1 Conditional Expectation

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $(X_i)_{i\in I}$ be a collection of random variables defined on this space. Then we define $\sigma(X_i : i \in I) \subseteq \mathcal{F}$ to be the smallest σ-algebra such that all of the X_i are measurable, i.e

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

Definition. If $B \in \mathcal{F}$ has $\mathbb{P}(B) > 0$ then we define

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

for any $A \in \mathcal{F}$. Furthermore, if X is an integrable random variable we define

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

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say a σ-algebra $\mathcal{G}$ is countably generated if there exist $(B_i)_{i\in I}$ pairwise disjoint (with I countable) such that $\bigcup_{i\in I} B_i = \Omega$ and $\mathcal{G} = \sigma(B_i : i \in I)$.

Let X be an integrable random variable and $\mathcal{G}$ a countably generated σ-algebra. We want to define $X' = \mathbb{E}[X|\mathcal{G}]$. So define

$$X'(\omega) = \mathbb{E}[X|B_i] \text{ whenever } \omega \in B_i.$$

Or equivalently,

$$X'(\omega) = \sum_{i\in I} \mathbb{E}[X|B_i]\mathbb{1}(\omega \in B_i)$$

where we use the convention that $\mathbb{E}[X|B_i] = 0$ if $\mathbb{P}(B_i) = 0$. Then X' is indeed $\mathcal{G}$-measurable (note $\mathcal{G}$ is the set of $\bigcup_{j\in J} B_j$ for $J \subseteq I$).

Note that for any $G \in \mathcal{G}$ we have $\mathbb{E}[X\mathbb{1}(G)] = \mathbb{E}[X'\mathbb{1}(G)]$. Also

$$\mathbb{E}[|X'|] \le \mathbb{E}\left[\sum_{i\in I} \mathbb{E}[|X||B_i]\mathbb{1}(B_i)\right] = \sum_{i\in I} \mathbb{E}[|X||B_i]\mathbb{P}(B_i) = \mathbb{E}|X| < \infty$$

so X' is integrable.

Theorem (Monotone convergence theorem)**.** *Let $(X_n)_{n\ge 1}$ be a sequence of non-negative random variables with $X_n \uparrow X$ as $n \to \infty$ almost-surely. Then $\mathbb{E}X_n \uparrow \mathbb{E}X$ as $n \to \infty$.*

Proof. See Part II Probability & Measure. □

Theorem (Dominated convergence theorem)**.** *Let $(X_n)_{n\ge 1}$ be a sequence of random variables with $X_n \to X$ as $n \to \infty$ almost-surely and $|X_n| \le Y$ almost-surely for some Y integrable. Then $\mathbb{E}X_n \to \mathbb{E}X$ as $n \to \infty$.*

Proof. See Part II Probability & Measure. □

Definition (L^p)**.** Let $p \in [1, \infty]$ and f be a measurable function. Define the *L^p-norm*

$$||f||_p = (\mathbb{E}[|f|^p])^{1/p} \text{ for } p \in [1, \infty)$$
$$||f||_\infty = \inf\{\lambda : |f| \le \lambda \text{ a.e}\}.$$

Furthermore write $f \sim g$ if $f = g$ almost-everywhere. Then define the L^p-space $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{f : ||f||_p < \infty\}/\sim$.

Theorem ($\mathcal{L}^2$ is a Hilbert space)**.** *$\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with inner product $\langle U, V\rangle = \mathbb{E}[UV]$. For a closed subspace $\mathcal{H}$, if $f \in \mathcal{L}^2$ there exists a unique $g \in \mathcal{H}$ with $||f - g||_2 = \inf\{||f - h||_2 : h \in \mathcal{H}\}$ and $\langle f - g, h\rangle = 0$ for all $h \in \mathcal{H}$. g is called the orthogonal projection of f on $\mathcal{H}$.*

Proof. See Part II Probability & Measure. □

Theorem (Conditional expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra, $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists an integrable random variable Y satisfying*

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

(b) for all $A \in \mathcal{G}$, $\mathbb{E}[X\mathbb{1}(A)] = \mathbb{E}[Y\mathbb{1}(A)]$.

Moreover Y is unique, in the sense that if Y' also satisfies (a) and (b), then $Y = Y'$ almost-surely. We call Y a version of the conditional expectation of X given $\mathcal{G}$. We write $Y = \mathbb{E}[X|\mathcal{G}]$ almost-surely. If $\mathcal{G} = \sigma(Z)$ for a random variable Z, then we write $\mathbb{E}[X|Z] = \mathbb{E}[X|\mathcal{G}]$.

Remark. (b) could be replaced by $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all Z bounded and $\mathcal{G}$-measurable.

Proof. First we show uniqueness. Suppose Y and Y' both satisfy (a) and (b) and let $A = \{Y > Y'\} \in \mathcal{G}$. Then

$$\mathbb{E}[Y\mathbb{1}(A)] = \mathbb{E}[Y'\mathbb{1}(A)] \Rightarrow \mathbb{E}[(Y-Y')\mathbb{1}(A)] = 0 \Rightarrow \mathbb{P}(Y > Y') = 0 \Rightarrow Y \leq Y' \text{ a.s}$$

and similarly $Y \geq Y'$ a.s.

Now we show existence. First assume $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $\mathcal{L}^2(\mathcal{F})$. Hence

$$\mathcal{L}^2(\mathcal{F}) = \mathcal{L}^2(\mathcal{G}) \oplus \mathcal{L}^2(\mathcal{G})^\perp$$

so we can write $X = Y + Z$ for $Y \in \mathcal{L}^2(\mathcal{G})$ and $Z \in \mathcal{L}^2(\mathcal{G})^\perp$. Define $\mathbb{E}[X|\mathcal{G}] = Y$, so Y is $\mathcal{G}$-measurable and for all $A \in \mathcal{G}$

$$\mathbb{E}[X\mathbb{1}(A)] = \mathbb{E}[Y\mathbb{1}(A)] + \underbrace{\mathbb{E}[Z\mathbb{1}(A)]}_{=0} = \mathbb{E}[Y\mathbb{1}(A)].$$

We claim that if $X \geq 0$ almost-surely, then $Y \geq 0$ almost-surely. Indeed, let $A = \{Y < 0\} \in \mathcal{G}$ so $0 \leq \mathbb{E}[X\mathbb{1}(Y < 0)] = \mathbb{E}[Y\mathbb{1}(Y < 0)] \leq 0$ which implies $\mathbb{P}(Y < 0) = 0$.

Assume now that $X \geq 0$ almost-surely. Define $X_n = X \wedge n \leq n$, so $X_n \in \mathcal{L}^2$ for all n. Let $Y_n = \mathbb{E}[X_n|\mathcal{G}]$. Then X_n is an increasing sequence and by the above claim, Y_n is also an increasing sequence almost-surely. Define $Y = \limsup_{n\to\infty} Y_n$, so Y is $\mathcal{G}$-measurable. Also $Y =\uparrow \lim_{n\to\infty} Y_n$ almost-surely. For any $A \in \mathcal{G}$ we have

$$\mathbb{E}[X\mathbb{1}(A)] = \lim_{n\to\infty} \mathbb{E}[X_n\mathbb{1}(A)] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbb{1}(A)] = \mathbb{E}[Y\mathbb{1}(A)]$$

by the Monotone Convergence Theorem.

Finally, for general X write $X = X^+ - X^-$ and define $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]$. □

Remark. From the last proof we can see that we can define $\mathbb{E}[X|\mathcal{G}]$ for $X \geq 0$ without assuming integrability of X. It satisfies all the conditions apart from integrability.

Definition. Let $(\mathcal{G}_n)_{n\geq 1}$ be sub σ-algebras of $\mathcal{F}$. We call them *independent* if whenever $G_i \in \mathcal{G}_i$ and $i_1 < i_2 < \ldots < i_k$ we have

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

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

Properties of conditional expectation

Let $X, Y \in \mathcal{L}^1$, $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra. Then

1. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$ (take $A = \Omega$);
2. If X is $\mathcal{G}$-measurable then $\mathbb{E}[X|\mathcal{G}] = X$ almost-surely (X clearly satisfies the conditions);
3. If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ almost-surely;
4. If $X \geq 0$ almost-surely then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost-surely;
5. For $\alpha, \beta \in \mathbb{R}$, $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$ almost-surely;
6. $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ almost-surely.

Recall:

Theorem (Fatou's Lemma)**.** *If $X_n \geq 0$ for all n almost-surely, then*

$$\mathbb{E}[\liminf_{n\geq 1} X_n] \leq \liminf_{n\geq 1} \mathbb{E}X_n.$$

Proof. See Part II Probability & Measure. □

Theorem (Jensen's Inequality)**.** *If X is integrable, $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, then*

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]).$$

We consider any analogues of our convergence theorems for conditional expectation.

Theorem (Conditional Monotone Convergence Theorem)**.** *Suppose $X_n \geq 0$ for all n and $X_n \uparrow X$ almost-surely as $n \to \infty$. Let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$ almost-surely.*

Remark. Note that $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$ in the almost-sure sense, as these are random variables.

Proof. Let $Y_n = \mathbb{E}[X_n|\mathcal{G}]$ almost-surely. Then Y_n is increasing. Set $Y = \limsup_{n\geq 1} Y_n$. Since Y_n is $\mathcal{G}$-measurable, Y is $\mathcal{G}$-measurable. Also $Y =\uparrow \lim_{n\geq 1} Y_n$ almost-surely. We need to show $\mathbb{E}[Y\mathbb{1}(A)] = \mathbb{E}[X\mathbb{1}(A)]$ for all $A \in \mathcal{G}$. This follows from the usual Monotone Convergence Theorem as

$$\mathbb{E}[Y\mathbb{1}(A)] = \lim_{n\geq 1} \mathbb{E}[Y_n\mathbb{1}(A)] = \lim_{n\geq 1} \mathbb{E}[X_n\mathbb{1}(A)] = \mathbb{E}[X\mathbb{1}(A)].$$

$\square$

Theorem (Conditional Fatou's Lemma)**.** *Let $(X_n)_{n\geq 1}$ be a non-negative sequence of random variables. Then*

$$\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \text{ almost-surely.}$$

Proof. Note that $\inf_{k\geq n} X_k \uparrow \liminf_{n\to\infty} X_n$ so by the conditional MCT

$$\lim_{n\to\infty} \mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}].$$

We also have

$$\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] \leq \mathbb{E}[X_k|\mathcal{G}] \quad \forall k \geq n \text{ almost-surely.}$$

Which implies

$$\mathbb{E}[\inf_{k\geq n} X_k|\mathcal{G}] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{G}] \quad \forall k \geq n \text{ almost-surely}$$

since k takes countable values (intersection of countable sets of full measure also has full measure). Now taking limits as $n \to \infty$ we are done. $\square$

Theorem (Conditional Dominated Convergence Theorem)**.** *Suppose $X_n \to X$ almost-surely, $|X_n| \leq Y$ almost-surely with Y integrable. Then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ almost-surely.*

Proof. We apply the Conditional Fatou's Lemma. Indeed $-Y \leq X_n \leq Y$ so $X_n + Y \geq 0$ and $Y - X_n \geq 0$ for all n. By Conditional Fatou's Lemma

$$\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[X+Y|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty}(X_n+Y)] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$$

and

$$\mathbb{E}[Y|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty}(Y - X_n)|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}] + \liminf_{n\to\infty} \left(-\mathbb{E}[X_n|\mathcal{G}]\right).$$

Hence $\limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \leq \mathbb{E}[X|\mathcal{G}]$ and $\liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \geq \mathbb{E}[X|\mathcal{G}]$ almost-surely.

$\square$

Theorem (Conditional Jensen's Inequality)**.** *Let X be integrable, $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function such that $\varphi(X)$ is integrable or $\varphi(X) \geq 0$. Then $\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}])$ almost-surely.*

Proof. We claim that $\varphi(x) = \sup_{i\in\mathbb{N}}(a_i x + b_i)$, $a_i, b_i \in \mathbb{R}$.

Then $\varphi(X) = \sup_{i\in\mathbb{N}}(a_i X + b_i)$. So

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \sup_{n\geq 1}(a_i \mathbb{E}[X|\mathcal{G}] + b_i) \quad \forall i \in \mathbb{N} \text{ almost-surely.}$$

□

Note. We need the supremum in the claim to be over a countable set so we can preserve the almost-sue property of an inequality.

Corollary. *For all $p \in [1, \infty)$ we have*

$$||\mathbb{E}[X|\mathcal{G}]||_p \leq ||X||_p.$$

Proof. Apply conditional Jensen ($x \mapsto x^p$ is convex). □

Theorem (Tower property)**.** *Let X be integrable and $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ sub σ-algebras. Then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] \text{ almost-surely.}$$

Proof. $\mathbb{E}[X|\mathcal{H}]$ is certainly $\mathcal{H}$-measurable so it remains to check

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbb{1}(A)] \quad \forall A \in \mathcal{H}.$$

But since $A \in \mathcal{G}$ whenever $A \in \mathcal{H}$ we have

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

□

Proposition. Let $X \in \mathcal{L}^1$, $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra, Y bounded and $\mathcal{G}$-measurable. Then

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}] \text{ almost-surely.}$$

Proof. $Y\mathbb{E}[X\mathcal{G}]$ is certainly $\mathcal{G}$-measurable. Also for any $A \in \mathcal{G}$

$$\mathbb{E}[XY\mathbb{1}(A)] = \mathbb{E}[X \underbrace{(Y\mathbb{1}(A))}_{\substack{\text{bounded,}\\ \mathcal{G}\text{-measurable}}}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}](Y\mathbb{1}(A))].$$

□

Definition. Let $\mathcal{A}$ be a collection of sets. It is called a *π-system* if whenever $A, B \in \mathcal{A}$ we have $A \cap B \in \mathcal{A}$.

Recall

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

Proof. See Part II Probability & Measure. □

Theorem. *Let $X \in \mathcal{L}^1$, $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ sub σ-algebras. Assume $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$. Then*

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}] \text{ almost-surely.}$$

Proof. We need to show $\mathbb{E}[X\mathbb{1}(F)] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(F)]$ for all $F \in \sigma(\mathcal{G}, \mathcal{H})$. Define $\mathcal{A} = \{A \cap B : A \in \mathcal{G}, B \in \mathcal{H}\}$. This is a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$. If $F = A \cap B$, $A \in \mathcal{G}, B \in \mathcal{H}$ then

$$\begin{aligned}
\mathbb{E}[X\mathbb{1}(A \cap B)] &= \mathbb{E}[\underbrace{(X\mathbb{1}(A))}_{\sigma(X,\mathcal{G})\text{measurable}} \mathbb{1}(B)] \\
&= \mathbb{E}[X\mathbb{1}(A)]\mathbb{P}(B) \\
&= \mathbb{E}[\underbrace{\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)}_{\substack{\mathcal{G}\\ \text{measurable}}}]\mathbb{P}(B) \\
&= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(A)\mathbb{1}(B)].
\end{aligned}$$

Assume $X \geq 0$. Define $\mu(F) = \mathbb{E}[X\mathbb{1}(F)]$ and $\nu(F) = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}(F)]$ for $F \in \sigma(\mathcal{G}, \mathcal{H})$. Then $\mu = \nu$ on $\mathcal{A}$ by the above and $\mu(\Omega) = \nu(\Omega) < \infty$. Therefore $\mu = \nu$ on $\sigma(\mathcal{G}, \mathcal{H})$. □

Definition. We say $(X_1, \ldots, X_n) \in \mathbb{R}^n$ has the *Gaussian distribution* iff for all $a_1, \ldots, a_n \in \mathbb{R}$

$$a_1X_1 + \ldots + a_nX_n$$

has the Gaussian distribution in $\mathbb{R}$.

