18.676: Stochastic Calculus

Lecturer: Professor Nike Sun

Notes by: Andrew Lin

Spring 2020

Introduction

Most of the logistical information is on the class website at [1], including an official class summary and many references to relevant papers and textbooks. Here are the main points for us: there will be homework roughly once every two weeks. The first two are already posted, and they'll be due February 12 and February 24 (submitted in class). Grading is weighted 55 percent for homework, 20 percent per exam, and 5 percent for attendance. Office hours are Monday 2–4 in Professor Sun's office, 2-432.

18.675 is a prereq, so we should talk to Professor Sun if we haven't taken that class. We will be using [2] as our main textbook.

1 February 3, 2020

Today, we'll begin with an informal overview of the topics covered in this class. We'll start with some basic reminders: the **standard Gaussian density**

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

should be burned into our head, and the variable $Z \sim N(0,1)$ is distributed according to this density. We should know that if X, X_i are **symmetric random signs** (± 1 with equal probability), and $S_n = \sum_{i=1}^n X_i$, then $\frac{S_n}{\sqrt{n}}$ converges in distribution to $N(0,1)$. We should also know how to prove this, either using the central limit theorem or by direct combinatorial calculation (because S_n is a scaling of the binomial distribution).

Next, we can consider the **simple random walk** on the integers, which gives us a process $(S_n)_{n \geq 0}$ (where n is a time index): since $\frac{S_n}{\sqrt{n}}$ converges in distribution to a Gaussian, this means that over time n, the walk typically covers a distance on the order of $\sqrt{n}$. So if we rescale time by n and rescale space by $\sqrt{n}$, we get a process

$$X^{(n)}(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}.$$

For any fixed t, we still have the central limit theorem as before, which tells us that $X^{(n)}(t) \xrightarrow{d} N(0,t)$. But one idea throughout this class will be that we don't need to consider a single t: the entire **process** $X^{(n)} = \left(X^{(n)}(t)\right)_{t \geq 0}$ converges in distribution to $(B_t)_{t \geq 0}$, something called a **Brownian motion**, as $n \to \infty$. Note that we haven't defined a Brownian motion yet, and we haven't described the topology in which this converges in distribution. But we'll do everything more formally later on.

In short, here are some of the main goals of this class:

- Formal construction of Brownian motion,
- Convergence of some natural processes (like simple random walk), which we can think of as a "functional CLT,"
- Calculations with Brownian motion (stochastic calculus).

For now, though, we'll keep surveying some more ideas from the course: we're going to talk a bit about Itô's formula and give an application to the conformal invariance of planar Brownian motion.

Example 1

First, we'll describe some properties of our Brownian motion B_t given our informal definition above.

We should have $B_0 = 0$, and for any $0 \le s \le t$, we should have $B_t - B_s \sim N(0, t-s)$. Also, for any $0 \le s_1 \le t_1 \le s_2 \le t_2$, $B_{t_1} - B_{s_1}$ and $B_{t_2} - B_{s_2}$ should be independent (because they correspond to disjoint parts of the random walk). And in fact, just these properties actually suffice to characterize Brownian motion completely.

Example 2

Next, we'll do a conceptual overview of Itô's formula. Consider a process that evolves as

$$dX_t = \mu_t dt + \sigma_t dB_t.$$

Informally, we can think of writing this as

$$X_{t+dt} - X_t = \mu_t dt + \sigma_t \cdot N(0, dt).$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a twice-differentiable function. If X_t followed a deterministic smooth trajectory, then we would know how $f(X_t)$ evolves, since we just have $df(X_t) = f'(X_t)dX_t$. But if we expand the stochastic version out, we instead find that

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2 \\ &= f'(X_t)\left(\mu_t dt + \sigma_t dB_t\right) + \frac{f''(X_t)}{2}\left(\mu_t dt + \sigma_t dB_t\right)^2. \end{aligned}$$

Because dB_t is on the order of $\sqrt{dt}$, it dominates the $\mu_t dt$, so we can replace $\left(\mu_t dt + \sigma_t dB_t\right)^2$ with just $\sigma_t^2 (dB_t)^2 = \sigma_t^2 dt \cdot N(0,1)^2$. What Itô's formula says is basically that we can actually ignore the fluctuations in the $N(0,1)^2$ term if we take many measurements, and so that just disappears from the expression (it's 1 on average). Thus,

$$\begin{aligned} df(X_t) &= f'(X_t)\left(\mu_t dt + \sigma_t dB_t\right) + \frac{f''(X_t)}{2}\sigma_t^2 dt \\ &= \boxed{\left(f'(X_t)\mu_t + \frac{f''(X_t)\sigma_t^2}{2}\right)dt + f'(X_t)\sigma_t dB_t}, \end{aligned}$$

and we've now separated the contribution into a **drift** and a **stochastic** term.

Using this, we can do an application to planar Brownian motion – first, we'll review a bit of complex analysis. If we have a function $f : \mathbb{C} \to \mathbb{C}$ or $f : D \to \mathbb{C}$ for some open D, then f is **holomorphic** or **complex differentiable** at $z \in \mathbb{C}$ if the complex derivative

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h} \in \mathbb{C}$$

exists. (Being complex differentiable is much stronger than being differentiable in $\mathbb{R}^2$ because we approach 0 in all directions in the complex plane.) If we think of our function as going from $\mathbb{R}^2 \to \mathbb{R}^2$, where $z = x + iy$ and $f = u + iv$

(so that u, v are real-valued functions and x, y are real numbers), then f is holomorphic at z if the limits from the real axis and imaginary axis are the same, meaning that we require

$$\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}, \quad \frac{1}{i}\frac{\partial f}{\partial y} = \frac{1}{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$$

to be equal. Thus, we have the **Cauchy-Riemann equations**

$$u_x = v_y, \quad u_y = -v_x.$$

One useful thing to know is that the Laplacian of the real part of f is

$$\Delta u = u_{xx} + u_{yy} = v_{xy} - v_{yx} = 0,$$

which means that the real part of any complex differentiable function is harmonic (and so is the imaginary part by an analogous calculation).

Example 3

Consider a two-dimensional (standard) Brownian motion (X_t, Y_t) – this just means that X_t and Y_t are independent standard one-dimensional Brownian motions. Alternatively, we can take $Z_t = X_t + iY_t$ to be a standard Brownian motion in $\mathbb{C}$.

Suppose we have a **conformal** map $f : D \to D'$ (which means that f is holomorphic and has a holomorphic inverse $f^{-1} : D' \to D$). Again, let $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$.

Question 4. *How does $f(Z_t)$ evolve if Z_t stops when it hits the boundary of D?*

We'll need a two-dimensional version of Itô's formula for this, but the same Taylor expansion idea works:

$$du(Z_t) = \begin{bmatrix} u_x(Z_t) & u_y(z_t) \end{bmatrix} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix} + \frac{1}{2} \begin{bmatrix} dX_t & dY_t \end{bmatrix} \begin{bmatrix} u_{xx}(Z_t) & u_{xy}(Z_t) \\ u_{yx}(Z_t) & u_{yy}(Z_t) \end{bmatrix} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix}.$$

When we expand this out, we get cross terms like $dX_t \cdot dY_t$, which look like $dt \cdot N(0,1)\tilde{N}(0,1)$. If we add up many of these, they cancel out and become negligible, so we don't have to worry about those: this means the second term will only have the diagonal term contributions

$$\frac{1}{2}\left(u_{xx}(dX_t)^2 + u_{yy}(dY_t)^2\right),$$

but now we can replace $(dX_t)^2$ and $(dY_t)^2$ with dt by the same argument as above, and now $u_{xx} + u_{yy} = 0$ because u is harmonic. So the entire second-order term actually vanishes, and we're just left with (now doing the same calculations for $v(Z_t)$)

$$\boxed{\begin{bmatrix} du(Z_t) \\ dv(Z_t) \end{bmatrix} = \begin{bmatrix} u_x(Z_t) & u_y(Z_t) \\ v_x(Z_t) & v_y(Z_t) \end{bmatrix} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix}}.$$

The 2-by-2 matrix on the right-hand side can be thought of as

$$G = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \begin{bmatrix} u_x & u_y \\ -u_y & u_x \end{bmatrix} = \sqrt{u_x^2 + u_y^2}(\text{rotation matrix}),$$

but $\sqrt{u_x^2 + u_y^2}$ is just the modulus of $f'(z) = u_x + iv_x$, and the determinant of the rotation matrix has to be 1 (we have a conformal map). So we can conclude that if $Z_t = X_t + iY_t$ is a standard Brownian motion in $D \subseteq \mathbb{C}$, and

$f : D \to D'$ is conformal, then as long as Z_t is in D, $f(Z_t)$ evolves via

$$\boxed{\begin{bmatrix} du(Z_t) \\ dv(Z_t) \end{bmatrix} = |f'(Z_t)| O(Z_t) \begin{bmatrix} dX_t \\ dY_t \end{bmatrix}},$$

where $O(Z_t)$ is a 2×2 rotation matrix. Note that the standard bivariate Gaussian $N(0, I_{2\times 2})$ is rotationally invariant (spherically symmetric), so it is reasonable to believe that standard Brownian motion is also rotationally invariant: if O is a 2×2 orthogonal matrix and Z is a BM in $\mathbb{R}^2$, then so is OZ.

We can also consider scaling: if B_t is a Brownian motion, then σB_t is equal in distribution to $B_{\sigma^2 t}$. This is clearly true for any fixed t because they're both Gaussian, but in fact the idea is that we actually can't tell the two images (of the sample paths) apart. So the Brownian motion is a **self-similar fractal**.

Remark 5. *Note that in the formula we've derived, the scale factor and the rotation depend on the given time. (This means, for example, that when $|f'(Z_t)|$ is big, the process runs faster.) So the* **process** *itself is not conformally invariant, but the* **trace** *(the image of the motion) is indeed conformally invariant.*

With the rest of the time today, we'll give an example of a question that this class will let us answer:

Example 6

Consider a simple random walk on a grid with ε spacing on $(\varepsilon\mathbb{Z}) \times (\varepsilon\mathbb{Z}_{\geq 0})$. Suppose that we start our walk near $(0, y)$ and stop when we hit the horizontal axis. What is the law of the hitting location (the x-coordinate)?

We'll think about this problem when ε is small, since we can approximate this walk with a Brownian motion in the upper-half (complex) plane. One way to approach this is to map the problem into an easier domain: the function $f(z) = \frac{i-z/y}{i+z/y}$ maps $\mathbb{H}$ conformally into the unit disk $\mathbb{D}$, and our new starting point is now the origin. Since we only care about the hitting location (and not the time), and the Brownian motion is spherically symmetric, the **hitting location must be uniform on** $\partial\mathbb{D}$, which means the distribution on $\mathbb{H}$ can be easily recovered.

More explicitly, for any interval $[a, b] \in \mathbb{R}$, we can calculate the probability of hitting within that interval explicitly – the angle covered near x is proportional to $\frac{|f'(x)|}{2\pi} = \frac{y}{\pi(x^2+y^2)}$ (the factor of 2π coming from the length of the boundary of the unit disk), meaning that

$$\begin{aligned} \mathbb{P}_{\mathbb{H}}(Z_\tau \in [a, b]) &= \mathbb{P}_{\mathbb{D}}(\tilde{Z}_\sigma \in f([a, b])) \\ &= \int_a^b \frac{y\, dx}{\pi(x^2 + y^2)}. \end{aligned}$$

This last integrand is also called the **Poisson kernel** for $\mathbb{H}$ (we'll write it as $P_y(x) = \frac{y}{\pi(x^2+y^2)}$), because it's closely connected to the Dirichlet problem on $\mathbb{H}$, which goes as follows. If we're given a nice function $b : \mathbb{R} \to \mathbb{R}$ and we want to know the harmonic interpolation of b to $\mathbb{H}$, the answer is given by

$$h(x, y) = \mathbb{E}[b(Z_\tau)|Z_0 = x + iy];$$

that is, we start a Brownian motion at $x + iy$ and find the expected value of b when it hits the boundary. (We can prove this by looking at a finite graph or with direct calculation, and we'll talk about it more later.) But explicitly, this can be written as an integral

$$\mathbb{E}[b(Z_\tau + x)|Z_0 = iy] = \int_{\mathbb{R}} P_y(s) b(s + x) ds.$$

Since P_y is symmetric, we can replace s with $-s$, and thus this expected value becomes a convolution $(b * P_y)(x)$ with the Poisson kernel. This is something we'll see come up in the future as well!

2 February 5, 2020

Today, we'll start to formalize some of the ideas from yesterday's informal overview – specifically, we'll be starting with Gaussian processes and Gaussian spaces. (We'll follow the textbook pretty closely for now.) All random variables will live on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 7

A ***d*-dimensional Gaussian vector** is an $\mathbb{R}^d$-valued random variable X such that $\langle X, u\rangle$ is a one-dimensional Gaussian variable for any $u \in \mathbb{R}^d$.

This is somewhat fancier than other definitions and doesn't depend on a choice of basis. In addition, this definition doesn't specify that $\langle X, u\rangle$ and $\langle X, v\rangle$ need to be jointly Gaussian for vectors u, v, but we'll see that it is a consequence of the definition.

Proposition 8

The law of X is uniquely determined by the mean vector $\mu = \mathbb{E}[X] \in \mathbb{R}^d$ and the covariance matrix $\Sigma = \mathbb{E}[(X-\mu)(X-\mu)^T] \in \mathbb{R}^{d\times d}$.

Proof. Take any $\theta \in \mathbb{R}^d$. By definition, $\langle X, \theta\rangle$ is Gaussian with some mean and variance, but we can compute those from μ and Σ:

$$\mathbb{E}[\langle X, \theta\rangle] = \langle \mathbb{E}[X], \theta\rangle = \langle \mu, \theta\rangle$$

by linearity of expectation, and

$$\mathrm{Var}(\langle X, \theta\rangle) = \mathrm{Cov}(\langle X, \theta\rangle, \langle X, \theta\rangle) = \theta^T \Sigma \theta$$

because covariance is bilinear. This means that we know the full distribution of $\langle X, \theta\rangle$, which tells us the characteristic function

$$\phi_X(\theta) = \mathbb{E}\left[\exp\left(i\langle X, \theta\rangle\right)\right] = \exp\left(i\langle \mu, \theta\rangle - \frac{1}{2}\theta^T\Sigma\theta\right),$$

and as we saw in 18.675, knowing all of these characteristic functions is enough to determine the law of X. □

Because of this, we'll use the notation $X \sim N(\mu, \Sigma)$, where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ is a symmetric positive semidefinite matrix (because variance is always nonnegative). Any such matrix has a Cholesky factorization $\Sigma = AA^T$, where A is a $d \times r$ matrix, and then if Z is standard normal in r dimensions (that is, iid standard Gaussian in each component), then we have

$$\mu + AZ \sim N(\mu, \Sigma).$$

(Indeed, we can check that this has the right mean and variance by expanding out the matrix multiplication for AZ and using that $\mathbb{E}[Z_a Z_b] = \delta_{ab}$.) In particular, we can write $\Sigma = AA^T$ for an invertible square matrix A if and only if Σ has full rank, and in this case only we can use the change of variables formula to find that X has density

$$f(x) = \frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{(2\pi)^{d/2}|\det\Sigma|^{1/2}}.$$

(If $r < d$, then the law of X is supported on a subspace of $\mathbb{R}^d$ of lower dimension, so it has no density.)

One important fact is that for Gaussians, being independent and being uncorrelated are the same thing:

Lemma 9
If $X \sim N(\mu, \Sigma)$ is a d-dimensional Gaussian vector, then X_i are mutually independent if and only if Σ is diagonal.

In contrast, there's the standard non-example where we consider $Z \sim N(0,1)$ and $\varepsilon \sim \text{Unif}(\{\pm 1\})$. Then the covariance between εZ and Z is zero, but the two variables aren't independent because they always have the same absolute value (basically, this goes wrong because the two variables aren't **jointly** Gaussian).

Proof. The forward direction is easy (independent implies uncorrelated). For the other direction, suppose Σ is a diagonal matrix. Then the characteristic function

$$\phi_X(\theta) = \prod_{j=1}^{d} \exp\left(i\theta_j - \frac{1}{2}\sigma_{jj}\theta_j^2\right)$$

has no cross-terms, so the characteristic polynomial factorizes and thus the different components are independent as desired. □

With this, we'll move on to the idea of a **Gaussian space** on $(\Omega, \mathcal{F}, \mathbb{P})$. (For the rest of today, we'll assume Gaussians are **centered**, meaning they have mean zero.) Recall that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all ($\mathbb{R}$-valued) random variables on our probability space with finite second moment – this is a Hilbert space with inner product

$$\langle X, Y \rangle_{L^2(\Omega,\mathcal{F},\mathbb{P})} = \int_\Omega X(\omega)Y(\omega)d\mathbb{P}(\omega) = \mathbb{E}[XY].$$

Definition 10
A **(centered) Gaussian space** is a closed linear subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ containing only centered Gaussian variables.

Example 11
Take $X \sim N(0, \Sigma)$ in $\mathbb{R}^d$. Then the span of the coordinate random variables $\{X_1, \cdots, X_d\}$ is a Gaussian space. Meanwhile, a non-example is the span of $Z \sim N(0,1)$ and εZ (where ε is the random sign as before); this doesn't work because $Z + \varepsilon Z$ is zero half the time and thus not Gaussian.

Gaussian spaces are important because they **turn probability into geometry**; for example, independence of Gaussian variables becomes orthogonality in the space:

Theorem 12
Let $H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ be a centered Gaussian space, and let $(H_\alpha)_{\alpha \in I}$ be linear subspaces of H. Then the σ-fields $(\sigma(H_\alpha))_{\alpha \in I}$ are independent if and only if the H_α are pairwise orthogonal.

We can read the book for this – it's really a fancier version of Lemma 9. A related point is that conditional expectation among Gaussians corresponds to orthogonal projection in a Gaussian space. By the way, the first hypothesis is important here – we want that the H_α are all **subspaces of a single centered Gaussian space** to ensure that they are jointly Gaussian.

Problem 13

Suppose we have a bivariate normal (two-dimensional Gaussian) given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a & b \\ b & c \end{bmatrix} \right),$$

where we assume that the covariance matrix is positive definite. What is the law of Y conditional on X?

The standard trick used here is to find some θ such that $Y - \theta X$ is independent from X. Here, $Y - \theta X$ is also jointly Gaussian with X and Y, so we just need to find what value of θ satisfies

$$0 = \text{Cov}(Y - \theta X, X) = b - \theta a,$$

meaning we should set $\theta = \frac{b}{a}$. Then we can break Y up into a "parallel" and a "perpendicular" part as

$$Y = \frac{b}{a} X + \left(Y - \frac{b}{a} X \right),$$

where the first term is in $\sigma(X)$ and the second term (independent of X) is a Gaussian with variance

$$\text{Var}\left(Y - \frac{b}{a} X \right) = \text{Cov}\left(Y - \frac{b}{a} X, Y \right) = c - \frac{b^2}{a}.$$

Putting everything together, the conditional distribution is given by

$$\boxed{Y|X \sim N\left(\frac{bX}{a}, c - \frac{b^2}{a} \right)}.$$

More generally (as an exercise for us), if

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \right),$$

where X lives in a k-dimensional space and C lives in an ℓ-dimensional space, then we can check that $Y|X$ is distributed as $N(B^T A^{-1} X, C - B^T A^{-1} B)$.

Theorem 14

Let $H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ be a centered Gaussian space, and let $K \subseteq H$ be a closed subspace. Then for any $X \in H$, we have

$$X|\sigma(K) = N(\pi_K(X), \mathbb{E}[(X - \pi_k(X))^2]),$$

where π_K denotes the orthogonal projection of X onto K.

In geometric terms, the mean is the "parallel" part, and the variance is the "perpendicular" part. (Also, $\pi_K(X)$ is measurable with respect to $\sigma(K)$, so it is indeed known once we condition on the sigma-algebra.) In contrast, notice that for a general $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ (not necessarily Gaussian), we only know that

$$\mathbb{E}(X|\sigma(K)) = \pi_{L^2(\Omega, \sigma(K), \mathbb{P})}(X).$$

In particular, $\sigma(K)$ is generally very big – if K is the span of some variable Z, then $\sigma(K)$ is the set of all measurable functions of Z. So this theorem tells us that we can project onto a much smaller subspace in the Gaussian case.

Problem 15 (Kalman filter; on homework)

Suppose we have independent Gaussians $\varepsilon_n \sim N(0, \sigma^2)$ and $\eta_n \sim N(0, \delta^2)$ and we have some true unknown state of a system which evolves over time:

$$0 = X_0 \to X_1 \to X_2 \to \cdots \to X_n, \quad X_{n+1} = a_n X_n + \varepsilon_{n+1}.$$

Suppose we're given a noisy observation Y_i at each time satisfying $Y_n = cX_n + \eta_n$. (Assume that we know the values of a_n and c and σ^2 and δ^2.) Our goal is to find $\mathbb{E}[X_n | Y_1, \cdots, Y_n]$.

One approach is to define a Gaussian space H which is the span of the ε_i and η_i up to some time n. (In other words, this is all of the noise going into the system up to time n.) From the way this is designed, all of the X_i and Y_i are linear combinations of the εs and ηs, so we always stay in the Gaussian space throughout the evolution. Then if we want $\mathbb{E}[X_n | Y_1, \cdots, Y_n]$, we're doing an orthogonal projection – in particular, X_n must be a linear combination of Y_1 up to Y_n.

We'll next discuss Gaussian processes, which are generalizations of Gaussian vectors:

Definition 16

Let I be an arbitrary (possibly uncountable) index set. A collection of random variables $(X_t)_{t \in I}$ is a **(centered) Gaussian process** if any finite linear combination of the X_ts is a one-dimensional Gaussian. The **Gaussian space generated by** $(X_t)_{t \in I}$ is the closure of the linear span of the X_ts. Define the **covariance function** $\Gamma : I \times I \to \mathbb{R}$ via $\Gamma(s, t) = \mathrm{Cov}(X_s, X_t)$.

The covariance function Γ is symmetric and positive semidefinite, meaning that $\sum_{s,t} \theta(s)\theta(t)\Gamma(s, t) \geq 0$ for any θ which is nonzero for **finitely many values** in I (otherwise, this statement may not make any sense), since this expression is just the variance of $\sum_s \theta(s) X_s$.

The natural follow-up question is whether there necessarily exists a Gaussian process with a given (symmetric, positive semidefinite) covariance function. The answer is yes, and this follows basically from the Kolmogorov extension theorem (which we proved in 18.675, so we won't do now). The most important example for us will be the construction of Brownian motion based on a Gaussian process with index set $\mathbb{R}_{\geq 0}$, where the function takes the form

$$\Gamma(s, t) = \mathrm{Cov}(B_s, B_t) = \min(s, t)$$

(because if WLOG $s < t$, then $B_t = B_s + (B_t - B_s)$, and the second term is independent of B_s). But we don't want to immediately cite the Kolmogorov extension theorem now, because we want to make sure that with probability 1, our process is continuous in t – that is, for all $\omega \in (\Omega, \mathcal{F}, \mathbb{P})$, $B_t(\omega)$ is continuous in t. Specifically, the theorem will give us a measure on $(\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu)$, but that's not really the space we want to use – we want the space of continuous functions instead. So we'll come back to this a little later.

On our homework, though, there's a different construction of Brownian motion based on the construction of "white noise," which is what we'll discuss for the rest of class. The heuristic idea is that on every "pixel" of space, we see an independent Gaussian random variable, so we get "snow on a TV screen." Here's a more formal definition:

Definition 17

Let $(E, \mathcal{E})$ be a measurable space, and let μ be a σ-finite measure on $(E, \mathcal{E})$. Then a **Gaussian white noise** on $(E, \mathcal{E})$ **with intensity** μ is a linear isometry $G : L^2(E, \mathcal{E}, \mu) \to H$, where $H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a centered Gaussian space.

An **isometry** preserves the inner product, so $\langle f, g\rangle = \langle G(f), G(g)\rangle$. (In other words, the covariance between $G(f)$ and $G(g)$ should be the same as the inner product between f and g.) So for any "patch" of size dx around x, we assign a Gaussian variable $N(0, \mu(dx))$, and we do this for all points independently. So if we take a subset $A \subseteq E$, we can think of the white noise being distributed as the "sum of all of these small Gaussian random variables:"

$$G(1_A) = G(A) \sim N(0, \mu(A)).$$

This is an isometry because A and B being disjoint means we have independence $G(A) \perp\!\!\!\perp G(B)$, so

$$0 = \langle 1_A, 1_B\rangle = \mathrm{Cov}(G(1_A), G(1_B)).$$

In the informal language of "adding up infinitesimal Gaussian random variables," we can write $G(f) = \sum_{x\in E} f(x)Z_x$ where $Z_x = N(0, \mu(dx))$, which helps us see that

$$\mathrm{Cov}(G(f), G(g)) = \mathrm{Cov}\left(\sum_x f(x)Z_x, \sum_x g(x)Z_x\right) = \sum_x f(x)g(x)\mu(dx)$$

(only the diagonal terms emerge), which gives exactly the inner product between f and g. But of course, this doesn't exactly make sense without the formal definition.

However, even when we have this definition, we still need to ask whether we can actually construct such an object. The last question on our homework asks us to construct an explicit white noise G and in fact construct Brownian motion from that via $B_t = G(1_{[0,t]})$. This Brownian motion will have the right covariance properties, and we can show that B is continuous almost surely – this is actually the historically older construction of Brownian motion.

We'll finish by contrasting all of this with something else we might have seen: compare this Gaussian white noise of $N(0, \mu(dx))$ to the **Poisson random measure**, where for each patch of dx, we assign a Bernoulli random variable of parameter $\mu(dx)$. Then taking subsets $A \subseteq E$, we'll get a **normal distribution** in the white noise case with variance $\mu(A)$, but a **Poisson distribution** of parameter $\mu(A)$ in the Poisson random measure case.

3 February 10, 2020

Last time, we talked about the general definitions of finite-dimensional Gaussian vectors, Gaussian spaces, Gaussian processes, and Gaussian white noise. Today, we'll go through about the construction of Brownian motion, but we'll need to cover a few preliminaries first.

Recall that if we have a covariance function $\Gamma : I \times I \to \mathbb{R}$ which is symmetric and positive semidefinite, then there exists a Gaussian process $(X_t)_{t\in I}$ with covariance function $\mathrm{Cov}(X_s, X_t) = \Gamma(s, t)$. (We will often be working with $I = [0, \infty)$.) We showed this with the **Kolmogorov extension theorem** – informally, the idea is that to define a measure on $\mathbb{R}^I$, we need to be able to write down the joint law for any finite subset of I in a consistent way, and then the Kolmogorov extension theorem gives us the measure $(\mathbb{R}^I, \mathcal{B}_{\mathbb{R}}^{\otimes I}, \nu)$. But in our case, if we're given the values of $\Gamma(s, t)$, then Γ yields a covariance matrix for every finite collection of points in I.

Definition 18

In the special case where $I = [0, \infty)$ and $\Gamma(s, t) = \min(s, t)$, the resulting Gaussian process $(X_t)_{t\geq 0}$ is called a **pre-Brownian motion**.

Here's a quick connection to the material from the end of last lecture:

Proposition 19

Let $(X_t)_{t\geq 0}$ be a real-valued stochastic process (any collection of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ indexed by t). Then the following are equivalent:

- X is a pre-Brownian motion,
- We can express $X_t = G([0,t]) = G(1_{[0,t]})$, where G is a Gaussian white noise on $I = [0,\infty)$ with intensity equal to the Lebesgue measure.

Proof. The backwards direction follows directly from the definition of a Gaussian white noise – G is defined to be an isometry, so we do get the correct covariance function. For the forward direction, we're given a pre-Brownian motion, and we need to construct an isometry. If f is a step function of the form

$$f(t) = \sum_{i=1}^{n} a_i 1(t_{i-1}, t_i],$$

then we define

$$G(f) = \sum_{i=1}^{n} a_i (X_{t_i} - X_{t_{i-1}}).$$

We can check that G is an isometry on this class of step functions – specifically, if $h(t) = \sum_{i=1}^{n} b_i 1\{t \in (t_{i-1}, t_i]\}$ (without loss of generality we can assume the two functions have the same break points t_i), then the covariance can be computed as

$$\langle G(f), G(h)\rangle = \sum_{i=1}^{n} a_i b_i (t_i - t_{i-1}) = \langle f, h\rangle$$

because distinct increments of X are independent by the definition of a pre-Brownian motion. So we do have an isometry G from the step functions to the Gaussian space H spanned by X, and now we just need to define G on all of L^2. But the step functions are dense in $L^2([0,\infty))$, so we can extend G by finding step functions f_n that converge in L^2 to a general $f \in L^2([0,\infty))$. Because f_n converge to f in L^2, they form a Cauchy sequence (in L^2); since G preserves distances, $G(f_n)$ is a Cauchy sequence in the Gaussian space. Since the Gaussian space is a subspace of L^2, this means $G(f_n)$ will converge in L^2 to a limit $G(f)$, and this gives us the isometry we want. □

As a reminder, we're going to construct a specific white noise G which guarantees continuity of sample paths. The generic definition we have here doesn't contain such a guarantee, because the Kolmogorov extension theorem gives us a process $X = (X_t)_{t\geq 0}$ which is just some random element of the space $(\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu)$ for $I = [0,\infty)$. This sigma-algebra contains events of the form

$$\{X_{t_1} \in A_1, X_{t_2} \in A_2, \cdots, X_{t_n} \in A_n\},$$

where A_i are Borel subsets of the real line, as well as an event like

$$\{X_t = 0 \quad \forall t \in \mathbb{Q}\},$$

which is a countable intersection of the events above. On the other hand, events that are **not** measurable are things like

$$\{X_t = 0 \quad \forall t \in I\},$$

because an uncountable intersection of events need not be measurable, and similarly

$$\{X_t \text{ continuous in } t\}, \quad \{X_t \text{ measurable in } t\}$$

also require us to know about X_t on uncountably many values of t. So the **probability space isn't rich enough** to capture properties like continuity – here is an example of something that can go wrong:

Example 20

Let X be a Gaussian process with $\Gamma(s,t) = \min\{s,t\}$ on $(\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu)$. We will introduce some additional randomness by augmenting the probability space to include the uniform random variable $U \sim \text{Unif}[0,1]$ independent of X. (This means we've moved to a larger space $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu) \otimes ([0,1], \mathcal{B}, \text{Leb})$, where we draw X and then independently draw U.) Now if we sample $\omega' \in \mathbb{R}^I$ and $u \in [0,1]$, we can define a new random variable

$$\tilde{X}_t(\omega', u) = X_t(\omega') + 1\{t = u\}.$$

If we define $X_t(\omega', u) = X_t(\omega')$, then X_t and $\tilde{X}_t$ are closely related – for any fixed t, $\mathbb{P}(X_t = \tilde{X}_t)$ are equal with probability 1, because any fixed time t has probability 0 of being equal to u. But X and $\tilde{X}$ aren't the same process, and in particular they can't both be continuous because we add a 1 to one of them at some random time. So X_t and $\tilde{X}_t$ are both Gaussian processes with the correct covariance, but we can't guarantee continuity even though we have the same finite-dimensional marginals.

Definition 21

Let $(X_t)_{t\in I}$ and $(\tilde{X}_t)_{t\in I}$ be two processes. We call $\tilde{X}$ a **modification** of X if $\mathbb{P}(X_t = \tilde{X}_t) = 1$ for any fixed time $t \in I$, and we say that X and $\tilde{X}$ are **indistinguishable** if

$$\left\{X_t \neq \tilde{X}_t \text{ for any } t \in I\right\} \subseteq N$$

for some measure-zero set N (this is just because the event may not actually be measurable).

Note that X_t and $\tilde{X}_t$ above are not indistinguishable, because they will always be different at **some** time $t = u$ with probability 1. The main goal of this lecture is to construct a modification of X which is continuous, but we'll have to change the probability space a little bit to do that. The actual construction is actually very concrete: we'll first define Brownian motion on $[0,1]$, looking at the dyadic points

$$D_n = \left\{0, \frac{1}{2^n}, \frac{2}{2^n}, \cdots, \frac{2^n - 1}{2^n}\right\}.$$

These sets D_i are nested in each other ($D_0 \subseteq D_1 \subseteq \cdots$), and their infinite union $D = \bigcup_{n=0}^{\infty} D_n$ is countable and dense in $[0,1]$. At a very high level, D is a countable dense subset, and we'll only look at the process X on D. Then for any other value not in D, we'll define the process using continuity, and we just need to show that we do end up with a continuous process.

Lemma 22

Let $f : D \to \mathbb{R}$ be a function that satisfies

$$\left| f\left(\frac{i}{2^n}\right) - f\left(\frac{i-1}{2^n}\right) \right| \leq \frac{K}{2^{n\alpha}}$$

for all $n \geq 1$ and $1 \leq i \leq 2^n - 1$ for some constant K and some $\alpha > 0$. Then f satisfies a similar type of estimate for all points in D: we have $|f(s) - f(t)| \leq K'|s-t|^\alpha$ for all $s, t \in D$ with $K' = \frac{2K}{1-2^{-\alpha}}$. (In other words, f is an **α-Hölder function** on D.)

(This is completely deterministic – there's no probability going on here, and note that f is defined on D only.)

Proof. Without loss of generality, let $s < t$. Then there is some integer p such that

$$\frac{1}{2^p} \leq t - s \leq \frac{1}{2^{p-1}},$$

which means s and t are either in adjacent $\frac{1}{2^p}$ blocks or separated by one block. Either way, let s_0 be the smallest point in D_p larger than s and t_0 be the largest point smaller than t. Then, the idea is that "the best way to get from s to t should use the largest jumps, because the small jumps don't give a good estimate," so formally we can write

$$s = s_0 - \sum_{t=1}^{n} \frac{\delta_\ell}{2^{p+\ell}}, \quad \delta_\ell \in \{0, 1\}$$

(looking at the steps of size $\frac{1}{2^{p+1}}$, $\frac{1}{2^{p+2}}$, and so on and always taking them if we can), and similarly

$$t = t_0 + \sum_{\ell=1}^{n} \frac{\eta_\ell}{2^{p+\ell}}, \quad \eta_\ell \in \{0, 1\}.$$

In the worst case, we will need all of these steps, which means that by the assumption in our lemma we have

$$|f(s) - f(t)| \leq \frac{K}{2^{p\alpha}} + 2\sum_{\ell \geq 1} \frac{K}{2^{(p+\ell)\alpha}} \leq \frac{K'}{2^{p\alpha}} \leq K'|s-t|^\alpha,$$

as desired. □

Lemma 23

Suppose that $(X_t)_{t\in[0,1]}$ is **any** stochastic process satisfying $\mathbb{E}[|X_s - X_t|^q] \leq C|s-t|^{1+\varepsilon}$ for all $s, t \in [0, 1]$. Then for all $\alpha \in \left(0, \frac{\varepsilon}{q}\right)$, there exists a $K_\alpha(\omega) < \infty$ such that

$$\left| X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega) \right| < \frac{K_\alpha(\omega)}{2^{n\alpha}}$$

for all $n \geq 1$ and $1 \leq i \leq 2^n - 1$.

This lemma basically says that the conditions from the previous lemma hold, except now we have a random K.

Proof. Let A_n be the event $\{\omega : \left|X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)\right| \geq \frac{1}{2^{n\alpha}}$ for any $1 \leq i \leq 2^n - 1\}$ (the numerator of 1 here is good enough for the calculations). By a union bound (over all 2^n possibilities) and Markov's inequality, we have that

$$\mathbb{P}(A_n) \leq 2^n \cdot 2^{n\alpha q} \cdot \mathbb{E}\left[|X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)|^q\right] \leq 2^n 2^{n\alpha q} C \left(\frac{1}{2^n}\right)^{1+\varepsilon} = \frac{C}{2^{n(\varepsilon - \alpha q)}},$$

where the middle inequality comes from our assumption. But by Borel-Cantelli, $\mathbb{P}(A_n \text{ i.o.}) = 0$ because $\mathbb{P}(A_n)$ forms a geometric series. So if ω lies outside the measure-zero event $\{A_n \text{ i.o.}\}$, then we must have an $N(\omega)$ such that $\omega \notin A_n$ for all $n > N(\omega)$. That means that with probability 1, the quantity **inside the supremum** of

$$\sup_{n\geq 1}\left\{\max_{1\leq i\leq 2^n-1}\left|X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)\right|\cdot 2^{n\alpha}\right\}$$

is at most 1 for all $n \geq N(\omega)$, meaning that

$$\sup_{n\geq 1}\left\{\max_{1\leq i\leq 2^n-1}\left|X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)\right|\cdot 2^{n\alpha}\right\} \leq \max\left\{1, \max_{n<N(\omega)}\left\{\max_{1\leq i\leq 2^n-1}\left|X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)\right|\cdot 2^{n\alpha}\right\}\right\}.$$

The right-hand side is now some finite number for each ω (because the max is being taken over a finite set), and in fact that is the $K_\alpha(\omega)$ from the lemma statement. □

Lemma 24

Under the same assumptions as Lemma 23, there is a modification $\tilde{X}$ of X whose sample paths are continuous. In fact, the sample paths will be α-Hölder continuous for all $\alpha \in \left(0, \frac{\varepsilon}{q}\right)$.

