = 2^{-n}\lceil 2^n T \rceil$. We first prove the statement for $T_n$. We let
$$B_t^{(k)} = B_{t+k/2^n} - B_{k/2^n}$$
This is then a standard Browninan motion independent of $\mathcal{F}_{k/2^n}^+$ by the simple Markov property. Let
$$B_*(t) = B_{t+T_n} - B_{T_n}.$$
Let $\mathcal{A}$ be the $\sigma$-algebra on $\mathcal{C} = C([0,\infty), \mathbb{R}^d)$, and $A \in \mathcal{A}$. Let $E \in \mathcal{F}_{T_n}^+$. The claim that $B_*$ is a standard Brownian motion independent of $E$ can be concisely captured in the equality
$$\mathbb{P}(\{B_* \in A\} \cap E) = \mathbb{P}(\{B \in A\})\mathbb{P}(E). \tag{$\dagger$}$$
Taking $E = \Omega$ tells us $B_*$ and $B$ have the same law, and then taking general $E$ tells us $B_*$ is independent of $\mathcal{F}_{T_n}^+$.

It is a straightforward computation to prove ($\dagger$). Indeed, we have
$$\mathbb{P}(\{B_* \in A\} \cap E) = \sum_{k=0}^\infty \mathbb{P}\left(\{B^{(k)} \in A\} \cap E \cap \left\{T_n = \frac{k}{2^n}\right\}\right)$$
Since $E \in \mathcal{F}_{T_n}^+$, we know $E \cap \{T_n = k/2^n\} \in \mathcal{F}_{k/2^n}^+$. So by the simple Markov property, this is equal to
$$= \sum_{k=0}^\infty \mathbb{P}(\{B^{(k)} \in A\})\mathbb{P}\left(E \cap \left\{T_n = \frac{k}{2^n}\right\}\right).$$

But we know $B_k$ is a standard Brownian motion. So this is equal to

$$\begin{aligned}&= \sum_{b=0}^{\infty} \mathbb{P}(\{B \in A\}) \mathbb{P}\left(E \cap \left\{T_n = \frac{k}{2^n}\right\}\right)\\ &= \mathbb{P}(\{B \in A\})\mathbb{P}(E).\end{aligned}$$

So we are done.

Now as $n \to \infty$, the increments of $B_*$ converge almost surely to the increments of $\tilde{B}$, since $B$ is continuous and $T_n \searrow T$ almost surely. But $B_*$ all have the same distribution, and almost sure convergence implies convergence in distribution. So $\tilde{B}$ is a standard Brownian motion. Being independent of $\mathcal{F}_T^+$ is clear. □

**Theorem** (Reflection principle)**.** Let $(B_t)_{T\geq 0}$ and $T$ be as above. Then the *reflected process* $(\tilde{B}_t)_{t\geq 0}$ defined by

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

is a standard Brownian motion.

*Proof.* By the strong Markov property, we know

$$B_t^T = B_{T+t} - B_T$$

and $-B_t^T$ are standard Brownian motions independent of $\mathcal{F}_T^+$. This implies that the pairs of random variables

$$P_1 = ((B_t)_{0\leq t\leq T}, (B_t)_{t\geq 0}^T), \quad P_2 = ((B_t)_{0\leq t\leq T}, (-B_t)_{t\geq 0}^T)$$

taking values in $\mathcal{C} \times \mathcal{C}$ have the same law on $\mathcal{C} \times \mathcal{C}$ with the product $\sigma$-algebra.

Define the concatenation map $\psi_T(X, Y) : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ by

$$\psi_T(X, Y) = X_t \mathbf{1}_{t<T} + (X_T + Y_{t-T})\mathbf{1}_{t\geq T}.$$

Assuming $Y_0 = 0$, the resulting process is continuous.