A process $(X_t)_{t\geq 0}$ is called a *Gaussian process* if $\forall t_1 < t_2 < \ldots < t_n$, the vector $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian random vector.

Example. Let (X, Y) be a Gaussian vector in $\mathbb{R}^2$. We want to compute $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$. Let $X' = \mathbb{E}[X|Y]$. Since X' is $\sigma(Y)$-measurable it follows X' is a measurable function of Y. So are looking for f Borel such that $\mathbb{E}[X|Y] = f(Y)$ almost-surely. Let $f(y) = ay + b$ for some $a, b \in \mathbb{R}$ to be determined.

Since $\mathbb{E}[X'] = \mathbb{E}[X]$ we have $a\mathbb{E}Y + b = \mathbb{E}X$. Also

$$\begin{aligned}\mathbb{E}[XY] = \mathbb{E}[X'Y] &\implies \mathbb{E}[(X - X')Y] = 0\\ &\implies \mathrm{Cov}(X - X', Y) = 0\\ &\implies \mathrm{Cov}(X, Y) = a\mathrm{Var}(Y)\end{aligned}$$

so we have determined a, b. We need to check that for any Z bounded and $\sigma(Y)$-measurable we have $\mathbb{E}[(X - X')Z] = 0$. Write $Z = g(Y)$ and note $\mathrm{Cov}(X - X', Y) = 0$, implying $X - X'$ is independent of Y. Therefore $\mathbb{E}[(X - X')g(Y)] = \mathbb{E}[X - X']\mathbb{E}[g(Y)] = 0$.

Example. Let (X, Y) be a random vector in $\mathbb{R}^2$ with joint density function $f_{X,Y}(x, y)$. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function such that $h(X)$ is integrable. We want to compute $\mathbb{E}[h(X)|Y]$. Note

$$\mathbb{E}[h(X)g(Y)] = \int_{\mathbb{R}^2} h(x)g(x)f_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y$$

and write

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\mathrm{d}x$$

for the density of Y. So (using the convention $0/0 = 0$)

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x, y)}{f_Y(y)}\mathrm{d}x \right) g(y)f_Y(y)\mathrm{d}y$$

define

$$\varphi(y) = \begin{cases} \int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}\mathrm{d}x & \text{if } f_Y(y) > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Then $\mathbb{E}[h(X)|Y] = \varphi(Y)$ almost-surely.

2 Martingales

2.1 Discrete-time Martingales

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a sequence of increasing sub σ-algebras of $\mathcal{F}$, $(\mathcal{F}_n)_{n\geq 0}$, $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. We call $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ a *filtered probability space*.

If $X = (X_n)_{n\geq 0}$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, define $\mathcal{F}_n^X = \sigma(X_k : k \leq n)$, the *natural filtration* associated with X. We say X is *adapted* to a filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable for all n. X is *integrable* if X_n is integrable for all n.

Definition. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. We say an integrable adapted process $X = (X_n)_{n\geq 0}$ is called a

- *martingale* if

$$\mathbb{E}[X_n|\mathcal{F}_m] = X_m \text{ almost-surely } \forall n \geq m.$$

- *super-martingale* if

$$\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m \text{ almost-surely } \forall n \geq m.$$

- *sub-martingale* if

$$\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m \text{ almost-surely } \forall n \geq m.$$

Remark. If X is a martingale with respect to $(\mathcal{F}_n)$, then it is also a martingale with respect to the natural filtration $(\mathcal{F}_n^X)$.

Example. Let (ξ_i) be a sequence of iid random variables with $\mathbb{E}[\xi_1] = 0$. Let $X_n = \xi_1 + \ldots + \xi_n$, $X_0 = 0$. This is a martingale. We have

$$\mathbb{E}[X_n|\mathcal{F}_{n-1}] = \xi_1 + \ldots + \xi_{n-1} + \mathbb{E}[\xi_n|\mathcal{F}_{n-1}] = \xi_1 + \ldots + \xi_{n-1}$$

by independence.

Example. Let (ξ_i) be a sequence of iid random variables with $\mathbb{E}[\xi_1] = 1$. Let $X_n = \prod_{i=1}^n \xi_i$, $X_0 = 1$. This is a martingale.

Definition. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. A *stopping time* T is a random variable $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ such that $\{T \leq n\} \in \mathcal{F}_n$ for all n.

Note. T being a stopping time is equivalent to $\{T = n\} \in \mathcal{F}_n$ for all n.

Examples.

- Constant times are trivial stopping times;
- Suppose $(X_n)_{n\geq 0}$ is an adapted process taking values in $\mathbb{R}$. For $A \in \mathcal{B}$ define $T_A = \inf\{n \geq 0 : X_n \in A\}$ (with the convention that $\inf \emptyset = \infty$). Then $\{T_A \leq n\} = \bigcup_{k\leq n}\{X_k \in A\} \in \mathcal{F}_n$, so T_A is a stopping time;
- In the setting above, let $L_A = \sup\{n \geq 0 : X_n \in A\}$. This is in general not a stopping time.

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

Proof. Follows directly from the definition. $\square$

Definition. If T is a stopping time, we define

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

If $(X_n)_{n\geq 0}$ is a process, write $X_T(\omega) = X_{T(\omega)}(\omega)$ whenever $T(\omega) < \infty$. We define the *stopped process* $X_t^T = X_{T\wedge t}$.

Proposition. Let S and T be stopping times and let X be an adapted process. Then

1. If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$;
2. $X_T \mathbb{1}(T < \infty)$ is $\mathcal{F}_T$-measurable;
3. X^T is adapted;
4. If X is integrable, then X^T is also integrable.

Proof.

1. Immediate from the definition;
2. Let $A \in \mathcal{B}(\mathbb{R})$. We need to show $\{X_T \mathbb{1}(T < \infty) \in \mathcal{A}\} \in \mathcal{F}_T$. Note that

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

3. $X_t^T = X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable so $\mathcal{F}_t$-measurable by (1).

4. We have

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

□

Theorem (Optional Stopping Theorem)**.** *Let (X_n) be a martingale.*

1. *If T is a stopping time, then X^T is also a martingale. In particular $\mathbb{E}[X_{T\wedge t}] = \mathbb{E}[X_0]$ for all t;*
2. *If $S \leq T$ are bounded stopping times then $\mathbb{E}[X_T|\mathcal{F}_S] = X_S$ almost-surely, and $\mathbb{E}[X_T] = \mathbb{E}[X_S]$;*
3. *If there exists an integrable random variable Y such that $|X_n| \leq Y$ for all n, and T is finite almost-surely then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$;*
4. *If there exists $M > 0$ such that $|X_{n+1} - X_n| \leq M$ for all n, and T is a stopping time with $\mathbb{E}T < \infty$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.*

Proof.

1. We need to show that for all t we have

$$\mathbb{E}[X_{T\wedge t}|\mathcal{F}_{t-1}] = X_{T\wedge(t-1)}$$

almost-surely. Indeed

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

using the fact that $\mathbb{1}(T \geq t)$ is $\mathcal{F}_{t-1}$-measurable;

2. Suppose $S \leq T \leq n$ and let $A \in \mathcal{F}_S$. We need to show $\mathbb{E}[X_T\mathbb{1}(A)] = \mathbb{E}[X_S\mathbb{1}(A)]$. Note

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

Hence

$$\begin{aligned}\mathbb{E}[X_T\mathbb{1}(A)] &= \mathbb{E}[X_S\mathbb{1}(A)] + \sum_{k=0}^{n}\mathbb{E}[(X_{k+1}-X_k)\underbrace{\mathbb{1}(S\le k<T)\mathbb{1}(A)}_{\in\mathcal{F}_k}] \\ &= \mathbb{E}[X_S\mathbb{1}(A)]\end{aligned}$$

since $\mathbb{E}[X_{k+1}|\mathcal{F}_k] = X_k$ almost-surely. Taking expectations gives $\mathbb{E}[X_T] = \mathbb{E}[X_S]$;

3. Example Sheet;

4. Example Sheet.

□

Note. Analogous results follow if (X_n) is instead a sub/super-martingale.

Corollary. *If X is a positive super-martingale, T is a stopping time, $T<\infty$ almost-surely, then $\mathbb{E}[X_T]\le\mathbb{E}[X_0]$.*

Proof. Fatou's lemma gives $\mathbb{E}[\liminf_t X_{T\wedge t}] \le \liminf_t \mathbb{E}[X_{T\wedge t}] \le \mathbb{E}[X_0]$. □

Example. Let $(\xi_i)_{i\ge0}$ be iid with $\mathbb{P}(\xi_0=1)=\mathbb{P}(\xi_0=-1)=1/2$. Define $X_0=0$ and $X_n=\sum_{i=1}^n\xi_i$ for $n\ge1$. Then $(X_n)_{n\ge0}$ is a martingale. Define $T=\inf\{n\ge0 : X_n=1\}$. Then $\mathbb{P}(T<\infty)=1$ and for all t we have $\mathbb{E}[X_{T\wedge t}]=0$, while $\mathbb{E}[X_T]=1$. Hence (4) from the previous theorem tells us $\mathbb{E}T=\infty$.

Example. Consider a SRW on $\mathbb{Z}$, $X_0 = 0$, $X_n = \sum_{i=1}^n \xi_i$ with $(\xi_i)_{i\geq 1}$ iid taking values ± 1 with equal probability. Define $T_c = \inf\{n \geq 0 : X_n = c\}$ and set $T = T_{-a} \wedge T_b$. What is $\mathbb{P}(T_{-a} < T_b)$?

We have that $X_n^T = X_{T\wedge n}$ is a martingale by the optional stopping theorem. Furthermore $|X_{n+1} - X_n| = 1$ for all n. Need to check $\mathbb{E}[T] < \infty$: consider blocks

- $\xi_1, \ldots, \xi_{a+b}$
- $\xi_{a+b+1}, \ldots, \xi_{2(a+b)}$
- $\xi_{2(a+b)+1}, \ldots, \xi_{3(a+b)}$
- $\vdots$

note that the probability the ξ_i in one of these blocks are all equal to either 1 or -1 is $2 \cdot 2^{-(a+b)}$. Hence $T \leq (a+b)\mathrm{Geo}(2 \cdot 2^{-(a+b)})$ and $\mathbb{E}T \leq (a+b)2^{a+b-1} < \infty$.

So applying the optional stopping theorem to T we have $\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0$. Hence $-a\mathbb{P}(T_{-a} < T_b) + b\mathbb{P}(T_b < T_{-a}) = 0$ and $\mathbb{P}(T_{-a} < T_b) + \mathbb{P}(T_b < T_{-a})$, which gives $\mathbb{P}(T_{-a} < T_b) = \frac{b}{a+b}$.

Martingale convergence theoem

Theorem (Almost-sure martingale convergence theorem)**.** *Let X be a supermartingale bounded in $\mathcal{L}^1$, i.e $\sup_{n\geq 0} \mathbb{E}|X_n| < \infty$. Then there exists a random variable $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$ where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$ such that $X_n \to X_\infty$ almost-surely as $n \to \infty$.*

Before we can prove this we will need some preliminary results.

Doob's upcrossing inequality

For a real sequence $(x_n)_{n\geq 0}$, for an interval $[a, b]$ we want to count the number of times (x_n) crosses below a or above b. Define $T_0(x) = 0$ and define for $k \geq 0$

$$\begin{aligned} S_{k+1}(x) &= \inf\{n \geq T_k(x) : x_n \leq a\} \text{ the } (k+1)\text{st downcrossing} \\ T_{k+1}(x) &= \inf\{n \geq S_{k+1}(x) : x_n \geq b\} \text{ the } (k+1)\text{st upcrossing.} \end{aligned}$$

Also let $N_n([a, b], x) = \sup\{k \geq 0 : T_k(x) \leq n\}$, the number of up crossings up to time N. Then as $n \to \infty$, $N_n([a, b], x) \uparrow N([a, b], x) = \sup\{k \geq 0 : T_k(x) < \infty\}$.

Lemma. *Let $x = (x_n)_{n\geq 0}$ be a real sequence. Then x converges in $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ if and only if for all $a < b$, $a, b \in \mathbb{Q}$ we have $N([a, b], x) < \infty$.*

Proof. If x converges then suppose there is $a < b$ with $N([a,b],x) = \infty$. Then

$$\liminf x_n \leq a < b \leq \limsup x_n$$

a contradiction.

