Advanced Probability

Cambridge Part III, Michaelmas 2022
Taught by Perla Sousi
Notes taken by Leonard Tomczak

Contents

1 Conditional Expectation

1.1 Some Recap

We recall some basic definitions:

A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a set, $\mathcal{F}$ a σ-algebra on Ω and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$, i.e. a measure with total mass $\mathbb{P}(\Omega) = 1$.

If $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, then the *conditional probability of* A *given* B is $\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

The *Borel σ-algebra* of $\mathbb{R}$, denoted $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by the open subsets of $\mathbb{R}$.

A (real valued) *random variable* on Ω is a map $X : \Omega \to \mathbb{R}$ that is $\mathcal{F} - \mathcal{B}$-measurable, i.e. $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$ (or equivalently just for all open subsets $B \subseteq \mathbb{R}$).

If $(X_i)_{i \in I}$ is a collection of random variables on Ω, then the σ-algebra generated by the X_i

$$\sigma(X_i : i \in I) = \sigma(\{X_i^{-1}(B) \mid i \in I, B \in \mathcal{B}(\mathbb{R})\})$$

is the smallest σ-algebra such that all $(X_i)_{i \in I}$ are measurable w.r.t. it.

Notation. If $A \in \mathcal{F}$, we denote the indicator function for A by $1(A)$.

We finally recall the definition of the expectation of a random variable:

(1) For simple non-negative random variables X, i.e. those of the form $X = \sum_{i=1}^{n} c_i 1(A_i)$ with $A_i \in \mathcal{F}$, $c_i \geq 0$, define $\mathbb{E}[X] := \sum_{i=1}^{n} c_i \mathbb{P}(A_i)$.

(2) If $X \geq 0$ is any non-negative random variable, let X_n be a sequence of non-negative simple random variables such that $X_n \nearrow X$ pointwise as $n \to \infty$, e.g. one can take $X_n(\omega) = \min(2^{-n} \lfloor 2^n X(\omega) \rfloor, n)$ for all ω. Then define $\mathbb{E}[X] := \lim_{n \to \infty} \mathbb{E}[X_n]$.

(3) Let X be any random variable. Let $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$ so that $X = X^+ - X^-$. If at least one of $\mathbb{E}[X^+]$ or $\mathbb{E}[X^-]$ is finite, define $\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

X is *integrable* if $\mathbb{E}[|X|] < \infty$.

Let $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ and X integrable. Then define $\mathbb{E}[X \mid A] := \frac{\mathbb{E}[X 1(A)]}{\mathbb{P}(A)}$.

1.2 Conditional Expectation in the Discrete Case

Let $(B_i)_{i\in I}$ be a countable collection of pairwise disjoint events $B_i \in \mathcal{F}$ such that $\bigcup_{i\in I} B_i = \Omega$. Let $\mathcal{G} = \sigma((B_i)_{i\in I})$ be the σ-algebra generated by the B_i. One easily checks that

$$\mathcal{G} = \{\bigcup_{j\in J} B_j \mid J \subseteq I\}.$$

Let X be an integrable random variable on Ω. We want to define the conditional expectation $\mathbb{E}[X \mid \mathcal{G}] = X' : \Omega \to \mathbb{R}$. We let

$$X' := \sum_{i\in I} \mathbb{E}[X \mid B_i] 1(B_i).$$

Here $\mathbb{E}[X \mid B_i]$ is defined as above in the case $\mathbb{P}(B_i) > 0$. If $\mathbb{P}(B_i) = 0$, we let $\mathbb{E}[X \mid B_i] = 0$.

X' has the following properties:

(1) X' is $\mathcal{G}$-measurable.

(2) For all $A \in \mathcal{G}$, $\mathbb{E}[X1(A)] = \mathbb{E}[X'1(A)]$.

Both are immediate from the definition. In the general case we will use these properties to define $\mathbb{E}[X \mid \mathcal{G}]$. Also note that

$$\mathbb{E}[|X'|] \le \sum_i \mathbb{E}[|X|1(B_i)] = \mathbb{E}[|X|] < \infty,$$

so X' is also integrable.

1.3 Conditional Expectation in General

Theorem 1.1 (Conditional expectation)**.** *Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. Then there exists an integrable random variable Y satisfying:*

(a) Y is $\mathcal{G}$-measurable.

(b) For all $A \in \mathcal{G}$, we have $\mathbb{E}[X1(A)] = \mathbb{E}[Y1(A)]$.

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

We call Y a version *of the* conditional expectation of X given $\mathcal{G}$ *and write $Y = \mathbb{E}[X \mid \mathcal{G}]$ a.s.*

Remark. Instead of (b) we could have asked that $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all bounded $\mathcal{G}$-measurable random variables Z.

Proof. **Uniqueness.** Let Y' satisfy (a) and (b). Consider $A = \{Y > Y'\}$. By (a), we have $A \in \mathcal{G}$, and by (b), $\mathbb{E}[(Y' - Y)1(A)] = 0$, hence $Y \le Y'$ a.s., and similarly $Y \ge Y'$ a.s.

Existence.

1. Assume first that $X \in \mathcal{L}^2(\mathcal{F}) = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Recall that these $\mathcal{L}^2$ spaces are Hilbert spaces. The space $\mathcal{L}^2(\mathcal{G}) = \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $\mathcal{L}^2(\mathcal{F})$. Hence $\mathcal{L}^2(\mathcal{F}) = \mathcal{L}^2(\mathcal{G}) + \mathcal{L}^2(\mathcal{G})^\perp$, so we can write $X = Y + Z$ where $Y \in \mathcal{L}^2(\mathcal{G})$ and $Z \in \mathcal{L}^2(\mathcal{G})$. We set $\mathbb{E}[X \mid \mathcal{G}] = Y$. Y is $\mathcal{G}$-measurable by definition. For (b) let $A \in \mathcal{G}$. Then $\mathbb{E}[X1(A)] = \mathbb{E}[Y1(A)] + \mathbb{E}[Z1(A)] = \mathbb{E}[Y1(A)]$ since $1(A) \in \mathcal{L}(\mathcal{G})$ and $Z \in \mathcal{L}^2(\mathcal{G})^\perp$.

   Note that if $X \ge 0$, then $\mathbb{E}[X \mid \mathcal{G}] \ge 0$ a.s. Indeed, if $A = \{Y < 0\}$, then $0 \le \mathbb{E}[X1(A)] = \mathbb{E}[Y1(A)] \le 0$, so $\mathbb{P}(A) = 0$, i.e. $Y \ge 0$ a.s.

2. Assume $X \ge 0$. Then let $X_n = X \wedge n$ for all n. Note that $X_n \in \mathcal{L}^2$, so that $Y_n := \mathbb{E}[X_n \mid \mathcal{G}]$ is defined by 1. Note that $X_n \nearrow X$ as $n \to \infty$ a.s. By the final remark in 1., Y_n is a.s. increasing. Let $Y = \limsup_{n\to\infty} Y_n$. Since the Y_n are $\mathcal{G}$-measurable, so is Y. By the a.s. increasing property of $(Y_n)_n$, we get $Y = \lim_{n\to\infty} Y_n$ a.s. Let $A \in \mathcal{G}$. Take the limit as $n \to \infty$ in $\mathbb{E}[Y_n 1(A)] = \mathbb{E}[X_n 1(A)]$ and use the monotone convergence theorem to get $\mathbb{E}[Y1(A)] = \mathbb{E}[X1(A)]$.

   Taking $A = \Omega$ in particular shows that $\mathbb{E}[Y] = \mathbb{E}[X] < \infty$.

3. Let X be any integrable random variable. Write $X = X^+ - X^-$ as usual and define $\mathbb{E}[X \mid \mathcal{G}] := \mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}]$.

□

The second step shows that if $X \ge 0$, but not necessarily integrable, we can still define the conditional expectation Y satisfying (a) and (b), but it need not be integrable.

We extend the notion of independence to σ-algebras and random variables:

Definition. *A sequence $\mathcal{G}_1, \mathcal{G}_2, \ldots$ of σ-algebras is called* independent *if for all $i_1, \ldots, i_k$ distinct and $G_i \in G_i$, we have*

$$\mathbb{P}(G_{i_1} \cap \cdots \cap G_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(G_{i_j}).$$

A random variable X is independent *of a σ-algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.*

Notation. If $\mathcal{G} = \sigma(Z)$, we write $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X \mid Z]$.

Some properties of conditional expectation:

Proposition 1.2. *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. Then*

(1) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$.

(2) *If X is $\mathcal{G}$-measurable, then* $\mathbb{E}[X \mid \mathcal{G}] = X$ *a.s.*

(3) *If X is independent of $\mathcal{G}$, then* $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$*.*

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

(5) $\mathbb{E}[X \mid \mathcal{G}]$ *is linear in* X*.*

(6) $|\mathbb{E}[X \mid \mathcal{G}]| \leq \mathbb{E}[|X| \mid \mathcal{G}]$ *a.s.*

There are analogues of classical measure theory theorems for conditional expection. Let X_n be a sequence of random variables.

Proposition 1.3 (Conditional monotone convergence)**.** *If* $X_n \geq 0$ *and* $X_n \nearrow X$ *as* $n \to \infty$ *a.s., then*

$$\mathbb{E}[X_n \mid \mathcal{G}] \nearrow \mathbb{E}[X \mid \mathcal{G}] \text{ a.s. as } n \to \infty.$$

Proof. $Y_n = \mathbb{E}[X_n \mid \mathcal{G}]$ is a.s. increasing. Let $Y = \limsup Y_n$. Then Y is $\mathcal{G}$-measurable and $Y = \lim_{n\to\infty} Y_n$ a.s. Clearly Y satisfies the properties defining $\mathbb{E}[X \mid \mathcal{G}]$ (using monotone convergence), so that $\mathbb{E}[X \mid \mathcal{G}] = Y$ a.s. □

Proposition 1.4 (Conditional Fatou)**.** *If* $X_n \geq 0$*, then*

$$\mathbb{E}[\liminf X_n \mid \mathcal{G}] \leq \liminf \mathbb{E}[X_n \mid \mathcal{G}] \text{ a.s.}$$

Proof. Note that $\inf_{k\geq n} X_k \nearrow \liminf X_k$ as $n \to \infty$. So by conditional monotone convergence, $\lim_{n\to\infty} \mathbb{E}[\inf_{k\geq n} X_k \mid \mathcal{G}] = \mathbb{E}[\liminf X_k \mid \mathcal{G}]$. For every n we have $\mathbb{E}[\inf_{k\geq n} E_k \mid \mathcal{G}] \leq \mathbb{E}[X_k \mid \mathcal{G}]$ a.s., so $\mathbb{E}[\inf_{k\geq n} E_k \mid \mathcal{G}] \leq \inf_{k\geq n} \mathbb{E}[X_k \mid \mathcal{G}]$ a.s. Now take limits. □

Proposition 1.5 (Conditional dominated convergence)**.** *Suppose* $X_n \to X$ *a.s.,* $|X_n| \leq Y$ *for all n where* $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$*. Then*

$$\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}] \text{ a.s. as } n \to \infty.$$

Proof. Note that $X_n + Y, Y - X_n \geq 0$ for all n, so by Fatou, $\mathbb{E}[X + Y \mid \mathcal{G}] \leq \liminf_n \mathbb{E}[X_n + Y \mid \mathcal{G}]$ a.s., so $\liminf \mathbb{E}[X_n \mid \mathcal{G}] \geq \mathbb{E}[X \mid \mathcal{G}]$ a.s. Similarly $\limsup \mathbb{E}[X_n \mid \mathcal{G}] \leq \mathbb{E}[X \mid \mathcal{G}]$ a.s. □

Proposition 1.6 (Conditional Jensen)**.** *Let* $X \in \mathcal{L}^1$ *and* $\varphi : \mathbb{R} \to (-\infty, \infty]$ *be a convex function such that* $\varphi(X) \geq 0$ *or* $\varphi(X)$ *is integrable. Then*

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \varphi(\mathbb{E}[X \mid \mathcal{G}]) \text{ a.s.}$$

Proof. We can write $\varphi(x) = \sup_{i\in\mathbb{N}}(a_i x + b_i)$ with $a_i, b_i \in \mathbb{R}$. Then for all $i \in \mathbb{N}$, $\varphi(X) \geq a_i X + b_i$, so $\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq a_i \mathbb{E}[X \mid \mathcal{G}] + b_i$ a.s. Then by countability we get $\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \sup_i (a_i \mathbb{E}[X \mid \mathcal{G}] + b_i) = \varphi(\mathbb{E}[X \mid \mathcal{G}])$ a.s. □

Consequence: For all $p \in [1, \infty)$, we have $\|\mathbb{E}[X \mid \mathcal{G}]\|_p \leq \|X\|_p$. Indeed,

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

Proposition 1.7 (Tower property)**.** *Let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be sub-σ-algebras. Then*

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}] \text{ a.s.}$$

Proof. We check that $\mathbb{E}[X \mid \mathcal{H}]$ satisfies the defining properties of the conditional expectation of $\mathbb{E}[X \mid \mathcal{G}]$ given $\mathcal{H}$. It is clearly $\mathcal{H}$-measurable. Let $A \in \mathcal{H}$. Then also $A \in \mathcal{G}$ and so

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{H}]1(A)] = \mathbb{E}[X1(A)] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1(A)].$$

□

Proposition 1.8 (Take out what is known)**.** *Let $X \in \mathcal{L}^1$ and let Y be a bounded $\mathcal{G}$-measurable random variable. Then*

$$\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}] \text{ a.s.}$$

Proof. $Y\mathbb{E}[X \mid \mathcal{G}]$ is clearly $\mathcal{G}$-measurable. Let $A \in \mathcal{G}$, then

$$\mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]1(A)] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}](Y1(A))] = \mathbb{E}[(XY)1(A)].$$

□

Theorem 1.9. *Let X be an integrable random variable in $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sub-σ-algebras. Assume that $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$. Then*

$$\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}] \text{ a.s.}$$

For the proof we recall some measure theory: A set $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is called a *π-system* if $\mathcal{A}$ is closed under finite intersections.

Theorem 1.10 (Uniqueness of extension)**.** *Let μ and ν be two measures on the same measurable space $(E, \mathcal{E})$. If $\mu(E) = \nu(E) < \infty$ and μ and ν agree on a π-system generating $\mathcal{E}$, then $\mu = \nu$.*

Proof of Theorem 1.9. $\mathbb{E}[X \mid \mathcal{G}]$ is obviously $\sigma(\mathcal{G}, \mathcal{H})$-measurable. Let $A \in \sigma(\mathcal{G}, \mathcal{H})$. We have to show that $\mathbb{E}[X1(A)] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1(A)]$. By writing $X = X^+ - X^-$, we may assume $X \geq 0$. Define the measures μ, ν by $\mu(A) = \mathbb{E}[X1(A)]$ and $\nu(A) = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1(A)]$ for $A \in \sigma(\mathcal{G}, \mathcal{H})$. Let $\mathcal{A} = \{A \cap B \mid A \in \mathcal{G}, B \in \mathcal{H}\}$. Then $\mathcal{A}$ is a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$. By the theorem on uniqueness of extension, it is enough to check that $\mu = \nu$ on $\mathcal{A}$ and $\mu(\Omega) = \nu(\Omega) < \infty$. The latter is immediate from the integrability of X. Let $A \cap B \in \mathcal{A}$ where $A \in \mathcal{G}, B \in \mathcal{H}$. Then $\mathbb{E}[X1(A \cap B)] = \mathbb{E}[(X1(A))1(B)]$. Now note that

$X1(A)$ is $\sigma(X, \mathcal{G})$-measurable, so $\mathbb{E}[(X1(A))1(B)] = \mathbb{E}[X1(A)]\mathbb{P}(B)$ since $\sigma(X, \mathcal{H})$ and $\mathcal{H}$ are independent. Then using the same reasoning again:

$$\mu(A \cap B) = \mathbb{E}[X1(A)]\mathbb{P}(B) = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1(A)]\mathbb{P}(B) = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1(A \cap B)] = \nu(A \cap B).$$

□

Remark. If we only required that $\mathcal{H}$ is independent of X and independent of $\mathcal{G}$, then the statement would be false.

1.4 Examples

1.4.1 Gaussian Distribution

Definition. *A random vector $(X_1, \ldots, X_n)$ with values in $\mathbb{R}^n$ is called a* Gaussian *if for all $a_1, \ldots, a_n \in \mathbb{R}$, $\sum_{i=1}^n a_i X_i$ has a Gaussian distribution.*

Let (X, Y) be a Gaussian vector in $\mathbb{R}^2$. We want to find $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$. We know that $\mathbb{E}[X \mid Y] = f(Y)$ a.s. for some function f as $\mathbb{E}[X \mid Y]$ is Y-measurable. We try $f(Y) = aY + b$ for some $a, b \in \mathbb{R}$ to be determined. Then $f(Y)$ is certainly Y-measurable. Letting $A = \Omega$ in the definition of conditional expectation we get the condition

$$\mathbb{E}[X] = a\mathbb{E}[Y] + b.$$

We also must have $\mathbb{E}[XY] = \mathbb{E}[f(Y)Y]$ (first note that $\mathbb{E}[X(Y \wedge n)] = \mathbb{E}[f(Y)(Y \wedge n)]$ and the get it from dominated convergence). So $\mathrm{Cov}(X - f(Y), Y) = \mathbb{E}[(X - f(Y))Y] = 0$, i.e.

$$\mathrm{Cov}(X, Y) = a\mathrm{Var}(Y).$$

These two conditions determine a, b uniquely. We now have to check that for these values we indeed have $aY + b = f(Y) = \mathbb{E}[X \mid Y]$. Let Z be bounded and Y-measurable. Since $X - aY - b, Y$ are jointly Gaussian and $\mathrm{Cov}(X - aY - b, Y) = 0$, we get that $X - f(Y)$ is independent from Y, hence also independent from Z and thus $\mathbb{E}[(X - f(Y))Z] = \mathbb{E}[X - f(Y)]\mathbb{E}[Z] = 0$. Therefore $\mathbb{E}[XZ] = \mathbb{E}[f(Y)Z]$ and so indeed $\mathbb{E}[X \mid Y] = f(Y)$.

1.4.2 Conditional Density

Let (X, Y) be a random variable in $\mathbb{R}^2$ with density $f_{X,Y}(x, y)$. If $h : \mathbb{R} \to \mathbb{R}$ is Borel such that $h(X)$ is integrable, we want to find $\mathbb{E}[h(X) \mid Y]$. Again, this is Y-measurable, so $\mathbb{E}[h(X) \mid Y] = \varphi(Y)$ for some function φ. Let g be a bounded measurable function. We want to determine φ such that $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[\varphi(Y)g(Y)]$. We have

$$\mathbb{E}[\varphi(Y)g(Y)] = \int_{\mathbb{R}} \varphi(y)g(y)f_Y(y)dy,$$

where f_Y is the density of Y, and

$$\begin{aligned}\mathbb{E}[h(X)g(Y)] &= \int_{\mathbb{R}}\int_{\mathbb{R}} h(X)g(y)f_{X,Y}(x,y)dxdy \\ &= \int_{\mathbb{R}} g(y)f_Y(y)\Big(\int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}dx\Big)dy\end{aligned}$$

where we set $0/0 = 0$. Hence we define $\varphi(y) := \int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}dx$. We let $f_{X|Y}(x \mid y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$. Thus, for this φ we have $\mathbb{E}[h(X) \mid Y] = \varphi(Y)$.

2 Discrete-time Martingales

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable space $(E, \mathcal{E})$. Some preliminary definitions:

A sequence $(X_n)_{n\geq 0}$ of random variables (on Ω and taking values in E) is called a *stochastic process.*

A *filtration* is an increasing family $(\mathcal{F}_n)_{n\geq 0}$ of sub-σ-algebras of $\mathcal{F}$.

Given a process $X = (X_n)_{n\geq 0}$, its *natural filtration* is defined to be $\mathcal{F}_n^X = \sigma(X_k \mid k \leq n)$.

A stochastic process $(X_n)_{n\geq 0}$ is called *adapted* to the filtration $(\mathcal{F}_n)_{n\geq 0}$ if X_n is $\mathcal{F}_n$-measurable for all n. Clearly, $(X_n)_{n\geq 0}$ is adapted to its natural filtration.

$X = (X_n)_{n\geq 0}$ is called *integrable* if X_n is integrable for all n.