Notice that $\psi_T$ is a measurable map, which we can prove by approximations of $T$ by discrete stopping times. We then conclude that $\psi_T(P_1)$ has the same law as $\psi_T(P_2)$. □

**Corollary.** Let $(B_t)_{T\geq 0}$ be a standard Brownian motion in $d = 1$. Let $b > 0$ and $a \leq b$. Let

$$S_t = \sup_{0\leq s\leq t} B_t.$$

Then

$$\mathbb{P}(S_t \geq b, B_t \leq a) = \mathbb{P}(B_t \geq 2b - a).$$

*Proof.* Consider the stopping time $T$ given by the first hitting time of $b$. Since $S_\infty = \infty$, we know $T$ is finite almost surely. Let $(\tilde{B}_t)_{t\geq 0}$ be the reflected process. Then

$$\{S_t \geq b, B_t \leq a\} = \{\tilde{B}_t \geq 2b - a\}.$$ □

**Corollary.** The law of $S_t$ is equal to the law of $|B_t|$.

*Proof.* Apply the previous process with $b = a$ to get

$$\begin{aligned}
\mathbb{P}(S_t \geq a) &= \mathbb{P}(S_t \geq a, B_t < a) + \mathbb{P}(S_t \geq a, B_t \geq a)\\
&= \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \geq a)\\
&= \mathbb{P}(B_t \leq a) + \mathbb{P}(B_t \geq a)\\
&= \mathbb{P}(|B_t| \geq a).
\end{aligned}$$

□

**Proposition.** Let $d = 1$ and $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then the following processes are $(\mathcal{F}_t^+)_{t\geq 0}$ martingales:

(i) $(B_t)_{t\geq 0}$

(ii) $(B_t^2 - t)_{t\geq 0}$

(iii) $\left(\exp\left(uB_t - \frac{u^2 t}{2}\right)\right)_{t\geq 0}$ for $u \in \mathbb{R}$.

*Proof.*

(i) Using the fact that $B_t - B_s$ is independent of $\mathcal{F}_s^+$, we know

$$\mathbb{E}(B_t - B_s \mid \mathcal{F}_s^+) = \mathbb{E}(B_t - B_s) = 0.$$

(ii) We have

$$\mathbb{E}(B_t^2 - t \mid \mathcal{F}_s^+) = \mathbb{E}((B_t - B_s)^2 \mid \mathcal{F}_s) - \mathbb{E}(B_s^2 \mid \mathcal{F}_s^+) + 2\mathbb{E}(B_t B_s \mid \mathcal{F}_s^+) - t$$

We know $B_t - B_s$ is independent of $\mathcal{F}_s^+$, and so the first term is equal to $\mathrm{var}(B_t - B_s) = (t - s)$, and we can simply to get

$$\begin{aligned}
&= (t - s) - B_s^2 + 2B_s^2 - t\\
&= B_s^2 - s.
\end{aligned}$$

(iii) Similar. □

## 5.2 Harmonic functions and Brownian motion

**Lemma.** Let $u : D \to \mathbb{R}$ be measurable and locally bounded. Then the following are equivalent:

(i) $u$ is twice continuously differentiable and $\Delta u = 0$.

(ii) For any $x \in D$ and $r > 0$ such that $B(x, r) \subseteq D$, we have

$$u(x) = \frac{1}{\mathrm{Vol}(B(x,r))} \int_{B(x,r)} u(y)\; \mathrm{d}y$$

(iii) For any $x \in D$ and $r > 0$ such that $B(x, r) \subseteq D$, we have

$$u(x) = \frac{1}{\mathrm{Area}(\partial B(x,r))} \int_{\partial B(x,r)} u(y)\; \mathrm{d}y.$$

*Proof.* IA Vector Calculus. □

**Theorem.** Let $(B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$, and $u : \mathbb{R}^d \to \mathbb{R}$ be harmonic such that

$$\mathbb{E}|u(x + B_t)| < \infty$$

for any $x \in \mathbb{R}^d$ and $t \geq 0$. Then the process $(u(B_t))_{t\geq 0}$ is a martingale with respect to $(\mathcal{F}_t^+)_{t\geq 0}$.