Conversely, if x doesn't converge we have $\liminf x_n < \limsup x_n$ so there are $a < b$ (with $a, b \in \mathbb{Q}$) with $\liminf x_n < a < b < \limsup x_n$ and hence $N([a,b],x) = \infty$. □

Now we can prove

Theorem (Doob's upcrossing inequality)**.** *Let X be a supermartingale and $a < b$. Then for all n,*

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

Proof. We have $(T_k)_{k\geq 0}$, $(S_k)_{k\geq 0}$ stopping times. Then

$$\sum_{k=1}^{n}(X_{T_k\wedge n} - T_{S_k\wedge n}) = \sum_{k=1}^{N_n([a,b],X)} \underbrace{(X_{T_k} - X_{S_k})}_{\geq b-a} + \underbrace{(X_n - X_{S_{N_n+1}})\mathbb{1}(S_{N_n+1} \leq n)}_{\geq (X_n - a)\vee 0 = -(X_n-a)^-}.$$

Note $T_k \wedge n, S_k \wedge n$ are stopping times with $T_k \wedge n \geq S_k \wedge n$. Then by the optional stopping theorem $\mathbb{E}[X_{T_k\wedge n}] \leq \mathbb{E}[X_{S_k\wedge n}]$. So taking expectations we have

$$0 \geq (b-a)\mathbb{E}[N_n] - \mathbb{E}[(X_n-a)^-].$$

□

Now we are ready to prove

Theorem (Almost-sure martingale convergence theorem)**.** *Let X be a supermartingale bounded in $\mathcal{L}^1$, i.e $\sup_{n\geq 0}\mathbb{E}|X_n| < \infty$. Then there exists a random variable $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$ where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$ such that $X_n \to X_\infty$ almost-surely as $n \to \infty$.*

Proof. Let $a, b \in \mathbb{Q}$ be such that $a < b$. Then

$$\begin{aligned}\mathbb{E}[N_n([a,b],X)] &\leq (b-a)^{-1}\mathbb{E}[(X_n-a)^-] \\ &\leq (b-a)^{-1}\mathbb{E}[|X_n| + a] \\ &\leq (b-a)^{-1}\left(\sup_{n\geq 0}\mathbb{E}[|X_n|] + 1\right).\end{aligned}$$

We know $N_n([a,b],X) \uparrow N([a,b],X)$ as $n \to \infty$, so by monotone convergence, $\mathbb{E}[N([a,b],X)] < \infty$. Set

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

so $\mathbb{P}(\Omega_0) = 1$ as the intersection of almost-sure events. On Ω_0, X converges by a previous lemma. Set

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

So X_∞ is $\mathcal{F}_\infty$-measurable, and $X_n \to X_\infty$ almost surely. Also

$$\mathbb{E}[|X_\infty|] = \mathbb{E}[\liminf_n |X_n|] \leq \liminf_n \mathbb{E}[|X_n|] < \infty$$

by Fatou. □

Corollary. *Let X be a positive super-martingale. Then X converges almost-surely.*

Proof. $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0]$. So apply the previous. □

Doob's inequalities

Theorem (Doob's maximal inequality)**.** *Let X be a non-negative submartingale. Set $X_n^* = \sup_{0\leq k\leq n} X_k$. Then for all $\lambda \geq 0$*

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

Proof. Let $T = \inf\{k \geq 0 : X_k \geq \lambda\}$. Then T is a stopping time and $\{X_n^* \geq \lambda\} = \{T \leq n\}$. By the optional stopping theorem we have $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n]$ and note

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

Therefore

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

□

Theorem. *Doob's $\mathcal{L}^p$-inequality Let $p > 1$ and let X be a martingale or a non-negative submartingale. Set $X_n^* = \sup_{0\leq k\leq n} |X_k|$. Then*

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

Proof. By Jensen's inequality it is enough to prove for X a non-negative submartingale. Let $k > 0$ and note

$$(y \wedge k)^p = \int_0^k px^{p-1}\mathbb{1}(y \geq x)\mathrm{d}x$$

so

$$\begin{aligned}\|X_n^* \wedge k\|_p^p &= \mathbb{E}[(X_n^* \wedge k)^p]\\ &= \mathbb{E}\left[\int_0^k px^{p-1}\mathbb{1}(X_n^* \geq x)\mathrm{d}x\right]\\ &= \int_0^k px^{p-1}\mathbb{P}(X_n^* \geq x)\mathrm{d}x && \text{(Fubini)}\\ &\leq \int_0^k px^{p-1}x^{-1}\mathbb{E}[X_n\mathbb{1}(X_n^* \geq x)]\mathrm{d}x && \text{(Doob's max inequality)}\\ &= \mathbb{E}\left[\int_0^k px^{p-2}\mathbb{1}(X_n^* \geq x)\mathrm{d}x X_n\right] && \text{(Fubini)}\\ &= \mathbb{E}\left[X_n \frac{p}{p-1}(X_n^* \wedge k)^{p-1}\right]\\ &\leq \frac{p}{p-1}\|X_n\|_p\|X_n^* \wedge k\|_p^{p-1}. && \text{(Hölder)}\end{aligned}$$

Therefore $\|X_n^* \wedge k\|_p \leq \frac{p}{p-1}\|X_n\|_p$. Taking $k \to \infty$ gives the result by monotone convergence.

$\square$

Theorem ($\mathcal{L}^p$-convergence theorems)**.** *Let X be a martingale, $p > 1$. The following are equivalent*

1. *X is bounded in $\mathcal{L}^p$, i.e $\sup_{n\geq 0}\|X_n\|_p < \infty$.*
2. *X converges almost-surely and in $\mathcal{L}^p$ to a limit $X_\infty \in \mathcal{L}^p$.*
3. *There exists $Z \in \mathcal{L}^p$ such that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ almost-surely.*

Proof. (1⇒2) If X is bounded in $\mathcal{L}^p$ then it its bounded in $\mathcal{L}^1$. Hence there exists X_∞ such that $X_n \to X_\infty$ almost-surely as $n \to \infty$. Furthermore

$$\mathbb{E}|X_\infty|^p = \mathbb{E}[\liminf_n |X_n|^p] \leq \liminf_n \mathbb{E}[|X_n|^p] < \infty \qquad \text{(Fatou)}$$

so $X_\infty \in \mathcal{L}^p$. Define $X_n^* = \sup_{0\leq k\leq n}|X_k|$, $X_\infty^* = \sup_{k\geq 0}|X_k|$. Then $|X_n - X_\infty| \leq 2X_\infty^*$ for all n. By dominated convergence it is enough to show $X_\infty^* \in \mathcal{L}^p$. Doob's $\mathcal{L}^p$ inequality gives

$$\|X_n^*\|_p \leq \frac{p}{p-1}\|X_n\|_p \leq \frac{p}{p-1}\sup \_n \geq 0\|X_n\|_p.$$

So by monotone convergence $\|X_\infty^*\|_p < \infty$.

(2⇒3) Set $Z = X_\infty$. Need to show $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ almost-surely. We have for $m \geq n$ that

$$\begin{aligned} \|X_n - \mathbb{E}[X_\infty|\mathcal{F}_n]\|_p &= \|\mathbb{E}[X_m|\mathcal{F}_n] - \mathbb{E}[X_\infty|\mathcal{F}_n]\|_p \\ &\leq \|X_m - X_\infty\|_p \qquad \text{(conditional Jensen)} \\ &\to 0 \text{ as } m \to \infty. \end{aligned}$$

(3⇒1) By conditional Jensen. $\square$

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

Corollary. *If $Z \in \mathcal{L}^p$, $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ almost-surely then $X_n \to \mathbb{E}[Z|\mathcal{F}_\infty]$ almost-surely and in $\mathcal{L}^p$, where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$.*

Proof. By the theorem we have $X_n \to X_\infty$ almost-surely and in $\mathcal{L}^p$. We need to show $X_\infty = \mathbb{E}[Z|\mathcal{F}_\infty]$ almost-surely.

- X_∞ is certainly $\mathcal{F}_\infty$-measurable.

- So we check that for all $A \in \mathcal{F}_\infty$ we have $\mathbb{E}[Z\mathbb{1}(A)] = \mathbb{E}[X_\infty \mathbb{1}(A)]$. Note that $\bigcup_{n\geq 0} \mathcal{F}_n$ is a π-system generating $\mathcal{F}_\infty$ so it suffices to check for A in this π-system. Indeed for such A, there exists $N \geq 0$ such that $A \in \mathcal{F}_N$. Now let $n \geq N$ so

$$\begin{aligned}\mathbb{E}[Z\mathbb{1}(A)] &= \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_N]\mathbb{1}(A)] \\ &= \mathbb{E}[X_N \mathbb{1}(A)] \to \mathbb{E}[X_\infty \mathbb{1}(A)] \text{ as } n \to \infty.\end{aligned}$$

□

Uniform integrability

Recall that a collection $(X_i)_{i\in I}$ of random variables is said to be *uniformly integrable* if

$$\sup_{i\in I} \mathbb{E}[|X_i||1(|X_i| > \alpha)] \to 0 \text{ as } \alpha \to \infty.$$

Equivalently, $(X_i)_{i\in I}$ is uniformly integrable (UI) if it is bounded in $\mathcal{L}^1$ and for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$ we have

$$\sup_{i\in I} \mathbb{E}[|X_i|\mathbb{1}(A)] < \varepsilon.$$

Remark. If $(X_i)_{i\in I}$ is bounded in $\mathcal{L}^p$ for $p > 1$ then it is uniformly integrable.

Lemma. *Let $(X_n)_{n\geq 1}, X$ be in $\mathcal{L}^1$ and $X_n \to X$ almost-surely as $n \to \infty$. Then $X_n \to X$ in $\mathcal{L}^1$ if and only if $(X_n)_{n\geq 1}$ is uniformly integrable.*

Proof. See Part II Probability & Measure. □

Theorem. *Let $X \in \mathcal{L}^1$. The family $\{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra$\}$ is uniformly integrable.*

Proof. We need to show that for all $\varepsilon > 0$, there exists λ large enough such that for any sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we have

$$\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]\mathbb{1}(|\mathbb{E}[X|\mathcal{G}]| > \lambda)] < \varepsilon.$$

Indeed

$$\begin{aligned}\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]\mathbb{1}(|\mathbb{E}[X|\mathcal{G}]| > \lambda)] &\leq \mathbb{E}[\mathbb{E}[|X||\mathcal{G}]\underbrace{\mathbb{1}(|\mathbb{E}[X|\mathcal{G}]| > \lambda)}_{\mathcal{G}\text{-meas}}] \\ &= \mathbb{E}[|X|\mathbb{1}(|\mathbb{E}[X|\mathcal{G}]| > \lambda)].\end{aligned}$$

Since $X \in \mathcal{L}^1$, there exists $\delta > 0$ such that if $A \in \mathcal{F}$ has $\mathbb{P}(A) < \delta$, then $\mathbb{E}[|X|\mathbb{1}(A)] < \varepsilon$. Then

$$\mathbb{P}(|\mathbb{E}[X|\mathcal{G}]| > \lambda) \leq \frac{\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|]}{\lambda} \leq \frac{\mathbb{E}|X|}{\lambda}.$$

So taking $\lambda = \mathbb{E}|X|/\delta$, we are done. □

Definition. $X = (X_n)_{n\geq 0}$ is called a *UI [sub/super] martingale* if it is a [sub/super] martingale and $(X_n)_{n\geq 1}$ is uniformly integrable.

Example. Let $X_1, X_2, \dots$ be iid with $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 2) = 1/2$. Set $Y_0 = 1$ and $Y_n = X_1 X_2 \dots X_n$ for $n \geq 1$, so $(Y_n)_{n\geq 0}$ is a martingale and $\mathbb{E}[Y_n] = 1$ for all n. But $Y_n \to 0$ almost surely.

Theorem. *Let X be a martingale. The following are equivalent*

- *X is UI;*
- *X converges almost surely in $\mathcal{L}^1$ to X_∞ as $n \to \infty$;*
- *There exists $Z \in \mathcal{L}^1$ such that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for all n almost-surely.*

Proof. (1⇒2) X is bounded in $\mathcal{L}^1$, so by the martingale convergence theorem X converges almost-surely to X_∞. Since X is also UI, $X_n \to X_\infty$ in $\mathcal{L}^1$ too.

(2⇒3) Set $Z = X_\infty$. We need to show $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ almost surely. Then for $m \geq n$

$$\begin{aligned}\|X_n - \mathbb{E}[X_\infty|\mathcal{F}_n]\|_1 &= \|\mathbb{E}[X_m - X_\infty|\mathcal{F}_n]\|_1 \\ &\leq \|X_m - X_\infty\|_1 \xrightarrow{m\to\infty} 0.\end{aligned}$$

(3⇒1) The previous theorem implies X is UI. □

Remark. As before we get $X_\infty = \mathbb{E}[Z|\mathcal{F}_\infty]$ almost-surely since $\mathcal{F}_\infty = \sigma(\mathcal{F}_n : n \geq 0)$.

Remark. If X were a UI super/sub-martingale, then we would get $\mathbb{E}[X_\infty|\mathcal{F}_n] \leq X_n$ or $\geq X_n$ respectively.

If X is UI with $X_n \to X_\infty$, and T is a stopping time then

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

Theorem (Optional Stopping Theorem for UI Martingales)**.** *Let X be a UI martingale and let S, T be stopping times with $S \leq T$. Then*

$$\mathbb{E}[X_T|\mathcal{F}_S] = X_S \textit{ almost-surely.}$$

Proof. We know $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ almost-surely since X is UI. It suffices to prove that for any stopping time T, $\mathbb{E}[X_\infty|\mathcal{F}_T] = X_T$ almost-surely. Indeed, then we will have

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

by the tower property since $\mathcal{F}_S \subseteq \mathcal{F}_T$.

So we just establish $\mathbb{E}[X_\infty|\mathcal{F}_T] = X_T$ almost-surely. First we show $X_T \in \mathcal{L}^1$. We have

$$\begin{aligned}
\mathbb{E}[|X_T|] &= \sum_{n\geq 0} \mathbb{E}[|X_n|\mathbb{1}(T=n)] + \mathbb{E}[|X_\infty|\mathbb{1}(T=\infty)] \\
&\leq \sum_{n\geq 0} \mathbb{E}[\mathbb{E}[|X_\infty||\mathcal{F}_n]\mathbb{1}(T=n)] + \mathbb{E}[|X_\infty|\mathbb{1}(T=\infty)] && \text{(Jensen)} \\
&= \sum_{n\geq 0} \mathbb{E}[|X_\infty|\mathbb{1}(T=n)] + \mathbb{E}[|X_\infty|\mathbb{1}(T=n)] \\
&= \mathbb{E}[|X_\infty|] < \infty.
\end{aligned}$$

We have that X_T is $\mathcal{F}_T$-measurable so we need to show that for all $B \in \mathcal{F}_T$, $\mathbb{E}[X_\infty \mathbb{1}(B)] = \mathbb{E}[X_T \mathbb{1}(B)]$. Indeed

$$\begin{aligned}
\mathbb{E}[X_T \mathbb{1}(B)] &= \sum_{n\geq 0} \mathbb{E}[X_n \underbrace{\mathbb{1}(T=n)\mathbb{1}(B)}_{\in \mathcal{F}_n}] + \mathbb{E}[X_\infty \mathbb{1}(B)\mathbb{1}(T=\infty)] \\
&= \sum_{n\geq 0} \mathbb{E}[X_\infty \mathbb{1}(T=n)\mathbb{1}(B)] + \mathbb{E}[X_\infty \mathbb{1}(B)\mathbb{1}(T=\infty)] \\
&= \mathbb{E}[X_\infty \mathbb{1}(B)].
\end{aligned}$$

□

Backwards martingales

Let $\mathcal{F} \supseteq \mathcal{G}_0 \supseteq \mathcal{G}_{-1} \supseteq \ldots$ be a decreasing family of sub-σ-algebras of $\mathcal{F}$. We call $X = (X_n)_{n\geq 0}$ a *backwards martingale* if $X_0 \in \mathcal{L}^1$ and for all $n \leq -1$,

$\mathbb{E}[X_{n+1}|\mathcal{G}_n] = X_n$ almost-surely.

By the tower property, $\mathbb{E}[X_0|\mathcal{G}_n] = X_n$ for all $n \leq 0$ almost-surely. Since $X_0 \in \mathcal{L}^1$, a backwards martingale is automatically UI.

Theorem. *Let X be a backwards martingale with $X_0 \in \mathcal{L}^p$ for $p \in [1, \infty)$. Then $X_n \to X_{-\infty}$ almost-surely and in $\mathcal{L}^p$, where $X_\infty = \mathbb{E}[X_0|\mathcal{G}_{-\infty}]$ for $\mathcal{G}_{-\infty} = \bigcap_{n \geq 0} \mathcal{G}_{-n}$.*

Proof. Set $\mathcal{F}_k = \mathcal{G}_{-n+k}$ for $0 \leq k \leq n$. This is an increasing filtration and $(X_{-n+k})_{0 \leq k \leq n}$ is a $(\mathcal{F}_k)$-martingale. Let $N_{-n}([a,b], X)$ be the number of upcrossings of $[a,b]$ between $-n$ and 0. Doob's upcrossing inequality gives

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

As before, we get $X_n \to X_{-\infty}$ as $n \to -\infty$ almost-surely. $X_{-\infty}$ is $\mathcal{G}_{-\infty}$-measurable (since it's $\mathcal{G}_{-n}$-measurable for all $n \geq 0$, so measurable by the intersection). Since $X_0 \in \mathcal{L}^p$, we have $X_n \in \mathcal{L}^p$ for all $n \leq 0$ by Jensen. Also $X_{-\infty} \in \mathcal{L}^p$ by Fatou.