Definition. *Let $(\mathcal{F}_n)_{n\geq 0}$ be a flitration on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $X = (X_n)_{n\geq 0}$ be an adapted and integrable process. X is a*

- martingale *if $\mathbb{E}[X_n \mid \mathcal{F}_m] = X_m$ a.s. for all $n \geq m$.*
- supermartingale *if $\mathbb{E}[X_n \mid \mathcal{F}_m] \leq X_m$ a.s. for all $n \geq m$.*
- submartingale *if $\mathbb{E}[X_n \mid \mathcal{F}_m] \geq X_m$ a.s. for all $n \geq m$.*

Remark. It follows from the tower property that if X is a martingale (or super-, sub-) w.r.t. some filtration, then it is also a martingale (super, sub) w.r.t. its natural filtration $(\mathcal{F}_n^X)$.

Example. Let $(\xi_i)_i$ be i.i.d. random variables with $\mathbb{E}[\xi_i] = 0$ for all i. Let $X_n = \sum_{i=1}^n \xi_i$, $X_0 = 0$. Then X is a martingale. For example the simple random walk on $\mathbb{Z}$ is of this form.

Example. Let $(\xi_i)_i$ be i.i.d. random variables with $\mathbb{E}[\xi_i] = 1$ for all i. Let $X_n = \prod_{i=1}^n \xi_i$, $X_0 = 1$. Then X_n is a martingale.

2.1 Stopping Times

Definition. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ be a filtered probability space. A random variable $T : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called a* stopping time *if for all $n \in \mathbb{N}$, $\{T \leq n\} \in \mathcal{F}_n$.*

Equivalently, for all $n \in \mathbb{N}$, $\{T = n\} \in \mathcal{F}_n$.

Examples.

- Constant times $T = n$ are stopping times.
- Let $X = (X_n)_{n\geq 0}$ be an adapted stochastic process. Let $A \in \mathcal{B}(\mathbb{R})$. $T = \inf\{n \geq 0 \mid X_n \in A\}$. Then the *entrance time* $T = \inf\{n \geq 0 \mid X_n \in A\}$ is a stopping time. Indeed, $\{T \leq n\} = \bigcup_{k=0}^{n} \underbrace{\{X_k \in A\}}_{\in \mathcal{F}_k} \in \mathcal{F}_n$.
- In the situation of the previous example, $L_A = \sup\{n \leq 100 \mid X_n \in A\}$ is (in general) not a stopping time.

Proposition 2.1. *Let $S, T, (T_n)_n$ be stopping times. Then $S \vee T, S \wedge T, \inf T_n$, $\sup T_n, \liminf T_n, \limsup T_n$ are also stopping times.*

Definition. *Let T be a stopping time. We define*

$$\mathcal{F}_T = \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_t \,\forall t\}.$$

Remark. If T is a constant stopping time $T = t$, then $\mathcal{F}_T = \mathcal{F}_t$.

Given a stopping time T and a process X_n, we define X_T by $X_T(\omega) = X_{T(\omega)}(\omega)$ when $T(\omega) < \infty$. We define the *stopped process* X^T by $X_t^T = X_{T\wedge t}$.

Proposition 2.2. *Let S, T be stopping times and $X = (X_n)_{n\geq 0}$ an adapted process. Then*

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

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

(c) X^T is adapted.

(d) If X is integrable, then X^T is integrable.

Proof.

(a) Immediate.

(b) Let $A \in \mathcal{B}(\mathbb{R})$. We have to show $\{X_T 1(T < \infty) \in A\} \cap \{T \leq t\} \in \mathcal{F}_t$ for all t. We have

$$\{X_T 1(T < \infty) \in A\} \cap \{T \leq t\} = \bigcup_{s=0}^{t} \underbrace{\{X_s \in A\}}_{\in \mathcal{F}_s \subseteq \mathcal{F}_t} \cap \underbrace{\{T = s\}}_{\in \mathcal{F}_s} \in \mathcal{F}_t.$$

(c) $X_t^T = X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable by (b). From (a), $\mathcal{F}_{T\wedge t} \subseteq \mathcal{F}_t$, so X_t^T is $\mathcal{F}_t$-measurable.

(d)

$$\mathbb{E}[|X_{T\wedge t}|] = \sum_{s=0}^{t-1} \mathbb{E}[|X_s| 1(T = s)] + \mathbb{E}[|X_t| 1(T > t-1)] \leq \sum_{s=0}^{t} \mathbb{E}[|X_s|] < \infty.$$

□

Theorem 2.3 (Optional Stopping Theorem (OST)). *Let $X = (X_n)_{n\geq 0}$ be a martingale.*

(1) If T is a stopping time, then X^T is also a martingale and $\mathbb{E}[X_{T\wedge t}] = \mathbb{E}[X_0]$ for all t.

(2) If $S \leq T \leq n$ are stopping times, then $\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S$ a.s. and $\mathbb{E}[X_T] = \mathbb{E}[X_S]$.

(3) Let Y be an integrable random variable and $|X_n| \leq Y$ for all n. Let T be a stopping time with $\mathbb{P}(T < \infty) = 1$. Then $\mathbb{E}[X_T] = \mathbb{E}[X_n]$

(4) Suppose there exists M such that $|X_{n+1} - X_n| \leq M$ for all n a.s. (i.e. X has bounded increments*) and T is a stopping time with $\mathbb{E}[T] < \infty$. Then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.*

Proof.

(1) By Proposition 2.2, X^T is again an integrable adapted process. By repeated use the tower property it suffices to prove $\mathbb{E}[X_{T\wedge t} \mid \mathcal{F}_{t-1}] = X_{T\wedge(t-1)}$ a.s. We have

$$\begin{aligned}
\mathbb{E}[X_{T\wedge t} \mid \mathcal{F}_{t-1}] &= \mathbb{E}\left[\sum_{s=0}^{t-1} \underbrace{X_s 1(T = s)}_{\mathcal{F}_{t-1}\text{-measurable}} \mid \mathcal{F}_{t-1}\right] + \mathbb{E}\left[X_t \underbrace{1(T > t-1)}_{\mathcal{F}_{t-1}\text{-measurable}} \mid \mathcal{F}_{t-1}\right] \\
&= \sum_{s=0}^{t-1} X_s 1(T = s) + 1(T > t-1)\mathbb{E}\left[X_t \mid \mathcal{F}_{t-1}\right] \\
&= \sum_{s=0}^{t-1} X_s 1(T = s) + 1(T > t-1) X_{t-1} \\
&= X_{T\wedge(t-1)}.
\end{aligned}$$

(2) Let $A \in \mathcal{F}_S$. We need to show that $\mathbb{E}[X_T 1(A)] = \mathbb{E}[X_S 1(A)]$. We have

$$\begin{aligned}
X_T &= (X_T - X_{T-1}) + \cdots + (X_{S+1} - X_S) + X_S \\
&= X_S + \sum_{k=0}^{n} (X_{k+1} - X_k) 1(S \leq k < T)
\end{aligned}$$

So

$$\mathbb{E}[X_T 1(A)] = \mathbb{E}[X_S 1(A)] + \sum_{k=0}^{n} \mathbb{E}[(X_{k+1} - X_k) 1(S \leq k < T) 1(A)]$$

Now note that $1(A) \cdot 1(S \leq k)$ is $\mathcal{F}_k$ measurable and $\{T > k\} \in \mathcal{F}_k$. So $1(A)1(S \leq k < T)$ is $\mathcal{F}_k$-measurable. Since $\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] = X_k$, we then get $\mathbb{E}[(X_{k+1} - X_k)1(S \leq k < T)1(A)] = 0$ and hence $\mathbb{E}[X_T 1(A)] = \mathbb{E}[X_S 1(A)]$.

(3) Exercise.

(4) Exercise.

□

Proposition 2.4. *Let X be a positive supermartingale and let T be a stopping time with $\mathbb{P}(T < \infty) = 1$. Then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

Proof. The previous theorem also holds for supermartingale if we replace equalities by suitable inequalities. Then apply Fatou's lemma. □

Example (Simple 1D random walk)**.** Let $(\xi_i)_i$ be i.i.d. random variables taking ± 1 equally likely. Let $X_0 = 0$, and $X_n = \sum_{i=1}^n \xi_i$. Take $T_x = \inf\{n \geq 0 \mid X_n = x\}$. T_1 is a stopping time with $\mathbb{P}(T_1 < \infty) = 1$. For all t, $\mathbb{E}[X_{T_1 \wedge t}] = \mathbb{E}[X_0] = 0$ by the theorem. But $\mathbb{E}[X_{T_1}] = 1$, so we can't (in general) get rid of the $\wedge t$ in part (1) of the theorem.

2.2 Gambler's Ruin

Let $a, b > 0$ be integers. Consider again the simple random walk X as in the last example. We let $T = T_{-a} \wedge T_b$. We want to determine $\mathbb{P}(T_{-a} < T_b)$, i.e. the probability that we reach $-a$ before b.

Note that $|X_{n+1} - X_n| = 1$. We need to check that $\mathbb{E}[T] < \infty$ in order to apply (4) of the OST. Then we would get $\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0$. On the other hand, we have

$$\mathbb{E}[X_T] = -a\mathbb{P}(T_{-a} < T_b) + b\mathbb{P}(T_b < T_{-a}).$$

Since also $\mathbb{P}(T_{-a} < T_b) + \mathbb{P}(T_b < T_{-a}) = 1$, we get

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

We only have to establish $\mathbb{E}[T] < \infty$. For this, note that if $\xi_i = 1$ for $a + b$ consecutive i with n being the last such i, then $T < n$. Hence, we can bound T from above by the first time there are $a + b$ consecutive i's with $\xi_i = 1$. We can further bound this by only considering blocks $1, \dots, a + b; a + b + 1, \dots, 2(a + b); \dots$. Note if $I \subseteq \mathbb{N}_0$ consists of $a + b$ consecutive indices, then $\mathbb{P}(\xi_i = 1 \text{ for all } i \in I) = 2^{-(a+b)} =: p$, hence

$$\mathbb{E}[T] \leq \sum_{k=1} k(a + p)(1 - p^{k-1})p = (a + b)2^{a+b}.$$

2.3 Martingale Convergence Theorem

Theorem 2.5. *Let X be a supermartingale bounded in $\mathcal{L}^1$, i.e. $\sup_{n \geq 0} \mathbb{E}[|X_n|] < \infty$. Then $X_n \to X_\infty$ a.s. as $n \to \infty$ for some $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$ where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geq 0)$.*

Corollary 2.6. *Let X be a positive supermartingale. Then X converges a.s. to an integrable limit.*

Proof. $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0]$ for all n. So the theorem applies. □

Let $x = (x_n)_{n\geq 0}$ be a real sequence. Let $a < b$. Set $T_0(x) = 0$ and inductively define

$$S_{k+1}(x) = \inf\{n \geq T_k(x) \mid x_n \leq a\},$$
$$T_{k+1}(x) = \inf\{n \geq S_{k+1}(x) \mid x_n \geq b\}.$$

Then define $N_n([a,b],x) = \sup\{k \geq 0 \mid T_k(x) \leq n\}$. It is the number of "upcrossings" up to time n. Let $N([a,b],x) = \sup\{k \geq 0 \mid T_k(x) < \infty\} = \sup_n N_n([a,b],x)$. It is the total number of upcrossings.

Lemma 2.7. *A real sequence $x = (x_n)_n$ converges in $\mathbb{R} \cup \{\pm\infty\}$ if and only if for all $a, b \in \mathbb{Q}$ with $a < b$,*

$$N([a,b],x) < \infty.$$

Proof. "$\Rightarrow$" Suppose there exist $a, b \in \mathbb{Q}$ with $a < b$ and $N([a,b],x) = \infty$. Then $\liminf x_n \leq a < b \leq \limsup x_n$, so x is not convergent.

"$\Leftarrow$" If x does not converge, then $\liminf x_n < \limsup x_n$. Thus we find $a, b \in \mathbb{Q}$ s.t. $\liminf x_n < a < b < \limsup x_n$. Then it is easily seen that $N([a,b],x) = \infty$. □

Theorem 2.8 (Doob's Upcrossing Inequality)**.** *Let X be a supermartingale and $a < b$ with $a, b \in \mathbb{R}$. Then*

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

Proof of Theorem 2.5. We want to apply the lemma above. By the upcrossing inequality,

$$(b-a)\mathbb{E}[N_n([a,b],X)] \leq \mathbb{E}[(X_n - a)^-] \leq |a| + \sup_{n\geq 0} \mathbb{E}[|X_n|] =: C.$$

As $N_n([a,b],X) \nearrow N([a,b],X)$ as $n \to \infty$, we get from monotone convergence

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

So $N([a,b],X) < \infty$ a.s. Let $\Omega_0 = \bigcap_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{N([a,b],X) < \infty\}$. We get $\mathbb{P}(\Omega_0) = 1$. By Lemma 2.7, we can define

$$X_\infty = \begin{cases} \lim_{n\to\infty} X_n & \text{on } \Omega_0, \\ 0 & \text{on } \Omega_0^c. \end{cases}$$

Then $X_n \to X_\infty$ a.s. as $n \to \infty$ and X_∞ is $\mathcal{F}_\infty$-measurable. Also $\mathbb{E}[|X_\infty|] = \mathbb{E}[\liminf |X_n|] \leq \liminf \mathbb{E}[|X_n|] < \infty$, so $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$. □

Proof of Theorem 2.8. Write T_k, S_k, N for $T_k(X), S_k(X), N_n([a,b],X)$. One easily sees that T_k, S_k are stopping times. By definition of T_k, S_k, we have $X_{T_k} - X_{S_k} \geq b-a$. We have

$$\sum_{k=1}^{n}(X_{T_k\wedge n} - X_{S_k\wedge n}) = \sum_{k=1}^{N}(X_{T_k} - X_{S_k}) + \sum_{k=N+1}^{n}(X_n - X_{S_k\wedge n})1(N<n).$$

Note that $S_{N+2} > n$ by definition, so this is

$$\sum_{k=1}^{N}(X_{T_k} - X_{S_k}) + (X_n - X_{S_{N+1}})1(S_{N+1} \leq n).$$

Since $S_k \wedge n \leq T_k \wedge n$, we get from the supermartingale version of the OST that $\mathbb{E}[X_{S_k\wedge n}] \geq \mathbb{E}[X_{T_k\wedge n}]$. Then

$$\begin{aligned} 0 \geq \mathbb{E}\left[\sum_{k=1}^{n}(X_{T_k\wedge n} - X_{S_k\wedge n})\right] &= \mathbb{E}\left[\sum_{k=1}^{N}(X_{T_k} - X_{S_k}) + (X_n - X_{S_{N+1}})1(S_{N+1}\leq n)\right] \\ &\geq (b-a)\mathbb{E}[N] + \mathbb{E}[(X_n - X_{S_{N+1}})1(S_{N+1}\leq n)] \\ &\geq (b-a)\mathbb{E}[N] - \mathbb{E}[(X_n-a)^-]. \end{aligned}$$

□

2.4 Doob's Inequalities

Theorem 2.9 (Doob's maximal inequality)**.** *Let X be a nonnegative submartingale. Let $X_n^* = \sup_{0\leq k\leq n} X_k$. For $\lambda \geq 0$ we have*

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

Proof. Define the stopping time $T = \inf\{k \geq 0 \mid X_k \geq \lambda\}$. The OST applied to $T\wedge n \leq n$ gives

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T 1(T\leq n)] + \mathbb{E}[X_n 1(T>n)]$$

Hence $\mathbb{E}[X_T 1(T\leq n)] \leq \mathbb{E}[X_n 1(T\leq n)]$. Now note that $\{T\leq n\} = \{X_n^* \geq \lambda\}$ and we get

$$\lambda\mathbb{P}(X_n^*\geq\lambda) \leq \mathbb{E}[X_T 1(T\leq n) = \mathbb{E}[X_n 1(X_n^*\geq\lambda)] \leq \mathbb{E}[X_n].$$

□

Theorem 2.10 (Doob's $\mathcal{L}^p$ inequality)**.** *Let X be a martingale or a nonnegative submartingale. Let $X_n^* = \sup_{0\leq k\leq n}|X_k|$. For $p>1$, we have*

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

Proof. It suffices to prove the second case, since by Jensen, if X is a martingale, then $|X|$ is a nonnegative submartingale. For $k \in \mathbb{N}$ we have

$$\begin{aligned}
\mathbb{E}[(X_n^* \wedge k)^p] &= \mathbb{E}\left[\int_0^k p x^{p-1} 1(X_n^* \geq x) dx\right] = \int_\Omega \int_0^k p x^{p-1} 1(X_n^*(\omega) \geq x) dx d\mathbb{P} \\
&= \int_0^k p x^{p-1} \mathbb{P}(X_n \geq x) dx \\
&\overset{\text{Theorem 2.9}}{\leq} \int_0^k p x^{p-2} \mathbb{E}[X_n 1(X_n^* \geq x)] dx \\
&= \int_0^k p x^{p-2} \int_\Omega X_n(\omega) 1(X_n^*(\omega) \geq x) d\mathbb{P} dx \\
&= \int_\Omega X_n(\omega)(X_n^*(\omega) \wedge k)^{p-1} \frac{p}{p-1} d\mathbb{P} \\
&= \frac{p}{p-1} \mathbb{E}[X_n (X_n^* \wedge k)^{p-1}] \\
&\overset{\text{Hölder}}{\leq} \frac{p}{p-1} \|X_n\|_p \|X_n^* \wedge k\|_p^{p-1}.
\end{aligned}$$

Rearranging gives $\|X_n^* \wedge k\|_p \leq \frac{p}{p-1} \|X_n\|_p$. Now let $k \to \infty$. □

2.5 L^p-convergence

Theorem 2.11. *Let X be a martingale, $p > 1$. TFAE:*

1. *X is bounded in $\mathcal{L}^p$,*
2. *X converges a.s. and in $\mathcal{L}^p$ to some $X_\infty \in \mathcal{L}^p$.*
3. *There exists a random variable $Z \in \mathcal{L}^p$ such that $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ a.s.*

Proof. “1. ⇒ 2.” If X is bounded in $\mathcal{L}^p$, then also in $\mathcal{L}^1$ (e.g. by Jensen or Hölder). So by Theorem 2.5 there exists $X_\infty \in \mathcal{L}^p$ such that $X_n \to X_\infty$ a.s. By Fatou, $\mathbb{E}[|X_\infty|^p] \leq \liminf \mathbb{E}[|X_n|^p] < \infty$ since X is bounded in $\mathcal{L}^p$. So $X_\infty \in \mathcal{L}^p$. By Doob's $\mathcal{L}^p$-inequality we have $\|X_n^*\|_p \leq \frac{p}{p-1} \|X_n\|_p$. Letting $n \to \infty$, we get $\|X_\infty^*\|_p \leq \frac{p}{p-1} \sup_n \|X_n\|_p$ where $X_\infty^* = \sup_{0 \leq k} |X_k|$. So $X_\infty^* \in \mathcal{L}^p$. Since $|X_n - X_\infty| \leq 2X_\infty^*$, we get $X_n \to X_\infty$ in $\mathcal{L}^p$ by dominated convergence.

“2. ⇒ 3.” Set $Z = X_\infty$. We have $Z \in \mathcal{L}^p$. For $m \geq n$, we have

$$\|X_n - \mathbb{E}[X_\infty \mid \mathcal{F}_n]\|_p = \|\mathbb{E}[X_m - X_\infty \mid \mathcal{F} - n]\|_p \leq \|X_m - X_\infty\|_p.$$

Letting $m \to \infty$ this goes to 0, so $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ a.s.

“3. ⇒ 1.” Jensen. □

A martingale of the form $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ for some $Z \in \mathcal{L}^p$ is called a *martingale closed in* $\mathcal{L}^p$.

Corollary 2.12. *Let $p > 1$, $Z \in \mathcal{L}^p$, $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$. Then $X_n \to X_\infty$ a.s. and in $\mathcal{L}^p$ where $X_\infty = \mathbb{E}[Z \mid \mathcal{F}_\infty]$ a.s. and $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$.*

Proof. We know from the theorem that $X_n \to X_\infty$ a.s. and in $\mathcal{L}^p$ for some $X_\infty \in \mathcal{L}^\infty$. We have to show $X_\infty = \mathbb{E}[Z \mid \mathcal{F}_\infty]$ a.s. We have to show $\mathbb{E}[X_\infty 1(A)] = \mathbb{E}[Z1(A)]$ for all $A \in \mathcal{F}_\infty$. Since $\bigcup_n \mathcal{F}_n$ is a π-system generating $\mathcal{F}_\infty$, it suffices to prove this for $A \in \mathcal{F}_N$ for some N. Then $\mathbb{E}[Z1(A)] = \mathbb{E}[X_N 1(A)] = \mathbb{E}[X_n 1(A)]$ for all $n \geq N$. We have $\mathbb{E}[X_n 1(A)] \to \mathbb{E}[X_\infty 1(A)]$ as $n \to \infty$. Then $\mathbb{E}[Z1(A)] = \mathbb{E}[X_\infty 1(A)]$. □

For $p = 1$, we need another condition.