**Lemma.** If $X$ and $Y$ are independent random variables in $\mathbb{R}^d$, and $X$ is $\mathcal{G}$-measurable. If $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is such that $f(X, Y)$ is integrable, then

$$\mathbb{E}(f(X, Y) \mid \mathcal{G}) = \mathbb{E}f(z, Y)|_{z=X}.$$

*Proof.* Use Fubini and the fact that $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$. □

*Proof of theorem.* Let $t \geq s$. Then

$$\begin{aligned}
\mathbb{E}(u(B_t) \mid \mathcal{F}_s^+) &= \mathbb{E}(u(B_s + (B_t - B_s)) \mid \mathcal{F}_s^+) \\
&= \mathbb{E}(u(z + B_t - B_s))|_{Z=B_s} \\
&= u(z)|_{z=B_s} \\
&= u(B_s).
\end{aligned}$$

□

**Theorem.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice continuously differentiable with bounded derivatives. Then, the processes $(X_t)_{t\geq 0}$ defined by

$$X_t = f(B_t) - \frac{1}{2}\int_0^t \Delta f(B_s)\, \mathrm{d}s$$

is a martingale with respect to $(\mathcal{F}_t^+)_{t\geq 0}$.

*Proof.* Follows from the mean value property of harmonic functions. □

**Corollary.** If $u$ and $u'$ solve $\Delta u = \Delta u' = 0$, and $u$ and $u'$ agree on $\partial D$, then $u = u'$.

*Proof.* $u - u'$ is also harmonic, and so attains the maximum at the boundary, where it is 0. Similarly, the minimum is attained at the boundary. □

**Lemma.** Let $C$ be an open cone in $\mathbb{R}^d$ based at 0. Then there exists $0 \leq a < 1$ such that if $|x| \leq \frac{1}{2^k}$, then

$$\mathbb{P}_x(T_{\partial B(0,1)} < T_C) \leq a^k.$$

*Proof.* Pick

$$a = \sup_{|x|\leq \frac{1}{2}} \mathbb{P}_x(T_{\partial B(0,1)} < T_C) < 1.$$

We then apply the strong Markov property, and the fact that Brownian motion is scale invariant. We reason as follows — if we start with $|x| \leq \frac{1}{2^k}$, we may or may not hit $\partial B(2^{-k+1})$ before hitting $C$. If we don't, then we are happy. If we are not, then we have reached $\partial B(2^{-k+1})$. This happens with probability at most $a$. Now that we are at $\partial B(2^{-k+1})$, the probability of hitting $\partial B(2^{-k+2})$ before hitting the cone is at most $a$ again. If we hit $\partial B(2^{-k+3})$, we again have a probability of $\leq a$ of hitting $\partial B(2^{-k+4})$, and keep going on. Then by induction, we find that the probability of hitting $\partial B(0, 1)$ before hitting the cone is $\leq a^k$. □

**Theorem.** Let $D$ be a bounded domain satisfying the Poincaré cone condition, and let $\varphi : \partial D \to \mathbb{R}$ be continuous. Let

$$T_{\partial D} = \inf\{t \geq 0 : B_t \in \partial D\}.$$

This is a *bounded* stopping time. Then the function $u : \bar{D} \to \mathbb{R}$ defined by

$$u(x) = \mathbb{E}_x(\varphi(B_{T_{\partial D}})),$$

where $\mathbb{E}_x$ is the expectation if we start at $x$, is the unique continuous function such that $u(x) = \varphi(x)$ for $x \in \partial D$, and $\Delta u = 0$ for $x \in D$.

*Proof.* Let $\tau = T_{\partial B(x,\delta)}$ for $\delta$ small. Then by the strong Markov property and the tower law, we have

$$u(x) = \mathbb{E}_x(u(x_\tau)),$$

and $x_\tau$ is uniformly distributed over $\partial B(x, \delta)$. So we know $u$ is harmonic in the interior of $D$, and in particular is continuous in the interior. It is also clear that $u|_{\partial D} = \varphi$. So it remains to show that $u$ is continuous up to $\bar{D}$.

So let $x \in \partial D$. Since $\varphi$ is continuous, for every $\varepsilon > 0$, there is $\delta > 0$ such that if $y \in \partial D$ and $|y - x| < \delta$, then $|\varphi(y) - \varphi(x)| \leq \varepsilon$.

Take $z \in \bar{D}$ such that $|z - x| \leq \frac{\delta}{2}$. Suppose we start our Brownian motion at $z$. If we hit the boundary before we leave the ball, then we are in good shape. If not, then we are sad. But if the second case has small probability, then since $\varphi$ is bounded, we can still be fine.