Now we need to show $X_n \to X_{-\infty}$ in $\mathcal{L}^p$. We have

$$\begin{aligned} |X_n - X_{-\infty}|^p &= |\mathbb{E}[X_0|\mathcal{G}_n] - \mathbb{E}[X_{-\infty}|\mathcal{G}_n]|^p \\ &\leq \mathbb{E}[|X_0 - X_{-\infty}|^p|\mathcal{G}_n] \end{aligned}$$

hence by a previous result, $(|X_n - X_{-\infty}|^p)_n$ is a UI family. Since $X_n \to X_{-\infty}$ almost-surely, we have $\mathcal{L}^1$ convergence of $|X_n - X_{-\infty}|^p$, i.e $\mathcal{L}^p$ convergence of the X_n.

Finally we need to show $X_{-\infty} = \mathbb{E}[X_0|\mathcal{G}_{-\infty}]$ almost-surely. Let $A \in \mathcal{G}_{-\infty} = \bigcap_{n \leq 0} \mathcal{G}_n$ so $A \in \mathcal{G}_n$ for all $n \leq 0$. Then $\mathbb{E}[X_n \mathbb{1}(A)] = \mathbb{E}[X_0 \mathbb{1}(A)]$ for all n. Since $X_n \to X_{-\infty}$ in $\mathcal{L}^1$ we have $\mathbb{E}[X_{-\infty}\mathbb{1}(A)] = \mathbb{E}[X_0 \mathbb{1}(A)]$ and so $X_{-\infty} = \mathbb{E}[X_0|\mathcal{G}_{-\infty}]$. □

Applications of martingales

Theorem (Kolmogorov's 0-1 Law)**.** *Let $(X_n)_{n\geq 0}$ be iid and $\mathcal{F}_n = \sigma(X_k : k \geq n)$ be the tail σ-algebra. Take $\mathcal{F}_\infty = \bigcap_{n\geq 0} \mathcal{F}_n$. Then $\mathcal{F}_\infty$ is trivial, i.e for all $A \in \mathcal{F}_\infty$ we have $\mathbb{P}(A) \in \{0,1\}$.*

Proof. Let $A \in \mathcal{F}_\infty$ and let $\mathcal{G}_n = \sigma(X_k : k \leq n)$ be the natural filtration of the X_n, and $\mathcal{G}_\infty = \sigma(\mathcal{G}_n : n \geq 0)$. Note $(\mathbb{E}[\mathbb{1}(A)|\mathcal{G}_n])_{n\geq 0}$ is a martigale and $\mathbb{E}[\mathbb{1}(A)|\mathcal{G}_n] \to \mathbb{E}[\mathbb{1}(A)|\mathcal{G}_\infty]$ almost-surely. Since $A \in \mathcal{F}_\infty$, we have $A \in \mathcal{F}_{n+1}$ and $\mathcal{G}_n$ is independent of $\mathcal{F}_{n+1}$ by independence of the X_n. So $\mathbb{E}[\mathbb{1}(A)|\mathcal{G}_n] = \mathbb{P}(A)$ almost-surely. Since $\mathcal{F}_\infty \subseteq \mathcal{G}_\infty$ we have $A \in \mathcal{G}_\infty$, we have $\mathbb{E}[\mathbb{1}(A)|\mathcal{G}_\infty] = \mathbb{1}(A)$ almost-surely. Therefore $\mathbb{P}(A) = \mathbb{1}(A)$ almost-surely, so $\mathbb{P}(A) \in \{0,1\}$. □

Theorem (Strong Law of Large Numbers)**.** *Let (X_i) be an iid sequence in $\mathcal{L}^1$ with $\mu = \mathbb{E}[X_1]$. Define $S_n = X_1 + \ldots + S_n$. Then $\frac{S_n}{n}$ converges almost-surely and in $\mathcal{L}^1$ to μ as $n \to \infty$.*

Proof. Define $\mathcal{G}_n = \sigma(S_n, S_{n+1}, \ldots) = \sigma(S_n, X_{n+1}, \ldots)$. For $n \leq -1$ let $M_n = \frac{S_{-n}}{-n}$. We will show $(M_n)_{n\leq -1}$ is a backwards martingale with respect to $(\mathcal{G}_{-n})_{n\leq -1}$. We have

$$\mathbb{E}\left[M_{m+1}|\mathcal{G}_{-m}\right] = \mathbb{E}\left[\frac{S_{-m-1}}{-m-1}|\mathcal{G}_{-m}\right].$$

Take $n = -m$ so this becomes

$$\begin{aligned}
\mathbb{E}\left[\frac{S_{n-1}}{n-1}|\mathcal{G}_n\right] &= \mathbb{E}\left[\frac{S_{n-1}}{n-1}|S_n, X_{n+1}, \ldots\right] \\
&= \mathbb{E}\left[\frac{S_n - X_n}{n-1}|S_n\right] && \text{(independence)} \\
&= S_n - \frac{\mathbb{E}[X_n|S_n]}{n-1} \\
&= S_n - \frac{S_n}{n-1} \\
&= \frac{S_n}{n} \\
&= M_m
\end{aligned}$$

where we used the fact $\mathbb{E}[X_k|S_n] = \mathbb{E}[X_1|S_n]$ for all $k \in [n]$. Hence we have a backwards martingale, so $\frac{S_n}{n} \to Y$ almost-surely and in $\mathcal{L}^1$ for some Y by the Backwards Martingale Theorem.

To finish, we need to show $Y = \mu$ almost-surely. We have

$$Y = \lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{X_{k+1} + \ldots + X_{k+n}}{n} \text{ for all } k.$$

Hene Y is $\sigma(X_{k+1}, \ldots)$ measurable for all k. Hence Y is $\bigcap_{k\geq 0} \sigma(X_{k+1}, \ldots)$-measurable, so by Kolmogorov's 0-1 law Y is almost-surely constant. Since S_n/n converges to Y in $\mathcal{L}^1$, $\lim_{n\to\infty} \mathbb{E}[S_n/n] = \mu = \mathbb{E}Y = Y$. □

Theorem (Radon-Nikodym Theorem)**.** *Let $\mathbb{P}$ and Q be two probability measures on the space $(\Omega, \mathcal{F})$. Suppose $\mathcal{F}$ is countably generated, i.e there exist $(F_n)_{n\geq 1}$ such that $\mathcal{F} = \sigma(F_n : n \geq 1)$. The following are equivalent*

- *$Q \ll \mathbb{P}$, i.e for all $A \in \mathcal{F}$, $\mathbb{P}(A) = 0$ implies $Q(A) = 0$. We say Q is absolutely continuous with respect to $\mathbb{P}$;*
- *For all $\varepsilon > 0$, there exists $\delta > 0$ such that if $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$ then $Q(A) < \varepsilon$;*
- *There exists a non-negative random variable X such that $Q(A) = \mathbb{E}[X\mathbb{1}(A)]$ for all $A \in \mathcal{F}$.*

Note. The general case where $\mathcal{F}$ is not necessarily countably generated follows from this (see Williams).

Remark. X as in (3) is called a *version of the Radon-Nikodym derivative of Q with respect to $\mathbb{P}$*. We write $X = \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}$ on $\mathcal{F}$ almost-surely.

Proof. (1⇒2) If 2 doesn't hold, there exists $\varepsilon > 0$ such that for all n there exists $A_n \in \mathcal{F}$ with $\mathbb{P}(A_n) \leq 1/n^2$ and $Q(A_n) \geq \varepsilon$. Then $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$, so by Borel-Cantelli we see $\mathbb{P}(A_n \text{ i.o}) = 0$. Hence by (1), $Q(A_n \text{ i.o}) = 0$. Since

$$\{A_n \text{ i.o}\} = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k \implies Q(A_n \text{ i.o}) = \lim_{n\to\infty} Q\left(\bigcup_{k\geq n} A_k\right) \geq \varepsilon$$

we have a contradiction.

(2⇒3) Define

$$\mathcal{A}_n = \{H_1 \cap \ldots \cap H_n : H_i = F_i \text{ or } H_i = F_i^c \ \forall i\}$$

and $\mathcal{F}_n = \sigma(\mathcal{A}_n)$. Note the elements of $\mathcal{A}_n$ are disjoint and define $X_n(\omega) = \sum_{A\in\mathcal{A}_n} \frac{Q(A)}{\mathbb{P}(A)} \mathbb{1}(\omega \in A)$. If $A \in \mathcal{F}_n$ we have $\mathbb{E}[X_n\mathbb{1}(A)] = Q(A) = \mathbb{E}[X_{n+1}\mathbb{1}(A)]$. Hence (X_n) is an $(\mathcal{F}_n)$-martingale.

We have $\mathbb{E}[X_n] = Q(\Omega) = 1$, so (X_n) is an $\mathcal{L}^1$-bounded martingale and $X_n \to X_\infty$ almost-surely as $n \to \infty$. Furthermore, $\mathbb{P}(X_n \geq \lambda) \leq \frac{1}{\lambda}$ by Markov's inequality, so for any $\varepsilon > 0$, taking $\delta > 0$ as in (2) and setting $\lambda = 1/\delta$ we have

$$\mathbb{E}[X_n\mathbb{1}(X_n \geq \lambda)] = Q(X_n \geq \lambda) < \varepsilon$$

and so (X_n) is UI. Hence $X_n \to X_\infty$ in $\mathcal{L}^1$. Define $\tilde{Q}(A) = \mathbb{E}[X_\infty\mathbb{1}(A)]$ for all $A \in \mathcal{F}$. Then if $A \in \bigcup_{n\geq 0} \mathcal{F}_n$, $A \in \mathcal{F}_n$ for some n and

$$Q(A) = \mathbb{E}[X_n\mathbb{1}(A)] = \mathbb{E}[X_\infty\mathbb{1}(A)] = \tilde{Q}(A).$$

Since $\bigcup_{n\geq 0} \mathcal{F}_n$ is a π-system generating $\mathcal{F}$, $Q = \tilde{Q}$ on $\mathcal{F}$.

(3⇒1) Trivial. □

Continuous-time Processes

So far, we have considered sequences of random variables $(X_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Equivalently, we have a map $X : (\omega, n) \to X_n(\omega)$. It follows that this map is actually measurable with respect to the product σ-algebra $\mathcal{F} \otimes \mathcal{P}(\mathbb{N})$. Our random variables will be taking values in $E = \mathbb{R}^d$.

We call $(X_t)_{t\in\mathbb{R}^+}$ a *stochastic process* if for all t, X_t is a random variable. However, the map $X : (\omega, t) \mapsto X_t(\omega)$ is not necessarily measurable on $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$.

Proposition. If for all $\omega \in \Omega$, $(0, 1] \to \mathbb{R}^d$ defined by $t \mapsto X_t(\omega)$ is continuous, then $X : (\omega, t) \mapsto X_t(\omega)$ is $\mathcal{F} \otimes \mathcal{B}((0, 1])$-measurable.

Proof. By continuity,

$$X_t(\omega) = \lim_{n\to\infty} \sum_{i=0}^{2^n-1} \mathbb{1}(t \in (k2^{-n}, (k + 1)2^{-n}])X_{k2^{-n}}(\omega).$$

Hence X is measurable as a limit of measurable functions. □

It is enough (and unless stated otherwise we will always assume) that X is right-continuous and admits left-limits almost-everywhere. We call such processes *càdlàg*.

A *filtration* is an increasing family of σ-algebras $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, $\mathcal{F}_t \subseteq \mathcal{F}_{t'}$ for all $t \leq t'$. We say X is *adapted* if X_t is $\mathcal{F}_t$-measurable for all t. A random variable $T : \Omega \to [0, \infty]$ is called a *stopping time* if for all t, $\{T \leq t\} \in \mathcal{F}_t$.

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

For $A \in \mathcal{B}(\mathbb{R})$, $T_A = \inf\{t \geq 0 : X_t \in A\}$ is <u>not</u> always a stopping time. We have

$$\{T_A \leq t\} = \bigcup_{s\leq t} \{X_s \in A\}$$

which is not necessarily in $\mathcal{F}_t$ as we have an uncountable union.

Example. Let

$$J = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$

and

$$X_t = \begin{cases} t & 0 \leq t \leq 1 \\ 1 + J(t - 1) & t > 1 \end{cases}.$$

Let $A = (1, 2)$, then $\{T_A \leq 1\} \notin \mathcal{F}_1$.

We also define the *stopped process* $X^T_t = X_{T\wedge t}$.

Proposition. Let S and T be stopping times and X a càdlàg adapted process. Then

1. If $S \le T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$;
2. $S \wedge T$ is a stopping time;
3. $X_T \mathbb{1}(T < \infty)$ is $\mathcal{F}_T$-measurable;
4. X^T is adapted.

Proof. (1) and (2) are obvious and (4) follows from (3) since $X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable and $\mathcal{F}_{T\wedge t} \subseteq \mathcal{F}_t$. So we just prove (3).

We claim a random variable Z is $\mathcal{F}_T$-measurable if and only if $Z\mathbb{1}(T \le t)$ is $\mathcal{F}_t$-measurable for all t. Indeed, if Z is $\mathcal{F}_T$-measurable then this is immediate by definition of $\mathcal{F}_T$.

Conversely, suppose $Z\mathbb{1}(T \le t)$ is $\mathcal{F}_t$-measurable for all t. If $Z = c\mathbb{1}(A)$ for some $A \in \mathcal{F}$ it is clear. This extends to simple $Z = \sum_{i=1}^n c_i \mathbb{1}(A_i)$, $c_i > 0$, $A_i \in \mathcal{F}$. So writing $Z \ge 0$ as a limit of simple functions $2^{-n}\lfloor 2^n Z\rfloor \wedge n$, we are done.