Definition. *A collection of random variables $(X_i)_{i \in I}$ is called* uniformly integrable *(UI) if* $\sup_{i \in I} \mathbb{E}[|X_i| 1(|X_i| > \alpha)] \to 0$ *as* $\alpha \to \infty$.

Equivalently, $(X_i)_{i \in I}$ is UI if $(X_i)_{i \in I}$ if $(X_i)_{i \in I}$ is uniformly bounded in $\mathcal{L}^1$ and for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $\mathbb{P}(A) < \delta$, then $\sup_{i \in I} \mathbb{E}[|X_i| 1(A)] \leq \varepsilon$.

Lemma 2.13. *Given $X_n, X \in \mathcal{L}^1$ for $n \geq 1$, we have*

$$X_n \xrightarrow{\mathcal{L}^1} \mathcal{L} \iff \begin{cases} X_n \to X \text{ in probability} & \text{and} \\ (X_n) \text{ is uniformly integrable} \end{cases}$$

Proof. See undergraduate probability lecture notes or book by Williams. □

Note: If (X_n) is bounded in $\mathcal{L}^p$ with $p > 1$, then it is uniformly integrable.

Theorem 2.14. *Suppose $X \in \mathcal{L}^1$. Then the family*

$$\{\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G} \text{ is a sub-}\sigma\text{-algebra of } \mathcal{F}\}$$

us uniformly integrable.

Proof. We have to show that for all $\varepsilon > 0$ there exists K_0 such that for all $K \geq K_0$,

$$\sup_{\mathcal{G}} \mathbb{E}[|\mathbb{E}[X \mid \mathcal{G}]| 1(|\mathbb{E}[X \mid \mathcal{G}]| > K)] \leq \varepsilon.$$

We have

$$\begin{aligned} \mathbb{E}[|\mathbb{E}[X \mid \mathcal{G}]| 1(|\mathbb{E}[X \mid \mathcal{G}]| > K) &\leq \mathbb{E}[\mathbb{E}[|X| \mid \mathcal{G}] 1(\underbrace{|\mathbb{E}[X \mid \mathcal{G}]| \geq K}_{\mathcal{G}\text{-measurable}}) \\ &\leq \mathbb{E}[|X| \cdot 1(|\mathbb{E}[X \mid \mathcal{G}]| \geq K)] \end{aligned}$$

If $X \in \mathcal{L}^1$, it is easy to check that for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $\mathbb{P}(A) < \delta$, then $\mathbb{E}[|X|\mathbb{P}(A)] \leq \varepsilon$. We have

$$\mathbb{P}(|\mathbb{E}[X \mid \mathcal{G}]| \geq K) \leq \frac{1}{K}\mathbb{E}[|X|]$$

by Markov. Thus we take $K_0 = \frac{\mathbb{E}[|X|]}{\delta}$. Then for all $K \geq K_0$, we have $\mathbb{P}(|\mathbb{E}[X \mid \mathcal{G}]| \geq K) \leq \delta$ and so

$$\begin{aligned}\mathbb{E}[|\mathbb{E}[X \mid \mathcal{G}]| \cdot 1(|\mathbb{E}[X \mid \mathcal{G}]| \geq K)] &\leq \mathbb{E}[|X| \cdot 1(|\mathbb{E}[X \mid \mathcal{G}]| \geq K)] \\ &\leq \varepsilon.\end{aligned}$$

□

Definition. $(X_n)_{n\geq 0}$ *is called a* UI (uniformly integrable) martingale *if it is a martingale and* $(X_n)_{n\geq 0}$ *is UI.*

Theorem 2.15. *Let X be a martingale. TFAE:*

(1) X is UI.

(2) (X_n) converges a.s. and in $\mathcal{L}^1$ to some X_∞.

(3) There exists $Z \in \mathcal{L}^1$ such that $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ a.s. for all $n \geq 0$.

Proof. "(1) ⇒ (2)" Since X is bounded in $\mathcal{L}^1$, by the martingale convergence theorem we get that $X_n \to X_\infty$ a.s. as $n \to \infty$ for some $X_\infty \in \mathcal{L}^1$. Then also $X_n \to X$ in probability, hence in $\mathcal{L}^1$ by the lemma above.

"(2) ⇒ (3)" Set $Z = X_\infty$. We have to show $X_n = \mathbb{E}[X_\infty \mid \mathcal{F}_n]$. The same proof as in Theorem 2.11 works.

"(3) ⇒ (1)" Theorem 2.14. □

Remark. As before, if $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ with $Z \in \mathcal{L}^1$, then $X_\infty = \mathbb{E}[Z \mid \mathcal{F}_\infty]$ a.s.

Example. Let $X_1, X_2, \ldots$ be i.i.d. such that $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 2) = \frac{1}{2}$. Take $Y_n = X_1 X_2 \cdots X_n$. Then (Y_n) is a martingale. Then $Y_n \to 0$ a.s. as $n \to \infty$. But $\mathbb{E}[Y_n] = 1$ for all n, so it does not converge in $\mathcal{L}^1$ and it is not a UI martingale.

Let T be a stopping time and X a UI martingale. Then we can define

$$X_T = \sum_{n=0}^{\infty} X_n 1(T = n) + X_\infty 1(T = \infty)$$

Theorem 2.16. *Let X be a UI martingale and $S \leq T$ be stopping times. Then*

$$\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S \text{ a.s.}$$

Proof. X_n converges to X_∞ a.s. and in $\mathcal{L}^1$. It suffices to prove that $X_T = \mathbb{E}[X_\infty \mid \mathcal{F}_T]$ a.s. Indeed, then we would get (applying this also to S)

$$\mathbb{E}[X_T \mid \mathcal{F}_S] = \mathbb{E}[\mathbb{E}[X_\infty \mid \mathcal{F}_T] \mid \mathcal{F}_S] = \mathbb{E}[X_\infty \mid \mathcal{F}_S] = X_S.$$

First note that X_T is in $\mathcal{L}^1$. Indeed, from $X_n = \mathbb{E}[X_\infty \mid \mathcal{F}_n]$ we get $|X_n| \le \mathbb{E}[|X_\infty| \mid \mathcal{F}_n]$ and then

$$\begin{aligned}
\mathbb{E}[|X_T|] &= \sum_{n=0}^{\infty} \mathbb{E}[|X_n| \cdot 1(T=n)] + \mathbb{E}[|X_\infty| \cdot 1(T=\infty)] \\
&\le \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{E}[|X_\infty| \mid \mathcal{F}_n] \cdot 1(T=n)] + \mathbb{E}[|X_\infty| \cdot 1(T=\infty)] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[|X_\infty| \cdot 1(T=n)] + \mathbb{E}[|X_\infty| \cdot 1(T=\infty)] \\
&= \mathbb{E}[|X_\infty|].
\end{aligned}$$

Let $A \in \mathcal{F}_T$. We need to show $\mathbb{E}[X_\infty 1(A)] = \mathbb{E}[X_T 1(A)]$. Since $A \in \mathcal{F}_T$, we have $\{T = n\} \cap A \in \mathcal{F}_n$ for all n, and then

$$\begin{aligned}
\mathbb{E}[X_T 1(A)] &= \sum_{n=0}^{\infty} \mathbb{E}[X_n \cdot 1(T=n)1(A)] + \mathbb{E}[X_\infty \cdot 1(T=\infty)1(A)] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[X_\infty \cdot 1(T=n)1(A)] + \mathbb{E}[X_\infty \cdot 1(T=\infty)1(A)] \\
&= \mathbb{E}[X_\infty 1(A)].
\end{aligned}$$

□

2.6 Backwards martingales

Suppose $\ldots \supseteq \mathcal{G}_{-2} \subseteq \mathcal{G}_{-1} \subseteq \mathcal{G}_0$ is a decreasing family of sub-σ-algebras. We call $(X_n)_{n\le 0}$ a *backwards martingale* if $X_0 \in \mathcal{L}^1$ and (X_n) is adapted to (G_n) and $\mathbb{E}[X_{n+1} \mid \mathcal{G}_n] = X_n$ a.s. for all $n \le -1$. Note that in this case $X_n = \mathbb{E}[X_0 \mid \mathcal{G}_0]$ a.s. for all $n \le -1$, so (X_n) is UI by Theorem 2.14.

Theorem 2.17 (Backwards Martingale Convergence Theorem)**.** *Let $X_0 \in \mathcal{L}^p$, for some $p \ge 1$. Then (X_n) converges a.s. and in $\mathcal{L}^p$ as $n \to -\infty$ to a random variable $X_{-\infty}$ which satisfies $X_{-\infty} = \mathbb{E}[X_0 \mid \mathcal{G}_{-\infty}]$ a.s. where $G_{-\infty} = \bigcap_{n\le 0} \mathcal{G}_N$.*

Proof. Let $a < b$ and $N_{-n}([a,b], X)$ be the number of upcrossings of $[a,b]$ between times $-n$ and 0. Set $\mathcal{F}_k = \mathcal{G}_{-n+k}$ for $0 \le k \le n$. This is an increasing filtration and $(X_{-n+k})_{0\le k\le n}$ is a martingale w.r.t. $(\mathcal{F}_k)$. Then $N_{-n}([a,b], X)$ is the number of upcrossings of $[a,b]$

by $(X_{-n+k})_{0\le k\le n}$ between times 0 and n. Doob's upcrossing inequality for martingales, Theorem 2.8, gives

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

Take the limit as $n\to\infty$ and use monotone convergence to get $(b-a)\mathbb{E}[N_{-\infty}([a,b],X)] < \infty$. So in particular, $N_{-\infty}([a,b],X)<\infty$ a.s. and so like before we get that X_n converges a.s. as $n\to-\infty$ to some $X_{-\infty}$. Then $X_{-\infty}$ is $\mathcal{G}_{-\infty}$-measurable since the $\mathcal{G}_n$ are decreasing. Since $X_0\in\mathcal{L}^p$, we get $X_n\in\mathcal{L}^p$ for all n by Jensen. Then $X_{-\infty}\in\mathcal{L}^p$ by Fatou. Next we show $\mathcal{L}^p$-convergence. We have

$$|X_n - X_{-\infty}|^p = |\mathbb{E}[X_n \mid \mathcal{G}_n] - \mathbb{E}[X_{-\infty}\mid \mathcal{G}_n]|^p \le \mathbb{E}[|X_0-X_{-\infty}|^p \mid \mathcal{G}_n].$$

By Theorem 2.14 this is a UI family. Then also $|X_n - X_{-\infty}|^p$ is UI and since it convergences to 0 a.s., it also converges in $\mathcal{L}^1$ by Lemma 2.13. So $X_n\to X_{-\infty}$ in $\mathcal{L}^p$ as $n\to-\infty$.

Finally, we have to show $X_{-\infty}=\mathbb{E}[X_0\mid\mathcal{G}_\infty]$ a.s. We prove this using the definition of conditional expectation. Let $A\in\mathcal{G}_\infty$. Then $A\in\mathcal{G}_n$ and so $\mathbb{E}[X_0 1(A)]=\mathbb{E}[X_n 1(A)]$ for all $n\le -1$. Since $X_n\to X_{-\infty}$ in $\mathcal{L}^1$, we get $\mathbb{E}[X_0 1(A)] = \mathbb{E}[X_{-\infty}1(A)]$. □

2.7 Applications

2.7.1 Kolmogorov's 0-1 Law

Theorem 2.18 (Kolmogorov's 0-1 law)**.** *Let $X_1, X_2, \dots$ be independent random variables and define $\mathcal{F}_n = \sigma(X_k : k\ge n)$ and $\mathcal{F}_\infty = \bigcap_{n\ge 0}\mathcal{F}_n$ (the* tail σ-algebra*). Then $\mathcal{F}_\infty$ is trivial, i.e. for all $A\in\mathcal{F}_\infty$, $\mathbb{P}(A)\in\{0,1\}$.*

Proof. Let $\mathcal{G}_n=\sigma(X_k : k\le n)$. Let $A\in\mathcal{F}_\infty$. Consider $\mathbb{E}[1(A)\mid\mathcal{G}_n]$. This is a martingale (w.r.t. $\mathcal{G}_n$) and as $n\to\infty$ it converges a.s. to $\mathbb{E}[1(A)\mid\mathcal{G}_\infty]$ where $\mathcal{G}_\infty=\sigma(\mathcal{G}_n : n\ge 0)$. Clearly, $\mathcal{F}_\infty\subseteq\mathcal{G}_\infty$, so $A\in\mathcal{G}_\infty$ and then $\mathbb{E}[1(A)\mid\mathcal{G}_\infty]=1(A)$ a.s. Since $A\in\mathcal{F}_\infty$, $A\in\mathcal{F}_{n+1}=\sigma(X_k : k\ge n+1)$. Now $\mathcal{F}_{n+1}$ is independent from $\mathcal{G}_n$ by assumption. Hence $\mathbb{E}[1(A)\mid\mathcal{G}_n]=\mathbb{E}[1(A)]=\mathbb{P}(A)$ a.s. So $\mathbb{P}(A)=1(A)$ a.s., which can only happen if $\mathbb{P}(A)\in\{0,1\}$. □

2.7.2 Strong Law of Large Numbers

Theorem 2.19 (Strong law of large numbers)**.** *Let $X_1, X_2, \dots$ be i.i.d. random variables in $\mathcal{L}^1$ and set $\mu=\mathbb{E}[X_1]$. Define $S_0=0$ and $S_n = X_1+\dots+X_n$. Then $\frac{S_n}{n}\to\mu$ a.s. and in $\mathcal{L}^1$ as $n\to\infty$.*

Proof. Define $\mathcal{G}_n=\sigma(S_n, S_{n+1},\dots)=\sigma(S_n, X_{n+1},\dots)$. Set $\mathcal{F}_n=\mathcal{G}_{-n}$ and define $M_n=\frac{S_{-n}}{-n}$ for $n\le -1$. We will prove that M is a backwards martingale w.r.t. $\mathcal{F}$. Let $m=-n$. Then

$$\mathbb{E}[M_{n+1}\mid\mathcal{F}_n] = \mathbb{E}\Big[\frac{S_{-n-1}}{-n-1}\Big|\mathcal{G}_{-n}\Big]$$

$$
\begin{aligned}
&= \mathbb{E}\Big[\frac{S_{m-1}}{m-1}\Big| S_m, X_{m+1}, \dots\Big] \\
&= \mathbb{E}\Big[\frac{S_m - X_m}{m-1}\Big| S_m, X_{m+1}, \dots\Big] \\
&= \frac{S_m}{m-1} - \frac{1}{m-1}\mathbb{E}[X_m \mid S_m, X_{m+1}, \dots] \\
&\overset{\text{Theorem 1.9}}{=} \frac{S_m}{m-1} - \frac{1}{m-1}\mathbb{E}[X_m \mid S_m]
\end{aligned}
$$

By symmetry (the X_n are i.i.d.), $\mathbb{E}[X_m \mid S_m] = \mathbb{E}[X_k \mid S_m]$ for all $k = 1, \dots, m$. So $m\mathbb{E}[X_1 \mid S_m] = \sum_{k=1}^m \mathbb{E}[X_k \mid S_m] = \mathbb{E}[S_m \mid S_m] = S_m$, hence $\mathbb{E}[X_m \mid S_m] = \frac{S_m}{m}$ a.s., and then

$$
\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \frac{S_m}{m-1} - \frac{1}{m-1}\frac{S_m}{m} = \frac{S_m}{m} = M_n \text{ a.s.}
$$

So by the convergence theorem we get that $\frac{S_n}{n}$ converges a.s. and in $\mathcal{L}^1$ to a limit variable Y. We have

$$
Y = \lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{X_{k+1} + \cdots + X_n}{n},
$$

so Y is $\sigma(X_k, X_{k+1}, \dots)$-measurable for all k. So Y is $\bigcap_k \sigma(X_k, X_{k+1}, \dots)$-measurable. This is the tail σ-algebra, hence trivial by Kolmogorov's 0-1 law. Then there exists $c \in \mathbb{R}$ such that $\mathbb{P}(Y = c) = 1$. Since $\frac{S_n}{n} \to Y$ in $\mathcal{L}^1$ and $\mathbb{E}[\frac{S_n}{n}] = \mu$, we get $c = \mu$. □

2.7.3 Kakutani's Product Martingale Theorem

Theorem 2.20 (Kakutani's product martingale theorem)**.** *Let $X_1, X_2, \dots$ be independent nonnegative random variables with $\mathbb{E}[X_n] = 1$ for all n. Define $M_0 = 1$, $M_n = X_1 X_2 \cdots X_n$ for $n \in \mathbb{N}$. Then (M_n) converges a.s. to some M_∞ as $n \to \infty$. Set $a_n = \mathbb{E}[\sqrt{X_n}]$. Then $0 < a_n \le 1$ for all n.*

(1) If $\prod_{n=1}^\infty a_n > 0$, then $M_n \to M_\infty$ in $\mathcal{L}^1$ and $\mathbb{E}[M_\infty] = 1$.

(2) If $\prod_{n=1}^\infty a_n = 0$, then $M_\infty = 0$ a.s.

Proof. (M_n) is a martingale, nonnegative and $\mathbb{E}[M_n] = 1$. So (M_n) is a martingale bounded in $\mathcal{L}^1$, so $M_n \to M_\infty$ a.s. as $n \to \infty$ for some M_∞. The inequality $a_n \le 1$ is immediate from Cauchy-Schwarz. Set $N_n = \frac{\sqrt{X_1 \cdots X_n}}{a_1 \cdots a_n}$. N is a nonnegative martingale and $\mathbb{E}[N_n] = 1$. Hence again $N_n \to N_\infty$ a.s. as $n \to \infty$ for some N_∞. We have $M_n = N_n^2\big(\prod_{i=1}^n a_i\big)^2 \le N_n^2$. Note that $\mathbb{E}[N_n^2] = \big(\prod_{i=1}^n a_i\big)^{-2}$, so

$$
\sup_{n\ge 0} \mathbb{E}[N_n^2] = \Big(\prod_{i=1}^\infty a_i\Big)^{-2}.
$$

(1) If $\prod_{a_i} > 0$, then $\sup_{n\ge0} \mathbb{E}[N_n^2] < \infty$. We show that M is UI, then we get $M_n \to M_\infty$ in $\mathcal{L}^1$ from Theorem 2.15. Note that $M_k \le \sup_{n\ge0} M_n$, so it is enough to prove that

$\sup_{n\geq 0} M_n \in \mathcal{L}^1$. We have

$$\begin{aligned}\mathbb{E}[\sup_{k\leq n} M_k] &\leq \mathbb{E}[\sup_{k\leq n} N_k^2] \\ &\overset{2.10}{\leq} 4\mathbb{E}[N_n^2] \\ &\leq 4\sup_{n\geq 0}\mathbb{E}[N_n^2].\end{aligned}$$

Thus by monotone convergence, $\mathbb{E}[\sup_{n\geq 0} M_n] \leq 4\sup_{n\geq 0}\mathbb{E}[N_n^2] < \infty$.

(2) If $\prod a_i = 0$, then $M_n = N_n^2\big(\prod_{i=1}^n a_i\big)^2 \to 0$ a.s. as the product goes to 0 and $N_n \to N_\infty$ a.s. as $n \to \infty$.

□

2.7.4 Radon-Nikodym Theorem

Theorem 2.21 (Radon-Nikodym). *Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on $(\Omega, \mathcal{F})$. Suppose that $\mathcal{F}$ is countably generated, i.e. $\mathcal{F} = \sigma(F_n : n \in \mathbb{N})$ where $F_n \subseteq \Omega$. TFAE:*

(a) For all $A \in \mathcal{F}$, if $\mathbb{P}(A) = 0$, then $\mathbb{Q}(A) = 0$ (i.e. $\mathbb{Q}$ is absolutely continuous *w.r.t. $\mathbb{P}$, written $\mathbb{Q} \ll \mathbb{P}$).*

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

(c) There exists a nonnegative random variable X such that $\mathbb{Q}(A) = \mathbb{E}[X1(A)]$ for all $A \in \mathcal{F}$ (expectation taken w.r.t. $\mathbb{P}$).

X is called the *Radon-Nikodym derivative* of $\mathbb{Q}$ w.r.t. $\mathbb{P}$, denoted $X = \frac{d\mathbb{Q}}{d\mathbb{P}}$ a.s.

Remark. By scaling this extends to finite measures. It also extends to σ-finite measures by splitting Ω into sets of finite measure. One can also lift the countably generated assumption of $\mathcal{F}$.