Pick a cone $C$ as in the definition of the Poincaré cone condition, and assume we picked $\delta$ small enough that $C \cap B(x, \delta) \cap D = \emptyset$. Then we have

$$\begin{aligned}
|u(z) - \varphi(x)| &= |\mathbb{E}_z(\varphi(B_{T_{\partial D}})) - \varphi(x)| \\
&\leq \mathbb{E}_z|\varphi(B_{T_{\partial D}} - \varphi(x))| \\
&\leq \varepsilon \mathbb{P}_z(T_{B(x,\delta)} > T_{\partial D}) + 2 \sup \|\varphi\| \mathbb{P}_z(T_{\partial D} > T_{\partial B(x,\delta)}) \\
&\leq \varepsilon + 2\|\varphi\|_\infty \mathbb{P}_z(T_{B(x,\delta)} \leq T_C),
\end{aligned}$$

and we know the second term $\to 0$ as $z \to x$. ☐

## 5.3 Transience and recurrence

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

- If $d = 1$, then $(B_t)_{t\geq 0}$ is *point recurrent*, i.e. for each $x, z \in \mathbb{R}$, the set $\{t \geq 0 : B_t = z\}$ is unbounded $\mathbb{P}_x$-almost surely.
- If $d = 2$, then $(B_t)_{t\geq 0}$ is *neighbourhood recurrent*, i.e. for each $x \in \mathbb{R}^2$ and $U \subseteq \mathbb{R}^2$ open, the set $\{t \geq 0 : B_t \in U\}$ is unbounded $\mathbb{P}_x$-almost surely. However, the process does not visit points, i.e. for all $x, z \in \mathbb{R}^d$, we have

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

- If $d \geq 3$, then $(B_t)_{t\geq 0}$ is *transient*, i.e. $|B_t| \to \infty$ as $t \to \infty$ $\mathbb{P}_x$-almost surely.

*Proof.*

– This is trivial, since $\inf_{t\geq 0} B_t = -\infty$ and $\sup_{t\geq 0} B_t = \infty$ almost surely, and $(B_t)_{t\geq 0}$ is continuous.

– It is enough to prove for $x = 0$. Let $0 < \varepsilon < R < \infty$ and $\varphi \in C_b^2(\mathbb{R}^2)$ such that $\varphi(x) = \log|x|$ for $\varepsilon \leq |x| \leq R$. It is an easy exercise to check that this is harmonic inside the annulus. By the theorem we didn't prove, we know

$$M_t = \varphi(B_t) - \frac{1}{2}\int_0^t \Delta\varphi(B_s)\,\mathrm{d}s$$

is a martingale. For $\lambda \geq 0$, let $S_\lambda = \inf\{t \geq 0 : |B_t| = \lambda\}$. If $\varepsilon \leq |x| \leq R$, then $H = S_\varepsilon \wedge S_R$ is $\mathbb{P}_X$-almost surely finite. Then $M_H$ is a bounded martingale. By optional stopping, we have

$$\mathbb{E}_x(\log|B_H|) = \log|x|.$$

But the LHS is

$$\log\varepsilon\mathbb{P}(S_\varepsilon < S_R) + \log R\mathbb{P}(S_R < S_\varepsilon).$$

So we find that

$$\mathbb{P}_x(S_\varepsilon < S_R) = \frac{\log R - \log|x|}{\log R - \log\varepsilon}. \tag{$*$}$$

Note that if we let $R \to \infty$, then $S_R \to \infty$ almost surely. Using $(*)$, this implies $\mathbb{P}_X(S_\varepsilon < \infty) = 1$, and this does not depend on $x$. So we are done.

To prove that $(B_t)_{t\geq 0}$ does not visit points, let $\varepsilon \to 0$ in $(*)$ and then $R \to \infty$ for $x \neq z = 0$.