Now we show $X_T\mathbb{1}(T \le t)$ is $\mathcal{F}_t$ measurable for all t. Since

$$X_T\mathbb{1}(T \le t) = X_T\mathbb{1}(T < t) + \underbrace{X_t\mathbb{1}(T = t)}_{\mathcal{F}_t\text{-measurable}}$$

it suffices to show $X_T\mathbb{1}(T < t)$ is $\mathcal{F}_t$-measurable. Define $T_n = 2^{-n}\lceil 2^n T\rceil$. These are stopping times, since

$$\begin{aligned}\{T_n \le t\} = \{\lceil 2^n T\rceil \le 2^n t\} &= \{2^n T \le \lfloor 2^n t\rfloor\} \\ &= \{T \le 2^{-n}\lfloor 2^n t\rfloor\} \in \mathcal{F}_{2^{-n}\lfloor 2^n t\rfloor} \subseteq \mathcal{F}_t.\end{aligned}$$

By the càdlàg property, $X_T\mathbb{1}(T < t) = \lim_{n\to\infty} X_{T_n\wedge t}\mathbb{1}(T < t)$. T_n takes values in $\mathcal{D}_n = \{k2^{-n} : k \in \mathbb{N}\}$. Note

$$X_{T_n\wedge t}\mathbb{1}(T < t) = \sum_{\substack{d\in\mathcal{D}_n \\ d\le t}} \underbrace{X_d\mathbb{1}(T_n = d)\mathbb{1}(T < t)}_{\mathcal{F}_t\text{-measurable}} + \underbrace{X_t\mathbb{1}(T_n = t)\mathbb{1}(T < t)}_{\mathcal{F}_t\text{-measurable}}.$$

Hence $X_T\mathbb{1}(T < t)$ is $\mathcal{F}_t$-measurable as a limit of $\mathcal{F}_t$-measurable functions. □

Proposition. Let X be continuous and adapated, and let A be a closed set. Then $T_A = \inf\{t \ge 0 : X_t \in A\}$ is a stopping time.

Proof. It suffices to show

$$\{T_A \le t\} = \left\{\inf_{\substack{s\in\mathbb{Q} \\ s\le t}} d(X_s, A) = 0\right\}$$

where $d(x, A) = \inf_{a \in A} |x - a|$. Suppose $T_A = s \leq t$. Then there exists a sequence $(s_n)_{n \geq 1}$ with $s_n \downarrow s$ such that $X_{s_n} \in A$ by definition of T_A. Since A is closed, this means $d(X_{s_n}, A) = 0$. By continuity $X_{s_n} \to X_s$ as $n \to \infty$, so $d(X_s, A) = 0$, implying $X_s = X_{T_A} \in A$. By continuity of X and d, there exists a sequence $(q_n)_{n \geq 1}$ of rationals with $q_n \uparrow s$ such that $d(X_{q_n}, A) \to 0$, and hence $\inf_{\substack{s \in \mathbb{Q} \\ s \leq t}} d(X_s, A) = 0$.

If $\inf_{\substack{s \in \mathbb{Q} \\ s \leq t}} d(X_s, A) = 0$, then there is a sequence $(s_n)_{n \geq 1}$ of rationals with $s_n \leq t$ such that $d(X_{s_n}, A) \to 0$ as $n \to \infty$. So there is a convergent subsequence s_{n_k} of s_n, converging to some $s \leq t$ such that $d(X_{s_{n_k}}, A) \to 0$. Thus by continuity $d(X_s, A) = 0$, and since A is closed, $X_s \in A$ and $T_A \leq t$. $\square$

Define $\mathcal{F}_{t+} = \bigcap_{s > t} \mathcal{F}_s$, a σ-algebra. If for all t, $\mathcal{F}_{t+} = \mathcal{F}_t$, we say that $(\mathcal{F}_t)$ is right-continuous.

Proposition. Let X be a continuous process and A be an open set. Then $T_A = \inf\{t \geq 0 : X_t \in A\}$ is a stopping time with respect to $(\mathcal{F}_{t+})$

Proof. We need to show that for all t, $\{T_A \leq t\} \in \mathcal{F}_{t+}$. Note

$$\{T_A < s\} = \bigcup_{\substack{q \in \mathbb{Q} \\ q < s}} \underbrace{\{X_q \in A\}}_{\in \mathcal{F}_s} \in \mathcal{F}_s.$$

Also

$$\{T_A \leq t\} = \bigcap_n \{T_A < t + 1/n\} \in \mathcal{F}_{t+1/n} \ \forall n$$

so $\{T_A \leq t\} \in \mathcal{F}_{t+}$. □

A stochastic process $(X_t)_{t\geq 0}$ takes values in $\{f : f : \mathbb{R}_+ \to E\}$ (for us usually, $E = \mathbb{R}^d$). Denote by $C(\mathbb{R}_+, E)$ the space of continuous $f : \mathbb{R}_+ \to E$ and by $D(\mathbb{R}_+, E)$ the space of càdlàg $f : \mathbb{R}_+ \to E$.

We will endow C, D with the product σ-algebra that makes all projections $\pi_t : f \mapsto f(t)$ measurable for all t. This σ-algebra is generated by the cylinder sets $\{\bigcap_{s\in J}\{f_s \in A_s\} : J \text{ finite }, A_s \in \mathcal{B}\}$. For A in the product σ-algebra we write $\mu(A) = \mathbb{P}(X \in A)$, and we call μ the law of X.

For every finite subset $J \subseteq \mathbb{R}_+$, write μ_J for the law of $(X_t, t \in J)$. The measures $(\mu_J)_J$ are called the finite dimensional marginals of X. The (μ_J) completely characterise the law of X. This follows because $\{\bigcap_{s\in J}\{X_s \in A_s\} : J \text{ finite}, A_s \in \mathcal{B}(\mathbb{R}_+)\}$ is a π-system generating the σ-algebra on which μ is determined by the (μ_J).

Example. Let $X_t = 0$ for all $t \in [0, 1]$, let $U \sim \text{Uniform}[0, 1]$ and $X'_t = \mathbb{1}(U = t)$ for $t \in [0, 1]$. Both of these have the same finite dimensional marginals, but the two processes are different, since $\mathbb{P}(X_t = 0 \ \forall t \in [0, 1]) = 1$ while $\mathbb{P}(X'_t = 0 \ \forall t \in [0, 1]) = 0$. However $\mathbb{P}(X_t = X'_t) = 1$ for all $t \in [0, 1]$.

Definition. Let X and X' be two processes on $(\Omega, \mathcal{F}, \mathbb{P})$. We say X' is *a version of* X if $\mathbb{P}(X_t = X'_t) = 1$ for all t.

Definition. Given our filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, set $\mathcal{N}$ to be the collection of sets of measure 0. Define

$$\tilde{\mathcal{F}}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N}).$$

If for all t, $\mathcal{F}_t = \tilde{\mathcal{F}}_t$, we say that $(\mathcal{F}_t)$ satisfies the *usual conditions* (u.c).

Theorem (Martingale regularisation theorem)**.** *Let $(X_t)_{t\geq 0}$ be a martingale with respect to $(\mathcal{F}_t)$. Then there exists a càdlàg process $\tilde{X}$ satisfying $\mathbb{P}(X_t = \mathbb{E}[\tilde{X}_t|\mathcal{F}_t]) = 1$ for all t, and $\tilde{X}$ is a martingale with respect to $(\tilde{F}_t)$.*

Remark. If $(\mathcal{F}_t)$ satisfies the usual conditions, then $\tilde{X}$ is a càdlàg version of X.

Lemma. *Let $f : \mathbb{Q}_+ \to \mathbb{R}$ be such that for all $I \subseteq \mathbb{Q}_+$ bounded, f is bounded on I, and for any $a, b \in \mathbb{Q}$ with $a < b$,*

$$N([a,b], I, f) < \infty$$

where $N([a,b], I, f)$ is equal to

$$\sup\{n \geq 0 : \exists 0 < s_1 < t_1 < \ldots < s_n < t_n,\ s_i, t_i \in I,\ f(s_i) < a,\ f(t_i) > b\}.$$

Then for all $t \in \mathbb{R}_+$, the limits

$$\lim_{\substack{s \uparrow t \\ s \in \mathbb{Q}_+}} f(s) \text{ and } \lim_{\substack{s \downarrow t \\ s \in \mathbb{Q}_+}} f(s)$$

exist, and are finite.

Proof. Let $s_n \downarrow t$, $(f(s_n))$ will converge by the finite upcrossing proof. Let $s_n \downarrow t$ and $t_n \downarrow t$ and combine them to get a decreasing sequence, implying $\lim f(s_n) = \lim f(t_n)$. Since f is bounded, these limits are finite. □

Proof of martingale regularisation theorem. We aim to define $\tilde{X}_t = \lim_{\substack{s \downarrow t \\ s \in \mathbb{Q}_+}} X_s$ on a set of measure 1 and $\tilde{X}_t = 0$ otherwise. Steps:

1. Show that the limit exists and is finite;
2. Show $\tilde{X}$ is $\tilde{\mathcal{F}}$-measurable and $X_t = \mathbb{E}[\tilde{X}_t | \mathcal{F}_t]$ almost-surely;
3. Show $\tilde{X}$ satisfies the martingale property;
4. Show $\tilde{X}$ is càdlàg.

Step 1: Let I be a bounded subset of $\mathbb{Q}_+$. We need to show $\mathbb{P}(\sup_{t \in I} |X_t| < \infty) = 1$. We have $\sup_{t \in I} |X_t| = \sup_{\substack{J \subseteq I \\ \text{finite}}} \sup_{t \in J} |X_t|$. Let $J = \{j_1, \ldots, j_n\} \subseteq I$ with $j_1 < \ldots < j_n$ and $K > \sup I$. Then

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

by Doob's maximal inequality (since $(X_t)_{t \in J}$ is a discrete-time martingale). Taking limits as $J \uparrow I$, we have $\lambda \mathbb{P}(\sup_{t \in I} |X_t| \geq \lambda) \leq \mathbb{E}[|X_K|]$. Hence $\mathbb{P}(\sup_{t \in I} |X_t| < \infty) = 1$.

For $M \in \mathbb{N}$, define $I_M = \mathbb{Q}_+ \cap [0, M]$. Then

$$\mathbb{P}\left(\bigcap_{M \in \mathbb{N}} \{ \sup_{t \in I_M} |X_t| < \infty \} \right) = 1.$$

Let $a, b \in \mathbb{Q}$ be such that $a < b$, $I \subseteq \mathbb{Q}_+$ bounded. Then

$$N([a,b], I, X) = \sup_{\substack{J \subseteq I \\ J \text{ finite}}} N([a,b], J, X).$$

Write $J = \{a_1, \dots, a_n\}$ where $a_1 < a_2 < \dots < a_n$ and take $K > \sup I$. Then (X_{a_i}) is a discrete-time martingale so

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

by Doob's upcrossing inequality. By monotone convergence we get

$$(b-a)\mathbb{E}[N([a,b],I,X)] < \infty.$$

Let $M \in \mathbb{N}$ and $I_M = \mathbb{Q}_+ \cap [0,M]$. Define

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

Then on Ω_0, by the previous lemma we have that $\lim_{\substack{s\downarrow t \\ s\in\mathbb{Q}}} X_s$ exists. Furthermore we have $\mathbb{P}(\Omega_0) = 1$.

So define

$$\tilde{X}_t = \begin{cases} \lim_{\substack{s\downarrow t \\ s\in\mathbb{Q}}} X_s & \text{on } \Omega_0 \\ 0 & \text{otherwise} \end{cases}.$$

Step 2: Then $\tilde{X}_t$ is measurable with respect to $\tilde{\mathcal{F}}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N})$ by definition.

Let $t_n \downarrow t$, $t_n \in \mathbb{Q}$. Then $\tilde{X}_t = \lim_{n\to\infty} X_{t_n}$ almost-surely. Note $(X_{t_n})_{n\geq 0}$ is a backwards martingale with respect to $(\mathcal{F}_{t_n})_{n\geq 0}$. Hence by the backwards martingale convergence theorem, (X_{t_n}) converges almost surely and in $\mathcal{L}^1$. So

$$X_t = \mathbb{E}[X_{t_n}|\mathcal{F}_t] \to \mathbb{E}[\tilde{X}_t|\mathcal{F}_t] \text{ in } \mathcal{L}^1.$$

Hence $X_t = \mathbb{E}[\tilde{X}_t|\mathcal{F}_t]$ almost-surely.

Step 3: First we'll show $\mathbb{E}[X_t|\mathcal{F}_{s+}] = \tilde{X}_s$ almost-surely whenever $s < t$.

Claim: for any random variable X and $\mathcal{G}$ a sub-σ-algebra we have $\mathbb{E}[X|\sigma(\mathcal{G},\mathcal{N})] = \mathbb{E}[X|\mathcal{G}]$. Proof: exercise (consider a suitable π-system).

So if we can show $\mathbb{E}[X_t|\mathcal{F}_{s+}] = \tilde{X}_s$, the martingale follows from this claim immediately since the claim says $\mathbb{E}[X_t|\tilde{\mathcal{F}}_s] = \mathbb{E}[X_t|\mathcal{F}_{s+}] = \tilde{X}_s$ almost-surely, and then we can apply the tower law.

Now we show $\mathbb{E}[X_t|\mathcal{F}_{s+}] = \tilde{X}_s$ for $s < t$. Indeed let $s_n \downarrow s$, $s_n \in \mathbb{Q}_+$, $s_0 < t$. Then $\mathbb{E}[X_t|\mathcal{F}_{s_n}]$ is a backwards martingale, so converges a.s and in $\mathcal{L}^2$ to $\mathbb{E}[X_t|\mathcal{F}_{s+}]$. But $\mathbb{E}[X_t|\mathcal{F}_{s_n}] = X_{s_n}$ and $X_{s_n} \to \tilde{X}_s$ a.s, so $\tilde{X}_s = \mathbb{E}[X_t|\mathcal{F}_{s+}]$ a.s.

Step 4: We will show $\tilde{X}$ is right continuous. Suppose not, then there exists $\omega \in \Omega_0$ and some t such that $\tilde{X}(\omega)$ is not right-continuous at t, i.e there exists $s_n \downarrow t$ and some $\varepsilon > 0$ such that $|\tilde{X}_{s_n} - \tilde{X}_t| \geq \varepsilon$ for all n. By the definition of $\tilde{X}$, there exists (s_n') such that $s_n' > s_n$ for all n, $s_n' \in \mathbb{Q}_+$, and $s_n' \downarrow t$ with $|\tilde{X}_{s_n} - X_{s_n'}| \leq \varepsilon/2$. So $|X_{s_n'} - \tilde{X}_t| \geq \varepsilon/2$, a contradiction since $s_n' \downarrow t$ and $s_n' \in \mathbb{Q}_+$. $\square$

Example. Let ξ, η be independent, taking values ± 1 with equal probability. Define

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

and $\mathcal{F}_t = \sigma(X_s : s \leq t)$. Then X is a $\mathcal{F}$-martingale. We have that $\tilde{X}$ satisfies $X_t = \mathbb{E}[\tilde{X}_t | \mathcal{F}_t]$. So we can see

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

Noting that $\mathcal{F}_1 = \sigma(\xi)$ and $\mathcal{F}_t = \sigma(\xi, \eta)$ for $t > 1$. Clearly $\tilde{X}$ is càdlàg with respect to $\tilde{F}$, and note $\mathcal{F}_{1+} = \sigma(\xi, \eta)$. $\mathcal{F}$ is not right-continuous so $\tilde{X}$ is not a version of X.

Theorem (Almost-sure Martingale Convergence Theorem)**.** *Let X be a càdlàg martingale bounded in $\mathcal{L}^1$. Then $X_t \to X_\infty$ almost-surely as $t \to \infty$ with $X_\infty \in \mathcal{L}^1(\mathcal{F}_\infty)$.*

Proof. Take $I_M = \mathbb{Q}_+ \cap [0, M]$ we have (considering sequences and Doob's upcrossing inequality)

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

hence $N([a,b], \mathbb{Q}_+, X) < \infty$ almost-surely for all $a < b$. Define

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

so $\mathbb{P}(\Omega_0) = 1$ and on Ω_0, $\lim_{\substack{q \to \infty \\ q \in \mathbb{Q}}} X_q$ exists and is finite. Define $X_\infty = \lim_{\substack{q \to \infty \\ q \in \mathbb{Q}}} X_q$ on Ω_0.

