Math 6720. Probability Theory II

Taught by Lionel Levine

Notes by Linus Setiabrata

This is a second course in probability theory. Please let me know if you spot any mistakes! There are probably lots of typos. Things in [blue font square brackets] are personal comments. Things in [red font square brackets] are (important) announcements.

Notes for Math 6710. Probability Theory I, from last semester, can be found here: 6710. We are citing the version from Dec 9, 2019.

Theorem numbering unchanged since May 12, 2020. Last compiled August 31, 2020.

Contents

1 Conditional Expectation

1.1 Jan 21, 2020

[Our professor is Lionel Levine. His office MLT 438. Our TA is Hannah Cairns; we'll meet her next week. For the time being, Prof. Levine's office hours are on Mondays, 2:00–3:00; Hannah's are on Wednesdays, 3:00–5:00. Office hours start next week. Send Prof. Levine an email if you want to meet him this week.]

We will mostly follow Durrett's *Probability: theory and examples* (5th edition), as well as Williams's *Probability with martingales*. Topics, roughly, are:

- Martingales, discrete time
- Brownian motion, continuous time. These are some kind of universal object that contains a lot of things in it, for example:
- Martingales in continuous time (and how they're embedded in Brownian motion)
- Ergodic theory, stationary sequences

There will be problem sets, roughly one a week (due Thursdays, starting Jan 30), as well as presentations towards the end of the semester. These will consist of 3–5 students presenting 1 topic.

Conditional expectation. (See Durrett 4.1, or Williams Ch. 9)

Let's consider a probability space $(\Omega, \mathcal{F}_0, \mathbb{P})$ and a random variable $X\colon (\Omega, \mathcal{F}_0) \to (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel sets. Suppose that

$$\mathbb{E}|X| \stackrel{\text{def}}{=} \int_\Omega |X|\, d\mathbb{P} < \infty.$$

Definition 1.1.1 (Conditional expectation)**.** Given a σ-field $\mathcal{F} \subseteq \mathcal{F}_0$, (a version of) the *conditional expectation* $\mathbb{E}(X|\mathcal{F})$ is any random variable Y satisfying:

(i) $Y \in m\mathcal{F}$ ("Y is $\mathcal{F}$-measurable", i.e. $Y^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$)

(ii) For any $A \in \mathcal{F}$,

$$\int_A X\, d\mathbb{P} = \int_A Y\, d\mathbb{P}.$$

△

Remark 1.1.2. Note that Y is integrable, and in particular that $\mathbb{E}|Y| \le \mathbb{E}|X|$. △

Proof. Let $A = \{Y > 0\} = \{\omega \in \Omega\colon Y(\omega) > 0\}$. Note that $A = Y^{-1}(0,\infty) \in \mathcal{F}$. Also, $|Y| = Y\mathbb{1}_A + (-Y)\mathbb{1}_{A^c}$. Taking expectations, $\mathbb{E}|Y| = \mathbb{E}(Y\mathbb{1}_A) + \mathbb{E}((-Y)\mathbb{1}_{A^c})$. Now property (ii) of Definition 1.1.1 says

$$\mathbb{E}(Y\mathbb{1}_A) = \int_A Y\, d\mathbb{P} = \int_A X\, d\mathbb{P} = \mathbb{E}(X\mathbb{1}_A).$$

Similarly, $\mathbb{E}((-Y)\mathbb{1}_{A^c}) \le \mathbb{E}(|X|\mathbb{1}_{A^c})$. Adding these together, we obtain $\mathbb{E}|Y| \le \mathbb{E}|X|$. □

Remark 1.1.3. The random variable Y, if it exists, is unique up to measure zero, that is, if Y and Y' both satisfy (i) and (ii), then $Y = Y'$ a.s.. (In other words, $\mathbb{P}(Y = Y') = 1$.) △

Proof. Fix $\varepsilon > 0$. Let $A_\varepsilon = \{Y - Y' \ge \varepsilon\}$. Then

$$\int_{A_\varepsilon} (Y - Y')\, d\mathbb{P} \ge \varepsilon \mathbb{P}(A_\varepsilon).$$

On the other hand,

$$\int_{A_\varepsilon} (Y - Y')\, d\mathbb{P} = \int_{A_\varepsilon} Y\, d\mathbb{P} - \int_{A_\varepsilon} Y'\, d\mathbb{P} = \int_{A_\varepsilon} X\, d\mathbb{P} - \int_{A_\varepsilon} X\, d\mathbb{P} = 0.$$

It follows that $\mathbb{P}(A_\varepsilon) = 0$ for all $\varepsilon > 0$. To finish, we observe that

$$\{Y > Y'\} = \bigcup_{\varepsilon>0} \{Y > Y' + \varepsilon\} = \bigcup_{\varepsilon>0} A_\varepsilon = \bigcup_{n\geq 1} A_{2^{-n}}$$

is a countable union of measure zero sets, and $\mathbb{P}(Y > Y') = 0$. Likewise, $\mathbb{P}(Y' > Y) = 0$ as well. □

We'll use some nontrivial measure theory to prove that Y exists.

Definition 1.1.4 (Absolute continuity)**.** Let μ, ν be σ-finite measures on $(\Omega, \mathcal{F})$. We say ν is *absolutely continuous* with respect to μ if

$$\mu(A) = 0 \implies \nu(A) = 0$$

for all $A \in \mathcal{F}$. We denote this by $\nu \ll \mu$. △

(Notice that this notion is asymmetric.)

Example 1.1.5. Consider:

1. Let $X \sim N(0, 1)$ and let $\nu = \nu_X$ be the distribution of X. Denote by

   $$\nu(A) = \mathbb{P}(X \in A) = \int_A e^{-x^2/2}\, d\lambda(x),$$

   where λ is the Lebesgue measure on $\mathbb{R}$. Then $\nu \ll \lambda$, since if $\lambda(A) = 0$ then $\nu(A)$ is an integral over a measure zero set. More generally, any random variable with a density gives rise to a distribution that is absolutely continuous with respect to λ.

2. Let $Y \sim \mathrm{Be}(\frac{1}{2})$, so $\mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = \frac{1}{2}$. Then $\nu_Y \not\ll \lambda$, since $\nu_Y(\{0\}) = \frac{1}{2}$ whereas $\lambda(\{0\}) = 0$. More generally, any random variable with an atom gives rise to a distribution that is *not* absolutely continuous with respect to λ.

3. There are random variables with no atoms whose distribution is *not* absolutely continuous with respect to λ [cf. [HW 1, Ex 4]]. Indeed, let

   $$W = \sum_{n\geq 1} \frac{\beta_n}{3^n},$$

   where the $\beta_n \sim 2\mathrm{Be}(\frac{1}{2})$ are independent. Indeed, W is supported on the Cantor set. (Its distribution function is the Cantor-Lebesgue function.)

△

Here's the measure theory that will help us prove the existence of conditional expectations:

Theorem 1.1.6 (Radon–Nikodym)**.** *If $\nu \ll \mu$, then there exists $f \in m\mathcal{F}$ with $f \geq 0$ such that*

$$\nu(A) = \int_A f\, d\mu$$

for all $A \in \mathcal{F}$. Sometimes this is denoted $f = \frac{d\nu}{d\mu}$, and f is called the Radon-Nikodym derivative.

(Note that f is unique up to measure zero.)

We won't prove this right now. (Williams manages to avoid using Theorem 1.1.6, and then uses martingales to prove Theorem 1.1.6. We won't follow this route, though.)

Theorem 1.1.7. *The conditional expectation $Y = \mathbb{E}(X|\mathcal{F})$ exists.*

Proof. Let us assume first that $X \geq 0$. Let $\mu = \mathbb{P}$ as a measure on $(\Omega, \mathcal{F})$ (not $\mathcal{F}_0$ (!)), and for every $A \in \mathcal{F}$, let

$$\nu(A) = \int_A X \, d\mathbb{P}.$$

Both ν, μ are measures on $(\Omega, \mathcal{F})$ and $\nu \ll \mu$. (We used nonnegativity of X to conclude that ν is an honest measure.) The Radon–Nikodym Theorem (Theorem 1.1.6) asserts the existence of $Y \in m\mathcal{F}$ such that $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$ for all $A \in \mathcal{F}$. This Y is (a version of) the conditional expectation $\mathbb{E}(X|\mathcal{F})$.

For general X, write $X = X^+ - X^-$ for nonnegative X^+ and X^-. We have conditional expectations $Y_1 = \mathbb{E}(X^+|\mathcal{F})$ and $Y_2 = \mathbb{E}(X^-|\mathcal{F})$ and $Y = Y_1 - Y_2$ serves as a $\mathbb{E}(X|\mathcal{F})$. □

Note in general that $Y \neq X$, because X might not be $\mathcal{F}$-measurable. (Only when X is $\mathcal{F}$-measurable is $Y = X$ possible.)

Remark 1.1.8. Let's discuss the intuitive meaning of $\mathbb{E}(X|\mathcal{F})$. This is a random variable, not just a real number. Then, $\mathbb{E}(X|\mathcal{F})(\omega)$ is the best guess of the value of $X(\omega)$, given all information in $\mathcal{F}$. Last semester we briefly discussed σ-fields $\mathcal{F}$ as information, namely, as answers to yes–no questions of the form "is $\omega \in A$?", where $A \in \mathcal{F}$ (see [6710, Definition 2.4.6] or [6710 HW 2, Ex 2(iv)]).

Let's consider extreme cases. We said before that if $X \in m\mathcal{F}$ then $\mathbb{E}(X|\mathcal{F}) = X$ a.s..

On the other hand, suppose X and $\mathcal{F}$ are independent. This means that $\sigma(X)$ and $\mathcal{F}$ are independent, so $\mathbb{P}(\{X \in B\} \cap A) = \mathbb{P}(X \in B)\mathbb{P}(A)$ for all $A \in \mathcal{F}$ and $B \in \mathcal{B}$; equivalently, that X and Y are independent random variables for all $Y \in m\mathcal{F}$.

If X and $\mathcal{F}$ are independent, $\mathbb{E}(X|\mathcal{F})$ is the constant random variable $\mathbb{E}X$ (a.s., of course). To see this, observe that for any $A \in \mathcal{F}$,

$$\int_A X \, d\mathbb{P} = \mathbb{E}(X \mathbb{1}_A) = (\mathbb{E}X)(\mathbb{E}\mathbb{1}_A) = (\mathbb{E}X)\mathbb{P}(A) = \int_A (\mathbb{E}X) \, d\mathbb{P}. \qquad \triangle$$

1.2 Jan 23, 2020

Today we'll connect conditional expectation to the elementary definition we learnt a long time ago. Specifically, we'll discuss the random variable $\mathbb{E}(X|Z)$, where X and Z take finitely many values, denoted $X_1, \ldots, X_m$ and $Z_1, \ldots, Z_n$ respectively.

Traditionally, for an event B with $\mathbb{P}(B) > 0$ we have the quantity

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

Specifically, in our setting

$$\mathbb{P}(\{X = x_i\}|\{Z = z_j\}) = \frac{\mathbb{P}(\{X = x_i \text{ and } Z = Z_j\})}{\mathbb{P}(Z = z_j)}.$$

Thus we may define

$$\mathbb{E}(X|\{Z = z_j\}) = \sum_i x_i \mathbb{P}(\{X = x_i\}|\{Z = z_j\})$$

Let us denote the sum above by Y_j. We now define the random variable

$$\mathbb{E}(X|Z)(\omega) \stackrel{\text{def}}{=} Y_j \text{ whenever } \omega \in \{Z = z_j\}$$

Claim 1.2.1. *The equality of random variables*

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

holds.

(The left side is as defined earlier, whereas the right side is in the sense of Definition 1.1.1. Recall also that $\sigma(Z) = \{\{Z \in B\} \colon B \subseteq \{Z_1, \ldots, Z_n\}\} = \{\text{disjoint unions of } \{Z = z_j\}\}$.)

Proof of Claim 1.2.1. Note that $Y = \mathbb{E}(X|Z)$ is constant on each event $\{Z = z_j\}$, hence

$$Y^{-1}(Y_j) = \{Z = z_j\} \implies Y \in m\mathcal{F}.$$

This verifies property (i) of Definition 1.1.1. To verify property (ii), it suffices to check for $A = \{Z = z_j\}$ that

$$\int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P}.$$

Since we observed Y is constant on $A = \{Z = z_j\}$ we have

$$\begin{aligned}
\int_A Y \, d\mathbb{P} &= Y_j \mathbb{P}(Z = z_j) \\
&= \sum_i x_i \mathbb{P}(\{X = x_i\}|\{Z = z_j\}) \mathbb{P}(Z = z_j) \\
&= \sum_i x_i \mathbb{P}(\{X = x_i \text{ and } Z = z_j\}) \\
&= \sum_i x_i \mathbb{P}(X \mathbb{1}_A = x_i) \\
&= \mathbb{E}(X \mathbb{1}_A) = \int_A X \, d\mathbb{P}.
\end{aligned}$$

□

Remark 1.2.2. Recall (see e.g. [6710 HW 2, Ex 4]) that if $Y \in m(\sigma(Z))$ then $Y = f(Z)$ for some measurable function $f \colon \mathbb{R} \to \mathbb{R}$. In particular, $\mathbb{E}(X|Z)$ is a measurable function of Z. It's the best guess for $X(\omega)$, given the value of $Z(\omega)$. △

Proposition 1.2.3. *Some properties of* $\mathbb{E}(X|\mathcal{F})$*:*

1. $\mathbb{E}(aX + bY|\mathcal{F}) = a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F})$ *a.s..*
2. *If* $X \le Y$ *then* $\mathbb{E}(X|\mathcal{F}) \le \mathbb{E}(Y|\mathcal{F})$ *a.s..*
3. *If* $X_n \ge 0$ *and* $X_n \uparrow X$ *with* $\mathbb{E}X < \infty$ *then* $\mathbb{E}(X_n|\mathcal{F}) \uparrow \mathbb{E}(X|\mathcal{F})$*.*
4. *Jensen: If* $\varphi\colon \mathbb{R} \to \mathbb{R}$ *is convex,* $\mathbb{E}|X| < \infty$*, and* $\mathbb{E}|\varphi(X)| < \infty$*, then*
$$\varphi(\mathbb{E}(X|\mathcal{F})) \le \mathbb{E}(\varphi(X)|\mathcal{F}) \quad \text{a.s..}$$
5. *[HW 1, Ex 1] If* $\mathcal{F}_1 \subseteq \mathcal{F}_2$*, then*
$$\mathbb{E}\big[\mathbb{E}(X|\mathcal{F}_1)\big|\mathcal{F}_2\big] = \mathbb{E}(X|\mathcal{F}_1) = \mathbb{E}\big[\mathbb{E}(X|\mathcal{F}_2)\big|\mathcal{F}_1\big] \quad \text{a.s.,}$$
so "the smaller sigma field wins".
6. *If* $X \in m\mathcal{F}$*,* $\mathbb{E}|Y| < \infty$*,* $\mathbb{E}|XY| < \infty$*, then*
$$\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F}) \quad \text{a.s..} \tag{1}$$

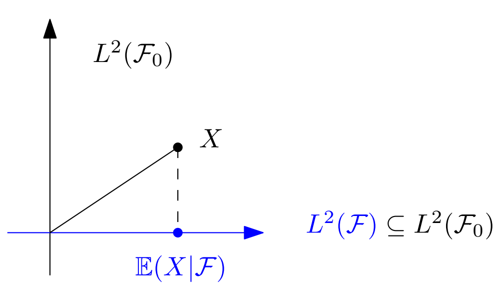
(Part 6 above is often a key step in many proofs. These properties can be found in the back cover of Williams (!).)

Proof of part 6 of Proposition 1.2.3. It's easy to check condition (i), namely, that $X\mathbb{E}(Y|\mathcal{F}) \in m\mathcal{F}$.
To check condition (ii), we apply the usual four-step machine:
If $X = \mathbb{1}_B$ for $B \in \mathcal{F}$, then for all $A \in \mathcal{F}$ we have
$$\int_A \mathbb{1}_B \mathbb{E}(Y|\mathcal{F})\, d\mathbb{P} = \int_{A\cap B} \mathbb{E}(Y|\mathcal{F})\, d\mathbb{P} = \int_{A\cap B} Y\, d\mathbb{P} = \int_A \mathbb{1}_B Y\, d\mathbb{P}.$$
Now if $X = \sum_{i=1}^k \mathbb{1}_{B_i}$ is simple, use linearity of Equation (1).
Now if $X, Y \ge 0$ and we have a sequence of simple functions $X_n \uparrow X$, then by part 3 of Proposition 1.2.3 we obtain
$$\int_A X\mathbb{E}(Y|\mathcal{F}) \underset{\text{MCT}}{=} \lim_{n\to\infty} \int_A X_n \mathbb{E}(Y|\mathcal{F})\, d\mathbb{P} = \lim \int_A X_n\, d\mathbb{P} \underset{\text{MCT}}{=} \int_A XY\, d\mathbb{P},$$
where we applied monotone convergence theorem twice.
Finally, we may write $X = X^+ - X^-$ and $Y = Y^+ - Y^-$ as a difference of nonnegative functions and apply linearity. □

As promised, let's describe $\mathbb{E}(X|\mathcal{F})$ as orthogonal projection in $L^2(\Omega, \mathcal{F}_0, \mathbb{P}) = \{Y \in m\mathcal{F}_0 \colon \mathbb{E}Y^2 < \infty\}$.
If $F \subseteq \mathcal{F}_0$, then $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a closed subspace of the Hilbert space $L^2(\Omega, \mathcal{F}_0, \mathbb{P})$.

Claim 1.2.4. *If* $X \in L^2(\mathcal{F}_0)$*, then* $\mathbb{E}(X|\mathcal{F})$ *is the point in* $L^2(\mathcal{F})$ *closest to* X*.*

For a [rare!] picture:



Proof. If $Z \in L^2(\mathcal{F})$ then

$$Z\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(ZX|\mathcal{F}).$$

(Note of course that $\mathbb{E}|ZX| < \mathbb{E}Z^2 + \mathbb{E}X^2 < \infty$, so we may apply item 6 in Proposition 1.2.3.) Thus

$$\mathbb{E}\big[Z\mathbb{E}(X|\mathcal{F})\big] = \mathbb{E}\big[\mathbb{E}(XZ)\big|\mathcal{F}\big] = \mathbb{E}(XZ)$$

and hence

$$\mathbb{E}\big(Z(X - \mathbb{E}(X|\mathcal{F}))\big) = 0.$$

Since L^2 is endowed with the inner product $\langle X, Y \rangle = \mathbb{E}(XY)$, we have shown that Z is orthogonal to $X - \mathbb{E}(X|\mathcal{F})$. In particular, for $Y \in L^2(\mathcal{F})$ we have

$$\mathbb{E}(X - Y)^2 = \mathbb{E}(X - \mathbb{E}(X|\mathcal{F}))^2 + \mathbb{E}Z^2 \tag{2}$$

is minimized when $Z = 0$. Thus $Y = \mathbb{E}(X|\mathcal{F})$. $\square$

2 Martingales

2.2 Jan 23, 2020

Definition 2.2.1. A *filtration* is a sequence of σ-fields $(\mathcal{F}_n)_{n\geq 0}$ with $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$. △

Definition 2.2.2. A *martingale* (sometimes abbreviated *MG*) relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$ is a sequence of random variables $(X_n)_{n\geq 0}$ satisfying for all $n \geq 0$ the conditions:

1. $\mathbb{E}|X_n| < \infty$
2. $X_n \in m\mathcal{F}_n$
3. $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ a.s.. △

Although we require equality in part 3 of the definition, replacing it with $\leq$ gives the notion of a *supermartingale*, whereas replacing it with $\geq$ gives the notion of a *submartingale*.

Example 2.2.3. We may consider a simple random walk $X_n = \xi_1 + \cdots + \xi_n$ for independent random variables ξ_i with $\xi_i = \pm 1$ with probability $\frac{1}{2}$.

Then $(X_n)_{n\geq 1}$ is a martingale relative to the filtration

$$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n) = \sigma(X_1, \dots, X_n).$$

This is sometimes called the *natural filtration*, since it's the smallest filtration for which $(X_n)_{n\geq 1}$ is a martingale. (It is sometimes useful to consider filtrations other than the natural one.) [cf. [HW 1, Ex 6].] △

The first two conditions in Definition 2.2.2 are easy to check. To see the third one, observe that

$$\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) = \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = \mathbb{E}\xi_{n+1} = 0$$

where the second last equality follows from independence of ξ_{n+1}.

Example 2.2.4 (cf. [6710 HW6, Ex 2])**.** Let $(\xi_n)_{n\geq 1}$ be independent with $\mathbb{E}\xi_n = 1$ for all n. Let $X_n = \xi_1 \dots \xi_n$. As before $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n) = \sigma(X_1, \dots, X_n)$ wil be the natural filtration. Then

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

Since X_n is $\mathcal{F}_n$-measurable, we obtain

$$\mathbb{E}(X_n\xi_{n+1}|\mathcal{F}_n) = X_n\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = X_n\mathbb{E}\xi_{n+1} = X_n.$$ △

Example 2.2.5. Fix $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_n)$. Let $X_n = \mathbb{E}(X|\mathcal{F}_n)$. (There is an interpretation of this setup in financial terms.)

We may check that

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}\big(\mathbb{E}(X|\mathcal{F}_{n+1})\big|\mathcal{F}_n\big) = \mathbb{E}(X|\mathcal{F}_n) = X_n.$$

△

2.3 Jan 28, 2020

[There will be office hours today from 3–4pm at 438 MLT!]

Let's begin with another example of a martingale, which hopefully motivates the definition of *sub*- and *super*-martingale.

Example 2.3.1 (Random walks in higher dimensions)**.** Suppose $\mathbf{X}_n = \xi_1 + \cdots + \xi_n$, where $\xi \sim \text{Unif}(B(0,1))$ are i.i.d.. Here $B(0,1)$ is the unit ball in $\mathbb{R}^d$. Suppose also that $f : \mathbb{R}^d \to \mathbb{R}$ is *superharmonic*, which means its Laplacian is nonpositive, i.e.

$$\Delta f \stackrel{\text{def}}{=} \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2} \le 0. \tag{3}$$

One can show that $(f(\mathbf{X}_n))_{n\ge1}$ is a supermartingale. (This is Durrett Exercise 5.2.2 in the 4th edition.) [See [Durrett Exercise 5.4.3].]

We call f satisfying (3) *superharmonic* because if h is any harmonic function defined on the same domain D as f, and $h = f$ on ∂D, then $f \ge h$ on D. (This follows from the minimum principle, which says superharmonic functions attain their minimum on the boundary.) △

Lemma 2.3.2. *If $(X_n)_{n\ge1}$ is a supermartingale with respect to $(\mathcal{F}_n)_{n\ge1}$, then for all $n > m$ we have $\mathbb{E}(X_n|\mathcal{F}_m) \le X_m$ a.s..*

Proof. Write $n = m + k$ for $k \ge 1$. We have almost sure (in)equalities

$$\mathbb{E}(X_{m+k}|\mathcal{F}_m) = \mathbb{E}\big(\mathbb{E}(X_{m+k}|\mathcal{F}_{m+k-1})\big|\mathcal{F}_m\big) \le \mathbb{E}(X_{m+k-1}|\mathcal{F}_m)$$

since conditional expectation is monotone (Proposition 1.2.3, part 2). Then induction says $\mathbb{E}(X_{m+k-1}|\mathcal{F}) \le X_m$ a.s.. □

Remark 2.3.3. Note that $(X_n)_{n\ge1}$ is a supermartingale if and only if $(-X_n)_{n\ge1}$ is a submartingale. Furthermore, $(X_n)_{n\ge1}$ is a martingale if and only if $(X_n)_{n\ge1}$ is both a supermartingale and submartingale. △

Doob transform.

Definition 2.3.4. The random variables $(H_n)_{n\ge1}$ are said to be *predictable* (with respect to a filtration $(\mathcal{F}_n)_{n\ge0}$) if $H_n \in m\mathcal{F}_{n-1}$. △

Think of H_n as a betting strategy at time n, and think of X_n as net winnings at time n if you always bet $\$1$. On round n, you win $X_n - X_{n-1}$ if you bet $\$1$, so in particular you win $H_n(X_n - X_{n-1})$ if you bet $\$H_n$. Hence, our net winnings at time n, using gambling system H, is:

Definition 2.3.5. The *Doob transform* of H and X is

$$(H \cdot X)_n \stackrel{\text{def}}{=} \sum_{m=1}^{n} H_m(X_m - X_{m-1}).$$

△

Theorem 2.3.6 (You can't beat an unfavorable game)**.** *If $(X_n)_{n\ge0}$ is a supermartingale and $(H_n)_{n\ge0}$ is predictable with $0 \le H_n \le C$, i.e. nonnegative and bounded, then $(H \cdot X)_{n\ge0}$ is also a supermartingale.*

Proof. We have

$$\mathbb{E}\big((H\cdot X)_{n+1} - (H\cdot X)_n\big|\mathcal{F}_n\big) = \mathbb{E}\big(\underbrace{H_{n+1}}_{\in m\mathcal{F}_n}(X_{n+1} - X_n)\big|\mathcal{F}_n\big) = H_{n+1}\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) \le 0,$$

with the last inequality following from $H_{n+1} \ge 0$ and $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \le \mathbb{E}(X_n|\mathcal{F}_n)$ by supermartingaleness. □

Stopping times.

Recall that a random variable T taking values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ is a *stopping time* with respect to $\mathcal{F} = (\mathcal{F}_n)_{n\ge0}$ if $\{T = n\} \in \mathcal{F}_n$ for all $0 \le n < \infty$ (or if equivalently $\{T \le n\} \in \mathcal{F}_n$ for all $0 \le n < \infty$).

Example 2.3.7. If T is a stopping time, the sequence $H_n = \mathbb{1}_{\{T \geq n\}} = \mathbb{1}_{\{T \leq n-1\}^c} \in m\mathcal{F}_{n-1}$ is predictable. △

Thus we obtain

Corollary 2.3.8. *If $(X_n)_{n \geq 0}$ is a supermartingale and T is a stopping time, then $(X_{T \wedge n})_{n \geq 0}$ is also a supermartingale.*

(As usual, $T \wedge n$ means $\min(T, n)$.) In particular, Lemma 2.3.2 says

$$\mathbb{E}(X_{T \wedge n}) \leq \mathbb{E}X_{T \wedge 0} = \mathbb{E}X_0.$$

We'd like to show that $\mathbb{E}(X_T) \leq \mathbb{E}X_0$.

Suppose that $T < \infty$ a.s, so $X_{T \wedge n} \to X_T$ a.s. as $n \to \infty$. Does it follow that $\mathbb{E}(X_{T \wedge n}) \to \mathbb{E}(X_T)$? Not in general:

Example 2.3.9 (A common enemy to martingale proofs)**.** Consider a simple random walk $X_n = \sum_{i=1}^n \xi_i$ for i.i.d. ξ_i with $\mathbb{P}(\xi_i = \pm 1) = \frac{1}{2}$. Take $T = \inf\{n \colon X_n = 1\}$.
We know $T < \infty$ a.s. because simple random walks are recurrent, but also that $\mathbb{E}T = \infty$. △

Martingale convergence theorem.
Let $(X_n)_{n \geq 0}$ be a supermartingale, and let

$$\begin{aligned} U_N[a, b] &= \#\{\text{upcrossings of } [a, b] \text{ by time } N\} \\ &= \max\{k \colon \text{there are } 0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq N \text{ with } X_{s_i} < a, X_{t_i} > b \text{ for } i \in [k]\}. \end{aligned}$$

(Since supermartingales generally decrease, there shouldn't be so many of these.)
Let $Y = H \cdot X$ where $H_1 = \mathbb{1}_{\{X_0 < a\}}$ and $H_n = \mathbb{1}_{\{H_{n-1}=1, X_{n-1} \leq b\}} + \mathbb{1}_{\{H_{n-1}=0, X_{n-1} \leq a\}}$. (The first indicator says one should keep betting if X_n is below b, while the second indicator says one should start betting if X_n is below a.)
By definition,

$$Y_N \geq (b - a) U_N[a, b] - (X_N - a)^-,$$

where $(X_n - a)^- = (X_n - a)\mathbb{1}_{\{X_n < a\}}$. If $(X_n)_{n \geq 0}$ is a supermartingale, then so is $(Y_n)_{n \geq 0}$; hence $\mathbb{E}Y_N \leq 0$. Thus we arrive at the *upcrossing inequality*

$$(b - a)\mathbb{E}U_N[a, b] \leq \mathbb{E}(X_n - a)^-. \tag{4}$$

Corollary 2.3.10. *If $(X_n)_{n \geq 0}$ is a supermartingale bounded in L^1 (which means $\sup_n \mathbb{E}|X_n| < \infty$), then*

$$\mathbb{P}(U_\infty[a, b] = \infty) = 0.$$

(Here, $U_\infty[a, b] \stackrel{\text{def}}{=} \lim_{N \to \infty} U_N[a, b] \in \{0, 1, 2, \dots\} \cup \{+\infty\}$.)