Proof. Let E be the event that the estimate of Lemma 23 holds for D, meaning that E is the complement of the event $\{A_n \text{ i.o.}\}$ from the previous proof. Then by Lemma 22, we know that

$$|X_s(\omega) - X_t(\omega)| \leq K'_\alpha(\omega)|s-t|^\alpha$$

for all $\omega \in E$ and all $s, t \in D$. Now we define $\tilde{X}$ in the way we said we would (extending by continuity):

$$\tilde{X}_t(\omega) = \begin{cases} \lim_{s\to t, s\in D} X_s(\omega) & \text{if } \omega \in E, \\ 0 & \text{otherwise.} \end{cases}$$

(So if we're in the measure-zero situation where the estimates don't hold, then $\tilde{X}$ is zero for all time.) But now $\tilde{X}$ is an α-Hölder continuous function for **all** ω – it satisfies the same estimate with the same K':

$$|\tilde{X}_s - \tilde{X}_t| \leq K'_\alpha(\omega)|s-t|^\alpha$$

for all $s, t \in [0,1]$ (not just D). We do still need to check that $\tilde{X}$ is a modification of X – in principle, it seems like we may have changed a lot from X, because we've ignored its value everywhere except on a countable set. But remember that we have the assumption $\mathbb{E}\left[|X_s - X_t|^q\right] \leq C|s-t|^{1+\varepsilon}$, so as s converges to t, the right hand side goes to 0, meaning X_s goes to X_t in L^q and therefore also in **probability**. On the other hand, X_s goes **almost surely** to $\tilde{X}_t$ as $s \to t$ for $s \in D$ by definition. Therefore $X_t = \tilde{X}_t$ almost surely by uniqueness of limits, because they are both limits of X_s as s approaches t. □

The combination of these three lemmas is called the **Kolmogorov continuity lemma**. We'll take a closer look now at one of the bounds

$$\mathbb{P}(A_n) \leq 2^n 2^{n\alpha q}\cdot C\left(\frac{1}{2^n}\right)^{1+\varepsilon}$$

that we had in our proofs earlier. Notice that we've used the dyadic partitioning twice here – once to go from s to t, where $s, t \in D_n$ are separated by many intervals of $\frac{1}{2^n}$, and once in this Markov bound to use the same bound many times. A more naive way we could have done our bound is to say that because we have $|s-t| = \frac{L}{2^n}$ for some L and

some n, we can consider the probability that $|X_s - X_t| \le \left(\frac{L}{2^n}\right)^\alpha$ for all $s \in D_n$ and $t = s + \frac{L}{2^n}$. A union bound here would not use the fact that our intervals are overlapping, and we would arrive at the **weaker** result

$$\mathbb{P}\left(|X_s - X_t| \le \left(\frac{L}{2^n}\right)^\alpha \text{ for all } s \in D_n,\ t = s + \frac{L}{2^n}\right) \le 2^n \cdot \frac{C\left(\frac{L}{2^n}\right)^{1+\varepsilon}}{(L/2^n)^{\alpha q}} = \frac{C \cdot L^{1-(\alpha q - \varepsilon)}}{2^{n(\alpha q - \varepsilon)}}.$$

The $L^{1-(\alpha q-\varepsilon)}$ term is large here (possibly close to the order of L), and L is possibly on the order of 2^{n-1}, so our union bound has lost a lot from the overlapping intervals. So the point is that we should be reusing bounds that we already have and "taking the largest steps possible!"

We'll finish this class by applying these lemmas to Brownian motion, finally completing the construction that we want. We have

$$\mathbb{E}\left[|X_s - X_t|^q\right] = \mathbb{E}\left[N(0, |s-t|)^q\right] = |s-t|^{q/2}\mathbb{E}[|Z|^q]$$

because $X_s - X_t$ is a centered normal random variable with variance $t - s$. If we set $\frac{q}{2} = 1 + \varepsilon$, the above lemmas allow us to make a modification to X that is α-Hölder for $\alpha < \frac{\varepsilon}{q} = \frac{q/2-1}{q}$. This estimate holds for any positive q, so in particular if we take large q, this approaches $\frac{1}{2}$. So our modification is **just short of $\frac{1}{2}$-Hölder continuous**, and the obvious question is whether this is optimal. It turns out that the answer is yes, and this is the last part of our homework.

Thus, we have constructed a process with continuous sample paths and the correct covariance! Next time, we'll talk about the probability space that is "canonical" for this Brownian motion.

4 February 12, 2020

As a reminder, the website for this class contains links to some useful references – in particular, you can find these notes that you're reading now. (These shouldn't be considered an official resource, though, since they aren't being checked by the course instructors.)

Recall from last time that we defined a **pre-Brownian motion** to be a Gaussian process with covariance function $\Gamma(s,t) = \min(s,t)$, which allows us to define a **Brownian motion** to be a pre-Brownian motion with continuous sample paths. The idea is as follows: suppose $(X_t)_{t\in[0,1]}$ is a pre-Brownian motion on any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ rich enough to support such a process. The main content of last class was the Kolmogorov continuity lemma, which basically tells us that a **modification** $(B_t)_{t\in[0,1]}$ of X exists with sample paths α-Hölder for all $\alpha \in (0, \frac{1}{2})$. In other words, there exists a $K_\alpha(\omega) < \infty$ such that

$$|B_s(\omega) - B_t(\omega)| \le K_\alpha(\omega)|s-t|^\alpha \quad \forall s,t \in [0,1],$$

and in particular this tells us that B_t is also continuous. (We constructed this by only looking at the dyadic set, showing continuity there, and then taking limits.)

Our next question is what happens if we consider a pre-Brownian motion $(X_t)_{t\ge0}$ for all nonnegative t rather than just the interval $[0,1]$. The idea is that we can just apply the above results for $[i, i+1]$ for each integer i, which tells us that a modification $(B_t)_{t\ge0}$ exists which is **locally** α-Hölder (on every compact interval) and therefore also continuous; in particular, B satisfies the definition of a Brownian motion. Letting $I = [0,\infty)$, define the set

$$C(I) = \{\text{continuous functions } I \to \mathbb{R}\} \subset \{\text{all functions } I \to \mathbb{R}\} = \mathbb{R}^I.$$

The sigma-algebra we can place on $\mathbb{R}^I$ is $Q = \mathcal{B}^{\otimes I}$, so a natural sigma-algebra to place on $C(I)$ is

$$\mathcal{G} = Q|_{C(I)} = \{C(I) \cap A : A \in Q\}.$$

The Brownian motion B now gives us a measurable (exercise) mapping which induces a measure on $C(I)$: we can write this as

$$B : (\Omega, \mathcal{F}, \mathbb{P}) \to (C(I), \mathcal{G}, P = B_{\#}\mathbb{P}).$$

In other words, for all $A \in \mathcal{G}$, we assign it a measure $P(A) = \mathbb{P}(B^{-1}(A))$ where B^{-1} is the pre-image of the Brownian motion. While $(\Omega, \mathcal{F}, \mathbb{P})$ is not unique, because there can be all kinds of "extra randomness" in the probability space, we do have the following:

Proposition 25

The measure P (called the **Wiener measure**) is unique.

Proof. It suffices to show that the value of P is uniquely determined, and it's enough to check this for a pi-system which generates the sigma-algebra $\mathcal{G}$. Consider the probability of a "simple" event like

$$P(B_{t_1} \in [a_1, b_1], \cdots, B_{t_n} \in [a_n, b_n]).$$

Events of this type generate $\mathcal{G}$, and we can calculate the probability using the covariance matrix $\Gamma(t_i, t_j)$ (where i, j run from 1 to n) to be

$$\mathbb{P}\left(N(0, \Gamma(t_i, t_j)) \in \prod [a_i, b_i]\right).$$

But this is an explicit value we can find by calculating an integral, and this characterizes the value of $P(E)$ for all events E in the pi-system generating $\mathcal{G}$, so we're done. □

Fact 26

We defined a sigma-algebra $\mathcal{G}$ on $C(I)$ by restriction, but here's another way to characterize it. A natural topology to put on $C(I)$ is to say that f_n converges to f if f_n **converges to f uniformly on compact sets**; this topology is metrizable because we can define the metric

$$d(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min\left\{1, \sup_{t \in [0,n]} |f(t) - g(t)|\right\}$$

(in particular, $d(f, g) \leq 1$ for all f, g). Then $d(f_n, f)$ goes to 0 if and only if $f_n \to f$ locally uniformly, so the metric captures the topology.

Proposition 27

The Borel sigma-algebra of $C(I)$ in the d-topology, denoted $\mathcal{H}$, is the same as the sigma-algebra $\mathcal{G}$ above.

Proof. First, we show that $\mathcal{G} \subseteq \mathcal{H}$. It's enough to show that the events in $\mathcal{G}$ are in $\mathcal{H}$; consider events of the form (which generate $\mathcal{G}$)

$$\{B_{t_1} \in [a_1, b_1], \cdots, B_{t_n} \in [a_n, b_n]\}.$$

This set is closed with respect to the d-topology (if we have a sequence of functions in this set, then any limit point will also have that $B_{t_i} \in [a_i, b_i]$ for each i), and the Borel sigma-algebra contains all of the closed sets, so such sets are indeed in $\mathcal{H}$ and thus all of $\mathcal{G}$ is contained in $\mathcal{H}$.

To show the other direction, it's enough to show that an open ball in $\mathcal{H}$ is contained in $\mathcal{G}$. The set $\{h : d(f, h) < \varepsilon\}$ is contained in $\mathcal{G}$, because we can measure $d(f, g)$ by looking just at rational t (basically we're saying that while the

definition of d takes a supremum over the whole interval $[0, n]$, we can restrict to only a countable set of indices by continuity). Thus, this open ball is also measurable in $\mathcal{G}$, so $\mathcal{H} \subseteq \mathcal{G}$. □

So far, we've been just treating Brownian motion just as a function in t, but now we want to actually make the process "move forwards or backwards in time."

Definition 28

Define the sigma-algebras

$$\mathcal{F}_t = \sigma\left(B_s : s \le t\right), \quad \mathcal{F}_{t+} = \bigcap_{s:s>t} \mathcal{F}_s.$$

Intuitively, $\mathcal{F}_{t+}$ gives an "infinitesimal amount of information past time t." For example, if the process B had a right derivative $\lim_{h \downarrow 0} \frac{B_{t+h} - B_t}{h}$, then it would be measurable with respect to $\mathcal{F}_{t+}$ but not $\mathcal{F}_t$. (But we'll see soon that Brownian motion does **not** have a right derivative.) Recall that we've characterized Brownian motion in a way that ensures it has independent increments for disjoint time intervals (from the definition of the covariance function); this next result is a stronger version of that statement:

Proposition 29 (Markov property, simplest version)

Suppose that $(B_t)_{t \ge 0}$ is a Brownian motion. Then for any fixed time $s \ge 0$, the process $(B_{s+t} - B_s)_{t \ge 0}$ is a Brownian motion independent of $\mathcal{F}_s$.

Proof. Let $W_t = B_{s+t} - B_s$ be our new process. We must show that W_t is a pre-Brownian motion and that it has continuous sample paths – the latter is true because the sample paths are just subsets of the sample paths for B_t, and the former follows because it has the correct covariance function. Furthermore, because Brownian motion has independent increments, we can make the independence statement

$$(W_{t_1}, W_{t_2}, \cdots, W_{t_k}) \perp\!\!\!\perp (B_{r_1}, \cdots, B_{r_\ell})$$

for any $r_1, \cdots, r_\ell \le s$. A pi-lambda argument then tells us that $W \perp\!\!\!\perp \mathcal{F}_s$ as desired. □

Proposition 30 (Markov property, slight improvement)

Under the same setting as Proposition 29, we have $W \perp\!\!\!\perp \mathcal{F}_{s+}$.

Proof. Again, $(W_{t_1}, W_{t_2}, \cdots, W_{t_k})$ is independent of $\mathcal{F}_{s+}$ unless some of the times are 0. To get around that, note that we can write

$$(W_{t_1}, W_{t_2}, \cdots, W_{t_k}) = \lim_{\varepsilon \downarrow 0} (W_{t_1+\varepsilon}, W_{t_2+\varepsilon}, \cdots, W_{t_k+\varepsilon})$$

by continuity of Brownian motion. For any fixed $\varepsilon > 0$, this is independent of $\mathcal{F}_{s+}$, so the independence holds in the limit as well (because the limit is measurable with respect to those variables). □

This leads us to the following consequence:

Theorem 31 (Blumenthal 0-1 law)

For any event $A \in \mathcal{F}_{0+}$, we have $\mathbb{P}(A) \in \{0, 1\}$.

In words, A "only sees the Brownian motion at an infinitesimal time at the beginning," and the idea is that we can't produce something nontrivial that depends on this infinitesimal time.

Proof. By Proposition 30 with $s = 0$, we know that $\sigma(B_s : s \geq 0)$ is independent of $\mathcal{F}_{0+}$. But $\mathcal{F}_{0+}$ sits inside $\sigma(B_s : s \geq 0)$, so $\mathcal{F}_{0+}$ is independent of itself; thus for any $A \in \mathcal{F}_{0+}$, we have $\mathbb{P}(A)^2 = \mathbb{P}(A)$ and therefore $\mathbb{P}(A) \in \{0, 1\}$. □

This gives us a bit more to work with:

Proposition 32

Let B be a Brownian motion with $B_0 = 0$. Then the following hold:

1. B will cross 0 an infinite number of times almost surely. In other words, for all $\varepsilon > 0$, we have almost surely that
$$\sup(B_s : s \in [0, \varepsilon]) > 0, \quad \inf(B_s : s \in [0, \varepsilon]) < 0.$$
2. For any $a \in \mathbb{R}$, the hitting time $T_a = \inf\{t : B_t = a\}$ is finite almost surely.

For example, this tells us that it doesn't make sense to define "the first time B returns to 0."

Proof. For (1), for any $\varepsilon > 0$ define the event

$$A_\varepsilon = \{\sup(B_s : s \in [0, \varepsilon]) > 0\}$$

and let $A = \bigcap_{\varepsilon > 0} A_\varepsilon$. We can restrict to only looking at time-values s which are rational, so A_ε is a measurable set for any ε, and similarly this means $A = \bigcap_n A_{1/n}$ is also measurable. Now A is in $\mathcal{F}_{0+}$, because the events $A_{1/n} \in \mathcal{F}_{1/n}$ are decreasing and nested. So by the zero-one law, $\mathbb{P}(A)$ is either 0 or 1, and we want to show that it is 1. However, for any $\varepsilon > 0$, we have $\mathbb{P}(A_\varepsilon) \geq \frac{1}{2}$ (because Brownian motion is symmetric around 0). Since these events are decreasing as $\varepsilon \to 0$, in the limit we have

$$\mathbb{P}(A_\varepsilon) \geq \frac{1}{2} \overset{\varepsilon \to 0}{\Longrightarrow} \mathbb{P}(A) \geq \frac{1}{2} \implies \mathbb{P}(A) = 1.$$

(The other statement follows analogously.) Now (2) is actually a consequence of (1), since taking $\varepsilon = 1$ tells us that

$$1 = \mathbb{P}\left(\sup(B_s : s \in [0, 1]) > 0\right) = \lim_{\delta \downarrow 0} \mathbb{P}\left(\sup(B_s : s \in [0, 1]) \geq \delta\right)$$

by continuity. Now we can rescale space by $\frac{1}{\delta}$ and rescale time by $\frac{1}{\delta^2}$; the result is still a Brownian motion, which means that

$$1 = \lim_{\delta \downarrow 0} \mathbb{P}\left(\sup\left(B_s : s \in \left[0, \frac{1}{\delta^2}\right]\right) \geq 1\right) = \mathbb{P}\left(\sup_s B_s \geq 1\right)$$

again by continuity because $\frac{1}{\delta^2}$ goes to infinity. This means that B_s will hit height 1 at some point with probability 1, and we can scale again to show that B hits any height a almost surely (for instance, we can multiply space by 10 and multiply time by 100). □

In particular, if we run a Brownian motion for all time, it will not converge to anything, since it must hit height a, then height $-a$, and so on. That means it oscillates a lot, and thus Brownian motion is not particularly well-behaved compared to other processes we might have seen. To answer the question of "how regular Brownian motion is," we'll choose a few interesting results for this class that are interesting but skip the more specialized ones. Our first result will be important to stochastic calculus: we know that Brownian motion is $(\frac{1}{2} - \varepsilon)$-Hölder, and on homework 2 we'll

see that

$$\limsup_{h \downarrow 0} \frac{|B_{t+h} - B_t|}{\sqrt{2h \log(1/h)}} = 1 \text{ a.s.}.$$

This tells us that the $\frac{1}{2}$ exponent is basically correct, and this should generally make sense – we have that

$$|B_{t+h} - B_t| \sim \sqrt{h}|N(0,1)|,$$

so we should expect the increment to be on the order of $\sqrt{h}$ (and the above statement shows that sometimes it is bigger).

Definition 33

A function $f : [a,b] \to \mathbb{R}$ is of **bounded variation** (BV) if (taking a supremum over all sets of break points between a and b)

$$\sup_{p=\{t_i\}} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| < \infty.$$

For a partition $p = \{t_i\}$, we'll define $V_p^1(f) = \sum_{i=1}^n |f(t_i) - f(t_{i-1})|$. Sufficiently nice functions will always be of bounded variation – for example, a C^1 function f is of bounded variation because

$$V_p^1(f) \leq \int_a^b |f'(t)| dt.$$

However, **Brownian motion is not of bounded variation** – we can read this in the book or try this ourselves. (If we partition our interval $[0,1]$ into blocks of size ε, we have $\frac{1}{\varepsilon}$ intervals and each increment is on the order of $\sqrt{\varepsilon}$, so this gives us something very large.) So we'll need a more suitable measure of variation instead. For instance, consider an α-Hölder function f (meaning that $|f(s) - f(t)| \leq K|s-t|^\alpha$). Then one way we could measure its regularity is to consider

$$V_p^{1/\alpha}(f) = \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^{1/\alpha} \leq \sum_{i=1}^{n} K^{1/\alpha} |t_i - t_{i-1}| \leq K^{1/\alpha}(b-a) < \infty$$

(by the α-Hölder bound). This means that Brownian motion on a compact interval satisfies $V_p^{1/\alpha}(B) < \infty$ for all $\alpha < \frac{1}{2}$. But **this is not a good measure of variation either**: because we're raising $|t_i - t_{i-1}|$ to a power greater than 2, the sum tends to 0 as our partition gets finer. Specifically, for any fixed $\alpha \in (0, \frac{1}{2})$, we know the Brownian motion is actually γ-Hölder for $\gamma \in (\alpha, \frac{1}{2})$, so taking P to be a set of break points separated by ε, we find that

$$V_p^{1/\alpha}(B) \asymp \frac{1}{\varepsilon} (\varepsilon^\gamma)^{1/\alpha} \asymp \varepsilon^{\gamma/\alpha - 1},$$

which tends to 0. So raising the increments to any power larger than 2 doesn't work – this motivates raising to **exactly** the second power:

$$V_p^2(B) = \sum_{i=1}^{n} (B_{t_i} - B_{t_{i-1}})^2.$$

This turns out to actually tend to a non-trivial limit:

Proposition 34

Suppose that P_n is a subdivision of the interval $[0,t]$ for all n. Then $V_{P_n}^2(B)$ converges to t in L^2 as the mesh of P_n goes to 0.

Note that we have L^2 convergence but not almost sure convergence.

Proof. This is a simple variance calculation. First, break up the t term into the lengths of our intervals as

$$||V^2_{P_n}(B) - t||_2^2 = \mathbb{E}\left[\left(\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\right)^2\right].$$

Each term has mean 0, because the expected value of each squared increment is the length of the interval. In addition, the increments are independent, so if we expand the square and take expectations, the cross-terms go away and we're left with

$$||V^2_{P_n}(B) - t||_2^2 = \sum_{i=1}^{n}(t_i - t_{i-1})^2\mathbb{E}[(Z^2-1)^2]$$

where Z is a standard normal. Now the expectation term is just some constant, meaning we can bound this L^2 norm from above by $\mathbb{E}[(Z^2-1)^2] \cdot \text{mesh}(P_n)\sum_{i=1}^{n}(t_i - t_{i-1}) = \mathbb{E}[(Z^2-1)^2] \cdot \text{mesh}(P_n)t$, which goes to 0 as $n \to \infty$ by assumption. □

This result is important, because the limit $V^2_{P_n}(B)$ has to do with sum of squares of Brownian motion on small intervals, which is a good way of understanding an expression like $\int_0^t (dB_s)^2$ (as we'll do later on in the class).

For our next result, consider the natural filtration $\mathcal{F}_t = \sigma(B_s : s \le t)$ as before, and define $\mathcal{F}_\infty = \sigma(B_s : s \ge 0)$ to be the sigma-algebra of "everything we know about the Brownian motion." Now let T be a **stopping time** (meaning that $\{T \le t\} \in \mathcal{F}_t$ for all t), and define the **stopping-time sigma-algebra**

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \le t\} \in \mathcal{F}_t\}.$$

In words, $\mathcal{F}_T$ captures everything we know about the process up until our random stopping time.

Proposition 35 (Markov property, strong version)
Let T be a stopping time. Then $(B_{T+t} - B_T)_t$ is a Brownian motion, and it is independent of $\mathcal{F}_T$ under the measure $\mathbb{P}$ **where we condition on** T **being finite**.

We can read the proof in the book (it involves approximating the vector of Brownian motion values with dyadic rationals). One of the most important applications for this is the **reflection principle**: suppose we want to compute the value

$$\mathbb{P}\left(B_t > a \text{ for some } t \in [0,1]\right).$$

The idea is to consider a height $b < a$ and ask about the probability that we do exceed a but end below b at time 1. To answer that question, we can **reflect** the Brownian motion after the stopping time when we hit a, which shows that we just need to end above $a - (b-a) = 2a - b$. In other words, if $S_t = \sup(B_s : s \le t)$,

$$\mathbb{P}(S_1 \ge a, B_1 < b) = \mathbb{P}(S_1 \ge a, B_1 \ge 2a - b) = \mathbb{P}(B_1 \ge 2a - b).$$

But B_1 is just a standard normal, so this tells us everything we might want to know about the supremum process! We'll cover this in more detail next time.

5 February 18, 2020

At the end of last time, we discussed the reflection principle briefly, and we'll elaborate on that discussion now. Recall the **strong Markov property** for Brownian motion, which tells us that for any stopping time T, $(B_{T+t} - B_T)_{t \ge 0}$ is

a Brownian motion independent of $\mathcal{F}_T$ (the stopping-time sigma-algebra) if we condition on T being finite. This is useful, for example, if we want to consider the supremum process

$$S_t = \sup\{B_s : 0 \le s \le t\}$$

which is nondecreasing in t.

Theorem 36

For any $a > 0$ and $b \in (-\infty, a]$, we have

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

The point is that this gives us a joint distribution between S_t and B_t in terms of something that is completely explicit, since B_t is just a normal random variable with mean 0 and variance t.

Proof. Apply the strong Markov property to the stopping time $\tau = \inf\{s \ge 0 : B_s \ge a\}$ (the first time we hit a) to see that $W_s = (B_{\tau+s} - B_\tau)_{s \ge 0}$ is a Brownian motion independent of the process up to time τ. Now we can explicitly write down

$$\mathbb{P}(S_t \ge a, B_t \le b) = \mathbb{P}(\tau \le t, W_{t-\tau} \le -(a-b))$$

because both sides tell us the probability of hitting a at some point and then going down by $(a-b)$ to get back below b. Using the reflection principle and noting that W is symmetric in law, the previous probability can also be written as

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

But $B_t \ge 2a - b$ means in particular that $S_t \ge B_t \ge a$, so we can drop the $S_t \ge a$ assumption and get the stated result. □

In particular, this means the joint density of S_t and B_t is given by $-\frac{\partial^2}{\partial a \partial b}\mathbb{P}(B_t \ge 2a - b)$ (negative sign because we have $S_t \ge a$ but $B_t \le b$), and this density is supported on $\{a \ge 0,\ b \le a\}$. And now we can take this density and integrate out the b to find the marginal law of S_t, but there's a faster way: decompose as

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

because the law of B_t is symmetric. So this means that for any fixed t, S_t **is equally distributed as** $|B_t|$. However, do note that the processes

$$(S_t)_{t \ge 0}, \quad (|B_t|)_{t \ge 0}$$

are not identically distributed, because S_t is increasing while B_t returns to 0 infinitely many times.

Remark 37. *On our homework, we'll discover a bit more: it turns out that S_t and $|B_t|$ are both equally distributed as $S_t - B_t$ for any fixed t, and $(S_t - B_t)$ is also equally distributed as $|B_t|$ as a* **process**.

Note that this calculation also gives us the law of the first time we hit a: if $\sigma_a = \inf\{t \ge 0 : B_t = a\}$, then

$$\sigma_a \overset{d}{=} \inf\{t \ge 0 : B_{t/a^2} = 1\} = a^2 \sigma_1,$$

but we can get the exact distribution via the calculation

$$\mathbb{P}(\sigma_a \ge t) = \mathbb{P}(S_t \le a) = \mathbb{P}(|B_t| \le a) = \mathbb{P}\left(|Z| \le \frac{a}{\sqrt{t}}\right) = \mathbb{P}\left(\frac{a^2}{Z^2} \ge t\right).$$

where Z is a standard normal. Comparing the left and right sides, we see that σ_a is equally distributed as $\frac{a^2}{Z^2}$, so in particular $\mathbb{E}[\sigma_a]$ is infinite for all $a > 0$. Taking $a = 1$, this means that σ_1 is distributed as $\frac{1}{Z^2}$. We can then calculate the density of σ_1 explicitly (left as an exercise), and find that for large t,

$$\mathbb{P}(\sigma > t) \asymp \frac{1}{\sqrt{t}},$$

which decays quite slowly. But now we can compare this with the $\tau = \inf\{t : |B_t| \geq 1\}$ (so now asking for the first time we hit either 1 or -1); it will turn out that

$$\mathbb{P}(\tau > t) \leq \exp(-\Omega(t)),$$

which decays much more quickly. To understand this exponential decay intuitively, suppose that we run Brownian motion for a very long time and we want it to stay confined in the interval $[-1, 1]$. Looking at successive intervals of length 1, there is always some positive chance it leaves in each time interval and conditioned on staying inside $[-1, 1]$ this probability is bounded from below. But again, we'll be more precise on the homework.

This is all we'll cover from chapter 2 of Le Gall – now we'll move on to chapter 3, which discusses **continuous-time martingales**. Unfortunately, it's pretty boring: under mild assumptions, all the properties from discrete-time martingales hold. So we'll go through this fairly quickly – if we took 18.675 in a different semester where we didn't cover martingales in such detail, we might have to do reading on our own.

Brownian motion B_t is an example of a continuous-time martingale, and here's another example to keep in mind as well: let ζ, ζ_i be iid exponential random variables, and let

$$N_t = \max\left\{n : \sum_{i=1}^{n} \zeta_i \leq t\right\}.$$

Then N_t is distributed as $\text{Pois}(t)$, and it is an integer-valued process with right-continuous sample paths (but discontinuous jumps). It will turn out $N_t - t$ is a continuous-time martingale as well, so our formalism should be able to study it. (Generally, we'll be assuming that the processes we are studying are right-continuous.) Throughout this chapter, everything will live on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a **filtration**

$$(\mathcal{F}_t)_{0 \leq t \leq \infty} : \quad \mathcal{F}_s \subseteq \mathcal{F}_t \quad \forall s \leq t.$$

We'll call $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ a **filtered probability space**.

Definition 38

A process $(X_t)_{t \geq 0}$ is **adapted** to a filtration $\mathcal{F}_t$ if $X_t \in \mathcal{F}_t$ for all t.

We'll also reiterate the following definition from earlier:

Definition 39

A random variable $\tau : \Omega \to [0, \infty]$ is a **stopping time** if $\{\tau \leq t\} \in \mathcal{F}_t$ for all t (that is, whether we stop doesn't depend on what happens in the future). The ***σ*-field of the past up to *τ*** is

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \forall t\}.$$

So an event that only depends on time up to τ can be rephrased as "only needing information up to t if $\tau \leq t$." We should read all of the basic facts about filtrations and stopping times on our own (sections 3.1 and 3.2 of our

book). An example of what we'll see is that if σ, τ are both stopping times, then their minimum $\sigma \wedge \tau$ and maximum $\sigma \vee \tau$ are also stopping times, and

$$\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau.$$

(This specific fact is useful for one of the questions on our homework.)

Definition 40

Let $(\Omega, \mathcal{F}, (\mathcal{F})_+, \mathbb{P})$ be a filtered probability space, and let $(X_t)_{t\geq 0}$ be a real-valued process. Then (X_t) is a **submartingale** if the following properties hold:

- $X_t \in \mathcal{F}_t$ for all t (in other words, the process is **adapted** to the filtration),
- $\mathbb{E}[|X_t|]$ is finite for all t (the variable is **integrable**),
- For every $0 \leq s \leq t$, we have $X_s \leq \mathbb{E}[X_t|\mathcal{F}_s]$; we similarly say that (X_t) is a **supermartingale** if we flip the inequality.

A martingale is both a submartingale and a supermartingale, meaning that $\mathbb{E}[X_t]$ is constant for a martingale, nondecreasing for a submartingale, and nonincreasing for a supermartingale.

Example 41

Let $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ be any integrable random variable. Then we can check that $X_t = \mathbb{E}[Z|\mathcal{F}_t]$ is a martingale.

The fact that $X_s = \mathbb{E}[X_t|\mathcal{F}_s]$ follows from basic properties of the conditional expectation, and each X_t is integrable because

$$\mathbb{E}[|X_t|] = \mathbb{E}\left[|\mathbb{E}[Z|\mathcal{F}_t]|\right] \leq \mathbb{E}\left[\mathbb{E}[|Z||\mathcal{F}_t]\right] = \mathbb{E}[|Z|] < \infty$$

by Jensen's inequality.

Remark 42. *The standard Brownian motion B_t is a martingale (it satisfies all properties in the definition), but there is no random variable Z such that $B_t = \mathbb{E}[Z|\mathcal{F}_t]$ because $\mathbb{E}[|B_t|] = t^{1/2}\mathbb{E}|N(0,1)|$ is unbounded as $t \to \infty$.*

There are some other important martingales based on Brownian motion as well: we can check that $B_t^2 - t$ is a martingale, as is $\exp\left(\theta B_t - \frac{\theta^2 t}{2}\right)$. We'll use the rest of this lecture to prove some basic results about general martingales:

Proposition 43

Let X_t be a (sub/super)martingale. Then for any $t < \infty$, we have

$$\sup\{\mathbb{E}[|X_s|] : 0 \leq s \leq t\} < \infty.$$

We didn't have to prove this in the discrete case, because we only had a finite number of variables to consider between 0 and t, and we know that $\mathbb{E}[|X_s|]$ is finite at any given time s.

Proof. Without loss of generality, say that X is a submartingale. Then $(X_t)_+ = \max\{X_t, 0\}$ is a submartingale (because $f(x) = \max(x, 0)$ is a convex nondecreasing function), so $\mathbb{E}[(X_t)_+]$ is nondecreasing. Therefore, for any $s \leq t$,

$$\mathbb{E}[|X_s|] = \mathbb{E}[2(X_s)_+ - X_s] \leq \mathbb{E}[2(X_t)_+ - X_0].$$

This bound holds uniformly over s, so the supremum of $\mathbb{E}[|X_s|]$ must indeed be finite. □

This next fact is a weak version of the optional stopping theorem for discrete-time submartingales, and we saw this in 18.675:

Lemma 44

Let X_n be a **discrete-time** submartingale, and let τ be a bounded stopping time such that $\tau \leq n$ almost surely. Then $\mathbb{E}[X_0] \leq \mathbb{E}[X_\tau] \leq \mathbb{E}[X_n]$.

Proof. Consider the stopped process $Y_k = X_{k \wedge \tau}$. This is also a submartingale, and

$$\mathbb{E}[X_0] = \mathbb{E}[Y_0] \leq \mathbb{E}[Y_n] = \mathbb{E}[X_{n \wedge \tau}] = \mathbb{E}[X_\tau]$$

because $\tau \leq n$ almost surely. For the other inequality, write

$$\mathbb{E}[X_\tau] = \sum_{k=0}^{n} \mathbb{E}[1\{\tau = k\} X_k] \leq \sum_{k=0}^{n} \mathbb{E}[1\{\tau = k\} \mathbb{E}[X_n | \mathcal{F}_k]]$$

by the submartingale condition, and now we can put $1\{\tau = k\}$ inside the conditional expectation because it is measurable with respect to $\mathcal{F}_k$. Thus,

$$\mathbb{E}[X_\tau] \leq \sum_{k=0}^{n} \mathbb{E}[\mathbb{E}[1\{\tau = k\} X_n | \mathcal{F}_k]] = \sum_{k=0}^{n} \mathbb{E}[1\{\tau = k\} X_n] = \mathbb{E}[X_n],$$

as desired. □

Similarly, we also have the following result:

Proposition 45 (Maximal inequality, discrete version)

Let Y_n be a discrete-time (sub/super)martingale. Then for all $\lambda \geq 0$,

$$\lambda \cdot \mathbb{P}\left(\max_{0 \leq k \leq n} |Y_k| \geq \lambda \right) \leq \mathbb{E}[|Y_0| + 2|Y_n|].$$

This tells us that we have control of the entire trajectory up to time n just by knowing something about the process at the beginning and end.

Proof. Without loss of generality, assume Y_n is a supermartingale. Let A be the event that $\max_{0 \leq k \leq n} |Y_k| \geq \lambda$, and consider the stopping time

$$\tau = \min\{k : |Y_k| \geq \lambda \text{ or } k = n\} \in \{0, 1, \cdots, n\}.$$

Recall that the notation $\mathbb{E}[X; A]$ means $\mathbb{E}[X \cdot 1\{X \in A\}]$. We have

$$\lambda \mathbb{P}\left(\max_{0 \leq k \leq n} |Y_k| \geq \lambda \right) \leq \mathbb{E}[|Y_\tau|; A],$$

because $|Y_\tau| \geq \lambda$ whenever the event A occurs. The right-hand side can now be decomposed as

$$\mathbb{E}[|Y_\tau|; A] \leq \mathbb{E}[|Y_\tau|] = \mathbb{E}\left[Y_\tau + 2(Y_\tau)_-\right],$$

and now since Y_τ is a supermartingale and $(Y_\tau)_-$ is a submartingale, we can upper bound this quantity by $\mathbb{E}[Y_0 + 2(Y_n)_-]$, which is at most the right-hand side. □

Remark 46. *We can get a small improvement in Proposition 45 if we assume that X is a martingale. Then we can do the decomposition before removing the indicator on A:*

$$\lambda\mathbb{P}(\max_{k\le n}|X_k|\ge\lambda)\le\mathbb{E}[|X_\tau|;A]=\mathbb{E}[(X_\tau)_+ + (X_\tau)_-;A],$$

and because X is a martingale, $(X_\tau)_+$ and $(X_\tau)_-$ are **both** *submartingales, so this is actually bounded by $\mathbb{E}[|X_n|;A]$.*

We can now generalize this result to continuous-time martingales:

Proposition 47 (Maximal inequality, continuous version)
Let X_t be a (sub/super) martingale with **right-continuous sample paths** (this is a regularity condition). Then

$$\lambda\mathbb{P}\left(\sup_{s\le t}|X_s|\ge\lambda\right)\le\mathbb{E}[|X_0|+2|X_t|]$$

Proof. Fix t. Then any sequence $0=t_0<t_1<t_2<\cdots<t_m=t$ gives a discrete-time (sub/super)martingale $(X_{t_k})_k$, so by Proposition 45, we have

$$\lambda\mathbb{P}\left(\max_{0\le k\le m}|X_{t_k}|\ge\lambda\right)\le\mathbb{E}[|X_0|+2|X_t|].$$

Now consider a sequence of such time sequences $D_m\uparrow D$, where $D_1=\{t_0=0,t_1=t\}$ and the D_ms are nested and increase to a countable dense subset D in $[0,t]$. We find that

$$\lambda\mathbb{P}\left(\sup_{s\in[0,t]\cap D}|X_s|\ge\lambda\right)\le\mathbb{E}[|X_0|+2|X_t|],$$

and right-continuity allows us to replace $[0,t]\cap D$ with $[0,t]$, yielding the desired result. □

Next, we'll prove a few more inequalities that are weaker but easier to package and remember:

Proposition 48 (Doob's L^p inequality, discrete version)
Let X_n be a discrete-time martingale. Then for all $p>1$ and finite n, we have (letting $C_p=\frac{p}{p-1}$)

$$\left|\left|\max_{0\le k\le n}|X_k|\right|\right|_p\le C_p||X_n||_p.$$

Proposition 49 (Doob's L^p inequality, continuous version)
Let X_t be a martingale with right-continuous sample paths. Then for all $p>1$ and finite t, we have (letting $C_p=\frac{p}{p-1}$)

$$\left|\left|\sup_{0\le s\le t}|X_s|\right|\right|_p\le C_p||X_t||_p.$$

Here, the discrete version again implies the continuous version by the same argument as above, so we'll just prove the discrete version.