Then for all $\varepsilon > 0$ there exists q_0 such that $|X_q - X_\infty| \leq \varepsilon/2$ for all $q \in (q_0, \infty) \cap \mathbb{Q}$. Let $t > q_0$, then there exists $q > t$ with $q \in \mathbb{Q}$ such that $|X_t - X_q| \leq \varepsilon/2$ by right-continuity. Hence $|X_t - X_\infty| \leq \varepsilon$. $\square$

Theorem (Doob's Maximal Inequality)**.** *Let X be a càdlàg martingale, $X_t^* = \sup_{s \leq t} |X_s|$. Then for all $\lambda \geq 0$,*

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

Proof. We have

$$\sup_{s \leq t} |X_s| = \sup_{s \in \{t\} \cup (\mathbb{Q}_+ \cap [0,t))} |X_s|$$

and use the beginning of the proof of the martingale regularisation theorem. $\square$

Note. The $\mathcal{L}^p$ convergence theorems etc hold in the same way for continuous càdlàg martingales.

Theorem (Optional Stopping Theorem)**.** *Let X be a càdlàg UI martingale. Then for all $S \leq T$ stopping times,*

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

Proof. Let $T_n = 2^{-n}\lceil 2^n T\rceil$ and $S_n = 2^{-n}\lceil 2^n S\rceil$, so $T_n \downarrow T$ and $S_n \downarrow S$. We need to show that for $A \in \mathcal{F}_S$ we have $\mathbb{E}[X_T \mathbb{1}(A)] = \mathbb{E}[X_S \mathbb{1}(A)]$. We have $X_{T_n} \to X_T$ and $X_{S_n} \to X_S$ almost-surely by right-continuity. We have $X_{T_n} = \mathbb{E}[X_\infty|\mathcal{F}_{T_n}]$ by a discrete result, so (X_{T_n}) is UI, giving $X_{T_n} \to X_T$ and $X_{S_n} \to X_S$ in $\mathcal{L}^1$ as well. By the discrete optional stopping theorem, $\mathbb{E}[X_{T_n}|\mathcal{F}_{S_n}] = X_{S_n}$ a.s. For $A \in \mathcal{F}_S$ we have $A \in \mathcal{F}_{S_n}$ (check), so

$$\mathbb{E}[X_{T_n}\mathbb{1}(A)] = \mathbb{E}[X_{S_n}\mathbb{1}(A)]$$

and taking limits we're done. $\square$

Proposition (Kolmogorov's continuity criterion)**.** Let $\mathcal{D}_n = \{k2^{-n} : 0 \leq k \leq 2^n\}$, $\mathcal{D} = \bigcup_{n\geq 0} \mathcal{D}_n$. Let $(X_t)_{t\in\mathcal{D}}$be a stochastic process taking real values. Suppose there exists $\varepsilon > 0$, $p > 0$, $c > 0$ such that

$$\mathbb{E}[|X_t - X_s|^p] \leq c|t-s|^{1+\varepsilon} \ \forall s, t \in \mathcal{D}.$$

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

Proof. Note that

$$\mathbb{P}(|X_{k2^{-n}} - X_{(k+1)2^{-n}}| \geq 2^{-n\alpha}) \leq 2^{n\alpha p} c 2^{-n(1+\varepsilon)} \tag{Markov}$$

so

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

and therefore summing over n and applying Borel-Cantelli,

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

Hence

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

for some random variable M. Now we show there exists M' with $|X_t - X_s| \leq M'|t-s|^\alpha$. Suppose $s < t$, $s, t \in \mathcal{D}$ and let r be the unique integer such that $2^{-(r+1)} < t - s \leq 2^{-r}$. Then there exists k such that $s < k2^{-(r+1)} < t$, and set $a = k2^{-(r+1)}$. Then $t - a \leq 2^{-r}$ so

$$\begin{aligned} t - a &= \sum_{j\geq r+1} \frac{x_j}{2^j}, \ x_j \in \{0,1\} \\ a - s &= \sum_{j\geq r+1} \frac{y_j}{2^j}, \ y_j \in \{0,1\}. \end{aligned}$$

We can write $[s,t)$ as a disjoint union of dyadic intervals, each of them having length some 2^{-n} for $n \geq r+1$, and each interval of length 2^{-n} will appear at most twice. Hence

$$\begin{aligned}
|X_t - X_s| &\leq \sum_{\substack{d,n \\ d \text{ endpoint of} \\ \text{dyadic interval} \\ \text{of length } 2^{-n}}} |X_d - X_{d+2^{-n}}| \\
&\leq \sum_{d,n} 2^{-n\alpha} M \\
&\leq 2M \sum_{n \geq r+1} 2^{-n\alpha} \\
&= 2M \frac{2^{-(r+1)\alpha}}{1-2^{-\alpha}} \\
&\leq \frac{2M}{1-2^{-\alpha}} |t-s|^\alpha.
\end{aligned}$$

□

Weak convergence

Let (M, d) be a metric space endowed with its Borel σ-algebra.

Definition. Let (μ_n) be a sequence of probability measures on M. We say (μ_n) converges weakly to a measure μ, writing $\mu_n \Rightarrow \mu$ as $n \to \infty$, if $\mu_n(f) \to \mu(f)$ for all $f : M \to \mathbb{R}$ continuous and bounded.

Remark. Taking $f = 1$ gives $\mu(f) = 1$, so μ is necessarily a probability measure.

Example. Let (x_n) be a sequence in M with $x_n \to x$. Then $\delta_{x_n} \Rightarrow \delta_x$. Indeed $\delta_{x_n}(f) = f(x_n) \to f(x) = \delta_x(f)$ by continuity of f.

Example. Let $M = [0, 1]$, endowed with the Borel σ-algebra. Then defining $\mu_n = \frac{1}{n} \sum_{0 \leq k \leq n} \delta_{k/n}$, we have that μ_n converges weakly to the Lebesgue measure. Indeed,

$$\mu_n(f) = \frac{1}{n} \sum_{0 \leq k \leq n} f(k/n)$$

which is the Riemann sum of f, converging to $\int_0^1 f(x) \mathrm{d}x$.

Example. Let $\mu_n = \delta_{1/n}$. Then $\mu_n \Rightarrow \delta_0$. Take $A = (0, 1)$, so $\mu_n(A) = 1$ for all n, but $\mu(A) = 0$, so $\mu_n(A) \not\Rightarrow \mu(A)$.

Theorem. *Let (μ_n) be a sequence of probability measures on (M, d). The following are equivalent:*

1. $\mu_n \Rightarrow \mu$*;*
2. *For all open $G \subseteq M$, $\liminf_n \mu_n(G) \geq \mu(G)$;*

3. *For all closed $A \subseteq M$, $\limsup_n \mu_n(A) \leq \mu(A)$;*
4. *If A has $\mu(\partial A) = 0$, then $\mu_n(A) \to \mu(A)$.*

Proof.

- (1⇒2) Let G be open with $G^c \neq \emptyset$ (empty case is trivial). Then for $M > 0$ define $f_M(x) = 1 \wedge (Md(x, G^c)) \leq \mathbb{1}(x \in G)$. Also $f_M(x) \uparrow \mathbb{1}(x \in G)$ as $M \to \infty$ and f_M is bounded and continuous. Hence $\mu_n(f_M) \to \mu(f_M)$ as $n \to \infty$. Also

$$\liminf_n \mu_n(G) \geq \liminf_n \mu_n(f_M) = \mu(f_M) \to \mu(f)$$

where the last limit follows by monotone convergence.

- (2 $\iff$ 3) Follows by taking complements.
- (2,3⇒4) We have

$$0 = \mu(\partial A) = \mu(\overline{A} \setminus \mathrm{int}(A))$$

so $\mu(\overline{A}) = \mu(A) = \mu(\mathrm{int}(A))$. Then by 2,3 we have

$$\begin{aligned}\liminf_n \mu_n(\mathrm{int}(A)) &\geq \mu(\mathrm{int}(A)) = \mu(A)\\ \limsup_n \mu_n(\overline{A}) &\leq \mu(\overline{A}) = \mu(A).\end{aligned}$$

- (4⇒1) We need to show $\mu_n(f) \to \mu(f)$ for all f continuous and bounded. Let $K > \sup |f|$ and suppose $f \geq 0$. Note

$$\begin{aligned}\mu_n(f) = \int_M f(x)\mathrm{d}\mu_n(x) &= \int_M \left(\int_0^K \mathbb{1}(t \leq f(x))\mathrm{d}x\right)\mathrm{d}\mu_n(x)\\ &= \int_0^K \mu_n(f \geq t)\mathrm{d}t.\end{aligned}$$

It suffices to show $\mu_n(f \geq t) \to \mu(f \geq t)$ for almost-all t by dominated convergence. Note that $\{f \geq t\} = f^{-1}([t, \infty))$ is closed by continuity of f, so $\overline{\{f \geq t\}} = \{f \geq t\}$. Hence $\partial\{f \geq t\} \subseteq \{f = t\}$. We claim there exist at most a countable number of t such that $\mu(f = t) > 0$.

Indeed note $\{t : \mu(f = t) > 0\} = \bigcup_n \{t : \mu(f = t) \geq 1/n\}$ and $\{t : \mu(f = t) \geq 1/n\}$ has cardinality at most n, so $\{t : \mu(f = t) > 0\}$ is countable as a countable union of countable sets.

□

Consider the case $M = \mathbb{R}$ and let μ be a probability measure on $\mathbb{R}$. We define the distribution function of μ, $F_\mu : \mathbb{R} \to [0,1]$ by $F_\mu(x) = \mu((-\infty, x])$.

Proposition. Let (μ_n) be a sequence of probability measures on $\mathbb{R}$. The following are equivalent

(a) $\mu_n \Rightarrow \mu$ as $n \to \infty$;

(b) $F_{\mu_n}(x) \to F(x)$ for all $x \in \mathbb{R}$ such that F_μ is continuous at x.

Proof. (a⇒b) Let x be a continuity point of F_μ. Then $F_{\mu_n}(x) = \mu_n((-\infty, x])$. Note that

$$\mu(\partial(-\infty, x]) = \mu(\{x\}) = \mu((-\infty, x]) - \lim_{n\to\infty}((-\infty, x - 1/n]) = 0$$

by continuity of F_μ at x. By a previous proposition this implies $\mu_n((-\infty, x]) \to \mu((-\infty, x])$.

(b⇒a) Let G be an open set in $\mathbb{R}$. Then $G = \bigcup_k (a_k, b_k)$ for disjoint intervals (a_k, b_k). Then

$$\begin{aligned} \liminf_n \mu_n(G) &= \liminf_n \sum_k \mu_n((a_k, b_k)) \\ &\geq \sum_j \liminf_n \mu_n((a_k, b_k)). \qquad \text{(Fatou)} \end{aligned}$$

So it suffices to show $\liminf_n \mu_n((a,b)) \geq \mu((a,b))$ for all $a < b$. We have $\mu_n((a,b)) = F_{\mu_n}(b-) - F_{\mu_n}(a)$. F_μ is non-decreasing, so it has at most a countable number of discontinuities. Hence there exist a', b' continuity points of F_μ with $a < a' < b' < b$. Then

$$\mu_n((a,b)) \geq F_{\mu_n}(b') - F_{\mu_n}(a') \to F_\mu(b') - F_\mu(a')$$

so $\liminf_n \mu_n((a,b)) \geq F_\mu(b') - F_\mu(a')$. By density of the continuity points, we can take $a_n \downarrow a$ and $b_n \uparrow b$ sequences of continuity points of F_μ, to conclude $\liminf_n \mu_n((a,b)) \geq F_\mu(b-) - F_\mu(a) = \mu((a,b))$.

□

Definition. Let (X_n) be a sequence of random variables taking values in (M,d), with each X_n defined on $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$. We say that X_n *converges weakly*, or *in distribution* to a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ if $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$.

Proposition.

1. If $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$ then $X_n \Rightarrow X$;
2. If $X_n \Rightarrow c$ for c constant, then $X_n \xrightarrow{\mathbb{P}} c$.

Example (CLT)**.** Let (X_n) be iid, $\mathbb{E}X_1 = m$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then defining $S_n = \sum_{i=1}^n X_i$ we have

$$\frac{S_n - nm}{\sqrt{n\sigma^2}} \Rightarrow \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Definition (Tightness)**.** Let (M, d) be a metric space. A sequence of probability measures (μ_n) on M is called *tight* if for all $\varepsilon > 0$, there exists $K \subseteq M$ compact such that $\sup_n \mu_n(K^c) < \varepsilon$.

Remark. If the metric space M is compact, then any sequence of probability measures is tight.

Theorem (Prokhorov)**.** *Let (μ_n) be a tight sequence of probability measures. Then there exists a subsequence (n_k) and a probability measure such that $\mu_{n_k} \Rightarrow \mu$ as $k \to \infty$.*

Proof. We only give a proof in the case $M = \mathbb{R}$. Let $\mathbb{Q} = (x_n)_{n\geq 1}$ be an enumeration of $\mathbb{Q}$. Let $F_n = F_{\mu_n}$ and note $(F_n(x_1))_{n\geq 1}$ is a sequence in $[0, 1]$ so it has a convergent subsequence $F_{n_k^{(1)}} \to F(x_1)$. Now $(F_{n_k^{(1)}}(x_2))_{n\geq 2}$ is also a sequence in $[0, 1]$ so continuing like this, there exists a subsequence $(n_k^{(j)})$ such that $F_{n_k^(j)}(x_j) \to F(x_j)$ for all $j \in \mathbb{N}$.

Now taking the sequence $m_k = n_k^{(k)}$ we have $F_{m_k}(x) \to F(x)$ for all $x \in \mathbb{Q}$. Each F_{m_k} is non-decreasing, so F is non-decreasing as well. Define $F(x) = \lim_{\substack{q\downarrow x \\ q\in\mathbb{Q}}} F(q)$ so F is right-continuous and non-decreasing, so F has left limits and is càdlàg.

Let x be a continuity point of F. We need to show $F_{m_k}(x) \to F(x)$. Then there exist $s_1 < x < s_2$ with $s_1, s_2 \in \mathbb{Q}$ and $|F(s_i) - F(x)| < \varepsilon/2$ for $i = 1, 2$. Hence

$$F(x) - \varepsilon < F(s_1) - \varepsilon/2 \leq F_{m_k}(s_1) \leq F_{m_k}(x) \leq F_{m_k}(s_2) \leq F(s_2) + \varepsilon/2 < F(x) + \varepsilon$$

for all k large enough, so indeed $F_{m_k}(x) \to F(x)$.