– It is enough to consider the case $d = 3$. As before, let $\varphi \in C_b^2(\mathbb{R}^3)$ be such that

$$\varphi(x) = \frac{1}{|x|}$$

for $\varepsilon \leq x \leq 2$. Then $\Delta\varphi(x) = 0$ for $\varepsilon \leq x \leq R$. As before, we get

$$\mathbb{P}_x(S_\varepsilon < S_R) = \frac{|x|^{-1} - |R|^{-1}}{\varepsilon^{-1} - R^{-1}}.$$

As $R \to \infty$, we have

$$\mathbb{P}_x(S_\varepsilon < \infty) = \frac{\varepsilon}{x}.$$

Now let

$$A_n = \{B_t \geq n \text{ for all } t \geq B_{T_{n^3}}\}.$$

Then

$$\mathbb{P}_0(A_n^c) = \frac{1}{n^2}.$$

So by Borel–Cantelli, we know only finitely of $A_n^c$ occur almost surely. So infinitely many of the $A_n$ hold. This guarantees our process $\to \infty$. $\square$

## 5.4 Donsker's invariance principle

**Theorem** (Donsker's invariance principle)**.** Let $(X_n)_{n\geq 0}$ be iid random variables with mean 0 and variance 1, and set $S_n = X_1 + \cdots + X_n$. Define

$$S_t = (1 - \{t\})s_{\lfloor t\rfloor} + \{t\}S_{\lfloor t\rfloor+1}.$$

where $\{t\} = t - \lfloor t\rfloor$.

Define

$$(S_t^{[N]})_{t\geq 0} = (N^{-1/2}S_{t\cdot N})_{t\in[0,1]}.$$

As $(S_t^{[N]})_{t\in[0,1]}$ converges in distribution to the law of standard Brownian motion on $[0, 1]$.

**Theorem** (Skorokhod embedding theorem)**.** Let $\mu$ be a probability measure on $\mathbb{R}$ with mean 0 and variance $\sigma^2$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ on which there is a standard Brownian motion $(B_t)_{t\geq 0}$ and a sequence of stopping times $(T_n)_{n\geq 0}$ such that, setting $S_n = B_{T_n}$,

(i) $T_n$ is a random walk with steps of mean $\sigma^2$

(ii) $S_n$ is a random walk with step distribution $\mu$.

**Lemma.** Let $x, y > 0$. Then

$$\mathbb{P}_0(T_{-x} < T_y) = \frac{y}{x+y}, \quad \mathbb{E}_0T_{-x} \wedge T_y = xy.$$

*Proof sketch.* Use optional stopping with $(B_t^2 - t)_{t\geq 0}$. □

*Proof of Skorokhod embedding theorem.* Define Borel measures $\mu^\pm$ on $[0, \infty)$ by

$$\mu^\pm(A) = \mu(\pm A).$$

Note that these are not probability measures, but we can define a probability measure $\nu$ on $[0, \infty)^2$ given by

$$\mathrm{d}\nu(x, y) = C(x + y)\;\mathrm{d}\mu^-(x)\;\mathrm{d}\mu^+(y)$$

for some normalizing constant $C$ (this is possible since $\mu$ is integrable). This $(x+y)$ is the same $(x+y)$ appearing in the denominator of $\mathbb{P}_0(T_{-x} < T_y) = \frac{y}{x+y}$. Then we claim that any $(X, Y)$ with this distribution will do the job.

We first figure out the value of $C$. Note that since $\mu$ has mean 0, we have

$$C\int_0^\infty x\;\mathrm{d}\mu^-(x) = C\int_0^\infty y\;\mathrm{d}\mu^+(y).$$

Thus, we have

$$\begin{aligned}
1 &= \int C(x + y)\;\mathrm{d}\mu^-(x)\;\mathrm{d}\mu^+(y)\\
&= C\int x\;\mathrm{d}\mu^-(x)\int\;\mathrm{d}\mu^+(y) + C\int y\;\mathrm{d}\mu^+(y)\int\;\mathrm{d}\mu^-(x)\\
&= C\int x\;\mathrm{d}\mu^-(x)\left(\int\;\mathrm{d}\mu^+(y) + \int\;\mathrm{d}\mu^-(x)\right)\\
&= C\int x\;\mathrm{d}\mu^-(x) = C\int y\;\mathrm{d}\mu^+(y).
\end{aligned}$$

We now set up our notation. Take a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a standard Brownian motion $(B_t)_{t\geq 0}$ and a sequence $((X_n, Y_n))_{n\geq 0}$ iid with distribution $\nu$ and independent of $(B_t)_{t\geq 0}$.