Proof of Proposition 48. With the same event A as before, we take the inequality

$$\boxed{\lambda\mathbb{P}\left(\max_{k\le n}|X_k|\ge\lambda\right)\le\mathbb{E}[|X_n|;A]}$$

from Remark 46. We wish to bound the L^p norm of $S_n = \max_{0\le k\le n}|X_k|$, which we will do using the formula

$$\mathbb{E}[(S_n)^p] = \int_0^\infty \mathbb{P}(S_n^p \ge t)dt = \int_0^\infty py^{p-1}\mathbb{P}(S_n \ge y)dy$$

(last step by a change of variables $t = y^p$). Plugging in the boxed inequality yields

$$\mathbb{E}[(S_n)^p] \le \int_0^\infty py^{p-1}\left[\frac{\mathbb{E}(|X_n|; S_n \ge y)}{y}\right]dy,$$

and by Fubini's theorem, we can change the order of integration to rewrite this as $\mathbb{E}\left[|X_n|\int_0^{S_n} py^{p-2}dy\right]$, where the indicator that $S_n \ge y$ is accounted for by replacing ∞ with S_n. Integrating directly and then using Hölder's inequality, we find that

$$\mathbb{E}[(S_n)^p] \le \frac{p}{p-1}\mathbb{E}\left[|X_n|S_n^{p-1}\right] \le \frac{p}{p-1}||X_n||_p||(S_n)^{p-1}||_{\frac{p}{p-1}} = \frac{p}{p-1}||X_n||_p||S_n||_p^{p-1}.$$

We can now divide through by $||S_n||_p^{p-1}$ to get the desired result if the norm is finite, and otherwise we use a standard truncation argument to finish the proof. □

However, it's important to remember that there's no L^p inequality for $p = 1$. For example, consider the simple random walk X_n on the integers starting from $X_0 = 1$, let τ be the first time n such that $X_n = 0$, and let $M_n = X_{n\wedge\tau}$. This yields a nonnegative martingale with $\mathbb{E}[M_n] = 1$, but we can't control the maximum of this process – indeed, if we define $S_n = \max_{k\le n} M_k$, then $||S_n||$ is unbounded as $n \to \infty$. To see this, it's enough to show that $||S_\infty||$ has infinite expectation – the probability that S_∞ is at least a is the chance that a random walk hits a before 0, which is $\frac{1}{a}$ (by the optional stopping theorem for discrete-time martingales, for example). This is not summable over a, so the expectation is indeed infinite.

6 February 19, 2020

Yesterday, we defined what it means for $(X_t)_{t\ge0}$ to be a continuous-time (sub/super)martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. We then showed the maximal inequality

$$\lambda\mathbb{P}\left(\sup_{0\le s\le t}|X_s| \ge \lambda\right) \le \mathbb{E}\left[|X_0| + 2|X_t|\right],$$

assuming that the process X has right-continuous sample paths. (There's a slightly stronger result for martingales, which is that $\lambda\mathbb{P}(A) \le \mathbb{E}[|X_t|; A]$ for an event of the type $A = \sup_{0\le s\le t}|X_s| \ge \lambda$.) We also used this to show the L^p inequality by integrating the previous result to find that

$$\left|\left|\sup_{0\le s\le t} X_s\right|\right|_p \le \frac{p}{p-1}||X_t||_p$$

for all $p > 1$ and $t \in [0, \infty)$. This requires right-continuous sample paths – if we don't assume that fact, we actually only prove that

$$\lambda\mathbb{P}\left(\sup_{s\in[0,t]\cap D}|X_s| \ge \lambda\right) \le \mathbb{E}\left[|X_0| + 2|X_t|\right]$$

for a countable dense set D. We'll build off of this today: in particular, we'll show that under mild conditions, a (sub/super)martingale has a right-continuous modification.

Definition 50

A function f is **right continuous with left limits** (also **rcll** or **càdlàg** for short) if for all $t \geq 0$, $f(t) = \lim_{s \downarrow t} f(s)$, and for all $t > 0$, $\lim_{s \uparrow t} f(s)$ exists.

The main idea is that martingales can't oscillate too much, so we can guarantee existence of limits. We'll start with a deterministic result by controlling **upcrossing numbers**: for any subset $I \subseteq [0, \infty)$ and any $a < b$, denote $U^f_{a,b}(I)$ to be the maximum k such that there exist times $s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k$ suh that $f(s_i) \leq a$ and $f(t_i) \geq b$ for all i. (In other words, this is the maximum number of times that we go from a to b.) Basically, if we can control upcrossing numbers, we have some regularity control:

Lemma 51

Let D be a countable dense subset of $[0, \infty)$, and suppose we have a function $f : D \to \mathbb{R}$ that is **locally bounded**, meaning that $\sup\{|f(t)| : t \in [0, T] \cap D\} < \infty$ for all $T \in D$. Also, suppose that $U^f_{a,b}(D \cap [0, T]) < \infty$ for all $T \in D$ and for all rational $a < b$ (to avoid issues with measurability). Then f has all of its left and right limits, and the function

$$g(t) = f(t+) = \lim_{s \downarrow t, s \in D} f(s)$$

is rcll.

Proof. Take any $t \geq 0$, and suppose for the sake of contradiction that the (WLOG) right limit $\lim_{s \downarrow t, s \in D} f(s)$ does not exist. This means that the lim sup and lim inf of this limit are different, so there exist rational a, b with

$$\liminf_{s \downarrow t, s \in D} f(s) < a < b < \limsup_{s \downarrow t, s \in D} f(s).$$

But this means the function f must cross between a and b infinitely many times, which is a contradiction with the assumption that $U^f_{a,b}([0, T] \cap D)$ is finite. Showing that the function g is rcll follows from a similar argument. □

This basically tells us that we need to control upcrossing numbers, and we can do so using the following idea:

Definition 52

Let X_n be an adapted discrete process, meaning that $X_n \in \mathcal{F}_n$ for all n, and let H_n be a **previsible** process, meaning that $H_n \in \mathcal{F}_{n-1}$ for all n. Then the **Doob transform** is defined via

$$(H \cdot X)_n = \sum_{k=1}^{n} H_k (X_k - X_{k-1}).$$

Notably, if X is a supermartingale and H is a **nonnegative bounded** previsible process, then $H \cdot X$ is also a supermartingale. We can check this from the definition, but what it's really saying is that if there is a gambling system X where we can't win, even with a betting strategy H (to tell us how much to bet in the next game), we can't game the expected gain $H \cdot X$.

Lemma 53 (Doob's upcrossing inequality)

Let X_n be a discrete supermartingale. Then the expected number of upcrossings satisfies

$$\mathbb{E}\left[U^X_{a,b}([0, n])\right] \leq \frac{\mathbb{E}[(X_n - a)_-]}{b - a}.$$

Proof. We'll study the value Y we get off of betting on X only during upcrossings. More formally, let $Y = H \cdot X$, where

$$H_j = 1\{j \in (\sigma_i, \tau_i]\} \text{ for some } i$$

and where σ_i and τ_i are stopping times corresponding to the ith time X_n is below a and above b, respectively. Note that H_j is in $\mathcal{F}_{j-1}$, because

$$\{j \in (\sigma_i, \tau_i]\} = \{\sigma_i \leq j-1\} \cap \{\tau_i \leq j-1\}^c,$$

and both of these events are $\mathcal{F}_{j-1}$-measurable by the definition of a stopping time. Thus $Y = H \cdot X$ is a supermartingale with $Y_0 = 0$, so $\mathbb{E}[Y_n] \leq 0$. On the other hand, we know that Y_n gets a contribution from the number of completed up-crossings, and then there's an extra term from the end where we start an up-crossing but we go very far down at the end. In the worst case we can lose $(X_n - a)$, so

$$Y_n \geq (b-a)U_{a,b}^X([0,n]) - (X_n - a)_-.$$

Taking expectations of both sides, we have

$$0 \geq \mathbb{E}[Y_n] \geq (b-a)\mathbb{E}[U_{a,b}^X([0,n])] - \mathbb{E}[(X_n - a)_-],$$

and rearranging gives the desired bound. □

Corollary 54

Let X_t be a supermartingale and let D be a countable dense subset of $[0, \infty)$. Then there is a probability-zero event N such that for all $\omega \notin N$, the function $t \mapsto X_t(\omega)$ satisfies Lemma 51.

Proof. The first property (locally bounded) follows by the maximal inequality

$$\lambda \mathbb{P}\left(\sup_{s \in [0,t] \cap D} |X_s| \geq \lambda\right) \leq \mathbb{E}(|X_0| + 2|X_t|) < \infty$$

and taking $\lambda \to \infty$ (which shows that we cannot have a positive probability of going off to infinity). The second property (number of upcrossings is bounded) follows from Lemma 53 plus an approximation argument. Indeed, for any finite subset D_n of D, Lemma 53 tells us that

$$\mathbb{E}\left[U_{a,b}^X([0,t] \cap D_n)\right] \leq \frac{\mathbb{E}[(X_t - a)_-]}{b-a}$$

because X_t is a supermartingale. Now taking D_n nested and increasing to D should yield an increasing number of upcrossings, but we'll always be uniformly bounded by the right-hand side, so even in the limit we must have finitely many upcrossings. □

Remember that our goal is to turn our process X_t into a modification $\tilde{X}_t$ such that $\mathbb{P}(\tilde{X}_t = X_t) = 1$ for all t. What we've proved so far suggests that we should do this by **taking limits from the right**. However, while we know that $t \mapsto X_t(\omega)$ has left and right limits, we still need to check that taking limits from the right to get an rcll function actually yields a modification of X_t. For example, let f be a deterministic nonincreasing function and let $X_t = f(t)$. Then this is a supermartingale, but if f is not right-continuous, then there's no way for us to modify it to get right-continuous sample paths.

So from here, we'll need to use two facts: first of all, we showed last time that if X_t is a (sub/super)martingale, then $\sup\{\mathbb{E}[|X_s|] : 0 \leq s \leq t\}$ is finite for all finite t. Also, we'll need some theory of **backwards (sub/super)martingales**,

which are indexed by $\mathbb{Z}_{\leq 0}$ instead of $\mathbb{Z}_{\geq 0}$. In such a process, we have sigma-algebras $\mathcal{F}_{-3} \subset \mathcal{F}_{-2} \subset \mathcal{F}_{-1} \subset \cdots$, and a supermartingale now satisfies an inequality of the form $Y_{-10} \geq \mathbb{E}[Y_{-9}|\mathcal{F}_{-10}]$.

Proposition 55

If Y_n is a backwards (sub/super)martingale and $\sup_n[|Y_n|_1] < \infty$, then Y_n converges to a finite limit $Y_{-\infty}$ almost surely and in L^1.

Proof. If $(\cdots, Y_{-3}, Y_{-2}, Y_{-1}, Y_0)$ is a backwards supermartingale, then we can apply the Doob upcrossing inequality for any finite $n \in \mathbb{Z}_{\leq 0}$ to find

$$\mathbb{E}\left[U^Y_{a,b}([n,0])\right] \leq \frac{\mathbb{E}[X_0 - a)]}{b-a}.$$

Now we can take the limit as $n \to -\infty$, and the total number of upcrossings will be uniformly bounded by the finite quantity on the right-hand side. This means that for all rational a, b, $U^Y_{a,b}([-\infty, 0]) < \infty$, almost surely, so Y_n must converge almost surely to a limit $Y_{-\infty}$ (or else it would oscillate between two rational numbers infinitely often). The L^1 convergence is a uniform integrability argument, which is trickier for supermartingales than for martingales (which we did in 18.675) – we should read this on our own. □

Theorem 56

Let X_t be a supermartingale and let D be a countable dense subset. Then there is some probability-zero event N such that $X_t(\omega)$ has left and right limits for all $\omega \notin N$. Furthermore,

$$Y_t(\omega) = \begin{cases} X_{t+}(\omega) = \lim_{s \downarrow t, s \in D} X_s(\omega) & \text{if the limit exists,} \\ 0 & \text{otherwise} \end{cases}$$

is a supermartingale with the filtration $\mathcal{G}_t = \mathcal{F}_{t+}$, and $X_t \geq \mathbb{E}[Y_t|\mathcal{F}_t]$ with equality if the map $t \mapsto \mathbb{E}[X_t]$ is right-continuous (this is the mild condition).

Proof. The first part (left and right limits) follows directly from Corollary 54 and Lemma 51. To check that Y_t is a supermartingale, it's clear that $Y_t \in \mathcal{G}_t$ by the limit definition, and for any $s_k \downarrow t$, X_{s_k} is a backwards supermartingale. This backwards supermartingale is bounded in L^1, because the supremum of $\mathbb{E}[|X_s|]$ is bounded on finite time intervals, so we can use Proposition 55 to show that X_{s_k} converges almost surely and in L^1 to Y_t – in particular, Y_t is indeed in L^1 (which shows integrability).

Next, the supermartingale condition for X_t shows that $X_t \geq \mathbb{E}[X_{s_k}|\mathcal{F}_t]$, and X_{s_k} converges in L^1 to Y_t, so this converges to $\mathbb{E}[Y_t|\mathcal{F}_t]$ as $k \to \infty$, showing the desired inequality. Finally for the equality case, if $t \to \mathbb{E}[X_t]$ is right continuous, then $\mathbb{E}[X_t] = \lim_{s \downarrow t} \mathbb{E}[X_s]$. We can switch the limit and expectation by the L^1 convergence, so we in fact have

$$\mathbb{E}[X_t] = \mathbb{E}\left[\lim_{s \downarrow t} X_s\right] = \mathbb{E}[Y_t].$$

This means that we know both that $X_t \geq \mathbb{E}[Y_t|\mathcal{F}_t]$ and that $\mathbb{E}[X_t] = \mathbb{E}[Y_t]$, but these can only hold if the former inequality is actually an equality. Thus $X_t = \mathbb{E}[Y_t|\mathcal{F}_t]$ almost surely as desired.

Finally, we still need to show that Y_t is actually a supermartingale. Let $s_n \downarrow s$ and $t_n \downarrow t$ be chosen so that $s < t$ and $s_n < t_n$ for all n. Then for any event $A \in \mathcal{G}_s$, we have

$$\boxed{\mathbb{E}[Y_s; A]} = \lim_{k \to \infty} \mathbb{E}[X_{s_k}; A]$$

by the backwards supermartingale L^1 convergence, but now we may bound this from below by $\lim_{k\to\infty}\mathbb{E}[X_{t_k};A] = \boxed{\mathbb{E}[Y_t;A]}$, where in the last step we've again switched the limit and expectation. This shows that we do have a supermartingale. □

We can now put everything together for the main result:

Theorem 57
Suppose we have a **right-continuous filtration** (meaning that $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t) and assume that $\mathcal{F}_t$ is **complete** (meaning it contains the null sets). Let X_t be a supermartingale with **right-continuous mean** $\mathbb{E}[X_t]$. Then X has an rcll modification $\tilde{X}$ which is also a supermartingale with respect to $\mathcal{F}_t$.

Proof. We define

$$\tilde{X}_t(\omega) = \begin{cases} Y_t(\omega) & \text{if } \omega \notin N, \\ 0 & \text{otherwise,} \end{cases}$$

where N is the set from Corollary 54. (This means that we're allowed to look into the whole future and see if things go wrong – it's just an issue with measurability.) So now $\tilde{X}_t \in \mathcal{F}_t$ because $\mathcal{F}_t$ is right-continuous and complete, and it is also a supermartingale because Y_t is a supermartingale and $\tilde{X}_t = Y_t$ almost surely.

It remains to show that X is a modification of $\tilde{X}$. By the last part of Theorem 56 (using $\tilde{X}$ instead of Y), we have $X_t = \mathbb{E}[\tilde{X}_t|\mathcal{F}_t]$ because the mean is right-continuous, but $\tilde{X}_t$ is measurable with respect to $\mathcal{F}_{t+} = \mathcal{F}_t$ and thus $X_t = \tilde{X}_t$ almost surely, as desired. □

Basically, with a sufficiently rich filtration and with the mild condition that the deterministic function $\mathbb{E}[X_t]$ is right-continuous, we get some nice results.

Remark 58. *It's necessary to assume that $\mathbb{E}[X_t]$ is right-continuous: as mentioned above, a counterexample otherwise is the deterministic process $X_t = f(t)$ for a non-right-continuous f. It's also necessary to have the assumption of right-continuity: consider $\Omega = \{\pm 1\}$, let $\mathbb{P}$ be the uniform measure on Ω, and define $X_t(\omega) = \omega 1\{t > 1\}$. In words, this means that X starts off as 0 and then jumps to a random bit at time 1, so it is a martingale. However, the filtration generated by X is trivial until $t = 1$ and then jumps to the complete sigma-algebra, so the filtration is not right-continuous, and indeed there is no modification of X that is rcll.*

Next time, we'll talk about the optional stopping theorem for continuous martingales, and that will be all from chapter 3.

7 February 24, 2020

Last time, we discussed sample path regularity for continuous-time (sub/super)martingales: it was somewhat technical, so let's review the main points. If D is a countable dense subset in $[0,\infty)$, we found that for any (sub/super)martingale X_t, its left and right limits

$$X_{t+}(\omega) = \lim_{s\downarrow t, s\in D} X_s(\omega), \quad X_{t-}(\omega) \lim_{s\uparrow t, s\in D} X_s(\omega)$$

exist, and if the mapping $t \to \mathbb{E}[X_t]$ is right-continuous, we actually have $X_t = \mathbb{E}[X_{t+}|\mathcal{F}_t]$. Then if $\mathcal{F}_+$ is right-continuous and complete (contains the null sets), then X has a modification $\tilde{X}$ satisfying $\tilde{X}_t = X_{t+}$ except on a null set, and this modification has sample paths which are **rcll** (right-continuous with left limits). This will allow us to generally assume right-continuous sample paths in most of our future discussion.

Today, the discussion will focus on **optional stopping theorems**, primarily for martingales. The main feature is that we can extend the statement $X_s = \mathbb{E}[X_t|\mathcal{F}_s]$ for fixed $s \leq t$ to **random** times; our goal is to prove that if $\sigma \leq \tau$ are both stopping times, then $X_\sigma = \mathbb{E}[X_\tau|\mathcal{F}_\sigma]$. There are lots of applications of this, many of which are on our homework, but here's a simple example:

Example 59

Let B_t be a Brownian motion started from 0, and let $\tau = \tau_a \wedge \tau_b$ for $a < 0 < b$. Then an optional stopping theorem will tell us (because B_t is a martingale) that

$$0 = B_0 = \mathbb{E}[B_\tau] = a\mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_a > \tau_b),$$

allowing us to explicitly calculate the probability $p = \frac{b}{b-a}$ that we hit a before we hit b.

Note, though, that optional stopping theorems **do not hold** without further restrictions:

Example 60

Let τ be the first time that Brownian motion hits 1. We've shown previously that τ is finite almost surely, but $0 = B_0 \neq \mathbb{E}[B_\tau] = 1$.

We do have a reason to believe that the optional stopping theorem should continue to hold, though: we know that if X_t is a martingale, then $\mathbb{E}[X_t]$ is constant in t, and if τ is a stopping time, then $X_{t\wedge\tau}$ is a martingale. (There is a small caveat: we know the stopped process is a martingale in discrete time, but we haven't actually proved this for continuous time yet.) But then this means $\mathbb{E}[X_{t\wedge\tau}]$ is also constant, so

$$\mathbb{E}[X_0] = \mathbb{E}[X_{0\wedge\tau}] = \mathbb{E}[X_{t\wedge\tau}]$$

for all finite t. And then taking $t \to \infty$ should yield (for any finite stopping time τ) that

$$\mathbb{E}[X_0] = \lim_{t\to\infty} \mathbb{E}[X_{t\wedge\tau}] \stackrel{?}{=} \mathbb{E}\left[\lim_{t\to\infty} X_{t\wedge\tau}\right] = \mathbb{E}[X_\tau],$$

so we've reduced this to the usual L^1 convergence question: "can we swap the limit and expectation?". We'll start with an almost-sure (pointwise) convergence result:

Proposition 61

Let X_t be a (sub/super)martingale with right-continuous sample paths, and suppose that $\sup_t ||X_t||_1 < \infty$ (meaning our process is bounded in L^1). Then X_t converges almost surely to $X_\infty \in L^1$ (though L^1 convergence may not occur).

Proof. Without loss of generality we can assume that X_t is a supermartingale (otherwise multiply it by -1). Then we know by the upcrossing inequality that

$$\mathbb{E}\left[U_{a,b}([0,T] \cap D)\right] \leq \frac{\mathbb{E}[(X_T - a)_-]}{b-a},$$

and by the monotone convergence theorem, because the left hand side is increasing as we take $T \to \infty$, we have

$$\mathbb{E}[U_{a,b}(D)] \leq \sup_t \frac{\mathbb{E}[|X_t - a|]}{b-a} < \infty.$$

This means that $U_{a,b}$ is finite for all rational $a < b$ almost surely, and thus the same is true over all rational $< b$. In particular, this means X_t must converge (or else there would be a rational sandwiched between the liminf and limsup). Then Fatou's lemma tells us (because $X_\infty = \liminf_{t\to\infty} X_t$) that

$$\mathbb{E}[|X_\infty|] \le \liminf_{t\to\infty} \mathbb{E}[|X_t|] < \infty,$$

so X_∞ is in L^1 as desired. □

Again, it's important to remember that X_t does not always converge in L^1 to X_∞ under these conditions. Another example that is good to keep in mind is

$$X_t = \exp(B_t - t/2),$$

which is a nonnegative martingale with $\mathbb{E}[X_t] = 1$ for all t. However, B_t will be much smaller than $\frac{t}{2}$ as $t \to \infty$, so X_t converges almost surely to $X_\infty = 0$.

Proposition 62

Let X_t be a (sub/super)martingale with right-continuous sample paths, and suppose we know that $\sup_t ||X_t||_p < \infty$ for some $p > 1$. Then $X_t \to X_\infty$ almost surely and in L^p, and in particular it also converges in L^1.

Proof. This proof is the same as in the discrete case. Recall Doob's L^p inequality, which tells us that

$$\left|\left|\sup_{s\le t} |X_s|\right|\right|_p \le \frac{p}{p-1} ||X_t||_p.$$

Taking $t \to \infty$, the left side is nondecreasing in t, and the right-hand side stays bounded by assumption, so $S = \sup_{t\ge 0} |X_t|$ is in L^p. The conditions assumed here are strictly stronger than in the previous proposition, so we know already that $X_n \to X_\infty$ almost surely. But now by the dominated convergence theorem, we have

$$\lim_{t\to\infty} \mathbb{E}(|X_t - X_\infty|^p) \to 0,$$

since $|X_t - X_\infty|^p$ is dominated by $(2S)^p$, which we've shown is in L^1. □

This will help us with some but not all of the cases we're interested in, and in fact we have a precise characterization of when L^1 convergence occurs. (And the proofs now will be a bit more complicated than in the discrete time case.)

Definition 63

A collection of random variables $\{X_i\}_{i\in I}$ is **uniformly integrable** (u.i.) if

$$\lim_{M\to\infty} \left(\sup_{i\in I} \mathbb{E}[|X_i|; |X_i| \ge M] \right) \to 0.$$

As a trivial example, let $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and assume that all of our random variables satisfy $|X_i| \le Z$. Then the $|X_i|$s are uniformly integrable because $\mathbb{E}[|Z|; |Z| \ge M]$ goes to 0. A less trivial example is to consider the collection of random variables

$$X_\mathcal{G} = \mathbb{E}[Z|\mathcal{G}],$$

where $\mathcal{G}$ is any sub-σ-field of $\mathcal{F}$ (showing this is uniformly integrable is a good exercise). Finally, note that being uniformly integrable is stronger than being bounded in L^1 (also useful to think about on our own).

Theorem 64

Suppose we have a collection of random variables X_n indexed by **integer** n which converge in probability to X_∞. Then the following are equivalent:

1. The $\{X_n\}$ are uniformly integrable,
2. X_n converges in L^1 to X_∞,
3. $\mathbb{E}[|X_n|]$ converges to $\mathbb{E}[|X_\infty|]$.

(We can see [3] for the proof, but we also proved this last semester in 18.675.) As a word of caution, (1) implies (2) implies (3) if we have a **real-indexed** process X_t, but (3) does not imply (1).

Definition 65

A martingale X_t is **closed** if there is some $Z \in L^1$ such that $X_t = \mathbb{E}[Z|\mathcal{F}_t]$ for all t.

Theorem 66

Let X_t be a right-continuous martingale. Then the following are equivalent:

1. X is closed,
2. $\{X_t\}$ is uniformly integrable,
3. X_t converges almost surely and in L^1 as $t \to \infty$.

Again, remember that the L^p condition was sufficient, while these assumptions are both necessary and sufficient.

Proof. (1) implies (2) because $\mathbb{E}[Z|\mathcal{G}]$ is always a uniformly integrable family. To show that (2) implies (3), note that uniformly integrable implies $\sup_t ||X_t||_1 < \infty$, so Proposition 61 tells us that X_t converges almost surely to X_∞, and Theorem 64 tells us that X_t converges in L^1 to X_∞ as well. Finally, to show that (3) implies (1), note that we have $X_t = \mathbb{E}[X_u|\mathcal{F}_t]$ for all $t \le u < \infty$. But X_u converges to X_∞ in L^1 by assumption, so we can pass the limit through the integral and find that $X_t = \mathbb{E}[X_\infty|\mathcal{F}_t]$. Therefore taking $Z = X_\infty$ shows that the martingale is closed. □

This result was the main step in proving the discrete-time optional stopping theorem, but the continuous-time case makes a few things more complicated. Remember that if X_n is **discrete** and adapted to $\mathcal{F}_n$ and τ is a stopping time, then we define the stopping-time sigma-algebra

$$X_\tau \in \mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \le n\} \in \mathcal{F}_n \quad \forall n\}.$$

Then to check that X_τ is actually measurable with respect to $\mathcal{F}_\tau$, we just needed to check if $\{X_\tau \in B\} \cap \{\tau \le n\} \in \mathcal{F}_n$. But this is just a finite union of events

$$\{\tau \le n\} \cap \left(\bigcup_{k=1}^{n} \{X_k \in B, \tau = k\} \right),$$

which is indeed in $\mathcal{F}_n$. But we immediately get measurability issues if we use continuous time because we'd have to take an infinite union, so claiming that $X_\tau \in \mathcal{F}_\tau$ does actually require some regularity assumptions.

Proposition 67

Suppose X_t is adapted to $\mathcal{F}_t$ and has right-continuous sample paths, and let τ be a stopping time. Then $X_\tau \in \mathcal{F}_\tau$.

Proof. We'll prove that X_τ is a composition of two maps. Fix $t > 0$, and consider the map $F : \Omega \times [0, t] \to \mathbb{R}$ sending $F(\omega, s) = X_s(\omega)$. We claim that this map is measurable with respect to $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$ (where $\mathcal{B}$ denotes the Borel sigma-algebra) – this property is called being **progressive**, and it's stronger than being adapted. To show this, we'll approximate F with something that has this property: we split our interval $[0, t]$ into blocks of length $\frac{t}{n}$, and define

$$F^{(n)}(\omega, s) = X_{(\lceil s \cdot n/t \rceil)/(n/t)}(\omega)$$

(meaning that we take the value of X at the right edge of the block). Each $F^{(n)}$ is measurable, because the preimage of any set B can be written as

$$(F^{(n)})^{-1}(B) = \left[\bigcup_{k=1}^{n} \left\{ \{X_{kt/n} \in B\} \times \left(\frac{(k-1)t}{n}, \frac{kt}{n} \right] \right\}\right] \cup (\{X_0 \in B\} \times \{0\}),$$

which is just a collection of rectangles. As we take $n \to \infty$, $F^{(n)}$ converges to F by right-continuity (because $(\lceil s \cdot n/t \rceil)/(n/t) \downarrow s$ as n gets large), and the pointwise limit of measurable functions is measurable. So now we want to show that $X_\tau \in \mathcal{F}_\tau$, which means that we need to show that

$$\{X_\tau \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$$

for all t. For any fixed t, define $G(\omega) = (\omega, t \wedge \tau(\omega))$, and take F from above. Then

$$(\Omega, \mathcal{F}_t) \xrightarrow{G} (\Omega \times [0, t], \mathcal{F}_t \otimes \mathcal{B}_{[0,t]}) \xrightarrow{F} (\mathbb{R}, \mathcal{B}_\mathbb{R})$$

is a composition of two measurable maps, and notably

$$\{X_\tau \in B\} \cap \{\tau \leq t\} = \{\tau \leq t\} \cap \{\omega : F(G(\omega)) \in B\},$$

because $t \wedge \tau(\omega) = \tau$ on the event $\tau \leq t$. And now $\tau \leq t$ is in $\mathcal{F}_t$ by the definition of a stopping time, and similarly $\{\omega : F(G(\omega)) \in B\}$ is measurable because it's the preimage of a measurable set. Thus we have the desired result. □

Let's look now at another result from discrete-time that we have to be more careful about in the continuous-time case. In discrete time, the stopped process $X_{n\wedge\tau}$ can be written as the sum

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

which is the same as the H-transform $X_0 + (H \cdot X)_n$, where $H_k = 1\{\tau \geq k\}$ is $\mathcal{F}_{k-1}$-measurable. Since each X_n is integrable by definition of a martingale, $X_{n\wedge\tau}$ will be in L^1 for all n (since it's a sum over $n + 1$ terms each in L^1); from there, we can check that $X_{n\wedge\tau}$ is a martingale using the H-transform property. But both of these facts are less obvious in the continuous-time case, and we have to do a bit more.

Theorem 68

Let X_t be a **uniformly integrable** martingale with **right-continuous sample paths**, and let σ, τ be stopping times such that $\sigma \leq \tau$. Then X_σ, X_τ are both in L^1, and $X_\sigma = \mathbb{E}[X_\tau | \mathcal{F}_\sigma]$.

Proof. We'll sketch the proof here: **assume that we know the discrete-time result already**. Approximate the stopping times by

$$\sigma_n = \frac{\lceil \sigma \cdot 2^n \rceil}{2^n}, \quad \tau_n = \frac{\lceil \tau \cdot 2^n \rceil}{2^n}.$$

Clearly $\sigma_n \downarrow \sigma, \tau_n \downarrow \tau, \sigma_n \leq \tau_n$, and σ_n, τ_n are stopping times for all n (because they're larger than σ, τ respectively, so we "know to stop" at some time later). Then the discrete time optional stopping theorem tells us that $X_{\sigma_n} = \mathbb{E}[X_{\sigma_0} | \mathcal{F}_{\sigma_n}]$, so the X_{σ_n} are uniformly integrable. Because X_{σ_n} converge almost surely to X_σ by right-continuity, that means that X_{σ_n} also converges in L^1 to X_σ. Similarly, X_{τ_n} converges almost surely and in L^1 to X_τ. But (by the discrete-time result) we have

$$X_{\sigma_n} = \mathbb{E}[X_{\tau_n} | \mathcal{F}_{\sigma_n}],$$

and any event $A \in \mathcal{F}_\sigma$ is contained in all $\mathcal{F}_{\sigma_n}$ and therefore in their intersection. Thus $\mathbb{E}[X_{\sigma_n}; A] = \mathbb{E}[X_{\tau_n}; A]$ for all events $A \in \mathcal{F}_\sigma$, and taking $n \to \infty$ yields (by L^1 convergence) $\mathbb{E}[X_\sigma; A] = \mathbb{E}[X_\tau; A]$. This is exactly the definition of the conditional expectation $X_\sigma = \mathbb{E}[X_\tau | \mathcal{F}_\sigma]$. □

With this, we can finally prove the result that we're after:

Corollary 69

Let X_t be a martingale with right-continuous sample paths, and let τ be a stopping time. Then (1) $X_{t \wedge \tau}$ is a martingale, and (2) if $\{X_t\}$ is uniformly integrable, then $\{X_{t \wedge \tau}\}$ is also uniformly integrable with $X_{t \wedge \tau} = \mathbb{E}[X_\tau | \mathcal{F}_t]$.

Proof. We prove (2) first. We know that τ and $t \wedge \tau$ are both stopping times, so Theorem 68 tells us that $X_\tau, X_{t \wedge \tau}$ are in L^1 and that $X_{t \wedge \tau} = \mathbb{E}[X_\tau | \mathcal{F}_{t \wedge \tau}]$. Now we wish to show that for all A in $\mathcal{F}_t$,

$$\mathbb{E}[X_{t \wedge \tau}; A] = \mathbb{E}[X_\tau; A].$$

On the event that $\tau \leq t$, the expressions inside the expectations are identical, so $\boxed{\mathbb{E}[1_A 1_{\tau \leq t} X_{t \wedge \tau}] = \mathbb{E}[1_A 1_{\tau \leq t} X_\tau]}$. On the other hand, on the event that $\tau > t$, $A \cap \{\tau > t\}$ is in $\mathcal{F}_t$, and it is also in $\mathcal{F}_\tau$ (by the definition of the stopping time sigma-algebra), so it is in their intersection, which is $\mathcal{F}_{t \wedge \tau}$. In other words,

$$\boxed{\mathbb{E}[1_A 1_{\{\tau > t\}} X_{t \wedge \tau}] = \mathbb{E}[1_A 1_{\{\tau > t\}} X_\tau]},$$

by using $X_{t \wedge \tau} = \mathbb{E}[X_\tau | \mathcal{F}_{t \wedge \tau}]$ and bringing $1_A 1_{\tau > t}$ inside the inner expectation. Adding together the boxed expressions shows the result.

Now to show (1) from (2), we know that $X_{t \wedge \tau}$ is measurable with respect to $\mathcal{F}_{t \wedge \tau} \subseteq \mathcal{F}_t$, so our process is indeed adapted to the filtration. Also, Theorem 68, $X_{t \wedge \tau}$ is in L^1, so it is integrable. Thus, we just need to show that $X_{s \wedge \tau} = \mathbb{E}[X_{t \wedge \tau} | \mathcal{F}_s]$. If we fix any finite t and define $Y_s = X_{s \wedge t}$ (so we run the process only for a finite time), then Y_s is uniformly integrable because it is closed – there is a single random variable X_t with conditional expectations $Y_s = \mathbb{E}[X_t | \mathcal{F}_{s \wedge \tau}]$. So applying (2), we see that $Y_\tau = X_{\tau \wedge t}$ is in L^1, and $Y_s = \mathbb{E}[Y_t | \mathcal{F}_s]$. In terms of the original process, this means that $X_{s \wedge \tau} = \mathbb{E}[X_{t \wedge \tau} | \mathcal{F}_s]$, showing the martingale condition as desired. □

8 February 26, 2020

Today, we're starting with Chapter 4 of our textbook, which discusses **continuous semimartingales**. In short, these are processes of the form $X_t = M_t + A_t$ where M_t is a **continuous local martingale** and A_t is a **finite variation process**. We haven't defined any of these terms or classified any of these processes yet, which will be the topic of the

next few classes. But before that, we'll take a step back and look at where things are headed (because the last few lectures have been a bit technical).

Example 70

Consider the standard Brownian motion B_t, which is a continuous martingale. Suppose that we're interested in studying how $X_t = f(B_t)$ evolves for some smooth function f.

Heuristically, we can use Itô's formula, which tells us something of the form

$$dX_t = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt$$

(where the main idea is that we've replaced the $(dB_t)^2$ term with dt). In particular, this means X is not a martingale unless f is linear, which makes sense: a linear scaling $\mu + \sigma B_t$ should still be a martingale, but otherwise we get a drift term caused by the curvature of our function f''. (Intuitively, a positively curved function f makes $f(B_{t+dt})$ larger than the linear approximation $f(B_t) + f'(B_t)dB_t$.) So Itô's formula gives us a decomposition

$$X_t = \left[\int_0^t f'(B_s)dB_s\right] + \left[\int_0^t \frac{1}{2}f''(B_s)ds\right],$$

which turns out to exactly be the process $M_t + A_t$ we'll be constructing. So this definition is really coming out of **manipulations of Brownian motion**.

Throughout this discussion, it might be useful to keep the discrete-time picture in mind. Say that a process X_n is adapted to $\mathcal{F}_n$, where $X_0 = 0$ and $\mathbb{E}|X_n| < \infty$ for all n. Then we can decompose

$$X_n = \left[\sum_{i=1}^n (X_i - X_{i-1} - \mathbb{E}[X_i - X_{i-1}|\mathcal{F}_{i-1}]\right] + \left[\sum_{i=1}^n \mathbb{E}[X_i - X_{i-1}|\mathcal{F}_{i-1}]\right],$$

upon which the first term corresponds to the martingale M_n and the second term to the finite variation process A_n. It may not be entirely clear how this A_n relates to the A_t above, though: to make that more apparent, imagine that our process X_n is a function $f(S_n)$, where $S_n = \varepsilon \sum_{i=1}^n \zeta_i$, $\zeta_i \sim \text{Unif}\{\pm 1\}$ is a random walk with step size ε. Then

$$A_{n+1} - A_n = \mathbb{E}[f(S_n + \varepsilon\zeta_{n+1}) - f(S_n)|\mathcal{F}_n],$$

and now we can Taylor expand in ε to write this as

$$f'(S_n)\mathbb{E}[\zeta_{n+1}|\mathcal{F}_n] + \frac{1}{2}\varepsilon^2 f''(S_n)\mathbb{E}[\zeta_{n+1}^2|\mathcal{F}_n] + \cdots$$

(taking out the constants and derivatives because they're measurable with respect to S_n). But the ζs are iid symmetric random signs, so this simplifies to

$$A_{n+1} - A_n = \frac{1}{2}\varepsilon^2 f''(S_n) + o(\varepsilon^2),$$

which now looks identical to the A_t term above. So to summarize, we want to understand this decomposition in the continuous case: chapters 4 and 5 will help us with a formal characterization of this class of processes. Along the way, we'll prove Itô's formula, which essentially tells us that **the image of a continuous semimartingale under a smooth map is another continuous semimartingale**: that is, $h(M_t + A_t) = \tilde{M}_t + \tilde{A}_t$ for sufficiently nice h. Heuristically, the idea is that if $X_t = M_t + A_t$, then

$$\begin{aligned} dh(M_t + A_t) &= h'(X_t)dX_t + \frac{1}{2}h''(X_t)(dX_t)^2 \\ &= h'(X_t)(dM_t + dA_t) + \frac{1}{2}h''(X_t)\left((dM_t)^2 + 2dM_t dA_t + (dA_t)^2\right). \end{aligned}$$

A step of the martingale dM_t is "like a step of the Brownian motion with some variance," so dM_t is a Gaussian step with variance of order $\sqrt{dt}$, and dA_t is something deterministic of order dt. So we can toss the last two terms, and we're just left with the three terms of highest order:

$$dh(M_t + A_t) = h'(X_t)dM_t + \left(h'(X_t)dA_t + \frac{1}{2}h''(X_t)(dM_t)^2\right),$$

where the $h'(X_t)$ term will be the martingale part and the rest is the finite variation part. We don't really know how to integrate any of these terms yet, so we have a lot to understand – we'll start with the $h'(X_t)dA_t$ term. Our first task is to understand **what a finite variation process is and how we take an integral against it**.