Proof of Corollary 2.3.10. Apply monotone convergence theorem to obtain

$$(b - a)\mathbb{E}U_\infty[a, b] \leq |a| + \sup_n \mathbb{E}|X_n| < \infty.$$ □

We can now state the martingale convergence theorem:

Theorem 2.3.11 (Doob's Martingale Convergence Theorem)**.** *Let $(X_n)_{n \geq 0}$ be a supermartingale bounded in L^1. Then*

$$\mathbb{P}(\lim_{n \to \infty} X_n \text{ exists in } \mathbb{R}) = 1.$$

2.4 Jan 30, 2020

We had stated the martingale convergence theorem yesterday:

Theorem 2.4.1 (Doob's Martingale Convergence Theorem)**.** *Let $(X_n)_{n\geq 0}$ be a supermartingale bounded in L^1. Then*

$$\mathbb{P}(\lim_{n\to\infty} X_n \text{ exists in } \mathbb{R}) = 1.$$

Proof. Let's define the event

$$\begin{aligned}\Lambda &\overset{\text{def}}{=} \{\omega\colon X_n(\omega) \text{ does not converge in } [-\infty,\infty]\}\\ &= \{\omega\colon \liminf X_n(\omega) < \limsup X_n(\omega)\}\\ &= \bigcup_{\substack{a<b\\ a,b\in\mathbb{Q}}} \underbrace{\{\liminf X_n(\omega) < a \text{ and } \limsup X_n(\omega) > b\}}_{\text{call this } \Lambda_{a,b}}.\end{aligned}$$

Recall that the random variables

$$U_n[a,b] = \#\{\text{upcrossings of } [a,b] \text{ by time } n\}$$

are monotone increasing in n, hence converges to a limit $U_n[a,b] \uparrow U_\infty[a,b]$. In Corollary 2.3.10 we showed $\mathbb{P}(U_\infty[a,b] = \infty) = 0$.

Since $\Lambda_{a,b} \subseteq \{U_\infty[a,b] = \infty\}$, it follows that $\mathbb{P}(\Lambda_{a,b}) = 0$ for all $a < b$, and hence $\mathbb{P}(\Lambda) = 0$.

It is left to rule out the cases that the limit is $\pm\infty$. Note that we may define $X_\infty \overset{\text{def}}{=} \lim X_n$, which exists in $[-\infty,\infty]$. Fatou's lemma says

$$\mathbb{E}|X_\infty| \leq \liminf \mathbb{E}|X_n| < \sup \mathbb{E}|X_n| < \infty.$$

In particular, $\mathbb{P}(X_\infty \in \mathbb{R}) = 1$. ☐

We can squeeze out a little more:

Corollary 2.4.2. *If $(X_n)_{n\geq 0}$ is a supermartingale with $X_n \geq 0$ for all n, then $X_n \to X_\infty$ a.s. and $\mathbb{E}X_\infty \leq \mathbb{E}X_0$.*

Proof. Since $(X_n)_{n\geq 0}$ is a supermartingale, we have $\mathbb{E}X_n \leq \mathbb{E}X_0$ for all n (e.g. by Lemma 2.3.2). Thus $\sup \mathbb{E}|X_n| < \infty$. Then the martingale convergence theorem (Theorem 2.4.1) says $X_n \to X_\infty$ a.s.. Furthermore, Fatou says $\mathbb{E}X_\infty \leq \liminf \mathbb{E}X_n \leq \mathbb{E}X_0$. ☐

Example 2.4.3 (Enemy, cf. Example 2.3.9)**.** Let Y_n be a simple random walk and let $T = \inf\{n\colon Y_n = 1\}$. Let $X_n = Y_{n\wedge T}$. Both $(Y_n)_{n\geq 0}$ and $(X_n)_{n\geq 0}$ are martingales. However, they're not bounded in L^1.

It turns out that a limit $X_n \to X_\infty$ exists a.s. and in particular $X_\infty = 1$ a.s.. Corollary 2.4.2 fails to hold for $(X_n)_{n\geq 0}$, since $1 = \mathbb{E}X_\infty > \mathbb{E}X_0 = 0$. △

Lemma 2.4.4. *Let $X = (X_n)_{n\geq 0}$ be a martingale with $|X_{n+1} - X_n| \leq M < \infty$ for all n. Let*

$$\begin{aligned}C &\overset{\text{def}}{=} \{\lim X_n \text{ exists in } \mathbb{R}\}\\ D &\overset{\text{def}}{=} \{\limsup X_n = +\infty \text{ and } \liminf X_n = -\infty\}.\end{aligned}$$

Then $\mathbb{P}(C \cup D) = 1$.

(The condition $|X_{n+1} - X_n| \leq M < \infty$ is sometimes called *bounded increments*. This does not imply X is bounded in L^1; a counterexample is the simple random walk. The simple random walk does not satisfy martingale convergence, i.e. $\mathbb{P}(C) \neq 1$, but it is recurrent and hence $\mathbb{P}(D) = 1$.)

Proof. The proof is in [Durrett, Lem 4.3.1]. It's a dense, short proof, and it's a good exercise to decode it!

The idea is if $\liminf X_n > -\infty$ then we may shift to get a nonnegative martingale, which converges a.s.. ☐

Theorem 2.4.5 (Improved Borel-Cantelli Lemma 2)**.** *Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration, and let $A_1, A_2, \ldots$ be events with $A_n \in \mathcal{F}_n$. Then*

$$\{A_n \text{ i.o.}\} = \left\{ \sum_{n\geq 1} \mathbb{P}(A_n|\mathcal{F}_{n-1}) = \infty \right\} \quad \text{a.s..}$$

Remark 2.4.6 (Borel-Cantelli 2)**.** Let's see how Theorem 2.4.5 implies the usual Borel-Cantelli 2 (cf. [6710, Lemma 5.13.2]). If A_n are independent and $\sum \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n|\mathcal{F}_{n-1}) = \mathbb{P}(A_n)$. Hence $\mathbb{P}(A_n \text{ i.o.}) = 1$. △

Proof of Theorem 2.4.5. Let

$$X_n = \sum_{m=1}^{n} \Big(\mathbb{1}_{A_m} - \mathbb{P}(A_m|\mathcal{F}_{m-1}) \Big).$$

We claim that this is a martingale with respect to $(\mathcal{F}_n)_{n\geq 1}$ that has bounded increments. The latter part is easy to see; that X_n is a martingale is a special case of a general recipe to construct martingales (Remark **??**), namely the case $Y_m = \mathbb{1}_{A_m}$.

Since X_n is a martingale we may apply Lemma 2.4.4, so $\mathbb{P}(C \cup D) = 1$. If $\omega \in D$ then

$$\sum \mathbb{1}_{A_m}(\omega) = \infty \quad \text{and} \quad \sum \mathbb{P}(A_m|\mathcal{F}_{m-1})(\omega) = \infty.$$

Conversely if $\omega \in C$ then

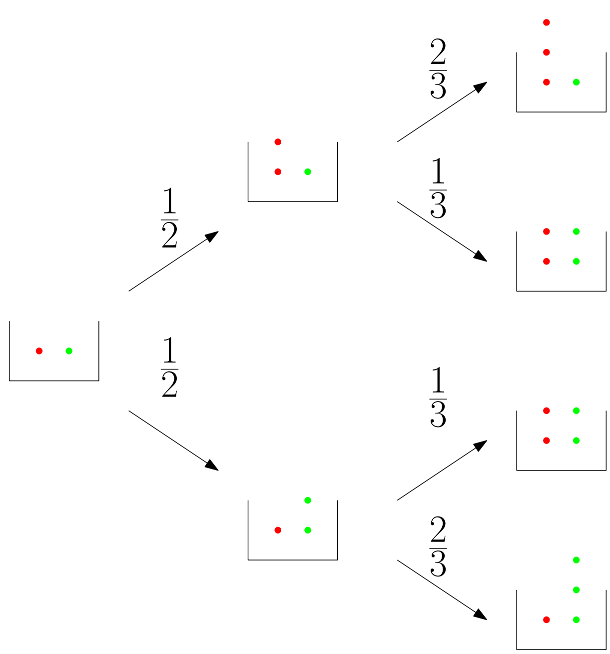
$$\sum \mathbb{1}_{A_m}(\omega) = \infty \iff \sum \mathbb{P}(A_m|\mathcal{F}_{m-1})(\omega) = \infty.$$ □

Remark 2.4.7 (A general recipe to construct martingales)**.** namely, given any sequence of random variables Y_n with $\mathbb{E}|Y_n| < \infty$, let

$$X_n = \sum_{m=1}^{n} \Big(Y_m - \mathbb{E}(Y_m|\mathcal{F}_{m-1}) \Big).$$

Then $X_{n+1} - X_n = Y_{n+1} - \mathbb{E}(Y_{n+1}|\mathcal{F}_n)$. Conditioning both sides on $\mathcal{F}_n$, the right hand side becomes zero. It's a common trick, given a stochastic process Y_n "in the wild", to decompose it into a "martingale part" X_n and a "compensating term", and analyze the two parts separately. △

Example 2.4.8 (Polya's urn)**.** Suppose we have a large urn with r red balls and g green balls. Start with $r = g = 1$. At each time n, pick a ball uniformly at random and put it back along with an extra ball of the same color:



This can be thought of as a model for reinforcement; once a red ball is picked, the whole model is skewed towards redness. Let

$$X_n \stackrel{\text{def}}{=} \frac{\#\{\text{red balls at time n}\}}{n+2}.$$

So if there are i green balls and j red balls at time n, then

$$X_{n+1} = \begin{cases} \frac{j+1}{i+j+1} & \text{with probability } \frac{j}{i+j} \\ \frac{j}{i+j+1} & \text{with probability } \frac{i}{j+1} \end{cases}$$

We can compute

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \frac{j+1}{i+j+1}\frac{j}{i+j} + \frac{j}{i+j+1}\frac{i}{i+j} = \frac{j(i+j+1)}{(i+j)(i+j+1)} = X_n.$$

[The above is an equality on the set $\{X_n = \frac{j}{i+j}\}$, but it holds for every such set and hence holds on Ω.] Thus X_n is a nonnegative martingale and $X_n \to X_\infty$ a.s. for some limit, by Corollary 2.4.2. What is the distribution of X_∞?

We're going to prove that $X_\infty \sim \text{Unif}(0,1)$. Indeed, let $G_n = \#\{\text{green balls at time } n\}$. We may compute directly that

$$\mathbb{P}\Big(G_i = i+1 \text{ for } i = 1,\dots,m \text{ and then } G_{m+1} = \cdots = G_n = m+1\Big) = \frac{m!(n-m)!}{(n+1)!}.$$

But this probability is independent of the red-green drawing order (e.g. the probability of drawing it green-green-red-red is the same as red-green-red-green). It follows that

$$\mathbb{P}(G_n = m+1) = \binom{n}{m}\frac{m!(n-m)!}{(n+1)!} = \frac{1}{n+1}$$

for all $m = 0,\dots,n$. Then

$$\frac{G_n}{n+2} \xrightarrow{\text{d}} \text{Unif}(0,1)$$

and

$$X_n = 1 - \frac{G_n}{n+2} \xrightarrow{\text{d}} X_\infty \sim \text{Unif}(0,1). \qquad \triangle$$

Optional stopping/sampling.

Theorem 2.4.9 (Optional Stopping). *Let $(X_n)_{n\geq 0}$ be a submartingale. Let T be a stopping time which is bounded, so $0 \leq T \leq n$. Then $\mathbb{E}X_0 \leq \mathbb{E}X_T \leq \mathbb{E}X_n$.*

Proof. We've seen that $\mathbb{E}X_0 \leq \mathbb{E}X_T$, using the gambler strategy (Doob transform with $H_k = \mathbb{1}_{\{k\leq T\}} = \mathbb{1}_{\{T>k-1\}} \in m\mathcal{F}_{k-1}$).

On the other hand, when we Doob transform with $H_k = \mathbb{1}_{\{k>T\}} = \mathbb{1}_{\{k-1\geq T\}} \in m\mathcal{F}_{k-1}$, then

$$(H\cdot X)_m - (H\cdot X)_0 = \sum_{k=1}^{m} \mathbb{1}_{\{k>T\}}(X_k - X_{k-1}) = X_m - X_{T\wedge m}.$$

Since $H \cdot X$ is a submartingale (Theorem 2.3.6), we obtain

$$0 \leq \mathbb{E}((H\cdot X)_n - (H\cdot X)_0) = \mathbb{E}(X_n - X_{T\wedge n}) = \mathbb{E}X_n - \mathbb{E}X_T.$$

□

Theorem 2.4.10 (Doob's submartingale inequality). *Let X_n be a submartingale, and let $M_n = \max_{0\leq k\leq n} X_n^+$. Then*

$$\mathbb{P}(M_n \geq \lambda) \leq \frac{\mathbb{E}X_n^+}{\lambda}.$$

(cf. Markov, which says $\lambda\mathbb{P}(X_n \geq \lambda) \leq \mathbb{E}X_n\mathbb{1}_{\{X_n>\lambda\}} \leq \mathbb{E}X_n^+$. Theorem 2.4.10 is a strengthening because we are controlling the max of the X_n^+ as opposed to just the X_n. Roughly, submartingaleness says the maximum M_n is not much larger than the current X_n^+)

2.5 Feb 4, 2020

Using the π-λ theorem ([6710, Thm 1.2.11]) would've made [HW 1, Ex 2] easier. As a reminder, this states that if $\mathcal{A} \subseteq \mathcal{F}$ is a π-system contained in a σ-field, and $\mathcal{L}$ is a λ-system with $\mathcal{A} \subseteq \mathcal{L} \subseteq \mathcal{F}$, then $\sigma(\mathcal{A}) \subseteq \mathcal{L}$. (So for [HW 1, Ex 2], the claim is that $\mathcal{L} \stackrel{\text{def}}{=} \{A \in \mathcal{F} \colon \mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)\}$ really is a λ-system.)

We stated Doob's submartingale inequality last time:

Theorem 2.5.1 (Doob's submartingale inequality)**.** *Let X_n be a submartingale, and let $M_n = \max_{0 \le k \le n} X_n^+$. Then*

$$\mathbb{P}(M_n \ge \lambda) \le \frac{\mathbb{E}X_n^+}{\lambda}.$$

Proof. Consider the stopping time $T \stackrel{\text{def}}{=} \min\{k \colon X_k \ge \lambda \text{ or } k = n\}$. Let $A \stackrel{\text{def}}{=} \{M_n \ge \lambda\} = \{X_T \ge \lambda\}$. So Markov ([6710, Lem 3.8.4]) gives

$$\lambda \mathbb{P}(A) \le \mathbb{E}(X_T \mathbb{1}_A).$$

Furthermore, T is bounded, so optional stopping (Theorem 2.4.9) says $\mathbb{E}X_n \ge \mathbb{E}X_T \ge \mathbb{E}X_0$. Since

$$\mathbb{E}X_T = \mathbb{E}(X_T\mathbb{1}_A + X_T\mathbb{1}_{A^c}) = \mathbb{E}X_T\mathbb{1}_A + \mathbb{E}X_n\mathbb{1}_{A^c},$$

the fact that $\mathbb{E}X_T \le \mathbb{E}X_n$ implies

$$\lambda\mathbb{P}(A) \le \mathbb{E}X_T\mathbb{1}_A \le \mathbb{E}X_n - \mathbb{E}X_n\mathbb{1}_{A^c} = \mathbb{E}X_n\mathbb{1}_A \le \mathbb{E}X_n^+.$$

□

Corollary 2.5.2 (Kolmogorov maximal inequality, cf. [6710, Thm 7.25.5])**.** *Let $S_n = \xi_1 + \cdots + \xi_n$, where ξ_i are independent and $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = \sigma_i^2 < \infty$ for all i. Then*

$$\mathbb{P}(\max_{1\le k\le n} |S_k| \ge y) \le \frac{\operatorname{var}(S_n)}{y^2}.$$

Proof. The claim is that S_n^2 is a submartingale. Indeed,

$$\mathbb{E}(S_{n+1}^2|\mathcal{F}_n) \ge \mathbb{E}(S_{n+1}|\mathcal{F}_n)^2 = S_n^2 \quad \text{a.s.},$$

by conditional Jensen (Proposition 1.2.3, part 4). Since S_n^2 is a submartingale, we may apply Doob's submartingale inequality (Theorem 2.5.1), and for $\lambda = x^2$ we obtain

$$\mathbb{P}(\max_{1\le k\le n} |S_k| \ge x) = \mathbb{P}(\max_{1\le k\le n} S_k^2 \ge x^2) \le \frac{\mathbb{E}S_n^2}{x^2} = \frac{\operatorname{var}(S_n)}{x^2}.$$

□

In Corollary 2.3.8, we saw the following result: Let $(X_n)_{n\ge0}$ be a supermartingale, and let T be a stopping time (both with respect to a filtration $(\mathcal{F}_n)_{n\ge0}$). Then

$$\mathbb{E}X_{T\wedge n} \le \mathbb{E}X_0 \text{ for all } n. \tag{5}$$

We wondered when $\mathbb{E}X_T \le \mathbb{E}X_0$.

Theorem 2.5.3 (Doob's optional stopping theorem)**.** *We have:*

1. *If T is bounded, so $\mathbb{P}(T > k) = 0$ for some $k \in \mathbb{N}$, then $\mathbb{E}X_T \le \mathbb{E}X_0$.*
2. *If X is bounded, so $|X_n| \le k$ for all n, and $\mathbb{P}(T < \infty) = 1$, then $\mathbb{E}X_T \le \mathbb{E}X_0$.*
3. *If $\mathbb{E}T < \infty$ and X has bounded increments ($|X_{n+1} - X_n| \le k$ for all $n \in \mathbb{N}$), then $\mathbb{E}X_T \le \mathbb{E}X_0$.*

Proof. For part 1, apply Equation (5) for $n = k$. Then $T \wedge n = T$.

For part 2, observe that $X_{T\wedge n} \to X_T$ a.s., so $\mathbb{E}X_{T\wedge n} \to \mathbb{E}X_T$ by bounded convergence theorem. Then Equation (5) does the trick.

For part 3, observe that

$$X_{T\wedge n} - X_0 = \sum_{k=1}^{T\wedge n} (X_k - X_{k-1}),$$

so

$$|X_{T\wedge n} - X_0| \le \sum_{k=1}^{T\wedge n} |X_k - X_{k-1}| \le kT.$$

It follows that

$$|X_{T\wedge n}| \le \underbrace{|X_0| + kT}_{\in L^1},$$

so we may apply dominated convergence to obtain $\mathbb{E}X_{T\wedge n} \to \mathbb{E}X_T$. □

Let's remind ourselves of the counterexample, showing that $\mathbb{E}X_T \le \mathbb{E}X_0$ does not necessarily hold in general:

Example 2.5.4 (Enemy, cf. Example 2.3.9, 2.4.3)**.** Let Y_n be a simple random walk and let $T = \inf\{n \colon Y_n = 1\}$. Let $X_n = Y_{n\wedge T}$. [Missed this part in lecture. But I think $1 = \mathbb{E}X_T > \mathbb{E}X_0 = 0$, right?] △

Branching Process

Let $(Z_n)_{n\ge 0}$ denote a population at generation n, with $Z_0 = 1$ and

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n+1} \qquad \text{for } n \ge 0,$$

where $(\xi_{i,n})_{i,n\ge 1}$ are i.i.d. $\mathbb{N}$-valued random variables. In particular there are numbers $p_k \ge 0$ with $\mathbb{P}(\xi_{i,n} = k) = p_k$. With this setup in place, a question one could ask is whether the population will die out almost surely or whether it has a chance of living forever.

Lemma 2.5.5. *Let $\mathcal{F}_n = \sigma(\xi_{i,m})_{i\ge 1, 1\le m\le n}$. Then $\frac{Z_n}{\mu^n}$ is a martingale with respect to $(\mathcal{F}_n)_{n\ge 0}$, where $\mu = \mathbb{E}\xi_{i,n}$.*

Proof. We certainly have $Z_n \in m\mathcal{F}_n$. Now fix $k \in \mathbb{N}$ and let $A_k = \{Z_n = k\} \in m\mathcal{F}_n$. Then

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n)\mathbb{1}_{A_k} = \mathbb{E}(Z_{n+1}\mathbb{1}_{A_k}|\mathcal{F}_n) = \mathbb{E}\left(\sum_{i=1}^{k} \xi_{i,n+1}\middle|\mathcal{F}_n\right)\mathbb{1}_{A_k} = \mathbb{E}\left(\sum_{i=1}^{k} \xi_{i,n+1}\right)\mathbb{1}_{A_k} = k\mu\mathbb{1}_{A_k}.$$

So we may sum over k to get

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \sum_{k=0}^{\infty} \mathbb{E}(Z_{n+1}\mathbb{1}_{A_k}|\mathcal{F}_n) = \sum_{k=0}^{\infty} k\mu\mathbb{1}_{A_k} = \mu Z_n. \tag{6}$$

The last equality comes from the fact that

$$Z_n = \sum_{k=0}^{\infty} k\mathbb{1}_{A_k},$$

which follows from the definition of A_k.

Equation (6) guarantees that

$$\mathbb{E}\left(\frac{Z_{n+1}}{\mu^{n+1}}\middle|\mathcal{F}_n\right) = \frac{Z_n}{\mu^n},$$

which proves the lemma. □

Since $\frac{Z_n}{\mu^n} \geq 0$ is a nonnegative martingale, $\frac{Z_n}{\mu^n} \to W$ for some finite limit by the martingale convergence theorem (Corollary 2.4.2). Note that

$$\mathbb{E}\left(\frac{Z_n}{\mu^n}\right) = \mathbb{E}\left(\frac{Z_0}{\mu^0}\right) = 1$$

so that $\mathbb{E}Z_n = \mu^n$. If $\mu < 1$, then $\mathbb{P}(Z_n \geq 1) \leq \mathbb{E}Z_n = \mu^n \downarrow 0$. Let $e = \mathbb{P}(\{Z_n = 0 \text{ eventually}\})$. Then

$$e = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} (1 - \mathbb{P}(Z_n \geq 1)) = 1,$$

so $W = 0$ a.s..

If $\mu = 1$, then Z_n itself is a martingale. Then $Z_n \to W$ a.s.., and $\mathbb{P}(Z_n = W \text{ eventually}) = 1$. If $p_1 = 1$ then $Z_n = 1$ for all n, and $W = 1$. Otherwise, $p_1 < 1$ and hence $p_0 > 0$. In this case

$$\mathbb{P}(Z_n = k \text{ for all } n \geq N) \leq \underbrace{\mathbb{P}(Z_{N+1} = k | Z_N = k)}_{<1,\text{ since } p_0 > 0} \underbrace{\mathbb{P}(Z_{N+2} = k | Z_{N+1} = k)}_{<1,\text{ since } p_0 > 0} \cdots = 0.$$

Again, $e = 1$ and $W = 0$ a.s..

If $\mu > 1$, define the generating function

$$\varphi(t) = \sum_{k\geq 0} p_k t^k.$$

Next lecture, we'll prove

Theorem 2.5.6. *The number e is the unique fixed point of φ in $[0, 1)$.*

(In particular, $e \neq 1$.)

Note that the theorem doesn't say anything about W, and in particular this theorem is consistent with $W = 0$, since Z_n might be subexponential. We will later investigate the behavior of W.

2.6 Feb 6, 2020

We discussed branching processes $(Z_n)_{n\geq 0}$ with $Z_0 = 1$; we saw in Lemma 2.5.5 that $(Z_n/\mu^n)_{n\geq 0}$ is a martingale, where $\mu = \sum_{k\geq 0} kp_k$ is the mean number of offspring per individual.

We showed last time that if $\mu \leq 1$ and $p_1 < 1$ then $e \overset{\text{def}}{=} \mathbb{P}(\{Z_n = 0 \text{ eventually}\}) = 1$. Today, we consider the case $\mu > 1$. We may consider the generating function

$$\varphi(t) = \sum_{k\geq 0} p_k t^k,$$

so that

$$\varphi'(t) = \sum_{k\geq 1} kp_k t^{k-1} \geq 0 \text{ for } t \geq 0 \quad \text{ and } \quad \varphi''(t) = \sum_{k\geq 2} k(k-1)p_k t^{k-2} > 0 \text{ for } t > 0.$$

So the function φ is increasing and convex; furthermore

$$\begin{aligned}
\varphi(1) &= \sum_{k\geq 0} p_k = 1, \\
\varphi'(1) &= \sum_{k\geq 0} kp_k = \mu > 1, \\
\varphi(0) &= p_0 < 1.
\end{aligned}$$

Note that there is a unique $f \in [0, 1)$ so that $\varphi(f) = f$. Existence is the intermediate value theorem on φ' [In particular, $\varphi'(0) < 1$]; uniqueness is convexity.

Theorem 2.6.1 (cf. Theorem 2.5.6)**.** *The number $e = \mathbb{P}(\{Z_n = 0 \text{ eventually}\})$ is equal to the fixed point $f \in [0, 1)$ of φ.*

Proof. Let $e_m = \mathbb{P}(Z_m = 0)$. Then $e_m \uparrow e$. We have

$$e_{m+1} = \sum_{k\geq 0} \mathbb{P}(Z_{m+1} = 0 | Z_1 = k) \underbrace{\mathbb{P}(Z_1 = k)}_{p_k}$$

We claim that $\mathbb{P}(Z_{m+1} = 0 | Z_1 = k) = e_m^k$. This is because each of the k individuals at Z_1 define independent branching processes of length m, and for $Z_{m+1} = 0$ we need each independent branching process to go extinct. We obtain

$$e_{m+1} = \sum_{k\geq 0} e_m^k p_k = \varphi(e_m).$$

It follows that

$$\varphi(e) = \varphi(\lim e_m) = \lim \varphi(e_m) = \lim e_{m+1} = e.$$

By uniqueness of $f \in [0, 1)$, it follows that either $e = f$ or $e = 1$.

We rule out $e = 1$ as follows. Observe that $e_0 = \mathbb{P}(Z_0 = 0) = 0 \leq f$. Then $e_1 = \varphi(e_0) \leq \varphi(f) = f$, because φ is increasing. In particular, by induction we have $e_m \leq f$ for all m. It follows that $e \leq f$. □

We can understand the situtation better using martingales. In particular, recall (by Lemma 2.5.5) that $(Z_n/\mu^n)_{n\geq 0}$ is a nonnegative martingale, hence converges

$$\frac{Z_n}{\mu^n} \to W \quad \text{a.s.},$$

by Corollary 2.4.2. In the case $\mu \leq 1$, then $e = 1$ and $W = 0$ a.s.. The question, which motivates the following martingale theory, is:

Question 2.6.2. *If $\mu > 1$ is it possible that $\mathbb{P}(W > 0) > 0$?*

Orthogonality of martingale increments.
Let $(X_n)_{n\geq 0}$ be a martingale with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$ with $\mathbb{E}X_n^2 < \infty$ for all n.

Lemma 2.6.3. *For all $k < n$ and $Y \in m\mathcal{F}_k$ with $\mathbb{E}Y^2 < \infty$, we have*

$$\mathbb{E}((X_n - X_k)Y) = 0.$$

Proof. We have $\mathbb{E}|(X_n - X_k)Y| < \infty$ by Cauchy-Schwarz. Then

$$\mathbb{E}((X_n - X_k)Y) = \mathbb{E}\big[\mathbb{E}((X_n - X_k)Y|\mathcal{F}_k)\big],$$

by the tower rule (cf. [HW 1, Ex 1]). Since $Y \in m\mathcal{F}_k$,

$$\mathbb{E}\big[\mathbb{E}((X_n - X_k)Y|\mathcal{F}_k)\big] = \mathbb{E}\big[Y\, \underbrace{\mathbb{E}(X_n - X_k|\mathcal{F}_k)}_{=0 \text{ a.s.}}\big] = 0.$$ □

Example 2.6.4. Let us take $Y = X_i - X_j$ for some $i < j < k < n$. Then Lemma 2.6.3 says

$$\mathbb{E}((X_n - X_k)(X_i - X_j)) = 0. \tag{7}$$

In a random walk, the increments $X_n - X_{n-1}$ are independent (by definition). A special case of Equation (7), specifically the case $k = n - 1$ and $i = j - 1$, says

$$\mathbb{E}((X_n - X_{n-1})(X_j - X_{j-1})) = 0,$$

so that the increments $X_n - X_{n-1}$ of *any* martingale are uncorrelated.

In general, independent increments implies martingaleness, which in turn implies uncorrelated increments. Although having uncorrelated increments is formally much weaker than being a martingale, it is often simpler to verify martingaleness only to use it to say that increments are uncorrelated. △

Lemma 2.6.5 (cf. HW 3)**.** *For $k < n$, we have*

$$\mathbb{E}((X_n - X_k)^2|\mathcal{F}_k) = \mathbb{E}(X_n^2|\mathcal{F}_k) - X_k^2.$$

Martingales bounded in L^2. (See Williams Ch. 12)
Let $(M_n)_{n\geq 0}$ be a martingale adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Suppose

$$\sum_{n\geq 1} \mathbb{E}(M_n - M_{n-1})^2 < \infty. \tag{8}$$

Theorem 2.6.6. *Assume Equation (8) holds. Then:*

1. $\sup_n \mathbb{E}M_n^2 < \infty$,
2. *M_n converges a.s.,*
3. *M_n converges in L^2.*

Proof. For part 1, observe that

$$\mathbb{E}M_n^2 = \mathbb{E}M_0^2 + \sum_{k=1}^{n} \mathbb{E}(M_k - M_{k-1})^2,$$

since we may write $M_n = M_0 + (M_1 - M_0) + \cdots + (M_n - M_{n-1})$, and the cross terms $\mathbb{E}((M_k - M_{k-1})M_j)$ vanish for $j < k$ (Lemma 2.6.3). Then

$$\mathbb{E}M_n \leq \mathbb{E}M_0^2 + \sum_{k\geq 1} \mathbb{E}(M_k - M_{k-1})^2 < \infty.$$

Observe also that

$$\sup \mathbb{E}|M_n| \leq \sup(\mathbb{E}M_n^2)^{1/2} < \infty$$

by part 1, so the martingale convergence theorem (Theorem 2.4.1) applies and says

$$M_n \to M_\infty \quad \text{a.s.}$$

for some finite limit M_∞. Then Fatou says

$$\begin{aligned}\mathbb{E}(M_n - M_\infty)^2 &\le \liminf_{k\to\infty} \mathbb{E}(M_n - M_{n+k})^2 \\ &= \liminf_{k\to\infty} \sum_{j=n+1}^{n+k} \mathbb{E}(M_j - M_{j-1})^2 \\ &= \sum_{j\ge n+1} \mathbb{E}(M_j - M_{j-1})^2 \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

where the sum tends to zero because of Equation (8). □

Theorem 2.6.7 (Doob decomposition)**.** *Any submartingale $(X_n)_{n\ge0}$ can be uniquely written as $X_n = M_n + A_n$ where $(M_n)_{n\ge0}$ is a martingale and $(A_n)_{n\ge0}$ is an increasing predictable sequence with $A_0 = 0$.*

Proof. If we had a decomposition, we would need

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = \mathbb{E}(M_n|\mathcal{F}_{n-1}) + \mathbb{E}(A_n|\mathcal{F}_{n-1}) = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n.$$

This would recursively define A_n, namely, we would need

$$A_n - A_{n-1} = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1}. \tag{9}$$

Since $A_0 = 0$, this defines $(A_n)_{n\ge1}$. Then we would need

$$M_n = X_n - A_n,$$

and we would need to check that A_n is predictable and that M_n is a martingale.