Define

$$\mathcal{F}_0 = \sigma((X_n, Y_n), n = 1, 2, \ldots), \quad \mathcal{F}_t = \sigma(\mathcal{F}_0, \mathcal{F}_t^B).$$

Define a sequence of stopping times

$$T_0 = 0, T_{n+1} = \inf\{t \geq T_n : B_t - B_{T_n} \in \{-X_{n+1}, Y_{n+1}\}\}.$$

By the strong Markov property, it suffices to prove that things work in the case $n = 1$. So for convenience, let $T = T_1, X = X_1, Y = Y_1$.

To simplify notation, let $\tau : C([0, 1], \mathbb{R}) \times [0, \infty)^2 \to [0, \infty)$ be given by

$$\tau(\omega, x, y) = \inf\{t \geq 0 : \omega(t) \in \{-x, y\}\}.$$

Then we have

$$T = \tau((B_t)_{t\geq 0}, X, Y).$$

To check that this works, i.e. (ii) holds, if $A \subseteq [0, \infty)$, then

$$\mathbb{P}(B_T \in A) = \int_{[0,\infty)^2} \int_{C([0,\infty),\mathbb{R})} \mathbf{1}_{\tau(\omega,x,y)\in A} \;\mathrm{d}\mu_B(\omega)\;\mathrm{d}\nu(x, y).$$

Using the first part of the previous computation, this is given by

$$\int_{[0,\infty)^2} \frac{x}{x+y} \mathbf{1}_{y\in A}\; C(x + y)\;\mathrm{d}\mu^-(x)\;\mathrm{d}\mu^+(y) = \mu^+(A).$$

We can prove a similar result if $A \subseteq (-\infty, 0)$. So $B_T$ has the right law.

To see that $T$ is also well-behaved, we compute

$$\begin{aligned}
\mathbb{E}T &= \int_{[0,\infty)^2} \int_{\mathbb{C}([0,1],\mathbb{R})} \tau(\omega, x, y)\;\mathrm{d}\mu_B(\omega)\;\mathrm{d}\nu(x, y)\\
&= \int_{[0,\infty)^2} xy\;\mathrm{d}\nu(x, y)\\
&= C \int_{[0,\infty)^2} (x^2 y + yx^2)\;\mathrm{d}\mu^-(x)\;\mathrm{d}\mu^+(y)\\
&= \int_{[0,\infty)} x^2\;\mathrm{d}\mu^-(x) + \int_{[0,\infty)} y^2\;\mathrm{d}\mu^+(y)\\
&= \sigma^2.
\end{aligned}$$

□

*Proof of Donsker's invariance principle.* Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then by Brownian scaling,

$$(B_t^{(N)})_{t\geq 0} = (N^{1/2} B_{t/N})_{t\geq 0}$$

is a standard Brownian motion.

For every $N > 0$, we let $(T_n^{(N)})_{n\geq 0}$ be a sequence of stopping times as in the embedding theorem for $B^{(N)}$. We then set

$$S_n^{(N)} = B^{(N)}_{T_n^{(N)}}.$$

For $t$ not an integer, define $S_t^{(N)}$ by linear interpolation. Observe that

$$((T_n^{(N)})_{n\geq 0}, S_t^{(N)}) \sim ((T_n^{(1)})_{n\geq 0}, S_t^{(1)}).$$

We define

$$\tilde{S}_t^{(N)} = N^{-1/2} S_{tN}^{(N)}, \quad \tilde{T}_n^{(N)} = \frac{T_n^{(N)}}{N}.$$

Note that if $t = \frac{n}{N}$, then

$$\tilde{S}_{n/N}^{(N)} = N^{-1/2} S_n^{(N)} = N^{-1/2} B^{(N)}_{T_n^{(N)}} = B_{T_n^{(N)}/N} = B_{\tilde{T}_n^N}. \tag{$*$}$$

Note that $(\tilde{S}_t^{(N)})_{t\geq 0} \sim (S_t^{(N)})_{t\geq 0}$. We will prove that we have convergence in probability, i.e. for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{0\leq t<1} |\tilde{S}_t^{(N)} - B_t| > \delta\right) = \mathbb{P}(\|\tilde{S}^{(N)} - B\|_\infty > \delta) \to 0 \text{ as } N \to \infty.$$