Today, we'll start with the case where A_t is some deterministic function of time and X_t is some deterministic process. The reason for eliminating the ω here is that t will "be our ω."

Definition 71

Let $(\Omega, \mathcal{F})$ be a measurable space (not the probability space of our process). A **finite signed measure** on $(\Omega, \mathcal{F})$ is a function $\alpha : \mathcal{F} \to \mathbb{R}$ such that α is countably additive, meaning that if we have disjoint sets $A_i \in \mathcal{F}$, then

$$\alpha\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \alpha(A_i),$$

where the sum must be absolutely convergent.

Example 72

If α_+, α_- are (actual nonnegative) measures on $(\Omega, \mathcal{F})$, then $\alpha = \alpha_+ - \alpha_-$ is a signed measure. Also, if μ is a measure on $(\Omega, \mathcal{F})$ such that $\int_\Omega |h|\, d\mu < \infty$, then $\nu(E) = \int_E h\, d\mu$ is also a signed measure – in fact, $h = \frac{d\nu}{d\mu}$ is the Radon-Nikodym derivative.

It turns out that we can only get a signed measure by writing $\alpha = \alpha_+ - \alpha_-$ as the difference of two measures. Let's see how to show this:

Definition 73

Let α be a signed measure on $(\Omega, \mathcal{F})$. Then $A \in \mathcal{F}$ is **positive** if $\alpha(B) \geq 0$ for all $B \subseteq A$ (similarly **negative** if $\alpha(B) \leq 0$), and A is a **null set** if $\alpha(B) = 0$ for all $B \subseteq A$.

Theorem 74 (Hahn decomposition theorem)

For any signed measure α on $(\Omega, \mathcal{F})$, there is a bipartition $\Omega = \Omega_+ \sqcup \Omega_-$ such that Ω_+ is positive and Ω_- is negative. This decomposition is essentially unique – for any other decomposition $\Omega = B_+ \sqcup B_-$, the sets $B_+ \cap \Omega_-, B_- \cap \Omega_+$ must be null sets.

We should refer to the textbook from 18.675, [3], for the proof – it's about a paragraph long.

Theorem 75 (Jordan decomposition theorem)

Any signed measure can be uniquely written as $\alpha = \alpha_+ - \alpha_-$, where α_+, α_- are measures on $(\Omega, \mathcal{F})$.

Proof. Let $\alpha_+(E) = \alpha(E \cap \Omega_+)$ and $\alpha_-(E) = \alpha(E \cap \Omega_-)$ – we can check that this is unique. □

We can now connect this to the idea of a finite variation process:

Definition 76

A continuous function $a : [0, T] \to \mathbb{R}$ is **of finite/bounded variation** (FV or BV) if there exists a signed measure α on $([0, T], \mathcal{B}_{[0,T]})$ such that $a(t) = \alpha([0, t])$ for all $0 \leq t \leq T$.

By the Jordan decomposition, this means we can write

$$a(t) = \alpha_+([0, t]) + \alpha_-([0, t])$$

where the two terms on the right hand side are nondecreasing functions of t (because α_+, α_- are actual measures). We will sometimes refer to these as $a_+(t)$ and $a_-(t)$. Then the measure $\mu = |\alpha| = \alpha_+ + \alpha_-$ is the **total variation measure** of a, and we define $v(t) = \mu([0, t])$ to be the "total variation of a on the interval $[0, t]$." So now if we want to define an integral against the function $a(s)$, one natural way to do so is to define it in terms of our measure:

$$\int_0^T f(s)da(s) = \int_0^T f(s)\alpha(ds),$$

and similarly $\int_0^T f(s)|da(s)| = \int_0^t f(s)\mu(ds)$. It's important to emphasize that in both of these equations, **the left-hand side is new notation**, while the right-hand side is just a Lebesgue integral which we know how to compute. In order for these integrals to be well-defined, we just need to make sure that f is measurable and absolutely integrable (that is, $\int_0^T |f(s)|\mu(ds) < \infty$). Here are two simple properties of this integral:

- By Jensen's inequality, we have $\left|\int_0^T f(s)da(s)\right| \leq \int_0^T |f(s)||da(s)|$.
- The function $b(t) = \int_0^t f(s)da(s)$ is also of finite variation – this will be important for something like Itô's formula. Indeed, we can define $b(t) = \beta([0, t])$, where β is the signed measure $\beta(E) = \int_E f(s)\alpha(ds)$ (this should look like the Radon-Nikodym derivative equation). Now decomposing into positive and negative parts for both f and α, the explicit formula for β is that

$$\beta(E) = \int_E (f_+(s)\alpha_+(ds) + f_-(s)\alpha_-(ds)) - \int_E (f_+(s)\alpha_-(ds) + f_-(s)\alpha_+(ds)).$$

So to recap, we're considering the space $(\Omega, \mathcal{F}) = ([0, t], \mathcal{B})$, and we have a signed measure α corresponding to a decomposition $\Omega = \Omega_+ \sqcup \Omega_-$. This yields a corresponding decomposition $\alpha = \alpha_+ - \alpha_-$ and $\mu = \alpha_+ + \alpha_-$. Note that α_+ is absolutely continuous with respect to μ (because if a set has zero measure under μ, then it also has zero measure under α_+) and similarly, $\alpha_- \ll \mu$. Furthermore, we can write down the Radon-Nikodym derivatives using the defining properties: we have

$$\frac{d\alpha_+}{d\mu} = 1_{\Omega_+}, \quad \frac{d\alpha_-}{d\mu} = 1_{\Omega_-}$$

(because we want $\alpha_+(E) = \int_E \frac{d\alpha_+}{d\mu} d\mu$ to be the integral over only Ω_+, and similar logic for α_-); call these two functions h_+ and h_-. Then the corresponding finite-variation process $a : [0, t] \to \mathbb{R}$ can be written as

$$a(s) = \alpha([0, s]) = \int_0^s h(s)\mu(ds), \text{ where } h = \frac{d\alpha}{d\mu} = h_+ - h_-.$$

Decomposing as $a(s) = a_+(s) - a_-(s)$, we see that the total variation is $v(s) = a_+(s) + a_-(s)$. Therefore,

$$a_+(s) = \frac{1}{2}(v(s) + a(s)), \quad a_-(s) = \frac{1}{2}(v(s) - a(s)).$$

Remember that our eventual goal is to understand the quantity $\int_0^t h'(X_s)dA_s$. Once we introduce randomness back in the process, our processes A and X will usually be correlated, so we need to ask how we can calculate the integral

(for example, whether we can do it using time-steps like a Riemann integral). The first step is to look at a simpler quantity:

Lemma 77

Let $a : [0, t] \to \mathbb{R}$ be a function of finite variation. Then

$$\int_0^t |da(s)| = \sup\{V_P(a) : P \text{ subdivision of } [0, t]\},$$

where P is of the form $\{0 = t_0 < t_1 < \cdots < t_p = t\}$ and $V_P(a) = \sum_{i=1}^p |a(t_i) - a(t_{i-1})|$. In addition, if we have an increasing set of subdivisions $P_1 \subseteq P_2 \subseteq \cdots$ and the mesh of P_n goes to zero, then $V_{P_n}(a) \to \int_0^t |da(s)|$.

Basically, we can approximate total variation by a discrete subdivision. This proof is similar to the Radon-Nikodym theorem, but we'll present a self-contained argument here:

Proof. We know that $V_P(a) \le \int_0^t |da(s)|$ for all P by Jensen, so it suffices to show the convergence result (in other words, show that we do approach $\int_0^a |da(s)|$). By scaling, we can assume without loss of generality that $\mu([0, t]) = 1$ so that $([0, t], \mathcal{B}, \mu)$ is a probability space. Let $\mathcal{G}_n$ be the sigma-algebra generated by the intervals of P_n – for each fixed n, $\mathcal{G}_n$ is a finite collection, and the $\mathcal{G}_n$ are nondecreasing because the P_n are nondecreasing. Then the sigma-algebra σ generated by the union of the $\mathcal{G}_n$s is just the Borel sigma-algebra on $[0, t]$.

By the Hahn decomposition, we can write $X(t) = h(t) = 1\{t \in \Omega_+\} - 1\{t \in \Omega_-\}$. X is a random variable on $([0, t], \mathcal{B}, \mu)$, and we have the filtration $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \cdots \subseteq \mathcal{B}$, which gives us the **closed** martingale (implying uniformly integrable as well) $X_n = \mathbb{E}[X|\mathcal{G}_n]$. This means that X_n converges to X almost surely and in L^1, but because $\mathcal{G}_n$ is finite, X_n is **piecewise constant** (on the intervals of P_n). Then (notation meaning here that the break points depend on n) we can write

$$\mathbb{E}\left[X_n; [t_{i-1}^{(n)}, t_i^{(n)}]\right] = X_n(t_{i-1}^{(n)}) \cdot \mu([t_{i-1}^{(n)}, t_i^{(n)}])$$

(where technically it's possible to modify X at any point, but the point is that it's the measure of the interval times the value at any given point in that interval). But we also know that on this same interval (recalling that X and h are the same) we have

$$\mathbb{E}[X; [t_{i-1}, t_i]] = \int_{t_{i-1}}^{t_i} h(s)\mu(ds) = \alpha([t_{i-1}, t_i]) = a(t_i) - a(t_{i-1}).$$

But these two expectations should be equal, because $X_n = \mathbb{E}[X|\mathcal{G}_n]$, so **the value of X_n on $[t_{i-1}, t_i]$ is equal to $\frac{a(t_i)-a(t_{i-1})}{\mu([t_{i-1},t_i])}$.** Now because X_n converges to X in L^1, $\mathbb{E}[|X_n|]$ converges to $\mathbb{E}[|X|]$. X_n is piecewise constant, so we can write

$$\mathbb{E}[|X_n|] = \sum_{i=1}^{|P_n|} \mu([t_{i-1}, t_i]) \cdot \left|\frac{a(t_i) - a(t_{i-1})}{\mu([t_{i-1}, t_i])}\right| = \sum_{i=1}^{|P_n|} |a(t_i) - a(t_{i-1})| = V_{P_n}(a).$$

Meanwhile, $|X|$ is 1 almost surely (it's either 1 or -1 almost everywhere), so

$$\mathbb{E}[|X|] = 1 = \mu([0, t]) = \int_0^t |da(s)|.$$

Thus as $n \to \infty$, the convergence $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$ yields the result. □

Corollary 78

Let a be a function of finite variation and $f : [0,t] \to \mathbb{R}$ be a continuous function. Then if $P_1 \subseteq P_2 \subseteq \cdots$ are subdivisions of $[0,t]$ with mesh going to 0, then

$$\sum_{i=1}^{|P_n|} f(t_{i-1}^{(n)}) \left(a(t_i^{(n)}) - a(t_{i-1}^{(n)}) \right) \to \int_0^t f(s) da(s),$$

where again the t_i implicitly depend on n.

This is useful because we have more practice looking at things like the left hand side, so we'll be able to simplify calculations.

Proof. We start by trying to define the Stieltjes integral: let $t_i^{(n)}$ be our break points corresponding to the subdivision P_n, and define $f^{(n)}(t) = f(t_{i-1}^{(n)})$ for all $t \in [t_{i-1}^{(n)}, t_i^{(n)})$. The $f^{(n)}$ are bounded uniformly, because $||f^{(n)}||_\infty \leq ||f||_\infty < \infty$, and $f^{(n)}$ converges to f pointwise because f is continuous. But then the left hand side of the equation is $\int_0^t f^{(n)} d\alpha$, which converges to $\int_0^t f d\alpha$ by the dominated convergence theorem. □

Remark 79. *We say that a function $a : [0,\infty) \to \mathbb{R}$ has finite variation if a has finite variation on any compact interval $[0,t]$.*

9 March 2, 2020

Solutions for the first two homework assignments are posted on Stellar now; the next homework is due on Monday, and we'll have office hours 3-5 on Thursday instead of today.

Recall that a function $a : [0,T] \to \mathbb{R}$ is of **finite/bounded variation** if there exists a signed measure $\alpha = \alpha_+ - \alpha_-$ on $[0,T]$ such that $a(t) = \alpha([0,t]) = a_+(t) - a_-(t)$ (noting that $a(0) = 0$). Last time, we defined the integral

$$\int_0^t f(s) ds = \int_0^t f(s) \alpha(ds),$$

which is well-defined as long as we have the absolute integrability condition $\int_0^t |f(s)||da(s)| < \infty$. We noted that the integral $\int_0^t f(s) da(s)$, as a function of t, is also of finite variation, and we noticed that if f is continuous and we have a sequence of subdivisions $P_0 \subseteq P_1 \subseteq \cdots$ of $[0,t]$ with mesh going to zero, then the discrete approximations $\sum_{i=1}^{|P_n|} f(t_{i-1}^{(n)}) \left(a(t_i^{(n)}) - a(t_{i-1}^{(n)}) \right)$ converge to the integral $\int_0^t f(s) da(s)$.

Everything last time was deterministic, so we'll add randomness now. We'll be on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ for the rest of this lecture, and for the rest of this chapter we'll assume that we have continuous sample paths.

Definition 80

An adapted process A_t (to our filtered probability space) is a **finite variation process** if all sample paths have finite variation. If all sample paths are nondecreasing in t, then we say that A_t is an **increasing process**.

Recall that a process H_t is **progressive** if $F(\omega, s) = H_s(\omega)$ is measurable with respect to the sigma-algebra $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$. In general, this is stronger than being adapted, but any **continuous** adapted process will be progressive.

Proposition 81

Let A be a finite-variation process, and let H be a progressive process. Suppose that $\int_0^t |H_s(\omega)||dA_s(\omega)|$ is finite for all finite t. Then $(H \cdot A)_t = \int_0^t H_s dA_s$ is a well-defined finite variation process.

We'll skip over this proof for now, but the fact that this is an FV process follows from the result above that $\int_0^t f(s)da(s)$ is FV. What we need to check is that this process is measurable with respect to $\mathcal{F}_t$ –we should read this on our own, and this is where we use the progressive condition.

Remember that our original goal was to study the class of continuous semi-martingales of the form $X_t = A_t + M_t$, where A is FV and M is a **(local) martingale**. We'll spend some time now on the latter object.

Definition 82

A **continuous local martingale** M_t on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is an adapted continuous process for which there exists a sequence of stopping times (τ_n) such that

- $\tau_n(\omega) \uparrow \infty$ for all ω,
- The stopped processes $(M_{t\wedge\tau_n} - M_0)$ are uniformly integrable martingales for all n (we often say that τ_n **reduces** M).

Example 83

Let B_t be a Brownian motion in $\mathbb{R}^3$, and define $M_t = \frac{1}{|B_t|} = (B_{t,1}^2 + B_{t,2}^2 + B_{t,3}^2)^{-1/2}$. This is not a true martingale, but it is a local martingale. (We'll talk more about the difference in a future lecture.)

Note that all continuous martingales are continuous local martingales, because we can just take $\tau_n = n$: we know that $X_s = \mathbb{E}[X_n|\mathcal{F}_s]$ for all $s \leq n$, so the variables $(M_{t\wedge\tau_n} - M_0)$ are just conditional expectations of X_n and thus must be uniformly integrable. This same conditional expectation argument actually implies that **we can actually leave out "uniformly integrable" from the definition**.

Remark 84. *The optional stopping theorem tells us that a stopped uniformly integrable martingale is still uniformly integrable, so any stopping time τ of a continuous local martingale M gives us a continuous local martingale $M_{t\wedge\tau}$.*

Note that one set of stopping times to consider (which we can **always** use) is

$$\tau_n = \inf\{t : |M_t - M_0| = n\}.$$

The idea is that at any finite n, all of the stopped processes $M_{t\wedge\tau_n}$ are bounded, so we have uniform integrability.

Proposition 85

Let M be a **nonnegative** continuous local martingale with $M_0 \in L^1$ (we need to add this condition because it's no longer true that M_t needs to be integrable in general). Then M is a true supermartingale but not necessarily a martingale.

Proof. Let $N_t = M_t - M_0$, and let τ_n be the reducing sequence for M. Then for all $s \leq t$, we have

$$N_{s\wedge\tau_n} = \mathbb{E}[N_{t\wedge\tau_n}|\mathcal{F}_s]$$

because τ_n creates a uniformly integrable martingale, so the optional stopping theorem holds. Since M_0 is integrable, we can add it back to both sides to find

$$M_{s\wedge\tau_n} = \mathbb{E}[M_{t\wedge\tau_n}|\mathcal{F}_s].$$

Taking limits on both sides, $M_{s\wedge\tau_n}$ converges to M_s by continuity, while the right hand side satisfies

$$\lim_{n\to\infty} \mathbb{E}[M_{t\wedge\tau_n}|\mathcal{F}_s] \geq \mathbb{E}\left[\liminf_{n\to\infty} M_{t\wedge\tau_n}|\mathcal{F}_s\right] = \mathbb{E}[M_t|\mathcal{F}_s]$$

by Fatou's lemma and continuity, so we get the supermartingale inequality. □

Theorem 86

If X is both a continuous local martingale and a finite variation process, then $X = 0$.

This is a useful result to have, because this result gives us uniqueness of the decomposition $X_t = M_t + A_t$.

Proof. Since X is a finite variation process, we know that $X_0 = 0$ and $X_t = \int_0^t dX_s$. Then because X is also a continuous local martingale, we can define

$$\tau_n = \inf\{t : \int_0^t |dX_s| = n\}$$

to be the first time the total variation of X exceeds n. Letting $X_t^{(n)} = X_{t\wedge\tau_n}$, we know that $|X_t^{(n)}| \leq n$ for all t (because we stop the process before it can change by more than n), which means that $X^{(n)}$ is bounded and therefore uniformly integrable for each n. Fix n and let $N = X^{(n)}$ for notation (note that N is a martingale). We have

$$\mathbb{E}[(N_{s_2} - N_{s_1})(N_{t_2} - N_{t_1})] = 0$$

for all $s_1 \leq s_2 \leq t_1 \leq t_2$ (because when we condition on $\mathcal{F}_{t_1}$, the martingale condition tells us this expectation is zero, and then taking another expectation gives us 0 overall). This means that we can break up the sum as

$$\mathbb{E}[N_t^2] = \mathbb{E}\left[\sum_i (N_{t_i} - N_{t_{i-1}})^2\right] = \sum_i \mathbb{E}[(N_{t_i} - N_{t_{i-1}})^2],$$

but the fact that $\int_0^t |dX_s| = n$ means that summing squares of increments will give us something small. Specifically,

$$\mathbb{E}[N_t^2] \leq \mathbb{E}\left[\left(\sup_i |N_{t_i} - N_{t_{i-1}}|\right) \cdot \sum_i |N_{t_i} - N_{t_{i-1}}|\right].$$

The sum is at most n by definition, and the first term goes to 0 as the mesh of $\{t_i\}$ goes to zero. Because the whole expression inside the expectation is bounded, we know that this goes to 0 by the dominated convergence theorem. Thus $\mathbb{E}[N_t^2] = 0$, meaning that $N_t = X_t^{(n)} = X_{t\wedge\tau_n}$ is identically zero for all n, which can only happen if $X = 0$. □

From here on, we assume that $\mathcal{F}_t$ is complete.

Theorem 87

Let M be a continuous local martingale. Then there exists an increasing (finite variation) process $A_t = \langle M, M\rangle_t$ called the **quadratic variation**, such that $M_t^2 - A_t$ is a continuous local martingale. A is unique up to null sets, and if $P_1 \subseteq P_2 \subseteq \cdots$ is a subdivision of $[0, t]$, then $V_{P_n}^2(M)$ converges to A_t as $n \to \infty$.

Note that A is the formalization of what we labeled as $\int_0^t (dM_s)^2$ in our earlier heuristic arguments. This is a really important quantity: as motivation for why we care about this, if M is a continuous local martingale and $A_t = \langle M, M\rangle_t$, then M_t is equal in distribution to a Brownian motion indexed by A_t. So **a continuous local martingale is just a time change of Brownian motion** (and in the special case where $\langle M, M\rangle_t = t$, M is just a standard Brownian motion).

Start of proof. We'll prove this **assuming that M is a bounded continuous martingale**. (Generalizing to continuous local martingales takes very little work.) Again, we assume $M_0 = 0$ for simplicity.

For motivation, first consider a discrete-time martingale M_n. We can write

$$M_n^2 = \sum_{i=1}^{n} \left(M_i^2 - M_{i-1}^2 - \mathbb{E}[M_i^2 - M_{i-1}^2|\mathcal{F}_{i-1}]\right) + \sum_{i=1}^{n} \mathbb{E}[M_i^2 - M_{i-1}^2|\mathcal{F}_{i-1}],$$

where the first summation is a martingale. So a natural idea is to take a discrete approximation for our continous-time martingale, and hope that the remainder converges to our process A. It turns out that we don't even need the expectations – we can take $A^{(n)} = \sum \left(M_{t_i} - M_{t_{i-1}}\right)^2$. More formally (in continuous time), M is a continuous bounded martingale, so we can take our increasing subdivisions of $[0, T]$ (with mesh going to zero) and define

$$A_t^{(n)} = \sum_{i=1}^{|P_n|} \left(M_{t\wedge t_i^{(n)}} - M_{t\wedge t_{i-1}^{(n)}}\right)^2.$$

Because $M_t^2 = \sum_{i=1}^{|P_n|} \left(M^2_{t\wedge t_i^{(n)}} - M^2_{t\wedge t_{i-1}^{(n)}}\right)$, subtracting the two and expanding yields

$$N_t^{(n)} = M_t^2 - A_t^{(n)} = 2\sum_{i=1}^{|P_n|} M_{t\wedge t_{i-1}^{(n)}} \left(M_{t\wedge t_i^{(n)}} - M_{t\wedge t_{i-1}^{(n)}}\right),$$

which is a bounded martingale (by assumption, since M is a bounded martingale).

Lemma 88

Suppose that $|M_t| \le C$ for all $t \le T$, and suppose that $M_0 = 0$. Then

$$\mathbb{E}\left[\left(\sum_i (M_{t_i} - M_{t_{i-1}})^2\right)^2\right] \le 12C^4.$$

(Without the assumption that $M_0 = 0$, the right-hand side is instead $48C^4$.)

Proof of lemma. Expanding out the left hand side, we get diagonal terms for $i = j$ and off-diagonal terms otherwise:

$$\mathbb{E}\left[\left(\sum_i (M_{t_i} - M_{t_{i-1}})^2\right)^2\right] = \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^4] + 2\sum_{i<j} \mathbb{E}\left[(M_{t_i} - M_{t_{i-1}})^2 (M_{t_j} - M_{t_{j-1}})^2\right].$$

Simplify the first term by pulling out some factors (and bounding them by C), and compute the sum over $j > i$ in the second term to get

$$\mathbb{E}\left[\left(\sum_i (M_{t_i} - M_{t_{i-1}})^2\right)^2\right] \le (2C)^2 \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2] + 2\sum_i \mathbb{E}\left[(M_{t_i} - M_{t_{i-1}})^2 (M_T - M_{t_i})^2\right],$$

where we've used the orthogonality of disjoint intervals (since $M_T - M_{t_i} = \sum_{j>i} M_{t_j} - M_{t_{j-1}}$). Now bound the blue

term in terms of C to show that

$$\mathbb{E}\left[\left(\sum_i (M_{t_i} - M_{t_{i-1}})^2\right)^2\right] \leq \left[(2C)^2 + 2(2C)^2\right] \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2] = 12C^2 \cdot \mathbb{E}[M_T^2]$$

again using orthogonality (here's where we gain the factor of 4 when $M_0 = 0$ isn't assumed), and this last expression is at most $12C^2 \cdot C^2 = 12C^4$, as desired. □

Lemma 89

If we still assume that M_t is bounded, then the sequence $(N_T^{(n)})$ is Cauchy in L^2 (as $n \to \infty$).

Proof of lemma. Suppose $m \leq n$, so that the subdivision P_m is contained in the subdivision P_n. Let $P_n = (s_j : 1 \leq j \leq |P_m|)$ and $P_m = (t_i : 1 \leq i \leq |P_n|)$; because P_n is a refinement of P_m, we can also index the t_is as $t_{i,k}$, where the first index i tells us that we are in the range $[s_{i-1}, s_i]$. Now the expected covariance can be calculated by writing out the definition of $N_T^{(n)}$:

$$\frac{1}{4}\mathbb{E}\left[N_T^{(m)} N_T^{(n)}\right] = \mathbb{E}\left[\sum_{i,j} M_{s_{i-1}}(M_{s_i} - M_{s_{i-1}}) M_{t_j}(M_{t_j} - M_{t_{j-1}})\right].$$

If the time intervals are disjoint, the contribution to the expectation is zero, so we only get a contribution when the t increments are inside the s increments and thus this expectation reduces to a sum over subincrements

$$\sum_{i,k} \mathbb{E}\left[M_{s_{i-1}}(M_{s_i} - M_{s_{i-1}}) M_{t_{i,k}}(M_{t_{i,k}} - M_{t_{i,k-1}})\right].$$

Each term here involves the times $s_{i-1} \leq t_{i,k-1} \leq t_{i,k} \leq s_i$, so we can further break up the increment $M_{s_i} - M_{s_{i-1}}$ and the only term that remains is the "middle one" (by conditioning at an appropriate time). Thus we have

$$\frac{1}{4}\mathbb{E}[N_T^{(m)} N_T^{(n)}] = \sum_{i,k} \mathbb{E}\left[M_{s_{i-1}}(M_{t_{i,k}} - M_{t_{i,k-1}}) M_{t_{i,k}}(M_{t_{i,k}} - M_{t_{i,k-1}})\right].$$

So now if we want to show that our sequence $N_T^{(n)}$ is Cauchy in L^2, we want to calculate the L^2 distance. Expanding and then simplifying the terms to make the sum line up, we find that

$$\begin{aligned}
\frac{1}{4}\mathbb{E}\left(N_T^{(m)} - N_T^{(n)}\right)^2 &= \sum_i \mathbb{E}[M_{s_{i-1}}^2 (M_{s_i} - M_{s_{i-1}})^2] - 2(\text{cross term above}) + \sum_{i,k} \mathbb{E}\left[M_{t_{i,k-1}}^2 (M_{t_{i,k}} - M_{t_{i,k-1}})^2\right] \\
&= \sum_{i,k} \mathbb{E}\left[(M_{s_i} - M_{t_{i,k-1}})^2 (M_{t_{i,k}} - M_{t_{i,k-1}})^2\right] \\
&\leq \mathbb{E}\left[\sup_{i,k} \left|M_{s_i} - M_{t_{i,k-1}}\right|^4\right]^{1/2} \mathbb{E}\left[\left(\sum_{i,k} (M_{t_{i,k}} - M_{t_{i,k-1}})^2\right)^2\right]^{1/2},
\end{aligned}$$

where the last step is by Cauchy-Schwarz. The second term is bounded by our previous lemma, and the first term converges to 0 again by dominated convergence theorem, so we do have Cauchy convergence in L^2. □

We'll finish up the proof next time, but this lemma is the main point – knowing that the process is Cauchy allows us to use the maximal inequality to get convergence. □

10 March 4, 2020

Today, we'll continue the proof that we started last time. The main idea is that we have a continuous local martingale M_t and want to decompose

$$M_t^2 = (\text{continuous local martingale}) + (\text{finite variation process}).$$

We know that this is unique, because any continuous local martingale that is of finite variation must be identically zero. Specifically, our goal is to show that (1) there exists a finite variation process $A_t = \langle M, M\rangle_t$ (called the **quadratic variation** of M) such that (2) $M_t^2 - A_t$ is a local martingale, and (3) if P_n is an increasing subdivision of $[0, t]$, we have

$$\sum_{i=1}^{|P_n|} \left(M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}}\right)^2 \xrightarrow{p} \langle M, M\rangle_t.$$

Continuation of proof of Theorem 87. Remember that we're proving this result assuming that M is bounded and that $M_0 = 0$. For any subdivision of $[0, T]$ and any $t \in [0, T]$, we can decompose

$$M_t^2 = \sum_{i=1}^{|P_n|} \left(M^2_{t\wedge t_i^{(n)}} - M^2_{t\wedge t_{i-1}^{(n)}}\right),$$

which can then further be decomposed as

$$\boxed{M_t^2 = \sum_{i=1}^{|P_n|} \left(M_{t\wedge t_i^{(n)}} - M_{t\wedge t_{i-1}^{(n)}}\right)^2 + 2\sum_{i=1}^{|P_n|} M_{t\wedge t_{i-1}^{(n)}} \left(M_{t\wedge t_i^{(n)}} - M_{t\wedge t_{i-1}^{(n)}}\right)}.$$

The first term, which we denote $A_t^{(n)}$, is supposed to approach our finite variation process, and we showed last time that the second term, which we denote $N_t^{(n)}$, is a martingale. We showed last time that $N_T^{(n)}$ is Cauchy in L^2 (as $n \to \infty$) for any finite T, as long as M is bounded. So now by Doob's L^2 inequality, there is some constant C such that

$$\sup_{0\le t\le T} \left|\left|N_t^{(n)} - N_t^{(m)}\right|\right|_2 \le C \left|\left|N_T^{(n)} - N_T^{(m)}\right|\right|_2.$$

As $m, n \to \infty$, the right-hand side goes to zero – because we have a martingale, having control of the endpoint gives us control of the entire process in the form of a kind of uniform convergence. In particular, we can find $n_k \to \infty$ (by extracting a subsequence) so that

$$\sup_{0\le t\le T} \left|\left|N_t^{(n_k)} - N_t^{(n_{k+1})}\right|\right|_2 \le \frac{1}{2^k},$$

which implies that (because $\mathbb{E}[|X|] = ||X||_1 \le ||X||_2$)

$$\mathbb{E}\left[\sum_{k=1}^{\infty} \sup_{0\le t\le T} \left|N_t^{(n_k)} - N_t^{(n_{k+1})}\right|\right] < \infty.$$

This means that the quantity inside the expectation is finite almost surely, so the $N^{(n_k)}$ converge uniformly on $[0, T]$ outside of a null set Ω_0. We can thus define

$$Y_t(\omega) = \begin{cases} \lim_{k\to\infty} N_t^{(n_k)}(\omega) & \omega \notin \Omega_0, \\ 0 & \text{otherwise.} \end{cases}$$

This process is adapted to $\mathcal{F}_t$, and we just need to check that Y_t is a martingale. But we've shown that $N_t^{(n_k)}$ converges

to Y_t almost surely and in L^2, so because each $N_t^{(n_k)}$ is a martingale, the statement $\mathbb{E}[Y_t|\mathcal{F}_s] = Y_s$ follows by taking a limit and passing the integral through (using L^2 convergence).

So now we can construct our process $A_t = M_t^2 - Y_t$ by taking the (uniform) limit (on $[0,T]$) of the $A_t^{(n)}$s, because M_t^2 is fixed and $N_t^{(n)}$ converges to Y_t in the boxed equation above. While it is not necessarily true that $A_t^{(n)}$ is an increasing process, we do know that each $A^{(n)}$ is nondecreasing on the set $\{t_i^{(n)}\}$. The set of t_is is dense in the limit, so continuity tells us that the limit A is indeed a **nondecreasing process** on $[0,T]$.

Finally, repeating this process for all integers $T \geq 1$ yields a collection of processes $M_t^2 = A_t^{(T)} + Y_t^{(T)}$ for $t \leq T$. We just need to show that all of these are compatible of each other, which follows from the uniqueness claim we made earlier. In particular, the stopped process $M_{t\wedge T}^2 - A_{t\wedge T}^{(T)}$ is a martingale, and if $T' \geq T$, then $M_{t\wedge T}^2 - A_{t\wedge T}^{(T')}$ is also a martingale. So if we subtract these, $A_t^{(T)}$ and $A_t^{(T')}$ must agree up to time T (because the difference is both a local martingale and a finite variation process), so this means A_t is indeed well-defined.

It just remains to check the final claim, but we have already proven that

$$\sum_{i=1}^{|P_n|} \left(M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}}\right)^2 = A_T^{(n)} = M_T^2 - N_T^{(n)}.$$

As $n \to \infty$, this converges in L^2 to $M_T^2 - Y_T$, which is exactly A_T, so in particular the convergence in probability follows as well. □

All of the discussion above proves the theorem in the case where M is bounded and $M_0 = 0$. Extending to the general case is easy, and we can read the details of that on our own.

Example 90

Let B be a standard Brownian motion. We've shown that $B_t^2 - t$ is a martingale, so $\langle B, B\rangle_t = t$.

In a future lecture, we'll see the converse as well, which tells us that a continuous local martingale with $\langle B, B\rangle_t$ is a Brownian motion. In fact, if M_t is both a continuous martingale and a Gaussian process, then $M_t^2 - \mathbb{E}[M_t^2]$ is a martingale, so we have $\langle M, M\rangle_t = \mathbb{E}[M_t^2]$. So in both of these cases the quadratic variation is deterministic, but in general it can be random.

Theorem 91

Suppose that M is a continuous local martingale, and $M_0 \in L^2$. Then the following are equivalent:

- M is a true martingale bounded in L^2,
- $\mathbb{E}[\langle M, M\rangle_\infty]$ is finite.

If these properties hold, then $M_t^2 - \langle M, M\rangle_t$ is also a true, uniformly integrable martingale.

Based on our work above, it's natural to claim that $M_t^2 - \langle M, M\rangle_t$ is a martingale and thus the expectations of the two terms are equal for all t. However, the main difficulty is that we only know that we have a local martingale.

Lemma 92

If M is a continuous local martingale such that $|M_t| \leq Z \in L^1$ for all t, then M is a uniformly integrable martingale.

Proof of lemma. By definition, there are stopping times τ_n such that $M_{t\wedge\tau_n}$ is a uniformly integrable martingale for all n. Thus, we have

$$M_{s\wedge\tau_n} = \mathbb{E}[M_{t\wedge\tau_n}|\mathcal{F}_s]$$

for all s, t, and since $M_{t\wedge\tau_n}$ is a continuous process, it converges to M_t as $n \to \infty$. By assumption, $|M_t|$ is dominated by Z for **all** t, which means $M_{t\wedge\tau_n}$ is also dominated by Z for all n and thus the collection is uniformly integrable. Thus $M_{t\wedge\tau_n}$ converges in L^1 to M_t as $n \to \infty$, meaning that by the dominated convergence theorem we have $M_s = \mathbb{E}[M_t|\mathcal{F}_s]$ and thus we do have a martingale. Finally, again the condition that $|M_t| \leq Z$ guarantees that we have a uniformly integrable process. □

Proof of Theorem 91. We can assume $M_0 = 0$ without loss of generality (because all of the quantities here only depend on increments of M). To show the forward direction, first note that $S = \sup_{t\geq 0} |M_t|$ is in L^2 by the Doob L^2 inequality. Define the stopping time

$$\sigma_n = \inf\{t : \langle M, M\rangle_t = n\};$$

then $Y_t = M_t^2 - \langle M, M\rangle_t$ is a local martingale, so the stopped version $Y_{t\wedge\sigma_n}$ is also a local martingale. But

$$|Y_{t\wedge\sigma_n}| = \left|M_{t\wedge\sigma_n}^2 - \langle M, M\rangle_{t\wedge\sigma_n}\right| \leq S^2 + n,$$

which is in L^1 because $S \in L^2$. Thus, by Lemma 92, the stopped process $Y_{t\wedge\sigma_n}$ is a uniformly integrable martingale (because it's dominated by $S^2 + n$), and thus $\mathbb{E}[M_{t\wedge\tau_n}^2] = \mathbb{E}[\langle M, M\rangle_{t\wedge\sigma_n}]$. By assumption, M is a true martingale and thus the left-hand side is uniformly bounded by $\mathbb{E}[S^2]$, while the right-hand side increases to $\mathbb{E}[\langle M, M\rangle_t]$ by the monotone convergence theorem. Thus taking $n \to \infty$ and then $t \to \infty$ yields the result.

On the other hand, assume the total quadratic variation is finite. Define the stopping times

$$\tau_n = \inf\{t : |M_t| = n\};$$

similarly $Y_t = M_t^2 - \langle M, M\rangle_t$ is a local martingale, so $Y_{t\wedge\tau_n}$ is also a local martingale. And now

$$|Y_{t\wedge\tau_n}| = \left|M_{t\wedge\tau_n}^2 - \langle M, M\rangle_{t\wedge\tau_n}\right| \leq n^2 + \langle M, M\rangle_\infty$$

is in L^1 by assumption, so $Y_{t\wedge\tau_n}$ is a uniformly integrable martingale again by Lemma 92. This means that

$$\mathbb{E}[M_{t\wedge\tau_n}^2] = \mathbb{E}[\langle M, M\rangle_{t\wedge\tau_n}] \leq \mathbb{E}[\langle M, M\rangle_\infty] < \infty,$$

so M is bounded in L^2 (by using Fatou's lemma as $n \to \infty$). It remains to show that we have a martingale – we already know that $\mathbb{E}[M_{t\wedge\tau_n}|\mathcal{F}_s] = M_{s\wedge\tau_n}$. For each t, the collection $M_{t\wedge\tau_n}$ is bounded in L^2, and we'll prove below in Proposition 93 that this implies uniform integrability. Therefore we can pass the limit through the integral as $n \to \infty$ to find that $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$, as desired. □

It remains only to prove the following:

Proposition 93

If $\{X_i\}$ are bounded in L^p for some $p > 1$, then $\{X_i\}$ is uniformly integrable.