Since $A_n - A_{n-1} \in m\mathcal{F}_{n-1}$, it follows that $(A_n)_{n\ge0}$ is predictable. Furthermore since $(X_n)_{n\ge0}$ is a submartingale, $A_n - A_{n-1} \ge 0$, so A_n is increasing. Then $(M_n)_{n\ge0}$ is a martingale, because

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - A_n \overset{(9)}{=} X_{n-1} - A_{n-1} = M_{n-1}$$

□

Let $(M_n)_{n\ge0}$ be a martingale with $\mathbb{E}M_n^2 < \infty$ for all n, and suppose $M_0 = 0$. Then $(M_n^2)_{n\ge0}$ is a submartingale with Doob decomposition

$$M_n^2 = N_n + A_n,$$

where

$$A_n - A_{n-1} = \mathbb{E}(M_n^2|\mathcal{F}_n) - M_{n-1}^2 = \mathbb{E}(M_n^2 - M_{n-1}^2|\mathcal{F}_n) = \mathbb{E}((M_n - M_{n-1})^2|\mathcal{F}_n),$$

with the last equality from orthogonality of increments. Note that

$$\mathbb{E}M_n^2 = \mathbb{E}A_n,$$

e.g. by writing $M_n = \sum_{j=1}^n (M_j - M_{j-1})$, squaring, and taking expectations. So

$$\sup \mathbb{E}M_n^2 < \infty \quad \text{if and only if} \quad \mathbb{E}A_\infty < \infty,$$

where $A_\infty = \lim_{n\to\infty} A_n$ is a $\mathbb{R}_{\ge0} \cup \{+\infty\}$-valued random variable.

Theorem 2.6.8. *The random variable $\lim_{n\to\infty} M_n(\omega)$ exists for almost every ω such that $A_\infty(\omega) < \infty$, i.e.*

$$\{\omega\colon A_\infty(\omega) < \infty\} \subseteq \{\omega\colon \lim M_n(\omega) \text{ exists}\} \cup N$$

for some event N with $\mathbb{P}(N) = 0$.

The A_n's of a square of a martingale are sufficiently important that they have a name and special notation, which varies from book to book. We'll use:

Definition 2.6.9. The A_n's are called the *quadratic variation* of the martingale $(M_n)_{n\ge0}$. They are denoted

$$(\langle M\rangle_n)_{n\ge0} \overset{\text{def}}{=} (A_n)_{n\ge0}.$$

△

2.7 Feb 11, 2020

Last time, we considered a martingale $(M_n)_{n\geq 0}$ adapted to $(\mathcal{F}_n)_{n\geq 0}$, with $\mathbb{E}M_n^2 < \infty$ for all n. If $\sum_{k\geq 1}\mathbb{E}(M_k - M_{k-1})^2 < \infty$, then we showed $\sup_n \mathbb{E}M_n^2 < \infty$ and M_n converges a.s. and in L^2 (Theorem 2.6.6). If $M_n = X_1 + \cdots + X_n$ where the X_i are independent with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = \sigma_i^2 < \infty$, then Theorem 2.6.6 says:

Corollary 2.7.1. *The random variable $\sum_{i=1}^{\infty} X_i$ converges almost surely.*

Proof. Note that $(M_n)_{n\geq 1}$ is a martingale with respect to $\sigma(X_1, \ldots, X_{n-1})$. Then $\mathbb{E}(M_k - M_{k-1})^2 = \mathbb{E}X_k^2 = \sigma_k^2$, so $\sum_{k\geq 1}\mathbb{E}(M_k - M_{k-1})^2 < \infty$ by assumption. Thus M_n converges a.s., by Theorem 2.6.6. □

A strengthening of Corollary 2.7.1 is the Kolmogorov 3-series theorem ([6710, Thm 7.25.6]).

Example 2.7.2. Let's consider a supercritical branching process (so $\mu > 1$), call it $Z_n = \sum_{k=1}^{Z_{n-1}} \xi_{i,n}$. As usual Z_n denote the population of the nth generation, and the $\xi_{i,n}$ are independent $\mathbb{Z}$-valued random variables with $\mathbb{P}(\xi_{i,n} = k) = p_k$. We assume $\mu = \sum kp_k > 1$ and $\sigma^2 = \mathrm{var}(\xi_{i,n}) < \infty$.

Recall that $\mathbb{E}(Z_n|\mathcal{F}_{n-1}) = \mu Z_{n-1}$ (Lemma 2.5.5) and we'll show in [HW 3] that $\mathrm{var}(Z_n|\mathcal{F}_{n-1}) = \sigma^2 Z_{n-1}$. So we have a martingale $M_n = \frac{Z_n}{\mu^n}$. Does Theorem 2.6.6 apply?

Well, we have

$$\mathbb{E}((M_n - M_{n-1})^2|\mathcal{F}_{n-1}) = \frac{1}{\mu^{2n}}\mathbb{E}((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}) = \frac{1}{\mu^{2n}}\mathrm{var}(Z_n|\mathcal{F}_{n-1}) = \frac{1}{\mu^{2n}}\sigma^2 Z_{n-1}.$$

Taking expectations,

$$\mathbb{E}\big(\mathbb{E}((M_n - M_{n-1})^2|\mathcal{F}_{n-1})\big) = \mathbb{E}\left(\frac{1}{\mu^{2n}}\sigma^2 Z_{n-1}\right) = \frac{\sigma^2}{\mu^{2n-(n-1)}} \downarrow 0,$$

since $\mu > 1$, and Theorem 2.6.6 does apply, and $M_n \to M_\infty$ a.s. and in L^2. (Although we knew the a.s. convergence from martingale convergence (Theorem 2.3.11), we now have convergence in L^2 as well.) Since

$$1 = \mathbb{E}M_0 = \mathbb{E}M_n \to \mathbb{E}M_\infty,$$

it follows that $\mathbb{E}M_\infty = 1$, and $\mathbb{P}(M_\infty = 0) < 1$. This answers Question 2.6.2 (!), at least under the assumption that $\sigma^2 < \infty$. △

What if $\sum \mathbb{E}(M_k - M_{k-1})^2 = \infty$? Let

$$\langle M\rangle_n = \sum_{k=1}^{n}\mathbb{E}((M_k - M_{k-1})^2|\mathcal{F}_{k-1})$$

denote the quadratic variation of M. It is a predictable and increasing process, so $\langle M\rangle_n \uparrow \langle M\rangle_\infty$ increases to a random variable taking values in $[0, \infty]$.

Theorem 2.7.3 (L^2 strong law)**.** *We have*

$$\{\langle M\rangle_\infty = \infty\} \overset{\text{a.s.}}{\implies} \left\{\frac{M_n}{\langle M\rangle_n} \to 0\right\}.$$

(The notation $A \overset{\text{a.s.}}{\implies} B$ means $\mathbb{P}(A \cap B^c) = 0$.)

Example 2.7.4. Suppose $M_n = X_1 + \cdots + X_n$ where X_i are independent mean 0 variance 1 random variables. In this case,

$$\langle M\rangle_n = \sum_{k=1}^{n}\mathbb{E}(X_k^2) = n.$$

So Theorem 2.7.3 says $\frac{M_n}{n} \to 0$ a.s.. We knew this as the strong law of large numbers (!). △

The intuition for Theorem 2.7.3 is roughly that $\langle M \rangle_n$ is the best measure for "total elapsed time", in the sense that the most natural measurement occurs when variance is introduced at a constant rate, and that $\langle M \rangle_n$ measures the variance at n. [cf. conditional expectation as orthogonal projection (Claim 1.2.4)...?]

Proof of Theorem 2.7.3. We can assume $M_0 = 0$. We use the *Kronecker Lemma*, which says that if $b_n \in (0, \infty)$ with $b_n \uparrow \infty$, and $x_n \in \mathbb{R}$ such that $\sum_{n\geq 1} \frac{x_n}{b_n}$ converges, then $\lim_{n\to\infty} \frac{x_1+\cdots+x_n}{b_n} = 0$.

We set up a Doob transform. The idea is to bet $f(\langle M_k \rangle)^{-1} = H_k$ on round k, where $f \geq 1$, increasing, and

$$\int_0^\infty \frac{1}{f(t)^2}\, dt < \infty$$

(think, for example, $f(t) = (1+t)^a$ for $a > \frac{1}{2}$). Note that H_k is predictable, so

$$W_n = (H \cdot M)_n = \sum_{k=1}^n (M_k - M_{k-1}) H_k$$

is a martingale. Furthermore,

$$\mathbb{E}((W_k - W_{k-1})^2 | \mathcal{F}_{k-1}) = H_k^2 \mathbb{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}) = H_k^2 (\langle M \rangle_k - \langle M \rangle_{k-1}).$$

The martingale $(W_n)_{n\geq 0}$ has a quadratic variation $\langle W \rangle_n \uparrow \langle W \rangle_\infty$ and

$$\langle W \rangle_\infty = \sum_{k\geq 1} \frac{\langle M \rangle_k - \langle M \rangle_{k-1}}{f(\langle M \rangle_k)^2} \leq \sum_{k\geq 1} \int_{\langle M \rangle_{k-1}}^{\langle M \rangle_k} \frac{dt}{f(t)^2} = \int_0^\infty \frac{dt}{f(t)^2} < \infty.$$

Then $W_n \to W_\infty < \infty$ by Theorem 2.6.6. We can apply Kronecker with $x_n = \langle M \rangle_n - \langle M \rangle_{n-1}$ and $b_n = f(\langle M \rangle_n) \uparrow \infty$. Observe that

$$W_n = \sum_{k=1}^n \frac{x_k}{b_k} \text{ converges as } n \to \infty,$$

so Kronecker says $\frac{x_1+\cdots+x_n}{b_n} \to 0$. [Proof to be finished Thursday.] □

Theorem 2.7.5 (Lévy). *Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration and let $B_1, B_2, \ldots$ be a sequence of events with $B_n \in \mathcal{F}_n$ for all n. Let*

$$Z_n = \sum_{k=1}^n \mathbb{1}_{B_k}, \quad Z_n \uparrow Z_\infty$$

count the number of events that occur by time n. Also let

$$Y_n = \sum_{k=1}^n \xi_k, \quad \text{where} \quad \xi_k = \mathbb{P}(B_k | \mathcal{F}_{k-1}) = \mathbb{E}(\mathbb{1}_{B_k} | \mathcal{F}_{k-1})$$

be the "running forecast of Z_n". Then:

1. $Y_\infty < \infty \overset{\text{a.s.}}{\Longrightarrow} Z_\infty < \infty$
2. $Y_\infty = \infty \overset{\text{a.s.}}{\Longrightarrow} Z_n / Y_n \to 1$.

Theorem 2.7.5 contains both Borel Cantelli lemmas as corollaries:

- (Borel-Cantelli 1): If $\sum_{k\geq 1} \mathbb{P}(B_k) < \infty$, then $\mathbb{E} Y_\infty < \infty$, so $Y_\infty < \infty$ a.s., and part 1 of Theorem 2.7.5 says $Z_\infty < \infty$ a.s., so only finitely many B_k occur.
- (Borel-Cantelli 2): If B_k are independent, then $\xi_k = \mathbb{P}(B_k)$ a.s.. Let $\mathcal{F}_k = \sigma(B_1, \ldots, B_k)$. If $\sum \mathbb{P}(B_\infty) = \infty$, then $Y_\infty = \infty$ a.s., and part 2 of Theorem 2.7.5 says $Z_\infty = \infty$ a.s., so infinitely many B_k occur.

Proof of Theorem 2.7.5. Since $(Z_n)_{n\geq 0}$ is a submartingale, it has a Doob decomposition (Theorem 2.6.7)

$$Z_n = M_n + Y_n,$$

where Y_n is as given in the theorem statement. Then

$$\langle M\rangle_n - \langle M\rangle_{n-1} = \mathbb{E}((M_n - M_{n-1})^2|\mathcal{F}_{n-1}) = \mathbb{E}((\mathbb{1}_{B_n} - \xi_n)^2|\mathcal{F}_{n-1}) = \text{var}(\mathbb{1}_{B_n}|\mathcal{F}_{n-1}) = \xi_n(1-\xi_n) \leq \xi_n.$$

Summing the inequality over n, we obtain

$$\langle M\rangle_n \leq \xi_1 + \cdots + \xi_n = Y_n.$$

Since $Y_\infty < \infty \overset{\text{a.s.}}{\Longrightarrow} \langle M\rangle_\infty < \infty$, we conclude that $\lim M_n$ exists in $\mathbb{R}$, and hence $\lim Z_n$ exists in $\mathbb{R}$. In other words, $Z_\infty < \infty$. This is part 1 of the theorem.

Now observe that

$$\{Y_\infty = \infty, \langle M\rangle_\infty < \infty\} \overset{\text{a.s.}}{\Longrightarrow} \{\lim M_n \text{ exists in } \mathbb{R}\},$$

so

$$\frac{Z_n}{Y_n} = \frac{M_n}{Y_n} + \frac{Y_n}{Y_n} \to 0 + 1.$$

On the other hand,

$$\{Y_\infty = \infty = \langle M\rangle_\infty\} \overset{\text{a.s.}}{\Longrightarrow} \frac{M_n}{\langle M\rangle_n} \to 0$$

which implies $\frac{M_n}{Y_n} \to 0$ and hence $\frac{Z_n}{Y_n} \to 1$. In total, we've verified part 2 of Theorem 2.7.5. □

2.8 Feb 13, 2020

Let's discuss uniform integrability:

Definition 2.8.1. A set of random variables $\{X_i\}_{i\in I}$ is *uniformly integrable* (U.I.) if

$$\lim_{M\to\infty} \sup_{i\in I} \mathbb{E}(|X_i| \mathbb{1}_{\{|X_i|>M\}}) = 0. \qquad \triangle$$

This condition can be hard to check with bare hands, but we'll see some sufficient conditions later. Uniform integrability has nice consequences:

Remark 2.8.2. Let $\{X_i\}_{i\in I}$ be uniformly integrable. Then for sufficiently large M, the supremum is less than 1. It follows that

$$\sup_{i\in I} \mathbb{E}|X_i| = \sup_{i\in I}(\mathbb{E}|X_i|\mathbb{1}_{\{|X_i|\le M\}} + \mathbb{E}|X_i|\mathbb{1}_{\{|X_i|>M\}}) \le M+1 < \infty.$$

The converse does not hold, i.e. $\sup \mathbb{E}|X_i| < \infty$ does not imply uniform integrability. We may take the standard example $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}, \lambda)$ where λ is the Lebesgue measure. Set $X_n = n\mathbb{1}_{[0,\frac{1}{n})}$. Then $\mathbb{E}X_n = 1$ for all n. But also

$$\mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}) = 1 \qquad \text{for all } n > M,$$

so $(X_n)_{n\in\mathbb{N}}$ is not uniformly integrable. (The limit is equal to 1.) $\triangle$

Let's discuss some sufficient conditions for uniform integrability.

Lemma 2.8.3.

1. *If $|X_i| \le Y$ for all $i \in I$, and $\mathbb{E}Y < \infty$, then $\{X_i\}_{i\in I}$ is uniformly integrable.*
2. *If $\sup_{i\in I} \mathbb{E}|X_i|^p < \infty$ for some $p > 1$, then $\{X_i\}_{i\in I}$ is uniformly integrable.*

Proof. We verify part 1. Observe that

$$\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) \le \mathbb{E}(Y\mathbb{1}_{\{|X_i|>M\}}) \le \mathbb{E}(Y\mathbb{1}_{\{Y>M\}}) \to 0.$$

We next verify part 2. Note that if $x \ge M > 0$ then $x \le M^{1-p}x^p$. So

$$\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) \le \mathbb{E}(M^{1-p}|X_i|^p\mathbb{1}_{\{|X_i|>M\}}) \le M^{1-p}\underbrace{\sup \mathbb{E}|X_i|^p}_{\text{finite}} \to 0 \qquad \text{as } M \to \infty. \qquad \square$$

More nontrivial than Lemma 2.8.3 is

Theorem 2.8.4. *Let $X \in L^1(\Omega, \mathcal{F}_0, \mathbb{P})$. Then $\{\mathbb{E}(X|\mathcal{F})\}_{\mathcal{F}\subseteq\mathcal{F}_0}$ is uniformly integrable.*

By $\{\mathbb{E}(X|\mathcal{F})\}$ we mean the family which contains every version of every $\mathbb{E}(X|\mathcal{F})$; here, $\mathcal{F}$ runs over all sub-σ-fields of $\mathcal{F}_0$.

To prove Theorem 2.8.4 we use

Lemma 2.8.5. *If $\mathbb{E}|X| < \infty$ then for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\mathbb{E}|X|\mathbb{1}_A < \varepsilon$ for all $A \in \mathcal{F}_0$ with $\mathbb{P}(A) \le \delta$.*

[There was a comment about the lemma was saying that the measure $\nu(A) \overset{\text{def}}{=} \mathbb{E}(|X|\mathbb{1}_A)$ is absolutely continuous with respect to $\mu(A) \overset{\text{def}}{=} \mathbb{P}(A)$.]

Proof. Otherwise, there exist events $A_1, A_2, \ldots$ with $\mathbb{P}(A_n) \le 2^{-n}$ and $\mathbb{E}|X|\mathbb{1}_{A_n} > \varepsilon$ for all n.

By Borel-Cantelli 1, $\mathbb{P}(\{A_n \text{ i.o.}\}) = 0$, so $|X|\mathbb{1}_{A_n} \to 0$ a.s.. Dominated convergence says $\mathbb{E}|X|\mathbb{1}_{A_n} \to 0$. $\square$

Proof of Theorem 2.8.4. Fix $\varepsilon > 0$ and choose δ from Lemma 2.8.5, i.e. choose δ so that

$$\mathbb{P}(A) \le \delta \implies \mathbb{E}|X|\mathbb{1}_A \le \varepsilon \quad \text{for all } A \in \mathcal{F}_0.$$

Now choose M so that $\frac{1}{M}\mathbb{E}|X| < \delta$. If Y is any version of $\mathbb{E}(X|\mathcal{F})$ then Jensen's inequality says

$$Y \le \mathbb{E}\big(|X|\big|\mathcal{F}\big) \quad \text{a.s..}$$

So

$$M\mathbb{P}(|Y| > M) \le \mathbb{E}|Y| \le \mathbb{E}|X|,$$

and $\mathbb{P}(|Y| > M) < \delta$. Define $A = \{|Y| > M\} \in \mathcal{F}$. We have

$$|Y|\mathbb{1}_A \le \mathbb{E}\big(|X|\big|\mathcal{F}\big) = \mathbb{E}\big(|X|\mathbb{1}_A\big|\mathcal{F}\big).$$

Taking expectations of both sides,

$$\mathbb{E}(|Y|\mathbb{1}_A) \le \mathbb{E}(|X|\mathbb{1}_A) < \varepsilon.$$

This verifies uniform integrability. ∎

Uniform integrability and L^1 convergence.

Theorem 2.8.6. *Let $X_n \to X$ in probability. The following are equivalent:*

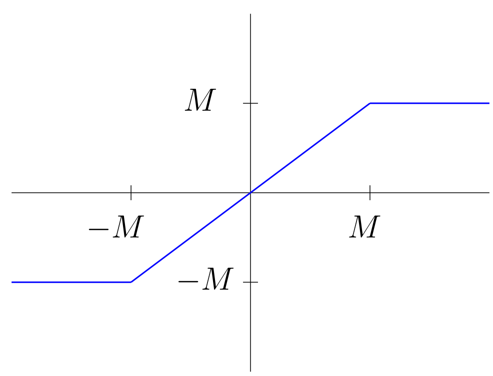
1. *$\{X_n\}_{n\ge 0}$ is uniformly integrable*
2. *$X_n \to X$ in L^1 (i.e., $\mathbb{E}|X_n - X| \to 0$)*
3. *$\mathbb{E}|X_n| \to \mathbb{E}|X|$.*

Proof. Let's prove that item 1 implies item 2. Let

$$\varphi_M : \mathbb{R} \to [-M, M]$$

$$\varphi_M(x) = \begin{cases} -M & \text{if } x < -M \\ x & \text{if } x \in [-M, M] \\ M & \text{if } x > M \end{cases}$$

with graph that looks like



Note that for all Y, we have

$$|\varphi_M(Y) - Y| = (|Y| - M)^+ \le |Y|\mathbb{1}_{\{|Y|>M\}}. \tag{10}$$

Then

$$\begin{aligned} |X_n - X| &\le |X_n - \varphi_M(X_n)| + |\varphi_M(X_n) - \varphi_M(X) + \mathbb{E}|\varphi_M(X) - X| \\ &\underset{(10)}{\le} \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}) + \mathbb{E}|\varphi_M(X_n) - \varphi_M(X)| + \mathbb{E}(|X|\mathbb{1}_{\{|X|>M\}}). \end{aligned}$$

The first summand satisfies $\mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}) < \varepsilon$ by uniform integrability of X_n.

The third summand satisfies $\mathbb{E}(|X|\mathbb{1}_{\{|X|>M\}}) < \varepsilon$ by Fatou: since $|X_{n_k}| \to |X|$ a.s., Fatou says $\mathbb{E}|X| \leq \sup_n \mathbb{E}|X_n| < \infty$ and its truncations $\mathbb{E}(|X|\mathbb{1}_{\{|X|>M\}})$ can be made as small as desired.

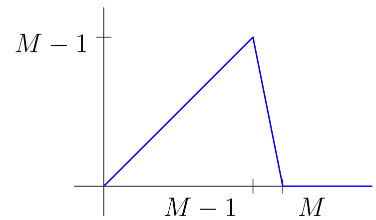
Finally, the second summand satisfies $\mathbb{E}|\varphi_M(X_n) - \varphi_M(X)| < \varepsilon$ because $X_n \xrightarrow{\mathbb{P}} X$ implies $\varphi_M(X_n) \xrightarrow{\mathbb{P}} \varphi_M(X)$, and bounded convergence theorem says $\mathbb{E}|\varphi_M(X_n) - \varphi_M(X)| \to 0$.

This completes the proof of item 1 implying item 2.

Let's show that item 2 implies item 3. Observe that $\mathbb{E}|X_n - X| \to 0$ and $\mathbb{E}|X| \leq \mathbb{E}|X_n - X| + \mathbb{E}|X_n| < \infty$. Then Jensen says

$$\big|\mathbb{E}|X| - \mathbb{E}|X_n|\big| \leq \mathbb{E}\big||X| - |X_n|\big| \leq \mathbb{E}|X - X_n| \to 0.$$

Finally, let's show item 3 implies item 1. Let $\psi_M\colon [0,\infty) \to [0, M]$ be given by the graph



and defined piecewise linearly by the formula

$$\psi_M(x) = \begin{cases} x & \text{if } x \leq M-1 \\ (M-x)(M-1) & \text{if } M-1 \leq x \leq M \\ 0 & \text{if } x > M \end{cases}$$

Bounded convergence says

$$\mathbb{E}\psi_M(|X_n|) \to \mathbb{E}\psi_M(|X|) \qquad \text{as } n \to \infty. \tag{11}$$

Since $X \in L^1$, dominated convergence says

$$\mathbb{E}\psi_M(|X|) \to \mathbb{E}|X| \qquad \text{as } M \to \infty \tag{12}$$

Then, for $n \geq n_0$ sufficiently large,

$$\mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}) \leq \mathbb{E}|X_n| - \mathbb{E}\psi_M(|X_n|) \overset{(11)}{<} \mathbb{E}|X| - \mathbb{E}\psi_M(|X|) + \varepsilon \overset{(12)}{<} 2\varepsilon \qquad \text{for } M \geq M_0.$$

(Here, the assumption $\mathbb{E}|X_n| \to \mathbb{E}|X|$ is also used to conclude $\mathbb{E}|X_n| - \mathbb{E}\psi_M(|X_n|) < \mathbb{E}|X| - \mathbb{E}\psi_M(|X|) + \varepsilon$.) Take M larger if needed so that

$$\mathbb{E}|X| - \mathbb{E}\psi_M(|X|) + \varepsilon < 2\varepsilon$$

also holds for $n = 1, \ldots, n_0 - 1$. It follows that $(X_n)_{n\geq 0}$ is uniformly integrable. □

L^1 convergence theorems for martingales.

Lemma 2.8.7. *If $(X_n)_{n\geq 0}$ is a martingale adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$ and $X_n \to X$ in L^1, then $X_n = \mathbb{E}(X|\mathcal{F}_n)$ for all $n \geq 0$.*

Proof. For $m > n$ we have $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$, so for $A \in \mathcal{F}_n$ we have

$$\mathbb{E}(X_m|\mathcal{F}_n)\mathbb{1}_A = X_n\mathbb{1}_A.$$

This gives

$$\mathbb{E}(X_n\mathbb{1}_A) = \mathbb{E}(X_m\mathbb{1}_A) \to \mathbb{E}(X\mathbb{1}_A). \tag{13}$$

Note that because $X_m \to X$ in L^1,

$$|\mathbb{E}(X_m\mathbb{1}_A) - \mathbb{E}(X\mathbb{1}_A)| \leq \mathbb{E}|X_m\mathbb{1}_A - X\mathbb{1}_A| \leq \mathbb{E}|X_m - X| \to 0$$

so $X_m\mathbb{1}_A \to X\mathbb{1}_A$ in L^1 for all $A \in \mathcal{F}$. With this in mind, Equation (13) implies

$$\mathbb{E}(X|\mathbb{1}_A) = \mathbb{E}(X_n\mathbb{1}_A)$$

for all $A \in \mathcal{F}_n$. It follows that $X_n = \mathbb{E}(X|\mathcal{F}_n)$ a.s., by definition of conditional expectation. □

2.9 Feb 18, 2020

[OH will be moved to Tuesday, 3–4 at 438 MLT this week.]
Last time we showed

Lemma 2.9.1 (cf. Lemma 2.8.7)**.** *If $(X_n)_{n\geq 0}$ is a martingale adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$ and $X_n \to X$ in L^1, then $X_n = \mathbb{E}(X|\mathcal{F}_n)$ for all $n \geq 0$.*

Using this, let's prove

Theorem 2.9.2. *Let $(X_n)_{n\geq 0}$ be a martingale adapted to $(\mathcal{F}_n)_{n\geq 0}$. The following are equivalent:*

1. *$\{X_n\}$ is uniformly integrable*
2. *X_n converges a.s. and in L^1*
3. *X_n converges in L^1*
4. *There is X with $\mathbb{E}|X| < \infty$ such that $X_n = \mathbb{E}(X|\mathcal{F}_n)$ a.s.*

Proof. We first show item 1 implies item 2. In Remark 2.8.2 we showed that uniform integrability implies $\sup_n \mathbb{E}|X_n| < \infty$. This implies $X_n \to X$ a.s. by martingale convergence theorem. Then by Theorem 2.8.6, $X_n \to X$ in L^1.

That item 2 implies item 3 is trivial.

That item 3 implies item 4 is Lemma 2.9.1.

That item 4 implies item 1 follows because $\{\mathbb{E}(X|\mathcal{F}) \colon \mathcal{F} \subseteq \mathcal{F}_0\}$ is uniformly integrable (Theorem 2.8.4). □

Theorem 2.9.3. *Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration, and let*

$$\mathcal{F}_\infty = \sigma\Big(\bigcup_{n\geq 0} \mathcal{F}_n\Big).$$

Let $\mathbb{E}|X| < \infty$. Then $\mathbb{E}(X|\mathcal{F}_n) \to \mathbb{E}(X|\mathcal{F}_\infty)$ a.s. and in L^1.

Corollary 2.9.4 (Lévy 0-1 law)**.** *If $A \in \mathcal{F}_\infty$, then $\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) \to \mathbb{1}_A$ a.s. and in L^1.*

This is just a special case of Theorem 2.9.3, when $X = \mathbb{1}_A$ for $A \in \mathcal{F}_\infty$. Although this looks innocent, it implies the Kolmogorov 0-1 law:

Corollary 2.9.5 (Kolmogorov 0-1 law, cf. [6710, Thm 7.25.4])**.** *Let $Y_1, Y_2, \ldots$ be independent random variables and let*

$$A \in \mathcal{T} \overset{\text{def}}{=} \bigcap_{n\geq 1} \sigma(Y_n, Y_{n+1}, Y_{n+2}, \ldots).$$

Then $\mathbb{P}(A) \in \{0, 1\}$.

Proof. Let's show that the Kolmogorov 0-1 law follows from the Lévy 0-1 law. Let $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$; observe that $\mathcal{F}_n \perp \sigma(Y_{n+1}, Y_{n+2} \ldots) \supseteq \mathcal{T}$. Since A is independent of $\mathcal{F}_n$, we obtain

$$\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) = \mathbb{E}\mathbb{1}_A = \mathbb{P}(A).$$

The Lévy 0-1 law says $\mathbb{P}(A) = \mathbb{1}_A$ a.s., where $\mathbb{P}(A)$ is interpreted as the constant function. So $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$. □

Proof of Theorem 2.9.3. Let $X_n = \mathbb{E}(X|\mathcal{F}_n)$. Note that $(X_n)_{n\geq 0}$ is a uniformly integrable martingale. Thus Theorem 2.9.2 says $X_n \to X_\infty$ a.s. and in L^1. It remains to show that $X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$ a.s..