Proof. "*(a) ⇒ (b)*" Suppose (b) does not hold. Then there exists $\varepsilon > 0$ such that for all n there is $A_n \in \mathcal{F}$ with $\mathbb{P}(A_n) \leq 2^{-n}$ and $\mathbb{Q}(A_n) \geq \varepsilon$. By Borel-Cantelli $\mathbb{P}(A_n$ happens infinitely often$) = 0$. But $\mathbb{Q}(A_n \text{ i.o.}) \geq \varepsilon$ as $\{A_n \text{ i.o.}\} = \bigcap_k \bigcup_{n\geq k} A_n$. This contradicts (a).

"*(b) ⇒ (c)*" Define $\mathcal{F}_n = \sigma(F_k : k \leq n)$. Set $\mathcal{A}_n = \{H_1 \cap \cdots \cap H_n \mid H_i = F_i \text{ or } H_i = F_i^c\}$. The sets in $\mathcal{A}_n$ are disjoint and $\mathcal{F}_n = \sigma(\mathcal{A}_n)$. Define

$$X_n(\omega) = \sum_{A\in\mathcal{A}_n} \frac{\mathbb{Q}(A)}{\mathbb{P}(A)} 1(\omega \in A),$$

where we set $\frac{0}{0} = 0$. Let $B \in \mathcal{F}_n$. Then $\mathbb{E}[X_n 1(B)] = \mathbb{Q}(B)$ (the expectation is taken w.r.t. $\mathbb{P}$). Also $\mathbb{E}[X_{n+1}1(B)] = \mathbb{Q}(B) = \mathbb{E}[X_n 1(B)]$ and X_n is $\mathcal{F}_n$-measurable, so $\mathbb{E}[X_{n+1} \mid \mathcal{F}] = X_n$, so $(X_n)_n$ is a martingale. Also $\mathbb{E}[X_n] = \mathbb{Q}(\Omega) = 1$. By the martingale convergence theorem, $X_n \to X_\infty$ a.s. as $n \to \infty$ for some X_∞. We want $\mathcal{L}^1$-convergence, so we will show that (X_n) is UI. So we want to show $\sup_{n\geq 0} \mathbb{E}[X_n 1(X_n \geq \lambda)] \to 0$ as $\lambda \to \infty$. We have $\mathbb{E}[X_n 1(X_n \geq \lambda)] = \mathbb{Q}(X_n \geq \lambda)$. By Markov we have $\mathbb{P}(X_n \geq \lambda) \leq \frac{\mathbb{E}[X_n]}{\lambda} = \frac{1}{\lambda}$. Let $\varepsilon > 0$. Take $\delta > 0$ as in (b) and let $\lambda_0 = \frac{1}{\delta}$. Then for $\lambda \geq \lambda_0$ we have $\mathbb{P}(X_n \geq \lambda) \leq \delta$ and thus $\mathbb{Q}(X_n \geq \lambda) \leq \varepsilon$. So (X_n) is UI and so $X_n \to X_\infty$ in $\mathcal{L}^1$. Then $\mathbb{E}[X_\infty] = 1$ and for all $A \in \mathcal{F}_n$, $\mathbb{Q}(A) = \mathbb{E}[X_n 1(A)] = \mathbb{E}[X_\infty 1(A)]$. So the probability measure $\widetilde{\mathbb{Q}}$, defined by $\widetilde{\mathbb{Q}}(B) = \mathbb{E}[X_\infty 1(B)]$ for $B \in \mathcal{F}$, agrees with $\mathbb{Q}$ on the π-system $\bigcup_{n\geq 0} \mathcal{F}_n$ that generates $\mathcal{F}$. Then $\widetilde{\mathbb{Q}} = \mathbb{Q}$ on $\mathcal{F}$.

"$(c) \Rightarrow (a)$" obvious. $\square$

3 Continuous-Time Processes

So we considered a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a sequence $(X_n)_{n\in\mathbb{N}_0}$ of random variables. Now we want to study processes $(X_t)_{t\in\mathbb{R}_+}$ where for each $t \in \mathbb{R}_+$, $\omega \mapsto X_t(\omega)$ is a random variable. Again we define filtrations $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ as sub-σ-algebras satisfying $\mathcal{F}_t \subseteq \mathcal{F}_s$ for all $t \leq s$. A function $T : \Omega \to [0, \infty]$ is called a stopping time if for all $t \in [0, \infty]$, $\{T \leq t\} \in \mathcal{F}_t$.

In the case of discrete-time processes, $T_A = \min\{n \geq 0 : X_n \in A\}$ (where $A \in \mathcal{B}(\mathbb{R})$) is a stopping time as $\{T_A \leq n\} = \bigcup_{k\leq n}\{X_k \in A\}$.

Now for continuous-time processes, if we define $T_A = \inf\{t \geq 0 : X_t \in A\}$, it is not so clear wether $\{T_A \leq t\} = \bigcup_{s\leq t}\{X_s \in A\}$ is in $\mathcal{F}_t$ and in fact this need not be the case, see the example after Proposition 3.2.

For discrete-time processes, consider $X : (\omega, n) \mapsto X_n(\omega)$ as a map $\Omega \times \mathbb{N}_0 \to \mathbb{R}$. Then X is measurable w.r.t. $\mathcal{F} \otimes \mathcal{P}(\mathbb{N}_0)$.

For continuous-time, the map $\Omega \times \mathbb{R}_+ \to \mathbb{R}$, $(\omega, t) \mapsto X_t(\omega)$ need not be measurable w.r.t. $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$.

To avoid these problems, we require our process to satisfy some regularity conditions:

Suppose X is continuous in t, i.e. for all ω, $t \mapsto X_t(\omega)$ is continuous. Then we can write $X_t(\omega) = \lim_{n\to\infty} X_{2^{-n}\lceil 2^n t\rceil}(\omega)$. Now for all n, $(\omega, t) \mapsto X_{2^{-n}\lceil 2^n t\rceil}(\omega)$ is measurable w.r.t. $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$ as $2^{-n}\lceil 2^n t\rceil$ takes on only countably many values. Then X is also measurable as the limit of measurable functions. We can actually consider X which is only right continuous. We will also require that the left limits $\lim_{t\to t_0^-} X_t(\omega)$ exist for all ω.

We call functions f that are right continuous and whose left limit exist *cadlag*. We write $C(\mathbb{R}_+, E), D(\mathbb{R}_+, E)$ for the space of continuous/cadlag functions $\mathbb{R}_+ \to E$ for suitable sets E (e.g. $E = \mathbb{R}$). Continuous and cadlag functions are uniquely determined by their values in a countable set. We endow these spaces with the product σ-algebra, i.e. the smallest σ-algebra that makes the projections $\pi_t : f \mapsto f(t)$ measurable for all t.

The stopped σ-algebra $\mathcal{F}_T$, where T is a stopping time, is $\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t \,\forall t\}$. If X is cadlag, we define as before $X_T(\omega) = X_{T(\omega)}(\omega)$ when $T(\omega) < \infty$ and the stopped process $X_t^T = X_{T\wedge t}$.

Proposition 3.1. *Let X be a cadlag adapted process and S, T stopping times. Then*

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

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

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

(4) X^T is adapted.

Proof.

(1) Clear.

(2) Clear.

(3) We need to show that for all t, $X_T \cdot 1(T \leq t)$ is $\mathcal{F}_t$-measurable. Let $T_n = 2^{-n}\lceil 2^n T \rceil$. Then $T_n \searrow T$ as $n \to \infty$. T_n takes values in $D_n = \{k2^{-n:k\in\mathbb{N}_0}\} \cup \{\infty\}$. T_n is a stopping time:

$$\{T_n \leq t\} = \{\lceil 2^n T \rceil \leq 2^n t\} = \{T \leq 2^{-n}\lfloor 2^n t \rfloor\} \in \mathcal{F}_{2^{-n}\lfloor 2^n t \rfloor} \subseteq \mathcal{F}_t.$$

We have $X_T 1(T \leq t) = X_T 1(T < t) + X_t 1(T = t)$. Clearly, $X_t 1(T = t)$ is $\mathcal{F}_t$-measurable. By the cadlag property and $T_n \searrow T$,

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

Now $X_{T_n\wedge t} 1(T < t) = \sum_{d\in D_n, d\leq t} X_d 1(T_n = d) + X_t 1(T_n > t) 1(T < t)$. Again $X_t 1(T_n > t) 1(T < t)$ is $\mathcal{F}_t$-measurable as T_n is a stopping time. Also each $X_d 1(T_n = d)$ is $\mathcal{F}_d$ measurable and $\mathcal{F}_d \subseteq \mathcal{F}_t$, so we are done.

(4) By (3), $X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable and $\mathcal{F}_{T\wedge t} \subseteq \mathcal{F}_T$.

□

Proposition 3.2. *If X is a continuous adapted process and A a closed set in $\mathbb{R}$, then*

$$T_A = \inf\{t \geq 0 \mid X_t \in A\}$$

is a stopping time.

Proof. We need to show that for all t, $\{T_A \leq t\} \in \mathcal{F}_t$. We will prove $\{T_A \leq t\} = \{\inf_{s\in\mathbb{Q}, s\leq t} d(X_s, A) = 0\}$. From this the claim is immediate. Suppose $T_A = s \leq t$. Then there is a sequence $s_n \searrow s$ as $n \to \infty$ such that $X_{s_n} \in A$ for all n. By continuity, $X_{s_n} \to X_s$ as $n \to \infty$. Then $d(X_{s_n}, A) = 0$ for all n, and hence $d(X_s, A) = 0$. Let q_n be a sequence of rational numbers such that $q_n \nearrow T_A$ as $n \to \infty$. Then $q_n \leq t$ and $d(X_{q_n}, A) \to d(X_{T_A}, A) = 0$. So $\inf_{s\in\mathbb{Q}, s\leq t} d(X_s, A) = 0$.

Conversely, suppose $\inf_{s\in\mathbb{Q}, s\leq t} d(X_s, A) = 0$. Then there is a sequence $s_n \in \mathbb{Q}$ with $s_n \leq t$ and $d(X_{s_n}, A) \to 0$ as $n \to \infty$. By passing to a subsequence we may assume that s_n converges to some $s \leq t$. Then $0 = d(X_{s_n}, A) \to d(X_s, A)$ as $n \to \infty$, so $d(X_s, A) = 0$ and then $X_s \in A$ as A is closed. So $T_A \leq s \leq t$. □

Example. Let $\xi = \pm 1$ equally likely. Define X by

$$X_t = \begin{cases} t & \text{if } 0 \le t \le 1, \\ 1 + \xi \cdot (t-1) & \text{if } t > 1. \end{cases}$$

Consider $A = (1,2)$, $T_A = \inf\{t \ge 0 : X_t \in A\}$. Let $\mathcal{F}_t = \sigma(X_s : s \le t)$ be the natural filtration. Then T_A is *not* a stopping time as $\{T_A \le 1\} \notin \mathcal{F}_1$.

Definition. *Let* $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ *be a filtration. Define*

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s.$$

If $\mathcal{F}_{t+} = \mathcal{F}_t$*, we call* $\mathcal{F}$ right-continuous.

Proposition 3.3. *Let A be an open set, X a cadlac adapted process. Then* $T_A = \inf\{t \ge 0 : X_t \in A\}$ *is an* $(\mathcal{F}_{t+})$*-stopping time.*

Proof. We need to show that $\{T_A \le t\} \in \mathcal{F}_{t+}$ for all t. We have $\{T_A \le t\} = \bigcap_{n \ge 1} \{T_A < t + \frac{1}{n}\}$ and $\{T_A < t + \frac{1}{n}\} = \bigcup_{s \in \mathbb{Q}, s < t + \frac{1}{n}} \{X_s \in A\}$ since A is open and X is cadlag. Since $\{X_s \in A\} \in \mathcal{F}_{t+\frac{1}{n}}$ for $s < t + \frac{1}{n}$, we get $\{T_A \le t\} \in \mathcal{F}_{t+\frac{1}{n}}$ for all n, hence $\{T_A \le t\} \in \mathcal{F}_{t+}$. □

Let $X = (X_t)_{t \in \mathbb{R}_+}$ be a process. On the set of functions $\{f : \mathbb{R}_+ \to E\}$ (for us $E = \mathbb{R}$) we defined the product σ-algebra, i.e. the smallest σ-algebra for which all the projections are measurable. Given A in this σ-algebra, we define $\mu(A) = \mathbb{P}(X \in A)$. μ is called the *law* of the process.

Given a finite subset $J \subseteq \mathbb{R}_+$, we define μ_J to be the law of $(X_t, t \in J)$. These are called *finite dimensional marginals.* The family $(\mu_J : J \subseteq \mathbb{R}_+, \text{ finite})$ uniquely determines the law of the process. Indeed, $\{\bigcap_{s \in J} \{X_s \in A_s\} : J \subseteq \mathbb{R}_+ \text{ finite}, A_s \in \mathcal{B}(\mathbb{R})\}$ is a π-system generating the product σ-algebra.

Definition. *Let* $(X_t)_{t \ge 0}, (X'_t)_{t \ge 0}$ *be processes on* $(\Omega, \mathcal{F}, \mathbb{P})$*. We say X' is a* version *of X if $X_t = X'_t$ a.s. for all t.*

If X' is a version of X, they both have the same finite dimensional marginals.

Example. Suppose $X_t = 0$ for all $t \in [0,1]$. Let $U \sim \mathcal{U}[0,1]$ and define $X'_t = 1(U = t)$ for $t \in [0,1]$. Clearly $X'_t = 0$ a.s. for all t, i.e. X' is a version of X. But

$$\mathbb{P}(X'_t = 0 \, \forall t) = 0 \neq 1 = \mathbb{P}(X_t = 0 \, \forall t).$$

The finite dimensional marginals are Dirac masses at 0.

Definition. *Let* $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ *be a filtered probability space. Define* $\widetilde{\mathcal{F}}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N})$ *where $\mathcal{N}$ is the set of measure-0 events of $\mathcal{F}$. We say $(\mathcal{F}_t)$ satisfies the* usual conditions *if $\widetilde{\mathcal{F}}_t = \mathcal{F}_t$ for all t.*

Theorem 3.4 (Martingale regularisation theorem)**.** *Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)$-martingale. Then there exists a cadlag martingale $(\widetilde{X}_t)$ w.r.t. $(\widetilde{\mathcal{F}}_t)$ such that $X_t = \mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_t]$ a.s. for all $t \geq 0$. If $(\mathcal{F}_t)$ satisfies the usual conditions, then $\widetilde{X}$ is a cadlag version of X.*

Lemma 3.5. *Let $f : \mathbb{Q}_+ \to \mathbb{R}$ be such that for all bounded $I \subseteq \mathbb{Q}_+$, $a, b \in \mathbb{Q}$ with $a < b$, f is bounded on I and $N([a,b], I, f) < \infty$ where*

$$N([a,b], I, f) = \sup\{n \geq 0 : \exists\, 0 \leq s_1 < t_1 < \cdots < s_n < t_n \in I,\ f(s_i) < a, f(t_i) > b\, \forall i\}$$

is the total number of upcrossings of $[a,b]$ by f in I, then

$$\lim_{\substack{s \nearrow t \\ s \in \mathbb{Q}_+}} f(s), \qquad \lim_{\substack{s \searrow t \\ s \in \mathbb{Q}_+}} f(s)$$

exist and are finite for all $t \in \mathbb{R}_+$.

Proof. Let $s_n \searrow t$. Then $(f(s_n))$ converges by Lemma 2.7 and the assumptions on f. Mixing two such sequences show that the limit is independent of the sequence s_n. Hence $\lim_{\substack{s \searrow t \\ s \in \mathbb{Q}_+}} f(s)$ exists. Similarly the other limit exists. The finiteness of the limits follows from the boundedness of f. □

Proof of Theorem 3.4. Overview:

(1) **Goal.** Define $\widetilde{X}_t = \lim_{\substack{s \searrow t \\ s \in \mathbb{Q}_+}} X_s$. So we want to prove the limit exists a.s. and is finite.

(2) Show $X_t = \mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_t]$ a.s. for all t

(3) Martingale property of $\widetilde{X}$

(4) Cadlag property of $\widetilde{X}$.

Start with (1). Let $I \subseteq \mathbb{Q}_+$ be bounded. To check the conditions in Lemma 3.5, we first want to show that $\mathbb{P}(\sup_{t\in I} |X_t| < \infty) = 1$. Let $J = \{j_1 < \cdots < j_n\} \subseteq I$. Since $(X_j)_{j \in J}$ is a martingale, Doob's maximal inequality, Theorem 2.9, gives

$$\lambda \mathbb{P}(\sup_{t \in J} |X_t| > \lambda) \leq \mathbb{E}[|X_{j_n}|]$$

for any $\lambda > 0$. If we choose $K > \sup I$, this is $\leq \mathbb{E}[|X_k|]$. Letting $J \nearrow I$, we get $\lambda \mathbb{P}(\sup_{t\in I} |X_t| > \lambda) \leq \mathbb{E}[|X_k|]$. Taking $\lambda \to \infty$ gives $\mathbb{P}(\sup_{t \in I} |X_t| < \infty) = 1$. Next we want to consider the upcrossing property. Doob's upcrossing inequality, Theorem 2.8 gives

$$(b-a)\mathbb{E}[N([a,b], J, X)] \leq \mathbb{E}[(X_{j_n} - a)^-] \leq \mathbb{E}[(X_k - a)^-].$$

Taking the sup over all finite $J \subseteq I$ gives $N([a,b], I, X) < \infty$ a.s. Next take $I_M = \mathbb{Q}_+ \cap [0, M]$ for $M \in \mathbb{N}$. Define

$$\Omega_0 = \bigcap_{M \in \mathbb{N}_0} \Big(\{\sup_{t \in I_M} |X_t| < \infty\} \cap \bigcap_{\substack{a<b \\ a,b \in \mathbb{Q}}} \{N([a,b], I_M, X) < \infty\} \Big).$$

Then $\mathbb{P}(\Omega_0) = 1$ and on Ω_0, the limits

$$X_{t+} := \lim_{\substack{s \downarrow t \\ s \in \mathbb{Q}_+}} X_s, \quad X_{t-} := \lim_{\substack{s \uparrow t \\ s \in \mathbb{Q}_+}} X_s$$

exist and are finite. Now define

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

Then $\widetilde{X}$ is $\widetilde{\mathcal{F}}$-adapted.

Next we do (2). Let $t_n \searrow t$ with $t_n \in \mathbb{Q}_+$. Then $\widetilde{X}_t = \lim_{n\to\infty} X_{t_n}$ a.s. Note that $(X_{t_n})_n$ is a backwards martingale as t_n is decreasing. So by Theorem 2.17, $X_{t_n} \to \widetilde{X}_t$ also in $\mathcal{L}^1$. Then $X_t = \mathbb{E}[X_{t_n} \mid \mathcal{F}_t] \xrightarrow{\mathcal{L}^1} \mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_t]$, so $X_t = \mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_t]$ a.s.

Now (3). We need to show that for $s < t$, $\mathbb{E}[\widetilde{X}_t \mid \widetilde{\mathcal{F}}_s] = \widetilde{X}_s$ a.s. Let $s_n \searrow s$ where $s_n \in \mathbb{Q}_+$ such that $s_0 < t$. Then $\widetilde{X}_s = \lim_{n\to\infty} X_{s_n} = \lim_{n\to\infty} \mathbb{E}[X_t \mid \mathcal{F}_{s_n}]$ Now, $\mathbb{E}[X_t \mid \mathcal{F}_{s_n}]$ is another backwards martingale, so $\mathbb{E}[X_t \mid \mathcal{F}_{s_n}] \to \mathbb{E}[X_t \mid \mathcal{F}_{s+}]$ a.s. and in $\mathcal{L}^1$ as $n \to \infty$. So $\widetilde{X}_s = \mathbb{E}[X_t \mid \mathcal{F}_{s+}] = \mathbb{E}[\mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_t] \mid \mathcal{F}_{s+}] = \mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_{s+}] = \mathbb{E}[\widetilde{X}_t \mid \widetilde{\mathcal{F}}_s]$[1].