Proof. Let $q = \frac{p}{p-1}$, so that $\frac{1}{p} + \frac{1}{q} = 1$. By Hölder's inequality,

$$\mathbb{E}\left[|X_i|; |X_i| \geq M\right] \leq ||X_i||_p ||1\{|X_i| \geq M\}||_q.$$

But the $||X_i||_p$ are uniformly bounded by some constant C, and the second term is just $\mathbb{P}(|X_i| \geq M)^{1/q}$. Thus this simplifies by Markov's inequality to

$$\mathbb{E}\left[|X_i|; |X_i| \geq M\right] \leq C\left(\frac{\mathbb{E}(|X_i|^p)}{M^p}\right)^{1/q} \leq \frac{C \cdot C^{p/q}}{M^{p/q}}.$$

This goes to 0 uniformly over i as $M \to \infty$, so we've verified the desired uniform integrability condition. □

We'll finish by wrapping up some minor points from chapter 4:

Definition 94

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space. If M, N are both continous local martingales, then their **quadratic covariation** is

$$\langle M, N\rangle_t = \frac{1}{2}\left(\langle M+N, M+N\rangle_t - \langle M, M\rangle_t - \langle N, N\rangle_t\right).$$

We say that M, N are **orthogonal** if their covariation is zero.

For example, if we take two independent Brownian motions B_1, B_2, then we can check that $\langle B_1, B_2\rangle = 0$.

Theorem 95 (Kunita-Watanabe)

Let M, N be continuous local martingales, and let H, K be **measurable processes**, meaning that the map $(t, \omega) \to H_t(\omega)$ (resp. $K_t(\omega)$) is measurable on $\mathcal{F} \otimes \mathcal{B}_{[0,\infty)}$. Then

$$\int_0^\infty |H_s K_s||d\langle M, N\rangle_s| \leq \left(\int_0^\infty H_s^2 d\langle M, M\rangle_s\right)^{1/2}\left(\int_0^\infty K_s^2 d\langle N, N\rangle_s\right)^{1/2}.$$

(In particular, it is sufficient for H and K to be adapted and continuous.) This is a kind of analogy to the Cauchy-Schwarz inequality. We won't go through the proof in full here, but the main idea is that the left-hand side can be approximated by

$$\sum_{i=1}^{|P_n|}\left|H_{t_{i-1}^{(n)}} K_{t_{i-1}^{(n)}}\left(M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}}\right)\left(N_{t_i^{(n)}} - N_{t_{i-1}^{(n)}}\right)\right|,$$

and by ordinary Cauchy-Schwarz this is bounded from above by

$$\left(\sum_i H_{t_{i-1}^{(n)}}^2\left(M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}}\right)^2\right)^{1/2}\left(\sum_j K_{t_{j-1}^{(n)}}^2\left(N_{t_j^{(n)}} - N_{t_{j-1}^{(n)}}\right)^2\right)^{1/2}.$$

But as $n \to \infty$, these two expressions converge to the left and right sides of the theorem statement, as desired.

Definition 96

A process X_t is a **continuous semimartingale** if it can be written as $X_t = M_t + A_t$, where M_t is a continuous local martingale and A_t is a finite variation process. If we have two such processes $X_t = M_t + A_t$ and $Y_t = M_t' + A_t'$ on the same space, then we define their **covariation** to be $\langle X, Y\rangle_t = \langle M, M'\rangle_t$.

And this definition should make sense, because the sum

$$\sum_{i=1}^{n}(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}})(Y_{t_i^{(n)}} - Y_{t_{i-1}^{(n)}})$$

will converge to $\langle M, M'\rangle_t$ as our subdivision P_n grows finer (in other words, the finite variation part doesn't contribute to the covariation or quadratic variation).

11 March 9, 2020

We have an exam on Thursday evening in 2-449 (**future note**: this did not end up happening). This may be the last thing we do in person before we get quarantined, but the class is not too big, so this is probably okay.

Today, we'll start discussing **stochastic integration** (from chapter 5). We're going to start seeing nice applications of the theory we've been developing, and we'll start with a bit of review. Recall that for a local martingale M_t, subtracting off the quadratic variation $\langle M, M\rangle_t$ (an increasing, finite variation process) from M^2 yields a local martingale $M_t^2 - \langle M, M\rangle_t$. We showed that if $M_0 \in L^2$, then M is bounded in L^2 if and only if $\mathbb{E}[\langle M, M\rangle_\infty]$ is finite; in this case, $M_t^2 - \langle M, M\rangle_t$ will in fact be a uniformly integrable (true) martingale because

$$\sup_t |M_t^2 - \langle M, M\rangle_t| \le \left(\sup_t |M_t|^2\right) + \langle M, M\rangle_\infty,$$

where the first term is integral by Doob's L^2 inequality and the second term is finite by assumption. (We wouldn't be asked to show Doob's L^2 inequality on the exam, but we would be expected to be able to reproduce the above argument.) At the end of last lecture, we also defined the **bracket**

$$\langle M, N\rangle = \frac{1}{2}\left(\langle M+N, M+N\rangle - \langle M, M\rangle - \langle N, N\rangle\right),$$

which in particular lets us calculate

$$M_tN_t - \langle M, N\rangle_t = \frac{1}{2}\left(M_t+N_t)^2 - \langle M+N, M+N\rangle_t\right) - \frac{1}{2}\left(M_t^2 - \langle M, M\rangle_t\right) - \frac{1}{2}\left(N_t^2 - \langle N, N\rangle_t\right).$$

Each term on the right is a local martingale, so the left side is also a local martingale, and if M, N are bounded in L^2, so is $M+N$. In such a situation, all three terms on the right side are uniformly integrable martingales, so $M_tN_t - \langle M, N\rangle_t$ is a uniformly integrable martingale as well, and in particular this means that $\boxed{\mathbb{E}[M_\infty N_\infty] = \mathbb{E}[\langle M, N\rangle_\infty]}$ if $M_0 = N_0 = 0$.

Our goal today is to take a semimartingale $X_t = A_t + M_t$ and a class of processes H_t and define the stochastic integral

$$(H\cdot X)_t = (H\cdot A)_t + (H\cdot M)_t = \int_0^t H_s dA_s + \int_0^t H_s dM_s.$$

We've already seen how to compute the first integral – A corresponds to a signed measure, so this is just the Lebesgue integral of H_s against that signed measure. The notation suggests that this should be a continuous-time version of the Doob transform – recall that in the discrete case, we defined

$$(H\cdot M)_n = \sum_{i=1}^n H_i(M_i - M_{i-1}).$$

This object was a martingale for appropriate H, and analogously the continuous-time version $\int_0^t H_s dM_s$ will turn out to be a **local** martingale.

The reason we talk about everything in L^2 is that this is an important case: we'll define the stochastic integral for martingales bounded in L^2 today, and extending to local martingales will be done next time.

Definition 97

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, let $\mathbb{H}^2$ denote the space of L^2-bounded martingales.

On this space $\mathbb{H}^2$, we can apply the boxed identity from above: since $\mathbb{E}[M_\infty N_\infty] = \mathbb{E}[\langle M, N\rangle_\infty]$, we define the **scalar product on $\mathbb{H}^2$** via

$$(M, N)_{\mathbb{H}^2} = \mathbb{E}[\langle M_\infty, N_\infty\rangle].$$

In particular, if $\mathbb{E}[\langle M_\infty, M_\infty\rangle] = 0$, then M must be identically zero, so this is indeed a norm. With this norm, $\mathbb{H}^2$ is a **Hilbert space** – to prove this, we need to check that it's complete, meaning that a Cauchy sequence (with respect to this norm) converges. We won't do this in much detail because it's similar to what we've already seen, but the point is that if we have $\lim_{m,n\to\infty} ||M^n - M^m||_{\mathbb{H}^2} \to 0$, then we can use the Doob L^2 inequality to show uniform convergence of the M^ns.

Recall that a process H_t is **progressive** if the function $F(\omega, t) = H_t(\omega)$ restricted to $\Omega \times [0, t]$ is measurable with respect to $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$. An equivalent characterization is to say that F is measurable with respect to $\mathcal{P}$, where $A \in \mathcal{P}$ if and only if the process $X_t(\omega) = 1\{(\omega, t) \in A\}$ is progressive. The point here is that there is a sigma-field $\mathcal{P}$ that is equivalent to being progressive, but it's not defined in a very useful way.

Definition 98

For an L^2-bounded martingale $M \in \mathbb{H}^2$, let $L^2(M)$ be the space of all progressive processes H such that

$$\mathbb{E}\left[\int_0^\infty H_t^2 d\langle M, M\rangle_t\right] < \infty.$$

Remember that $\langle M, M\rangle_t$ is of finite variation, so the inner integral is just a Lebesgue integral. Note that $L^2(M)$ is equivalent to a standard L^2 space $L^2(M) = L^2(\Omega \times [0,\infty), \mathcal{P}, \nu)$ where the measure ν is defined via

$$\nu(A) = \mathbb{E}\left[\int_0^\infty 1\{(\omega, t) \in A\} d\langle M, M\rangle_t\right] < \infty.$$

This means that $L^2(M)$ is also a Hilbert space with all of the corresponding nice properties, so we won't need to prove those again. In particular, we have an inner product

$$(H, K)_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_t K_t d\langle M, M\rangle_t\right].$$

From here, the idea is that for any $M \in \mathbb{H}^2$, we will define the stochastic integral with respect to M as an L^2 **isometry** $\mathcal{J}^M : L^2(M) \to \mathbb{H}^2$ which maps a martingale H via

$$H \mapsto \mathcal{J}^M(H) = H \cdot M = \left(\int_0^t H_s dM_s\right)_{t\geq 0}.$$

It's helpful to write out what the isometry is directly – if it needs to preserve scalar products, we must have

$$\begin{aligned}(H, K)_{L^2 M} &= \mathbb{E}\left[\int_0^\infty H_t K_t d\langle M, M\rangle_t\right] = (H \cdot M, K \cdot M)_{\mathbb{H}^2} = \mathbb{E}\left[(H \cdot M)_\infty (K \cdot M_\infty)\right] \\ &= \mathbb{E}\left[\left(\int_0^\infty H_t dM_t\right)\left(\int_0^\infty K_t dM_t\right)\right].\end{aligned}$$

In the particular case $H = K$, this is called the **Itô isometry**. To define such an isometry, we can first define it on a dense subspace of $L^2(M)$ and extend by continuity to the entire space:

Definition 99

Let $\mathcal{E}$ be the space of **elementary processes** of the form $H_t = \sum_{i=1}^p H_{(i)} 1\{t \in (t_i, t_{i+1}]\}$, where each $H_{(i)} \in \mathcal{F}_{t_i}$ is almost surely bounded by some constant.

These processes are pretty simple: we have some set of deterministic times, and on each time interval we put a (measurable) random variable. We have $\mathcal{E} \subseteq L^2(M)$ because of the boundedness condition, and note that $\mathcal{E}$ makes no reference to M at all.

Proposition 100

$\mathcal{E}$ is dense in $L^2(M)$ with respect to the scalar product on $L^2(M)$.

Proof. In an ordinary setting, we know that simple functions are dense in L^2, but this argument is a bit more complicated because we need to account for the sigma-algebra. It's enough to verify that if $K \in L^2(M)$ with $K \perp \mathcal{E}$, then $K = 0$. (Then the usual Hilbert space theory gives us denseness.) If K is orthogonal to $\mathcal{E}$, this means that for all $H \in \mathcal{E}$,

$$0 = (H, K)_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_t K_t d\langle M, M\rangle_t\right].$$

We will deduce that $X_t = \int_0^t K_s d\langle M, M\rangle_s$ is identically zero. First, we claim that X is well-defined as a finite-variation process – for this, we just need absolute integrability. Specifically, by Cauchy-Schwarz we have

$$\mathbb{E}\left[\int_0^\infty |K_s| d\langle M, M\rangle_s\right] \le \sqrt{\mathbb{E}\left[\int_0^\infty |K_t|^2 d\langle M, M\rangle_t\right] \mathbb{E}\left[1 \cdot d\langle M, M\rangle_t\right]} < \infty,$$

where we know the first term in the square root is finite because of the inner product on $L^2(M)$, and similarly the second term is finite because M is in $\mathbb{H}^2$. The integral on the left-hand side upper bounds $|X_t|$, so this implies that $X_t \in L^1$ for all t. Now define

$$H_r(\omega) = F_{(s)} 1\{r \in (s, t]\},$$

where $F_{(s)}$ is some bounded random variable that is measurable with respect to $\mathcal{F}_s$. We've assumed that $K \perp \mathcal{E}$, so

$$0 = (H, K)_{L^2(M)} = \mathbb{E}\left[F_{(s)} \int_s^t K_r d\langle M, M\rangle_r\right] = \mathbb{E}[F_{(s)}(X_t - X_s)]$$

by definition of X_t. Since we already proved that $X_t \in L^1$, this last calculation tells us that X is a martingale. But it is also a finite variation process, so we must have $X = 0$. This implies that $K = 0$ **as an element of** $L^2(M)$ (meaning that it is zero except on a set of measure zero with respect to $L^2(M)$), completing the proof. □

Theorem 101

Let $M \in \mathbb{H}^2$ be an L^2-bounded martingale. For any $H \in \mathcal{E}$ of the form $H_t = \sum_{i=1}^p H_{(i)} 1\{t \in (t_i, t_{i+1}]\}$, define

$$\mathcal{J}^M(H)_t = \sum_{i=1}^p H_{(i)}(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}).$$

Then $\mathcal{J}^M$ defines an isometry from $\mathcal{E}$ (with the $L^2(M)$ scalar product) into $\mathbb{H}^2$, so it extends to an isometry from $L^2(M)$ into $\mathbb{H}^2$.

(The idea here is that integrating $1\, dM_t$ should give us back the original martingale, so integrating elementary processes gives us increments of the martingale.)

Proof. We know that $\mathcal{J}^M(H)$ is in $\mathbb{H}^2$ (because $H_{(i)}$ are bounded and the stopped Ms are in $\mathbb{H}^2$), and the quadratic variation process can be computed as

$$\langle \mathcal{J}^M(H), \mathcal{J}^M(H)\rangle_t = \sum_{i=1}^p H_{(i)}^2 \left(\langle M, M\rangle_{t\wedge t_{i+1}} - \langle M, M\rangle_{t\wedge t_i}\right)$$

(this is only easy because H takes on a very simple form.) But this sum is also equal to $\int_0^t H_s^2 d\langle M, M\rangle_s$ by the definition of the Lebesgue integral for this simple-like function. So now we can check the isometry property on $\mathcal{E}$: by the definition of the $\mathbb{H}^2$ scalar product and our observation above, we have

$$\begin{aligned}(\mathcal{J}^M(H), \mathcal{J}^M(H))_{\mathbb{H}^2} &= \mathbb{E}\left[\sum_i H_{(i)}^2 \left(\langle M, M\rangle_{t_{i+1}} - \langle M, M\rangle_{t_i}\right)\right] \\ &= \mathbb{E}\left[\int_0^\infty H_s^2 d\langle M, M\rangle_s\right] = (H, H)_{L^2(M)}.\end{aligned}$$

So $\mathcal{J}^M$ is an isometry on $\mathcal{E}$, and we can extend this to $L^2(M)$ by continuity. □

Remark 102. *We should have checked at some point that $\mathcal{J}^M(H)$ doesn't depend on the representation of H, since H is written down in an explicit form and there might be different ways to do so. But the idea is that we have $||\mathcal{J}^M(H - H')||_{\mathbb{H}^2} = ||H - H'||_{L^2(M)}$, and we say that H and H' are the same if the right-hand side is zero.*

We'll spend the rest of the time on a useful result for Itô's formula, which says in words that "the stochastic integral commutes with the bracket."

Proposition 103
Let $M \in \mathbb{H}^2$ and let $H \in L^2(M)$. Then $H \cdot M$ is the unique element of $\mathbb{H}^2$ such that $\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle$ for all $N \in \mathbb{H}^2$.

In particular, applying this result twice tells us that

$$\langle H \cdot M, K \cdot N\rangle = HK \cdot \langle M, N\rangle$$

as long as everything is well-defined (meaning $M, N \in \mathbb{H}^2$, $H \in L^2(M)$, $K \in L^2(N)$); more explicitly, this is stating that

$$\left\langle \int_0^t H_s dM_s, \int_0^t K_s dN_s \right\rangle = \int_0^t H_s K_s d\langle M, N\rangle_s.$$

Start of proof. Consider first an elementary process $H \in \mathcal{E}$. In this case, we know that

$$(H \cdot M)_t = \sum_i H_{(i)}(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}),$$

where each term is a martingale. The covariation of this with N is just

$$\langle H \cdot M, N\rangle_t = \sum_i H_{(i)} \left(\langle M, N\rangle_{t\wedge t_{i+1}} - \langle M, N\rangle_{t\wedge t_i}\right) = \int_0^t H_s d\langle M, N\rangle_s = (H \cdot \langle M, N\rangle)_\infty,$$

so the identity holds in that case. For a general H, we can approximate H by elementary processes $H^n \in \mathcal{E}$, and we wish to show that

$$\langle H \cdot M, N\rangle = \lim_{n\to\infty} \langle H^n \cdot M, N\rangle = \lim_{n\to\infty} (H^n \cdot \langle M, N\rangle)_\infty = H \cdot \langle M, N\rangle.$$

It just remains to justify the limits on the left and right. We proved the Kunita-Watanabe inequality last time, which says that

$$\left|\int H_s K_s d\langle M, N\rangle_s\right| \le \left|\int H_s^2 d\langle M, M\rangle_s\right|^{1/2} \left|\int K_s^2 d\langle N, N\rangle_s\right|^{1/2}.$$

Thus for any $X \in \mathbb{H}^2$, we have by Cauchy-Schwarz and Kunita-Watanabe (setting M, N to both be M and H, K to be X, N) that

$$\mathbb{E}\left[\langle X, N\rangle\right] \le \mathbb{E}\left[\langle X, N\rangle\langle X, N\rangle\right]^{1/2} \le \mathbb{E}\left[\langle X, X\rangle\right]^{1/2} \mathbb{E}\left[\langle N, N\rangle\right]^{1/2} = ||X||_{\mathbb{H}^2} ||N||_{\mathbb{H}^2},$$

meaning that taking the quadratic covariation with N is continuous with respect to the $\mathbb{H}^2$ norm. Thus, the left limit is justified by taking $X = (H - H^n) \cdot M$ and noting that $H^n \cdot M$ converges to $H \cdot M$ in the $\mathbb{H}^2$ metric.

For the right limit, apply Kunita-Watanabe and Cauchy-Schwarz again with $H = X \in L^2(M)$ and $K = 1$ to find

$$\mathbb{E}\left[(X \cdot \langle M, N\rangle)_\infty\right] \leq ||X||_{L^2(M)} ||N||_{\mathbb{H}^2}.$$

The proof is therefore completed by a similar continuity argument by taking $X = H - H^n$, which we've shown converges to zero in $L^2(M)$. □

In summary, all of the properties are easy to prove when H is an elementary processes, and then we just need to take appropriate limits.

12 March 11, 2020

This will be the last class we have in person – we'll continue after spring break on Zoom. If we don't have efficient high-speed internet, we should let Professor Sun know by email. We started talking about stochastic integration last time: if $\mathbb{H}^2$ denotes the (Hilbert) space of L^2-bounded martingales, we defined a scalar product on $\mathbb{H}^2$ via $(M, N)_{\mathbb{H}^2} = \mathbb{E}[M_\infty N_\infty] = \mathbb{E}[\langle M, N\rangle_\infty]$. We also defined the Hilbert spaces

$$L^2(M) = \left\{\text{progressive } H : ||H||_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_t^2 d\langle M, M\rangle_t\right] < \infty\right\}.$$

Then for any $M \in \mathbb{H}^2$, we can define the stochastic integral, which is an isometry $\mathcal{J}^M$ from $L^2(M) \to \mathbb{H}^2$ sending H to $H \cdot M$. One property of $H \cdot M$ is that it is the unique element in $\mathbb{H}^2$ such that for all $N \in \mathbb{H}^2$, we have

$$\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle. \tag{1}$$

From now on, we'll use the notation $\langle M\rangle = \langle M, M\rangle$. Our next goal is to **extend stochastic integration to local martingales** (and therefore to semimartingales). We'll start with a few remarks: if M is a local martingale and τ is a stopping time, then the stopped process $M_t^\tau = M_{t\wedge\tau}$ is also a local martingale (this follows straightforwardly from the definitions). In particular, $M_t^2 - \langle M\rangle_t$ is a local martingale, $(M^\tau)_t^2 - \langle M\rangle_{t\wedge\tau}$ is also a local martingale. But there is a unique process that yields a local martingale when we subtract off from $(M^\tau)^2$, namely $\langle M^\tau\rangle$. Thus

$$\langle M^\tau\rangle = \langle M\rangle_{t\wedge\tau} \implies \langle M^\tau\rangle = \langle M\rangle^\tau.$$

Similarly, if M, N are both local martingales, then we also have

$$\langle M^\tau, N^\tau\rangle = \langle M, N^\tau\rangle = \langle M, N\rangle^\tau. \tag{2}$$

This last fact is a bit harder to prove, but we can read it on our own.

Lemma 104

If $M \in \mathbb{H}^2$ and $H \in L^2(M)$, then for all stopping times τ, we have

$$(H \cdot M)^\tau = (H \cdot 1_{[0,\tau]}) \cdot M = H \cdot (M^\tau).$$

Proof. Applying Eq. (2) and then Eq. (1), note that for any $N \in \mathbb{H}^2$, we have

$$\langle (H \cdot M)^\tau, N\rangle_t = \langle H \cdot M, N\rangle_{t\wedge\tau} = (H \cdot \langle M, N\rangle)_{t\wedge\tau}.$$

But the right-hand side is just the integral of H against a finite-variation process, stopped at some time τ, so we can also write it as $((H \cdot 1_{[0,\tau]}) \cdot \langle M, N\rangle)_t$. Thus, $(H \cdot M)^\tau$ satisfies the characterizing property Eq. (1) for $(H \cdot 1_{[0,\tau]}) \cdot M$, so we've shown the first equality.

For the second equality, we similarly have that

$$\langle H \cdot (M^\tau), N\rangle = H \cdot \langle M^\tau, N\rangle = H \cdot (\langle M, N\rangle^\tau).$$

Again, this is the integral of H against a finite-variation process, so we can write it as $((H \cdot 1_{[0,\tau]}) \cdot \langle M, N\rangle)_t$. But again this means we've shown the characterizing property, so the result follows. □

To extend our stochastic integral definition, we'll need to be a bit more general:

Definition 105

For a local martingale M, define the space $\boldsymbol{L^2_{\text{loc}}(M)}$ via

$$L^2_{\text{loc}}(M) = \left\{\text{progressive } H : \int_0^t H_s^2 d\langle M\rangle_s < \infty \text{ for all finite } t \text{ a.s.}\right\}.$$

Recall that to be in $L^2(M)$, we integrate from 0 to ∞ and require the expectation to be finite, but here we only need the integral to be **almost surely** finite. The main goal of today is to prove the following result:

Theorem 106

Let M be a local martingale and let $H \in L^2_{\text{loc}}(M)$. Then there exists a unique local martingale $H \cdot M$ with initial value 0 such that for all local martingales N, we have

$$\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle.$$

Once we prove this, it will make sense to write $(H \cdot M)_t = \left(\int_0^t H_s dM_s\right)_t$.

Proof. Let $M_0 = 0$ without loss of generality (since nothing depends on the initial value). We'll start by talking about how to construct this process: we want to extend our definition from the L^2-bounded case, so we define the stopping times

$$\tau_n = \inf\left\{t \geq 0 : \int_0^t (1 + H_s^2) d\langle M\rangle_s \geq n\right\}.$$

We can check that τ_n goes to ∞ almost surely: we assumed H^2 is in $L^2_{\text{loc}}(M)$ and $\langle M\rangle$ is a finite variation process, so neither term in the integrand gets large too fast. Furthermore, the total quadratic variation of the stopped process M^{τ_n} will be (almost surely) at most n, so $\mathbb{E}[\langle M\rangle_\infty]$ is finite and therefore $M^{\tau_n} \in \mathbb{H}^2$. Similarly, we have $H \in L^2(M^{\tau_n})$ because by time τ_n, the integral $\int H_s^2 d\langle M\rangle_s$ is bounded by n.

This implies immediately that for each n, $X^n = H \cdot (M^{\tau_n})$ is well-defined and is an element of $\mathbb{H}^2$. We wish to show that for any $m > n$, X^m and X^n are **consistently** defined, but we have

$$(X^m)^{\tau_n} = (H \cdot M^{\tau_m})^{\tau_n} = H \cdot ((M^{\tau_m})^{\tau_n}) = H \cdot M^{\tau_n}$$

by definition, then by Lemma 104, and then because $\tau_n < \tau_m$. This right-hand side is X^n, so do indeed have consistency and in particular there is some X such that $X^{\tau_n} = X^n$ for all n. Furthermore, each stopped X^n is an L^2-bounded martingale (meaning it is uniformly integrable), so X is a local martingale, as desired.

Next, we need to check that this $X = H \cdot M$ satisfies the desired property $\langle H \cdot M, N\rangle = H \cdot \langle M, N\rangle$. Assume without loss of generality that $N_0 = 0$. We'll basically use the fact that the desired property holds at stopping times – define

$\sigma_n = \inf\{t \geq 0 : |N_t| = n\}$ and let $\gamma_n = \sigma_n \wedge \tau_n$. Notice that N^{σ_n} is a local martingale bounded by n, so it is a true martingale in $\mathbb{H}^2$. Then we have

$$\begin{aligned}
\langle H \cdot M, N\rangle^{\gamma_n} &= \langle (H \cdot M)^{\tau_n}, N^{\sigma_n}\rangle \text{ (by Eq. (2))} \\
&= \langle H \cdot (M^{\tau_n}), N^{\sigma_n}\rangle \text{ (by Lemma 104)} \\
&= H \cdot \langle M^{\tau_n}, N^{\sigma_n}\rangle \text{ (by Eq. (1))} \\
&= H \cdot (\langle M, N\rangle)^{\gamma_n} \\
&= (H \cdot \langle M, N\rangle)^{\gamma_n},
\end{aligned}$$

where the last step comes from the definition of integrating against a finite variation process. But now taking $n \to \infty$ yields the desired property for $X = H \cdot M$, since both τ_n and σ_n go to infinity as $n \to \infty$.

Finally, we show that X is indeed unique: if there is some $\tilde{X}$ such that $\langle \tilde{X}, N\rangle = H \cdot \langle M, N\rangle = \langle X, N\rangle$, then $\langle X - \tilde{X}, N\rangle = 0$ for all local martingales N. But take $N = \langle X - \tilde{X}\rangle$ to show that we have a local martingale $X - \tilde{X}$ with quadratic variation 0, which can only happen if $X - \tilde{X} = 0$. □

Notice that Lemma 104 now also extends to the case where M is a **local** martingale and H is in $L^2_{\text{loc}}(M)$, because we've now proved the analogous arguments for local martingales.

We'll next discuss something useful for our next homework. Recall that if $M \in \mathbb{H}^2$ (the L^2-bounded case) and $H \in L^2(M)$, we have two nice properties: first of all, $\mathbb{E}\left[\int_0^t H_s dM_s\right] = 0$ (since $H \cdot M = \int_0^t H_s dM_s$ is also a martingale). We can also calculate the second moment in the following way: we have

$$\mathbb{E}\left[\int_0^t H_s d\langle M\rangle_s\right] = ||H \cdot 1_{[0,t]}||^2_{L^2(M)} = ||(H \cdot 1_{[0,t]}) \cdot M||^2_{\mathbb{H}^2} = ||(H \cdot M)^t||^2_{\mathbb{H}^2},$$

first by definition, then by the Itô isometry, and finally by Lemma 104. And since the $\mathbb{H}^2$ norm is just the expectation of the martingale's eventual value, this is exactly $\mathbb{E}[((H \cdot M)_t)^2] = \mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right]$ (the second moment we're after).

It's important to note that **these equations don't necessarily hold if we're in the general situation where M is a local martingale**. However, they do hold under restricted conditions, for example if $\mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right]$ is finite. This is because $X = (H \cdot M)^t$ has total quadratic variation $\langle X\rangle_\infty = \mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right]$ (by Proposition 103), meaning that if we assume the right-hand side is finite, then $X \in \mathbb{H}^2$. In general, if we're interested in the second moment of $\int_0^t H_s dM_s$, we have an **upper bound**

$$\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] \leq \mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right], \tag{3}$$

where if the right hand side is finite, then we have equality. (And otherwise this is a vacuous inequality anyway.)

With this, we can now make the stochastic integral definition for semimartingales: if we have a process $X = A + M$, we want to define $H \cdot X = H \cdot A + H \cdot M$, and we want to do this for some reasonably large class of H-processes on which both stochastic integrals will exist:

Definition 107

A progressive process H is **locally bounded** if $\sup_{s \leq t} |H_s| < \infty$ for all finite t.

For any such process, we know that for any t,

$$\int_0^t |H_s||dA_s| \leq \left(\sup_{s \leq t} |H_s|\right) \cdot \int_0^t |dA_s| < \infty,$$

so $(H \cdot A)_t$ is well-defined. Similar reasoning shows that $\int_0^t H_s^2 d\langle M\rangle_s < \infty$, so $H \in L^2_{\text{loc}}(M)$ for any local martingale M and thus $H \cdot M$ will be well-defined as well. This stochastic integral $H \cdot X$ is now also a semimartingale, and it's already written in its canonical decomposition ($H \cdot A$ is the finite variation part, and $H \cdot M$ is the local martingale part).

We'll use the remaining time to prove a useful convergence property, which is a dominated convergence type result for stochastic integrals.

Proposition 108

Let $X = A + M$ be a semimartingale, and suppose that H^n, H, K are locally bounded with $K \geq 0$. Fix some $t > 0$, and suppose that for all $s \leq t$, we have $H_s^n \to H_s$ as $n \to \infty$ and $|H_s^n| \leq K_s$ for all n. If $\int_0^t K_s |dA_s|$ and $\int_0^t K_s^2 d\langle M\rangle_s$ are both finite almost surely, then $(H^n \cdot X)_t$ converges in probability to $(H \cdot X)_t$.

Proof. By the usual dominated convergence theorem (since the finite-variation integral is just a Lebesgue integral and everything's dominated by K), we already have

$$\int_0^t H_s^n dA_s \overset{\text{a.s.}}{\to} \int_0^t H_s dA_s$$

To show the other part, let $\tau_k = t \wedge \inf\left\{r \geq 0 : \int_0^r K_s^2 d\langle M\rangle_s \geq k\right\}$. Eq. (3) implies that

$$\mathbb{E}\left[\left(\int_0^{\tau_k} (H_s^n - H_s) dM_s\right)^2\right] \leq \mathbb{E}\left[\int_0^{\tau_k} (H_s^n - H_s)^2 d\langle M\rangle_s\right],$$

and we'll show that this right-hand side goes to zero by two applications of the dominated convergence theorem. First of all, inside the expectation, the Hs are both dominated by K, so the integrand is dominated by $4K^2$. By definition of the stopping time, the integral of this up to τ_k will still be finite, so using the dominated convergence theorem when integrating against $d\langle M\rangle_s$ shows that the quantity inside the expectation goes to zero almost surely. Then we can use dominated convergence theorem again to show that the whole expectation converges to zero, since the integral is uniformly dominated by $4k^2$ again by the definition of τ_k.

To prove the statement we're after, now notice that

$$\mathbb{P}\left(\left|\int_0^t (H_s^n - H_s) dM_s\right| \geq \varepsilon\right) \leq \mathbb{P}\left(\left|\int_0^{\tau_k} (H_s^n - H_s) dM_s\right| \geq \varepsilon\right) + \mathbb{P}(\tau_k \neq t).$$

The second term goes to zero as $k \to \infty$ (since $\int_0^r K_s^2 d\langle M\rangle_s$ is finite almost surely by assumption), and then the first term goes to infinity as $n \to \infty$ because the integral inside converges to 0 in L^2. Thus the convergence in probability is proven. □

Corollary 109

Let X be a semimartingale, and let H be an adapted, continuous process (in particular, this means it will be locally bounded and also progressive). Then for a sequence of subdivisions $P_n = (t_i^{(n)})$ of $[0, t]$ with mesh going to 0, we have

$$\sum_{i=1}^{|P_n|} H_{t_{i-1}^{(n)}} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}\right) \overset{p}{\to} \int_0^t H_s dX_s.$$

If X were a finite-variation process, it wouldn't matter whether we take $H_{t_{i-1}^{(n)}}$ or $H_{t_i^{(n)}}$ in the sum on the left-hand side. **But it matters here in the local martingale case**, and we can check that directly. For example, if $H = X$, then

this corollary tells us that

$$\sum_{i=1}^{|P_n|} X_{t_{i-1}^{(n)}} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) \xrightarrow{p} \int_0^t X_s dX_s,$$

but if we replace $(i-1)$ with i, we will instead have

$$\sum_{i=1}^{|P_n|} X_{t_i^{(n)}} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) = \sum_{i=1}^{|P_n|} X_{t_{i-1}^{(n)}} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) + \sum_{i=1}^{|P_n|} \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right)^2 \xrightarrow{p} \int_0^t X_s dX_s + \langle X \rangle_t.$$

However, we can add these two statements together and find that

$$\sum_{i=1}^{|P_n|} \left(X_{t_i^{(n)}} + X_{t_{i-1}^{(n)}} \right) \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) = \sum_{i=1}^{|P_n|} \left(X^2_{t_i^{(n)}} - X^2_{t_{i-1}^{(n)}} \right) = X_t^2 - X_0^2,$$

so $X_t^2 - X_0^2 = 2\int_0^t X_s dX_s + \langle X \rangle_t$. This is actually a special case of **Itô's formula**, which tells us (in a special case) that for sufficiently smooth f, we have

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s.$$

We can understand this statement in full now: if f is twice continuously differentiable, then f' is a continuous, adapted process, so the first term on the right hand side is a semimartingale (it's an integral of an adapted process against a semimartingale), and the second term is a finite variation term.

We'll end here for now – all of the cool applications of stochastic calculus will unfortunately have to be done via the internet. Because we have two canceled lectures, we'll be asked (if possible) to read the section about Itô's formula. As a reminder, the midterm tomorrow is canceled, and class resumes again after spring break. And from 3:30-5:00, Professor Sun will be in room 2-175 for general advising hours.

13 March 30, 2020

Our first midterm will now be on Thursday during the usual timeslot – if this doesn't work for us, we should email Professor Sun. (The system will put us in a virtual waiting room, and we will be "admitted" into the internet office.) There will be no lecture on Wednesday due to the exam, so we'll have office hours during the lecture timeslot instead.

First of all, for a quick review of stochastic integration, we should scroll down to the bottom of the course webpage and read the "Review of stochastic integration" section. The main points are that we can define stochastic integration as an isometry in the L^2-bounded case by approximating with elementary processes, and then in general we can use the characterization $\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$ if M and N are local martingales. We'll start today with Itô's formula, and we're going to state it in multiple dimensions here:

Theorem 110 (Itô's formula)

Let $X_t = (X_t^1, \cdots, X_t^p)$ be a process that evolves in $\mathbb{R}^p$ such that each X_t^i is a continuous semimartingale. Then for a twice continuously differentiable function $F : \mathbb{R}^p \to \mathbb{R}$,

$$F(X_t) - F(X_0) = \sum_{i=1}^p \int_0^t \partial_i F(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^p \int_0^t \partial_{ij} F(X_s) d\langle X^i, X^j \rangle_s$$

is also a continuous semimartingale with the above decomposition, meaning the first term is a local martingale and the second term is a finite variation process.

Remember that the first term is integration with respect to a semimartingale, and the second is essentially a Lebesgue integral, so we do understand all of the individual components of this statement. We won't prove this during class – we should read the proof in Le Gall on our own.

Remark 111. *This formula holds even if F is defined only on an open set $U \subseteq \mathbb{R}^p$, as long as $t \leq \tau_\varepsilon = \inf\{t : dist(X_t, U^t) \leq \varepsilon\}$ for some $\varepsilon > 0$. In other words, we need to stay at least ε away from the boundary so that we can define a function $\tilde{F}$ on all of $\mathbb{R}^p$ consistent with F.*

Proposition 112
If M is a local martingale, then $\mathcal{E}(\lambda M)_t = \exp\left(\lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t\right)$ is also a local martingale for all $\lambda \in \mathbb{C}$.

Proof. Applying Itô's formula with $X_t = \lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t$ (so that $\mathcal{E}(\lambda M)_t = \exp(X_t)$) and $F(x) = e^x$, we have

$$d\mathcal{E}_t = \exp(X_t)dX_t + \frac{1}{2}\exp(X_t)d\langle X\rangle_t,$$

because the derivatives of the exponential are just the exponential itself. Substituting in the values of X_t and $\langle X\rangle_t$ (noting that the quadratic variation comes only from the local martingale part), we have

$$d\mathcal{E}_t = \exp(X_t)\left[\lambda dM_t - \frac{\lambda^2}{2}d\langle M\rangle_t + \frac{1}{2}\lambda^2\langle M\rangle_t\right],$$

because X_t gets its quadratic variation only from the local martingale part (λM_t). But the last two terms cancel, so we only have a local martingale term and thus $\mathcal{E}(\lambda M)_t$ is indeed a local martingale. □

Theorem 113 (Lévy's characterization of Brownian motion)
If X is a continuous adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ taking values in $\mathbb{R}^d$, then the following are equivalent:

- X is a Brownian motion with respect to $\mathcal{F}_t$,
- The X^is are continuous local martingales with $\langle X^i, X^j\rangle_t = t \cdot 1\{i = j\}$.