Lemma 2.8.7 says $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$ for all $n < \infty$. In particular, for all $A \in \mathcal{F}_n$ we have

$$\mathbb{E}X\mathbb{1}_A = \mathbb{E}X_\infty\mathbb{1}_A.$$

Note that $\cup_{n\geq 0}\mathcal{F}_n$ is a π-system containing Ω and generating $\mathcal{F}_\infty$. So $\mathbb{E}(X|\mathcal{F}_\infty) = \mathbb{E}(X_\infty|\mathcal{F}_\infty)$ a.s., by [HW 1, Ex 2]. Then $\mathbb{E}(X_\infty|\mathcal{F}_\infty) = X_\infty$ a.s. because $X_\infty \in m\mathcal{F}_\infty$. □

Backwards martingales.

Let $(\mathcal{F}_n)_{n\le 0}$ be a filtration, so

$$\cdots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{n-1} \subseteq \mathcal{F}_0.$$

Definition 2.9.6. We say $(X_n)_{n\le 0}$ is a *backwards martingale* if

1. $\mathbb{E}|X_n| < \infty$ for all n,
2. $X_n \in m\mathcal{F}_n$ for all n,
3. $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ a.s., for all $n \le -1$.

△

Theorem 2.9.7. *For any backwards martingale $(X_n)_{n\le 0}$, the limit*

$$X_{-\infty} \lim_{n\to-\infty} X_n$$

exists a.s. and the convergence is in L^1.

Proof. Fix $a < b \in \mathbb{R}$ and let U_n be the number of upcrossings of $[a,b]$ by $X_{-n}, X_{-n+1}, \dots, X_0$. (Recall that an upcrossing is (s,t) with $s < t$, $X_s < a$, and $X_t > b$.)

Note that $U_n \uparrow U_\infty$, so $\mathbb{E}U_n \uparrow \mathbb{E}U_\infty$. The upcrossing inequality says

$$\mathbb{E}U_n \le \frac{\mathbb{E}(X_0 - a)^+}{b-a} < \infty \text{ for all } n,$$

and in particular $\mathbb{E}U_\infty$ satisfies the same inequality. Thus U_∞ is a.s. finite, and by the same proof as the martingale convergence theorem (Theorem 2.3.11) we obtain

$$\mathbb{P}(\lim_{n\to-\infty} X_n \text{ exists in } \mathbb{R}) = 1.$$

This gives the a.s. convergence.

Let's show the L^1 convergence. For all $n \le 0$ we have $\mathbb{E}(X_0|\mathcal{F}_n) = X_n$, by the martingale property. In particular the $(X_n)_{n\ge 0}$ is uniformly integrable. In light of the a.s. convergence of X_n, the L^1 convergence of the X_n follows (e.g. by Theorem 2.8.6). □

Theorem 2.9.8 (Lévy's downward theorem)**.** *Let $(X_n)_{n\ge 0}$ be a backwards martingale. Let $X_{-\infty} = \lim_{n\to-\infty} X_n$ and let $\mathcal{F}_{-\infty} = \cap_{n\le 0}\mathcal{F}_n$. Then*

$$X_{-\infty} = \mathbb{E}(X_0|\mathcal{F}_{-\infty}) \quad \text{a.s..}$$

Proof. Note that $X_{-\infty} \in m\sigma((X_k)_{k\le n}) \subseteq \mathcal{F}_n$ for all $n \le 0$. This implies that $X_{-\infty} \in m\mathcal{F}_{-\infty}$. If $A \in \mathcal{F}_{-\infty}$¡ then since $X_n = \mathbb{E}(X_0|\mathcal{F}_n)$ a.s. and $A \in \mathcal{F}_n$ we have

$$\mathbb{E}X_n \mathbb{1}_A = \mathbb{E}X_0 \mathbb{1}_A.$$

On the other hand $\mathbb{E}X_n\mathbb{1}_A \to \mathbb{E}X_{-\infty}\mathbb{1}_A$ since $X_n \to X_{-\infty}$ in L^1 (same yoga as in Lemma 2.8.7), and we obtain

$$\mathbb{E}X_{-\infty}\mathbb{1}_A = \mathbb{E}X_0\mathbb{1}_A.$$

Hence $X_{-\infty} = \mathbb{E}(X_0|\mathcal{F}_{-\infty})$ a.s. □

Corollary 2.9.9. *If $\mathbb{E}|Y| < \infty$ and $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ as $n \downarrow -\infty$, then $\mathbb{E}(Y|\mathcal{F}_n) \to \mathbb{E}(Y|\mathcal{F}_{-\infty})$.*

Here, the notation $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ means that

$$\mathcal{F}_0 \supseteq \mathcal{F}_{-1} \supseteq \mathcal{F}_{-2} \supseteq \dots \quad \text{and} \quad \bigcap_{n\le 0} \mathcal{F}_n = \mathcal{F}_{-\infty}.$$

Proof. Let $X_n = \mathbb{E}(Y|\mathcal{F}_n)$. This is a backwards martingale, so $X_n \to X_{-\infty}$ a.s. and in L^1 for some limit $X_{-\infty}$. It follows that

$$X_{-\infty} = \mathbb{E}(X_0|\mathcal{F}_{-\infty}) = \mathbb{E}(\mathbb{E}(Y|\mathcal{F}_0)|\mathcal{F}_{-\infty}) = \mathbb{E}(Y|\mathcal{F}_{-\infty}) \quad \text{a.s.},$$

because the smaller σ-algebra wins. □

Backwards martingales allow us to prove the sharp strong law of large numbers. We've seen various weaker forms of this, such as when fourth moments are bounded [6710, Thm 5.13.4] and when second moments are bounded (Theorem 2.7.3). But now we can prove:

Theorem 2.9.10. *Let $\xi_1, \xi_2, \ldots$ be i.i.d. with $\mathbb{E}|\xi_i| < \infty$. Let $S_n = \xi_1 + \cdots + \xi_n$. Then*

$$\frac{S_n}{n} \to \mathbb{E}\xi_1 \quad \text{a.s.}.$$

Proof. For $n \geq 0$, define

$$X_{-n} \overset{\text{def}}{=} \frac{S_n}{n} \quad \text{and} \quad \mathcal{F}_{-n} \overset{\text{def}}{=} \sigma(S_n, S_{n+1}, \ldots).$$

We want to show this is a backwards martingale, so we need to understand the random variable $\mathbb{E}(X_{-n}|\mathcal{F}_{-n-1})$.

Observe that

$$\mathbb{E}(\xi_j|\mathcal{F}_{-n-1}) = \mathbb{E}(\xi_k|\mathcal{F}_{-n-1}) \quad \text{a.s.},$$

for all $j, k \leq n$. This is by symmetry: note that $\xi_j \overset{\text{d}}{=} \xi_k$, and for $N \geq n+1$ the random variable S_N depends on ξ_j and ξ_k only via $\xi_j + \xi_k$. Then

$$\begin{aligned}
\mathbb{E}(\xi_{n+1}|\mathcal{F}_{-n-1}) &= \frac{1}{n+1}\sum_{j=1}^{n+1} \mathbb{E}(\xi_j|\mathcal{F}_{n-1}) \\
&= \mathbb{E}\left(\frac{1}{n+1}\sum_{j=1}^{n+1} \xi_j \middle| \mathcal{F}_{-n-1}\right) \\
&= \mathbb{E}\left(\frac{S_{n+1}}{n+1} \middle| \mathcal{F}_{-n-1}\right) \\
&= \frac{S_{n+1}}{n+1}, \quad \text{a.s.}.
\end{aligned}$$

Let's come back to trying to understand $\mathbb{E}(X_{-n}|\mathcal{F}_{-n-1})$. Note that

$$\mathbb{E}(X_{-n}|\mathcal{F}_{-n-1}) = \mathbb{E}\left(\frac{S_{n+1} - \xi_{n+1}}{n} \middle| \mathcal{F}_{-n-1}\right) = \frac{S_{n+1}}{n} - \frac{1}{n}\mathbb{E}(\xi_{n+1}|\mathcal{F}_{-n-1}) = \frac{S_{n+1}}{n} - \frac{1}{n}\frac{S_{n+1}}{n+1} = X_{-n-1}.$$

We've verified that $(X_n)_{n\leq 0}$ is a backwards martingale. It follows that $\lim \frac{S_n}{n}$ exists a.s. and in L^1 (Theorem 2.9.7). But

$$\lim \frac{S_n}{n} = \mathbb{E}(X_{-1}|\mathcal{F}_{-\infty}) = \mathbb{E}(\xi_1|\mathcal{F}_{-\infty}).$$

Note that $\lim \frac{S_n}{n} \in m\mathcal{T}$, so by the Kolmogorov 0-1 law (Corollary 2.9.5) we have

$$\mathbb{P}\left(\lim \frac{S_n}{n} = x\right) \in \{0, 1\} \quad \text{for all } x \in \mathbb{R},$$

and $\lim \frac{S_n}{n}$ is an a.s. constant. Since $\mathbb{E}(\xi_1|\mathcal{F}_{-\infty})$ is an a.s. constant, this constant must be $\mathbb{E}(\xi_1)$. In total, we've verified

$$\lim \frac{S_n}{n} = \mathbb{E}(\xi_1).$$

□

2.10 Feb 20, 2020

[I was out of town. I am grateful for Jake Wasserstein's and Sara Venkatraman's notes, from which I copied.]

We begin with

Theorem 2.10.1 (Doob's L^2 inequality)**.** *If $M_n \geq 0$ is a submartingale with $\mathbb{E}M_n^2 < \infty$ for all n, then*

$$\mathbb{E}(\max(M_0, M_1, \dots, M_n)^2) \leq 4\mathbb{E}M_n^2,$$

and $\mathbb{E}(\sup M_n^2) \leq 4 \sup \mathbb{E}M_n^2$.

Proof. See Durrett 4.4, or Williams Ch. 14. □

Let $X_n \geq 0$ be independent random variables with $\mathbb{E}X_n = 1$. Then $M_n = X_1 \dots X_n$ is a martingale. Note that by the martingale convergence theorem (Theorem 2.3.11), $\lim M_n$ exists a.s. in $\mathbb{R}$.

Example 2.10.2. If $X_n \sim \text{Unif}(0, 2)$, then $\lim M_n = 0$ a.s. (cf. [6710 HW6, Ex 2]). △

When is $\lim M_n$ not equal to zero a.s.?

Theorem 2.10.3 (Kakutani)**.** *Let $X_n \geq 0$ be independent random variables with $\mathbb{E}X_n = 1$ for all n. Then $M_n = X_1 \dots X_n$ is a martingale. Let $a_n = \mathbb{E}\sqrt{X_n} \leq \sqrt{\mathbb{E}X_n} = 1$. (The inequality holds by Jensen.) Then:*

1. *If $\prod_{n\geq 1} a_n > 0$ then $M_n \to M_\infty$ in L^1, and hence $\mathbb{E}M_\infty = \mathbb{E}M_n = 1$, so $\mathbb{P}(M_\infty > 0) > 0$.*
2. *If $\prod_{n\geq 1} a_n = 0$, then $\mathbb{P}(M_\infty = 0) = 1$.*

Proof. Let

$$N_n = \prod_{i=1}^{n} \frac{X_i^{1/2}}{a_i}.$$

We claim that N_n is a martingale. Indeed, we compute

$$\mathbb{E}(N_{n+1}|\mathcal{F}_n) = \mathbb{E}\left(\frac{X_{n+1}^{1/2}}{a_{n+1}} N_n \middle| \mathcal{F}_n\right) = 1 \cdot N_n,$$

so $(N_n)_{n\geq 1}$ is a martingale. Since $N_n \geq 0$ there is an a.s. limit $N_n \to N_\infty$.

Let's now show the first item. Note that

$$M_n = a_1^2 \dots a_n^2 N_n^2 \leq N_n^2,$$

so

$$\mathbb{E}N_n^2 = \frac{\mathbb{E}X_1 \dots \mathbb{E}X_n}{a_1^2 \dots a_n^2} = \prod_{i=1}^{n} \frac{1}{a_i^2} < \left(\prod_{i\geq 1} \frac{1}{a_i}\right)^2 < \infty,$$

because we assumed $\prod_{n\geq 1} a_n > 0$. Because M_n is dominated by $\sup M_n$, and

$$\mathbb{E}(\sup M_n) \leq \mathbb{E}(\sup N_n)^2 \leq 4 \sup \mathbb{E}N_n^2 \leq \frac{4}{\prod_{n\geq 1} a_n^2} < \infty,$$

it follows that M_n is uniformly integrable and $M_n \to M_\infty$ in L^1.

To show item 2, we compute

$$M_\infty = \lim M_n = \lim(a_1^2 \dots a_n^2) N_n^2 = \lim(a_1^2 \dots a_n^2) \cdot \underbrace{\lim N_n^2}_{\text{a.s. finite}} = 0.$$ □

Two stopping times.

Definition 2.10.4. If $(\mathcal{F}_n)_{n\geq 0}$ is a filtration and L is a stopping time, then σ-algebra $\mathcal{F}_L$ consists of the sets A such that $A \cap \{L = \ell\} \in \mathcal{F}_\ell$ for all $\ell \in \mathbb{N}$. Note that $Y_L \in m\mathcal{F}_L$. △

Theorem 2.10.5. *Let $(Y_n)_{n\geq 0}$ be a uniformly integrable submartingale adapted to $(\mathcal{F}_n)_{n\geq 0}$. Let $L \leq M$ be stopping times, and assume M is finite a.s.. Then*

$$Y_L \leq \mathbb{E}(Y_M|\mathcal{F}_L) \qquad \text{a.s..}$$

Proof. Let $A \in \mathcal{F}_L$ and let $N = L\mathbb{1}_A + M\mathbb{1}_{A^c} \leq M$. One can check that N is a stopping time.

Let $X_n = Y_{M\wedge n}$. Lemma 2.10.6 will show that X_n is uniformly integrable. By uniform integrability, we know that $X_n \to X_\infty$ a.s. and $\mathbb{E}X_n \to \mathbb{E}X_\infty = \mathbb{E}Y_M$, because $X_\infty = Y_M$. Now,

$$\mathbb{E}Y_N = \mathbb{E}Y_{N\wedge M} = \mathbb{E}X_N \leq \mathbb{E}X_\infty = \mathbb{E}Y_M$$

with the last inequality from the forthcoming Theorem 2.10.7. Since $N = M$ on A^c and $Y_N = Y_N\mathbb{1}_A + Y_N\mathbb{1}_{A^c}$, we have

$$\mathbb{E}(Y_L\mathbb{1}_A) = \mathbb{E}(Y_N\mathbb{1}_A) \leq \mathbb{E}(Y_M\mathbb{1}_A) = \mathbb{E}(\mathbb{E}(Y_M|\mathcal{F}_L)\mathbb{1}_A).$$

Let $A_\varepsilon = \{Y_L - \mathbb{E}(Y_M|\mathcal{F}_L) > \varepsilon\} \in \mathcal{F}_L$. By the above inequality,

$$\varepsilon\mathbb{P}(A_\varepsilon) \leq \mathbb{E}(Y_L - \mathbb{E}(Y_m|\mathcal{F}_L)) \leq 0,$$

so $\mathbb{P}(A_\varepsilon) = 0$. □

Lemma 2.10.6. *If $(X_n)_{n\geq 0}$ is a uniformly integrable submartingale and N is a stopping time, then $(X_{N\wedge n})_{n\geq 0}$ is uniformly integrable.*

(The lemma holds even if $\mathbb{P}(N = \infty) > 0$; we use the convention that $\infty \wedge n = n$ for all $n \in \mathbb{N}$.)

Proof. Note that $0 \leq N \wedge n \leq n$, and $(X_n^+)_{n\geq 0}$ is also a uniformly integrable submartingale (because $x \mapsto x^+$ is convex). We already know that

$$\mathbb{E}X_{N\wedge n}^+ \leq \mathbb{E}X_n^+,$$

so $\sup \mathbb{E}X_{N\wedge n}^+ \leq \sup \mathbb{E}X_n^+ < \infty$, since $(X_n)_{n\geq 0}$ is uniformly integrable. The martingale convergence theorem says $X_{N\wedge n} \to X_N$ a.s. and $\mathbb{E}|X_N| < \infty$. We obtain

$$\mathbb{E}(|X_{N\wedge n}|\mathbb{1}_{\{|X_{N\wedge n}|>k\}}) = \mathbb{E}(|X_N|\mathbb{1}_{\{|X_N|>k, N\leq n\}}) + \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>k, N>n\}}) < \varepsilon + \varepsilon,$$

because $\mathbb{E}|X_N| < \infty$ and $(X_n)_{n\geq 0}$ is uniformly integrable. □

Theorem 2.10.7. *If $(X_n)_{n\geq 0}$ is a uniformly integrable submartingle, with $X_n \to X_\infty$ a.s., and N is a stopping time, then $\mathbb{E}X_0 \leq \mathbb{E}X_N \leq \mathbb{E}X_\infty$.*

Proof. We have, for every n,

$$\mathbb{E}X_0 \leq \mathbb{E}X_{N\wedge n} \leq \mathbb{E}X_n.$$

But $X_n \to X_\infty$ a.s. and in L^1 by uniform integrability, and $X_{N\wedge n} \to X_N$ a.s. and in L^1 by uniform integrability guaranteed by Lemma 2.10.6. We obtain

$$\mathbb{E}X_0 \leq \mathbb{E}X_N \leq \mathbb{E}X_\infty.$$ □

De Finetti's Theorem

Let $\Omega = S^{\mathbb{N}}$ and $\mathcal{F} = \mathcal{S}^{\mathbb{N}}$; let $X_1, X_2, \ldots$ be random variables taking values in $(S, \mathcal{S})$ such that $X_n(\omega) = \omega_n$. [Here, $\omega = (\omega_1, \omega_2, \ldots) \in S^{\mathbb{N}}$.]

Definition 2.10.8. The random variables $(X_n)_{n\geq 1}$ is *exchangeable* if for all $n \in \mathbb{N}$ and all permutations $\pi \in \mathfrak{S}_n$,

$$(X_1, \ldots, X_n) \stackrel{\mathrm{d}}{=} (X_{\pi(1)}, \ldots, X_{\pi(n)}).$$ △

For example, if $X_1, X_2, \ldots$ are i.i.d. then they are exchangeable.

Example 2.10.9. Mint a coin that comes up heads with probability $U \sim \text{Unif}(0,1)$, then flip it:

$$\mathbb{P}(X_n = 1|U) = U \quad \text{and} \quad \mathbb{P}(X_n = 0|U) = 1 - U \quad \text{for all } n.$$

Assume flips are independent conditioned on U, so that

$$\mathbb{P}(X_1 = b_1, \ldots, X_n = b_n|U) = \mathbb{P}(X_1 = b_1|U) \ldots \mathbb{P}(X_n = b_n|U) = U^m(1-U)^{n-m},$$

where $m = b_1 + \cdots + b_n$. These X_n are exchangeable but not independent. △

Definition 2.10.10. The event $A \subseteq \Omega = S^{\mathbb{N}}$ is *exchangeable* if $\pi A = A$ for all $n \in \mathbb{N}$ and all $\pi \in \mathfrak{S}_n$. △

Define

$$\mathcal{E} \stackrel{\text{def}}{=} \{\text{exchangeable events } A\}.$$

Note that

$$\mathcal{E} \supsetneq \mathcal{T} = \bigcap_{n \geq 0} \sigma(X_{n+1}, X_{n+2}, \ldots).$$

As an example,

Example 2.10.11. We have $\{X_n > 0 \text{ for all } n\} \in \mathcal{M}$, but $\{X_n > 0 \text{ for all } n\} \notin \mathcal{T}$. △

Theorem 2.10.12 (Hewitt-Savage 0-1 law)**.** *If $(X_n)_{n\geq 0}$ are i.i.d., then $\mathbb{P}(A) \in \{0,1\}$ for all $A \in \mathcal{E}$.*

Example 2.10.13. Consider the coin example from Example 2.10.9. We have $U \in m\mathcal{E}$. The strong law of large numbers says

$$\frac{X_1 + \cdots + X_n}{n} \to U \quad \text{a.s..}$$

△

Theorem 2.10.14 (De Finetti)**.** *If $(X_n)_{n\geq 0}$ is exchangeable then conditioned on $\mathcal{E}$, the random variables $(X_n)_{n\geq 1}$ are i.i.d..*

3 Brownian Motion

3.11 Feb 27, 2020

Sources include Durrett Ch. 7, as well as Mörters-Peres *Brownian Motion* and Baudoin *Diffusion processes and stochastic calculus*.

Definition 3.11.1. A *continuous time stochastic process* is a collection of random variables $(X_t)_{t\in[0,\infty)}$ indexed by $t \in [0,\infty)$, all defined on the same probability space $(\Omega, \mathcal{F} < \mathbb{P})$. △

Definition 3.11.2. The continuous time stochastic process $(X_t)_{t\geq 0}$ has *independent increments* if for all $t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent. △

Definition 3.11.3. The random variables $(B_t)_{t\geq 0}$ is a *Brownian motion* if

1. It has independent increments,
2. $B_{s+t} - B_s \sim N(0,t)$ for all $s,t > 0$,
3. $t \mapsto B_t$ is almost surely continuous.

If $B_0 = 0$, then we additionally say that $(B_t)_{t\geq 0}$ is a *standard Brownian motion*. △

Thus there are two views of $B(t,w) \colon [0,\infty) \times \Omega \to \mathbb{R}$. One view is through the map

$$t \mapsto (\omega \mapsto B(t,w)),$$

i.e. for each t we have a random variable $\omega \mapsto B(t,\omega)$. Another view is through the map

$$\omega \mapsto (t \mapsto B(t,\omega)),$$

i.e. for each $\omega \in \Omega$ we have a continuous function $[0,\infty) \to \mathbb{R}$. (Think of these functions as a squiggly graph; for each $\omega \in \Omega$ we have a bunch of squiggly graphs.)

In other words, the second viewpoint is to think of B as a function

$$f \colon \Omega \to C[0,\infty) = \{\text{continuous functions } [0,\infty) \to \mathbb{R}\}.$$

With this viewpoint it is natural to endow $C[0,\infty)$ with a σ-algebra. The natural one to endow it with is

$$\sigma\{A_{t,y} \colon t \in [0,\infty), y \in \mathbb{R}\}, \qquad \text{where} \qquad A_{t,y} = \{f \in C[0,\infty) \colon f(t) \geq y\}. \tag{14}$$

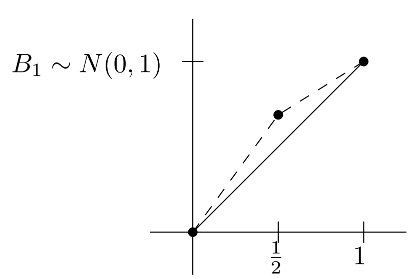
Let's pause to consider existence of brownian motions. The condition $B_{s+t} - B_s \sim N(0,t)$ is crucial, and it turns out that in general, if we replace $N(0,t)$ with other random variables they might not exist.

Lévy's construction of Brownian motion.

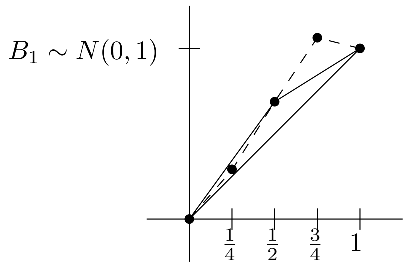
The idea is to first construct standard Brownian motions $B(d)$ for d a dyadic rational,

$$d = \left\{\frac{k}{2^n} \colon 0 \leq k \leq 2^n\right\}.$$

We know $B_0 = 0$ and $B_1 \sim N(0,1)$. We begin by linearly interpolating B_0 and B_1, and then we perturb B at $\frac{1}{2}$ (in some way to be made precise later), as below:



Then we perturb B at $\frac{1}{4}$ and $\frac{3}{4}$, as below:



To perturb it, we use a lemma (which was [6710, HW 9, Ex 1])

Lemma 3.11.4. *If $(X,Y) \sim N(0, \sigma^2 I_2)$ then $(X+Y, X-Y) \sim N(0, 2\sigma^2 I_2)$.*

Roughly, the lemma follows from rotational invariance of the normal distribution.
Given $d \in D_n \setminus D_{n-1}$, if we've already constructed $B(d')$ for each $d' \in D_{n-1}$, we set

$$B(d) = \frac{B(d+2^{-n}) + B(d-2^{-n})}{2} + \frac{Z_d}{2^{(n+1)/2}},$$

where $(Z_d)_{d \in \cup_n D_n}$ are i.i.d. $N(0,1)$ random variables. Then note that if

$$X = \frac{B(d+2^{-n}) + B(d-2^{-n})}{2} \sim N\left(0, \frac{2 \cdot 2^{-n}}{4}\right)$$

and

$$Y = \frac{Z_d}{2^{(n+1)/2}} \sim N\left(0, \frac{1}{2^{n+1}}\right)$$

then Lemma 3.11.4 says

$$X + Y = B(d+2^{-n}) - B(d) \quad \text{ and } X - Y = B(d) - B(d-2^{-n})$$

are independent $N(0, 2 \cdot \frac{1}{2^{n+1}})$ random variables.

We need to show that the limit is continuous (item 3 in Definition 3.11.3). Let

$$F_n(t) = \begin{cases} 2^{-(n+1)/2} Z_t, & t \in D_n \setminus D_{n-1} \\ 0 & t \in D_{n-1} \\ \text{linear} & \text{otherwise} \end{cases}$$

and define

$$B(t) = \sum_{n \geq 0} F_n(t).$$

Note that

$$B(d) = \sum_{k=0}^{n} F_k(d)$$

is a finite sum for each $d \in D_n$.

Claim 3.11.5. *The series defining $B(t)$ is a.s. uniformly convergent.*

Because the uniform limit of continuous functions is continuous, we'll get $\mathbb{P}(t \mapsto B(t) \text{ is continuous}) = 1$.

Proof of Claim 3.11.5. We have

$$\mathbb{P}(|Z_d| \geq c\sqrt{n}) < e^{-(c\sqrt{n})^2/2}$$

so

$$\sum_{d \in D_n} \mathbb{P}(|Z_d| \geq c\sqrt{n}) < (2^n + 1) e^{-c^2 n/2} < \lambda^n$$

for some $\lambda < 1$. On the other hand

$$\mathbb{P}(\max_{d\in D_n} |Z_n| \geq c\sqrt{n}) \leq \sum_{d\in D_n} \mathbb{P}(|Z_d| \geq c\sqrt{n}) < \lambda^n,$$

so summing over n we obtain

$$\sum_{n\geq 0} \mathbb{P}(\max_{d\in D_n} |Z_n| \geq c\sqrt{n}) < \sum_{n\geq 0} \lambda^n < \infty.$$

Borel-Cantelli now says

$$\mathbb{P}(\max_{d\in D_n} |Z_n| \geq c\sqrt{n} \text{ i.o.}) = 0.$$

Thus with probability 1, there exists $N < \infty$ so that for all $n > N$ the estimate $\|F_n\|_\infty \leq c\sqrt{n}2^{-n/2}$ holds. □

So far we've defined Brownian motion on $[0,1]$. To extend from $[0,1]$ to $[0,\infty)$, we let $B_1, B_2, \ldots$ be independent copies of $B(t)_{t\in[0,1]}$. For $t \in [0,\infty)$, write $t = n + u$ for $n \in \mathbb{N}$ and $u \in [0,1)$. Set

$$B(t) = \sum_{i=1}^{n} B_i(1) + B_{n+1}(u).$$

After extending from the dyadics to all of $\mathbb{R}$ by continuity, it's not so bad to verify that $B(s+t) - B(s) \sim N(0,t)$. This completes the construction of a Brownian motion.

Definition 3.11.6. A stochastic process $(X_t)_{t\geq 0}$ is called a *Gaussian* if for every $t_0 < \cdots < t_n$, the random vector $(X_{t_0}, \ldots, X_{t_n})$ is Gaussian. △

(We say the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is a multivariate distribution, so $\mathbf{X} \sim N(\mu, \Sigma)$.)

Definition 3.11.7. We say a Gaussian is *centered* if its mean is zero. △

Brownian motion $(B_t)_{t\geq 0}$ is a centered Gaussian process: the joint distribution of $B_{t_0}, \ldots, B_{t_n}$ is

$$\begin{bmatrix} B_{t_0} \\ \vdots \\ B_{t_n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} B_{t_0} \\ B_{t_1} - B_{t_0} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{bmatrix}$$

and the vector $(B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}})$ consists of independent normals.

Definition 3.11.8. We say $(X_t)_{t\geq 0} \stackrel{\text{fdd}}{=} (Y_t)_{t\geq 0}$ if X and Y have the same finite-dimensional distributions, i.e. for all $t_0 < \cdots < t_n$ we have

$$(X_{t_0}, \ldots, X_{t_n}) \stackrel{\text{d}}{=} (Y_{t_0}, \ldots, Y_{t_n}).$$

△

To specify the finite dimensional distributions of a centered Gaussian process $(X_t)_{t\geq 0}$ it suffices to specify $\mathbb{E}(X_s X_t)$ for $s \leq t$.

Remark 3.11.9. Iff $\mathbf{X}, \mathbf{Y}$ are a.s. continuous random paths $\Omega \to C[0,\infty)$ and $\mathbf{X} \stackrel{\text{fdd}}{=} \mathbf{Y}$, then $\mathbf{X} \stackrel{\text{d}}{=} \mathbf{Y}$, i.e. $\mathbb{P}(\mathbf{X} \in A) = \mathbb{P}(\mathbf{Y} \in A)$ for all $A \in \mathcal{F}' = \sigma(A_{t,y})$ (see Equation (14)). △

Proof. This is because $\mathbf{X} \stackrel{\text{fdd}}{=} \mathbf{Y}$ implies $(X_t)_{t\in\mathbb{Q}} \stackrel{\text{d}}{=} (Y_t)_{t\in\mathbb{Q}}$, which in turn implies $\mathbf{X} \stackrel{\text{d}}{=} \mathbf{Y}$ by continuity. □

Note also that if $(B_t)_{t\geq 0}$ is a Brownian motion and $s \leq t$, then

$$\mathbb{E}(B_s B_t) = \mathbb{E}(B_s(B_s + (B_t - B_s))) = \mathbb{E}B_s^2 + \mathbb{E}(B_s(B_t - B_s)) = s + 0, \tag{15}$$

since $B_s \sim N(0,s)$ and B_s is independent of $B_t - B_s$. This leads to another equivalent definition of Brownian motion, which we'll explore next time.