Finally we do (4). We show that $\widetilde{X}$ is right continuous. If not, there is some $\omega \in \Omega_0$ and $t \in \mathbb{R}_+$ and a sequence (s_n) such that $s_n \searrow t$ and $|\widetilde{X}_{s_n} - \widetilde{X}_t| > \varepsilon$. By definition of $\widetilde{X}$ on Ω_0 there exists a sequence (s'_n) of rationals such that $s'_n > s_n$ for all n and $s'_n \searrow t$ and $|\widetilde{X}_{s_n} - X_{s'_n}| < \frac{\varepsilon}{2}$. But then $|\widetilde{X}_t - X_{s'_n}| \geq \frac{\varepsilon}{2}$ contradicting the definition of $\widetilde{X}$ on Ω_0. The existence of left limits is done similarly (exercise). □

Example. Let ξ, η be independent random variables that attain the values ± 1 equally likely. Define

$$X_t = \begin{cases} 0 & t < 1, \\ \xi & t = 1, \\ \xi + \eta & t > 1. \end{cases}$$

X_t is not right continuous. Take $\mathcal{F}_t = \sigma(X_s : s \leq t)$. Then $\mathcal{F}_1 = \sigma(\xi)$ and $\mathcal{F}_{1+} = \sigma(\xi, \eta)$. (X_t) is a martingale w.r.t. $(\mathcal{F}_t)$. Define

$$\widetilde{X}_t = \begin{cases} 0 & t < 1, \\ \xi + \eta & t \geq 1. \end{cases}$$

Then $\widetilde{X}$ is cadlag and it is a martingale w.r.t. $(\mathcal{F}_{t+})$. Also $X_t = \mathbb{E}[\widetilde{X}_t \mid \mathcal{F}_t]$ a.s. for all t.

But $\widetilde{X}$ is not a version of X, because $\mathbb{P}(X_1 \neq \widetilde{X}_1) > 0$.

[1] If X is any random variable and $\mathcal{G}$ any sub-σ-algebra, we have $\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{N})] = \mathbb{E}[X \mid \mathcal{G}]$ a.s.

3.1 Convergence theorems and inequalities

Theorem 3.6 (A.s. martingale convergence theorem)**.** *Let (X_t) be a cadlag martingale which is bounded in $\mathcal{L}^1$, i.e.* $\sup_{t\geq 0}\mathbb{E}[|X_t|]<\infty$. *Then (X_t) converges a.s. to $X_\infty \in \mathcal{L}^1$.*

Proof. Let $I_M = \mathbb{Q}_+ \cap [0,M]$. As in the proof of Theorem 3.4 we get from Doob's upcrossing inequality, Theorem 2.8,

$$(b-a)\mathbb{E}[N([a,b],I_M,X)] \leq a + \sup_{t\geq 0}\mathbb{E}[|X_t|].$$

Taking $M\to\infty$. Hence $N([a,b],\mathbb{Q}_+,X)<\infty$ a.s. Define

$$\Omega_0 = \bigcap_{\substack{a<b\\ a,b\in\mathbb{Q}}} \{N([a,b],\mathbb{Q}_+,X)<\infty\}.$$

Then $\mathbb{P}(\Omega_0)=1$ and on Ω_0, X_q converges as $q\to\infty$, $q\in\mathbb{Q}_+$. Denote the limit by X_∞ (and set it to 0 on Ω_0^c). Then $X_q\to X_\infty$ a.s. as $q\to\infty, q\in\mathbb{Q}_+$. We already get $X_\infty\in\mathcal{L}^1$ by Fatou. We need to show $X_t\to\infty$ a.s. for $t\to\infty$. The follows easily from the cadlag property, indeed, given $\varepsilon>0$, there exists $q_0\in\mathbb{Q}_+$ such that for all $q>q_0$ with $q\in\mathbb{Q}$, we have $|X_q-X_\infty|<\frac{\varepsilon}{2}$. By right continuity for all $t>q_0$ there is $q\in\mathbb{Q}_+$ such that $q>t$ such that $|X_t-X_q|<\frac{\varepsilon}{2}$. So for $t>q_0$, we get $|X_t-X_\infty|<\varepsilon$. □

Theorem 3.7 (Doob's maximal inequality)**.** *Let (X_t) be a cadlag martingale, $X_t^*=\sup_{s\leq t}|X_s|$. Then for all $\lambda>0$,*

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

Proof. By the cadlag property, $X_t^*=\sup_{s\leq t}|X_s|=\sup_{s\in\mathbb{Q}_+\cap[0,t]\cup\{t\}}|X_s|$. The rest follows as in the proof of Theorem 3.4 □

Theorem 3.8 (Doob's $\mathcal{L}^p$-inequality)**.** *Let $p>1$, X a cadlag martingale. Then*

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

Proof. Exactly as in the discrete case, Theorem 2.10. □

Theorem 3.9 ($\mathcal{L}^p$ martingale convergence)**.** *Let X be a cadlag martingale, $p>1$. TFAE*

(1) X is bounded in $\mathcal{L}^p$.

(2) X converges a.s. and in $\mathcal{L}^p$ so some $X_\infty\in\mathcal{L}^p$.

(3) There exists $Z\in\mathcal{L}^p$ such that $X_t=\mathbb{E}[Z\mid\mathcal{F}_t]$ a.s. for all t.

Proof. Exactly as in the discrete case, Theorem 2.11 □

Theorem 3.10 (UI martingale convergence)**.** *Let X be a cadlag martingale. TFAE*

(1) X is UI.

(2) X converges a.s. and in $\mathcal{L}^1$ to some $X_\infty \in \mathcal{L}^1$.

(3) There exists $Z \in \mathcal{L}^1$ such that $X_t = \mathbb{E}[Z \mid \mathcal{F}_t]$ a.s. for all t.

Proof. Exactly as in the discrete case, Theorem 2.15 □

If these conditions are satisfied and T is a stopping time, we can define X_T by $X_T(\omega) = X_{T(\omega)}(\omega)$ where now the case $T(\omega) = \infty$ is included.

Theorem 3.11 (Optional stopping theorem)**.** *Let X be a cadlag UI martingale $S \leq T$ stopping times. Then*

$$\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S \text{ a.s.}$$

Proof. We need to show that for all $A \in \mathcal{F}_S$,

$$\mathbb{E}[X_T 1(A)] = \mathbb{E}[X_S 1(A)].$$

Let $T_n = 2^{-n}\lceil 2^n T\rceil$ and $S_n = 2^{-n}\lceil 2^n S\rceil$. Then $T_n \searrow T$ and $S_n \searrow S$ as $n \to \infty$. By the right continuity of X, we get $X_{T_n} \to X_T$ a.s. and $X_{S_n} \to X_S$ a.s. as $n \to \infty$. X is UI, so $X_{T_n} = \mathbb{E}[X_\infty \mid \mathcal{F}_{T_n}]$. Note that also $(X_{T_n}), (X_{S_n})$ are UI, so $X_{S_n} = \mathbb{E}[X_{T_n} \mid \mathcal{F}_{S_n}]$ by the discrete optional stopping theorem, Theorem 2.16. Since $A \in \mathcal{F}_S$, also $A \in \mathcal{F}_{S_n}$ for all n. So $\mathbb{E}[X_{T_n} 1(A)] = \mathbb{E}[X_{S_n} 1(A)]$. Let $n \to \infty$. By the UI property, $X_{T_n} \to X_T, X_{S_n} \to X_S$ in $\mathcal{L}^1$, so we get $\mathbb{E}[X_T 1(A)] = \mathbb{E}[X_S 1(A)]$. □

3.2 Kolmogorov's continuity criterion

Write $\mathcal{D}_n = \{k2^{-n} : 0 \leq k \leq 2^n\}$, $\mathcal{D} = \bigcup_{n \geq 0} \mathcal{D}_n$.

Theorem 3.12. *Let $(X_t)_{t \in \mathcal{D}}$ be a stochastic process such that there exist $c < \infty, p, \varepsilon > 0$ such that*

$$\mathbb{E}[|X_t - X_s|^p] \leq c|t - s|^{1+\varepsilon}$$

for all $s, t \in \mathcal{D}$. Then for every $\alpha \in (0, \frac{\varepsilon}{p})$, the process $(X_t)_{t \in \mathcal{D}}$ is a.s. α-Hölder continuous, i.e. there exists a random variable M_α with $M_\alpha < \infty$ a.s. such that $|X_t - X_s| \leq M_\alpha |t - s|^\alpha$ for all $t, s \in \mathcal{D}$.

This implies that there exists an α-Hölder continuous process $\widetilde{X}$ on $[0, 1]$ such that $\widetilde{X}_t = X_t$ a.s. for all $t \in \mathcal{D}$.

Proof. We first prove the Hölder continuity on dyadics of the same level and then extend it to all of $\mathcal{D}$.

We have

$$\mathbb{P}(|X_{k2^{-n}} - X_{(k+1)2^{-n}}| \geq 2^{-n\alpha}) \leq 2^{n\alpha p}\mathbb{E}[|X_{k2^{-n}} - X_{(k+1)2^{-n}}|] \leq 2^{n\alpha p}c2^{-n(1+\varepsilon)}.$$

Taking the union of these 2^n events, we get:

$$\mathbb{P}(\max_{0\leq k<2^n} |X_{k2^{-n}} - X_{(k+1)2^{-n}}| \geq 2^{-n\alpha}) \leq 2^n \cdot 2c^{-n((1+\varepsilon)-\alpha p)} = c2^{-n(\varepsilon-\alpha p)}.$$

If $\alpha < \frac{\varepsilon}{p}$, then by Borel-Cantelli,

$$\max_{0\leq k<2^n} |X_{k2^{-n}} - X_{(k+1)2^{-n}}| \leq 2^{-n\alpha} \text{ for all } n \text{ sufficiently large a.s.}$$

Then

$$\sup_{n\geq 0} \frac{\max_{0\leq k<2^n} |X_{k2^{-n}} - X_{(k+1)2^{-n}}|}{2^{-n\alpha}} \leq M < \infty$$

where M is a random variable with $M < \infty$ a.s. Next we want to show that for all $t, s \in \mathcal{D}$, $|X_t - X_s| \leq M'|t-s|^\alpha$ for some M'. Let $s < t$, $s, t \in \mathcal{D}$. Let r be the unique integer such that $2^{-(r+1)} \leq t - s < 2^{-r}$. Let k be such that $s < k2^{-(r+1)} < t$ and set $z = k2^{-(r+1)}$. Then $t - z = \sum_{j\geq r+1} \frac{x_j}{2^j}$ where $x_j \in \{0,1\}$ for all j. Also $z - s = \sum_{j\geq r+1} \frac{y_j}{2^j}$ for some $y_j \in \{0,1\}$. So we can decompose $[s,t) \cap \mathcal{D}$ into a disjoint union of dyadic intervals of lengths 2^{-n} with $n \geq r+1$. Any given length will appear in at most two intervals. Then

$$\begin{aligned}
|X_t - X_s| &\leq \sum_{\substack{d,n \text{ such that} \\ d, d+2^{-n} \text{ are} \\ \text{endpoints in this decomp.}}} |X_d - X_{d+s^{-n}}| \\
&\leq \sum_{d,n} M2^{-n\alpha} \\
&\leq 2M \sum_{n\geq r+1} 2^{-n\alpha} = 2M\frac{2^{-\alpha(r+1)}}{1-2^{-\alpha}}
\end{aligned}$$

Set $M' = \frac{2M}{1-2^{-\alpha}}$. We get

$$|X_t - X_s| \leq M' 2^{-\alpha(r+1)} \leq M'|t-s|^\alpha$$

This shows that $(X_t)_{t\in\mathcal{D}}$ is α-Hölder continuous a.s.

For the last part: On the event (of probability 1) that X is Hölder continuous, we set $\widetilde{X}_t = \lim_{n\to\infty} X_{t_n}$ where $t_n \in \mathcal{D}$ and $t_n \to t$ as $n \to \infty$. On the complement (of probability 0) set $\widetilde{X}_t = 0$ for all $t \in [0,1]$. Then $\widetilde{X}$ is α-Hölder continuous and $\widetilde{X}_t = X_t$ a.s. for $t \in \mathcal{D}$. □

4 Weak convergence

Let (M,d) be a metric space, endowed with its Borel σ-algebra. We will usually consider random variables with values in M.

Definition. *Let (μ_n) be a sequence of probability measures on (M,d). We say that (μ_n)* converges weakly *to μ and write $\mu_n \Rightarrow \mu$ if for all continuous, bounded functions $f : M \to \mathbb{R}$, we have $\int f d\mu_n \to \int f d\mu$.*

Note that by taking $f \equiv 1$, we get $\mu(M) = 1$, so μ is also a probability measure.

Example. If $x_n \to x$ in (M,d), then $\delta_{x_n} \Rightarrow \delta_x$.

Example. Let $M = [0,1]$ with the euclidean metric. Let $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{\frac{k}{n}}$. Then $\int f d\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} f\left(\frac{k}{n}\right)$ is a Riemann sum of f. So if f is continuous, then $\int f d\mu_n \to \int_0^1 f(x)dx$, so μ_n converges weakly to the Lebesgue measure on $[0,1]$.

Example. Let $M = [0,1]$ and $x_n = \frac{1}{n}$, $\mu_n = \delta_{x_n}$. Take $A = (0,1)$. Then $\mu_n(A) = 1$ and $\mu_n \Rightarrow \delta_0$, but $\delta_0(A) = 0$, so $\mu_n(A) \not\to \mu(A)$.

Theorem 4.1. *Let (μ_n) be a sequence of probability measures on (M,d). TFAE:*

(1) $\mu_n \Rightarrow \mu$.

(2) $\liminf \mu_n(G) \geq \mu(G)$ *for all open* $G \subseteq M$.

(3) $\limsup \mu_n(C) \leq \mu(C)$ *for all closed* $C \subseteq M$.

(4) If $A \in \mathcal{B}(M)$ *is such that* $\mu(\partial A) = 0$, *then* $\mu_n(A) \to \mu(A)$.

Proof. "(1) $\Rightarrow$ (2)" Let $K \in \mathbb{N}$. Define $f_K : M \to \mathbb{R}$ by $f_K(x) = 1 \wedge (Kd(x, G^c))$. Then f_K is continuous, bounded and $f_K \nearrow 1(G)$ as $K \to \infty$. Since $\mu \Rightarrow \mu$, we get $\int f_K d\mu_n \to \int f_M d\mu$ as $n \to \infty$. Then $\liminf_n \mu_n(G) \geq \liminf_n \mu_n(f_K) = \mu(f_K)$. By monotone convergence, $\mu(f_K) \to \mu(G)$ as $K \to \infty$. So $\liminf_n \mu_n(G) \geq \mu(G)$.

"(2) $\Leftrightarrow$ (3)" obvious.

"(2), (3) $\Rightarrow$ (4)" Let A be such that $\mu(\partial A) = 0$. Since $\partial A = \overline{A} \setminus A^\circ$, we get $\mu(\overline{A}) = \mu(A) = \mu(A^\circ)$. Applying (2) and (3) gives

$$\liminf_n \mu_n(A) \geq \liminf_n \mu_n(A^\circ) \geq \mu(A^\circ) = \mu(A),$$
$$\limsup_n \mu_n(A) \leq \limsup_n \mu_n(\overline{A}) \leq \mu(\overline{A}) = \mu(A)$$

This proves $\mu_n(A) \to \mu(A)$.

"(4) $\Rightarrow$ (1)" Suppose f continuous, bounded and positive. The general case follows by writing $f = f^+ - f^-$. We need to show that $\mu_n(f) \to \mu(f)$. Suppose $f \le K$ for some $K \in \mathbb{R}_+$. Then

$$\begin{aligned}\mu_n(f) &= \int_M f(x)d\mu_n(x) = \int_M \Big(\int_0^\infty 1(f(x) \ge t)dt\Big)d\mu_n(x)\\ &= \int_M \Big(\int_0^K (f(x) \ge t)dt\Big)d\mu_n(x)\\ &= \int_0^K \Big(\int_M 1(f(x) \ge t)d\mu_n(x)\Big)dt\\ &= \int_0^K \mu_n(\{x : f(x) \ge t\})dt\\ &= \int_0^K \mu_n(f \ge t)dt\end{aligned}$$

We have $\partial\{f \ge t\} = \{f \ge t\} \setminus (\{f \ge t\})^\circ \subseteq \{f \ge t\} \setminus \{f > t\} = \{f = t\}$. Claim: The set $\{t : \mu(f = t) > 0\}$ is countable. Indeed, we have $\{t : \mu(f = t) > 0\} = \bigcup_{n\ge1}\{t : \mu(f = t) \ge \frac{1}{n}\}$ and for $n \ge 1$ the set $\{t : \mu(f = t) \ge \frac{1}{n}\}$ can have at most n elements, as $\mu(M) = 1$. So $\mu(\partial\{f \ge t\}) = 0$ for Lebesgue a.s. t. By (4) and dominanted convergence, $\int_0^K \mu_n(f \ge t)dt \to \int_0^K \mu(f \ge t)dt$. By the same calculation as above, this last integral is $\mu(f)$. □

Definition. *Let μ be a Borel measure on $\mathbb{R}$. We define its* distribution function $F_\mu(x) := \mu((-\infty, x])$ *for $x \in \mathbb{R}$.*

Proposition 4.2. *Let $(\mu_n), \mu$ be probability measures on $\mathbb{R}$. TFAE:*

(i) $\mu_n \Rightarrow \mu$ as $n \to \infty$.

(ii) For all $x \in \mathbb{R}$ that are points of continuity of F_μ, we have $F_{\mu_n}(x) \to F(x)$ as $n \to \infty$.

Proof. "(i) $\Rightarrow$ (ii)" Let x be a continuity point of F_μ. We need to show $\mu_n((-\infty, x]) \to \mu((-\infty, x))$ as $n \to \infty$. By Theorem 4.1, it suffices to prove $\mu(\partial(-\infty, x]) = 0$. Indeed, we have

$$\begin{aligned}\mu(\partial(-\infty, x]) = \mu(\{x\}) &= \mu((-\infty, x]) - \lim_{n\to\infty} \mu((-\infty, x - \frac{1}{n}])\\ &= F_\mu(x) - \lim_{n\to\infty} F_\mu(x - \frac{1}{n}) = 0\end{aligned}$$

where the last equality follows since F_μ is continuous at x.

"(ii) $\Rightarrow$ (i)" We will prove that if G is an open subset of $\mathbb{R}$, then $\liminf_n \mu_n(G) \ge \mu(G)$. We can write $G = \bigcup_k (a_k, b_k)$ where (a_k, b_k) are disjoint interval. Fix such an interval $(a, b) = (a_k, b_k)$. We have $\mu_n(a, b) = \mu_n((-\infty, b)) - \mu_n((-\infty, a]) = F_{\mu_n}(b-) - F_{\mu_n}(a) \ge$

$F_{\mu_n}(b') - F_{\mu_n}(a')$ where $a < a' < b' < b$, and a', b' are continuity points of F_μ. Note that such a', b' exist as F_μ can only have at most countably many discontinuities since it is monotone. Then $\liminf_n \mu_n(a, b) \geq F_\mu(b') - F_\mu(a')$ by (ii). Take the limit as $b' \nearrow b$ and $a' \searrow a$ along continuity points of F_μ to get $\liminf_n \mu_n(a, b) \geq F_\mu(b-) - F_\mu(a) = \mu((a, b))$.

Then we have

$$\liminf_n \mu_n(G) = \liminf_{n\to\infty} \sum_k \mu_n((a_k, b_k)) \geq \sum_k \liminf_{n\to\infty} \mu_n((a_k, b_k)) \geq \sum_k \mu((a_k, b_k)) = \mu(G).$$

□

Definition. *Let $(X_n)_n$ be a sequence of random variables defined on perhaps different probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ taking values in the same metric space (M, d). We say that (X_n)* converges in distribution *to some other random variable X, written $X_n \xrightarrow{d} X$, if the law $\mathcal{L}(X_n)$ of X_n converges weakly to the law $\mathcal{L}(X)$ of X.*

Equivalently, for all continuous and bounded functions f,

$$\mathbb{E}_{\mathbb{P}_n}[f(X_n)] \to \mathbb{E}_{\mathbb{P}}[f(X)]$$

as $n \to \infty$.

Example. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mu = \mathbb{E}[X_1] < \infty$ and $\sigma^2 = \mathrm{Var}(X_1) < \infty$. Then $\frac{\sum_{i=1}^n X_i - \mu n}{\sigma\sqrt{n}}$ converges in distribution to a a $\mathcal{N}(0, 1)$ random variable as $n \to \infty$. This is the *Central Limit Theorem.*

Proposition 4.3.

(1) If X_n converges to X in probability, then X_n also converges to X in distribution.

(2) If X_n converges to a constant c in distribution, then X_n also converges in probability to c.

Proof. Example Sheet 3. □

4.1 Tightness

Definition. *A sequence of probability measures $(\mu_n)_{n\geq 0}$ on a metric space (M, d) is* tight, *if for all $\varepsilon > 0$, there exists a compact set $K \subseteq M$ such that*

$$\sup_{n\geq 0} \mu_n(M \setminus K) \leq \varepsilon.$$

Remark. If the metric space is compact, then every sequence is tight.