Proof. The forward direction is easy, since Brownian motion is a continuous local martingale and a Brownian motion in $\mathbb{R}^d$ is just d independent Brownian motions. For the reverse direction, we'll want to use the Fourier transform, so it's natural to consider the exponential martingale $\mathcal{E}(i\theta \cdot X)_t$ for $\theta \in \mathbb{R}^d$. By assumption, the quadratic variation is $\langle \theta \cdot X\rangle_t = |\theta|^2 t$, so

$$\mathcal{E}(i\theta \cdot X)_t = \exp\left(i\theta \cdot X_t - \frac{1}{2}i^2|\theta|^2 t\right),$$

Now $\exp(i\theta \cdot X_t)$ is bounded because it's only varying on the unit circle, and the remaining part $\exp\left(\frac{1}{2}|\theta|^2 t\right)$ is bounded **on any finite interval**. This means $\mathcal{E}(i\theta \cdot X)_t$ is actually a uniformly integrable martingale up to any finite time, so we can apply the optional stopping theorem $\mathbb{E}\left[\mathcal{E}(i\theta \cdot X_t) \mid \mathcal{F}_s\right] = \mathcal{E}(i\theta \cdot X_s)$ (for all $0 \leq s \leq t < \infty$). Plugging the definition in from above and rearranging terms yields

$$\mathbb{E}\left[\exp\left(i\theta \cdot (X_t - X_s)\right) \mid \mathcal{F}_s\right] = \exp\left(-\frac{|\theta|^2(t-s)}{2}\right).$$

But now the left hand side is the characteristic function of $X_t - X_s$ given $\mathcal{F}_s$, while the right hand side is the characteristic function of $N(0, (t-s)I_d)$. So conditioned on $\mathcal{F}_s$, $X_t - X_s$ has the correct normal distribution, which means that X has the same finite dimensional distributions as the Brownian motion in $\mathbb{R}^d$. Since we assumed that our sample paths are continuous, this indeed means X is a Brownian motion, as desired. □

We've mentioned our next result before in previous classes:

Theorem 114 (Dambis-Dubins-Schwarz)
Let M be a continuous local martingale, and suppose that $\langle M\rangle_\infty = \infty$ almost surely for simplicity. Then there exists a Brownian motion B such that almost surely, we have $M_t = B_{\langle M\rangle_t}$ for all $t \geq 0$.

If the quadratic variation is not infinite, then we just end up with a Brownian motion parameterized up to some time.

Proof. We're going to construct a B first and then show that it satisfies Lévy's characterization. Without loss of generality, we can assume that $M_0 = 0$. Write $A_t = \langle M\rangle_t$; since A is a nondecreasing process, we can define an "inverse" by defining $\tau_r = \inf\{t \geq 0 : A_t \geq r\}$. We now define $B_r = M_{\tau_r}$ – we will check that this is a Brownian motion with respect to the filtration $\mathcal{G}_r = \mathcal{F}_{\tau_r}$, and we just need to check the Lévy characterization.

We know that M^{τ_r} accumulates a total variation of r by definition, meaning it's a uniformly integrable martingale, which implies that $(M^{\tau_r})^2 - \langle M\rangle^{\tau_r}$ is also a uniformly integrable martingale. So by the optional stopping theorem, we have for all $r \geq s$ that

$$B_s = M_{\tau_s} = \mathbb{E}\left[M_{\tau_r} \mid \mathcal{F}_{\tau_s}\right] = \mathbb{E}[B_r \mid \mathcal{G}_s].$$

Also, $B_s^2 - s = (M_{\tau_s})^2 - \langle M\rangle_{\tau_s}$ is a uniformly integrable martingale, so we can again apply the optional stopping theorem to say that for all $r \geq s$,

$$B_s^2 - s = \mathbb{E}\left[(M_{\tau_r})^2 - \langle M\rangle_{\tau_r} \mid \mathcal{F}_{\tau_s}\right] = \mathbb{E}\left[B_r^2 - r \mid \mathcal{G}_s\right].$$

So B is a local martingale with quadratic variation $\langle B\rangle_s = s$, and now we just need to make sure B is continuous to apply Lévy's characterization. If A were **strictly** increasing, τ would be a continuous function – then $B_r = M_{\tau_r}$ is a composition of two continuous functions, so it must be continuous. So the only problem is that A_t may be constant on some interval $[\tau_r, \tau_{r+}]$, where $\tau_{r+} = \inf\{t : A_t > r\}$. If this interval is nontrivial, then $\tau_r < \tau_{r+}$, and we need to check if B is continuous at r. But if A is flat on this interval, M **must also be constant**, because any local martingale with constant quadratic variation does not evolve with time (this is a lemma that we need to check, but we can read the book for details). Thus B is continuous and thus it is indeed a Brownian motion. □

Example 115
Recall the example from our first lecture, where we considered a holomorphic function $f : \mathbb{C} \to \mathbb{C}$. (meaning that if we write $f = u + iv$, then $f' = u_x + iv_x = v_y - iu_y$ by the Cauchy-Riemann equations).

We can now check that f applied to a Brownian motion B_t yields another Brownian motion: specifically, we have

$$f(B_t) = \beta_{A_t}, \quad A_t = \int_0^t |f'(B_s)|^2 ds,$$

where β is a Brownian motino.

Remark 116. *We're going to skip two topics in the book for now, which are the Burkholder–Davis–Gundy inequality and the stochastic integral representation for martingales.*

For the rest of today, we'll discuss **Girsanov's theorem**, which will give us some practice working with all of the objects we've encountered so far. We did a lot of exercises involving change of measure, and we'll start with **exponential change of measure** here.

Example 117

Suppose we have a random variable X with moment generating function $m(\theta) = \mathbb{E}[e^{\theta X}] < \infty$. Define a new "tilted" probability measure with the Radon–Nikodym derivative

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{e^{\theta X}}{m(\theta)},$$

meaning that the probability of an event A is $\mathbb{P}_\theta(A) = \mathbb{E}\left[1_A \frac{d\mathbb{P}_\theta}{d\mathbb{P}}\right]$.

The reason for the normalizing constant $m(\theta)$ is that we want $\mathbb{P}_\theta(\Omega) = 1$. In particular, under this new measure, notice that

$$\mathbb{E}_\theta[X] = \mathbb{E}\left[X\frac{d\mathbb{P}_\theta}{d\mathbb{P}}\right] = \frac{\mathbb{E}[Xe^{\theta X}]}{\mathbb{E}[e^{\theta X}]} = \frac{m'(\theta)}{m(\theta)},$$

so shifting the distribution by $\mathbb{P}_\theta$ changes the mean as well.

Example 118

As an explicit example, let's exponentially tilt the standard normal $X \sim N(0,1)$ with $\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{e^{\theta X}}{\exp(\theta^2/2)}$.

Then the density of X under $\mathbb{P}_\theta$ is the product of the dentiy of X under $\mathbb{P}$ and the Radon-Nikodym derivative, which factors nicely as

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\cdot\exp\left(\theta x - \frac{\theta^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\theta)^2\right).$$

So X is now distributed as $N(\theta, 1)$ – in other words, this particular exponential tilt just moves the center of our distribution.

Example 119

Next, suppose $\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}\right)$ is bivariate normal and $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\theta Y - \frac{\theta^2}{2}\right)$.

We can repeat the calculation above, but another way to work through this is to use the characteristic functions – it suffices to calculate $\mathbb{E}_\mathbb{Q}\left[e^{itX}\right] = \mathbb{E}\left[e^{itX}e^{\theta Y}e^{-\theta^2/2}\right]$ for all real numbers t. This can be done by replacing Y with $aX + \sqrt{1-a^2}W$, where W is a standard normal independent to X (consistent with the covariance between X and Y). Plugging this in, we find that

$$\mathbb{E}_\mathbb{Q}\left[e^{itX}\right] = \exp\left(\frac{1}{2}(it+\theta a)^2 + \frac{\theta^2 t^2(1-a)^2}{2} - \frac{\theta^2}{2}\right) = \exp\left[-\frac{t^2}{2} + it\theta a\right],$$

where the first term comes from the characteristic function of X and the second term comes from the characteristic function of W. This means that X will now be distributed as $N(\theta a, 1)$ under $\mathbb{Q}$.

In both of those cases, we had a finite number of random variables, and now we'll think about a more general case:

Example 120

Consider a sequence of iid random variables $X_1, \cdots, X_n$ under $\mathbb{P}$, and define the change of measure $\frac{d\mathbb{Q}}{d\mathbb{P}} = \prod_{i=1}^n \frac{e^{\theta X_i}}{\mathbb{E}[e^{\theta X}]}$ by changing the measure for each X_i. Then the X_i are iid under $\mathbb{Q}$ as well, and now the process $S_k = \sum_{i=1}^k X_i$ is a random walk under both $\mathbb{P}$ and $\mathbb{Q}$ (just with different jump distributions).

Notice that for any finite n, we can also write this change of measure as

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\theta S_n - n \log m(\theta)\right),$$

but if we take $n \to \infty$, $\mathbb{Q}$ may not be absolutely continuous with respect to $\mathbb{P}$. Motivated by our earlier examples, consider a sequence of independent random variables distributed under $\mathbb{P}$ as

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a_i \\ a_i & 1 \end{bmatrix}\right).$$

Then we can consider the two processes $M_k = \sum_{i=1}^k \sigma_i X_i$ and $L_k = \sum_{i=1}^k \tau_i Y_i$, where the coefficients $\sigma_i, \tau_i \in \mathcal{F}_{i-1}$ can also be random. (So at each stage, we add a Gaussian increment times some random number which is measurable with respect to the past.) Note that M and L are both martingales, and now we can define the "exponential martingale for L"

$$D_k = \exp\left(L_k - \frac{1}{2}\sum_{i=1}^k \tau_i^2\right).$$

We can check that D_k is also a martingale – this should look very similar to the continuous exponential martingale we talked about earlier in class – and for any finite time, we can now define a tilt $\frac{d\mathbb{Q}}{d\mathbb{P}} = D_n$. By applying our above argument, $X_i \sim N(a_i\tau_i, 1)$, which means that M_k under $\mathbb{Q}$ behaves like M_k under $\mathbb{P}$ **plus an extra drift term** $\sum_{i=1}^k \sigma_i a_i \tau_i$.

In other words, we find that $M_k - \sum_{i=1}^k \sigma_i a_i \tau_i$ is a martingale under $\mathbb{Q}$ (while M_k itself is a martingale under $\mathbb{P}$). And this drift term is kind of a **measure of the covariation of M with L**: since M is the sum of $\sigma_i X_i$ and L is the sum of $\tau_i Y_i$, it makes sense that there is a covariation of $a_i\sigma_i\tau_i$. The point of Girsanov's theorem is to give a continuous-time version of this:

Theorem 121 (Girsanov's theorem, informal)
Let M and L be local martingales under $\mathbb{P}$, and define a change of measure via $\frac{d\mathbb{Q}}{d\mathbb{P}} = D_\infty = \mathcal{E}(L)_\infty$. Then $M - \langle M, L\rangle$ is a martingale under $\mathbb{Q}$.

The resemblance between this result and the discrete case should be clear. We should be a bit careful here – it's not always true that $\mathcal{E}(L)_\infty$ is a valid Radon-Nikodym derivative because of **absolute continuity**, and let's see how that can fail in the discrete case. Suppose we have probability measures μ, ν on $(\Omega, \mathcal{F})$ with $\mathcal{F}_n \uparrow \mathcal{F}$, and suppose that $\nu_n = \nu|_{\mathcal{F}_n}$ are absolutely continuous with respect to $\mu_n = \mu|_{\mathcal{F}_n}$ for all n. Then we know that $D_n = \frac{d\nu_n}{d\mu_n}$ is a martingale under μ, but it may not need to converge under ν, so we define $D_\infty = \limsup D_n$. We can then decompose ν into a continuous and singular part as

$$\nu(A) = \int_A D_\infty d\mu + \nu(A \cap \{D_\infty = \infty\}),$$

which just tells us that we may not have absolute continuity as long as there is a positive chance that D_n diverges in the limit.

Example 122
Let $\Omega = \prod_{i=1}^\infty \{0,1\}$, and let $\mathcal{F}_n = \{A \times \prod_{i=n+1}^\infty \{0,1\}, A \subseteq \{0,1\}^n\}$ be the set of events that only depend on the first n variables. Suppose that $\mu = \bigotimes_{i=1}^\infty \text{Ber}(p)$ and $\nu = \bigotimes_{i=1}^\infty \text{Ber}(q)$ for some $0 < p < q < 1$.

It's clear that μ and ν are not absolutely continuous with respect to each other: a sample $X \sim \mu$ looks like $(X_1, X_2, \cdots)$ where the X_i are iid Bernoulli with parameter p, and similarly $X \sim \nu$ looks like iid Bernoullis with

parameter q. In particular, $\frac{1}{n}\sum_{i=1}^n X_i$ converges to p under one measure and q under the other by the law of large numbers.

But in the discrete case, we know that μ_n is just the law of $(X_1, \cdots, X_n)$ under μ. Even if μ and ν are singular with respect to each other, we do have $\mu_n \ll \nu_n \ll \mu_n$ for all finite n – "seeing n bits doesn't tell us for sure whether it comes from an iid Bernoulli p or iid Bernoulli q." Indeed, we can calculate explicitly that

$$D_n = \frac{d\nu_n}{d\mu_n} = \prod_{i=1}^n \frac{e^{\theta X_i}}{m(\theta)},$$

where θ is chosen exactly so that the mean shifts from p to q, meaning it satisfies

$$\mathbb{E}_\theta[X_i] = q = \frac{\mathbb{E}\left[X_i e^{\theta X_i}\right]}{\mathbb{E}[e^{\theta X_i}]} = \frac{pe^\theta}{pe^\theta + 1 - p}.$$

Doing the algebra and substituting back in, we find that D_n concentrates around a specific point: we have

$$D_n = \frac{\left(\frac{q(1-p)}{p(1-q)}\right)^{S_n}}{\left(\frac{1-p}{1-q}\right)^n} \approx \left(\frac{q}{p}\right)^{nq}\left(\frac{1-q}{1-p}\right)^{n(1-q)} e^{o(n)} = \exp\left(nH(q|p)\right),$$

where $H(q|p)$ is the binary relative entropy. In particular, $H(q|p) = 0$ if $q = p$ and otherwise $H(q|p) > 0$, meaning that if $p \neq q$ then $D_n \to \infty$ and thus μ, ν are mutually singular.

Basically, we should remember that Girsanov's theorem doesn't work for arbitrary L, so we need to understand which L actually produce a valid change of measure in the continuous case. We'll discuss this more next time!

14 April 6, 2020

We'll finish the discussion of Girsanov's theorem today – we'll start by recalling last week's calculation. Suppose we're on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we have independent Gaussians of the bivariate distribution

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a_i \\ a_i & 1 \end{bmatrix}\right),$$

such that $\mathcal{F}_n$ is the sigma-algebra generated by these variables. If σ_i, τ_i are bounded random variables that are $\mathcal{F}_{i-1}$-measurable, we can define $M_k = \sum_{i=1}^k \sigma_i X_i$ and $L_k = \sum_{i=1}^k \tau_i Y_i$, which are martingales under $\mathbb{P}$. Considering the process up to some finite time n, we can calculate the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = D_n = \prod_{i=1}^n \exp\left(\tau_i Y_i - \frac{\tau_i^2}{2}\right) = \exp\left(L_n - \sum_{i=1}^n \frac{\tau_i^2}{2}\right).$$

Then by the martingale property, $\frac{d\mathbb{Q}}{d\mathbb{P}}\big|_{\mathcal{F}_k} = D_k$ for all $k \leq n$, and D_k is a discrete-time version of the exponential martingale $\mathcal{E}(L)_t = \exp\left(L_t - \frac{1}{2}\langle L\rangle_t\right)$ (remember that in the continuous case, this gives us a strictly-positive continuous local martingale). This D then helps us define a **change of measure**: we found last time that $X_i \sim N(a_i\tau_i, 1)$ is a shifted Gaussian under $\mathbb{Q}$ (conditioned on $\mathcal{F}_{i-1}$, so that we know the value of τ_i), so this means that M_k under $\mathbb{Q}$ looks like M_k under $\mathbb{P}$ but with an extra drift term $\sum_{i=1}^k \sigma_i\tau_i a_i$, which we can think of as a discrete-time version of the "covariation of M with L."

At the end of last lecture, we stated the informal version of Girsanov's theorem: if M and L are local martingales under $\mathbb{P}$, we can define a change of measure via $\frac{d\mathbb{Q}}{d\mathbb{P}} = D_\infty = \mathcal{E}(L)_\infty$, and $M - \langle M, L\rangle$ will be a local martingale under $\mathbb{Q}$. We'll formalize this today: assume we are working on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, where our

filtration $\mathcal{F}_t$ is right-continuous and complete.

Proposition 123
Suppose $\mathbb{Q} \ll \mathbb{P}$. Then $D_t = \frac{d\mathbb{Q}}{d\mathbb{P}}\big|_{\mathcal{F}_t}$ is a uniformly integrable martingale, so it has an rcll modification.

We'll skip the proof of this – the fact that it's a martingale is easy to check from the definition of conditional expectation, and the rcll modification comes from results of Chapter 3. The main point is that we'll be working with such rcll modifications from now on.

Lemma 124
Suppose D_t is a continuous local martingale with $D_0 = 1$ such that $D_t > 0$ for all t. Then we can write $D_t = \mathcal{E}(L)_t$ for some continuous local martingale L.

Proof. Apply Itô's formula to $Y_t = \log D_t$ to find

$$dY_t = \frac{1}{D_t} dD_t - \frac{1}{2D_t^2} d\langle D_t \rangle.$$

If we take L_t such that $dL_t = dY_t + \frac{1}{2D_t^2} d\langle D_t \rangle$ (to cancel out the finite variation term), then L is a local martingale with

$$dL_t = \frac{1}{D_t} dD_t \implies d\langle L \rangle_t = \frac{1}{D_t^2} d\langle D \rangle_t.$$

which we can substitute back in to find $dY_t = dL_t - \frac{1}{2} d\langle L \rangle_t$. Integrating this yields $\log D_t = Y_t = L_t - \frac{1}{2}\langle L \rangle_t$, and finally exponentiating both sides tells us that $D_t = \mathcal{E}(L)_t$, as desired. □

In particular, this proof tells us the explicit formula $L_t = \int_0^t \frac{1}{D_s} dD_s$.

Theorem 125 (Girsanov)
Assume that $\mathbb{Q} \ll \mathbb{P}$, and $D_t = \frac{d\mathbb{Q}}{d\mathbb{P}}\big|_{\mathcal{F}_t} = \mathcal{E}(L)_t$. Also assume that $\mathcal{F}_0$ is trivial, so $D_0 = 1$. If M is a continuous local martingale under $\mathbb{P}$, then $M - \langle M, L \rangle$ is a continuous local martingale under $\mathbb{Q}$.

This theorem essentially tells us that the class of martingales only changes by the drift term $\langle M, L \rangle$ – in particular, the quadratic variation of a continuous local martingale under $\mathbb{P}$ and under $\mathbb{Q}$ are the same.

Proof. Let X be any adapted process. We first claim that if $D \cdot X$ (the product, not the stochastic integral) is a continuous martingale under $\mathbb{P}$, then X is a continuous martingale under $\mathbb{Q}$. To check this, first we make sure X is in L^1 – indeed, $\mathbb{E}_{\mathbb{Q}}[|X_t|] = \mathbb{E}_P[D_t|X_t|] = \mathbb{E}_P[|D_t X_t|]$ (because D is positive), and this right-hand side is finite by assumption of $D \cdot X$ being a martingale. Now we check the martingale property: for any $s \leq t$ and any event $A \in \mathcal{F}_s$,

$$\mathbb{E}_{\mathbb{Q}}[X_t 1_A] = \mathbb{E}_{\mathbb{P}}[D_t X_t 1_A] = \mathbb{E}_{\mathbb{P}}[D_s X_s 1_A] = \mathbb{E}_{\mathbb{Q}}[X_s 1_A],$$

where we've used the definition of change of measure in the first and third equalities and the martingale property in the second. Thus $\mathbb{E}_{\mathbb{Q}}[X_t|\mathcal{F}_s] = X_s$ as desired. And similarly, we can show that if $D \cdot X$ is a continuous **local** martingale under $\mathbb{P}$, then X is a continuous **local** martingale under $\mathbb{Q}$.

We'll now apply this to $X = M - \langle M, L \rangle$: we want to show that this is a martingale under $\mathbb{Q}$, so it suffices to show that $D \cdot X$ is a martingale under $\mathbb{P}$. Remember that D evolves via the formula $dD_t = \mathcal{E}(L)_t dL_t = D_t dL_t$ (see the explicit expression from Lemma 124), so using Itô's formula (with the function $F(D, X) = D \cdot X$) we have

$$d(D \cdot X)_t = D_t dX_t + X_t dD_t + d\langle D, X \rangle_t$$

(only the mixed partial terms are nonzero for the function F, and the factors of 2 cancel). $X_t dD_t$ is already a martingale, so we don't need to expand it out further. However, we can plug in the formula for X_t and note that X is M plus a finite variation term, so this simplifies to

$$X_t dD_t + D_t(dM_t - d\langle M, L\rangle_t) + d\langle D, X\rangle_t = X_t dD_t + D_t(dM_t - d\langle M, L\rangle_t) + d\langle D, M\rangle_t.$$

Now $d\langle M, L\rangle_t = \frac{1}{D_t} d\langle M, D\rangle_t$, so the last two terms cancel, and we're just left with $d(D \cdot X)_t = X_t dD_t + D_t dM_t$. Since there is no finite variation term (and D_t is also a martingale), this indeed shows that $D \cdot X$ is a local martingale, completing the proof. □

Note that this is not the typical way that we apply Girsanov's theorem – often we start with a continuous local martingale L such that $L_0 = 0$ and $\langle L\rangle_\infty < \infty$ almost surely. We know that this means L_t converges almost surely to a limit L_∞, so $\mathcal{E}(L)_t$ is a continuous local martingale. In particular, because it is nonnegative, it is a supermartingale, and thus it converges almost surely to $\mathcal{E}(L)_\infty$ with $\mathbb{E}[\mathcal{E}(L)_\infty] \leq 1$ by Fatou's lemma. **If we have equality**, then $\mathcal{E}(L)_t = D_t$ is a uniformly integrable martingale (see our homework), and thus we can define $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(L)_\infty$ and apply Girsanov with this L. So we need to make sure L satisfies the condition $\mathbb{E}[\mathcal{E}(L)_\infty] = 1$ to make sure all of this is valid.

Fact 126

Theorem 5.23 in Le Gall gives a few criteria for this condition being satisfied. Specifically, if L is a continuous local martingale with $L_0 = 0$, then **Novikov's condition** $\mathbb{E}\left[\exp\left(\frac{1}{2}\langle L\rangle_\infty\right)\right] < \infty$ implies that L is a uniformly integrable martingale with $\mathbb{E}\left[\exp\left(\frac{1}{2}L_\infty\right)\right] < \infty$ (**Kazamaki's criterion**), which implies that $\mathcal{E}(L)$ is a uniformly integrable martingale.

We can read the proof on our own, but we'll instead focus on applications during class. Our first one will be to **constructing a solution for a stochastic differential equation**:

Example 127

Suppose we want to solve the differential equation $dX_t = b(t, X_t)dt + dB_t$, where b is a measurable function with $|b(t, x)| \leq g(t)$ for some g satisfying $\int_0^\infty g(t)^2 dt < \infty$.

Solution. Let X be a Brownian motion under $\mathbb{P}$, and let $L_t = \int_0^t b(s, X_s)dX_s$. Since X is a Brownian motion under $\mathbb{P}$, we have $\langle L\rangle_\infty = \int_0^\infty b(t, X_t)^2 dt$ (since $d\langle X\rangle_t = t$), and this is finite because it is bounded by $\int_0^\infty g(t)^2 dt$. Thus, Novikov's condition is satisfied, which means that we can define a new measure $\mathbb{Q}$ such that $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(L)_\infty$. Applying Girsanov's theorem now tells us that $B = X - \langle X, L\rangle$ is a local martingale under $\mathbb{Q}$.

Because B is X minus a finite variation process, Lévy's characterization tells us that B is a Brownian motion under $\mathbb{Q}$ because it has the correct quadratic variation. But $X = \langle X, L\rangle + B$ can be rewritten in differential form as $dX_t = b(t, X_t)dt + dB_t$ by plugging in the definition of L, and this is exactly what we wanted. □

Notice that the only assumption we needed is that $b(t, x)$ is measurable and bounded by an L^2 function $g(t)$ – no other regularity condition was required! Our next application is the **Cameron-Martin formula**:

Example 128

Let $L_t = \int_0^t g(s)dB_s$ for some deterministic function $g(s)$, and again define $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(L)_\infty$. Then $\tilde{B} = B_t - \langle B, L\rangle_t = B_t - \int_0^t g(s)ds$ will be a Brownian motion under $\mathbb{Q}$, as long as $\mathbb{E}[\mathcal{E}(L)_\infty] = 1$.

Thus is an **explicit change of measure** between a Brownian motion and a Brownian motion plus a deterministic function $h(t)$ of the form $h = -\int_0^t g(s)ds$. But note that not all functions h will work: for example, the law of $B_t + ct$ is not absolutely continuous with respect to the law of B_t, because $\frac{B_t}{t}$ goes to 0 almost surely while $\frac{B_t+ct}{t}$ goes to c almost surely (so the measures are in fact mutually singular). To study this in more detail, notice that if we define

$$A_t = \langle L \rangle_t = \int_0^t g(s)^2 ds,$$

then L_t behaves as β_{A_t}, where β is a Brownian motion. Since A_t is a **deterministic** time change, this means that L_t is distributed normally as $N(0, A_t)$, and thus **if** $\int_0^\infty g(t)^2 dt = \langle L \rangle_\infty$ is finite, then L_∞ is just distributed as $N(0, \langle L \rangle_\infty)$, so indeed we will have $\mathbb{E}[\mathcal{E}(L)_\infty = 1]$. (So here, we don't need Novikov's condition to see that this last condition holds, because we can calculate the law directly.) In other words, this means that the law of $(B_t + h(t))$ is absolutely continuous with respect to the law of B_t if and only if we can write $h(t) = \int_0^t g(s)ds$ such that $\int_0^\infty g(t)^2 dt < \infty$; such functions h form the **Cameron-Martin (CM) space**.

Example 129

We'll spend the remainder of this class discussing an application to the large deviations principle – this often goes under the name of **Schilder's theorem**.

First, we recall Cramér's theorem, which tells us about large deviations for the empirical mean $\frac{1}{n}\sum_{i=1}^n X_i$ of a random variable. Suppose that $m(\theta) = \mathbb{E}\left[e^{\theta X_i}\right]$ is finite for all $\theta \in \mathbb{R}$, allowing us to define the cumulant generating function $\kappa(\theta) = \log m(\theta)$. Cramér's theorem then tells us that for any $a > \mathbb{E}[X]$,

$$\frac{1}{n}\log \mathbb{P}\left(\frac{S_n}{n} \geq a\right) \to -I(a),$$

where $I(a) = \sup_\theta(\theta a - \kappa(\theta))$. In other words, the probability is exponentially decaying with rate given by this function I. We proved the upper bound by using Markov's inequality, and to show the lower bound, we used a change of measure. Specifically, choose θ so that $\mathbb{E}_\theta[X_i] = a + \varepsilon$, and define $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{\exp(\theta S_n)}{\exp(n\kappa(\theta)}$. Then the idea is that the tilted mean of X_i is slightly larger than a, so the event $\left\{\frac{S_n}{n} \geq a\right\}$ is now a typical event (since the sum of n iid terms with mean slightly larger than a is likely to give something larger than a). So

$$1 \approx \mathbb{Q}\left(\frac{S_n}{n} \geq a\right) \geq \mathbb{Q}\left(a \leq \frac{S_n}{n} \leq a + 2\varepsilon\right).$$

From there, we noticed that the Radon-Nikodym derivative is roughly constant on this event, so this is approximately

$$\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} 1\left\{a \leq \frac{S_n}{n} \leq a + 2\varepsilon\right\}\right] \approx \exp\left(\theta n a - n\kappa(\theta)\right) \mathbb{P}\left(\frac{S_n}{n} \in [a, a+\varepsilon]\right).$$

We're going to do something similar now, but with Brownian motion sample paths instead:

Theorem 130 (Schilder)

For simplicity, consider Brownian motion on the interval $[0, T]$. Let $C[0, T]$ be the space of continuous functions with the sup-norm $||\cdot||_\infty$, and let $W[0, T]$ be the subspace of functions in $C[0, T]$ that start at 0. Then for any $A \subseteq W[0, T]$,

$$-\Lambda(A^\circ) \leq \liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}(\sqrt{\varepsilon} B \in A) \leq \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}(\sqrt{\varepsilon} B \in A) \leq -\Lambda(\overline{A}),$$

where A° is the interior of A, $\overline{A}$ is the closure of A, and $\Lambda(A) = \inf_{h \in A} I(h)$, where $I(h) = \frac{1}{2}\int_0^T h'(t)^2 dt$ is analogous to the $I(a)$ of Cramér's theorem.

The key feature here is not the appearance of the liminf and limsup – instead, it's the new function $I(h)$ that is relevant. If we cared in Cramér's theorem about a general Borel set A, we'd have the liminf and limsup there, too.

Proof sketch for lower bound. Consider the set $A = \{f : ||f - h||_\infty < \delta\}$. We want to make an analogous argument as in Cramér, turning A into a typical event. Thus, consider the tilt

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\frac{1}{\sqrt{\varepsilon}}\int_0^T h'(t)dB_t - \frac{1}{2\varepsilon}\int_0^T h'(t)^2 dt\right).$$

Under this new measure, the Cameron-Martin formula tells us that being in A is basically like following a regular Brownian motion but with a drift term $\frac{h}{\sqrt{\varepsilon}}$, which is exactly what we want for B in the theorem statement. Thus $1 \approx \mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} 1_A\right]$, and now we can approximately evaluate the Radon-Nikodym derivative by taking its value at $B = \frac{h}{\sqrt{\varepsilon}}$ to find that

$$1 \approx \exp\left(\frac{1}{2\varepsilon}\int_0^T h'(t)^2 dt\right)\mathbb{P}(A) \implies \varepsilon \log \mathbb{P}(A) \approx -\frac{1}{2}\int_0^T h'(t)^2 dt = -I(h),$$

which is indeed the form of the desired inequality. □

If we're curious how we can extend this argument from a finite time interval to $[0, \infty)$, we can refer to the book of Deuschel and Stroock.

15 April 8, 2020

We haven't covered everything from chapter 5, but we'll hold off on stochastic differential equations for now and spend the next few lectures on **continuous-time Markov processes**, for which the theory goes beyond Brownian-type processes. (Most of this comes from Le Gall chapter 6, but we'll go a bit beyond that as well.) We'll start with a review of the discrete-time Markov chains:

Definition 131

A **discrete-time Markov process** or **Markov chain** on a finite state space $E = \{1, \cdots, k\}$ is a discrete E-valued process $(X_n)_{n\geq 0}$ specified by a **transition matrix** $P \in \mathbb{R}^{k\times k}$ with entries $p_{xy} = \mathbb{P}(X_{n+1} = y | X_n = x)$.

This matrix P can be thought of as a map from $\mathbb{C}^k$ to $\mathbb{C}^k$, meaning that it is a linear operator, and in particular it also acts on functions $f : E \to \mathbb{C}$ via

$$(Pf)(x) = \sum_y p_{x,y} f(y) = \mathbb{E}\left[f(X_{n+1}) | X_n = x\right].$$

(This is the view we'll take in the continuous-time case as well.) Here, P is a **stochastic matrix** – its rows sum to 1, so the constant vector $P1 = 1$ is a right eigenvector with eigenvalue 1. And if we have n steps of the chain (that is, if we only observe the state every nth step), the transition matrix is just P^n – this fact (across all n) goes by the name of the **Chapman-Kolmogorov equations**. Recall the following general result:

Theorem 132 (Perron-Frobenius)
If a Markov chain is irreducible (meaning there is a sequence of steps from any state to any other state) and aperiodic (the gcd of all directed cycle lengths is 1), then 1 is a simple eigenvalue (of multiplicity 1), and all other eigenvalues have modulus strictly less than 1. Then the associated left eigenvector π^* for the eigenvalue 1 (satisfying $\pi^* P = \pi^*$) has all positive entries, and it is the stationary distribution of the chain.

This can be thought of as a general theorem about matrices with real positive entries. The idea is that if we diagonalize $P = UDU^{-1}$, then the column vectors of U are the right eigenvectors u_i, and the rows of U^{-1} are the left eigenvectors v_i. But now we can write out

$$P^n = UD^nU^{-1} = \sum_{i=1}^{k} (d_i)^n u_i v_i^*.$$

And now since 1 is a simple eigenvalue but all of the other eigenvalues have modulus less than 1, $P^n\pi^*$ will approach $1\pi^*$ (the other contributions go away as $n \to \infty$).

For our purposes, what will be interesting is first generalizing the state space and then turning this into a continuous process. We'll begin by looking at a general **measurable** state space $(E, \mathcal{E})$:

Definition 133
A **Markov transition kernel** on a space $(E, \mathcal{E})$ is a map $Q : E \times \mathcal{E} \to [0, 1]$ such that

- for all $x \in E$, the function $Q(x, \cdot)$ is a probability measure on $(E, \mathcal{E})$, and
- for all $A \in \mathcal{E}$, the function $Q(\cdot, A)$ is measurable.

Here, $Q(x, \cdot)$ represents the law of the process at some future time, given that we're in state x right now, analogous tothe row vector of x in the transition matrix P. The second condition about $Q(\cdot, A)$ is more technical and comes from generalizing the equation $(Pf)(x) = \mathbb{E}[f(X_{n+1})|X_n = x]$ from above. Specifically, if we have a measurable function $f : E \to \mathbb{C}$, then we can define $Qf(x) = \int_E f(y)Q(x, dy)$ – this is the Lebesgue integral of f against the measure $Q(x, \cdot)$ – analogously to the discrete case. We just want that for any measurable bounded function f, the function Qf is also measurable and bounded, and this is where we use the measurability condition for $Q(\cdot, A)$. (Qf being measurable follows directly for indicator functions, and then we approximate with indicators in general.)

We'll let $B(E)$ denote the set of bounded measurable functions on E with the sup-norm $||f|| = ||f||_\infty$. Any Markov transition kernel Q maps $B(E)$ to itself – in fact, Q is a **contractive operator**, meaning that $||Qf|| \leq ||f||$, because Qf is an expectation of the function f and thus uniformly bounded by its sup-norm. Note that so far, E has had no conditions other than being a measurable space, but moving forward we'll require further regularity conditions (and point them out as they're needed).

Definition 134
A **transition semigroup** on a state space $(E, \mathcal{E})$ is a collection of transition kernels $(Q_t)_{t\geq 0}$ which satisfy the following conditions:

- For all $x \in E$, $Q_0(x, \cdot)$ (the law of where we go in zero time) is the Dirac measure δ_x.
- For all $s, t \geq 0$, $Q_{s+t} = Q_s Q_t$.
- For all measurable $A \in \mathcal{E}$, the map $(t, x) \to Q_t(x, A)$ is measurable with respect to $\mathcal{B}_{[0,\infty)} \otimes \mathcal{E}$.

The second condition here is the Chapman-Kolmogorov condition, and it's less trivial than before because now we're making sure that all of our Q_ts are consistent. (And here Q_sQ_t is defined via composition of operators on $B(E)$.)

Definition 135

A **Markov process with transition semigroup Q_t** on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is an adapted process $(X_t)_{t\geq 0}$ such that $\mathbb{E}[f(X_{s+t})|\mathcal{F}_s] = (Q_t f)(X_s)$ for all bounded measurable functions $f \in B(E)$ and all $s, t \geq 0$.

Since this process has restrictions, we should make sure it does actually exist. If E is nice enough (for example, if we have a Polish space), then existence of such a Markov process with the transition semigroup $(Q_t)_{t\geq 0}$ comes from the Kolmogorov extension theorem. (We have the uncountable index set is $[0, \infty)$, and we just need to specify consistent finite-dimensional distributions, but those come from the Q_ts.) We won't check the Kolmogorov extension theorem itself, but the point is just to make sure E is sufficiently nice.