3.12 Mar 3, 2020

[I was out of town for this lecture. I'm grateful for Will Gao's notes, from which I copied. I'm sorry it's late!]

[For [HW 4, Ex 1(e)]: change "2" to "3", i.e. prove $\mathbb{P}(X_n \geq \sqrt{3n \log n} \text{ i.o.}) = 0$.]

Claim 3.12.1. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Let $X(t) = \frac{1}{a}B(a^2 t)$ for fixed $a > 0$. Then $(X(t))_{t\geq 0}$ is a standard Brownian motion.*

In this sense, Brownian motion is a random fractal (when space/time are rescaled correctly).

Proof. Note that

$$X(t) - X(s) = \frac{1}{a}B(a^2 t) - \frac{1}{a}B(a^2 s) \sim \frac{1}{a}N(0, a^2(s-t)) \sim N(0, t-s).$$

Also, X has independent increments (since B does) and continuous sample path (since B does). Finally, $X(0) = \frac{1}{a}B(0) = 0$. □

Time inversion of Brownian motion

We also have

Claim 3.12.2. *Let $(B(t))_{t\geq 0}$ be a standard Brownian motion. Let*

$$Y(t) = \begin{cases} tB(\frac{1}{t}) & t > 0 \\ 0 & t = 0. \end{cases}$$

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

Proof. Let's check that $Y(t)$ is a Gaussian process with the same covariance as $B(t)$. Note that

$$(Y(t_0), \dots, Y(t_n)) = (t_0 B(1/t_0), \dots, t_n B(1/t_n))$$

has a multivariate normal distribution. Its covariance can be computed as follows: For $t, h > 0$ we have

$$\mathbb{E}(Y(t+h)Y(t)) = \mathbb{E}\left((t+h)B\left(\frac{1}{t+h}\right)tB\left(\frac{1}{t}\right)\right) = t = \mathbb{E}(B(t+h)B(t)),$$

by Equation (15). Hence, $(Y(t))_{t\geq 0} \overset{\text{fdd}}{=} (B(t))_{t\geq 0}$. It remains to check that $t \mapsto Y(t)$ is a.s. continuous. For $t > 0$, this follows since $t \mapsto B(t)$ is a.s. continuous.

Finally, because $(B(t))_{t\in\mathbb{Q}} \overset{\text{d}}{=} (Y(t))_{t\in\mathbb{Q}}$, we have

$$\lim_{\substack{t\in\mathbb{Q},\\ t\downarrow 0}} Y(t) = \lim_{\substack{t\in\mathbb{Q},\\ t\downarrow 0}} B(t) = B(0) = 0.$$

Thus $Y(t)$ is a standard Brownian motion, as claimed. □

Notice we have proven that

$$\lim_{s\uparrow\infty} \frac{1}{s}B(s) = \lim_{t\downarrow 0} tB(1/t) = \lim_{t\downarrow 0} Y(t) = 0.$$

Here's another proof of this fact. Let $n = \lfloor s \rfloor$. Then

$$B(s) = (B(1) - B(0)) + (B(2) - B(1)) + \cdots + (B(n) - B(n-1)) + (B(s) - B(n)) = \underbrace{Z_1 + \cdots + Z_n}_{\text{i.i.d. } N(0,1)} + Z',$$

where Z' is some independent $N(0, s-n)$. The strong law of large numbers says

$$\frac{Z_1 + \cdots + Z_n}{n} \to 0 \quad \text{a.s.},$$

so the claim follows by observing that $Z'/n \to 0$.

Note that we may have instead divided by $\sqrt{s}$; here $B(s)/\sqrt{s} \sim N(0,1)$ for all s, so $B(s)/\sqrt{s} \not\to 0$. In fact:

Lemma 3.12.3. *We have*

$$\limsup_{s\to\infty} \frac{B(s)}{\sqrt{s}} = +\infty \quad \text{a.s.,}$$
$$\liminf_{s\to-\infty} \frac{B(s)}{\sqrt{s}} = -\infty \quad \text{a.s..}$$

Proof. Let $X_n = B(n) - B(n-1) \sim N(0,1)$, so the X_n are independent. Note also that $B(n)/\sqrt{n} \sim N(0,1)$. Then,

$$\begin{aligned}
\mathbb{P}\left(\frac{B(n)}{\sqrt{n}} > c \text{ i.o.}\right) &= \mathbb{E}\mathbb{1}_{\{\frac{B(n)}{\sqrt{n}} > c \text{ i.o.}\}} \\
&= \mathbb{E}\mathbb{1}_{\cap_{N\geq 1}\cup_{n\geq N}\{\frac{B(n)}{\sqrt{n}} > c\}} \\
&= \mathbb{E}\limsup_{n\to\infty} \mathbb{1}_{\{\frac{B(n)}{\sqrt{n}} > c\}} \\
&\geq \limsup_{n\to\infty} \mathbb{E}\mathbb{1}_{\{\frac{B(n)}{\sqrt{n}} > c\}} \\
&= \limsup_{n\to\infty} \mathbb{P}\left(\frac{B(n)}{\sqrt{n}} > c\right) \\
&> 0.
\end{aligned}$$

We may now apply Theorem 2.10.12. The event

$$A = \left\{\frac{B(n)}{\sqrt{n}} > c \text{ i.o.}\right\}$$

is exchangeable under finite permutations of $X_1, X_2, \ldots$. Then

$$\mathbb{P}(A) \in \{0,1\}, \qquad \text{hence} \qquad \mathbb{P}(A) = 1.$$ □

A sharp result is given by the *law of the iterated logarithm* [cf. [HW 4, Ex 1]]:

$$\limsup_{s\to\infty} \frac{B(s)}{\sqrt{2s\log\log s}} = 1.$$

Nondifferentiability of Brownian motion

Definition 3.12.4. We define the *upper-right derivative* and *lower-right derivative* of a function f by

$$D^* f(t) \overset{\text{def}}{=} \limsup_{h\downarrow 0} \frac{f(t+h) - f(t)}{h}$$
$$D_* f(t) \overset{\text{def}}{=} \liminf_{h\downarrow 0} \frac{f(t+h) - f(t)}{h}$$

respectively. △

Theorem 3.12.5. *Fix $t \geq 0$. Then,*

$$\mathbb{P}(D^* B(t) = +\infty) = \mathbb{P}(D_* B(t) = -\infty) = 1.$$

Hence $\mathbb{P}(\{\omega\colon x \mapsto B(x) \text{ is differentiable at } x = t\}) = 0$.

Proof. Let $X(s) = B(t+s) - B(t)$. Then $(X(s))_{s\geq 0}$ is a standard Brownian motion. (This is not difficult to see.)

Now note that $(D^* B)(t) = (D^* X)(0)$, so we may assume $t = 0$.

Consider the standard Brownian motion

$$Y(s) = \begin{cases} sX(1/s) & s > 0 \\ 0 & s = 0 \end{cases}$$

(Claim 3.12.2 asserts Y is indeed a standard Brownian motion.)

Then

$$(D^*X)(0) = \limsup_{h\downarrow 0} \frac{X(h) - X(0)}{h} = \limsup_{n\to\infty} \frac{X(1/n)}{1/n} = \limsup_{n\to\infty} Y_n = \infty.$$

A similar argument holds for D_*. □

Remark 3.12.6. Actually, $t \mapsto B(t)$ is a.s. Hölder continuous for $c < 1/2$ and a.s. not Hölder continuous for $c \geq 1/2$. △

Theorem 3.12.7 (Paley-Wiener-Zygmund)**.** *We have*

$$\mathbb{P}(D^*B(t) - D_*B(t) = +\infty \text{ for all } t \in [0,1]) = 1.$$

In particular

$$\mathbb{P}(\{\omega\colon t \mapsto B(t) \text{ is differentiable for some } t = t_0 \in [0,1]\}) = 0.$$

This is an uncountable union over $t_0 \in [0,1]$, so it's a strengthening of Theorem 3.12.5.

Example 3.12.8. Let $Z_t = \{s > t\colon B(t) = B(s)\}$ and

$$A_t = \{t \in \overline{Z_t}\} = \{\text{there are } s_i \downarrow t \text{ with } B(s_n) = B(t) \text{ for all } n\}.$$

On Thursday we'll show that

$$\mathbb{P}(A_t) = 1 \quad \text{but} \quad \mathbb{P}\left(\bigcap_{t\in[0,1]} A_t\right) = 0.$$

For now let

$$L = \sup\{t < 1\colon B(t) = 1\}$$

and note that

$$\mathbb{P}(L < 1) \geq \mathbb{P}(B(1) \neq 1) = 1.$$

But $Z_L \sim (L, 1) \neq 0$. If $L < 1$, then A_L does not occur. [i'm very lost, sorry] △

3.13 Mar 5, 2020

Let's talk today about the Markov property of Brownian motion. (Durrett spends some time on Markov chains, but we skipped that for now.)

Suppose $(B(t))_{t\geq 0}$ is a Brownian motion. Then for all $s \geq 0$, we may consider the process

$$(B(t+s) - B(s))_{t\geq 0},$$

which is a standard Brownian motion that is independent of $(B(u))_{0\leq u\leq s}$.

Definition 3.13.1. A (continuous time) *filtration* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a family of σ-algebras $(\mathcal{F}(t))_{t\geq 0}$ with $\mathcal{F}(t) \subseteq \mathcal{F}$ and $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ for $s < t$.

A (continuous time) stochastic process $(X(t))_{t\geq 0}$ is *adapted* to $(\mathcal{F}(t))_{t\geq 0}$ if $X(t) \in m\mathcal{F}(t)$ for all $t \geq 0$. △

Example 3.13.2. A standard brownian motion $(B(t))_{t\geq 0}$ is adapted to $(\mathcal{F}^0(t))_{t\geq 0}$, where

$$\mathcal{F}^0(t) = \sigma(B(s) \colon 0 \leq s \leq t).$$

But we could take a slightly larger filtration: specifically, $(B(t))_{t\geq 0}$ is also adapted to $(\mathcal{F}^+(t))_{t\geq 0}$, where

$$\mathcal{F}^+(t) = \bigcap_{s>t} \mathcal{F}^0(s).$$

(Here, $\mathcal{F}^+$ can be thought of as looking infinitesimally into the future; it turns out that $\mathcal{F}^+(t) \supsetneq \mathcal{F}^0(t)$.)

Let's think about how much bigger $\mathcal{F}^+$ is. △

Theorem 3.13.3. *Let $X(t) = B(t+s) - B(s)$. Then $(X(t))_{t\geq 0}$ is independent of $\mathcal{F}^+(s)$.*

(The Markov property is the same statement but with $\mathcal{F}^0(s)$, so this is a strengthening.)

Proof. We're going to use continuity of B. Let $s_n = s + \frac{1}{n}$. Then

$$X(t) = \lim_{n\to\infty} (B(s_n + t) - B(s_n)) \quad \text{a.s.}$$

which exists since $t \mapsto B(t)$ is a.s. continuous. Since X is continuous, it is enough to check that the finite dimensional distributions of X are independent of $\mathcal{F}^+(s)$. But now

$$(X(t_1), \ldots, X(t_m)) = \lim_{n\to\infty} \underbrace{(B(s_n + t_1) - B(s_n), \ldots, B(s_n + t_m) - B(s_n))}_{\text{independent of } \mathcal{F}^+(s)}. \qquad \square$$

Anytime one proves an "independence from yourself" result, one gets a 0-1 law.

Corollary 3.13.4 (Blumenthal's 0-1 law)**.** *Let $(B(t))_{t\geq 0}$ be a Brownian motion with $B(0) = x$. Then $\mathbb{P}(A) \in \{0, 1\}$ for $A \in \mathcal{F}^+(0)$.*

Example 3.13.5. We have

$$A = \{\omega \colon t \mapsto B(t) \text{ is differentiable at } t = 0\} \in \mathcal{F}^+(0).$$

Last time we showed $\mathbb{P}(A) = 0$. △

Proof of Corollary 3.13.4. Take $s = 0$ in Theorem 3.13.3. Then $(B(t) - B(0))_{t\geq 0}$ is independent of $\mathcal{F}^+(0)$. It follows that $(B(t))_{t\geq 0}$ is independent of $\mathcal{F}^+(0)$.

Any event $A \in \mathcal{F}^+(0)$ satisfies $A \in \sigma(B(t))_{t\geq 0}$. It follows that A is independent of itself, and hence $\mathbb{P}(A)^2 = \mathbb{P}(A \cap A) = \mathbb{P}(A)$. $\square$

Corollary 3.13.6. *For any tail event*

$$A \in \mathcal{T} = \bigcap_{t\geq 0} \sigma(B(s) \colon s \geq t)$$

we have $\mathbb{P}(A) \in \{0, 1\}$.

Proof. The inversion

$$Y(t) = \begin{cases} tB(1/t) & \text{if } t > 0 \\ 0 & t = 0 \end{cases}$$

is a standard Brownian motion (Claim 3.12.2). Also,

$$\mathcal{T}_B = \mathcal{F}_Y^+(0).$$

□

Zero set of Brownian motion
Let B be a standard Brownian motion and let

$$\tau \stackrel{\text{def}}{=} \inf\{t > 0 \colon B(t) > 0\}$$
$$\sigma \stackrel{\text{def}}{=} \inf\{t > 0 \colon B(t) = 0\}.$$

One might expect $\sigma = \tau$, since otherwise $B(t)$ would obtain a local max at $t = \sigma$. In fact, more is true:

Theorem 3.13.7. *We have* $\mathbb{P}(\sigma = 0) = \mathbb{P}(\tau = 0) = 1$.

Proof. Observe that

$$\{\tau = 0\} = \bigcap_{n \geq 1} \{B(\varepsilon) > 0 \text{ for some } 0 < \varepsilon < 1/n\}.$$

In particular, $\{\tau = 0\} \in \mathcal{F}^+(0)$. Now fix $t > 0$ and observe that

$$\mathbb{P}(\tau \leq t) \geq \mathbb{P}(B(t) > 0) = \frac{1}{2},$$

since $B(t) \sim N(0, t)$. We conclude

$$\mathbb{P}(\tau = 0) = \lim_{t \downarrow 0} \mathbb{P}(\tau \leq t) \geq \frac{1}{2}.$$

Since $\{\tau = 0\} \in \mathcal{F}^+(0)$, Blumenthal's 0-1 law (Corollary 3.13.4) says $\mathbb{P}(\tau = 0) \in \{0, 1\}$. We conclude $\mathbb{P}(\tau = 0) = 1$.

Now let

$$\tau' \stackrel{\text{def}}{=} \inf\{t > 0 \colon B(t) < 0\}$$

and observe that $\mathbb{P}(\tau' = 0)$, either by repeating the proof or by using the scale invariance

$$\left(\frac{1}{a} B(a^2 t)\right)_{t \geq 0} \stackrel{\text{d}}{=} (B(t))_{t \geq 0}$$

for $a = -1$ (cf. Claim 3.12.1).

The intermediate theorem says $\sigma \leq \max(\tau, \tau')$ and the maximum is zero almost surely. It follows that $\mathbb{P}(\sigma = 0) = 1$. □

Example 3.13.8. Fix t and let

$$Z_t = \{s > t \colon B(t) = B(s)\},$$

so that Z_t is the zero set of the process

$$X(s) = (B(t + s) - B(t))_{s \geq 0}.$$

Let

$$A_t = \{t \in \overline{Z}_t\}.$$

We claimed last time that

$$\mathbb{P}(A_t) = 1 \quad \text{but} \quad \mathbb{P}\left(\bigcap_{t \in [0,1]} A_t\right) = 0.$$

(So for any fixed time, A_t almost surely happens, but there's almost surely a time when it fails.)

The claim that $\mathbb{P}(A_t) = 1$ since $\mathbb{P}(\sigma = 0) = 1$ for the standard brownian motion $X(s)$. Now let

$$L = \sup\{t < 1 \colon B(t) = 0\}.$$

Then note that $\mathbb{P}(L = 1) = \mathbb{P}(B(1) = 0) = 0$, since $B(1) \sim N(0, 1)$. On the other hand, $\mathbb{P}(A_L) = 0$. △

Right continuous filtrations

Definition 3.13.9. The filtration $(\mathcal{F}(t))_{t\geq 0}$ is *right-continuous* if

$$\mathcal{F}(t) = \bigcap_{\varepsilon \downarrow 0} \mathcal{F}(t+\varepsilon)$$

for all $t \geq 0$. △

For example, the filtrations (in the notation of Example 3.13.2) $\mathcal{F}^0$ is not right continuous but $\mathcal{F}^+$ is.

Lemma 3.13.10. *If $(\mathcal{F}(t))_{t\geq 0}$ is right-continuous and T is a random time such that $\{T < t\} \in \mathcal{F}(t)$ for all t, then T is a* stopping time *for $(\mathcal{F}(t))_{t\geq 0}$, i.e. $\{T \leq t\} \in \mathcal{F}(t)$ and $\{T = t\} \in \mathcal{F}(t)$ for all t.*

Proof. It's enough to prove $\{T \leq t\} \in \mathcal{F}(t)$, since $\{T = t\} = \{T \leq t\} \setminus \{T < t\}$.

Observe that

$$\{T \leq t\} = \bigcap_{n\geq 1} \left\{T < t + \frac{1}{n}\right\} \in \bigcap_{n\geq 1} \mathcal{F}\left(t + \frac{1}{n}\right) = \mathcal{F}(t)$$

with the last equality from right-continuity. □

Strong Markov Property of Brownian motion

Theorem 3.13.11. *Let B be a standard Brownian motion and let T be an $\mathcal{F}^+$-stopping time with $\mathbb{P}(T < \infty) = 1$. Define*

$$X(t) \stackrel{\text{def}}{=} B(T+t) - B(T).$$

Then $(X(t))_{t\geq 0}$ is a standard Brownian motion that is independent of $\mathcal{F}^+(T)$.

Recall that $\mathcal{F}^+(T)$ is the σ-algebra consisting of those sets A with $A \cap \{T = t\} \in \mathcal{F}^+(t)$ for all t.

Theorem 3.13.11 is usually false when T is not a stopping time:

Example 3.13.12.

1. Let $T = \sup\{t < 1 \colon B(t) = 0\}$.
2. Let $T = \text{argmax}\{B(t) \colon t \in [0, 1]\}$.

△

3.14 Mar 10, 2020

Last time we defined the σ-algebras

$$\mathcal{F}^0(u) = \sigma(B(s)\colon 0 \le s \le u)$$
$$\mathcal{F}^+(t) = \bigcap_{u>t} \mathcal{F}^0(u).$$

We asserted that

$$A = \{t \mapsto B(t) \text{ is differentiable at } 0\} \in \mathcal{F}^+(0) \setminus \mathcal{F}^0(0).$$

We may see this in HW 5.

We also discussed the Strong Markov Property of Brownian motion: if T is a stopping time (so $\{T \le t\} \in \mathcal{F}^+(t)$ for all t), then we define $\mathcal{F}^+(T) = \{A\colon A \cap \{T \le t\} \in \mathcal{F}^+(t)\}$. Then

Theorem 3.14.1 (cf. Theorem 3.13.11)**.** *If $\mathbb{P}(T < \infty) = 1$, then $(B(T+t) - B(t))_{t \ge 0}$ is a standard Brownian motion independent of $\mathcal{F}^+(T)$.*

Proof. We proved this for a fixed time $T = t_0$. To prove it in general, let $T_n = (k+1)2^{-n}$, where $k \in \mathbb{N}$ and $k2^{-n} \le T < (k+1)2^{-n}$.

Note that T_n is a stopping time, because

$$\{T_n \le t\} = \bigcup_{k \in \mathbb{N}} \{T_n \le t, k2^{-n} \le T < (k+1)2^{-n}\}.$$

We have $T_n \downarrow T$. Define the standard brownian motion

$$B_k(t) = B(t + k2^{-n}) - B(k2^{-n});$$

note that this is independent of $\mathcal{F}^+(k2^{-n})$ (this is the ordinary Markov property). Let's define

$$B_*(t) = B(t + T_n) - B(T_n).$$

Let $\mathcal{E} \in \mathcal{F}^+(T_n)$. We want to show

$$\mathbb{P}(\{B_* \in A\} \cap \mathcal{E}) = \mathbb{P}(\{B_* \in A\})\mathbb{P}(\mathcal{E}).$$

On the other hand,

$$\begin{aligned}
\mathbb{P}(\{B_* \in A\} \cap \mathcal{E}) &= \sum_{k \ge 0} \mathbb{P}(\{B_k \in A\} \cap \underbrace{\mathcal{E} \cap \{T_n = k2^{-n}\}}_{\in \mathcal{F}^+(k2^{-n})}) \\
&= \sum_{k \ge 0} \mathbb{P}(B_k \in A)\mathbb{P}(\mathcal{E} \cap \{T_n = k2^{-n}\}) \\
&= \sum_{k \ge 0} \mathbb{P}(B_0 \in A)\mathbb{P}(\mathcal{E} \cap \{T_n = k2^{-n}\}) \\
&= \mathbb{P}(B_0 \in A)\mathbb{P}(\mathcal{E}).
\end{aligned}$$

(In the second equality we used that $B_k \overset{\mathrm{d}}{=} B_0$.)

We've shown that for each n, $(B(t+T_n) - B(T_n))_{t \ge 0}$ is a standard Brownian motion independent of $\mathcal{F}^+(T_n) \supseteq \mathcal{F}^+(T)$. Then

$$B(t+T) - B(T) = \lim_{n \to \infty} (B(t + T_n) - B(T_n))$$

is independent of $\mathcal{F}^+(T)$. Furthermore, B is a Brownian motion because it is a pointwise limit of Brownian motions. $\square$

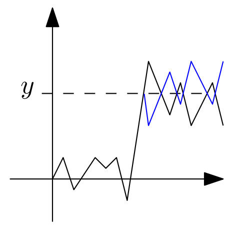
The Reflection Principle

Theorem 3.14.2. *If $(B(t))_{t\geq 0}$ is a standard Brownian motion and T is a stopping time, then*

$$B^*(t) = B(t)\mathbb{1}_{\{t\leq T\}} + (2B(T) - B(t))\mathbb{1}_{\{T>t\}}$$

is also a standard Brownian motion.

A good example is the stopping time given by $T = \{\inf t\colon B(t) > y\}$. Then the graph of $B^*(t)$ is obtained from that of $B(t)$ reflected along the horizontal axis $B = y$ starting at time T:



Proof of Theorem 3.14.2. By the strong Markov property, $(g_1(t))_{t\geq 0} \overset{\text{def}}{=} (B(t+T) - B(T))_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}^+(T)$. But so is $(g_2(t))_{t\geq 0} \overset{\text{def}}{=} (-B(t+T) + B(T))_{t\geq 0}$.

For $f, g \in \Omega = C[0,\infty)$ with $g(0) = 0$ and $T\colon \Omega \to [0,\infty)$, define $\mathrm{Glue}(f,g)\colon [0,\infty) \to \Omega$ by

$$\mathrm{Glue}(f,g)(t) = \begin{cases} f(t) & t \leq T = T(f) \\ f(T) + g(t-T) & t > T \end{cases}.$$

Now $B = \mathrm{Glue}(B, g_1)$ and $B^* = \mathrm{Glue}(B, g_2)$, and $B \overset{\text{d}}{=} B^*$. □

Corollary 3.14.3. *Let $T_y = \min\{t\colon B(t) = y\}$. Then $\mathbb{P}(T_y \leq t) = 2\mathbb{P}(B(t) \geq y)$.*

Proof. We have

$$\{T_y \leq t\} = \{B(t) \geq y\} \sqcup \{M(t) \geq y, B(t) < y\},$$

where

$$M(t) = \max_{0\leq s\leq t} B(s).$$

Then

$$\{T_y \leq t\} = \{B(t) \geq y\} \sqcup \{B^*(t) > y\} \implies \mathbb{P}(T_y \leq t) = \mathbb{P}(B(t) \geq y) + \mathbb{P}(B^*(t) > y) = 2\mathbb{P}(B(t) \geq y),$$

where the last equality is the reflection principle (Theorem 3.14.2). □

Theorem 3.14.4 (Lévy's Theorem). *Let $Y(t) = M(t) - B(t)$. Then $(Y(t))_{t\geq 0} \overset{\text{d}}{=} (|B(t)|)_{t\geq 0}$.*

Proof idea. We want to show that Y and $|B|$ have the same finite dimensional distributions, and use the fact that both Y and $|B|$ are both a.s. continuous.

Note that

$$\mathbb{P}(|B(t)| \geq y) = \mathbb{P}(\{B(t) \geq y\} \sqcup \{B(t) \leq -y\}) = 2\mathbb{P}(B(t) \geq y) = \mathbb{P}(T_y \leq t) = \mathbb{P}(M(t) \geq y),$$

with the penultimate equality from Corollary 3.14.3. This shows that $M(t) \overset{\text{d}}{=} |B(t)|$. From here we'd try to show $Y \overset{\text{fdd}}{=} |B|$. □

Note that Brownian motion is a continuous time martingale, since for $s < t$ we have

$$\begin{aligned}\mathbb{E}(B(t)|\mathcal{F}^+(s)) &= \mathbb{E}(B(s) + (B(t) - B(s))|\mathcal{F}^+(s)) \\ &= B(s) + \mathbb{E}(B(t) - B(s)|\mathcal{F}^+(s)) \\ &= B(s) + \mathbb{E}(\underbrace{B(t) - B(s)}_{\sim N(0,t-s)}) \\ &= B(s).\end{aligned}$$

Let's develop optional stopping for continuous time martingales.

Theorem 3.14.5. *Let $(X(t))_{t\geq 0}$ be a martingale adapted to $(\mathcal{F}(t))_{t\geq 0}$, and let $S \leq T$ be $\mathcal{F}$-stopping times with $\mathbb{P}(T < \infty) = 1$. Assume also that*

$$\mathbb{P}(t \mapsto X(t) \text{ is continuous}) = 1$$

and furthermore that there exists Y with $\mathbb{E}|Y| < \infty$ and $|X(t \wedge T)| \leq Y$ a.s.. Then

$$\mathbb{E}(X(T)|\mathcal{F}(S)) = X(S).$$

We could prove this by adapting the proof of the discrete time case to the continuous time case; alternatively, we can take discrete approximations and take a limit. We'll adapt the second method here:

Proof of Theorem 3.14.5. Fix $k \in \mathbb{N}$ and let $X_n = X(T \wedge n2^{-k})$. Then $(X_n)_{n\geq 0}$ is a discrete time martingale adapted to

$$(\mathcal{F}'(n))_{n\in\mathbb{N}} \text{ where } \mathcal{F}'(n) = \mathcal{F}(n2^{-k}).$$

Let $S' = \lfloor 2^k S \rfloor + 1$ and let $T' = \lfloor 2^k T \rfloor + 1$. These are discrete stopping times, and the discrete optional stopping (Theorem 2.4.9) implies the desired claim. Details to follow on Thursday. □

3.15 Mar 12, 2020

[Classes are online now.]

Let $(B(t))_{t\geq 0}$ be a standard Brownian motion. Fix two numbers $-a, b \in \mathbb{R}$ and consider the stopping time $T = \inf\{t \geq 0 \colon B(t) \in \{-a, b\}\}$.

Theorem 3.15.1.

1. *$\mathbb{P}(B(T) = b) = \frac{a}{a+b}$, and*
2. *$\mathbb{E}T = ab$.*

Proof. As in the proof for the discrete case, we're going to use optional stopping (but the continuous version in Theorem 3.14.5). For part 1, we'll consider the continuous-time martingale $B(t)$; for part 2, we'll consider the continuous-time martingale $B(t)^2 - t$. The technical part of the proof will be to verify that we can use optional stopping. For now, let's apply it and see what happens:

For part 1, optional stopping (Theorem 3.14.5) says $\mathbb{E}B(T) = 0$. On the other hand, for $p = \mathbb{P}(B(T) = b)$

$$\mathbb{E}B(T) = pb + (1-p)(-a),$$

and solving this gives $p = \frac{a}{a+b}$.

For part 2, optional stopping (Theorem 3.14.5) says $\mathbb{E}(B(T)^2 - T) = \mathbb{E}M(T) = \mathbb{E}M(0) = 0$. This gives

$$\mathbb{E}T = \mathbb{E}B(T)^2 = b^2 p + (-a)^2(1-p),$$

where as before $p = \mathbb{P}(B(T) = b) = \frac{a}{a+b}$. Solving this gives $\mathbb{E}T = ab$.

So why can we use optional stopping? The answer is Lemma 3.15.3 below. □

Wald's Lemma for Brownian Motion

Let $(B(t))_{t\geq 0}$ be a standard brownian motion and let T be a stopping time. Then $B(t) \sim N(0,t)$ implies $\mathbb{E}B(t) = 0$. When is $\mathbb{E}B(T) = 0$? This is not always true, for example:

Example 3.15.2 (Enemy, cf. Examples 2.3.9, 2.4.3, 2.5.4)**.** Let $T = T_1 = \inf\{t > 0 \colon B(t) = 1\}$. Then $\mathbb{P}(T_1 = \infty) = 0$; this follows, for example, from Lemma 3.12.3, or from using scale invariance. Note that $\mathbb{E}B(T_1) = 1$ since $B(T_1) = 1$ a.s.. (This is a familiar counterexample to us, from the discrete time case!) △

Similar to the discrete time case, we have

Lemma 3.15.3 (Wald's Lemma for Brownian motion)**.** *If $\mathbb{E}T < \infty$ then $\mathbb{E}B(T) = 0$.*

Proof. We first show that $\mathbb{E}T < \infty$ implies the condition

$$\text{there exists } Y \in L^1 \text{ such that } B(t \wedge T) \leq Y \text{ for all } t \geq 0. \tag{16}$$

[Apr 8: Should this say $|B(t \wedge T)| \leq Y$?] and later that condition (16) implies $\mathbb{E}B(T) = 0$. Sometimes, even when $\mathbb{E}T = \infty$ we can still verify the more technical condition (16), so it's good to remember this.