We already know that $\tilde{S}$ and $B$ agree at some times, but the time on $\tilde{S}$ is fixed while that on $B$ is random. So what we want to apply is the law of large numbers. By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n}|T_n^{(1)} - n| \to 0 \text{ as } n \to 0.$$

This implies that

$$\sup_{1\leq n\leq N} \frac{1}{N}|T_n^{(1)} - n| \to 0 \text{ as } N \to \infty.$$

Note that $(T_n^{(1)})_{n\geq 0} \sim (T_n^{(N)})_{n\geq 0}$, it follows for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{1\leq n\leq N}\left|\frac{T_n^{(N)}}{N} - \frac{n}{N}\right| \geq \delta\right) \to 0 \text{ as } N \to \infty.$$

Using $(*)$ and continuity, for any $t \in [\frac{n}{N}, \frac{n+1}{N}]$, there exists $u \in [T_{n/N}^{(N)}, T_{(n+1)/N}^{(N)}]$ such that

$$\tilde{S}_t^{(N)} = B_u.$$

Note that if times are approximated well up to $\delta$, then $|t - u| \leq \delta + \frac{1}{N}$.

Hence we have

$$\begin{aligned}\{\|\tilde{S} - B\|_\infty > \varepsilon\} \leq &\left\{\left|\tilde{T}_n^{(N)} - \frac{n}{N}\right| > \delta \text{ for some } n \leq N\right\}\\ &\cup \left\{|B_t - B_u| > \varepsilon \text{ for some } t \in [0,1], |t-u| < \delta + \frac{1}{N}\right\}.\end{aligned}$$

The first probability $\to 0$ as $n \to \infty$. For the second, we observe that $(B_t)_{T\in[0,1]}$ has uniformly continuous paths, so for $\varepsilon > 0$, we can find $\delta > 0$ such that the second probability is less than $\varepsilon$ whenever $N > \frac{1}{\delta}$ (exercise!).

So $\tilde{S}^{(N)} \to B$ uniformly in probability, hence converges uniformly in distribution. $\square$

# 6 Large deviations

**Lemma** (Fekete)**.** If $b_n$ is a non-negative sub-additive sequence, then $\lim_n \frac{b_n}{n}$ exists.

**Theorem** (Cramér's theorem)**.** For $a > \bar{x}$, we have

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

*Proof.* We first prove an upper bound. For any $\lambda$, Markov tells us

$$\begin{aligned}\mathbb{P}(S_n \geq an) &= \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda an}) \leq e^{-\lambda a_n} \mathbb{E} e^{\lambda S_n}\\ &= e^{-\lambda an} \prod_{i=1}^n \mathbb{E} e^{\lambda X_i} = e^{-\lambda an} M(\lambda)^n = e^{-n(\lambda a - \psi(\lambda))}.\end{aligned}$$

Since $\lambda$ was arbitrary, we can pick $\lambda$ to maximize $\lambda a - \psi(\lambda)$, and so by definition of $\psi^*(a)$, we have $\mathbb{P}(S_n \geq a_n) \leq e^{-n\psi^*(a)}$. So it follows that

$$\limsup \frac{1}{n} \log \mathbb{P}(S_n \geq a_n) \leq -\psi^*(a).$$

The lower bound is a bit more involved. One checks that by translating $X_i$ by $a$, we may assume $a = 0$, and in particular, $\bar{x} < 0$.

So we want to prove that

$$\liminf_n \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq \inf_{\lambda \geq 0} \psi(\lambda).$$

We consider cases:

- If $\mathbb{P}(X \leq 0) = 1$, then

  $$\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_i = 0 \text{ for } i = 1, \ldots n) = \mathbb{P}(X_1 = 0)^n.$$

  So in fact

  $$\liminf_n \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \log \mathbb{P}(X_1 = 0).$$

  But by monotone convergence, we have

  $$\mathbb{P}(X_1 = 0) = \lim_{\lambda \to \infty} \mathbb{E} e^{\lambda X_1}.$$

  So we are done.