Remark 136. *However, just like with the construction of Brownian motion and martingales, there is no guarantee of sample path regularity in this definition. We may talk a bit about this later on, but we'll focus on the aspects that are different from what we've already seen.*

Going back to the discrete-time case, notice that we can also define a semigroup $(Q_n)_{n\geq 0}$, but the object is less useful because it's just $(I, P, P^2, P^3, \cdots)$, so in particular it's specified by a single transition matrix P. So the first mystery is whether there is a "basic building block" analogous to P for Markov processes in continuous time which encodes information of the entire semigroup $(Q_t)_{t\geq 0}$. The answer is "generally yes," but the story is more complicated – this is only true for **Feller processes**, which we'll define later. To make that analogy, we'll need a few more concepts:

Definition 137

The **λ-resolvent** of a semigroup $(Q_t)_{t\geq 0}$ (for some $\lambda > 0$) is the operator $R_\lambda : B(E) \to B(E)$ such that

$$(R_\lambda f)(x) = \int_0^\infty e^{-\lambda t} Q_t f(x) dt.$$

We should think of this as the Laplace transform (in the time-coordinate) of the semigroup. Remember that $Q_t f$ is the expectation of f at time t, given that we're currently at x – since the exponential distribution $\text{Exp}(\lambda)$ has density $\lambda e^{-\lambda t}$, what this definition is really saying is that

$$(R_\lambda f)(x) = \frac{1}{\lambda}\mathbb{E}[f(X_{\tau_\lambda})|X_0 = x],$$

where τ_λ is a random time distributed according to $\text{Exp}(\lambda)$. (So when λ is large, we emphasize times close to 0, and vice versa.)

Lemma 138 (Resolvent equation)

For any $\lambda, \mu > 0$, we have $R_\lambda - R_\mu + (\lambda - \mu)R_\lambda R_\mu = 0$.

Proof. It's enough to prove this for $\lambda \neq \mu$ (otherwise this is clearly 0). The composition of the two resolvents is

$$\boxed{(R_\lambda(R_\mu f))(x)} = \int_0^\infty e^{-\lambda s} Q_s(R_\mu(f(x)))ds = \int_0^\infty e^{-\lambda s} Q_s \left(\int_0^\infty e^{-\mu t} Q_t f dt\right)(x) ds.$$

Expanding out the definition of Q_s, this is

$$\int_0^\infty e^{-\lambda s} \int_E \left(\int_0^\infty e^{-\mu t} Q_t f(y) dt \right) Q_s(x, dy) ds,$$

and now we can apply Fubini's theorem (because there are no integrability issues when we have bounded functions) to get

$$\int_0^\infty e^{-\lambda s} \int_0^\infty e^{-\mu t} \int_E Q_t f(y) Q_s(x, dy) dt ds.$$

Now the inner integral $\int_E Q_t f(y) Q_s(x, dy)$ means that we start at x and evolve for time s, ending up at y, and finally evaluate $Q_t f$. This means we are at state y and evolve at time t, and see what the value of f looks like there, so this inner integral all just evaluates to $Q_{s+t} f(x)$. Thus, this can all be rewritten as

$$\int_0^\infty e^{-\lambda s} e^{\mu s} \int_0^\infty e^{-\mu s} e^{-\mu t} Q_{s+t} f(x) dt ds,$$

where we've slipped in an $e^{\mu s - \mu s}$ to separate the two integrals. Setting $r = s + t$ turns this into

$$\int_0^\infty e^{-\lambda s} e^{\mu s} \int_s^\infty e^{-\mu r} Q_r f(x) dr ds,$$

which is a double integral over pairs (r, s). Changing the order of integration and then evaluating the inner integral yields

$$= \int_0^\infty e^{-\mu r} Q_r f(x) \int_0^r e^{-\lambda s} e^{\mu s} ds dr = \int_0^\infty Q_r f(x) \frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} dr,$$

which is exactly $\boxed{\frac{R(\mu) - R(\lambda)}{\lambda - \mu} f(x)}$; comparing this to the boxed expression above yields the result. □

This is mostly an algebraic manipulation – we won't use it today, but it will come up again in the next few lectures. The main idea is to become more familiar with the idea of composition of operators, and the key idea of the proof here was "composing" the operators using Chapman-Kolmogorov.

For our next step, we're going to need some more regularity – assume that our space E is metrizable, locally compact (meaning that around any point, we can find a compact set that contains a neighborhood of the point), and σ-compact (meaning E is a countable union of compact sets). In particular, this implies that E is a Polish space – examples of such spaces E include open subsets of $\mathbb{R}^d$, as well as much more general spaces. Let $\mathcal{E}$ be the Borel σ-field of E, and write $E = \bigcup_{n=1}^\infty K_n$ (where the K_n are **compact and nested** – this exists by assumption of being σ-compact).

Definition 139

A function $f : E \to \mathbb{R}$ **tends to zero at infinity** if $\lim_{n \to \infty} \sup_{x \in E \setminus K_n} |f(x)| = 0$.

Remember that a Polish space is defined to be separable (containing a countable dense subset) and completely metrizable (topologically homeomorphic to a complete metric space). For example, if we take $E = (0, 1)$, this is locally compact and σ-compact (it's the countable union of the sets $\left[\frac{1}{n}, 1 - \frac{1}{n}\right]$), so it is a Polish space. Note that while E is not a complete metric space, because the point $\frac{1}{n}$ doesn't converge to anything in E, it **is** instead completely metrizable, because $(0, 1)$ is topologically homeomorphic to $\mathbb{R}$. So being completely metrizable is a **topological** property – we don't need to put a complete metric on the space. (In the language of the definition, "tending to zero at infinity" for the interval $(0, 1)$ means that we tend to 0 at the endpoints 0 and 1.)

We'll let $C_0(E)$ denote the set of continuous real functions on E tending to zero at infinity. This is a subspace of

$B(E)$, which is a Banach space with the sup norm, so we'll also look at $C_0(E)$ with the sup norm. And we now have enough of the topological setup to define the class of processes that we want:

Definition 140

On a space E satisfying the above conditions, a **Feller semigroup** is a transition semigroup $(Q_t)_{t\geq 0}$ such that

- Q_t maps $C_0(E)$ into itself for all $t \geq 0$, and
- for all $f \in C_0(E)$, $||Q_t f - f|| \to 0$ as $t \to 0$.

A **Feller process** is a Markov process with a Feller semigroup.

Remember that $Q_t f(x)$ is the expected value of f at time t, given that we're at x at time 0. So the Feller property tells us that the process doesn't make a large jump in a small amount of time, since the values of $Q_t f$ are close to the corresponding values of f. However, discontinuous jumps are still allowed, and many natural Feller processes do have jumps – it's just that we're not likely to make a jump immediately at any particular time. (We should think of having a process that evolves continuously for some interval and then makes a jump occasionally.)

Definition 141

For a Feller semigroup $(Q_t)_{t\geq 0}$, let the **domain of L**, denoted $D(L)$, be the set of $f \in C_0(E)$ such that $\frac{Q_t f - f}{t}$ converges in $C_0(E)$ (in the sup-norm) as $t \downarrow 0$. Then we can define the **(infinitesimal) generator** L of $(Q_t)_{t\geq 0}$ to be the operator $D(L) \to C_0(E)$ such that

$$Lf = \lim_{t\downarrow 0} \frac{Q_t f - f}{t}.$$

We'll spend the rest of the lecture on a heuristic preview of the material for next week. We should think of L as the "derivative" of Q_t at time $t = 0$, but by the Chapman-Kolmogorov equations we know that

$$\frac{d}{dt}Q_t = \lim_{s\to 0} \frac{Q_{t+s} - Q_t}{s} = \lim_{s\downarrow 0} \frac{Q_s Q_t - Q_t}{s} = LQ_t.$$

We know that a real-valued function $q : [0, \infty) \to \mathbb{R}$ solving the differential equation $q' = \ell q$ with initial condition $q(0) = 1$ has unique solution $q(t) = e^{\ell t}$, and the Laplace transform of such a function is $r(\lambda) = \int_0^{-\infty} e^{-\lambda t} q(t) dt = \frac{1}{\lambda - \ell}$ for $\lambda > \ell$. **By analogy**, we might guess that Q_t is similarly an exponential of the form $Q_t = e^{tL} = \sum_{k\geq 0} \frac{(tL)^k}{k!}$, and that the resolvent is $(\lambda - L)^{-1}$. This isn't a rigorous argument, but it's our best guess, and our goal for next week will be to explore how valid this analogy is.

Example 142

Consider Brownian motion on $\mathbb{R}^d$ – we'll study it from the perspective of Feller processes.

For each t, $Q_t(x, \cdot)$ is the distribution $N(x, tI_{d\times d})$ (this is how the process evolves in time t when started from x), so the generator of Brownian motion looks like

$$Lf = \lim_{t\downarrow 0} \frac{\mathbb{E}\left[f(x + B_t)\right] - f(x)}{t}.$$

We don't actually have all of the tools needed to evaluate this rigorously, but if f is nice enough and we let our Brownian motion go for a small amount of time, B_t is small and thus we should be able to Taylor expand. This means that

$$Lf = \lim_{t\downarrow 0} \frac{1}{t}\mathbb{E}\left[\nabla f(x) \cdot B_t + \frac{1}{2} B_t^t (\text{Hess } f) B_t\right],$$

and B_t has mean 0 so the first term goes away in expectation. The Hessian of f is a $d \times d$ matrix, and in this case we only picking up the diagonal terms because the Brownian motion B_t has independent entries in all d dimensions. Thus, we'll get the trace of the Hessian, which is the Laplacian, $\boxed{Lf = \frac{1}{2}\Delta f(x)}$. This highlights the connection between Dirichlet theory and Brownian motion, which we'll explore more soon!

We can also calculate R_λ in this case, but the integral is more complicated: it turns out to be related to the Green kernel (the inverse of the Laplacian) when we take $\lambda \downarrow 0$, since we're saying in that case that the resolvent is the inverse of L.

16 April 13, 2020

Remark 143. *In the survey responses that we filled out, it was mentioned that some of us are having trouble seeing the slides during class – as a reminder, there's a Dropbox link at the bottom of the course webpage, which has slides basically synchronized with lecture.*

We'll be talking more about Feller processes today, as well as a special case of these processes which is particularly simple. For review, say we have a Markov process X_t with transition semigroup $(Q_t)_{t\geq 0}$ – recall that this means $\mathbb{Q}_t(x,\cdot) = \mathbb{P}(X_{s+t} \in \cdot | X_s = x, \mathcal{F}_s)$ for all $s,t \geq 0$ and $x \in E$. As discussed last time, we can view Q_t as an operator on $B(E)$: given a bounded function f, we define the function $Q_t f$ via $(Q_t f)(x) = \int f(y) Q_t(x,dy) = \mathbb{E}[f(X_{s+t})|C_s = x]$. The **Chapman-Kolmogorov equations** tell us that $Q_{s+t} = Q_s Q_t$, and we can define a Laplace transform $R_\lambda = \int_0^{-\infty} e^{-\lambda t} Q_t dt$. (Since $\lambda e^{-\lambda t}$ integrates to 1, λR_λ is a Markov kernel, meaning that multiplying by λ makes it properly normalized.) We previously showed the resolvent equation $R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$, which in particular shows that R_λ and R_μ commute.

Last time, we defined Feller processes to be those such that $||Q_t f - f|| \to 0$ as $t \downarrow 0$ for all $f \in C_0(E)$. We also defined the space $D(L)$ (the "domain" of L) to be the space of functions $f \in C_0(E)$ such that $\lim_{t\downarrow 0} \frac{Q_t f - f}{t}$ **exists in** $C_0(E)$ (with respect to the sup-norm topology). Our goal today is to understand how L determines the semigroup Q_t.

Proposition 144
Q_t and L commute with each other on $D(L)$ (the space on which they are both defined) for all t.

Proof. We first write out

$$Q_t L f = Q_t \left(\lim_{s \downarrow 0} \frac{Q_s f - f}{s} \right),$$

but now the limit is in the sup-norm and Q_t is an operator which is contractive (the norm of Qf is at most the norm of f). Therefore, we can also bring the Q_t inside the limit to get

$$\lim_{s\downarrow 0} \frac{Q_t Q_s f - Q_t f}{s} = \lim_{s \downarrow 0} \frac{Q_s (Q_t f) - (Q_t f)}{s} = L Q_t f,$$

with the middle equality by Chapman-Kolmogorov, and this is exactly what we wanted to prove. □

The next result we'll prove is a differential relation:

Proposition 145
For all $f \in D(L)$ and for all $t \geq 0$, we have $\int_0^t Q_s L f ds = Q_t f - f = \int_0^t L Q_s f ds$.

In other words, the time-derivative of $Q_t f$ is given by $Q_t L f$. The two equalities were historically proven at different times – they're known as the Kolmogorov forward and backward equations, respectively.

Proof. By the previous proposition, we just need to prove one of the two equalities. Fix a point $x \in E$, define $h(t) = Q_t f(x)$, and take the right-derivative of h – we see that

$$\lim_{s \downarrow 0} \frac{1}{s} \left[Q_{s+t} f(x) - Q_t f(x) \right] = L Q_t f(x).$$

But since Q_t is contractive, this convergence happens uniformly in x – more explicitly, we have

$$\left\| Q_t \left(\frac{1}{s}(Q_s - I) f \right) \right\| \leq \left\| \frac{1}{s}(Q_s - I) f \right\|,$$

and because the right-hand side converges uniformly, the left-hand side does too. So we in fact have $h'(t) = L Q_t f(x)$ for all x, and integrating in t yields the result that we want. □

We can explain the names of "forward" and "backward" equation a bit more now – for a large class of (**diffusive**) processes, L turns out to be a differential operator. (For example, we showed last time that $L = \frac{1}{2}\Delta$ for a Brownian motion.)

- For the Kolmogorov forward equation, suppose we know the initial distribution $X_0 \sim \mu$ and we want to know how the forward evolution affects the law of the process at some final time. We know that $X_t \sim \mu Q_t(dy) = \int_{x \in E} \mu(dx) Q_t(x, dy)$. But because we know that $\frac{d}{dt}(\mu Q_t) = (\mu Q_t) L$, this gives us a partial differential equation: if μQ_t has a density $p(t, x)$ in the x-coordinate at time t, then

  $$\partial_t p(t, x) = L_x^* p(t, x),$$

  and we can solve this PDE "forward in time" by using the initial conditions $p(0, x)$ that come from the initial density μ.

- Meanwhile, for the Kolmogorov backward equation, suppose we want to calculate the expected value $\mathbb{E}[f(X_T) | X_0 = x] = (Q_t f)(x)$. This time we'll make use of the other equation $\frac{d}{dt} Q_t f(x) = L Q_t f(x)$ – if we write $f(t, x) = Q_t f(x)$, the defining partial differential equation is now

  $$\partial_t f(t, x) = L_x f(t, x),$$

  which we can solve "backward in time" from the **final** condition $f(T, x) = f(x)$.

We'll now return to the connection from last time between the **resolvent** and the **generator**:

Proposition 146
For a semigroup Q_t, the **range** of R_λ, denoted $\mathcal{R} = \{R_\lambda f : f \in C_0(E)\}$, doesn't depend on λ. Also, $\mathcal{R}$ is a dense subset of $C_0(E)$ in the sup-norm topology.

Proof. The resolvent equation can be rewritten as $R_\mu = R_\lambda \left(I + (\lambda - \mu) R_\mu \right)$, so the range of R_μ is contained in the range of R_λ. But the roles of μ and λ are interchangeable here, so that means the range of R_λ is the same for all λ. To show that $\mathcal{R}$, we can consider $\lambda R_\lambda f$ for any function $f \in C_0(E)$, which can be explicitly written out as

$$\lambda R_\lambda f = \int_0^\infty \lambda e^{-\lambda t} Q_t f \, dt.$$

This can be thought of as waiting an exponential (random variable) amount of time and evolving f by that length, and this simplifies after a change of variables to $\int_0^\infty e^{-t} Q_{t/\lambda} f\, dt$. Now as $\lambda \to \infty$, $Q_{t/\lambda} f$ converges uniformly to f by the Feller property and e^{-t} is integrable. Thus by the dominated convergence theorem, this integral converges uniform to $\int_0^\infty e^{-t} f\, dt = f$, as desired (we can approximate any f by functions in $\mathcal{R}$ with vanishing sup-norm error). □

Theorem 147

For any Feller semigroup, $D(L) = \mathcal{R}$, and the two functions $R_\lambda : C_0(E) \to \mathcal{R}$ and $\lambda - L : D(L) \to C_0(E)$ are inverses of each other.

Proof. It suffices to show that (1) $(\lambda - L) R_\lambda g = g$ for all $g \in C_0(E)$, and (2) $R_\lambda(\lambda - L) f = f$ for all $f \in D(L)$ (this will also show that the domain and range line up). For (1), we know that

$$L R_\lambda g = \lim_{s \downarrow 0} \frac{1}{s} \left(Q_s R_\lambda g - R_\lambda g \right) = \lim_{s \downarrow 0} \frac{1}{s} \left(Q_s \left(\int_0^\infty e^{-\lambda t} Q_t g\, dt \right) - \left(\int_0^\infty e^{-\lambda t} Q_t g\, dt \right) \right),$$

We can use Fubini's theorem to move the Q_s inside the integral and then do a change of variables to simplify (after some rearranging, since replacing t with $s + t$ shifts the bound) to

$$\lim_{s \downarrow 0} \frac{1}{s} \left(e^{\lambda s} \int_0^\infty e^{-\lambda s} e^{-\lambda t} Q_{s+t} g\, dt - \int_0^\infty e^{-\lambda t} Q_t g\, dt \right) = \lim_{s \downarrow 0} \frac{1}{s} \left((e^{\lambda s} - 1) \int_0^\infty e^{-\lambda t} Q_t g\, dt - e^{\lambda s} \int_0^s e^{-\lambda t} Q_t g\, dt \right).$$

But this simplifies to $\lambda R_\lambda g - g$ as $s \downarrow 0$, and rearranging proves the claim. Now for (2), take any $f \in D(L)$ and apply the Kolmogorov forward equation to rewrite

$$\lambda R_\lambda f = \lambda \int_0^\infty e^{-\lambda t} Q_t f\, dt = \int_0^\infty \lambda e^{-\lambda t} \left(f + \int_0^t Q_s L f\, ds \right) dt.$$

But now the first term integrates out to f, and we can swap the order of integration on the second term, so this is also equal to

$$f + \int_0^\infty Q_s L f \int_s^\infty \lambda e^{-\lambda t}\, dt\, ds = f + \int_0^\infty e^{-\lambda s} Q_s L f\, ds = f + R_\lambda L f,$$

and again comparing to the original expression yields the result. □

Corollary 148

A Feller semigroup Q_t is uniquely determined by its generator L (though we do need to specify the domain $D(L)$ for which the limit is well-defined).

Proof. Let $g \in C_0(E)$ be a nonnegative function (which in particular means $Q_t g$ is nonnegative as well). Knowing the generator L tells us $R_\lambda g = (\lambda - L)^{-1} g$ for all λ. But since $R_\lambda g = \int_0^\infty e^{-\lambda t} (Q_t g)(x)\, dt$, this means we know the Laplace transform of $Q_t g(x)$ for all λ and thus know $Q_t g(x)$ itself. This characterizes Q_t for any nonnegative function, which is enough to characterize it for all of $C_0(E)$. □

Our description here is less explicit than in the discrete case, in which we just said that $Q_n = P^n$. So it's natural to ask if we have something like $Q_t = \exp(tL)$, and that's what we'll discuss next. But first, we'll do an example:

Example 149

Consider a Brownian motion in $\mathbb{R}^d$, for which we've already shown that $L = \frac{1}{2}\Delta$.

Since Q_t tells us about the probability of going from x to y, we can write down the semigroup explicitly as a Gaussian density:

$$Q_t(x, dy) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x-y|^2}{2t}\right) dy.$$

The resolvent is then

$$R_\lambda(x, dy) = \int_0^\infty e^{-\lambda t} Q_t(x, dy) = \int_0^\infty e^{-\lambda t} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x-y|^2}{2t}\right) dt dy,$$

and we know in general that this will be equal to $(\lambda - L)^{-1}$. The inverse of the Laplacian is the **Green kernel**, so it's natural to plug in $\lambda = 0$ in this example. The resolvent will not always be defined at $\lambda = 0$ (we can't always evaluate the Laplace transform at 0), but in this case the integral converges as long as $d > 2$ (since for large t the exponential term just approaches 1). Computing explicitly then indeed shows us that $R_\lambda(x, dy) = 2G(x, dy)$ at $\lambda = 0$. The reason this argument doesn't work well when $d = 1, 2$ is that the Brownian motion returns to each state infinitely often; it is still possible to get the classical Green kernel back, but we need to do some renormalization.

We'll now turn our attention to the question posed earlier about whether $Q_t = \exp(tL)$ by thinking about the following situation:

Definition 150

Let Y_n be a discrete-time Markov chain on a space $(E, \mathcal{E})$ with a transition kernel $P(x, dy)$. The canonical way to turn such a chain into a continuous time process is to let N_t be a Poisson process of some rate c (which means in particular that N_t is an integer distributed according to $\text{Pois}(ct)$). Then we construct a **pseudo-Poisson process** via $X_t = Y_{Nt}$.

Here, X and Y have the same trajectory up to a time change, and the only difference is that Y jumps at integer times while X jumps at random times τ_n (specifically, the increment $\tau_{n+1} - \tau_n$ are iid exponential random variables with rate c). Then the generator of such a process is related to the chance of making a jump in a time t, where t is small. But the probability that an exponential clock rings in time t is $1 - e^{-ct} \approx ct$, so

$$Lf(x) = \lim_{t \downarrow 0} \mathbb{E}\left[\frac{f(X_t) - f(x)}{t} \middle| X_0 = x\right] = c \int (f(y) - f(x)) P(x, dy).$$

In operator notation, this says that $\boxed{L = c(P - I)}$, so if we know the jump rate and transition kernel of a dsicrete Markov chain, we can find the corresponding L for the pseudo-Poisson process. The generator L is also a bounded operator, since $||Lf|| \leq 2c||f||$ from the integral representation above, so we can indeed define

$$\exp(tL)f = \sum_{k=0}^\infty \frac{(tL)^k f}{k!}.$$

We wish to show that e^{tL} is indeed the same as Q_t for the continuous-time process, and we can show this by writing out

$$e^{tL} = e^{tc(P-I)} = e^{-ctI} e^{ctP} = e^{-ct} \sum_{k=0}^\infty \frac{(ctP)^k}{k!} = \sum_{k=0}^\infty \left(\frac{e^{-ct}(ct)^k}{k!}\right) P^k.$$

But now we can plug in the mass function for the Poisson distribution – this right-hand side is exactly $\sum_{k=0}^\infty \mathbb{P}(N_t = k) P^k$. And this is exactly how the process should evolve in time t: we first figure out how many times we update the chain, and then we evolve via P that many times. So we do indeed have $Q_t = e^{tL}$ in this simple case.

We'll finish by discussing the **Yosida approximation theorem**. Suppose we're back in the general case with a Feller semigroup $(Q_t)_{t \geq 0}$ and a generator L. Since we have no control on the boundedness of L, we can't always define

$\exp(tL)$, but we can define a λ-approximation (remembering that λR_λ "doesn't evolve very much" for large λ)

$$L^{(\lambda)} = \lambda L R_\lambda = \lambda\left(\lambda - (\lambda - L)\right) R_\lambda = \boxed{\lambda(\lambda R_\lambda - I)},$$

where we've used the fact that $\lambda - L$ is the inverse of R_λ. (A similar argument shows that this operator can also be written as $\lambda R_\lambda L$.) But here, the boxed expression looks a lot like $L = c(P - I)$, so $L^{(\lambda)}$ is the generator of a pseudo-Poisson process with transition kernel λR_λ and rate λ. We'll let $Q_t^{(\lambda)} = \exp(tL^{(\lambda)})$ denote the semigroup with this generator $L^{(\lambda)}$.

Theorem 151 (Yosida approximation theorem)
For all $f \in D(L)$, $||L^{(\lambda)}f - Lf||$ converges to 0 in the sup-norm as $\lambda \to \infty$. In addition,

$$\left|\left|Q_t^{(\lambda)}f - Q_t f\right|\right| \le t\left|\left|L^{(\lambda)}f - Lf\right|\right|,$$

so the left-hand side converges on any bounded interval. In fact, for all $f \in C_0(E)$ (a larger set than $D(L)$), $||Q_t^{(\lambda)}f - Q_t f||$ converges to 0 on any bounded time interval.

So this λ-approximation idea gives us one way to approximate a Markov process, which is to use simple processes of the form $Q_t^{(\lambda)}$ and take $\lambda \to \infty$. But at the end of the day, we're interested in the process, not the semigroup, so we're curious about whether the processes $(Q_t^{(\lambda)})_{t\ge 0}$ converge in law to Q_t as well. It turns out that this holds somewhat generally, and we might talk about this later on.

17 April 15, 2020

We'll cover three separate topics today, concluding our discussion of Markov processes. Our first is **sample path regularity** for Markov processes, which will provide some nice connections with martingales. Last time, we discussed that the resolvent and generator are related for a Feller process, and the other reason that this resolvent is important is that we can construct a supermartingale with it:

Lemma 152
Let X_t be a Markov process with semigroup $(Q_t)_{t\ge 0}$ and resolvent R_λ. For any bounded nonnegative function $h \in B(E)$, the process $S_t = e^{-\lambda t}R_\lambda h(X_t)$ is a supermartingale for all $\lambda > 0$.

Here, we're using the function h to go from an abstract space E to the reals, so that we can do things like addition and subtraction.

Proof. Clearly S_t is a nonnegative process, and we can check that $S_t \in L^1$ for all t. Indeed, λR_λ is a bounded (Markov transition) operator, meaning that $||\lambda R_\lambda h|| \le ||h||$, so there are no integrability issues at any finite time t. To check the supermartingale property, we write

$$\mathbb{E}\left[S_{s+t}|\mathcal{F}_s\right] = e^{-\lambda(s+t)}\mathbb{E}\left[R_\lambda h(X_{s+t})|\mathcal{F}_s\right] = e^{-\lambda(s+t)}\mathbb{E}\left[\int_0^\infty e^{-\lambda r}Q_r\,dr\,h(X_{s+t})\middle|\mathcal{F}_s\right].$$

We can move the expectation inside the integral by Fubini's theorem and then evaluate that expectation using the definition of the semigroup, so this simplifies to

$$e^{-\lambda(s+t)}\int_0^\infty e^{-\lambda r}\mathbb{E}\left[Q_r h(X_{s+t})|\mathcal{F}_s\right]dr = e^{-\lambda(s+t)}\int_0^\infty e^{-\lambda r}Q_t Q_r h(X_s)dr.$$

Applying Chapman-Kolmogorov and then performing a change of variables, we end up with

$$e^{-\lambda s}\int_0^\infty e^{-\lambda(t+r)}Q_{t+r}h(X_s)dr = e^{-\lambda s}\int_t^\infty e^{-\lambda r}Q_r h(X_t)dr.$$

But now h is a nonnegative function, meaning the integrand is always nonnegative, and thus this last expression is bounded from above by $e^{-\lambda s}\int_0^\infty e^{-\lambda r}Q_r h(X_t)dr = e^{-\lambda s}R_\lambda h(X_s) = S_s$, verifying the supermartingale inequality. □

Earlier in the class (chapter 3), we proved that when our filtration is right-continuous and complete and X_t is a supermartingale such that $t \to \mathbb{E}[X_t]$ is right-continuous, X has an **rcll** modification X which is also a supermartingale. We'll take advantage of this to prove that Markov processes have a similar property:

Theorem 153

Suppose $\mathcal{F}_t$ is right-continuous and complete on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and X_t is a Feller process with semigroup Q_t. Then X has a modification $\tilde{X}$ which is also a Markov process with semigroup Q_t and rcll sample paths.

Proving this directly is not so straightforward – the crucial fact that we used to prove the original result about martingales was Doob's upcrossing inequality (which gave us control over right and left limits). Markov processes don't necessarily have upcrossings in generic spaces E, but this result shows that working with nice enough (Feller) processes still gives us enough control.

Proof sketch. We'll assume first that E is compact, so that $C_0(E) = C(E)$ (in other words, the semigroup is defined on all continuous functions). As a (topological) exercise, there exists a countable subset $\{f_n\} \subseteq C(E)$ which **separate** the points of E, meaning that for all $x \neq y$, there is some n such that $f_n(x) \neq f_n(y)$. Consider the countable set

$$\mathcal{H} = \{R_p f_n : p, n \in \mathbb{N}\}.$$

We showed last time that λR_λ converges to the identity as $\lambda \to \infty$, and thus $\mathcal{H}$ also separates the points of E (for any x, y that are distinct, we can find an f_n such that $f_n(x) \neq f_n(y)$, and then we can pick sufficiently large p so that $R_p f_n(x) \neq R_p f_n(y)$). Now for any $h = R_p f_n \in \mathcal{H}$, define the process

$$S_t^h = e^{-pt}h(X_t) = e^{-pt}R_p f_n(X_t).$$

This is a supermartingale by Lemma 152, and the Feller property tells us that the expected value $t \to \mathbb{E}[S_t^h]$ is right-continuous. Thus, our arguments from chapter 3 show that there is a modification of S_t^h which is rcll, and now we can simultaneously define the countably many modifications $\tilde{S}_t^h$ for all $h \in \mathcal{H}$. To finish, we take a countable dense subset $D \subseteq [0, \infty)$ and take the limits

$$\lim_{s \downarrow t, s \in D} X_s(\omega), \qquad \lim_{s \uparrow t, s \in D} X_s(\omega).$$

(Remember that the S_t^hs are real-valued, while the X_s are E-valued – the claim we're making is that these limits above exist **in** E.) Indeed, we would violate the rcll property for some supermartingale $\tilde{S}_t^h$ if these limits didn't exist: if there were two sequences $s_k \downarrow t$ and $\tilde{s}_k \downarrow t$ (taking sequence values in D), where $X_{s_k}(\omega) \to x$ and $X_{\tilde{s}_k}(\omega) \to y$, then letting h separate x and y. we would find that $\tilde{S}_t^h$ has a limit along both of these sequences, which is a contradiction. Thus, all of the limits do exist and X has the desired rcll modification.

In general, if E is not a compact space, we can use a one-point compactification. Applying the above argument to $E \cup \{\Delta\}$ gives an rcll modification $\tilde{X}$ on the larger state space, and we just need to show that $\tilde{X}$ does not visit this extra point Δ. This basically follows by using the Markov semigroup to verify that $e^{-t}h(\tilde{X}_t)$ has hitting times at $\frac{1}{n}$ going to infinity as $n \to \infty$. □

Theorem 154

Under the same setting as the theorem above, the process $\tilde{X}$ satisfies the strong Markov property.

We won't say anything more about this here; the main argument is that it's a Markov process, so it satisfies the simple Markov property, and then we can approximate random times by a discrete set of possibilities, where being a Feller process helps with continuity.

We'll now move on to our next topic, **Lévy processes**. These processes don't have to be continuous, so they aren't covered as much in our textbook, but there's a large field of research about them – we can see [4] for more information. In short, this is a specific class of Feller processes (including Brownian motion and Poisson processes) which are "spatially homogeneous."

Definition 155

A **Lévy process** is a real-valued process X_t with stationary and independent increments (meaning that for all $s \le t$, $X_t - X_s$ is independent of $\mathcal{F}_s$ and $X_t - X_s$ is equidistributed as X_{t-s}), such that X_t converges in probability to $X_0 = 0$ as $t \to 0$.

To understand how the spatial homogeneity point affects our situation here, we may write

$$Q_t(x, dy) = \mathbb{P}(X_{s+t} \in dy | X_s = x) = \mathbb{P}(X_{s+t} - X_s \in d(y - X_s) | X_s = x) = Q_t(d(y - x)),$$

since the conditioning does not affect our probability. In other words, this is a Markov process where deciding how we'll move at the next step is independent of our current position. There are a lot of known facts about these processes, and we'll talk today about the characterizing features of a Levy process. The characteristic function for X_t can be written as

$$\mathbb{E}\left[e^{i\theta X_t}\right] = \mathbb{E}\left[e^{i\theta(X_{t/2} + (X_t - X_{t/2}))}\right],$$

and the increments $X_{t/2}$ and $X_t - X_{t/2}$ are iid, so we can factor this. In fact, we can break this up into arbitrarily small chunks, and the idea is that we end up with a characteristic function of the form $\mathbb{E}\left[e^{i\theta X_t}\right] = e^{t\psi(\theta)}$. Such a process does not have a lot of degrees of freedom: for example, X_t is **infinitely divisible**, meaning

$$X_t = X_{t/k} + (X_{2t/k} - X_{t/k}) + \cdots + (X_t - X_{t-t/k}),$$

where all k terms on the right side are iid increments. The class of infinitely divisible processes is indeed somewhat restricted:

Theorem 156

A Lévy process is characterized by three numbers (a, σ^2, ν) (a drift term, a rate of diffusion, and a jump measure), where ν is a possibly infinite signed measure on $\mathbb{R} \setminus \{0\}$ such that $\int \min(1, x^2)\nu(dx) < \infty$.

Proof sketch. First of all, Lévy processes are Feller processes (we can see our textbook for this), so we can assume that we're working with an rcll modification X_t. Such a process X_t has countably many discontinuities, which are the points where $\Delta X_t = X_t - X_{t-} = 0$. Take the empirical measure of the jumps

$$\eta = \sum_t \delta_{(t, \Delta X_t)}$$

(we can think of this as drawing points in the t, X_t plane corresponding to the jumps), so that η is a random measure

on $[0,\infty)\times(\mathbb{R}\setminus\{0\})$. Because our process has stationary and independent increments, η must be independent on distinct time blocks, meaning that η must be a Poisson random measure with intensity $\mathbb{E}[\eta] = (\text{Leb})\otimes\nu$. So ν is the distribution of the jumps, and that accounts for the discontinuity.

From here, we want to remove the jumps and end up with a continuous process, and the idea is that what we end up with is basically a Brownian motion. Ideally, we'd subtract off $\sum_{s\le t}\Delta X_s$ – there's only countably many jumps – but we don't know if this sum is convergent. Instead, we subtract off all jumps that are large: the rcll property tells us that there are only **finitely** many large jumps, so we can define $J_t = \sum_{s\le t}\Delta X_s 1\{|\Delta X_s|>1\}$. Now let $\varepsilon\in(0,1]$, and define

$$M_t^\varepsilon = \sum_{s\le t}(\Delta X_s - \mathbb{E}\Delta X_s)1\{|\Delta X_s|\in(\varepsilon,1]\}.$$

This is also well-defined for any positive ε, and on our homework we'll show that M_t^ε has a well-defined limit M_t as $\varepsilon\to 0$. (We'll need the second-moment property on ν, and then we'll need the martingale L^2 inequality.) So now the process $Y = X - M - J$ is a continuous Lévy process which will turn out to be of the form $at+\sigma B_t$. The issue is that we've **assumed nothing about integrability** in the definition of a Lévy process, and we'll see how to fill in the details on the homework. □

Again, the significant point here is that Lévy processes only do two things: the continuous part evolves as a Brownian motion, and the discontinuous part has jumps evolving at a Poissonian rate.

Theorem 157

Let ξ be a real-valued random variable. Then the following are equivalent:

- There exists a Lévy process such that X_1 is equidistributed as ξ,
- The law of ξ is infinitely divisible,
- There exists a triangular array $\xi_{i,j}$ such that each row is an iid sequence of length m_n and $\sum_{j=1}^{m_n}\xi_{n,j}\xrightarrow{d}\xi$.

This is an important result, because one idea from last semester (Lindeberg-Feller) is that such a triangular array with mild conditions forces ξ to be a normal random variable. But we also know that there are non-Gaussian infinitely divisible random variables (such as the Cauchy, gamma, and Poisson distribution) in which we violate the Lindeberg-Feller conditions. In particular, in such situations the random variables in our array are "heavy-tailed," and this theorem is interesting because it covers **all** iid triangular arrays with row sums converging to a limit (in contrast with Lindeberg-Feller).

We'll close with a brief note about our final topic, **approximation of Markov chains**. Recall that last time, we had a Feller semigroup Q_t with generator L, and we mentioned that $L^{(\lambda)} = \lambda(\lambda R_\lambda - I)$ generates a pseudo-Poisson process with an associated semigroup $Q_t^{(\lambda)} = \exp(tL^{(\lambda)})$. The Yosida approximation theorem then stated that $Q_t^{(\lambda)}\to Q_t$ as $\lambda\to\infty$, and now if $X_t^{(\lambda)}$ is such a realization – that is, a process with semigroup $Q_t^{(\lambda)}$ – we want to know whether $X^{(\lambda)}$ converges in distribution to X.

One example this can help us understand is whether a simple random walk converges in distribution to Brownian motion. And another example comes from the fact that the average of n Cauchy random variables is Cauchy (we can show this by looking at the characteristic function): we may ask whether the process $X_t^{(n)} = S_{\lceil nt\rceil}/n$ (where S sums up some number of Cauchy random variables) converges as a distribution to a continuous-time analog (namely the Lévy process with X_1 distributed according to the Cauchy distribution), given that $X_t^{(n)}$ converges to $t\cdot(\text{Cauchy})$ for any fixed t.

The first issue we need to worry about is the topology for convergence in distribution – if we have an rcll random variable that can also have jumps, the **sup-norm topology is not a good choice** anymore. For example, the process

$X^{(n)}(t)$ which is 0 for some time $1 + \frac{1}{n}$ and then takes on the value 1 after that should converge to the process which jumps up to 1, but this isn't true in the sup-norm topology. Instead, we use the **Skorohod topology**, which allows us to slightly reparameterize the time. And from here, the idea is that **semigroup convergence is equivalent to weak convergence in the Skorohod topology** for Feller processes. It turns out that semigroup convergence is easier to show in these kinds of situations (we can check each Q_t on its own) than showing convergence in law directly.

18 April 22, 2020

We'll quickly finish discussing Markov processes today (chapter 6) and move on to some preparation for potential theory (chapter 7). Recall that we started by discussing Markov chains in discrete time in a finite state space, where the chain is completely specified by a transition matrix P. Under mild conditions (being irreducible and aperiodic), the Perron-Frobenius theorem guarantees convergence in law to a unique stationary distribution π^* (that is, $\pi^* P = \pi^*$). In continuous time, the dynamics of the system are now specified by a semigroup Q_t, and in a particularly nice class of processes known as Feller processes, we have a generator L that determines the semigroup, and we know that the λ-resolvent satisfies $R_\lambda = (\lambda - L)^{-1}$.