Theorem 4.4 (Prokhorov). *If (μ_n) is a tight sequence, then there is a subsequence (n_k) and a probability measure μ such that $\mu_{n_k} \Rightarrow \mu$ as $k \to \infty$.*

Proof. We give the proof in the case $M = \mathbb{R}$. Consider $F_n = F_{\mu_n}$. By a standard diagonal subsequence argument, we find a subsequence n_k such that $F_{n_k}(x)$ converges for all $x \in \mathbb{Q}$. Define $F(q)$ for $q \in \mathbb{Q}$ to be the limit $\lim_{k\to\infty} F_{n_k}(q)$. Since the F_{n_k} are non-decreasing, so is F. Then we can define F on $\mathbb{R}$ by $F(x) = \lim_{q \searrow x, q\in\mathbb{Q}} F(q)$. Then F is right continuous and by the monotonicity left limits of F exist, so F is cadlag. Next we have to prove that $F_{n_k}(t) \to F(t)$ for all continuity points t of F. Let t be a continuity point of F and $\varepsilon > 0$. We can find $s_1 < t < s_2$ such that $s_1, s_2 \in \mathbb{Q}$ and $|F(s_i) - F(t)| < \frac{\varepsilon}{2}$. For k large enough we then have

$$F(t) - \varepsilon < F(s_1) - \frac{\varepsilon}{2} < F_{n_k}(s_1) \leq F_{n_k}(t) \leq F_{n_k}(s_2) < F(s_2) + \frac{\varepsilon}{2} < F(t) + \varepsilon.$$

So $F_{n_k}(t) \to F(t)$ as $k \to \infty$ for all points t of continuity for F. It remains to show that F is the distribution function of some probability measure on $\mathbb{R}$.

Since (μ_n) is tight, for all $\varepsilon > 0$ there exists N such that $\sup_{n\geq 0} \mu_n([-N, N]^c) \leq \frac{\varepsilon}{2}$. We can pick N such that N and $-N$ are continuity points for F (as F has at most countably many discontinuities). We have $F_{n_k}(-N) \leq \frac{\varepsilon}{2}$ and $1 - F_{n_k}(N) \leq \frac{\varepsilon}{2}$. Then $F(-N) \leq \varepsilon$ and $1 - F(N) \leq \varepsilon$, so $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Define $\mu((a, b]) = F(b) - f(a)$. By Caratheodory's extension theorem we can extend μ to a Borel probability measure on $\mathbb{R}$ and $F = F_\mu$. $\square$

4.2 Characteristic functions

Let X be a random variable with values in $\mathbb{R}^d$. Recall that the *characteristic function* of X is defined to be

$$\varphi = \varphi_X : \mathbb{R}^d \longrightarrow \mathbb{C},$$
$$\varphi(u) = \varphi_X(u) = \mathbb{E}[e^{i\langle u,X\rangle}]$$

Note that $\varphi_X(0) = 1$ and φ_X is continuous.

Theorem 4.5. *The characteristic function uniquely determines the law of a random variable, i.e. if $\varphi_X = \varphi_Y$, then $\mathcal{L}(X) = \mathcal{L}(Y)$.*

Theorem 4.6 (Lévy's convergence theorem)**.** *Let $(X_n)_{n\geq 0}, X$ be random variables with values in $\mathbb{R}^d$. Then*

$$X_n \xrightarrow{d} X \quad \Longleftrightarrow \quad \varphi_{X_n}(u) \to \varphi_X(u),\ \forall u \in \mathbb{R}^d.$$

This follows from the following slightly stronger theorem.

Theorem 4.7 (Lévy)**.**

(1) Let $(X_n), X$ be random variables with values in $\mathbb{R}^d$. If $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$, then $\varphi_{X_n}(u) \to \varphi_X(u)$ for all $u \in \mathbb{R}^d$.

(2) Suppose (X_n) is a sequence of random variables with values in $\mathbb{R}^d$ satisfying that there is some function $\psi : \mathbb{R}^d \to \mathbb{C}$ such that $\psi(0) = 1$, ψ is continuous at 0, and $\varphi_{X_n}(u) \to \psi(u)$ as $n \to \infty$ for all $n \in \mathbb{R}^d$. Then there exists a random variable X such that $\psi = \varphi_X$ and $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ as $n \to \infty$.

Lemma 4.8. *Let X be a random variable with values in $\mathbb{R}^d$. Then for $K > 0$,*

$$\mathbb{P}(\|X\|_\infty \geq K) \leq C\Big(\frac{K}{2}\Big)^d \int_{[-K^{-1},K^{-1}]^d} (1 - \varphi_X(u))du$$

where $C = (1 - \sin 1)^{-1}$ (the RHS is real).

Proof. Let $\lambda > 0$. Let μ be the law of X. Then

$$\begin{aligned}
\int_{[-\lambda,\lambda]^d} \varphi_X(u)du &= \int_{[-\lambda,\lambda]^d} \mathbb{E}[e^{i\langle u,X\rangle}]du \\
&= \int_{[-\lambda,\lambda]^d} \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} d\mu(x)du \\
&= \int_{\mathbb{R}^d} \int_{[-\lambda,\lambda]^d} \prod_{j=1}^d e^{iu_jx_j} dud\mu(x) \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^d \Big(\frac{e^{i\lambda x_j} - e^{-i\lambda x_j}}{ix_j}\Big) d\mu(x) \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^d \Big(\frac{2\sin(\lambda x_j)}{x_j}\Big) d\mu(x).
\end{aligned}$$

So

$$\lambda^{-d} \int_{[-\lambda,\lambda]^d} (1 - \varphi_X(u))du = 2^d \int_{\mathbb{R}^d} \Big(1 - \prod_{j=1}^d \frac{\sin(\lambda x_j)}{\lambda x_j}\Big) d\mu(x)$$

Consider $f : \mathbb{R}^d \to \mathbb{R}$, $f(u) = \prod_{j=1}^d \frac{\sin u_j}{u_j}$. If $x \geq 1$, then $|\sin x| \leq x \sin 1$. Then if $\|u\|_\infty \geq 1$, then $|f(u)| \leq \sin 1$. Consequently, $1(\|u\|_\infty \geq 1) \leq C(1 - f(u))$ where $C = (1 - \sin 1)^{-1}$. Then

$$\begin{aligned}
\mathbb{P}(\|X\|_\infty \geq K) = \mathbb{P}\Big(\Big\|\frac{X}{K}\Big\|_\infty \geq 1\Big) &\leq C\Big(1 - \mathbb{E}\Big[f\big(\frac{X}{K}\big)\Big]\Big) \\
&= C \int_{\mathbb{R}^d} \Big(1 - \prod_{j=1}^d \frac{\sin \frac{X_j}{K}}{\frac{X_j}{k}}\Big) d\mu(x) \\
&= C(2\lambda)^{-d} \int_{[-\lambda,\lambda]^d} (1 - \varphi_X(u))du.
\end{aligned}$$

□

Proof of Theorem 4.7. (1) is easy. Let $f(x) = e^{i\langle u,x\rangle}$. Then f is continuous and bounded, so letting $\mu_n = \mathcal{L}(X_n)$ and $\mu = \mathcal{L}(X)$, we get $\mu_n(f) \to \mu(f)$ as $n \to \infty$ by the definition of weak convergence.

(2) We first show that (X_n) is a tight sequence. We need to show that for every $\varepsilon > 0$, there is K large enough such that $\sup_{n\geq 0} \mathbb{P}(\|X_n\|_\infty \geq K) \leq \varepsilon$. By the lemma, we have

$$\mathbb{P}(\|X_n\|_\infty \geq K) \leq C_d K^d \int_{[-K^{-1},K^{-1}]^d} (1 - \varphi_{X_n}(u))du$$

where $C_d = C2^{-d}$. We have $|1 - \varphi_{X_n}(u)| \leq 2$, so by the dominated convergence theorem, the RHS goes to $C_d K^d \int_{[-K^{-1},K^{-1}]^d}(1 - \psi(u))du$ as $n \to \infty$. Using that $\psi(0) = 1$ and ψ is continuous at 0, we can take K sufficiently large so that

$$\left| C_d K^d \int_{[-K^{-1},K^{-1}]^d} (1 - \psi(u))du \right| < \frac{\varepsilon}{2}.$$

Taking n large enough, we get

$$\mathbb{P}(\|X_n\|_\infty \geq K) \leq \varepsilon.$$

Taking K even larger, we then get this also for the finitely many remaining n, so

$$\sup_{n\geq 0} \mathbb{P}(\|X_n\|_\infty \geq K) \leq \varepsilon.$$

So (X_n) is tight. Then by Theorem 4.4, $\mathcal{L}(X_{n_k}) \Rightarrow \mathcal{L}(X)$ for some random variable X and a subsequence n_k. By (1), $\varphi_{X_{n_k}}(u) \to \varphi_X(u)$ as $k \to \infty$ for all $u \in \mathbb{R}^d$. Then $\psi = \varphi_X$. It remains to prove $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$. Suppose not. Then there is some subsequence (m_k) and a continuous, bounded function f and $\varepsilon > 0$, such that $|\mathbb{E}[f(X_{m_k})] - \mathbb{E}[f(X)]| > \varepsilon$ for all k. Applying the previous argument to X_{m_k} instead of X_n gives a contradiction. $\square$

5 Large deviations

Let $X_1, X_2, \dots$ be i.i.d. random variables and $\mathbb{E}[X_1] = \bar{x}$. Let $S_n = \sum_{i=1}^n X_i$. Then $\frac{S_n}{n} \to \bar{x}$ a.s. as $n \to \infty$, this is the strong law of large numbers, Theorem 2.19.

By the Central Limit Theorem, $\frac{S_n - n\bar{x}}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$ as $n \to \infty$, where $\sigma^2 = \mathrm{Var}(X_1) < \infty$, in other words if $Z \sim \mathcal{N}(0,1)$, then

$$\mathbb{P}(S_n \geq n\bar{x} + a\sigma\sqrt{n}) = \mathbb{P}\Big(\frac{S_n - n\bar{x}}{\sigma\sqrt{n}} \geq a\Big) \to \mathbb{P}(Z \geq a).$$

In this chapter we are interested in the probability $\mathbb{P}(S_n \geq na)$ where $a > \bar{x}$. By the CLT it goes to 0 as $n \to \infty$. But what are the asymptotics?

Let $X_1, X_2, \dots$ by i.i.d. with distribution $\mathcal{N}(0,1)$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\mathbb{P}\Big(\frac{|X_1|}{\sqrt{n}} \geq \delta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-\frac{x^2}{2}}\, dx.$$

We have $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\frac{|S_n|}{\sqrt{n}} \geq \delta) = -\frac{\delta^2}{2}$. So $\frac{S_n}{\sqrt{n}}$ is of order $\frac{1}{\sqrt{n}}$, but it can take relatively large values with an exponentially small probability $\approx e^{-\frac{\delta^2}{2}n}$.

Setup. Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}[X_1] = \bar{x}$, $S_n = \sum_{i=1}^n X_i$. Set $b_n = \mathbb{P}(S_n \geq an)$. Then $b_{n+m} = \mathbb{P}(S_{n+m} \geq a(n+m)) \geq b_n b_m$. So $\log b_{n+m} \geq \log b_n + \log b_m$, so $(-\log b_n)$ is a subadditive sequence. Exercise: Show that this property implies that $\lim_{n\to\infty} -\frac{1}{n}\log b_n$ exists and equals $\inf_n -\frac{\log b_n}{n}$.

So we know that $\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(S_n \geq an)$ exists.

Notation. Let $M(\lambda) = \mathbb{E}[e^{\lambda X_1}]$, $\psi(\lambda) = \log M(\lambda)$.

For $\lambda \geq 0$, we have

$$\mathbb{P}(S_n \geq na) = \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda na}) \leq e^{-\lambda na}\mathbb{E}[e^{\lambda S_n}] = e^{-\lambda na}(M(\lambda))^n = \exp(-n(a\lambda - \psi(\lambda))).$$

Define $\psi^*(a) = \sup_{\lambda \geq 0}(\lambda a - \psi(\lambda))$. Note that $\psi(0) = 0$, so $\psi^*(a) \geq 0$. Then

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

And so

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

Theorem 5.1 (Cramer)**.** *Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}[X_1] = \bar{x}$. Then*

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

for all $a > \bar{x}$.

Lemma 5.2. *The functions $M(\lambda), \psi(\lambda)$ are continuous in $D = \{\lambda : M(\lambda) < \infty\}$ and differentiable in D° with*

$$M'(\lambda) = \mathbb{E}[X_1 e^{\lambda X_1}] \quad \text{and} \quad \psi'(\lambda) = \frac{M'(\lambda)}{M(\lambda)}$$

for $\lambda \in D^\circ$.

Proof. Continuity follows from the dominated convergence theorem. Note that D is an interval, i.e. if $\lambda_1 < \lambda_2$ are in D, then any $\lambda \in (\lambda_1, \lambda_2)$ is also in D which follows from $e^{\lambda x} \leq e^{\lambda_1 x} + e^{\lambda_2 x}$. Let $\eta \in D^\circ$. Let $\delta > 0$ be small and $\varepsilon \in (-\delta, \delta)$. Then $\frac{M(\eta+\varepsilon)-M(\eta)}{\varepsilon} = \frac{\mathbb{E}[e^{(\eta+\varepsilon)X} - e^{\eta X}]}{\varepsilon}$. We have

$$\left| \frac{e^{(\eta+\varepsilon)x} - e^{\eta x}}{\varepsilon} \right| \leq e^{\eta x} \frac{e^{\delta|x|} - 1}{\delta}.$$

Now choose δ small enough so that the RHS has finite mean. Then we can apply dominated convergence and let $\varepsilon \to 0$. ☐

Proof of Theorem 5.1. Replace X_i by $\widetilde{X}_i = X_i - a$ and S_n by $\widetilde{S}_n = S_n - na$. Then $\mathbb{E}[\widetilde{X}_i] \leq 0$ and we have $\mathbb{P}(\widetilde{S}_n \geq 0) = \mathbb{P}(S_n \geq an)$. Also $\widetilde{M}(\lambda) = e^{-\lambda a} M(\lambda)$ and $\widetilde{\psi}(\lambda) = \psi(\lambda) - \lambda a$. So we need to show

$$-\frac{1}{n} \log \mathbb{P}(\widetilde{S}_n \geq 0) \to \widetilde{\psi}^*(0) = \sup_{\lambda \geq 0}(-\widetilde{\psi}(\lambda)).$$

We have already proved $\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\widetilde{S}_n \geq 0) \geq \widetilde{\psi}^*(0)$, so we only need to prove the reverse inequality, equivalently,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\widetilde{S}_n \geq 0) \geq \inf_{\lambda \geq 0} \widetilde{\psi}(\lambda).$$

We now assume $\bar{x} < 0$ and write X_n, S_n for $\widetilde{X}_n, \widetilde{S}_n$.

We can assume that $\mathbb{P}(X_1 > 0) > 0$, since if $\mathbb{P}(X_1 \leq 0) = 1$, then $\mathbb{P}(S_n \geq 0) = \mu(0)^n$, so $\frac{1}{n} \log \mathbb{P}(S_n \geq 0) \to \log \mu(0)$ and $\inf_{\lambda \geq 0} \psi(\lambda) \leq \lim_{\lambda\to\infty} \psi(\lambda) = \log \mu(0)$.

Case 1. Assume $M(\lambda) < \infty$ for all λ. Let $\mu = \mathcal{L}(X_1)$. Define a new probability measure μ_θ by $\frac{d\mu_\theta}{d\mu}(x) = \frac{e^{\theta x}}{M(\theta)}$, so $\mathbb{E}_\theta[f(X_1)] := \mathbb{E}_{\mu_\theta}[f(X_1)] = \int_{\mathbb{R}} \frac{e^{\theta x} f(x)}{M(\theta)} d\mu(x)$. Also

$$\mathbb{E}_\theta[F(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} F(x_1, \dots, x_n) \prod_{i=1}^{n} \frac{e^{\theta x_i} d\mu(x_i)}{M(\theta)}.$$

Define $g(\theta) = \mathbb{E}_\theta[X_1] = \frac{\mathbb{E}[X_1 e^{\theta X_1}]}{M(\theta)} = \psi'(\theta)$. Since $\mu(0,\infty) > 0$ and $\mu(-\infty, 0) > 0$ (from $\bar{x} < 0$), we get $\lim_{|\theta|\to\infty} \psi(\theta) = \infty$. Then there is some η such that $\psi(\eta) = \inf_\theta \psi(\theta)$. Then $\psi'(\eta) = 0$, so $g(\eta) = 0$.

Let $\varepsilon > 0$. We have

$$\begin{aligned}\mathbb{P}(S_n \geq 0) \geq \mathbb{P}(S_n \in [0, \varepsilon n]) &\geq \mathbb{E}\big[e^{\eta S_n - |\eta|\varepsilon n} \cdot 1(S_n \in [0,\varepsilon n])\big] \\ &= e^{-|\eta|\varepsilon n}\mathbb{E}\big[e^{\eta S_n} 1(S_n \in [0,\varepsilon n])\big] \\ &= e^{-|\eta|\varepsilon n}\mathbb{P}_{\mu_\eta}(S_n \in [0,\varepsilon n])(M(\eta))^n\end{aligned}$$

By the Central Limit Theorem[1] (using $\mathbb{E}_\eta[X_1] = 0$), $\mathbb{P}_\eta(S_n \in [0,\varepsilon n]) \to \frac{1}{2}$ as $n \to \infty$. Then

$$\frac{\log \mathbb{P}(S_n \geq 0}{n} \geq -|\eta|\varepsilon + \psi(\eta) + \frac{\log \mathbb{P}_\eta(S_n \in [0,\varepsilon n])}{n} \xrightarrow{n\to\infty} -|\eta|\varepsilon + \psi(\eta)$$

Taking $\varepsilon \to 0$, we get

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq \psi(\eta) \geq \inf_{\lambda\in\mathbb{R}} \psi(\lambda).$$

By Jensen, $\psi(\lambda) = \log \mathbb{E}[e^{\lambda x}] \geq \lambda \bar{x}$. So if $\lambda \leq 0$, since $\bar{x} < 0$, we get $\lambda\bar{x} \geq 0$. So $\inf_{\lambda \leq 0} \psi(\lambda) = 0$ and therefore $\inf_{\lambda\in\mathbb{R}} \psi(\lambda) = \inf_{\lambda\geq 0} \psi(\lambda)$. This proves the theorem in the case $M(\lambda) < \infty$ for all λ.

Case 2. (General case) Let $K \in \mathbb{N}$, $\mu_n := \mathcal{L}(S_n)$, $\nu = \mathcal{L}(X_1 \mid |X_1| \leq K), \nu_n = \mathcal{L}(S_n \mid \bigcap_{i=1}^n \{|X_i| \leq K\})$. We have

$$\mu_n([0,\infty)) \geq \nu_n([0,\infty))\mu([-K,K])^n.$$

Then

$$\frac{1}{n}\log \mu_n([0,\infty) \geq \frac{1}{n} \log \nu_n([0,\infty)) + \log \mu([-K,K]).$$

From the previous case we know

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log \nu_n([0,\infty)) &= \inf_{\lambda\geq 0} \Big(\log \int_{-\infty}^{\infty} e^{\lambda x} d\nu(x)\Big) \\ &= \inf_{\lambda\geq 0} \Big(\log \int_{-K}^{K} e^{\lambda x} d\mu(x) - \log \mu([-K,K])\Big).\end{aligned}$$

So

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n([0,\infty)) \geq \inf_{\lambda\geq 0} \Big(\underbrace{\log \int_{-K}^{K} e^{\lambda x} d\mu(x)}_{=:\psi_K(\lambda)} \Big) =: J_K.$$

[1]L.T.: Why can we apply CLT, i.e. why is $X_1 \in \mathcal{L}^2(\mu_\eta)$?

We have $\psi_K \nearrow \psi$ as $K \to \infty$ and also $J_K \nearrow J$ as $K \to \infty$ for some J. Then $J_K = \inf_{\lambda \geq 0} \psi_K(\lambda) \leq \psi(0) = 0$, so $J \leq 0$. Taking K large, we get $\mu([0, K]) > 0$, so $J_K > -\infty$, so $J > -\infty$. Consider for each $K \in \mathbb{N}$ the set $\{\lambda : \psi_K(\lambda) \leq J\}$. These sets are non-empty, compact (as ψ_K is continuous) and nested. So there is some $\lambda_0 \in \bigcap_{K \in \mathbb{N}} \{\lambda : \psi_K(\lambda) \leq J\}$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n([0, \infty)) \geq J \geq \lim_{K\to\infty} \psi_K(\lambda_0) = \psi(\lambda_0) \geq \inf_{\lambda \geq 0} \psi(\lambda).$$

□

6 Brownian motion

Definition. $B = (B_t)_{t\geq 0}$ *is called a* Brownian motion in $\mathbb{R}^d$ starting from $x \in \mathbb{R}^d$ *if*

(i) B *is a continuous process (recall this means that for all* ω, $t \mapsto B_t(\omega)$ *is continuous).*[1]

(ii) $B_0 = x$ *a.s.*

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

(iv) For any $s < t$, $B_t - B_s$ *is independent of* $\mathcal{F}_s^B = \sigma(B_u : u < s)$ *(we say that* B *has* independent increments*).*

If $x = 0$, B *is called a* standard Brownian motion.