Let

$$M_k = \max_{t \in [k, k+1]} |B(t) - B(k)|$$

and let

$$\begin{aligned} Y &= M_1 + \cdots + M_{\lceil T \rceil} \\ &= \sum_{k=1}^{\infty} \mathbb{1}_{\{T > k-1\}} M_k. \end{aligned}$$

Then

$$\begin{aligned} |B(t \wedge T)| &\leq |B(1) - B(0)| + |B(2) - B(1)| + \cdots + |B(t \wedge T) - B(\lfloor t \wedge T \rfloor)| \\ &\leq M_0 + M_1 + \cdots + M_{\lceil T \rceil} = Y. \end{aligned}$$

To complete the proof of condition (16) we should check that $\mathbb{E}Y < \infty$. To see this, note that

$$\mathbb{E}Y \underset{\text{MCT}}{=} \sum_{k=1}^{\infty} \mathbb{E}(\mathbb{1}_{\{T>k-1\}} M_k),$$

where $\mathbb{1}_{\{T>k-1\}} \in m\mathcal{F}^+(k+1)$ and $M_k \in \sigma(B(t) - B(k))_{t \geq k}$. So these random variables are independent, and

$$\mathbb{E}Y = \sum_k \mathbb{P}(T > k-1)\mathbb{E}M_k \leq (\mathbb{E}M_0)\mathbb{E}(T+1)$$

because $\mathbb{E}M_k = \mathbb{E}M_0$, and $\sum \mathbb{P}(T > k-1) \leq \mathbb{E}(T+1)$.

We showed last time ([Theorem 3.14.4, I think?]) that $\mathbb{E}M_0 = \mathbb{E}|B(1)| = \frac{2}{\pi} < \infty$.

That condition (16) implies $\mathbb{E}B(T) = 0$ follows from optional stopping, which says $\mathbb{E}B(T) = \mathbb{E}B(0) = 0$. □

3.16 Apr 7, 2020

[Zoom links are on Canvas! For the next several classes, we'll be following Mörters-Peres.]

Recall the setup for Wald's Lemma:

Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then $\mathbb{E}B_t = 0$ for all $t \in \mathbb{R}_{\geq 0}$, since $B_t \sim N(0,t)$. What about $\mathbb{E}B_T$ where T is a stopping time?

We saw in Example 3.15.2 that $T = \inf\{t\colon B_t \geq 1\}$ satisfies $\mathbb{P}(T < \infty) = 1$ and $B_T = 1$ a.s.. Thus $\mathbb{E}B_T = 1 \neq 0$.

Lemma 3.16.1 (Wald's Lemma 1, cf. Lemma 3.15.3)**.**

1. *If $\mathbb{E}T < \infty$ then $\mathbb{E}B_T = 0$.*
2. *If $B_{t\wedge T} \leq Y \in L^1$ for all $t \geq 0$, then $\mathbb{E}B_T = 0$.*

Corollary 3.16.2. *If $S \leq T$ are stopping times with $\mathbb{E}T < \infty$, then $\mathbb{E}(B(T) - B(S)|\mathcal{F}^+(S)) = 0$ a.s..*

Proof. By the strong Markov property (Theorem 3.13.11), $\widetilde{B}(t) \stackrel{\text{def}}{=} (B(t) - B(S))_{t\geq S}$ is a standard Brownian motion independent of $\mathcal{F}^+(S)$. Apply Wald's Lemma 1 (Lemma 3.16.1) to $\widetilde{B}$. ☐

Corollary 3.16.3. *If $S \leq T$ are stopping times with $\mathbb{E}T < \infty$, then $\mathbb{E}B(T)^2 = \mathbb{E}B(S)^2 + \mathbb{E}(B(T) - B(S))^2$*

Proof. Square both sides of the equality $B(T) = B(S) + (B(T) - B(S))$ to obtain

$$B(T)^2 = (B(S) + (B(T) - B(S)))^2.$$

Thus, we need to check that

$$\mathbb{E}\big[\mathbb{E}(B(S)(B(T) - B(S)))\big|\mathcal{F}^+(S)\big] = \mathbb{E}(B(S)(B(T) - B(S))) = 0.$$

Because $B(S)$ is $\mathcal{F}^+(S)$-measurable, we have

$$\mathbb{E}(B(S)(B(T) - B(S))|\mathcal{F}^+(S)) = B(S)\underbrace{\mathbb{E}(B(T) - B(S)|\mathcal{F}^+(S))}_{=0,\text{ by Corollary 3.16.2}} = 0.$$

It follows that

$$\mathbb{E}\big[\mathbb{E}(B(S)(B(T) - B(S)))\big|\mathcal{F}^+(S)\big] = 0.$$

☐

Let us build towards Wald Lemma 2. We begin with:

Lemma 3.16.4. *Let $Q(t) = B(t)^2 - t$. Then Q is a martingale.*

Proof. Indeed, for $s \leq t$, we have

$$\mathbb{E}(Q(t)|\mathcal{F}^+(s)) = \mathbb{E}\big[B(s)^2 + 2B(s)(B(t) - B(s)) + (B(t) - B(s))^2 - t|\mathcal{F}^+(s)\big]$$

We may estimate these terms separately: note that $\mathbb{E}(B(s)(B(t) - B(s))|\mathcal{F}^+(s)) = 0$ and that $B(t) - B(s) \sim N(0, t-s)$, hence $\mathbb{E}((B(t) - B(s))^2|\mathcal{F}^+(s)) = t - s$. It follows that

$$\mathbb{E}(Q(t)|\mathcal{F}^+(s)) = B(s)^2 + 0 + (t-s) - t = B(s)^2 - s = Q(s),$$

as desired. ☐

Lemma 3.16.5 (Wald Lemma 2)**.** *If $\mathbb{E}T < \infty$ then $\mathbb{E}B(T)^2 = \mathbb{E}T$.*

Proof. The idea is to use optional stopping time for the martingale $Q(t)$: then we'd have

$$\mathbb{E}(B(T)^2 - T) = \mathbb{E}Q(T) = \mathbb{E}Q(0) = \mathbb{E}(B(0)^2 - 0) = 0.$$

So we need to justify using optional stopping to get the second equality above.

Let $T_n = \inf\{t \colon |B(t)| \geq n\}$. Then

$$|Q(t \wedge T_n \wedge T)| \leq n^2 + T \in L^1.$$

Thus, we can apply optional stopping to obtain $\mathbb{E}Q(T_n \wedge T) = \mathbb{E}Q(0) = 0$. In particular, $\mathbb{E}B(T_n \wedge T)^2 = \mathbb{E}(T_n \wedge T)$.

Note that $T_n \wedge T \uparrow T$ as $n \uparrow \infty$. Corollary 3.16.3 implies $\mathbb{E}B(T)^2 \geq \mathbb{E}B(T_n \wedge T)^2 = \mathbb{E}(T_n \wedge T)$. Now monotone convergence says $\mathbb{E}(T_n \wedge T) \uparrow \mathbb{E}T$. This implies $\mathbb{E}T \leq \mathbb{E}B(T)^2$.

Fatou's lemma gives the other direction: because $B(T_n \wedge T) \to B(T)$, we have

$$\mathbb{E}B(T)^2 \leq \liminf \mathbb{E}(B(T_n \wedge T))^2 = \liminf \mathbb{E}(T_n \wedge T) = \mathbb{E}T.$$ □

The Wald lemmas are useful for the Skorohod embedding:

Theorem 3.16.6 (Skorohod embedding)**.** *Let X be a real-valued random variable with $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$. Then there exists a stopping time T such that $X \stackrel{\mathrm{d}}{=} B(T)$ and $\mathbb{E}X^2 = \mathbb{E}T$.*

(Without the second conclusion, the result is not so hard.)

We'll prove this next time.

Example 3.16.7. Consider the case where X takes values in $\{-a, b\}$, for $-a < 0 < b$. Then if $p = \mathbb{P}(X = -a)$, the condition $\mathbb{E}X = 0$ forces $p = \frac{b}{a+b}$.

A stopping time T satisfying the conclusion of the Skorohod embedding theorem is $T = \inf\{t \colon B_t \in \{-a, b\}\}$. Note that $\mathbb{E}B(T)^2 = \mathbb{E}T$ by Wald Lemma 2 (Lemma 3.16.5). We'll see in [HW 5] $\mathbb{E}T = ab$; it's easy to check $\mathbb{E}X^2 = ab$, so we obtain $\mathbb{E}T = \mathbb{E}X^2$. △

3.17 Apr 9, 2020

To build up to Skorohod embedding (Theorem 3.16.6), let us first discuss Dubbin's binary splitting. (It's kind of like playing "20 questions" to determine a random variable.)

Given a random variable X with $\mathbb{E}X = 0$, $\mathbb{E}X^2 < \infty$, let $X_0 \stackrel{\text{def}}{=} \mathbb{E}X = 0$, and let

$$\xi_0 \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } X \geq X_0 \\ -1 & \text{otherwise} \end{cases}$$

With the information of ξ_0, we can update our "best guess" for X, by setting

$$X_1 \stackrel{\text{def}}{=} \mathbb{E}(X|\xi_0) = \mathbb{E}(X|X \geq X_0)\mathbb{1}_{\{X \geq X_0\}} + \mathbb{E}(X|X < X_0)\mathbb{1}_{\{X < X_0\}}.$$

We iterate this process: define $\mathcal{F}_n = \sigma(\xi_0, \ldots, \xi_{n-1})$ and $X_n = \mathbb{E}(X|\mathcal{F}_n)$. Then we define

$$\xi_n \stackrel{\text{def}}{=} \begin{cases} +1 & \text{if } X \geq X_n \\ -1 & \text{otherwise} \end{cases}$$

One would expect that as $n \to \infty$, our "best guesses" X_n should converge to X. This is true:

Lemma 3.17.1. *The random variables X_n converge to X a.s. and in L^2.*

Proof. Note that X_n is a martingale. Furthermore,

$$\mathbb{E}X_n^2 = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_n)^2) \leq \mathbb{E}X^2$$

because conditional expectation is projection in L^2 (see Claim 1.2.4, or Equation (2) in the proof)

The Lévy upward theorem says that $X_n \to X_\infty = \mathbb{E}(X|\mathcal{F}_\infty)$ a.s. and in L^2, by the L^2 martingale convergence theorem ([Theorem 2.6.6?]), where $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

The claim is that $X_\infty = X$ a.s.. Indeed, if $X(\omega) < X_\infty(\omega)$, then there exists $N(\omega)$ such that $X(\omega) < X_n(\omega)$ for all $n > N(\omega)$. This implies $\xi_n = -1$ for all $n > N(\omega)$. This sounds quite improbable:

Indeed, define $Y_n = \xi_n(X - X_{n+1})$; for $n > N(\omega)$ we have $Y_n = |X - X_{n+1}|$. It follows that $Y_n \to |X - X_\infty|$. Now we compute

$$\begin{aligned} \mathbb{E}Y_n &= \mathbb{E}(\mathbb{E}(\xi_n(X - X_{n+1})|\mathcal{F}_{n+1})) \\ &= \mathbb{E}(\xi_n \underbrace{\mathbb{E}(X - X_{n+1}|\mathcal{F}_{n+1})}_{=0}) \end{aligned}$$

Also, the family $\{Y_n\}$ is uniformly integrable, since

$$\mathbb{E}Y_n^2 = \mathbb{E}(X - X_{n+1})^2 \leq \mathbb{E}X^2 < \infty.$$

We conclude $0 = \mathbb{E}Y_n = \mathbb{E}|X - X_\infty|$, hence $X = X_\infty$ a.s.. ∎

Last time we stated

Theorem 3.17.2 (Skorohod embedding)**.** *Let X be a real-valued random variable with $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$. Then there exists a stopping time T such that $X \stackrel{\text{d}}{=} B(T)$ and $\mathbb{E}X^2 = \mathbb{E}T$.*

Last time we used $T_{-a,b} = \inf\{t \colon B(t) \in \{-a, b\}\}$ to embed the random variable

$$X = \begin{cases} -a & \text{with probability } \frac{b}{a+b} \\ b & \text{otherwise} \end{cases}$$

In the general case we'll use Dubbins binary splitting:

Proof of Theorem 3.17.2. For round one, let $-a = \mathbb{E}(X|X < 0)$ and $b = \mathbb{E}(X|X \geq 0)$. Let $T_1 = T_{-a,b}$. In [HW 5, Ex 1] we show that

$$\mathbb{E}T_1 = ab \qquad \text{and } B(T_1) = \begin{cases} -a & \text{with probability } \frac{b}{a+b} \\ b & \text{with probability } \frac{a}{a+b} \end{cases}$$

and so

$$X_1 = \begin{cases} -a & \text{on } \{X < 0\} \\ b & \text{on } \{X \geq 0\} \end{cases}$$

is equal in distribution to $B(T_1)$. Furthermore, $\mathbb{E}X_1^2 = \mathbb{E}T_1$.

Now let $-a_1 = \mathbb{E}(X|X < X_1)$ and $b_1 = \mathbb{E}(X|X \geq X_1)$. We may define

$$X_2 = \begin{cases} X_1 - a_1 & \text{on } \{X < X_1\} \\ X_1 + b_1 & \text{on } \{X \geq X_1\} \end{cases}$$

and

$$\begin{aligned} T_2 &= \inf\{t > T_1 \colon B(t) \in \{X_1 - a_1, X_1 + b_1\}\} \\ &= \inf\{t \geq T_1 \colon B(t) - B(T_1) \in \{-a_1, b_1\}\}. \end{aligned}$$

By the strong Markov property, $B(t) - B(T_1) \in \{-a_1, b_1\}$, this is a standard Brownian motion independent of $\mathcal{F}(T_1)$. We see that $B(T_2) \stackrel{\mathrm{d}}{=} X_2$, and

$$\begin{aligned} \mathbb{E}T_2 &= \mathbb{E}T_1 + a_1 b_1 \\ &= \mathbb{E}X_1^2 + \mathbb{E}(X_2 - X_1)^2 \\ &= \mathbb{E}X_2^2, \end{aligned}$$

where the last equality is due to orthogonality of martingale increments.

In general, we obtain stopping times T_n with $B(T_n) \stackrel{\mathrm{d}}{=} X_n$, and

$$\begin{aligned} \mathbb{E}T_n &= \mathbb{E}T_1 + \mathbb{E}(T_2 - T_1) + \cdots + \mathbb{E}(T_n - T_{n-1}) \\ &= \mathbb{E}X_1^2 + \mathbb{E}(X_2 - X_1)^2 + \cdots + \mathbb{E}(X_n - X_{n-1})^2 \\ &= \mathbb{E}X_n^2. \end{aligned}$$

The stopping times $T_n \uparrow T$ so $\mathbb{E}T_n \uparrow \mathbb{E}T$ by the monotone convergence theorem. On the other hand, $\mathbb{E}T_n = \mathbb{E}X_n^2 \to \mathbb{E}X^2$, so $\mathbb{E}T = \mathbb{E}X^2 < \infty$. (In particular, we have $T < \infty$ a.s..) □

Azema-Yor proof of Theorem 3.17.2. Given $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, let $\psi(x) = \mathbb{E}(X|X \geq x)$ for $x \in \mathbb{R}$. Let

$$\tau = \inf\{t \colon M(t) \geq \psi(B(t))\},$$

where $M(t) - \sup\{B(s) \colon 0 \leq s \leq t\}$.

One can show that $\mathbb{E}\tau = \mathbb{E}X^2$ and $B(\tau) \stackrel{\mathrm{d}}{=} X$. □

Embedding random walks in Brownian motion

Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$. Let $S_n = X_1 + \cdots + X_n$.

Lemma 3.17.3. *There exist stopping times $T_1 \leq T_2 \leq \ldots$ for standard Brownian motion $(B(t))_{t\geq 0}$ such that $\mathbb{E}T_n = n$, and*

$$(S_n)_{n\in\mathbb{N}} \stackrel{\mathrm{d}}{=} (B(T_n))_{n\in\mathbb{N}}.$$

Proof. Iterate the Skorohod embedding: take a stoppign time T_1 so that $X_1 \stackrel{\mathrm{d}}{=} B(T_1)$ and $\mathbb{E}T_1 = \mathbb{E}X_1^2 = 1$. The strong Markov property implies that $(B(t) - B(T_1))_{t\geq T_1}$ is a standard Brownian motion independent of $\mathcal{F}(T_1)$. By Skorohod, there is T_2 so that $X_2 \stackrel{\mathrm{d}}{=} B(T_2) - B(T_1)$, with $\mathbb{E}(T_2 - T_1) = \mathbb{E}X_2^2 = 1$. In general, we have stopping times T_n with $X_n \stackrel{\mathrm{d}}{=} B(T_n) - B(T_{n-1})$, with $\mathbb{E}(T_n - T_{n-1}) = 1$. It follows that

$$(S_n)_{n\in\mathbb{N}} \stackrel{\mathrm{d}}{=} (B(T_n))_{n\in\mathbb{N}}.$$

□

Corollary 3.17.4 (Central limit theorem for i.i.d. random variables)**.** *We have*

$$\frac{S_n}{n} \stackrel{\mathrm{d}}{\to} B(1) \sim N(0,1).$$

Proof. We have

$$\frac{T_n}{n} \to 1 \quad \text{a.s.}$$

by the strong law of large numbers, because the increments are i.i.d. with mean 1. Then

$$\frac{S_n}{\sqrt{n}} \stackrel{\mathrm{d}}{=} \frac{B(T_n)}{\sqrt{n}} = \underbrace{\frac{B(n)}{\sqrt{n}}}_{\sim N(0,1)} - \underbrace{\frac{B(n) - B(T_n)}{\sqrt{n}}}_{\text{error term}}$$

We will check next time that the error term converges in distribution to zero. □

3.18 Apr 14, 2020

Last time, we were in the middle of embedding random walks in Brownian motion. We were about to prove Corollary 3.17.4:

Corollary 3.18.1 (Central limit theorem for i.i.d. random variables, cf. Corollary 3.17.4)**.** *We have*

$$\frac{S_n}{n} \xrightarrow{\mathrm{d}} B(1) \sim N(0,1).$$

Proof. Last lecture we had shown

$$\frac{S_n}{\sqrt{n}} \frac{S_n}{\sqrt{n}} \stackrel{\mathrm{d}}{=} \underbrace{\frac{B(n)}{\sqrt{n}}}_{\sim N(0,1)} - \underbrace{\frac{B(n)-B(T_n)}{\sqrt{n}}}_{\text{error term}},$$

so it suffices to show that the error term converges to 0 in distribution. It is convenient to denote

$$W_n(t) = \frac{B(nt)}{\sqrt{n}}.$$

For each fixed n, note that $(W_n(t))_{t\ge 0}$ is a standard Brownian motion by scale invariance. Then the error term is

$$\frac{B(T_n)-B(n)}{\sqrt{n}} = W_n\left(\frac{T_n}{n}\right) - W_n(1).$$

We claim that this converges to 0 in probability (hence, in distribution, as desired). Indeed,

$$\left\{\left|W_n\left(\frac{T_n}{n}\right) - W_n(1)\right| > \varepsilon\right\} \subseteq \{|W_n(t)-W_n(1)| > \varepsilon \text{ for some } t \in (1-\delta, 1+\delta)\} \cup \left\{\left|\frac{T_n}{n} - 1\right| > \delta\right\},$$

since either $T_n/n \in (1-\delta, 1+\delta)$ is close to 1 (in which case we are in the first event), or $|T_n/n - 1| > \delta$ (in which case we are in the second event).

Now we fix δ small to make

$$\mathbb{P}(\{|W_n(t)-W_n(1)| > \varepsilon \text{ for some } t \in (1-\delta, 1+\delta)\})$$

small, and then choose n to make $\mathbb{P}(|T_n/n - 1| > \delta)$ small. ☐

The same idea gives a strengthening of the central limit theorem, using uniform continuity of W_n on $[0,2]$. Specifically, we get

Lemma 3.18.2. *Let*

$$S_n^* \stackrel{\text{def}}{=} \frac{S(nt)}{\sqrt{n}},$$

where $S(nt)$ is a linear interpolation of the discrete time process S. Then

$$\sup_{t\in[0,1]} |S_n^*(t) - W_n(t)| \to 0 \quad \text{in probability.}$$

Weak convergence in a metric space

Let X_n, X be random variables taking values in a metric space (M, d). (The key examples for our purposes are $\mathbb{R}^m$ with Euclidean norm, and $C[0,1]$ with sup norm.)

We want to spell out what it means for $X_n \xrightarrow{\mathrm{d}} X$, called *weak convergence*, or *convergence in distribution*.

Theorem 3.18.3 (Portmanteau theorem)**.** *The following are equivalent:*

1. *$\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ for all bounded continuous $g\colon M \to \mathbb{R}$.*
2. *$\limsup_{n\to\infty} \mathbb{P}(X_n \in K) \le \mathbb{P}(X \in K)$ for all closed $K \subseteq M$.*

3. $\liminf_{n\to\infty} \mathbb{P}(X_n \in U) \geq \mathbb{P}(X \in U)$ *for all open* $U \subseteq M$.

4. $\lim_{n\to\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$ *for all borel* $A \subseteq M$ *such that* $\mathbb{P}(X \in \partial A) = 0$.

5. $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ *for all bounded measurable* $g\colon M \to \mathbb{R}$ *such that* $\mathbb{P}\{\omega\colon g \text{ discontinuous at } X(\omega)\} = 0$.

If any of these conditions hold, we write $X_n \xrightarrow{\text{d}} X$.

Donsker's Theorem (on weak convergence of random walks to Brownian motion)
Let $\{Y_n\}_{n\in\mathbb{N}}$ be i.i.d. with $\mathbb{E}Y_n = 0$ and $\mathbb{E}Y_n^2 = 1$. Let $S_n = Y_1 + \cdots + Y_n$. (The general case of Z_n with possibly nonzero mean and finite variance follows by letting

$$Y_n = \frac{Z_n - \mathbb{E}Z_n}{\sqrt{\text{Var}(Z_n)}},$$

so the conditions on Y are without loss of generality.) Let

$$S(t) \stackrel{\text{def}}{=} S_n + (t-n)Y_{n+1} \qquad \text{for } t \in [n, n+1]$$

be the linear interpolation of S_n. Rescale $S_n^*(t) = \frac{S(nt)}{\sqrt{n}}$ for $0 \leq t \leq 1$.

Theorem 3.18.4 (Donsker's theorem). *S_n^* converges weakly to a standard Brownian motion $(B(t))_{t\geq 0}$ on $(C[0,1], \sup)$.*

By definition, this means that

$$\mathbb{E}g(S_n^*) \to \mathbb{E}g(B)$$

for all continuous bounded $g\colon C[0,1] \to \mathbb{R}$. So let $F \in C[0,1]$ and let $g(F) = F(1)$. The fact that

$$\mathbb{E}S_n^*(1) = \mathbb{E}g(S_n^*) \to \mathbb{E}g(B) = \mathbb{E}B(1) = 0$$

reflects the fact that a sum of mean zero random variables is still mean zero.
Now let $\gamma\colon \mathbb{R} \to \mathbb{R}$ be any bounded continuous function, and let $g(F) = \gamma(F(1))$. Then

$$\mathbb{E}\gamma(S_n^*(1)) = \mathbb{E}g(S_n^*) \to \mathbb{E}g(B) = \mathbb{E}\gamma(B(1)).$$

It follows that $S_n^*(1) \xrightarrow{\text{d}} B(1) \sim N(0,1)$, which is a restatement of the central limit theorem.

Proof of Donsker's Theorem (Theorem 3.18.4). Fix a closed subset $K \subseteq C[0,1]$ and an $\varepsilon > 0$. Let

$$K_\varepsilon = \{F \in C[0,1]\colon \sup|F-G| \leq \varepsilon \text{ for some } G \in K\}.$$

Then $\mathbb{P}(S_n^* \in K) \leq \mathbb{P}(W_n \in K_\varepsilon) + \underbrace{\mathbb{P}(\sup|X_n - S_n^*| > \varepsilon)}_{\to 0 \text{ as } n\to\infty} = \mathbb{P}(B \in K_\varepsilon)$. Because

$$K = \bigcap_{\varepsilon>0} K_\varepsilon,$$

we have $\mathbb{P}(B \in K) = \lim_{\varepsilon\downarrow 0} \mathbb{P}(B \in K_\varepsilon)$, so

$$\limsup_{n\to\infty} \mathbb{P}(S_n^* \in K) \leq \limsup_{n\to\infty} \mathbb{P}(B \in K_\varepsilon)$$

for every ε, and we see that

$$\limsup_{n\to\infty} \mathbb{P}(S_n^* \in K) \leq \mathbb{P}(B \in K).$$

This verifies condition 2 of the Portmanteau theorem (Theorem 3.18.3) ☐

Let's consider the maximum of a random walk. As in the setup of Donsker's theorem let Y_n be i.i.d. random variables with $\mathbb{E}Y_1 = 0$ and $\mathbb{E}Y_1^2 = 1$. Let $S_n = Y_1 + \cdots + Y_n$ and $M_n = \max(S_1, \ldots, S_n)$. Then:

Theorem 3.18.5. *We have*

$$\frac{M_n}{\sqrt{n}} \xrightarrow{\mathrm{d}} \max_{t\in[0,1]} B(t) \stackrel{\mathrm{d}}{=} |B(1)|.$$

In particular, we have the same limit regardless of the distribution of the increments Y_n.

Proof. Let $\gamma\colon \mathbb{R} \to \mathbb{R}$ be bounded and continuous. We aim to show that

$$\mathbb{E}\gamma\left(\frac{M_n}{n}\right) = \mathbb{E}\gamma(\max_{t\in[0,1]} B(t)).$$

We apply Donsker's theorem (Theorem 3.18.4): let $g\colon C[0,1] \to \mathbb{R}$ be the function $g(F) = \gamma(\max_{t\in[0,1]} F(t))$; note that g is bounded since γ is, and g is continuous because γ is. Donsker (Theorem 3.18.4) says

$$\mathbb{E}\gamma\left(\max_{t\in[0,1]} \frac{S(tn)}{\sqrt{n}}\right) = \mathbb{E}g(S_n^*) \to \mathbb{E}g(B) = \mathbb{E}\gamma(\max_{t\in[0,1]} B(t))$$

Note that

$$\mathbb{E}\gamma\left(\max_{t\in[0,1]} \frac{S(tn)}{\sqrt{n}}\right) = \mathbb{E}\gamma\left(\max_{k=0}^{n} \frac{S_k}{\sqrt{n}}\right) = \mathbb{E}\gamma\left(\frac{M_n}{\sqrt{n}}\right);$$

the first equality is because $S(n)$ is a linear interpolation of S_n (hence its extrema occur at the "vertices" of the path $S(n)$). It follows that

$$\frac{M_n}{\sqrt{n}} \xrightarrow{\mathrm{d}} \max_{t\in[0,1]} B(t).$$

□

3.19 Apr 16, 2020

We'll talk about arcsine distributions today.

Definition 3.19.1. We say the random variable X is *arcsine distributed* if

$$\mathbb{P}(X \in A) = \int_A \frac{dx}{\pi\sqrt{x(1-x)}}$$

for all $A \subseteq [0,1]$. △

The distribution function $s \mapsto \mathbb{P}(X \le s)$ is

$$\int_0^s \frac{dx}{\pi\sqrt{x(1-x)}} = \frac{2}{\pi}\arcsin\sqrt{s},$$

hence the name.

Now let

$$M^* = \text{argmax}_{s\in[0,1]} B(s) = \sup\{s \le 1\colon B(s) = M(s)\},$$

where $M(t) = \max_{0\le s\le t} B(s)$.

Theorem 3.19.2 (The 1st arcsine law)**.** *The random variable M^* is arcsine distributed.*

Proof. Observe that

$$\begin{aligned}\mathbb{P}(M^* < s) &= \mathbb{P}(M(s) \ge \max_{t\in[s,1]} B(s)) \\ &= \mathbb{P}(\max_{u\in[0,s]} B(u) - B(s) \ge \max_{t\in[s,1]} B(t) - B(s)).\end{aligned}$$

We denote

$$B_1(u) \overset{\text{def}}{=} (B(u) - B(s))_{u\in[0,s]} \qquad \text{and} \qquad B_2(t) \overset{\text{def}}{=} (B(t) - B(s))_{t\in[s,1]},$$

where time in B_1 runs backwards from s to 0. The processes B_1 and B_2 are standard Brownian motions which are independent from each other, by the Markov property. It follows that

$$\mathbb{P}(M^* < s) = \mathbb{P}(M_1(s) \ge M_2(1-s)) = \mathbb{P}(|B(s)| \ge |B_2(1-s)|).$$

For $(Z_1, Z_2) \sim N(0, I_2)$, recall that $(B_1(s), B_2(s)) \overset{\text{d}}{=} (\sqrt{s}Z_1, \sqrt{1-s}Z_2)$, and so

$$\mathbb{P}(M^* < s) = \mathbb{P}(\sqrt{s}|Z_1| \ge \sqrt{1-s}|Z_2|) = \mathbb{P}(sZ_1^2 > (1-s)Z_2^2).$$

We now apply rotational symmetry of joint normal distributions. Specificaly, let

$$\left\{\begin{array}{c} Z_1 = R\cos\theta \\ Z_2 = R\sin\theta \end{array}\right\}$$

for θ uniformly distributed. With this notation, we continue the long string of equalities with

$$\mathbb{P}(M^* s) = \mathbb{P}(sR^2 > Z_2^2) = \mathbb{P}(s > (\sin\theta)^2) = \frac{4\arcsin\sqrt{s}}{2\pi}.$$

□

Corollary 3.19.3. *The random variable $L = \sup\{t \le 1\colon B(t) = 0\}$ is also arcsine distributed.*

(Note in particular that the distribution of L is symmetric about $\frac{1}{2}$.)