- Consider the case $\mathbb{P}(X_1 > 0) > 0$, but $\mathbb{P}(X_1 \in [-K, K]) = 1$ for some $K$. The idea is to modify $X_1$ so that it has mean 0. For $\mu = \mu_{X_1}$, we define a new distribution by the density

  $$\frac{\mathrm{d}\mu^\theta}{\mathrm{d}\mu}(x) = \frac{e^{\theta x}}{M(\theta)}.$$

  We define

  $$g(\theta) = \int x \, \mathrm{d}\mu^\theta(x).$$

We claim that $g$ is continuous for $\theta \geq 0$. Indeed, by definition,

$$g(\theta) = \frac{\int xe^{\theta x}\, \mathrm{d}\mu(x)}{\int e^{\theta x}\, \mathrm{d}\mu(x)},$$

and both the numerator and denominator are continuous in $\theta$ by dominated convergence.

Now observe that $g(0) = \bar{x}$, and

$$\limsup_{\theta \to \infty} g(\theta) > 0.$$

So by the intermediate value theorem, we can find some $\theta_0$ such that $g(\theta_0) = 0$.

Define $\mu_n^{\theta_0}$ to be the law of the sum of $n$ iid random variables with law $\mu^{\theta_0}$. We have

$$\mathbb{P}(S_n \geq 0) \geq \mathbb{P}(S_n \in [0, \varepsilon n]) \geq \mathbb{E} e^{\theta_0 (S_n - \varepsilon n)} \mathbf{1}_{S_n \in [0, \varepsilon n]},$$

using the fact that on the event $S_n \in [0, \varepsilon n]$, we have $e^{\theta_0 (S_n - \varepsilon n)} \leq 1$. So we have

$$\mathbb{P}(S_n \geq 0) \geq M(\theta_0)^n e^{-\theta_0 \varepsilon n} \mu_n^{\theta_0}(\{S_n \in [0, \varepsilon n]\}).$$

By the central limit theorem, for each fixed $\varepsilon$, we know

$$\mu_n^{\theta_0}(\{S_n \in [0, \varepsilon n]\}) \to \frac{1}{2} \text{ as } n \to \infty.$$

So we can write

$$\liminf_n \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq \psi(\theta_0) - \theta_0 \varepsilon.$$

Then take the limit $\varepsilon \to 0$ to conclude the result.

– Finally, we drop the finiteness assumption, and only assume $\mathbb{P}(X_1 > 0) > 0$. We define $\nu$ to be the law of $X_1$ condition on the event $\{|X_1| \leq K\}$. Let $\nu_n$ be the law of the sum of $n$ iid random variables with law $\nu$. Define

$$\begin{aligned}
\psi_K(\lambda) &= \log \int_{-K}^{K} e^{\lambda x}\, \mathrm{d}\mu(x) \\
\psi^\nu(\lambda) &= \log \int_{-\infty}^{\infty} e^{\lambda x}\, \mathrm{d}\nu(x) = \psi_K(\lambda) - \log \mu(\{|X| \leq K\}).
\end{aligned}$$

Note that for $K$ large enough, $\int x\, \mathrm{d}\nu(x) < 0$. So we can use the previous case. By definition of $\nu$, we have

$$\mu_n([0, \infty)) \geq \nu([0, \infty)) \mu(|X| \leq K)^n.$$

So we have

$$\begin{aligned}
\liminf_n \frac{1}{n} \log \mu([0, \infty)) &\geq \log \mu(|X| \leq K) + \liminf \log \nu_n([0, \infty)) \\
&\geq \log \mu(|X| \leq K) + \inf \psi^\nu(\lambda) \\
&= \inf_\lambda \psi_K(\lambda) \\
&= \psi_K^\lambda.
\end{aligned}$$

Since $\psi_K$ increases as $K$ increases to infinity, this increases to some $\mathcal{J}$ we have

$$\liminf_n \frac{1}{n} \log \mu_n([0, \infty)) \geq \mathcal{J}. \tag{$\dagger$}$$

Since $\psi_K(\lambda)$ are continuous, $\{\lambda : \psi_K(\lambda) \leq \mathcal{J}\}$ is non-empty, compact and nested in $K$. By Cantor's theorem, we can find

$$\lambda_0 \in \bigcap_K \{\lambda : \psi_K(\lambda) \leq \mathcal{J}\}.$$

So the RHS of (†) satisfies

$$\mathcal{J} \geq \sup_K \psi_K(\lambda_0) = \psi(\lambda_0) \geq \inf_\lambda \psi(\lambda). \qquad \square$$