If (B_t) is a standard Brownian motion, $U \sim \mathcal{U}[0,1]$, define

$$X_t = \begin{cases} B_t & t \neq U, \\ 0 & t = U. \end{cases}$$

Then $X = (X_t)_{t\geq 0}$ has the same law as a Brownian motion, but is discontinuous, so is not a Brownian motion.

6.1 Existence

Theorem 6.1 (Wiener's theorem)**.** *There exists a Brownian motion on some probability space.*

Proof. We first construct a Brownian motion for $d = 1$ and on $[0,1]$. Then we extend to $\mathbb{R}_+$ and then to $d \geq 1$.

First we will construct Brownian motion along dyadic rationals of $[0,1]$. Recall for $n \geq 0$, we set $\mathcal{D}_n = \{k2^{-n} : 0 \leq k \leq 2^n\}$ and $\mathcal{D} = \bigcup_{n\geq 0} \mathcal{D}_n$. Let $(Z_d, d \in \mathcal{D})$ be i.i.d. random variables with distribution $\mathcal{N}(0,1)$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Set $B_0 = 0$, $B_1 = Z_1$. Suppose we have constructed $(B_d, d \in \mathcal{D}_{n-1})$ satisfying (iii) and (iv) in the definition of Brownian motion. Let $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$. Let $d_- = d - 2^{-n}, d_+ = d + 2^n \in \mathcal{D}_{n-1}$. Define

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

[1] L.T.: It seems that later we only require B to be continuous a.s.

Then

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

By induction, $B_{d_+} - B_{d_-} \sim \mathcal{N}(0, 2^{-(n-1)})$ and Z_d is independent from this. We see that indeed $B_{d_+} - B_d, B_d - B_{d_-} \sim \mathcal{N}(0, 2^{-n})$. Let $N = \frac{B_{d_+} - B_{d_-}}{2}$ and $N' = \frac{Z_d}{2^{\frac{n+1}{2}}}$. Then $\mathrm{Var}(N) = \mathrm{Var}(N') = 2^{-(n+1)}$. Then $\mathrm{Cov}(N - N', N + N') = \mathrm{Var}(N) - \mathrm{Var}(N') = 0$. Since $(N - N', N + N')$ is a Gaussian vector with covariance 0, they are independent. Similarly one shows that $(B_d - B_{d-2^{-n}}, d \in \mathcal{D}_n)$ are independent.

So we have constructed $(B_d, d \in \mathcal{D})$ satisfying the properties (iii) and (iv). Let $t, s \in \mathcal{D}$. Then $B_t - B_s \sim \mathcal{N}(0, t - s)$, so if $N \sim \mathcal{N}(0, 1)$, then we have

$$\mathbb{E}[|B_t - B_s|^p] = |t - s|^{\frac{p}{2}} \mathbb{E}[|N|^p],$$

and $\mathbb{E}[|N|^p] < \infty$ for all p. By Kolmogorov's continuity criterion, we get that $(B_d, d \in \mathcal{D})$ is a.s. Hölder continuous for all $\alpha < \frac{1}{2}$. For all $t \in [0, 1]$, define $B_t = \lim_{i\to\infty} B_{d_i}$, where $d_i \in \mathcal{D}$, $d_i \to t$ as $i \to \infty$, on the event that $(B_d, d \in \mathcal{D})$ is α-Hölder. On the complement set $B_t = 0$.Then $(B_t)_{t\in[0,1]}$ is a.s. α-Hölder continuous.

We need to show that if $0 \le t_0 < t_1 < \cdots < t_k$, then $(B_{t_i} - B_{t_{i-1}})$ are independent and $\sim \mathcal{N}(0, t_i - t_{i-1})$ for all i. Let $t_i^{(n)} \in \mathcal{D}$ with $t_i^{(n)} \to t_i$ as $n \to \infty$ and $0 \le t_0^{(n)} < t_1^{(n)} < \cdots < t_k^{(n)}$. Then $(B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)}) \to B_{t_i} - B_{t_{i-1}}$ a.s. by continuity of B. We have

$$\mathbb{E}\Big[e^{i\sum_{j=1}^k u_j \left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)}\Big] = \prod_{j=1}^k e^{-\frac{u_j^2\left(t_j^{(n)} - t_{j-1}^{(n)}\right)}{2}}$$
$$\xrightarrow{n\to\infty} \prod_{j=1}^k e^{-\frac{u_j^2(t_j - t_{j-1})}{2}}$$

By Lévy's theorem, Theorem 4.7, we get $(B_{t_i} - B_{t_{i-1}})$ are independent and $\sim \mathcal{N}(0, t_i - t_{i-1})$.

Thus we constructed Brownian motion on $[0, 1]$. We now extend it to $\mathbb{R}_+$. Let $(B_t^i, i \in [0, 1])$ by i.i.d. standard Brownian motions. Define

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

It is easy to check that B has the desired properties.

In $\mathbb{R}^d$, let $B^1, B^2, \ldots, B^d$ be independent one-dimensional standard Brownian motions. Then set $B_t = (B_t^1, \ldots, B_t^d)$. □

6.2 Properties

Proposition 6.2. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$. Then*

(1) *(rotational invariance of Brownian motion) If U is an orthogonal $d \times d$-matrix, then $(UB_t)_{t\geq 0}$ is also a standard Brownian motion.*

(2) *(scaling invariance) For all $\lambda > 0$, $(\frac{B_{\lambda t}}{\sqrt{\lambda}})_{t\geq 0}$ is a standard Brownian motion.*

(3) *(simple Markov property) For all $s \geq 0$, $(B_{t+s} - B_s : t \geq 0)$ is a standard Brownian motion, independent of $\mathcal{F}_s^B = \sigma(B_u : u \leq s)$.*

Proof. Immediate from basic properties of the normal distribution. □

Theorem 6.3 (Inversion formula)**.** *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$. Define*

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

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

Proof. First we check that the finite dimensional marginals are Gaussian with the same mean and covariance as Brownian motion. The mean is clearly 0. For the covariance, let $s < t$. We need to show $\mathrm{Cov}(X_t, X_s) = \mathrm{Cov}(B_t, B_s)$. We have

$$\mathrm{Cov}(B_t, B_s) = \mathrm{Cov}(B_t - B_s, B_s) + \mathrm{Var}(B_s, B_s) = \mathrm{Var}(B_s) = s$$

and

$$\mathrm{Cov}(X_t, X_s) = \mathrm{Cov}(tB_{\frac{1}{t}}, sB_{\frac{1}{s}}) = ts\,\mathrm{Cov}(B_{\frac{1}{t}}, B_{\frac{1}{s}}) = s.$$

It remains to show that X is continuous in t. For $t > 0$, this is clear by continuity of B. So we only need to show $\lim_{t\searrow 0} X_t = 0$. $(X_t, t \geq 0, t \in \mathbb{Q})$ has the same law as $(B_t, t \geq 0, t \in \mathbb{Q})$, so $\lim_{t\searrow 0, t\in\mathbb{Q}} X_t = \lim_{t\searrow 0, t\in\mathbb{Q}} B_t = 0$ a.s. Since X is continuous at all $t > 0$, we get $\lim_{t\searrow 0} X_t = \lim_{t\searrow 0, t\in\mathbb{Q}} X_t = 0$ a.s.[2] □

The continuity at 0 gives:

Corollary 6.4. $\lim_{t\to\infty} \frac{B_t}{t} = 0$ *a.s.*

If we restrict t in the limit to $t \in \mathbb{N}$, this also follows from the strong law of large numbers. If we wanted to use this to prove the corollary one need to control the oscillation of B_t in intervals $[n, n+1]$. See Exercise 2.2. on Example Sheet 3.

Define $\mathcal{F}_s^+ = \bigcap_{t>s} \mathcal{F}_t^B$.

Theorem 6.5. *Let $s \geq 0$. Then $(B_{t+s} - B_s)_{t\geq 0}$ is independent of $\mathcal{F}_s^+$.*

[2] L.T.: Initially, this was not quite obvious to me (i.e. given a continuous function $f : (0,1) \to \mathbb{R}$ s.t. $\lim_{t\searrow 0, t\in\mathbb{Q}} f(t) = 0$, then $\lim_{t\searrow 0} f(t)$), but it turns out that this is very easily proved by contradiction.

Proof. Let $A \in \mathcal{F}_s^+$ and let $0 \le t_1, \dots, t_k$ and $F : (\mathbb{R}^d)^k \to \mathbb{R}$ continuous and bounded. We need to show that

$$\mathbb{E}[F(B_{t_1+s} - B_s, \dots, B_{t_k+s} - B_s)1(A)] = \mathbb{E}[F(B_{t_1+s} - B_s, \dots, B_{t_k+s} - B_s)]\mathbb{P}(A).$$

Let $s_n \searrow s$ as $n \to \infty$. By continuity of B and F,

$$\lim_{n\to\infty} F(B_{t_1+s_n} - B_{s_n}, \dots, B_{t_k+s_n} - B_{s_n}) = F(B_{t_1+s} - B_s, \dots, B_{t_k+s} - B_s).$$

The simple Markov property of B now gives

$$\mathbb{E}[F(B_{t_1+s_n} - B_{s_n}, \dots, B_{t_k+s_n} - B_{s_n})1(A)] = \mathbb{E}[F(B_{t_1+s} - B_s, \dots, B_{t_k+s} - B_s)]\mathbb{P}(A).$$

The claim follows e.g. from the dominated convergence theorem. □

Theorem 6.6 (Blumenthal's 0-1 law)**.** *$\mathcal{F}_0^+$ is trivial, i.e. for all $A \in \mathcal{F}_0^+$, $\mathbb{P}(A) \in \{0, 1\}$.*

Proof. Let $A \in \mathcal{F}_0^+ \subseteq \sigma(B_t, t \ge 0)$. By the previous theorem, A is independent of $\mathcal{F}_0^+$, so A is independent from itself, so $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ and the result follows. □

Theorem 6.7. *Let B be a standard Brownian motion in dimension $d = 1$. Define $\tau = \inf\{t > 0 : B_t > 0\}$ and $\sigma = \inf\{t > 0 : B_t = 0\}$. Then $\sigma = \tau = 0$ a.s.*

Proof. For all $n \in \mathbb{N}$, we have $\{\tau = 0\} = \bigcap_{k\ge n}\{\exists 0 < \varepsilon < \frac{1}{k} : B_\varepsilon > 0\} \in \mathcal{F}^B_{\frac{1}{n}}$. So $\{\tau = 0\} \in \mathcal{F}_0^+$ and thus $\mathbb{P}(\tau = 0) \in \{0, 1\}$ by Blumenthal's 0-1 law. Next we have

$$\mathbb{P}(\tau = 0) = \lim_{t\searrow 0} \mathbb{P}(\tau \le t) \ge \lim \mathbb{P}(B_t > 0) = \frac{1}{2},$$

so $\mathbb{P}(\tau = 0) = 1$.

Of course, we also have $\mathbb{P}(\inf\{t > 0 : B_t < 0\} = 0) = 1$. Then $\mathbb{P}(\sigma = 0) = 1$ by the intermediate value theorem. □

Proposition 6.8. *Let $(B_t)_{t\ge 0}$ be a standard Brownian motion in $d = 1$. Let $S_t = \sup_{s\le t} B_s, I_t = \inf_{s\le t} B_s$. Then*

(1) $\mathbb{P}(\forall\varepsilon > 0 : S_\varepsilon > 0) = \mathbb{P}(\forall\varepsilon > 0 : I_\varepsilon < 0) = 1$.

(2) $\sup_{t\ge 0} B_t = \infty$, $\inf_{t\ge 0} B_t = -\infty$ a.s.

Proof. Let $t_n \searrow 0$. Then $\{\forall\varepsilon > 0 : S_\varepsilon > 0\} \supseteq \{B_{t_n} > 0 \text{ for infinitely many } n\}$. This event is independent of $B_{t_1}, \dots, B_{t_{k-1}}$ for all k, so it is in $\mathcal{F}^B_{t_k}$ for all k, hence in $\mathcal{F}_0^+$. So $\mathbb{P}(B_{t_n} > 0 \text{ i.o.}) \in \{0, 1\}$. Furthermore,

$$\mathbb{P}(B_{t_n} > 0 \text{ i.o.}) = \mathbb{P}(\limsup\{B_{t_n} > 0\}) \ge \limsup \mathbb{P}(B_{t_n} > 0) = \frac{1}{2}.$$

Hence $\mathbb{P}(\forall \varepsilon > 0 : S_\varepsilon > 0) = 1$ by symmetry, the same holds for I_ε (or apply to $-B$). This proves (1).

For (2) note that $\sup_{t \geq 0} B_t = \sup_{t \geq 0} B_{\lambda t} \overset{d}{=} \sup_{t \geq 0} \sqrt{\lambda} B_t = \sqrt{\lambda} \sup_{t \geq 0} B_t$. So $S_\infty = \sup_{t \geq 0} B_t$ as the same distribution as aS_∞ for all $a > 0$. So $\mathbb{P}(S_\infty \geq x) = \mathbb{P}(aS_\infty \geq x) = \mathbb{P}(S_\infty \geq \frac{x}{a}) \to 1$ as $a \to \infty$, hence $\mathbb{P}(S_\infty \geq x) = 1$ for all x, so $S_\infty = \infty$ a.s. □

Proposition 6.9. *Let $(B_t)_{t \geq 0}$ be a standard Brownian motion in $\mathbb{R}^d$ and C a cone in $\mathbb{R}^d$ with origin at 0 and non-empty interior, so $C = \{tu : t > 0, u \in A\}$ where A is an subset of $S^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ with non-empty interior. Define $H_C = \inf\{t > 0 : B_t \in C\}$. Then $\mathbb{P}(H_C = 0) = 1$.*

Proof. As before we have $\{H_C = 0\} \in \mathcal{F}_0^+$, so $\mathbb{P}(H_C = 0) \in \{0, 1\}$. We have

$$\mathbb{P}(H_C = 0) = \lim_{t \searrow 0} \mathbb{P}(H_C \leq t) \geq \mathbb{P}(B_t \in C),$$

and $\mathbb{P}(B_t \in C) = \mathbb{P}(\sqrt{t} B_t \in C) = \mathbb{P}(B_1 \in C) > 0$ since C has non-empty interior. □

Theorem 6.10 (Strong Markov property)**.** *Let T be a stopping time with $\mathbb{P}(T < \infty) = 1$ and let B be a standard Brownian motion. Then $(B_{t+T} - B_T, t \geq 0)$ is a standard Brownian motion, independent of $\mathcal{F}_T^+$.*

Proof. Let $T_n = 2^{-n} \lceil 2^n T \rceil$. It is again a stopping time. Let $B_*(t) = B_{t+T_n} - B_{T_n}$. We will prove that B_* is a Brownian motion, independent of $\mathcal{F}_{T_n}^+$. B_* is clearly continuous in t. Let $E \in \mathcal{F}_{T_n}^+$. Define $B_t^{(k)} = B_{t+k2^{-n}} - B_{k2^{-n}}$ for $t \geq 0$. This is a standard Brownian motion (so has the same law as B), independent of $\mathcal{F}_{k2^{-n}}^+$. For any event A,

$$\begin{aligned}
\mathbb{P}(B_* \in A,\, E) &= \sum_{k=0}^{\infty} \mathbb{P}(B_* \in A,\, T_n = k2^{-n},\, E) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(B^{(k)} \in A,\, \underbrace{T_n = k2^{-n},\, E}_{\in \mathcal{F}_{k2^{-n}}^+}) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(B^{(k)} \in A) \mathbb{P}(E,\, T_n = 2k^{-n}) \\
&= \mathbb{P}(B \in A) \sum_{k=0}^{\infty} \mathbb{P}(E,\, T_n = 2k^{-n}) \\
&= \mathbb{P}(B \in A) \mathbb{P}(E).
\end{aligned}$$

So $\mathbb{P}(B_* \in A,\, E) = \mathbb{P}(B \in A)\mathbb{P}(E)$. Taking $E = \Omega$ gives $\mathbb{P}(B_* \in A) = \mathbb{P}(B \in A)$, so B_* has the same law as B, so B_* is a standard Brownian motion. We also get that it is independent of $\mathcal{F}_{T_n}^+$.

By continuity of Brownian motion, $B_{t+s+T} - B_{s+T} = \lim_{n\to\infty} B_{t+s+T_n} - B_{s+T_n}$. Since $B_{t+s+T_n} - B_{s+T_n} \sim \mathcal{N}(0,(t-s)I)$, we get $B_{t+s+T} - B_{s+T} \sim \mathcal{N}(0,(t-s)I)$[3].

Let $0 \le t_1, t_2, \dots, t_k$ and $A \in \mathcal{F}_T^+$. Let $F : (\mathbb{R}^d)^k \to \mathbb{R}$ be continuous and bounded. We have to show that

$$\mathbb{E}[F(B_{t_1+T} - B_T, \dots, B_{t_k+T} - B_T)1(A)] = \mathbb{E}[F(B_{t_1+T} - B_T, \dots, B_{t_k+T} - B_T)]\mathbb{P}(A).$$

Since $A \in \mathcal{F}_T^+$, we get $A \in \mathcal{F}_{T_n}^+$ for all n. Then

$$\mathbb{E}[F(B_{t_1+T_n} - B_{T_n}, \dots, B_{t_k+T_n} - B_{T_n})1(A)] = \mathbb{E}[F(B_{t_1+T_n} - B_{T_n}, \dots, B_{t_k+T_n} - B_{T_n})]\mathbb{P}(A)$$

Now let $n \to \infty$ and use dominated convergence. □

Remark. Let B be a one-dimensional Brownian motion. Take $\tau = \inf\{t \ge 0 : B_t = \max_{0\le s\le 1} B_s\}$. Then τ is not a stopping time:

(1) $\tau < 1$ a.s.,

(2) If τ were a stopping time, then by the strong Markov property, $(B_{t+\tau} - B_\tau)$ would be a Brownian motion. Then for $t > 0$ sufficiently small, $B_{t+\tau} - B_\tau$ would have to be negative (by definition of τ), a contradiction.

Theorem 6.11 (Reflection principle)**.** *Let B be a standard Brownian motion and let T be an a.s. finite stopping time. Define*

$$\widetilde{B}_t = B_t 1(t \le T) + (2B_T - B_t)1(t > T).$$

Then $\widetilde{B}$ is also a standard Brownian motion. We call $\widetilde{B}$ Brownian motion reflected at *T.*

Proof. Define $B_t^{(T)} = B_{t+T} - B_T$ for $t \ge 0$. This is a Brownian motion, independent of $\mathcal{F}_T^+$ by the strong Markov property. So $B^{(T)}$ is independent of $(B_t)_{0\le t\le T}$. Also $-B^{(T)}$ is a Brownian motion, independent of $\mathcal{F}_T^+$. So

$$((B_t)_{0\le t\le T},\, B^{(T)}) \overset{(d)}{=} ((B_t)_{0\le t\le T},\, -B^{(T)}).$$

Let f, g be continuous paths with $g(0) = 0$. Let

$$\psi_T(f,g)(t) = f_t 1(t \le T) + (f_T + g_{t-T})1(t > T)$$

be the concatenation of f, g. Let $\mathcal{A}$ be the product σ-algebra on $C = C([0,\infty))$. Then $\psi_T : C \times C \to C$ and ψ_T is measurable w.r.t. $\mathcal{A} \otimes \mathcal{A}$ and $\mathcal{A}$.

Then $\psi_T(B, B^{(T)}) = B$ and $\psi_T(B, -B^{(T)}) = \widetilde{B}$. It follows that $B \overset{d}{=} \widetilde{B}$. □

[3] L.T.: Why?