Proof. Let $\widetilde{B}(t) = M(t) - B(t)$; Lévy (Theorem 3.14.4) says that $(\widetilde{B}(t))_{t\geq 0} \stackrel{d}{=} (|B(t)|)_{t\geq 0}$. Now,

$$\begin{aligned} L = \sup\{t \leq 1 \colon |B(t)| = 0\} &\stackrel{d}{=} \sup\{t \leq 1 \colon \widetilde{B}(t) = 0\} \\ &= \sup\{t \leq 1 \colon M(t) = B(t)\} = M^*. \end{aligned}$$

□

Transfer of arcsine law from Brownian motion to random walks

Let $S_n = X_1 + \cdots + X_n$, where X_i are i.i.d. with $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = 1$. Let $N_n = \max\{k \leq n \colon S_k S_{k-1} \leq 0\}$ be the last sign change before time n.

Theorem 3.19.4. *We have*

$$\frac{N_n}{n} \xrightarrow{d} L = \sup\{t \leq 1 \colon B_t = 0\}.$$

Hence, $\mathbb{P}(N_n < sn) \to \frac{2}{\pi} \arcsin \sqrt{s}$ *as* $n \to \infty$.

Proof. We use Donsker's theorem (Theorem 3.18.4), i.e. we choose $g \colon C[0,1] \to \mathbb{R}$ so that $\mathbb{E}g(S_n^*) \to \mathbb{E}g(B)$. The naive guess of $g(F) = \max\{t \leq 1 \colon F(t) = 0\}$ turns out not to be continuous (if the root of F at $g(F)$ is tangent to the x-axis, incrementing $F(t)$ by ε makes g jump).

Thankfully, part 5 of the Portmanteau theorem (Theorem 3.18.3) says that the converngence

$$\mathbb{E}g(S_n^*) \to \mathbb{E}g(B)$$

holds even for bounded discontinuous g, provided $\mathbb{P}(g \text{ is discontinuous at } B) = 0$. It turns out g satisfies this condition: in Mörters-Peres they show that for

$$Y = \left\{ F \in C[0,1] \colon \begin{array}{c} F(1) \neq 0, \text{ and for all } z \in [0,1] \text{ such that } F(z) = 0 \text{ for all } \delta > 0, \\ F \text{ takes both positive and negative values in } (z-\delta, z+\delta) \end{array} \right\},$$

we have $\mathbb{P}(B \in Y) = 1$. (Note that g is continuous on Y.)

Composing with any bounded continuous $\gamma \colon \mathbb{R} \to \mathbb{R}$, we may apply Donsker to get

$$\mathbb{E}\gamma(g(S_n^*)) \to \mathbb{E}\gamma(g(B))$$

to conclude that $g(S_n^*) \xrightarrow{d} g(B)$. Since

$$g(S_n^*) = \frac{N_n}{n} + O\left(\frac{1}{n}\right) \qquad \text{and} \qquad g(B) = L,$$

the result follows. □

Again let $S_n = X_1 + \cdots + X_n$ with X_i i.i.d. with mean 0 and variance 1. Let $P_n = \sum_{k=1}^n \mathbb{1}_{\{S_k > 0\}}$ count the amount of time that S is positive. Then

Theorem 3.19.5. *We have*

$$\frac{P_n}{n} \xrightarrow{d} T = \lambda\{t \leq 1 \colon B(t) > 0\}$$

and $\lambda(t \leq 1 \colon B(t) > 0)$ *is arcsine distributed.*

(Here, λ is the Lebesgue measure.)

Hence

$$\mathbb{P}\left(\frac{P_n}{n} < s\right) \to \frac{2}{\pi} \arcsin \sqrt{s}. \tag{17}$$

Proof outline. One can first prove Equation (17) in the special case of a simple random walk $X_i \in \{\pm 1\}$, which can be handled combinatorially.

Then apply Donsker to see that T is arcsine distributed.

By Donsker again, one concludes that Equation (17) holds in general. □

3.20 Apr 21, 2020

We have been able to embed random walks in Brownian motion (see Lemma 3.17.3). In fact, one can embed discrete time martingales in Brownian motion.

Specifically, let $(S_n)_{n\geq 0}$ be a discrete time martingale and suppose $\mathbb{E}S_n^2 < \infty$ and $S_0 = 0$. Then:

Lemma 3.20.1. *There exist stopping times $0 = T_0 \leq T_1 \leq T_2 \leq \dots$ such that $(S_0, S_1, \dots, S_)) \overset{\mathrm{d}}{=} (B(T_0), B(T_1), \dots, B(T_k))$ for all $k \geq 0$, where $(B(t))_{t\geq 0}$ is a standard Brownian motion. Moreover, $\mathbb{E}(T_{n+1} - T_n | \mathcal{F}_n) = \mathbb{E}((S_{n+1} - S_n)^2)$.*

Proof. We iterate Skorohod (Theorem 3.16.6), as we did for random walks in Lemma 3.17.3. Specifically, since S_1 has $\mathbb{E}S_1 = 0$ and $\mathbb{E}S_1^2 < \infty$, there exists a stopping time T_1 with $B(T_1) \overset{\mathrm{d}}{=} S_1$ and $\mathbb{E}T_1 = \mathbb{E}S_1^2$.

By the strong Markov property (Theorem 3.13.11), given $T_1, \dots, T_{k-1}$, the process

$$\widetilde{B} \overset{\text{def}}{=} (B(t) - B(T_{k-1}))_{t\geq T_{k-1}}$$

is a standard Brownian motion independent of $\mathcal{F}^+(T_{k-1})$. By Skorohod (Theorem 3.16.6), there exists T_k with $\widetilde{B}(T_k) \overset{\mathrm{d}}{=} S_k - S_{k-1}$ and $\mathbb{E}T_k = \mathbb{E}(S_k - S_{k-1})^2$. □

Our goal is to use Lemma 3.20.1 to prove a central limit theorem for certain dependent random variables.

Definition 3.20.2. A *martingale difference array* $(X_{n,m})_{1\leq m\leq n}$ adapted to a filtration $(\mathcal{F}_{n,m})_{1\leq m\leq n}$ is a collection of random variables so that $X_{n,m}$ is $\mathcal{F}_{n,m}$-measurable and

$$\mathbb{E}(X_{n,m}|\mathcal{F}_{n,m-1}) = 0 \qquad \text{for all } m \leq n.$$

△

So for each n, the process

$$S_{n,m} = X_{n,1} + \cdots + X_{n,m}$$

is a martingale with respect to the filtration $(\mathcal{F}_{n,m})_{m\geq 1}$.

Now let

$$V_{n,k} = \sum_{j=1}^{k} \mathbb{E}(X_{n,j}^2|\mathcal{F}_{n,j-1})$$

denote the quadratic variation. We have:

Theorem 3.20.3. *Suppose there exists a sequence of real numbers $\varepsilon \downarrow 0$ such that:*

1. *$|X_{n,m}| < \varepsilon_n$ for all n,*
2. *$V_{n,\lfloor nt\rfloor} \to t$ in probability for all $t \in [0,1]$.*

Then $(S_{n,nt})_{t\in[0,1]} \overset{\mathrm{d}}{\to} (B(t))_{t\in[0,1]}$.

(The process $(S_{n,nt})_{t\in[0,1]}$ shall be the linear interpolation of the discrete process $S_{n,m} = X_{n,1} + \cdots + X_{n,m}$.)

Remark 3.20.4. Condition (i) can be weakened to

$$\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 \mathbb{1}_{\{|X_{n,m}>\varepsilon\}}] \to 0 \qquad \text{in probability}$$

as in Lindeberg-Feller (see [6710, Theorem 6.19.4]). In HW 6, we'll show that there exist a sequence of independent mean 0 variance 1 random variables X_n such that

$$\frac{S_n}{\sqrt{n}} \to 0 \qquad \text{in probability},$$

where $S_n = X_1 + \cdots + X_n$. △

Proof of Theorem 3.20.3. By Lemma 3.20.1, there exist stopping times $T_{n,k}$ so that

$$(S_{n1},\dots,S_{nn}) \overset{\mathrm{d}}{=} (B(T_{n1}),\dots,B(T_{nn})).$$

We claim that $T_{n,\lfloor nt\rfloor} \to t$ in probability for $t \in [0,1]$. Indeed, let $\tau_{n,m} = T_{n,m} - T_{n,m-1}$ (we set $T_{n,0} = 0$), and observe that

$$\mathbb{E}(X_{n,m}^2|\mathcal{F}_{n,m-1}) = \mathbb{E}(\tau_{n,m}|\mathcal{F}_{n,m-1}).$$

Summing over m, we see that

$$\sum_{m=1}^{\lfloor nt\rfloor} \mathbb{E}(\tau_{n,m}|\mathcal{F}_{n,m-1}) = V_{n,\lfloor nt\rfloor} \to t \qquad \text{in probability}$$

by condition (ii).

Now observe that

$$\begin{aligned}
\mathbb{E}(T_{n,nt} - V_{n,nt})^2 &= \mathbb{E}\left[\sum_{m=1}^{\lfloor nt\rfloor}(\tau_{n,m} - \mathbb{E}[\tau_{n,m}|\mathcal{F}_{n,m-1}])\right]^2 \\
&= \mathbb{E}\sum_{m=1}^{\lfloor nt\rfloor}\underbrace{(\tau_{n,m} - \mathbb{E}[\tau_{n,m}|\mathcal{F}_{n,m-1}])^2}_{<\mathbb{E}(\tau_{nm}^2|\mathcal{F}_{n,m-1})} \\
&\le C\varepsilon_n^2\mathbb{E}V_{n,n} \to 0,
\end{aligned}$$

where the last inequality follows from the chain of inequalities

$$\mathbb{E}(\tau_{nm}^2|\mathcal{F}_{n,m-1}) \le C\mathbb{E}(X_{n,m}^4|\mathcal{F}_{n,m-1}) \le C\varepsilon_n^2\mathbb{E}(X_{n,m}^2|\mathcal{F}_{n,m-1})$$

applied to each term in the summand. It follows that $\mathbb{E}T_{n,nt} \to t$ in probability, since $V_{n,nt}$ does too. □

Before we move on from Brownian motion, let us remark that Brownian motion lives in the intersection of three worlds (Gaussian processes, Martingales, and Markov processes); the families of random variables living in two but not three are also of interest and include

- Fractional Brownian motions, which have some memory of the past (i.e. some "momentum")
- Ornstein-Uhlenbeck processes, which have some restoring force
- Brownian bridges
- Lévy processes

4 Ergodic Theory

4.21 Apr 23, 2020

We begin by discussing stationary sequences:

Definition 4.21.1. A sequence of random variables $(X_0, X_1, \dots)$ is *stationary* if for all $k \geq 0$ and $m \geq 0$ we have $(X_0, X_1, \dots, X_m) \stackrel{\text{d}}{=} (X_k, X_{k+1}, \dots, X_{k+m})$. △

Example 4.21.2. Any set of i.i.d. random variables is stationary. △

Example 4.21.3. More generally, let $(X_n)_{n\geq 0}$ be a Markov chain with state space S (with S finite or countable) and transition matrix $(p(x,y))_{x,y\in S}$.

(Recall that a Markov chain is defined by

$$\mathbb{P}(X_{n+1} = y | \mathcal{F}_n) = \mathbb{P}(X_{n+1} = y | X_n) = p(X_n, y)$$

for all $n \geq 0$ and $y \in S$.)

Definition 4.21.4. A *stationary dsitribution* for p is a probability distribution π on S such that

$$\sum_{x\in S} \pi(x) p(x,y) = \pi(y) \qquad \text{for all } y \in S.$$ △

(This can be thought as a matrix equation, where $\pi = (\pi(x))$ is a row vector and $p = (p(x,y))$ is an $S \times S$ matrix; then $\pi p = \pi$.)

If $X_0 \sim \pi$ is a stationary distribution, then we claim $X_n \sim \pi$ for all $n \geq 1$, and $(X_n)_{n\geq 0}$ is a stationary sequence. (In HW 6, we'll show that there exists a sequence that is not stationary, but all marginals are equal.)

Indeed, to show $X_n \sim \pi$, note that

$$\mathbb{P}(X_1 = y) = \sum_{x\in S} \mathbb{P}(X_1 = y, X_0 = x) = \sum_{x\in S} \pi(x) p(x,y) = \pi(y).$$

Then we can induct. Then, to show $(X_n)_{n\geq 0}$ is stationary, we just note

$$\mathbb{P}((X_k, \dots, X_{k+m}) = (s_0, \dots, s_m)) = \pi(s_0) p(s_0, s_1) p(s_1, s_2) \dots p(s_{m-1}, s_m),$$

using the Markov property. This holds for every k, as desired. △

Example 4.21.5. Let's consider the Markov chain with state space $S = \{a, b\}$ and

$$p = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

so the stationary distribution is $\pi(a) = \pi(b) = \frac{1}{2}$. Then

$$(X_n)_{n\geq 0} = \begin{cases} (a, b, a, b, \dots) & \text{with probability } \frac{1}{2} \\ (b, a, b, a, \dots) & \text{with probability } \frac{1}{2} \end{cases}$$

△

Definition 4.21.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say $\varphi \colon \Omega \to \Omega$ is *measure-preserving* if $\mathbb{P}(\varphi^{-1} A) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$. △

Example 4.21.7. Let $\Omega = \mathbb{R}/\mathbb{Z}$ be the circle and let $\mathbb{P}$ be Lebesgue measure. Then:

1. The rotation $\varphi(\omega) = \omega + \theta \pmod 1$ by angle θ is measure preserving.
2. The map $\varphi(\omega) = 2\omega \pmod 1$ is measure preserving. (Note it's not true that $\mathbb{P}(\varphi A) = \mathbb{P}(A)$.)

△

Lemma 4.21.8. *Let $X \in m\mathcal{F}$. If φ is measure-preserving, then*

$$X_n(\omega) = X(\varphi^n \omega)$$

is a stationary sequence.

Proof. For any Borel set $B \subset \mathbb{R}^{m+1}$, let

$$A = \{\omega\colon (X_0(\omega), \dots, X_m(\omega)) \in B\}.$$

Then

$$\begin{aligned}
\mathbb{P}((X_k, \dots, X_{k+m}) \in B) &= \mathbb{P}(\omega\colon (X_0(\varphi^k\omega), \dots, X_m(\varphi^k\omega)) \in B) \\
&= \mathbb{P}(\omega\colon \varphi^k\omega \in A) \\
&= \mathbb{P}(\varphi^{-k}A) \\
&= \mathbb{P}(A).
\end{aligned}$$

□

Conversely, if $(Y_n)_{n\geq 0}$ is a stationary sequence, then by the Kolmogorov extension theorem ([6710, Thm. 4.10.11]) there exists a measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ such that $(Y_n)_{n\geq 0} \overset{\text{d}}{=} (X_n)_{n\geq 0}$, where $X_n((\omega_0, \omega_1, \dots)) = \omega_n$. In particular, if $\varphi(\omega_0, \omega_1, \dots)$ is the (measure-preserving) shift map, then $X_n = X_0 \circ \varphi^n$. Thus, every stationary sequence arises from Lemma 4.21.8.

(Here, $\mathbb{R}^{\mathbb{N}} = \{(\omega_0, \omega_1, \dots)\colon \omega_i \in \mathbb{R})\}$ and $\mathcal{B}^{\mathbb{N}}$ is the σ-field generated by finite codimensional rectangles $(a_0, b_0] \times \cdots \times (a_k, b_k] \times \mathbb{R} \times \mathbb{R} \times \dots$.)

Lemma 4.21.9. *If $(X_n)_{n\geq 0}$ is a stationary sequence and $g\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is measurable, then $Y_n = g(X_n, X_{n+1}, \dots)$ is a stationary sequence.*

Proof. For any $B \in \mathcal{B}^{\mathbb{N}}$, let $A = \{x \in \mathbb{R}^{\mathbb{N}}\colon (g(x_0, x_1, \dots), g(x_1, x_2, \dots), g(x_2, x_3, \dots), \dots) \in B\}$. Then

$$\begin{aligned}
\mathbb{P}(Y_0, Y_1, \dots) \in B) &= \mathbb{P}((X_0, X_1, \dots) \in A) \\
&= \mathbb{P}((X_k, X_{k+1}, \dots) \in A) \\
&= \mathbb{P}((Y_k, Y_{k+1}, \dots) \in B),
\end{aligned}$$

hence Y is stationary. □

Example 4.21.10. Let $(X_n)_{n\geq 0}$ be i.i.d. Bernoulli $\frac{1}{2}$ random variables, and let

$$g(x_0, x_1, \dots) = \frac{x_0}{2} + \frac{x_1}{4} + \frac{x_2}{8} + \dots.$$

This is the binary expansion of $\omega \sim \text{Unif}(0, 1)$. Upon shifting, we get

$$g(x_n, x_{n+1}, \dots) = \frac{x_n}{2} + \frac{x_{n+1}}{4} + \dots,$$

which is the binary expansion of $2^n\omega \pmod 1$. So this stationary sequence arises from the measure preserving map $\varphi(\omega) = 2\omega \pmod 1$ which we saw in part 2 of Example 4.21.7. △

Definition 4.21.11. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\varphi\colon \Omega \to \Omega$ be a measure preserving map, so $\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$. Let $X\colon \Omega \to \mathbb{R}$ and $X_n(\omega) = X(\varphi^n\omega)$. We say $A \in \mathcal{F}$ is *invariant* if $\varphi^{-1}(A) = A$ holds $\mathbb{P}$-a.s., that is to say, $\mathbb{P}(\varphi^{-1}A \triangle A) = 0$. △

(Here, we write $A \triangle B = (A \cap B^c) \cup (A^c \cap B)$ to denote the symmetric difference of the two sets.)
In HW 6, we'll show that:

1. $\mathcal{I} = \{\text{invariant } A \in \mathcal{F}\}$ is a σ-field, and

2. $X \in m\mathcal{I}$ if and only if $X(\varphi\omega) = X(\omega)$ holds $\mathbb{P}$-a.s..

Definition 4.21.12. The map φ is *ergodic* if $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathcal{I}$. Similarly, we say $(X_n)_{n\geq 0}$ is ergodic if the measure-preserving map φ it arises from is ergodic. △

Note that if φ is not ergodic, then there is $A \in \mathcal{I}$ with $0 < \mathbb{P}(A) < 1$ with $\varphi(A) = A$ and $\varphi(A^c) = A^c$, both holding $\mathbb{P}$-a.s.. So in this sense, the map is "not irreducible".

Example 4.21.13.

1. A sequence of i.i.d. random variables is ergodic, because for any $A \in \mathcal{I}$ there exists a tail event $T \in \mathcal{T}$ such that $A = T$ holds $\mathbb{P}$-a.s.. (The Kolmogorov 0-1 law says that every $A \in \mathcal{T}$ has $\mathbb{P}(A) \in \{0, 1\}$.)

2. The Markov chain $(X_n)_{n\geq 0}$ is ergodic if and only if it's irreudcible: we say x communicates with y if $p^n(x, y) > 0$ for some n; this generates an equivalence relation on S; irreducibility of the Markov chain means that there is only one equivalence class.

3. We will see next time that the rotation of the circle $\varphi(x) = x + \theta \pmod 1$ is ergodic if and only if $\theta \notin \mathbb{Q}$.

△

4.22 Apr 28, 2020

Let $\Omega = \mathbb{R}^{\mathbb{N}}$ and $\mathcal{F} = \mathcal{B}^{\mathbb{N}}$, and let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. Also let $X_0(\omega_0, \omega_1, \dots) = \omega_0$ and $\varphi\colon \Omega \to \Omega$ be the shift map, i.e. $\varphi(\omega_0, \omega_1, \dots) = (\omega_1, \omega_2, \dots)$. Define $X_n = X_0(\varphi^n \omega) = \omega_n$.

Definition 4.22.1. An event $A \in \mathcal{F}$ is called *strictly invariant* if $\varphi^{-1} A = A$. An event $A \in \mathcal{F}$ is called *invariant* if $\mathbb{P}((\varphi^{-1} A) \triangle A) = 0$. △

In HW 6, we'll show for any invariant $A \in \mathcal{F}$, there is a strictly invariant A' such that $\mathbb{P}(A \triangle A') = 0$.

Observe that strictly invariant events are tail events: we have $A = \varphi^{-1} A = \varphi^{-2} A = \dots$, and hence that

$$A = \varphi^{-n} A = \{\omega \in \Omega\colon (\omega_n, \omega_{n+1}, \dots) \in A\} \in \sigma(X_n, X_{n+1}, \dots).$$

Thus

$$A \in \mathcal{T} = \bigcap_{n \geq 1} \sigma(X_n, X_{n+1}, \dots).$$

Example 4.22.2. Not every tail event is invariant.

Let $(\xi_n)_{n \geq 1}$ be independent ± 1 random variables with $\mathbb{P}(\xi_n = \pm 1) = \frac{1}{2}$. Let $X_n = X_0 + \xi_1 + \dots + \xi_n \pmod{k}$, where $X_0 \sim \text{Unif}(\mathbb{Z}/k\mathbb{Z})$. Then $(X_n)_{n \geq 0}$ is a stationary sequence (in fact, it's a Markov chain and X_0 follows a stationary distribution).

Note that the event $A = \{X_0 \text{ is even}\}$ is not invariant, since $\varphi^{-1} A = \{X_1 \text{ is even}\} \neq A$. But if k is even, then X_0 is even if and only if X_{2n} is even for all n; in other words,

$$A = \{X_{2n} \text{ is even}\} \qquad \text{for all } n,$$

hence $A \in \mathcal{T}$. △

Theorem 4.22.3 (Birkhoff ergodic theorem)**.** *We have*

$$\frac{1}{n} \sum_{m=0}^{n-1} X(\varphi^m \omega) \to \mathbb{E}(X | \mathcal{I})(\omega) \qquad \text{a.s. and in } L^1.$$

(Here, $\mathcal{I}$ is the invariant σ-field.)

Example 4.22.4.

1. If $X_n(\omega) = X(\varphi^n \omega)$ is i.i.d., then $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathcal{I}$. Thus $\mathbb{E}(X|\mathcal{I}) = \mathbb{E}X$ a.s., and we get the strong law of large numbers.

2. If $(X_n)_{n \geq 0}$ is an irreducible Markov chain, one can prove that $\mathcal{I}$ is trivial (i.e. consists of probability 0 or 1 events). If the state space S is countable, then given $f\colon S \to \mathbb{R}$ satisfying

$$\sum_{x \in S} |f(x)| \pi(x) < \infty,$$

where π is the stationary distribution, then applying the ergodic theorem to stationary sequence $(f(X_n)_{n \geq 0}$ we get

$$\underbrace{\frac{1}{n} \sum_{m=0}^{n-1} f(X_m)}_{\text{time average over } m} \to \mathbb{E}(f(X_0)|\mathcal{I}) \qquad \text{a.s.,}$$

$$= \mathbb{E}(f(X_0)|\mathcal{I})$$

$$\underbrace{= \mathbb{E} f(X_0) = \sum_{x \in S} f(x) \pi(x)}_{\text{space average over } x}.$$

So the mantra of the ergodic theorem is that time averages are space averages.

3. Consider the rotation of the circle, i.e. let $\Omega = \mathbb{R}/\mathbb{Z}$ and let $\varphi(x) = x + \theta \pmod 1$, where θ is irrational. We claim φ is ergodic, i.e. that $\mathcal{I}$ is trivial. It's enough to show that if $f \in m\mathcal{I}$ then f is constant a.e..

   To show this, recall that any measurable $f\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ has a Fourier series

$$\sum_{k\in\mathbb{Z}} c_k e^{2\pi ikx} \to f(x) \qquad \text{in } L^2(\mathbb{R}/\mathbb{Z})$$

   and furthermore, the c_k are unique (by Fourier inversion). Now for $f \in m\mathcal{I}$, we have

$$\begin{aligned} f(\varphi x) &= \sum_{k\in\mathbb{Z}} c_k e^{2\pi ik(x+\theta \pmod 1)} \\ &= \sum_{k\in\mathbb{Z}} c_k e^{2\pi ik(x+\theta)} \\ &= \sum_{k\in\mathbb{Z}} (c_k e^{2\pi i\theta k}) e^{2\pi ikx}. \end{aligned}$$

   On the other hand,

$$f(\varphi x) = f(x) = \sum c_k e^{2\pi ikx} \qquad \text{for all } k,$$

   so by uniqueness of Fourier coefficients we obtain

$$c_k = c_k e^{2\pi i\theta k} \qquad \text{for all } k \in \mathbb{Z}.$$

   Since for $k \neq 0$ we have $e^{2\pi i\theta k} - 1 \neq 0$ by irrationality of θ, we see that $c_k = 0$ for $k \neq 0$, and hence f is constant.

   We remark that rational rotations are not ergodic: for $\theta = \frac{p}{q}$ for $p, q \in \mathbb{N}$, we may consider the finite union

$$A = \bigcup_{n\geq 0} \varphi^n(I)$$

   for a little interval I.

   Let us now apply the ergodic theorem to $\varphi(x) = x + \theta \pmod 1$, where θ is irrational. Given a Borel set $A \subseteq [0, 1)$, let $X = \mathbb{1}_A$ and let $X_n(\omega) = X(\varphi^n\omega)$. Then

$$\frac{1}{n}\sum_{m=0}^{n-1} X_m(\omega) \to \mathbb{E}(X|\mathcal{I}) = \mathbb{E}X = \lambda(A) \qquad \text{a.s.,} \tag{18}$$

   where λ is Lebesgue measure. Note that the left hand side of Equation (18)

$$\frac{1}{n}\sum_{m=0}^{n-1} X_m = \frac{1}{n}\#\{m < n \colon \omega + m\theta \pmod 1 \in A\}.$$

   In this example, the a.s. convergence in Equation (18) can be upgraded to pointwise convergence; this is exactly the claim of Weyl's equidistribution theorem.

△

If $f\colon \Omega \to \mathbb{R}$ is integrable and $\varphi\colon \Omega \to \Omega$ is measure-preserving, then

$$(\varphi_*\mathbb{P})(A) = \mathbb{P}(\varphi^{-1}A) = \mathbb{P}(A),$$

so $\varphi_*\mathbb{P} = \mathbb{P}$. In particular,

$$\int_\Omega f(\varphi\omega)\, d\mathbb{P} = \int_\Omega f(\omega) d(\varphi_*\mathbb{P}) = \int_\Omega f\, d\mathbb{P}.$$

Intuitively, φ "scrambles" Ω.

The next lemma is an important step in proving the ergodic theorem (Theorem 4.22.3). Let φ be measure preserving on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $X_n(\omega) = X(\varphi^n\omega)$. Let $S_n = X_0 + \cdots + X_{n-1}$ and $M_n = \max(0, S_1, \ldots, S_n)$, and let $A_n = \{M_n > 0\}$.

Lemma 4.22.5 (Maximal ergodic lemma)**.** *With notation as above, we have*

$$\mathbb{E}(X\mathbb{1}_{A_n}) \geq 0 \qquad \text{for all } n.$$

Proof. We have

$$\begin{aligned} S_{j+1}(\omega) &= X_0(\omega) + \cdots + X_j(\omega) \\ &= X(\omega) + X_0(\varphi\omega) + \cdots + X_{j-1}(\varphi\omega) \\ &= X(\omega) + S_j(\varphi\omega). \end{aligned}$$

In other words, we obtain

$$S_{j+1}(\omega) = X(\omega) + S_j(\varphi\omega) \leq X(\omega) M_k(\varphi\omega) \qquad \text{for all } k \geq j.$$

Thus

$$X(\omega) \geq S_{j+1}(\omega) - M_k(\varphi\omega) \qquad \text{for } j = 1, \ldots, k.$$

In fact the inequality also holds for $j = 0$, since $M_k \geq 0$. Now

$$\mathbb{E}(X\mathbb{1}_{A_k}) = \int_{A_k} X \, d\mathbb{P} \geq \int_{A_k} \max_{j=0}^{k-1} (S_{j+1}(\omega) - M_k(\varphi\omega)) \, d\mathbb{P}.$$

Since $\max_{j=0}^{k-1} S_{j+1}(\omega) \geq M_k(\omega)$ for $\omega \in A_k$, we have

$$\int_{A_k} \max_{j=0}^{k-1} (S_{j+1}(\omega) - M_k(\varphi\omega)) \, d\mathbb{P} \geq \int_{A_k} M_k(\omega) - M_k(\varphi\omega) \, d\mathbb{P} \geq \int_{\Omega} M_k(\omega) - M_k(\varphi\omega) \, d\mathbb{P},$$

since the integrand is nonpositive for $\omega \in \Omega \setminus A_k = \{M_k(\omega) = 0\}$. But note that φ is measure preserving, so

$$\int_{\Omega} M_k(\omega) - M_k(\varphi\omega) \, d\mathbb{P} = 0.$$

Chaining the inequalities together gives the desired claim. □

4.23 Apr 30, 2020

We're ready to prove the ergodic theorem today. We have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and a measure preserving $\varphi\colon \Omega \to \Omega$. The ergodic theorem says:

Theorem 4.23.1 (Birkhoff ergodic theorem, cf. Theorem 4.22.3)**.** *We have*

$$\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m\omega) \to \mathbb{E}(X|\mathcal{I})(\omega) \qquad \text{a.s. and in } L^1.$$

(Here, $\mathcal{I}$ is the invariant σ-field.)