Finally we show there is a probability measure μ with $F = F_\mu$. By tightness, for all $\varepsilon > 0$ there exists N large enough so that $-N, N$ are continuity points of F and $\sup_n \mu_n([-N, N]^c) \leq \varepsilon$. Hence $F(-N) \leq \varepsilon$ and $1 - F(N) \leq \varepsilon$ so $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. Define $\mu((a, b)) = F(b) - F(a)$. Then μ can be extended to the Borel σ-algebra by Caratheodory's extension theorem. □

Definition. Let X be a random variable with values in $\mathbb{R}^d$. The *characteristic function* of X is defined $\phi_X(u) = \mathbb{E}[e^{i\langle u,X\rangle}]$ for $u \in \mathbb{R}^d$. We have that

- ϕ_X is continuous, $\phi_X(0) = 1$;
- ϕ_X completely determines the law of X, i.e if $\phi_X(u) = \phi_Y(u)$ for all $u \in \mathbb{R}^d$ then $\mu_X = \mu_Y$.

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

$$\mathbb{P}(\|X\|_\infty \geq K) \leq C \left(\frac{K}{2}\right)^d \int_{[-K^{-1}, K^{-1}]^d} (1 - \phi_X(u)) \mathrm{d}u$$

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

Proof. Note

$$\begin{aligned}
\int_{[-\lambda,\lambda]^d} \phi_X(u) \mathrm{d}u &= \int_{[-\lambda,\lambda]^d} \left(\int \mathbb{R}^d \prod_{j=1}^d e^{iu_j x_j} \mathrm{d}\mu(x) \right) \mathrm{d}u \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^d \left(\int_{[-\lambda,\lambda]} e^{iu_j x_j} \mathrm{d}u_j \right) \mathrm{d}\mu(x) && \text{(Fubini)} \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^d \frac{e^{i\lambda x_j} - e^{-i\lambda x_j}}{ix_j} \mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^d \frac{2\sin(\lambda x_j)}{x_j} \mathrm{d}\mu(x) \\
&= (2\lambda)^d \int_{\mathbb{R}^d} \prod_{j=1}^d \frac{\sin(\lambda x_j)}{\lambda x_j} \mathrm{d}\mu(x).
\end{aligned}$$

Hence,

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

Take $f(u) = \prod_{j=1}^d \frac{\sin(u_j)}{u_j}$. We claim $|\sin(x)/x| \leq \sin(1)$ for $x \geq 1$, so if $\|u\|_\infty \geq 1$ we have $|f(u)| \leq \sin 1$. Hence

$$\mathbb{1}(\|u\|_\infty \geq 1) \leq C(1 - f(u)) \implies \mathbb{P}(\|X\|_\infty \geq K) \leq C\mathbb{E}\left[1 - f\left(\frac{X}{K}\right)\right].$$

$\square$

Theorem (Lévy's Convergence Theorem)**.** *Let $(X_n)_{n\geq 1}, X$ be random variables with values in $\mathbb{R}^d$. Then $\mu_{X_n} \Rightarrow \mu_X$ if and only if $\phi_{X_n}(u) \to \phi_X(u)$ for all $u \in \mathbb{R}^d$.*

We'll actually prove a stronger form:

Theorem (Lévy's Convergence Theorem)**.** *Let $(X_n)_{n\geq 1}, X$ be random variables with values in $\mathbb{R}^d$. Then*

1. *If $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$, then $\phi_{X_n}(\xi) \to \phi_X(\xi)$ for all $\xi \in \mathbb{R}^d$.*

2. *Suppose there exists $\psi : \mathbb{R}^d \to \mathbb{C}$ with $\psi(0) = 1$ and ψ is continuous at 0. Suppose $\phi_{X_n}(\xi) \to \psi(\xi)$ for all $\xi \in \mathbb{R}^d$. Then there exists a random variable X with $\psi = \phi_X$ and $\mu_{X_n} \Rightarrow \mu_X$.*

Proof.

1. Trivial as $x \mapsto e^{i\langle u,x\rangle}$ is bounded and continuous.

2. First we prove that (μ_{X_n}) is tight. By the previous lemma,

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

where $C = 2^{-d}(1 - \sin 1)^{-1}$. Also $|1 - \phi_{X_n}(u)| \leq 2$ for all u, n so by DCT

$$K^d \int_{[-K^{-1},K^{-1}]^d} (1 - \phi_{X_n}(u))\mathrm{d}u \xrightarrow{n\to\infty} K^d \int_{[-K^{-1},K^{-1}]} (1 - \psi(u))\mathrm{d}u.$$

Since ψ is continuous at 0 and $\psi(0) = 1$, taking K sufficiently large we get

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

Therefore $\mathbb{P}(\|X_n\|_\infty \geq K) \leq \varepsilon$ for n large enough. Taking K larger if necessary we have $\sup_{n\geq 0} \mathbb{P}(\|X_n\|_\infty \geq K) \leq \varepsilon$, so (μ_{X_n}) is tight. By Prokhorov's theorem, there is a subsequence (n_k) with $\mu_{X_{n_k}} \Rightarrow \mu_X$ for some random variable X. Therefore $\phi_X = \psi$.

Suppose μ_{X_n} did not converge weakly. Then there is a continuous and bounded f and a subsequence (m_k) such that $|\mathbb{E}[f(X_{m_k}) - f(X)]| > \varepsilon$ for all k. But (μ_{m_k}) is tight, so has a convergent subsequence, giving a contradiction since this limit must also be X.

□

Large Deviations

Let $(X_n)_{n\geq 1}$ be iid $\mathcal{N}(0,1)$. Let $\hat{S}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(0, 1/n)$. Let $\delta > 0$, then

1. $\mathbb{P}(|\hat{S}_n| \geq \delta) \xrightarrow{n\to\infty} 0$.

2. $\mathbb{P}(\sqrt{n}\hat{S}_n \in A) \xrightarrow{n\to\infty} \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x$ by the CLT (even if the X_i are general centred distributions).

3. Note $\mathbb{P}(|\hat{S}_n| \geq \delta) = 1 - \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x$ so $\frac{1}{n}\log \mathbb{P}(|\hat{S}_n| \geq \delta) \xrightarrow{n\to\infty} -\frac{\delta^2}{2}$.

So the "typical value" of $\hat{S}_n$ is of order $1/\sqrt{n}$ and it can take relatively large values ($> \delta$) with very small probability ($e^{-n\delta^2/2}$). Points 1 and 2 above are universal, while 3 depends on the distribution of the X_i.

Let $(X_n)_{n\geq 1}$ be iid with $\mathbb{E}[X_1] = \bar{x}$ and $S_n = X_1 + \ldots + X_n$. Let $a \in \mathbb{R}$, so

$$\mathbb{P}(S_{n+m} \geq a(n+m)) \geq \mathbb{P}(S_n \geq an)\mathbb{P}(S_m \geq am)$$

and if we define $b_n = -\log \mathbb{P}(S_n \geq an)$ we have

$$b_{n+m} \leq b_n + b_m.$$

Exercise: $\lim_{n\geq 1} \frac{b_n}{n}$ exists and $\lim_{n\geq 1} \frac{b_n}{n} = \inf_{n\geq 1} \frac{b_n}{n}$. Hence $-\frac{1}{n} \log \mathbb{P}(S_n \geq an) \xrightarrow{n\to\infty} I(a)$ for some $I(a)$. Let $M(\lambda) = \mathbb{E}[e^{\lambda X_1}]$ and $\psi(\lambda) = \log M(\lambda)$ so

$$\begin{aligned}
\mathbb{P}(S_n \geq an) = \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda a_n}) &\leq \mathbb{E}[e^{\lambda S_n}]e^{-\lambda na} \\
&= (\mathbb{E}[e^{\lambda X_1}])^n e^{-\lambda na} \\
&= \exp(-n(\lambda a - \psi(\lambda))).
\end{aligned}$$

Let $\psi^*(a) = \sup_{\lambda\geq 0}(\lambda a - \psi(\lambda)) \geq 0$ so

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

Theorem (Cramer's Theorem)**.** *Let* $(X_n)_{n\geq 1}$ *be an iid sequence of random variables with* $\mathbb{E}X_1 = \bar{x}$*. Let* $S_n = \sum_{i=1}^n X_i$*. Then*

$$-\frac{1}{n}\log \mathbb{P}(S_n \geq an) \xrightarrow{n\to\infty} \psi^*(a) \ \forall a \geq \bar{x}$$

where $\psi^*(a) = \sup_{\lambda\geq 0}(\lambda a - \psi(\lambda))$ *and* $\psi(\lambda) = \log \mathbb{E}[e^{\lambda X_1}]$*.*

First we prove a preliminary lemma.

Lemma. *Let* $M(\lambda) = \mathbb{E}[e^{\lambda X_1}]$ *and* $\psi(\lambda) = \log \mathbb{E}[e^{\lambda X_1}]$*. Then the functions* M *and* ψ *are continuous in* $D = \{\lambda : M(\lambda) < \infty\}$*, and differentiable in int*(D) *with*

$$M'(\lambda) = \mathbb{E}[X_1 e^{\lambda X_1}] \text{ and } \psi'(\lambda) = \frac{M'(\lambda)}{M(\lambda)} \quad \forall \lambda \in \mathit{int}(D).$$

Proof. Continuity follows by dominated convergence. For the derivative we have

$$\frac{M(\nu+\varepsilon) - M(\nu)}{\varepsilon} = \mathbb{E}\left[\frac{e^{(\nu+\varepsilon)X_1} - e^{\nu X_1}}{\varepsilon}\right]$$

and

$$\left|\frac{e^{(\nu+\varepsilon)X_1} - e^{\nu X_1}}{\varepsilon}\right| \leq e^{\nu X_1}\left|\frac{e^{\varepsilon X_1} - 1}{\varepsilon}\right|.$$

Let $\delta > 0$ be sufficiently small so that $\nu + \delta \in \mathrm{int}(D)$. Take $\varepsilon \in (-\delta, \delta)$. Then

$$\left|\frac{e^{\varepsilon X_1} - 1}{\varepsilon}\right| \leq \frac{e^{\delta |X_1|} - 1}{\delta}$$

so

$$\left|\frac{e^{(\nu+\varepsilon)X_1} - e^{\nu X_1}}{\varepsilon}\right| \leq e^{\nu X_1}\frac{e^{\delta |X_1|} - 1}{\delta}$$

and apply dominated convergence. □

Now we are ready to prove Cramer:

Proof of Cramer's Theorem. By a Chernoff bound we have

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

so we need to show

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

Replace each X_i with $\tilde{X}_i = X_i - a$ and write $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_i$, $\tilde{M}(\lambda) = \mathbb{E}[e^{\lambda \tilde{X}_1}] = e^{-a\lambda} M(\lambda)$, $\tilde{\psi}(\lambda) = \psi(\lambda) - a\lambda$. Then we want to show

$$\begin{aligned}\limsup_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(S_n \geq an) &= \limsup_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\tilde{S}_n \geq 0) \\ &\leq \tilde{\psi}^*(0)\end{aligned}$$

where $\tilde{\psi}^*(0) = \sup_{\lambda\geq 0}(-\tilde{\psi}(\lambda))$. So it suffices to show

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{P}(S_n \geq 0) \geq \inf_{\lambda\geq 0} \psi(\lambda)$$

whenever $\bar{x} \leq 0$. Write $\mu = \mu_{X_1}$ and assume that $M(\lambda) < \infty$ for all $\lambda \geq 0$. Define a new measure for all $\theta \geq 0$ by

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

so

$$\mathbb{E}_\theta[f(X_1)] = \int_{\mathbb{R}} \frac{e^{\theta x} f(x)}{M(\theta)} \mathrm{d}\mu(x).$$

Then if $X_1, \ldots, X_n \sim \mu$ are iid we have

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

Set $g(\theta) = \mathbb{E}_\theta[X_1] = \int_{\mathbb{R}} \frac{e^{\theta x} x}{M(\theta)} \mathrm{d}\mu(x) = \frac{M'(\theta)}{M(\theta)} = \psi'(\theta)$. We find θ such that $g(\theta) = 0$. Suppose $\mu((0,\infty)) = \mathbb{P}(X_1 > 0) > 0$. Then

$$\psi(\theta) = \log \mathbb{E}[e^{\theta X_1}] \implies \lim_{\theta\to\infty} \psi(\theta) = \infty$$

so there exists $\eta \geq 0$ such that $\psi(\eta) = \inf_{\lambda\geq 0} \psi(\lambda)$ and $\psi'(\eta) = 0$, i.e $g(\eta) = 0$. We have

$$\begin{aligned}\mathbb{P}(S_n \geq 0) \geq \mathbb{P}(S_n \in [0, \varepsilon n]) &\geq \mathbb{E}[e^{\eta S_n - \eta\varepsilon n} \mathbb{1}(S_n \in [0, \varepsilon n])] \\ &= e^{-\eta\varepsilon n}(M(\eta))^n \underbrace{\mathbb{P}_\eta(S_n \in [0, \varepsilon n])}_{\to \frac{1}{2} \text{ by CLT}}\end{aligned}$$

using the fact $\mathbb{E}_\eta[X_1] = 0$. Hence

$$\frac{1}{n}\log \mathbb{P}(S_n \geq 0) \geq -\eta\varepsilon + \log M(\eta) + \frac{\log \mathbb{P}_\eta(S_n \in [0, \varepsilon n])}{n}$$

so taking limits

$$\liminf_{n\to\infty} \log \mathbb{P}(S_n \geq 0) \geq \log M(\eta) - \eta\varepsilon \geq \inf_{\lambda\geq 0} \psi(\lambda) - \eta\varepsilon$$

so take $\varepsilon \to 0$. If $\mathbb{P}(X_1 > 0) = 0$, then

$$\mathbb{P}(S_n \geq 0) = (\mu(0))^n \implies \frac{1}{n}\log \mathbb{P}(S_n \geq 0) = \log \mu(0) \geq \inf_{\lambda \geq 0} \psi(\lambda)$$

since $\inf_{\lambda\geq 0}\psi(\lambda) \leq \lim_{\lambda\to\infty}\psi(\lambda) = \log\mu(0)$.

In the general case (not assuming $M(\lambda) < \infty$ for all $\lambda \geq 0$): let $K > 0$ and define $\nu = \mu_{X_1 | |X_1| \leq K}$, $\nu_n = \mu_{S_n | \bigcap_{i=1}^n \{|X_i| \leq K\}}$, $\mu = \mu_{X_1}$ and $\mu_n = \mu_{S_n}$. Then

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

and

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

Define $\psi_K(\lambda) = \log\int_{-K}^K e^{\lambda x}\mathrm{d}\mu(x)$ so

$$\log\int_{-\infty}^{\infty} e^{\lambda x}\mathrm{d}\nu(x) = \psi_K(\lambda) - \log\mu([-K,K])$$

therefore

$$\begin{aligned}
\frac{1}{n}\log\mu_n([0,\infty)) &\geq \frac{\log\nu_n([0,\infty))}{n} + \mu([-K,K]) \\
&\geq \inf_{\lambda\geq 0}\left(\log\int_{-\infty}^{\infty} e^{\lambda x}\mathrm{d}\nu(x)\right) + \log\mu([-K,K]) \\
&= \inf_{\lambda\geq 0}\psi_K(\lambda) := J_K.
\end{aligned}$$

Then $J_K \uparrow J$ as $k \to \infty$ for some J. There exists K large so that $J_K > -\infty$, so take K larger and so that $\mu((0,K)) > 0$. Then $J_K = \inf_{\lambda\geq 0}\psi_K(\lambda)$, which implies $J > -\infty$. Note $\{\lambda : \psi_K(\lambda) \leq T\}$ are compact (by continuity of ψ_K) nested subsets so there exists $\lambda_0 \in \bigcap_{K\in\mathbb{N}}\{\lambda : \psi_K(\lambda) \leq T\}$, so $\psi(\lambda_0) = \lim_K \psi_K(\lambda_0) \leq J$. $\square$

Brownian Motion

Definition. $(B_t)_{t\geq 0}$ is called a *Brownian motion* in $\mathbb{R}^d$ started from $x \in \mathbb{R}^d$ if (B_t) is a continuous process and

(i) $B_0 = x$ almost-surely;

(ii) For all $s < t$, $B_t - B_s \sim \mathcal{N}(0, (t-s)I_d)$;

(iii) (B_t) has independent increments, independent of B_0.