Corollary 6.12 (Reflection principle)**.** *Let B be a standard Brownian motion in $d = 1$ and let $a \le b$, $b > 0$. Define $S_t = \sup_{0 \le s \le t} B_s$. Then*

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

Proof. Let $T_b = \inf\{t \ge 0 : B_t = b\}$. Since $\sup_{t \ge 0} B_t = \infty$ a.s., we get $T_b < \infty$ a.s. Note that $\{S_t \ge b\} = \{T_b \le t\}$. Consider $\widetilde{B}$ reflected at T_b. We have $B_{T_b} = b$ by continuity of Brownian motion. Then

$$\mathbb{P}(S_t \ge b, B_t \le a) = \mathbb{P}(T_b \le t, B_t \le a) = \mathbb{P}(T_b \le t, \widetilde{B}_t \ge 2b - a).$$

Now $\{\widetilde{B}_t \ge 2b - a\} \subseteq \{T_b \le t\}$ as $a \le b$. Then

$$\mathbb{P}(T_b \le t, \widetilde{B}_t \ge 2b - a) = \mathbb{P}(\widetilde{B}_t \ge 2b - a) = \mathbb{P}(B_t \ge 2b - a).$$

□

Corollary 6.13. $S_t \overset{d}{=} |B_t|$.

Proof.

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

□

Corollary 6.14. *If $T_x = \inf\{t \ge 0 : B_t = x\}$, then $T_x \overset{d}{=} \left(\frac{x}{B_1}\right)^2$.*

6.3 Martingales for Brownian motion

Theorem 6.15. *Let B be a standard Brownian motion in $d = 1$. Then (B_t) and $(B_t^2 - t)$ are $(\mathcal{F}_t^+)$-martingales.*

Proof. Integrability is clear (by the Gaussian property). If $s < t$, then

$$\mathbb{E}[B_t \mid \mathcal{F}_s^+] = \mathbb{E}[B_t - B_s \mid \mathcal{F}_s^+] + \mathbb{E}[B_s \mid \mathcal{F}_s^+] = B_s \text{ a.s.}$$

by the strong Markov property. Similarly

$$\begin{aligned}\mathbb{E}[B_t^2 \mid \mathcal{F}_s^+] &= \mathbb{E}[(B_t - B_s)^2 + 2B_s(B_t - B_s) + B_s^2 \mid \mathcal{F}_s^+] \\ &= \mathbb{E}[(B_t - B_s)^2] + 0 + B_s^2 \\ &= t - s + B_s^2\end{aligned}$$

□

Corollary 6.16. *Let B be a standard Brownian motion in $d = 1$, $x, y > 0$. Then*

$$\mathbb{P}(T_y < T_{-x}) = \frac{x}{x+y}$$

and

$$\mathbb{E}[T_y \wedge T_{-x}] = xy.$$

Theorem 6.17. *Let B be a standard Brownian motion in $\mathbb{R}^d$, $u \in \mathbb{R}^d$. Then*

$$M_t^u = \exp\Big(\langle u, B_t\rangle - \frac{|u|^2 t}{2}\Big)$$

is an $(\mathcal{F}_t^+)$-martingale.

Proof. For integrability note that $\mathbb{E}[\exp(\langle u, B_t\rangle)] = \exp(\frac{|u|^2 t}{2})$. For $t > s$, we have

$$\begin{aligned}\mathbb{E}[\exp(\langle u, B_t\rangle) \mid \mathcal{F}_s^+] &= \mathbb{E}[\exp(\langle u, B_t - B_s\rangle)\exp(\langle u, B_s\rangle) \mid \mathcal{F}_s^+] \\ &= \exp(\langle u, B_s\rangle)\exp\Big(\frac{|u|^2 t}{2}\Big).\end{aligned}$$

□

Let $(S_n)_{n\geq 0}$ be a simple symmetric random walk on $\mathbb{Z}$. Given a function f, we want to modify $f(S_n)$ to get martingale. We have

$$\begin{aligned}\mathbb{E}[f(S_{n+1}) - f(S_n) \mid S_0, \dots, S_n] &= \frac{1}{2}f(S_n+1) + \frac{1}{2}f(S_n-1) - f(S_n) \\ &= \frac{1}{2}(f(S_n+1) - 2f(S_n) + f(S_n-1)) \\ &= \frac{1}{2}\widetilde{\Delta} f(S_n)\end{aligned}$$

where $\widetilde{\Delta} f(x) = f(x+1) - 2f(x) + f(x-1)$ is the discrete Laplacian. So

$$\Big(f(S_n) - \frac{1}{2}\sum_{k=0}^{n-1}\widetilde{\Delta} f(S_k)\Big)_{n\geq 0}$$

is a martingale.

Going from (S_n) to Brownian motion, we will replace $\widetilde{\Delta}$ by the Laplacian Δ.

Theorem 6.18. *Let (B_t) be a standard Brownian motion in $\mathbb{R}^d$, $d \geq 1$. Let $f(t,x)$: $\mathbb{R}_+ \times \mathbb{R}^d$ be C^1 in t and C^2 in x. Suppose everything is bounded. Then*

$$M_t = f(t, B_t) - f(0, B_0) - \int_0^t \Big(\frac{\partial}{\partial s} + \frac{1}{2}\Delta\Big) f(s, B_s)ds, \ t \geq 0$$

is an $(\mathcal{F}_t^+)$-martingale.

Proof. M_t is integrable since f and its derivatives are bounded. Next the martingale property. We have

$$M_{t+s} - M_s = f(t+s, B_{t+s}) - f(s, B_s) - \int_0^t \Big(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\Big) f(s+r, B_{s+r}) dr$$

So

$$\begin{aligned}
\mathbb{E}[f(t+s, B_{t+s}) \mid \mathcal{F}_s^+] &= \mathbb{E}[f(t+s, B_{t+s} - B_s + B_s) \mid \mathcal{F}_s^+] \\
&= \int_{\mathbb{R}^d} f(t+s, x+B_s) \mathbb{P}(B_{t+s} - B_s \in dx) \\
&= \int f(t+s, x+B_s) \mathbb{P}(B_t \in dx) \\
&= \int f(t+s, x+B_s) p_t(0,x) dx
\end{aligned}$$

where $p_t(x,y) = (2\pi t)^{-d/2} e^{-\frac{|x-y|^2}{2t}}$. By Fubini, we have

$$\begin{aligned}
\mathbb{E}\Big[\int_0^t \Big(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\Big) f(r+s, B_{r+s}) dr \Big| \mathcal{F}_s^+\Big] &= \int_0^t \mathbb{E}\Big[t\Big(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\Big) f(r+s, B_{r+s}) \Big| \mathcal{F}_s^+\Big] dr \\
&= \int_0^t \int_{\mathbb{R}^d} \Big(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\Big) f(r+s, x+B_s) p_r(0,x) dx dr \\
&= \lim_{\varepsilon \to 0} \int_\varepsilon^t \int_{\mathbb{R}^d} \Big(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\Big) f(r+s, x+B_s) p_r(0,x) dx dr
\end{aligned}$$

Integrating by parts, we have:

$$\begin{aligned}
&\int_\varepsilon^t \int_{\mathbb{R}^d} \Big(\frac{\partial}{\partial r} + \frac{1}{2}\Delta\Big) f(r+s, x+B_s) p_r(0,x) dx dr \\
&\quad = \int_{\mathbb{R}^d} f(s+t, x+B_s) p_t(0,x) - f(s+\varepsilon), B_s + x) p_\varepsilon(0,x) dx \\
&\quad - \int_{\mathbb{R}^d} \int_\varepsilon^t f(r+s, x+B_s) \frac{\partial}{\partial r} p_r(0,x) dr dx \\
&\quad + \int_\varepsilon^t \int_{\mathbb{R}^d} f(r+s, x+B_s) \frac{1}{2}\Delta p_r(0,x) dx dr
\end{aligned}$$

We have $\frac{\partial}{\partial} p_r = \frac{1}{2}\Delta p_r$ (Heat equation). So the last two integrals cancel. Then

$$\begin{aligned}
\mathbb{E}[M_{t+s} - M_s \mid \mathcal{F}_s^+] + f(s, B_s) &= \lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} f(s+\varepsilon, B_s + x) p_\varepsilon(0,x) dx \\
&= \lim_{\varepsilon \to 0} \mathbb{E}[f(s+\varepsilon, B_{s+\varepsilon}) \mid \mathcal{F}_s^+]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[f(s, B_s) \mid \mathcal{F}_s^+] \\
&= f(s, B_s)
\end{aligned}
$$

□

6.4 Transience and recurrence

If B is a Brownian motion with $B_0 = x$, we can write $B_t = x + \widetilde{B}_t$, where $\widetilde{B}$ is a standard Brownian motion. Write $\mathbb{P}_x$ for the probability measure to indicate that B starts at x.

Theorem 6.19. *Let B be a Brownian motion in $d \geq 1$.*

(1) If $d = 1$, then B is point-reccurrent, i.e. for all x, the set $\{t \geq 0 : B_t = x\}$ is unbounded a.s.

(2) If $d = 2$, then for all $x, z \in \mathbb{R}^d, \varepsilon > 0$, the set $\{t : |B_t - z| \leq \varepsilon\}$ is unbounded $\mathbb{P}_x$-a.s. (“neighborhood recurrent”*) However, $\mathbb{P}_0(\exists t > 0 : B_t = x) = 0$.*

(3) If $d \geq 3$, then $\mathbb{P}_0(|B_t| \to \infty$ as $t \to \infty) = 1$ (“transient”*)*

Proof. The $d = 1$ case is immediate, since $\limsup B_t = \infty, \liminf B_t = -\infty$ a.s.

Next let $d = 2$. It suffices to take $z = 0$. First we prove that $\{y : |y| \leq \varepsilon\}$ is hit with probability 1 under $\mathbb{P}_x$. Let $\varphi \in C_b^2(\mathbb{R}^2)$ be such that $\varphi(y) = \log|y|$ if $\varepsilon \leq |y| \leq R$. Let $T_\varepsilon = \inf\{t \geq 0 : |B_t| \leq \varepsilon\}$ and $T_R = \inf\{t \geq 0 : |B_t| = R\}$. We first want to determine $\mathbb{P}_x(T_\varepsilon < T_R)$. $\Delta\varphi = 0$ in the annulus $\varepsilon < |x| < R$. By Theorem 6.18, $M_t = \varphi(B_t) - \frac{1}{2}\int_0^t \Delta\varphi(B_s)ds$ is a martingale. Take $H = T_\varepsilon \wedge T_R$. Then $H < \infty$ a.s. and $M_{t\wedge H} = \varphi(B_{t\wedge H})$ is a bounded martingale. By the optional stopping theorem, $\mathbb{E}_x[\varphi(B_H)] = \log|x|$, so

$$
\mathbb{P}_x(T_\varepsilon < T_R)\log\varepsilon + \mathbb{P}_x(T_R < T_\varepsilon)\log R = \log|x|.
$$

Also $\mathbb{P}_x(T_\varepsilon < T_R) + \mathbb{P}_x(T_\varepsilon > T_R) = 1$, so we obtain

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

We have $\lim_{R\to\infty} T_R = \infty$ a.s., hence $\mathbb{P}_x(T_\varepsilon < \infty) = 1$, so $\mathbb{P}_x(|B_t| \leq \varepsilon$ for some $t > 0) = 1$.

Next,

$$
\begin{aligned}
\mathbb{P}_x(|B_t| \leq \varepsilon \text{ for some } t > n) &= \int \mathbb{P}_x(|B_t - B_n + y| \leq \varepsilon \text{ for } t > n)p_n(x, y)dy \\
&= \int \mathbb{P}_y(|B_t| \leq \varepsilon \text{ for } t > 0)p_n(x, y)dy = 1
\end{aligned}
$$

So $\{t \geq 0 : |B_t| \leq \varepsilon\}$ unbounded $\mathbb{P}_x$-a.s. Taking $\varepsilon \to 0$ in $(*)$ gives $\mathbb{P}_x(T_0 < T_R) = 0$, so $\mathbb{P}_x(T_0 < \infty) = 0$, in other words

$$
\mathbb{P}_x(\exists t > 0 : B_t = 0) = 0.
$$

This holds for all $x \neq 0$. So it remains to show

$$0 = \mathbb{P}_0(\exists t > 0 : B_t = 0)$$

We have

$$\begin{aligned}\mathbb{P}_0(\exists t > 0 : B_t = 0) &= \lim_{a \searrow 0} \mathbb{P}_0(\exists t > a : B_t = 0) \\ &= \lim_{a \searrow 0} \int \underbrace{\mathbb{P}_x(\exists t > 0 : B_t = 0)}_{=0} p_a(0, x) dx = 0\end{aligned}$$

Finally, let $d \geq 3$. By only considering the first three coordinates, we can assume $d = 3$. Then argue similarly as before, but with $f(y) = \frac{1}{y} = \frac{1}{|y|}$ in $\varepsilon \leq |y| \leq R$. Then we get

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

Define $A_n = \{|B_k| \geq n \text{ for all } t \geq T_{n^3}\}$. Then

$$\mathbb{P}_0(A_n^c) = \mathbb{P}_0(\exists t \geq 0 : |B_{t+T_{n^3}}| \leq n)$$

noting that $T_{n^3} < \infty$ $\mathbb{P}_0$-a.s. $(B_{t+T_{n^3}} - B_{T_{n^3}})_{t \geq 0}$ is a standard Brownian motion, independent from $B_{T_{n^3}}$ by the strong Markov property. Then

$$\mathbb{P}_0(A_n^c) = \mathbb{E}_0[\mathbb{P}_{B_{T_{n^3}}}(T_n < \infty)] = \frac{n}{n^3} = \frac{1}{n^2}$$

So $\sum \mathbb{P}_0(A_n^c) < \infty$, hence eventually we are in (A_n) a.s., so $|B_t| \to \infty$ a.s. as $t \to \infty$ □

6.5 Donsker's invariance principle

Recall that on $C([0,1], \mathbb{R})$ we have the sup-norm, given by $\|f\| = \sup_{f \in [0,1]} |f(t)|$. This turns $C([0,1], \mathbb{R})$ into a metric space so we can talk about convergence of probability measures. The product σ-algebra (i.e. the one that makes the coordinate maps measurable) coincides with the Borel σ-algebra.

Theorem 6.20 (Donsker's invariance principle). *Let $(X_n)_{n \geq 1}$ be a sequence of i.i.d. $\mathbb{R}$-valued random variables with $\mathbb{E}[X_1] = 0, \operatorname{Var}(X_1) = \sigma^2$. Set $S_0 = 0$, $S_n = X_1 + \cdots + X_n$ and define*

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

for $t > 0$. Then

$$S^{[N]} = \left(\frac{S_{Nt}}{\sqrt{\sigma^2 N}} \right)_{0 \leq t \leq t}$$

converges in distribution to a standard Brownian motion $(B_t)_{t \in [0,1]}$, i.e. for all continuous, bounded functions $F : C([0,1], \mathbb{R}) \to \mathbb{R}$,

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

as $N \to \infty$.

Theorem 6.21 (Skorokhod embedding). *Let μ be a distribution on $\mathbb{R}$ with mean* 0 *and variance σ^2. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a standard Brownian motion (B_t) is defined and a filtration $(\mathcal{F}_t)_{t\geq 0}$ and a sequence of stopping times $0 = T_0 \leq T_1 \leq \dots$ such that setting $S_n = B_{T_n}$, we have*

(1) (T_n) is a random walk with increments of mean σ^2,

(2) (S_n) is a random walk with increments distribution μ.

Proof. Define Borel measures $\mu_\pm$ on $[0, \infty)$, by $\mu_\pm(A) = \mu(\pm A)$ for $A \in \mathcal{B}([0, \infty))$. There exists a rich probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a standard Brownian motion (B_t) and i.i.d. $(X_n, Y_n)_{n\in\mathbb{N}}$ with law

$$\nu(dx, dy) = C(x + y)\mu_-(dx)\mu_+(dy)$$

where C is the constant such that $C \int_0^\infty x\mu(dx) = 1$. Set $T_0 = 0$. Inductively, define

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

Now check that this works. □

7 Poisson Random Measures

Recall that X is *Poisson distributed* with parameter $\lambda > 0$, written $X \sim \text{Poi}(\lambda)$ or $X \sim P(\lambda)$, if $\mathbb{P}(X = n) = e^{-\lambda}\frac{\lambda^n}{n!}$ when $n \in \mathbb{N}_0$. We write $X \sim \text{Poi}(0)$ if $X \equiv 0$ and $X \sim \text{Poi}(\infty)$ if $X \equiv \infty$.

Proposition 7.1.

(1) *(Addition property) Let (N_k) be independent random variables with distribution $P(\lambda_k)$. Then $\sum_k N_k \sim P(\sum \lambda_k)$.*

(2) *(Splitting/Thinning property) Let $N \sim P(\lambda)$ and let (Y_n) be i.i.d., independent of N, and let $\mathbb{P}(Y_i = j) =: p_j$ for $j = 1, \dots, k$. Let $N_j = \sum_{m=1}^{N} 1(Y_m = j)$. Then $N_1, \dots, N_k$ are independent and $N_i \sim P(\lambda p_i)$ for all $i = 1, \dots, k$.*

Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space. A *Poisson random measure* with intensity μ is a map

$$M : \Omega \times \mathcal{E} \to \mathbb{Z}_+ \cup \{\infty\}$$

satisfying for any disjoint sets (A_k) in $\mathcal{E}$:

(i) $M(\bigcup_k A_k) = \sum_k M(A_k)$,

(ii) $(M(A_k))_k$ are independent random variables,

(iii) For any k, $M(A_k) \sim P(\mu(A_k))$.

Here the dependence on Ω is omitted.

In other words, for each $\omega \in \Omega$, we get a measure $M(\omega, -)$ and for each $A \in \mathcal{E}$, we get a Poisson random variable $M(-, A)$.

Let $E^* = \{\mathbb{Z}_+ \cup \{\infty\}\text{-valued measures on } \mathcal{E}\}$. Define $X : E^* \times \mathcal{E} \to \mathbb{Z}_+ \cup \{\infty\}$ and for $A \in \mathcal{E}$, $X_A : E^* \to \mathbb{Z}_+ \cup \{\infty\}$ by $X(m, A) = X_A(m) = m(A)$ (i.e. canonical pairing). Let $\mathcal{E}^* = \sigma(X_A : A \in \mathcal{E})$.

Theorem 7.2. *There exists a unique probability measure μ^* on $(E^*, \mathcal{E}^*)$ such that under μ^*, X is a Poisson random measure with intensity μ.*

Proof. **Uniqueness.** Let $A_1, \dots, A_k$ be disjoint and $n_1, \dots, n_k \in \mathbb{Z}_+$. Then let $A^* = \{m \in E^* : m(A_1) = n_1, \dots, m(A_k) = n_k\}$. The collection of such sets is a π-system that generates the σ-algebra $\mathcal{E}^*$. For μ^* as in the statement of the theorem, we must have

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

so μ^* is uniquely determined.

Existence. First assume $\mu(E) < \infty$. Let $N \sim P(\mu(E))$ real valued and (Y_n) Ω-valued be independent (and independent of N) with law $\frac{\mu}{\mu(E)}$ (on some perhaps different probability space Ω). For all $A \in \mathcal{E}$, set

$$M(A) = \sum_{n=1}^{N} 1(Y_n \in A).$$

Let $A_1, \dots, A_k$ be disjoint. Define the real-valued random variable X_n by $X_n = k$ iff $Y_n \in A_j$. Then

$$M(A_i) = \sum_{n=1}^{N} 1(Y_n \in A_1) = \sum_{n=1}^{N} 1(X_n = j).$$

Now $\mathbb{P}(X_n = j) = \mathbb{P}(Y_n \in A_j) = \frac{\mu(A_j)}{\mu(E)}$. By the splitting probability we get that the $M(A_i)$ are independent and $M(A_j) \sim P(\mu(E)\frac{\mu(A_j)}{\mu(E)}) = P(\mu(A_j))$.

General case. Since E is σ-finite, there exist disjoint sets $(E_k)_{k\in\mathbb{N}}$ such that $\bigcup_k E_k = E$ and $\mu(E_k) < \infty$ for all k. Let $(M_k)_k$ be independent Poisson random measures with intensity $\mu|_{E_k}(-) := \mu(- \cap E_k)$. For $A \in \mathcal{E}$, define

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

We have $M_k(A \cap E_k) \sim \mathrm{Poi}(\mu(A \cap E_k))$. By the addition property, $M(A) \sim \mathrm{Poi}(\mu(A))$. The independence is clear by the independence of the (M_k). We can view M as a random variable $\Omega \to E^*$. Then take μ^* to be the law of M. □

Proposition 7.3. *Let M be a Poisson random measure of intensity μ. If A is such that $\mu(A) < \infty$, then given $M(A) = k$, we can write $M = \sum_{i=1}^{k} \delta_{X_i}$, where $X_1, \dots, X_k$ are i.i.d. and $X_i \sim \frac{\mu(-\cap A)}{\mu(A)}$.*