Claim 4.23.2. *It suffices to prove Theorem 4.23.1 for the special case* $\mathbb{E}(X|\mathcal{I}) = 0$ *a.s..*

Proof. Let $X' = X - \mathbb{E}(X|\mathcal{I})$ so that $\mathbb{E}(X'|\mathcal{I}) = 0$ a.s.. Since $\mathbb{E}(X|\mathcal{I})$ is $\mathcal{I}$-measurable, we have $\mathbb{E}(X|\mathcal{I})(\omega) = \mathbb{E}(X|\mathcal{I})(\varphi^m\omega)$ for each m. (There is a different null set for each m where this a.s. equality fails.)

We see that

$$\begin{aligned}\frac{1}{n}\sum_{m=0}^{n-1} X'(\varphi^m\omega) &= \frac{1}{n}\sum\Big(X(\varphi^m\omega) - \mathbb{E}(X|\mathcal{I})(\varphi^m\omega)\Big)\\ &\overset{\text{a.s.}}{=} \left(\frac{1}{n}\sum_{m=0}^{m-1} X(\varphi^m\omega)\right) - \mathbb{E}(X|\mathcal{I})(\omega),\end{aligned}$$

where the a.s. equality holds on the (full measure!) set where $\mathbb{E}(X|\mathcal{I})(\omega) = \mathbb{E}(X|\mathcal{I})(\varphi^m\omega)$ holds for all m simultaneously. ☐

Proof of Theorem 4.23.1. Let

$$S_n = \sum_{m=0}^{n-1} X(\varphi^m\omega) \qquad \text{and} \qquad \overline{X} = \limsup_{n\to\infty}\frac{S_n}{n}.$$

The goal is to show $\mathbb{P}(D) = 0$, where $D = \{\overline{X} > \varepsilon\}$. This suffices to show a.s. convergence since we'd get

$$\mathbb{P}\bigg(\limsup\frac{S_n}{n} > \varepsilon\bigg) = 0 \qquad \text{and} \qquad \mathbb{P}\bigg(\limsup\frac{-S_n}{n} > \varepsilon\bigg) = \mathbb{P}\bigg(\liminf\frac{S_n}{n} < -\varepsilon\bigg) = 0,$$

where in the right side above we applied the result to $-X$.

Note that $D = \{\overline{X} > \varepsilon\} \in \mathcal{I}$. Since $\overline{X}(\varphi\omega) = \limsup\frac{S_n(\varphi\omega)}{n}$ and

$$S_n(\varphi\omega) = \sum_{m=0}^{n-1} X(\varphi\varphi^m\omega) = \sum_{m=1}^{n} X(\varphi^m\omega) = S_{n+1}(\omega) - X(\omega).$$

It follows that

$$\overline{X}(\varphi\omega) = \limsup\frac{S_{n+1}(\omega)}{n}\frac{n}{n+1} - \limsup\frac{X(\omega)}{n} = \limsup\frac{S_{n+1}(\omega)}{n+1} = \overline{X}(\omega).$$

Now let $X^* = (X-\varepsilon)\mathbb{1}_D$ and $S_n^* = X^*(\omega) + X^*(\varphi\omega) + \cdots + X^*(\varphi^{n-1}\omega)$ and $M_n^* = \max(0, S_1^*, \ldots, S_n^*)$. Define the event $F_n = \{M_n^* > 0\} = \{S_k^* > 0 \text{ for some } k \in [n]\}$. Finally, let

$$\begin{aligned}F = \bigcup_{n\geq 1} F_n &= \left\{\frac{S_k^*}{k} > 0 \text{ for some } k \in \mathbb{N}\right\}\\ &= D \cap \left\{\frac{S_k}{k} > \varepsilon \text{ for some } k \in \mathbb{N}\right\}\\ &= D.\end{aligned}$$

By the maximal ergodic lemma (Lemma 4.22.5), we have

$$\mathbb{E}X^* \mathbb{1}_{F_n} \geq 0.$$

Because $\mathbb{E}|X^*| \leq \mathbb{E}|X| + \varepsilon$ and $X \in L^1$ we may apply dominated convergence to see that $\mathbb{E}X^* \mathbb{1}_F \geq 0$. But now $\mathbb{E}(X^* \mathbb{1}_F) = \mathbb{E}(X^* \mathbb{1}_D) = \mathbb{E}((X - \varepsilon)\mathbb{1}_D)$.

Since $D \in \mathcal{I}$, we have

$$\mathbb{E}((X-\varepsilon)\mathbb{1}_D) = \mathbb{E}\big[\mathbb{E}((X-\varepsilon)\mathbb{1}_D|\mathcal{I})\big] = \mathbb{E}\big[\mathbb{1}_D(\mathbb{E}(X|\mathcal{I}) - \varepsilon)\big] = -\varepsilon\mathbb{P}(D).$$

Hence $\mathbb{P}(D) = 0$. This shows a.s. convergence.

To show the L^1 convergence, we first fix $M > 0$ and write $X = Y + Z$ where $Y = X\mathbb{1}_{\{X \leq m\}}$ and $Z = X\mathbb{1}_{\{X > M\}}$. The point is that the convergence

$$\frac{1}{n}\sum_{m=0}^{n-1} Y(\varphi^m \omega) \to \mathbb{E}(Y|\mathcal{I}) \qquad \text{a.s.}$$

also holds in L^1 because we can apply bounded convergence theorem to Y. Furthermore,

$$\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} Z(\varphi^m\omega) - \mathbb{E}(Z|\mathcal{I})\right| \leq 2\mathbb{E}|Z| \to 0$$

as $M \to \infty$ by dominated convergence theorem (since $Z \leq X$). □

Theorem 4.23.3. *Let $\varphi\colon \Omega \to \Omega$ be a measure preserving map. Then φ is ergodic if and only if for all $A, B \in \mathcal{I}$ we have*

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{P}((\varphi^{-m}A) \cap B) \to \mathbb{P}(A)\mathbb{P}(B) \qquad \text{as } n \to \infty. \tag{19}$$

Proof. We first prove the backwards direction. Pick any $A \in \mathcal{I}$, so $\varphi^{-m}A = A$ a.s. for all m. Thus, Equation (19) implies

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{P}(\varphi^{-m}A \cap B) \to \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)^2.$$

On the other hand the left hand side is just

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{P}(A) = \mathbb{P}(A),$$

so $\mathbb{P}(A) \in \{0, 1\}$ and φ is ergodic.

Let's now prove the forwards direction. Suppose φ is ergodic. Then $\mathbb{E}(\mathbb{1}_A|\mathcal{I}) = \mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$.

Applying the ergodic theorem (Theorem 4.23.1) to the event $\mathbb{1}_A$, we have

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m\omega) \to \mathbb{P}(A) \qquad \text{a.s. and in } L^1.$$

Multiplying both sides of the convergence by $\mathbb{1}_B$, we conclude furthermore that

$$\mathbb{E}\left[\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{1}_A(\varphi^m\omega)\mathbb{1}_B(\omega)\right] \to \mathbb{E}(\mathbb{P}(A)\mathbb{1}_B(\omega)).$$

By the bounded convergence theorem, we conclude that

$$\frac{1}{n}\sum_{m=0}^{n-1} \mathbb{P}(\varphi^{-m}A \cap B) \to \mathbb{P}(A)\mathbb{P}(B).$$

There are exotic examples where $\mathbb{P}(\varphi^{-m}A \cap B) \to \mathbb{P}(A)\mathbb{P}(B)$ for all $A, B \in \mathcal{F}$ as $m \to \infty$; this is stronger than ergodicity of φ. □

4.24 May 5, 2020

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\varphi\colon \Omega \to \Omega$ be measure preserving. Let $A \in \mathcal{F}$.

Definition 4.24.1. We say $a \in A$ is *recurrent (with respect to A)* if $\varphi^n(a) \in A$ for some $n \geq 1$. △

Theorem 4.24.2 (Poincaré recurrence)**.** *Almost every point $a \in A$ is recurrent (with respect to A).*

Proof. Let

$$\begin{aligned} B &= \{b \in A\colon \varphi^n(b) \notin A \text{ for all } n \geq 1\} \\ &= A \setminus \bigcup_{n\geq 1} \varphi^{-n}(A). \end{aligned}$$

If $b \in B$ then $\varphi^n(B) \notin B$, so B is disjoint from $\varphi^{-n}(B)$ for all $n \geq 1$. In particular, $\varphi^{-k}(B)$ is disjoint from $\varphi^{-n}(B)$ for all $n > k$. So we have countably many disjoint sets $\varphi^{-k}(B)$; all of these sets have the same probability by measure-preservingness of φ. It follows that $\mathbb{P}(B) = 0$. □

Example 4.24.3 (Ehrenfest urn model)**.** We have two urns. There is a Markov chain on $\{0, 1, \ldots, N\}$, where state k represents k balls in the left urn and $N - k$ balls in the right urn. We pick a ball uniformly at random and move it to the other urn. This gives rise to transition matrix

$$P(k, k-1) = \frac{k}{N} \qquad \text{and} \qquad P(k, k+1) = \frac{N-k}{N},$$

with all other $P(i, j) = 0$.

Let's find the stationary distribution. Set $\pi(k) = \mathbb{P}(k \text{ balls in the left urn in the steady state})$. By definition, we have $\pi P = P$. [We got a few minutes in breakout rooms trying to solve this ourselves.]

It turns out that

$$\pi(k) = \frac{\binom{N}{k}}{2^N}.$$

We label balls $1, \ldots N$ and consider the state space $\{L, R\}^N$; the act of taking a ball and moving it to the other urn becomes a simple random walk on the hypercube $\{L, R\}^N$ (we are picking a coordinate at random and flipping it.)

The stationary distribution $\widetilde{\pi}(x_1, \ldots, x_n)$ is $\frac{1}{2^N}$ by symmetry. The original Markov chain is $Y_n = f(X_n)$, where $f(x_1, \ldots, x_n) = \#\{k\colon X_k = L\}$.

Now suppose we begin with the state N. How long does it take to return to the starting state? In other words, let $T = \min\{n \geq 1\colon Y_n = N\}$, so for example $\mathbb{P}(T = 1 | Y_0 = N) = 0$ and $\mathbb{P}(T = 2 | Y_0 = N) = \frac{1}{N}$. We want to compute $\mathbb{E}(T | Y_0 = N)$. To do this, we use Kac's theorem (Theorem 4.24.4) below. △

Let $\varphi\colon \Omega \to \Omega$ be measure preserving, ergodic, and invertible. (Invertibility is not such a strong assumption. Although the shift map $\varphi((\omega_n)_{n\in\mathbb{N}}) = (\omega_{n+1})_{n\in\mathbb{N}}$ is not invertible, the 2-sided shift map $\varphi((\omega_n)_{n\in\mathbb{Z}}) = (\omega_{n+1})_{n\in\mathbb{Z}}$ is.)

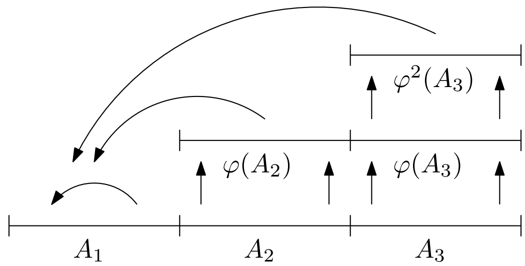
Let $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Let $M, N\colon \Omega \to \mathbb{N} \cup \{\infty\}$ be the random variables defined by

$$N(\omega) = \inf\{n \geq 1\colon \varphi^n(\omega) \in A\} \qquad \text{and} \qquad M(\omega) = \inf\{n \geq 0\colon \varphi^{-m}(\omega) \in A\}.$$

Theorem 4.24.4 (Kac's theorem on mean recurrence time)**.** *With the notation as above, we have*

$$\int_A N \, d\mathbb{P} = 1.$$

"Kakutani skyscraper". Let $A_n = \{a \in A\colon N(a) = n\}$. We construct a "skyscraper" whose ground (0-th) floor is $A = \sqcup_{k\geq 1} A_k$, and whose n-th floor is $\varphi^n(A) \setminus A = \sqcup_{k\geq 1} \varphi^n(A_{n+k})$ for $n \geq 1$:



All sets $\varphi^i(A_j)$ for $0 \leq i < j$ are disjoint since every $\omega^i(A_j)$ has $M(w) = i$ and $N(w) = j - i$. But M, N are a.s. finite by the Poincaré recurrence theorem (Theorem 4.24.2). So

$$\mathbb{P}(\Omega) = \sum_{0 \leq i < j} \mathbb{P}(\varphi^i(A_j)) = \sum_{0 \leq i < j} \mathbb{P}(A_j) = \sum_j j\mathbb{P}(A_j) = \int_A N \, d\mathbb{P}.$$

□

Example 4.24.5 (Example 4.24.3, cont.)**.** Let's apply Kac's theorem to a Markov chain $(Y_n)_{n\geq 0}$with countable state space S and stationary distribution π, for $A = \{a\}$ and $a \in S$. We have

$$1 = \int_A N \, d\mathbb{P} = \mathbb{E}(N\mathbb{1}_A) = \pi(a)\mathbb{E}(N|Y_0 = a).$$

It follows that

$$\mathbb{E}(N|Y_0 = a) = \frac{1}{\pi(a)}.$$

For the Ehrenfest urn,

$$\pi(a) = \frac{\binom{N}{a}}{2^N}, \qquad \text{so } \mathbb{E}(N|Y_0 = a) = \frac{1}{\pi(N)} = 2^N.$$

△

Let's consider ergodicity and dense orbits. Suppose Ω is a topological space with a countable base of open sets. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where $\mathcal{F}$ are the Borel sets of Ω. We assume further that $\mathbb{P}(U) > 0$ for all open U.

We say $\omega \in \Omega$ has a *dense orbit* under $\varphi\colon \Omega \to \Omega$ if $\{\omega, \varphi(\omega), \varphi^2(\omega), \dots\}$ is *dense* in Ω.

Theorem 4.24.6. *If φ is measure preserving and ergodic, then*

$$\mathbb{P}(\{\omega\colon \omega \text{ has a dense orbit under } \varphi\}) = 1.$$

Proof. Observe that ω has a dense orbit if and only if for all basic open sets U, there is n with $\varphi^n(\omega) \in U$. Thus, φ does not have a dense orbit if and only if there is a basic open U with $\varphi^n(\omega) \notin U$ for all n, i.e. $\omega \notin \varphi^{-n}U$ for all n. It follows that

$$\omega \in \bigcap_n (\varphi^{-n}U)^c$$

is in an invariant set, call it A, disjoint from U. Since A is invariant, we have $\mathbb{P}(A) \in \{0, 1\}$, but since $\mathbb{P}(U) > 0$, we conclude $\mathbb{P}(A) = 0$. □

4.25 May 7, 2020

We'll talk about Markov chain mixing, which will tell us about the rate of convergence in the ergodic theorem. It's a big topic, so today will be a little introduction.

Let $(X_n)_{n\geq 0}$ be a Markov chain on a finite or countable state space with transition matrix $p(x,y) = \mathbb{P}(X_{n+1} = y | X_n = x)$ and stationary distribution $\pi = \pi p$; here π is a row vector and p is a matrix.

If the chain is irreducible (i.e. for all $x, y \in S$ there's some n so that $p^n(x,y) > 0$) then the shift map is ergodic, and the ergodic theorem (Theorem 4.23.1) implies [][]

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \sum f(x)\pi(x) \qquad \text{a.s., as } n \to \infty. \tag{20}$$

Taking the case $f(x) = \mathbb{1}_y$, the convergence in Equation (20) becomes the convergence

$$\frac{1}{n}\#\{\text{visits to } y \text{ before time } n\} \to \pi(y).$$

How fast is this convergence? Can we remove the averaging?

Definition 4.25.1. A Markov chain $(X_n)_{n\geq 0}$ is called *aperiodic* if for all $x \in S$, we have

$$\gcd\{n \colon p^n(x,x) > 0\} = 1.$$

△

Example 4.25.2. The Markov chain on state space $S = \{x, y\}$ and transition matrix $p = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is not aperiodic, since the gcd is 2. △

Note that if $p(x,x) > 0$ for al x, then the chain is aperiodic. Given a Markov chain, there is an associated "lazy chain" with transition matrix $\frac{p+I}{2}$, so there is a $\frac{1}{2}$ chance of staying where you are. (This "slows" the Markov chain down.)

We have a following convergence theorem:

Theorem 4.25.3. *If p is both irreducible and aperiodic, then*

$$p^n(x,y) \to \pi(y) \qquad \text{as } n \to \infty$$

for all x.

How fast is this convergence? How far apart are the probability distributions $p^n(x,\cdot)$ and $\pi(\cdot)$?

The *total variation distance* can be thought of as a distance between random variables X and Y or between probability distributions μ and ν.

Definition 4.25.4. The *total variation difference* between two random variables X and Y with distributions μ and ν is defined by

$$\begin{aligned}\|\mu - \nu\|_{\text{TV}} \|X - Y\|_{\text{TV}} &= \inf\{\mathbb{P}\{X \neq Y\} \colon (X,Y) \text{ is a coupling of } X \text{ and } Y\} \\ &= \frac{1}{2}\sum_{x\in S} |\mu(x) - \nu(x)| \\ &= \sum_{x \colon \mu(x) > \nu(x)} \mu(x) - \nu(x) = \mu(A_*) - \nu(A_*) \\ &= \sup_A |\mu(A) - \nu(A)|,\end{aligned}$$

where $A_* = \{x \colon \mu(x) > \nu(x)\}$. △

We consider the distance from stationarity: let

$$d(t) \overset{\text{def}}{=} \max_{x\in S} \|p^t(x,\cdot) - \pi\|_{\text{TV}}.$$

(Note that we drop the variable y in p and π.) Also let

$$\overline{d}(t) = \max_{x,y\in S} \|p^t(x,\cdot) - p^t(y,\cdot)\|_{\text{TV}};$$

one can show that

Exercise: We have $d \le \overline{d} \le 2d$ and $\overline{d}(s)\overline{d}(t) \ge \overline{d}(s+t)$.

We define the *mixing time* as

$$t_{\text{mix}} \overset{\text{def}}{=} \min\{t\colon d(t) \le \varepsilon\}.$$

Then one can show

Exercise: We have $t_{\text{mix}}(\varepsilon) \le \lfloor \log_2(\frac{1}{\varepsilon})\rfloor \cdot t_{\text{mix}}(\frac{1}{4})$.

Consider the lazy random walk on the hypercube $V = \{0,1\}^n$, defined by transition matrix $p(\mathbf{x},\mathbf{y}) = 0$ unless $x_i = y_i$ for all or all but one $i \in \{1,\dots,n\}$, with

$$p(\mathbf{x},\mathbf{y}) = \begin{cases} \frac{1}{2} \text{ if } x_i = y_i \text{ for all } i \\ \frac{1}{2n} \text{ otherwise} \end{cases}$$

So half the time we are staying in place, and the other half we're doing a simple random walk. Note that $\pi(\mathbf{x}) = \frac{1}{2^n}$ for all $\mathbf{x} \in V$.

Now let N_t be the number of distinct coordinates chosen by time t. This N_t is a Markov chain, given by transition matrix

$$\mathbb{P}(N_{t+1} = k+1|N_t = k) = \frac{n-k}{n} \qquad \text{and} \qquad \mathbb{P}(N_{t+1} = k|N_t = k) = \frac{k}{n}.$$

Now let $\tau_k = \inf\{t\colon N_t = k\}$ (cf. *coupon collector problem*; see [6710, Example 5.12.2.2]). We have $\tau_{k+1} - \tau_k \sim \text{Geom}(\frac{n}{n-k})$. Thus,

$$\tau_n = \sum_{k=0}^{n-1}(\tau_{k+1} - \tau_k) \qquad \text{hence} \qquad \mathbb{E}\tau_n = \frac{n}{1} + \dots + \frac{n}{n} \approx n\log n.$$

We have

Lemma 4.25.5. *We have*

$$\mathbb{P}(\tau_n > n\log n + cn) < e^{-c}.$$

Proof. Let $A_i = \{\text{coordinate } i \text{ never chosen by time } n\log n + cn\}$. We have

$$\mathbb{P}\left(\bigcup A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i) = n\left(1 - \frac{1}{n}\right)^{n\log n + cn} \le ne^{-\log n - c} = e^{-c}.$$

□

At time τ_n (and any time $t > \tau_n$) we have $X_{\tau_n} \sim \pi$ It follows that $d(n\log n + cn) \le e^{-c}$, and it follows that $t_{\text{mix}}(\varepsilon) \le n\log n + \log(\frac{1}{\varepsilon})n$.

In fact, this bound can be improved to

$$t_{\text{mix}}(\varepsilon) \le \frac{1}{2}n\log n + \log\left(\frac{1}{\varepsilon}\right)n.$$

Consider the following thought experiment. Let $X = (x_1,\dots,x_n) \sim \pi = \text{Unif}(\{\cdot,1\})^n$ and $Y = (x_1',\dots,x_n')$ which are the same except k random coordinates are set to zero. Given $\{X,Y\}$, can you tell which one is which? More precisely, how large must $k = k(n)$ be so that you can tell with probability more than 0.51? [We got a few minutes in breakout rooms trying to solve this ourselves.]

One idea is to guess that whichever of $\{X,Y\}$ has more zeros is the spiked one. If $k \ll \sqrt{n}$ then this algorithm will get it wrong 50% of the time in the limit as $n \to \infty$. (We'll see next week that if $n - k \ll \sqrt{n}$ then we're already well mixed.)

Formalizing this, if you can't tell the difference between $X \sim \mu$ and $Y \sim \nu$ with probability more than $\frac{1}{2} + \varepsilon$, then $\|X - Y\|_{\text{TV}} \leq 2\varepsilon$. Indeed, we may define

$$A = \{\text{Your algorithm selects correctly } X \sim \mu \text{ and } Y \sim \nu \text{ instead of } X \sim \nu \text{ and } Y \sim \nu\}$$

and use that

$$\|X - Y\|_{\text{TV}} - \inf_A |\mu(A) - \nu(A)| < \left| \left(\frac{1}{2} + \varepsilon\right) - \left(\frac{1}{2} - \varepsilon\right) \right| = 2\varepsilon.$$

5 Presentations

5.26 May 12, 2020

[I missed a lot of key points, because I am slow. Sorry!]

Enlargement of Filtrations (Karen Grigorian)

A filtration $\mathcal{G} = (\mathcal{G}_t)_{t\geq 0}$ is called an *enlargement* of a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ if $\mathcal{F}_t \subset \mathcal{G}_t$ for all t.

A *semimartingale* is, roughly, the largest class of processes for which stochastic integration can be meaningful; formally, X is a semimartingale if $X = X_0 + M + V$, where M is an $\mathcal{F}$-local martingale and V is an $\mathcal{F}$-adapted process with finite variation and $M_0 = V_0 = 0$.

It's not true that every $\mathcal{F}$-martingale is a $\mathcal{G}$-martingale or even a $\mathcal{G}$-semimartingale. We are interested in conditions in $\mathcal{F}$ and $\mathcal{G}$ which would ensure:

$$\text{Every } \mathcal{F}\text{-martingale is a } \mathcal{G}\text{-martingale} \qquad (\mathcal{H})$$

or

$$\text{Every } \mathcal{F}\text{-martingale is a } \mathcal{G}\text{-semimartingale} \qquad (\mathcal{H}')$$

Stricker's theorem says that any $\mathcal{G}$-semimartingale is an $\mathcal{F}$-semimartingale; this is a far reaching generalization of a homework problem we worked on.

There are two main kinds of enlargement of filtrations:

- "Initial Enlargement": $\mathcal{G} = \mathcal{F} \vee \sigma(\zeta)$, where ζ is a random variable.
- "Progressive Enlargement": the smallest $\mathcal{G}$ that turns a given positive random variable τ into a stopping time.

[I might have missed the main theorem, which probably would've appeared here. Sorry!]

Example 5.26.1. Let B be a Brownian motion and let $\mathcal{F}$ its natural filtration; let $\mathcal{F}^{\sigma(B_1)} = \mathcal{F} \vee \sigma(B_1)$. Then $(\mathbb{E}(B_t|\mathcal{F}^{\sigma(B_1)}))_{t\geq 0}$ is a $\mathcal{F}^{\sigma(B_1)}$-semimartingale. △

Example 5.26.2. Let B be a standard Brownian motion under its natural filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$. Define $\mathcal{G} = (\mathcal{G}_t)_{t\geq 0}$ by $\mathcal{G}_t = \mathcal{F}_{t+\varepsilon}$ for some $\varepsilon > 0$. Then B is not a $\mathcal{G}$-semimartingale. This is because it can be shown that B is not a good $(\mathcal{G}, \mathbb{P})$-integrator, i.e. you cannot properly define a stochastic integral with respect to B; then use the fact that the good $(\mathcal{G}, \mathbb{P})$-integrators are precisely the $\mathcal{G}$-semimartingales. △

There is an application of these ideas to finance as follows. The high level idea is that enlargement of filtrations can lead to arbitrage opportunities. The first fundamental theorem of asset pricing says that for a semimartingale price process, having no arbitrage is equivalent to certain explicit probabilistic properties; if enlarging the filtration (the information set) destroys semimartingaleness, then there is likely to be arbitrage opportunities for insiders.

The Karhunen Loéve Theorem (Sara Venkatraman)

The theorem says that an L^2 continuous-time stochastic process $\{X_t\}_{t\geq 0}$ on $[0,1]$ can be expressed as an infinite linear combination of orthogonal functions. (This is analogous to representing a function on a bounded interval as a Fourier series.)

A continuous function $K : [0,1] \times [0,1] \to \mathbb{R}$ is a kernel if it is symmetric and positive semidefinite:

$$\sum_{i,j} K(x_i, x_j) c_i c_j \geq 0$$

for all $x_i \in [0,1]$ and $c_i \in \mathbb{R}$. Given a kernel K we obtain a linear operator

$$(T_K f)(x) \stackrel{\text{def}}{=} \int_0^1 K(x,s) f(s)\, ds$$

which has eigenvalues and eigenfunctions. That is, we search for f and λ with

$$\int_0^1 K(s,t) f(s)\, ds = \lambda f(t).$$

Theorem 5.26.3 (Mercer)**.** *The eigenfunctions* $\{e_k(t)\}_{k\geq 1}$ *of* T_K *are an orthonormal basis for* L^2 *and the eigenvalues* $\{\lambda_k\}_{k\geq 1}$ *are nonnegative. We can write* K *as*

$$K(s,t) = \sum_k \lambda_k e_k(s) e_k(t).$$

For an $L^2[0,1]$ stochastic process, the covariance function $K(s,t) := \text{cov}(X_s, X_t)$ is a kernel. Then we can expand X_t in the eigenbasis given by K. The Karhunen-Loéve theorem says:

Theorem 5.26.4 (Karhunen-Loéve)**.** *If* $\mathbb{E}X_t = 0$ *for all* $t \in [0,1]$, *then the coefficients* Z_k *in the expansion*

$$X_t = \sum_k Z_k e_k(t)$$

are given by

$$Z_k = \int_0^1 X_t e_k(t)\, dt,$$

and furthermore $\mathbb{E}Z_k = 0$, $\text{var}(Z_k) = \lambda_k$, $\text{cov}(Z_j Z_k) = 0$.

We have seen that for a standard Brownian motion, the covariance function $K(s,t) = \min(s,t)$. One can work out the eigenvalues/eigenfunctions explicitly to obtain

Corollary 5.26.5. *For every* $t \in [0,1]$, *we have*

$$B_t = \sqrt{2} \sum_{k=1}^{\infty} Z_k \frac{\sin((k-\frac{1}{2})\pi t)}{(k-\frac{1}{2})\pi}, \qquad \text{where } Z_k \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

Brownian motion and the heat equation (Emily Dautenhahn)

The heat equation is the PDE

$$u_t = \frac{1}{2}\Delta u$$

where $u(t,x)\colon (0,\infty)\times\mathbb{R}^n \to \mathbb{R}$ and Δ is the Laplacian. Solutions of the heat equation is a smooth (C^2) u satisfying the heat equation as well as the boundary condition $u(0,x) = f(x)$ for some bounded continuous f.

Theorem 5.26.6. *The above equation has a unique bounded solution*

$$u(t,x) = \mathbb{E}_x f(B_t) = \int_{\mathbb{R}^n} \frac{1}{(2\pi t)^{n/2}} \exp\left(\frac{-|x-y|^2}{2t}\right) f(y)\, dy.$$

We generalize the heat equation setup to include a dissipation term V, and replace the spatial domain $U = \mathbb{R}^n$ with an open bounded subset $U \subset \mathbb{R}^n$. [I missed exactly what the PDE was.]

Theorem 5.26.7 (Feynman-Kac Formula 1)**.** *Suppose* $V\colon \mathbb{R}^n \to \mathbb{R}$ *is bounded. Then* $u\colon [0,\infty)\times\mathbb{R}^n \to \mathbb{R}$ *defined by*

$$u(t,x) = \mathbb{E}_x\left[\exp\left(\int_0^t V(B(r))\, dr\right)\right]$$

solves the heat equation on $\mathbb{R}^n$ *with dissipation rate* V *and initial condition one.*

This formula follows from a direct, though involved, computation. (Take a Taylor expansion of the exponential, consider things termwise, etc.)

Theorem 5.26.8 (Feynman-Kac Formula 2)**.** *If* u *is a bounded and sufficiently smoth solution of the heat equation on* U, *with zero dissipation and continuous initial condition* g, *then*

$$u(t,x) = \mathbb{E}_x[g(B(t))\mathbb{1}_{\{t<\tau\}}],$$

where τ *is the first exit time from the domain* U.

This formula requires stochastic calculus and Itô's formula:

Proposition 5.26.9. *Let f be smooth. Then with probability 1, for all $t \geq 0$,*

$$f(t, B_t) - f(0, B_0) = \int_0^t \frac{\partial f}{\partial t}(s, B_s)\, ds + \sum_{i=1}^d \int_0^t D_i f(B_s)\, dB_s^i + \frac{1}{2}\sum_{i=1}^d \int_0^t D_{ii} f(B_s)\, ds.$$

This formulas will show that $M_s = u(t-s, B_s)$ is a martingale.