If $x_0 = 0$ we call (B_t) the *standard Brownian motion.*

Note. (ii) & (iii) uniquely characterise the law of (B_t).

Example. Let (B_t) be a standard Brownian motion in $\mathbb{R}$ and $U \sim \text{Unif}([0,1])$. Define

$$\tilde{B}_t = \begin{cases} B_t & t \neq U \\ 0 & t = U \end{cases}.$$

Then $\tilde{B}$ is almost-surely discontinuous, so even though it has the same finite dimensional distribution, it is not a Brownian motion.

Theorem (Weiner)**.** *There exists a Brownian motion on some probability space.*

Proof.

1. We first construct a Brownian motion in $d = 1$. We first construct in $[0,1]$, i.e $(B_t)_{t\in[0,1]}$ for $d = 1$. Let $\mathcal{D}_0 = \{0,1\}$, $\mathcal{D}_n = \{k2^{-n} : 0 \leq k \leq 2^n\}$ and $\mathcal{D} = \bigcup_{n\geq 0} \mathcal{D}_n$. We construct $(B_d)_{d\in\mathcal{D}}$ inductively. Let $(Z_d)_{d\in\mathcal{D}}$ be iid $\mathcal{N}(0,1)$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $\mathcal{D}_0 = \{0,1\}$ let $B_0 = 0$ and $B_1 = Z_1$. Now suppose we've constructed $(B_d)_{d\in\mathcal{D}_{n-1}}$ satisfying (ii) & (iii). For $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ let $d_- = d - 2^{-n}, d_+ = d + 2^{-n} \in \mathcal{D}_{n-1}$. Then set

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

so

$$\begin{aligned} B_d - B_{d-} &= \frac{B_{d_+} - B_{d_-}}{2} + \frac{Z_d}{2^{\frac{n+1}{2}}} \\ B_{d_+} - B_d &= \underbrace{\frac{B_{d_+} - B_{d_-}}{2}}_{:=N_d} - \underbrace{\frac{Z_d}{2^{\frac{n+1}{2}}}}_{:=N'_d}. \end{aligned}$$

So by induction $N_d \sim \mathcal{N}\left(0, \frac{d_+ - d_-}{4}\right) = \mathcal{N}(0, 2^{-n-1})$ and $N'_d \sim \mathcal{N}(0, 2^{-n-1})$. Also by induction N_d and N'_d are independent, so $B_d - B_{d-}$ and $B_{d_+} - B_d$

are Gaussian. To prove they are independent, we show $\mathrm{Cov}(N_d+N'_d, N_d-N'_d)=0$. Indeed

$$\mathrm{Cov}(N_d+N'_d, N_d-N'_d)=\mathrm{Var}(N_d)-\mathrm{Var}(N'_d)=0.$$

So we have checked $(B_d-B_{d-2^{-n}})_{d\in\mathcal{D}_n}$ are indepenent for consecutive intervals. If not consecutive, then express each increment as half the increment of the previous scale plus an independent Gaussian. We have so far constructed $(B_d)_{d\in\mathcal{D}}$ satisfying the assumptions. Note that for $d,q\in\mathcal{D}$, $p>0$

$$\mathbb{E}[|B_d-B_q|^p]=|d-q|^{p/2}\mathbb{E}[|N|^p], \text{ where } Z\sim\mathcal{N}(0,1).$$

And for all $p>0$, $\mathbb{E}|N|^p<|infty$. So by Kolmogorov's continuity criterion, we have that $(B_d)_{d\in\mathcal{D}}$ is almost-surely α-Hölder continuous for all $\alpha<1/2$. So we can extend to all of $[0,1]$. Set $B_t=\lim_{i\to\infty}B_{d_i}$, $d_i\in\mathcal{D}$, $d_i\to t$. It is immediate that $(B_t)_{t\in[0,1]}$ is almost-surely α-Hölder continuous for all $\alpha<1/2$.

We need to check (ii) and (iii). Let $0=t_0\le t_1\le\ldots\le t_k\le 1$. Then $(B_{t_i}-B_{t_{i-1}})_{i=1,\ldots,k}$ are independent Gaussians with variance $t_i-_{t_{i-1}}$. Let $0\le t_0^n\le t_1^n\le\ldots\le t_k^n$ be dyadic rationals with $t_0^n\to t_0,\ldots,t_k^n\to t_k$. Then by continuity

$$B_{t_j^n}-B_{t_{j-1}^n}\xrightarrow{n\to\infty}B_{t_j}-B_{t_{j-1}} \tag{$*$}$$

for all j almost-surely. Hence

$$\begin{aligned}\mathbb{E}\left[\exp\left(i\sum_{j=1}^k u_j(B_{t_j^n}-B_{t_{j-1}^n})\right)\right]&=\prod_{j=1}^k\exp\left(-\frac{u_j^2(t_j^n-t_{j-1}^n)}{2}\right)\\&\xrightarrow{n\to\infty}\prod_{j=1}^k\exp\left(-\frac{u_j^2(t_j-t_{j-1})}{2}\right).\end{aligned}$$

So by Levy's convergence theorem, since the limit is the characteristic function of independent $\mathcal{N}(0,t_j-t_{j-1})$ and since we have $(*)$, this forces the law of $(B_{t_j}-B_{t_{j-1}})_{j=1}^k$ to be independent $\mathcal{N}(0,t_j-t_{j-1})$. Hence $(B_t)_{t\in[0,1]}$ satisfies all the desired properties.

2. Take $\{(B_t^i)_{t\in[0,1]}\}_{i\in\mathbb{N}}$ to be independent Brownian motions. Then define $B_t=B_{t-\lfloor t\rfloor}^{\lfloor t\rfloor}+\sum_{i=0}^{\lfloor t\rfloor-1}B_1^i$.

3. For general d, let $(B_t^1)_{t\ge0},\ldots,(B_t^d)$ be independent 1-dimensional Brownian motions and set $B_t=(B_t^1,\ldots,B_t^d)$ and check this works.

□

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

(a) If U is an orthogonal matrix, then $UB = (UB_t)_{t\geq 0}$ is also a standard Brownian motion. In particular, $-B$ is a standard Brownian motion.

(b) For all $\lambda > 0$, $\left(\frac{B_{\lambda t}}{\sqrt{\lambda}}\right)_{t\geq 0}$ is also a standard Brownian motion.

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

Proof. Follows from definition of Brownian motion. $\square$

Properties of Brownian motion

Proposition (Time inversion)**.** Let B be a standard Brownian motion in one-dimension. Let

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

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

Proof. Let $t_1, \ldots, t_k > 0$. Then $(B_{t_1}, \ldots, B_{t_k})$ is a Gaussian random vector with zero mean and $\mathrm{Cov}(B_s, B_t) = s \wedge t$.

Need to check $(X_{t_1}, \ldots, X_{t_k})$ is Gaussian with zero mean and covariance as above. The vector is certainly Gaussian with zero mean. Furthermore

$$\mathrm{Cov}(X_{t_i}, X_{t_j}) = \mathrm{Cov}(t_1 B_{1/t_1}, t_j B_{1/t_j}) = t_i t_j \mathrm{Cov}(B_{1/t_i}, B_{1/t_j}) = t_i \wedge t_j.$$

Finally we show X is continuous. For $t > 0$ X is clearly continuous since B is. So it suffices to show $X_t \xrightarrow{t\to 0} 0$ almost-surely. Note $(X_t)_{t\in\mathbb{Q}_+} =^d (B_t)_{t\in\mathbb{Q}_+}$ since X, B have the same finite dimensional distribution. Hence $\lim_{\substack{t\downarrow 0 \\ t\in\mathbb{Q}_+}} X_t = \lim_{\substack{t\downarrow 0 \\ t\in\mathbb{Q}_+}} = 0$ almost-surely. Since $\mathbb{Q}_+$ is dense and X is continuous for $t > 0$ we conclude

$$\lim_{t\to 0} X_t = \lim_{t\to 0} B_t = 0$$

almost-surely. $\square$

Corollary. *Let B be a standard Brownian motion in one-dimension. Then $\frac{B_t}{t} \xrightarrow{t\to\infty} 0$ almost-surely.*

Proof. We have $\lim_{t\to\infty} \frac{B_t}{t} = \lim_{t\to 0} tB_{1/t} = 0$ almost-surely by the previous. $\square$

Definition. For $s \geq 0$ let $\mathcal{F}_s^+ = \bigcap_{t>s} \mathcal{F}_t^B$ (where $\mathcal{F}_t^B = \sigma(B_u : u \leq t)$ as before).

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

Proof. We need to show that if $t_1, \ldots, t_k \in \mathbb{R}_+$ and F is continuous and bounded on $(\mathbb{R}^d)^k$ for any $A \in \mathcal{F}_s^+$ we have

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

Indeed, if $s_n \downarrow s$ is strictly decreasing, by continuity we have

$$B_{t_1+s_n} - B_{s_n} \to B_{t_1+s} - B_s \text{ as } n \to \infty \text{ almost-surely.}$$

Hence

$$\begin{aligned}&\mathbb{E}[F(B_{t_1+s}-B_s,\dots,B_{t_k+s}-B_s)\mathbb{1}(A)]\\&=\lim_{n\to\infty}\mathbb{E}[F(B_{t_1+s_n}-B_{s_n},\dots,B_{t_k+s_n}-B_{s_n})\mathbb{1}(A)]\end{aligned}$$

by dominated convergence. Since $A\in\mathcal{F}_s^+$, we have $A\in\mathcal{F}_{s_n}^B$ for all n. Hence by the simple Markov property

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{E}[F(B_{t_1+s_n}-B_{s_n},\dots,B_{t_k+s_n}-B_{s_n})\mathbb{1}(A)]\\&=\lim_{n\to\infty}\mathbb{E}[F(B_{t_1+s_n}-B_{s_n},\dots,B_{t_k+s_n}-B_{s_n})]\mathbb{P}(A)\\&=\mathbb{E}[F(B_{t_1+s}-B_s,\dots,B_{t_k+s}-B_s)]\mathbb{P}(A)\end{aligned}$$

by dominated convergence. □

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

Proof. If $A\in\mathcal{F}_0^+$ then $A\in\sigma(B_t:t\geq 0)$. But $\sigma(B_t:t\geq0)$ is independent of $\mathcal{F}_0^+$ by the previous, so A is independent of itself and $\mathbb{P}(A)=\mathbb{P}(A\cap A)=\mathbb{P}(A)^2$. □

Theorem. *Let B be a standard Brownian motion in one-dimension. Then define $\tau=\inf\{t>0:B_t>0\}$ and $\sigma=\inf\{t>0:B_t=0\}$. Then $\mathbb{P}(\tau=0)=\mathbb{P}(\sigma=0)=1$.*

Proof. Note
$$\{\tau=0\}=\bigcap_{k\geq n}\underbrace{\{\exists 0<\varepsilon<1/k\text{ s.t }B_\varepsilon>0\}}_{\in\mathcal{F}^B_{1/n}}$$

for all n, so $\{\tau=0\}\in\mathcal{F}_0^+$. So $\mathbb{P}(\tau=0)\in\{0,1\}$ and it is enough to show $\mathbb{P}(\tau=0)>0$. For $t>0$ we have $\mathbb{P}(\tau\leq t)\geq\mathbb{P}(B_t>0)=1/2$ so taking the limit $t\downarrow 0$ we have $\mathbb{P}(\tau=0)\geq 1/2$ and so $\mathbb{P}(\tau=0)=1$.

Also note $\inf\{t>0:B_t<0\}=0$ almost-surely since $-B_t$ is also a standard Brownian motion. Since B is continuous, this means $\inf\{t>0:B_t=0\}=0$ almost-surely by the intermediate value theorem. □

Proposition. Let B be a standard Brownian motion in one dimension. Let $S_t=\sup_{s\leq t}B_s, I_t=\inf_{s\leq t}B_s$. Then

1. For all $\varepsilon>0$, $S_\varepsilon>0$ and $I_\varepsilon<0$ almost-surely;
2. $\sup_{t\geq0}B_t=\infty$ almost-surely and $\inf_{t\geq0}B_t=-\infty$ almost-surely.

Proof.

1. Let $t_n\downarrow 0$ as $n\to\infty$. Then $\{S_\varepsilon>0\}\supseteq\{B_{t_n}>0\text{ i.o}\}\in\mathcal{F}_0^+$. Hence $\mathbb{P}(B_{t_n}>0\text{ i.o})=\mathbb{P}(\limsup\{B_{t_n}>0\})\geq\limsup\mathbb{P}(B_{t_n}>0)=1/2$ by Fatou. Hence by Blumenthal's 0-1 law $\mathbb{P}(S_\varepsilon>0)=1$.

2. Note $\sup_{t\geq 0} B_t = \sup_{t\geq 0} B_{\lambda t} =^d \sqrt{\lambda}\sup_{t\geq 0} B_t$ for any $\lambda > 0$. So $S_\infty =^d \alpha S_\infty$ for any $\alpha > 0$. Hence $\mathbb{P}(S_\infty \geq x) = \mathbb{P}(S_\infty \geq 0) = 1$ for all x and so $\mathbb{P}(S_\infty = \infty) = 1$.

$\square$

Remark. (1) also follows immediately from the preceding proposition.

Proposition. Let B be a standard Brownian motion in $\mathbb{R}^d$ and let C be a cone with origin at 0 and non-empty interior, i.e $C = \{tu : t > 0, u \in A\}$ with $A \subseteq \mathbb{S}^1$ (unit sphere in $\mathbb{R}^d$). Let $H_C = \inf\{t > 0 : B_t \in C\}$. Then $\mathbb{P}(H_C = 0) = 1$.

Proof. We have $\{H_C = 0\} \in \mathcal{F}_0^+$. Also $\mathbb{P}(B_t \in C) = \mathbb{P}(B_1 \in C)$ by scale invariance of the Brownian motion and C. Also $\mathbb{P}(B_1 \in C) > 0$ since $\text{int}(C) \neq \emptyset$. Hence $\mathbb{P}(H_C \leq t) \geq \mathbb{P}(B_t \in C) > 0$. $\square$