A further subclass of these Feller processes is the space of psuedo-Poisson processes, where the generator L just looks like $c(P - I)$: in such a case, $Q_t = \exp(tL)$, and the Yosida approximation tells us that a Feller process can be approximated by these psuedo-Poisson processes. Finally, a different subclass of the Feller processes is the Lévy processes, which include standard Brownian motion B_t, the standard Poisson process N_t, and the **first passage time process** T_a (that is, the infimum t such that $B_t > a$). This last process $(T_a)_{a \geq 0}$ is indeed a Lévy process because it has independent increments, and we can completely characterize its behavior:

Fact 158

We have $T_a = \int_{[0,a]\times[0,\infty]} x\eta(dsdx)$, where η is a Poisson random measure with intensity Leb $\otimes \frac{1\{x>0\}}{\sqrt{2\pi}x^{3/2}}$.

This is a special case of the processes on our homework – it's good to see this worked out if we haven't seen anything like it before, and the details are on the online notes.

With this, we'll move on to **potential theory** – we'll first cover the subject in a discrete setting because there's enough going on already in that case. The central idea of what we'll be doing is **relating Markov chains to Dirichlet problems**.

Definition 159

A Markov chain Y_n evolving on a discrete state space V is **reversible** if there exists a symmetric function $c : V \times V \to [0, \infty)$ such that $p(x, y) = \frac{c(x,y)}{c(x)}$, where $c(x) = \sum_y c(x, y)$.

We'll only discuss reversible Markov chains here – any such chain can be described by a weighted graph $G = (V, E, c)$, where the vertices are the elements of the state space and the (undirected) edges are of the form (xy), where $c(x, y) > 0$. (For any edge e, $c(e)$ is called its **conductance**.) We can also define the **weighted adjacency matrix** $A = \{c(x, y)\}$ – assuming non self-loops, the matrix has zeros on the diagonal. If we let D be the diagonal matrix such that $D(x, x) = c(x)$ (this is like the degree of x in the unweighted case), then the transition matrix of the chain is $P = D^{-1}A$. For a function $u : V \to \mathbb{R}$, consider the function

$$Lu(x) = \mathbb{E}_x \left[u(Y_1) - u(Y_0)\right].$$

This is a discrete-time version of the generator, measuring the expected change in one step of the chain given that we started chain at x. But this is just $(Pu)(x) - u(x) = (P - I)u(x)$, and we call $L = P - I$ the **weighted graph Laplacian**. In particular, we can also write this out as $L = P - I = D^{-1}A - I = D^{-1}(A - D)$.

Remark 160. *To understand the name "graph Laplacian," consider the case where we have a random walk on $G = \mathbb{Z}^d$. Then x is a point in $\mathbb{Z}^d$, and there are $2d$ possibilities for where we can go, and thus*

$$Lu(x) = \frac{1}{2d}\sum_{i=1}^{d}(u(x+e_i) - u(x)) - (u(x) - u(x-e_i))\,.$$

This is a difference of first derivatives, so this behaves like a second derivative in the ith coordinate (summed over i), and thus L acts basically like $\frac{1}{2d}\sum_i \partial_i^2$.

From here on, we'll assume that G is a finite connected graph, and we'll be interested in **potential functions** (also called **harmonic functions**), which are essentially functions u such that $Lu = 0$. First note that if we actually require $Lu = 0$ to hold everywhere, then u must be constant. Indeed, take any $x \in \text{argmax}(u)$; since u is harmonic,

$$Lu(x) = 0 \implies u(x) = \mathbb{E}_x[u(Y_1)].$$

But $u(x)$ is the largest possible value of u, and the right hand side is a weighted average of values of u, so this means $u(y) = u(x)$ for all y adjacent to x. Continuing throughout the connected graph, this means that u is constant, as claimed – this is known as the **maximum principle**.

So to make things more interesting, we'll take some subset B of the vertices V, which we call the **boundary** B, and we won't require the function u to be harmonic on B. Then we get a more general maximum principle:

Lemma 161
Let G be a finite connected graph with boundary B and interior $U = V \setminus B$, and let $u : V \to \mathbb{R}$ be a function such that $(Lu)|_U = 0$ and $u|_B = 0$ (harmonic on the interior, zero on the boundary). Then $u = 0$.

Proof. Again consider $x \in \text{argmax}_{x\in U}(u(x))$. Since $u(x)$ is still a weighted average of us around it, we again have $u(y) = u(x)$ for all $y \sim x$. Continuing in this way, we eventually reach the boundary (where the value is fixed to be zero), and this tells us that everything must be zero. □

Definition 162
A **Dirichlet boundary value problem** consists of finding a harmonic function $u : V \to \mathbb{R}$ such that $(Lu)|_U = 0$ and $u|_B = f$ for some boundary condition f. If such a u exists, the solution is called the **harmonic interpolation** of f.

Note that if we have two solutions u', u'' with the same boundary data f, then $u' - u''$ is still a harmonic function but equal to zero on the boundary. Thus Lemma 161 tells us that $u' - u'' = 0$ – in other words, the harmonic interpolation must be unique. To show existence, we can write down a solution directly by defining

$$u(x) = \mathbb{E}_x\left[f(Y_\tau)\right],$$

where τ is the first hitting time of the boundary. This has the correct boundary conditions, because $\tau = 0$ for any $x \in B$, meaning that $u(x) = f(x)$, and we can check for ourselves that $Lu(x) = 0$ for all $x \in U$. (This means there is always a unique harmonic interpolation from the boundary to the rest of the graph, as long as both the boundary and the interior are nonempty. And the boundary does not need to correspond in any visual sense to an actual boundary.)

Example 163

The reason for the name "potential function" is that we can view $G = (V, E, c)$ as a wiring diagram of an **electrical network**. Think of the vertices V as nodes and edges E as wire connections with an associated number $c(e)$, representing the **electrical conductance** of e. (In particular, $r(e) = \frac{1}{c(e)}$ is the **electrical resistance** we might have seen in a physics class.)

There are two main facts to know about electrical networks:

- (Ohm's law) If we hold two vertices x and y at fixed voltages $v(x)$ and $v(y)$ (for instance, the two ends of a battery), then that imposed voltage difference creates an electrical current

$$i(x, y) = \frac{v(x) - v(y)}{r(xy)} = c(x)\frac{c(xy)}{c(x)}(v(x) - v(y)) = c(x)p(x, y)(v(x) - v(y)).$$

  (This is how we define the current i, and we'll take the convention that everything "flows downhill" from positive to negative voltage.) Now if we have a **voltage function** $v : V \to \mathbb{R}$ which tells us what the voltage is at each vertex, then the **net current** into the node x from its neighbors y is $(\text{div } i)(x) = \sum_{y\sim x} i(y, x)$. Substituting the previous expression in and using that $c(x)p(x, y) = c(y)p(y, x)$, we thus have

$$(\text{div } i)(x) = c(x)\sum_{y\sim x} p(x, y)(v(y) - v(x)) = c(x)Lv(x).$$

- (Kirchhoff's node law) If x is not connected to an external electrical source or sink, then the current into x is the same as the current out of x. (If electrons flow in, they also need to flow out.) In other words, this means that $\boxed{Lv(x) = 0}$ (the voltage function should be harmonic) **everywhere other than the sources and sinks**.

This means that the Dirichlet boundary value problem with boundary condition $f : B \to \mathbb{R}$ will take on the value of the voltage function $v : V \to \mathbb{R}$ if we impose voltages f on B. This is a nice view, because it's easy to calculate properties of electrical networks: for example, having two edges in parallel with conductance c_1 and c_2 is equivalent to a single edge with conductance $c_1 + c_2$, and having two edges in series with resistance r_1 and r_2 yields a single edge with resistance $r_1 + r_2$. So there are various rules for reducing a network, and sometimes this reduction allows us to take a complicated network and reduce to a single wire with an effective conductance (if we, for example, impose a voltage of 1 on one vertex and a voltage of 0 on another). We'll make this more formal:

Definition 164

Suppose our boundary $B \subset V$ is partitioned as $A \sqcup Z$, where A is the set of **sources** and Z is the set of **sinks**. If we have the boundary condition $f = 1_A$ (think of this as connecting the positive end of a battery to A and the negative end to Z), then define the **total current** and **effective resistance** via

$$I(A \to Z) = -\sum_{x\in A}(\text{div } i)(x) = \sum_{x\in Z}(\text{div } i)(x), \quad R_{\text{eff}}(A \to Z) = \frac{1}{I(A \to Z)} = \frac{1}{C_{\text{eff}}(A \to Z)}.$$

We can check that the effective resistance is the same if we switch A and Z, so we often represent the effective resistance with a double arrow $R_{\text{eff}}(A \leftrightarrow Z)$. This is interesting from a probabilistic point of view, because we can reframe quantities in terms of these conductances and resistances. For simplicity, take $A = \{a\}$ to be a single point, and define the **escape probability** from a to Z

$$\mathbb{P}(a \to Z) = \mathbb{P}_a\left(\inf\{n \geq 1 : Y_n \in Z\} < \inf\{n \geq 1 : Y_n = a\}\right).$$

(This is the probability that we escape to Z before returning to a.) To calculate this, condition on the first step of the chain and write

$$\mathbb{P}(a \to Z) = \sum_{y \sim a} \mathbb{P}_a(Y_1 = y)\mathbb{P}_y\left(\inf\{n \geq 0 : Y_n \in Z\} < \inf\{n \geq 0 : Y_n = a\}\right).$$

Letting τ be the first hitting time of $a \cup Z$, we can rewrite the last probability here as $\mathbb{E}_y[1_Z(Y_\tau)] = \mathbb{E}_y[1 - 1\{Y_\tau = a\}] = 1 - v(y)$, where v is the voltage function corresponding to $v(a) = 1$ and $v(Z) = 0$. Thus, we have

$$\mathbb{P}(a \to Z) = \sum_{y \sim a} p(a, y)(1 - v(y)) = \sum_{y \sim a} p(a, y)(v(a) - v(y)) = -Lv(a),$$

but as derived earlier, this means that

$$\mathbb{P}(a \to Z) = -\frac{(\text{div i})(a)}{c(a)} = \frac{I(a \to Z)}{c(a)} = \frac{C_{\text{eff}}(a \leftrightarrow Z)}{c(a)}.$$

So we have a nice equivalence between a probabilistic quantity and an electrical one!

In the time remaining today, we'll introduce the discrete Green kernel:

Definition 165

Suppose we divide our vertices again into a boundary $B \subset V$ and interior $U = V \setminus B$ – everything we define here depends on our choice of U, but we'll omit the subscript. Let $\tau = \inf\{n \geq 0 : Y_n \in B\}$ be the first time our chain leaves U. The **Green kernel** is defined by

$$G_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{\tau-1} 1\{Y_n = y\}\right].$$

This quantity $G_x(y)$ is the expected number of times our chain hits y when started from x, and it can also be rewritten as

$$G_x(y) = \sum_{n=0}^{\infty} \mathbb{E}_x\left[1\{Y_n = y\}; \tau > n\right] = \sum_{n=0}^{\infty} p_n(x, y),$$

where p_n is the probability that we hit y after n steps and haven't left U yet. We know that $c(x)p_n(x, y) = c(y)p_n(y, x)$ by reversibility, so the quantity $g_y(x) = \frac{G_x(y)}{c(y)} = \frac{G_y(x)}{c(x)}$ is symmetric in x and y; in particular, if G is the matrix with $G(x, y) = G_x(y)$, then $(D^{-1}G)(x, y) = g_x(y)$ is a symmetric matrix. The key identity we should keep in mind is that for any $x \in U$, we can condition on the first step of the chain, so

$$g_x(x) = \frac{G_x(x)}{c(x)} = \frac{1}{c(x)}\left(1 + \mathbb{E}_x[G_{Y_1}(x)]\right) = \frac{1}{c(x)}\left(1 + \frac{\mathbb{E}_x[G_x(Y_1)]c(x)}{c(Y_1)}\right) = \frac{1}{c(x)} + \mathbb{E}_x\left[g_x(Y_1)\right].$$

Thus, the left and right sides of this equation tell us that $Lg_x(x) = -\frac{1}{c(x)}$. However, we can check that for any other point $y \in U \setminus x$, we have $(Lg_x)(y) = 0$ (the same calculation goes through, but we don't get the $+1$ from the starting point, so there is no $\frac{1}{c(x)}$ term). This means that $Lg_x = -\frac{1_x}{c(x)}$, which can be rewritten as the matrix identity $\boxed{LG = -I}$ on U. So we can define G in the probabilistic way (with the expected number of visits), but it turns out to also be equal to the matrix inverse of the Laplacian.

Remark 166. *We should be a bit careful: remember that the definition of L does not depend on the choice of U, but in the identity above, we're restricting L to only contain the rows and columns corresponding to the vertices U.*

19 April 27, 2020

To give us a bit more time to think about the homework, the deadline is pushed to Thursday. To finish the class, there will be a test next Thursday, May 7, and a final problem set on Tuesday, May 12.

Last time, we started discussing potential theory for discrete space and time: we let $G = (V, E, c)$ be a weighted graph, where the weights determine the transition of the Markov chain. Defining the weighted adjacency matrix $A(i,j) = c(ij)$ and letting $D = \text{diag}(c(x) = \sum_y c(x,y)$ be the matrix which tracks the outdegree from each vertex, we find that $P = D^{-1}A$ is the transition matrix, and $L = P - I = D^{-1}(A - D)$. This was helpful for solving the Dirichlet problem, in which we have a vertex set $V = B \sqcup U$ and want to find a function $u : V \to \mathbb{R}$ such that $Lu = 0$ on U and $u = f$ on B. Such a function exists and is unique, and the answer is given by $u(x) = \mathbb{E}_x[f(Y_\tau)]$, where τ is the first hitting time of the boundary B.

We also introduced the Green kernel $G_U(x,y) = \mathbb{E}_x\left[\sum_{n=0}^{\tau-1} 1\{Y_n = y\}\right]$, which tracks the total number of visits to y – we found that $G_U = -(L_u)^{-1}$ can be written in terms of the Laplacian matrix (but only taking the rows and columns from U). Furthermore, we can actually use this to rewrite the solution u. Writing L in block form, the Dirichlet problem asks that

$$\begin{bmatrix} L_U & L_{UB} \\ L_{BU} & L_B \end{bmatrix} \begin{bmatrix} u_U \\ u_B \end{bmatrix} = \begin{bmatrix} 0 \\ * \end{bmatrix},$$

so we must have

$$L_U u_U + L_{UB} u_B = 0 \implies u_U = -(L_U)^{-1} L_{UB} f = G_U L_{UB} f.$$

Here, $K_U = G_U L_{UB}$ is also known as the **discrete Poisson kernel**, and we can write it out as $K_U(x,z) = \sum_y G_U(x,y) L_{UB}(y,z)$. But the only nonzero terms here come from $x, y \in U, z \in B$, so this can also be written as

$$K_U(x,z) = \sum_{y \in U} (G_U(x,y) - G_U(x,z)) p(y,z).$$

Here $G_U(x,z)$ is just zero, but the difference will become a derivative in the continuous analog. Indeed, the **continuous setting** is what we'll talk about today. We won't talk about things in full generality – we started with a general weighted graph in the discrete case, and it's possible to similarly use a general Feller process with infinitesimal generator L, which takes the place of the discrete Laplacian L. But in our case, we'll just discuss Brownian motion in $\mathbb{R}^d$, so we just have $L = \frac{1}{2}\Delta$.

Definition 167

Let $U \subseteq \mathbb{R}^d$ be an open subset. A function $u \in L^1_{\text{loc}}(U)$ (locally integrable, so in particular bounded on compact sets is strong enough) satisfies the **mean value property** on U if for all $x \in U$ and $r > 0$ such that $\overline{B_r(X)} \subset U$,

$$u(x) = \frac{1}{|B_r(x)|} \int_{B_r(x)} u(y)dy = \frac{1}{|S_r(x)|} \int_{S_r(x)} u(y)dy,$$

where $|B_r(x)|, |S_r(x)|$ denote the volume of the ball $B_r(x)$ and sphere $\partial B_r(x)$, respectively.

We'll make use of the following analysis fact:

Fact 168

A function $f \in L^1_{\text{loc}}(U)$ satisfies the mean value property on U if and only if f is harmonic on U, meaning that $\Delta f = \sum_{i=1}^n \partial_i^2 f = 0$.

(In particular, this implies that f is twice continuously differentiable and in fact smooth.) This will help us in our study of the continuous Dirichlet problem: we'll assume for simplicity that U is a bounded domain and our boundary condition is a continuous function $f : \partial U \to \mathbb{R}$. Our goal is then to find a continuous harmonic function $u : \overline{U} \to \mathbb{R}$ such that $\Delta u = 0$ on U and $u = f$ on ∂U. It turns out that the solution looks similar to the discrete case as long as we have some regularity condition:

Theorem 169

If U satisfies the "exterior cone condition," then the solution to the Dirichlet problem is $u(x) = \mathbb{E}_x[f(B_\tau)]$, where τ is the hitting time of the boundary (that is, $\tau = \inf\{t : B_t \notin U\}$).

We won't actually deal with domains that don't satisfy the exterior cone condition in this class, so we won't worry too much about that detail.

Proof sketch. Because U is a bounded domain, ∂U is compact. Since f is continuous and defined on a compact domain, it is bounded, and thus the function u we define above is bounded (in particular, it's definitely in L^1_{loc}). To show that u satisfies the boundary condition, we need the exterior cone condition (details here omitted). For the mean value property, if we consider a ball $B_r(x)$ and let σ be the hitting time of the boundary $S_r(x)$ when we start a Brownian motion from x, then the strong Markov property tells us that

$$\mathbb{E}_x[f(B_\tau)|\mathcal{F}_\sigma] = \mathbb{E}_{B_\sigma}[f(B_\tau)] = u(B_\sigma),$$

and plugging this in after using the law of iterated expectation yields

$$u(x) = \mathbb{E}\left[\mathbb{E}[f(B_\tau)|f(\sigma)]|B_0 = x\right] = \mathbb{E}[u(B_\sigma)|B_0 = x] = \frac{1}{|S_r(x)|}\int_{S_r(x)} u(y)dy,$$

since the distribution of the sphere hitting point is uniform. And the left and right sides of this equation yield the desired mean-value property. □

Next, we'll examine the continuous Green kernel: we can't exactly define a "number of visits" in $\mathbb{R}^d$, but we can just use a density instead.

Definition 170

Let $U \subseteq \mathbb{R}^d$, and let $p_t(x, y; U)$ be the transition kernel of Brownian motion that is killed upon exiting U (more explicitly, $\mathbb{P}_x(B_t \in A; \tau_u > t) = \int_A p_t(x, y; U)dy$ for all A). The **Green kernel** on U is $G_U(x, y) = \int_0^\infty p_t(x, y; U)dt$.

In other words, the total time that we expect to spend in a set A is

$$\mathbb{E}_x\left[\int_0^\infty 1\{B_t \in A\}dt\right] = \int_A G_U(x, y)dy.$$

This integral will be finite except maybe at $y = x$ because we have a bounded domain, but we won't worry too much about those details. We showed in the discrete case that $G_U = -(L_U)^{-1}$, meaning that the function $G_U(x, \cdot)$ is harmonic on $U \setminus \{x\}$ and $(LG_U)(x) = -1$. The first statement is still true in the continuous case, but the second no longer works because G_U is singular at x. Thus, we'll need to restate the "inverse" condition: let $P_{t,u}$ be the operator such that

$$P_{t,U}f(x) = \int_U p_t(x, y; U)f(y)dy.$$

Note that this expression is also equal to $\mathbb{E}_x[f(B_t); \tau_u > t]$. We then also define G_U to act on functions via integration:

$$G_U(f(x)) = \int_U G_U(x,y)f(y) = \int_0^\infty P_{t,U}f(x)dt.$$

Theorem 171

The Green kernel inverts the Laplacian, meaning that for any smooth function f with compact support in U, we have $-\frac{1}{2}\Delta_U G_U f = f$.

Proof sketch. Consider the quantity

$$\boxed{\frac{1}{t}\int_0^t P_{s,U}f(x)ds} = \frac{1}{t}\left(\int_0^\infty P_{s,U}f(x)ds - \int_t^\infty P_{s,U}f(x)ds\right).$$

The first integral on the right-hand side is $G_U(f(x))$, while the second can be rewritten as $P_{t,U}G_U(f(x))$ by Chapman-Kolmogorov, so the expression is equal to $\boxed{-\frac{1}{t}(P_{t,U} - I)G_U(f(x))}$. But now as $t \to 0$, the first boxed expression converges to f, while the $\frac{P_{t,U}-I}{t}$ in the final expression converges to the generator $\frac{1}{2}\Delta_U$ of the process. □

Everything we've been discussing so far has been using probabilistic quantities to say things about harmonic functions, but we can also work in reverse. In $\mathbb{R}^d$, we have the standard Green kernel

$$\Gamma(x,y) = \begin{cases} \frac{|x-y|^{2-d}}{(d-2)|S^{d-1}|} & d \neq 2, \\ \frac{1}{|S^1|}\log\frac{1}{|x-y|} & d = 2, \end{cases}$$

where $|S^{d-1}|$ is the volume of the standard sphere. We can check by directly computing the derivative that $\Gamma(x,\cdot)$ is always harmonic on $\mathbb{R}^d \setminus x$, so this helps us calculate probabilities for Brownian motion:

Example 172

Consider two balls of radius ε and R centered at the origin, and consider a Brownian motion at some x in the annulus U between the balls, stopped when it hits either boundary. We wish to compute the probability $\mathbb{P}_x(\tau_\varepsilon < \tau_R)$.

Letting $f : \partial U \to \mathbb{R}$ be the function which is 1 on the inner boundary ∂B_ε and 0 on the outer boundary ∂B_r, we are trying to compute $\mathbb{E}_x[f(B_\tau)]$. But we can construct the harmonic interpolation explicitly in this case: we have

$$u(x) = \mathbb{P}_x(\tau_\varepsilon < \tau_R) = \begin{cases} \frac{R^{2-d}-|x|^{2-d}}{R^{2-d}-\varepsilon^{2-d}} & d \neq 2, \\ \frac{\log R - \log|x|}{\log R - \log\varepsilon} & d = 2, \end{cases}$$

where we've used the fact that $\Gamma(x,\cdot)$ is harmonic – we can verify that the boundary conditions are indeed satisfied. So this tells us an exact probability for the Brownian motion hitting distance ε before distance R, and now if we take $R \to \infty$, the chance that $\tau_\varepsilon < \tau_R$ goes to 1 for $d = 1, 2$, but it goes to $\left(\frac{\varepsilon}{|x|}\right)^{d-2} < 1$ for $d \geq 3$. So Brownian motion starting at x always hits any ball not containing x in dimensions 1 or 2, but this doesn't necessarily occur in larger dimensions. In other words, Brownian motion is **recurrent** for $d = 1$ or 2 but **transient** otherwise.

Now, we can compare our $G_U(x,y)$ (defined for a bounded U) to the classical $\Gamma(x,y)$ (defined for the whole space). In $\mathbb{R}^d$, Γ inverts the Laplacian, meaning that integrating against a smooth, compactly supported test function f yields

$$\int_{\mathbb{R}} \Gamma(x,y)\Delta_y f(y)dy = f(x).$$

It turns out that we can actually take $U = \mathbb{R}^d$ in the definition of $G_U(x,y) = \int_0^\infty p_t(x,y;U)dt$ – we have that $\int_0^\infty p_t(x,y;U)dt = 2\Gamma(x,y)$ for $d \geq 3$ but not for $d = 1, 2$ (because the integral diverges). Nevertheless, we can still get a relation between the Green kernel (with occupation densities) and the classical kernel: we recenter by taking any fixed vector w of norm 1, and then we have

$$\int_0^\infty \left(p_t(x,y,\mathbb{R}^d) - p_t(x,x+w;\mathbb{R}^d)\right) dt = 2\Gamma(x,y).$$

Example 173

We'll finish by discussing the Feynman–Kac formula, doing a discrete-time version to give some intuition for the continuous-time version on our homework.

As before, let $G = (V,E,c)$ define a reversible Markov chain Y_n. Let $f : V \to \mathbb{R}$ and $w : V \to [0,\infty)$ be two functions, and define

$$u(n,x) = \mathbb{E}_x\left[f(Y_n)\prod_{k=0}^{n-1}\frac{1}{1+w(Y_k)}\right].$$

Then we can calculate $u(n+1,x)$ by conditioning on the first visit of the chain, so that

$$u(n+1,x) = \sum_y p(x,y)\frac{1}{1+w(x)}u(n,y) = \frac{1}{1+w(x)}\left(\sum_y p(x,y)(u(n,y)-u(n,x)) + u(n,x)\right).$$

But the sum over y in the right expression is just the discrete Laplacian $L_x u(n,x)$, meaning that after some rearranging, we arrive at

$$u(n+1,x) - u(n,x) = L_x u(n,x) - w(x)u(n+1,x).$$

This is a discrete PDE with initial condition $u(0,x) = f(x)$. In the continuous version of this result, we're similarly given two functions $f : \mathbb{R}^d \to \mathbb{R}$ and $w : \mathbb{R}^d \to [0,\infty)$ and defining

$$u(t,x) = \mathbb{E}_x\left[f(B_t)\exp\left(-\int_0^t w(B_s)ds\right)\right].$$

We can then find (see homework) that u solves the partial differential equation

$$\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\Delta_x u(t,x) - w(x)u(t,x)$$

again with initial condition $u(0,x) = f(x)$. The special case $w = 0$ just gives us the **heat equation** (heat diffuses like Brownian motion), and the solution in that case is $u(t,x) = \mathbb{E}_x[f(B_t)]$. And since $\frac{\partial u}{\partial t}$ should be zero at equilibrium (and thus u is a harmonic function), the only way for u to not be constant is if we have a nonconstant boundary condition.

20 April 29, 2020

We'll start with Chapter 8 of Le Gall today, discussing stochastic differential equations, existence and uniqueness of solutions, and the case where we have Lipschitz coefficients.

Definition 174

Let $\sigma, b : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be real-valued functions of time and space that are locally bounded and measurable. A **(weak) solution** of the stochastic differential equation (SDE)

$$E(\sigma, b) : \{dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt\}$$

(this is the usual informal notation) consists of the following:

- a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ (we'll assume the filtration $\mathcal{F}_t$ is complete),
- an $\mathcal{F}_t$ Brownian motion B_t, and
- an $\mathcal{F}_t$-adapted process X_t with continuous sample paths, such that

$$X_t = X_0 + \int_0^t \sigma(s, X_s)dB_s + \int_0^t b(s, X_s)ds.$$

Let $\Phi(X)_t$ denote the right-hand side of this last equation (it implicitly depends on $\Omega, \mathcal{F}, \mathcal{F}_t$, and B). Because σ and b are locally bounded, $\int_0^t \sigma(s, X_s)dB_s$ is a local martingale $M(X)_t$ and $\int_0^t b(s, X_s)ds$ a finite variation process $A(X)_t$, so we've had practice working with these types of objects already. If $X_0 = x \in \mathbb{R}$, then we say that X is a **solution for $E_x(\sigma, b)$**.

We'll focus on the one-dimensional case, though many results generalize to the multi-dimensional case. The way to think about this SDE is that it is the system governed by the ODE $\frac{df}{dt} = b(t, f(t))dt$ but with some additional noise σdB_t. We have existence and uniqueness of solutions for the ODE under mild conditions, so we'd like to establish an analogous idea for SDEs. However, things become a bit more complicated – in particular, there are a few notions of what a "solution" means here:

Definition 175

A weak solution of $E(\sigma, b)$ (as above) is defined in Definition 174. A **strong solution** also satisfies the additional condition that X_t is adapted to the Brownian filtration $\sigma(B_s : s \leq t) \subseteq \mathcal{F}_t$.

The idea is that X_t appears on both sides of the weak solution, so we may want to solve for an X_t where all of the randomness comes from the Brownian motion randomness alone.

Definition 176

Again take the definition of $E(\sigma, b)$ from above. We have **weak uniqueness** of solutions if all solutions of $E_x(\sigma, b)$ have the same law and **pathwise uniqueness** if given $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$, any two solutions X and Y with $X_0 = Y_0$ almost surely are indistinguishable.

Assuming that we have weak **existence** (so that we do have a weak solution), pathwise uniqueness is stronger than weak uniqueness. This isn't an obvious fact – we can read the book for an example where an SDE has weak uniqueness but not pathwise uniqueness (we can construct a probability space so that they are not indistinguishable). But showing that pathwise uniqueness implies weak uniqueness is the **Yamada-Watanabe theorem** – the idea is that indistinguishability requires us to look on a single probability space. If we are given $\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B$, and a starting point x, we have a unique solution X. But then if we have a different probability space and are given $\Omega', \mathcal{F}', \mathcal{F}'_t, \mathbb{P}', B', x$, we will also have a unique solution X', and the theorem tells us that X and X' will have the same law.

We'll first note down a technical result:

Fact 177 (Gronwell's lemma)

Let g be a nonnegative bounded function on $[0,t]$, and suppose there exist $a, b \geq 0$ such that

$$g(t) \leq a + b\int_0^t g(s)ds$$

for all $t \in [0,T]$. Then $g(t) \leq ae^{bt}$ for all $t \in [0,T]$.

This is a fact about deterministic functions – it's basically a calculus fact, so we'll omit the proof. Note that if we had equality, this would be easy, since we would have $g'(t) = bg(t)$ with initial condition $g(0) = a$, and this is uniquely solved by $g(t) = ae^{bt}$.

We'll be considering a class of processes where we can prove all of the things that we want: specifically, we'll assume our coefficient functions $\sigma, b : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ are continuous (jointly as a function of space and time) and K-Lipschitz in the space coordinate, meaning that for all $x, y \in \mathbb{R}$, we have

$$|\sigma(t,x) - \sigma(t,y)| \leq K|x-y|, \quad |b(t,x) - b(t,y)| \leq K|x-y|$$

Theorem 178

If σ, b are continuous and K-Lipschitz, then for all $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$ and all $x \in \mathbb{R}$, there exists a **strong solution** X for $E_x(\sigma, b)$ on the the probability space, and we have **pathwise uniqueness** of solutions (meaning any other solution Y is indistinguishable from X, so all solutions are strong).

Proof. We first show pathwise uniqueness. We are already given $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$, and suppose that X, Y are both solutions on this space with $X_0 = Y_0$ almost surely. (For this part, the starting point does not need to be fixed.) Our goal is to show that X and Y are indistinguishable. Let τ be the stopping time

$$\tau = \inf\{t \geq 0 : |X_t - X_0| \geq M \text{ or } |Y_t - Y_0| \geq M\}.$$

Consider the function $h(t) = \mathbb{E}\left[(X_{t\wedge\tau} - Y_{t\wedge\tau})^2\right]$. Since X and Y have the same starting point and we stop before moving more than M away from the starting point, h is bounded by $(2M)^2$. Now writing X and Y as a sum of the local martingale and FV parts and using the trivial inequality $(u+v)^2 \leq 2(u^2+v^2)$, we in fact have

$$h(t) \leq 2\mathbb{E}\left[\left(\int_0^{t\wedge\tau} (\sigma(s,X_s) - \sigma(s,Y_s))dB_s\right)^2\right] + 2\left[\left(\int_0^{t\wedge\tau} (b(s,X_s) - b(s,Y_s))ds\right)^2\right].$$

The first term is the expectation of the square of a stochastic integral, and recall that we've previously proven that $\mathbb{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] \leq \mathbb{E}\left[\int_0^t H_s^2 d\langle M\rangle_s\right]$. Using this and also applying Cauchy-Schwarz on the second term tells us that

$$h(t) \leq 2\mathbb{E}\left[\int_0^{t\wedge\tau} (\sigma(s,X_s) - \sigma(s,Y_s))^2 ds\right] + 2t\mathbb{E}\left[\int_0^{t\wedge\tau} (b(s,X_s) - b(s,Y_s))^2\right].$$

Now applying the Lipschitz condition, we have

$$h(t) \leq 2K^2(1+t)\mathbb{E}\left[\int_0^{t\wedge\tau} (X_s - Y_s)^2 ds\right] \leq 2K^2(1+t)\int_0^t h(s)ds \leq 2K^2(1+T)\int_0^t h(s)ds$$

for all $t \leq T$. Since h is a bounded nonnegative function, Gronwall's lemma tells us that $h = 0$ on $[0,T]$ (because the constant term is 0), so $X_{t\wedge\tau} = Y_{t\wedge\tau}$ almost surely for all $t \in [0,T]$. Now take $M \to \infty$ and $T \to \infty$ to show

that $X_t = Y_t$ almost surely for all t. By definition this doesn't mean we have indistinguishability, but in this case the assumption of continuity does imply that X and Y are indistinguishable, proving pathwise uniqueness.

Now we'll show existence of a strong solution, and the calculation here will be fairly similar to what we've just done. Remember that we're working with a given $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$, and a weak solution is just a solution of the fixed point equation $X = \Phi(X)$. This motivates the idea of iterating Φ: let X^0 be the constant process such that $(X^0)_t = x$, and define $X^n = \Phi^n(X^0)$ for all n. Our goal is to show that X^n converges to the fixed point solution satisfying $X = \Phi(X)$, and we'll do this by bounding the difference between X^n and X^{n+1}. Define

$$G_n(t) = \mathbb{E}\left[\sup_{s\le t}\left(X_s^{n+1} - X_s^n\right)^2\right].$$

By the same calculation strategy as before and writing X^{n+1} and X^n in terms of X^n and X^{n-1}, respectively, we find that

$$G_n(t) \le 2\mathbb{E}\left[\sup_{s\le t}\left(\int_0^t \sigma(s, X_s^n) - \sigma(s, X_s^{n-1})dB_s\right)^2\right] + 2\mathbb{E}\left[\sup_{s\le t}\left(\int_0^t b(s, X_s^n) - b(s, X_s^{n-1})ds\right)^2\right]$$

Now the first term can be controlled because it is the supremum of a local martingale, so we can apply Doob's L^2 inequality and then use the same inequality as before. Then also applying Cauchy-Schwarz to the second term yields

$$G_n(t) \le 8\mathbb{E}\left[\int_0^t \left(\sigma(s, X_s^n) - \sigma(s, X_s^{n-1})\right)^2 ds\right] + 2t\mathbb{E}\left[\int_0^t (b(s, X_s^n) - b(s, X_s^{n-1})^2 ds\right].$$

Finally, applying the Lipschitz assumption we find that

$$\boxed{G_n(t)} \le 2K^2(4+t)\mathbb{E}\left[\int_0^t (X_s^n - X_s^{n-1})^2 ds\right] \le \boxed{2K^2(4+T)\int_0^t G_{n-1}(s)ds}.$$

This means we have a bound for g_n in terms of g_{n-1}, and now we'll just do the base case

$$G_1(t) = \mathbb{E}\left[\sup_{s\le t}(X_s^1 - x)^2\right],$$

where $X_s^1 = \Phi(X^0)_s = x + \int_0^s \sigma(r, x)dB_r + \int_0^s b(r, x)dr$. Because σ, b are both locally bounded, $G_1(t)$ is bounded by some constant $c(T)$ for all $t \in [0, T]$ (notice that we're bounding the **expectation** of the second moment of $(X_s^1 - x)$, not the function itself). Then inductively integrating the boxed bound above yields

$$G_{n+1}(t) \le c(T)(2K^2(4+T))^n \frac{t^n}{n!}.$$

But this decays quickly because of the $n!$ in the denominator, which means that when we sum the sup-norm differences, we get

$$\mathbb{E}\left[\sum_n \sup_{t\le T}\left|X_t^{n+1} - X_t^n\right|\right] \le \sum_n \sqrt{G_n(T)} < \infty.$$

Thus, $\sum_n \sup_{t\le T}\left|X_t^{n+1} - X_t^n\right|$ is almost surely finite for any T, which means that X^n converges uniformly to X on $[0, T]$ (and in general any compact time interval). We know that X^1 is adapted to the Brownian filtration, and in general integrating against a Brownian motion still keeps things adapted, so X^n is adapted to the Brownian filtration for all n and thus the limit process X is adapted as well. So we just need to check that this is actually a (weak) solution to $X = \Phi(X)$, but we showed that $X^n = \Phi(X^{n-1})$ so taking $n \to \infty$ makes both X^n and X^{n-1} in the above equation converge to X. Thus it remains to check that

$$\Phi(X^n) = x + \int_0^t \sigma(s, X_s^{n-1})dB_s + \int_0^t b(s, X_s^{n-1})ds$$

converges to $\Phi(X)$. It's clear that the second term converges to $\int_0^t b(s, X_s)ds$ because b is continuous and X_s^{n-1} converges uniformly to X^n (so we can apply dominated convergence theorem). To argue that the first term converges correctly, we use the dominated convergence theorem on the difference $\int_0^t(\sigma(s, X_s) - \sigma(s, X_s^{n-1}))dB_s$ for stochastic integrals, where the dominating process is $D_s = K\left[\sum_n \sup_{r\le s} |X_r^n - X_r^{n-1}|\right]$. This finishes the verification and constructs our strong solution X. □

For the next result, recall that the Wiener measure is the law of Brownian motion started from 0 – in particular, it is a measure on continuous functions $C([0,\infty), \mathbb{R})$.

Theorem 179

Consider the space of functions $C([0,\infty), \mathbb{R})$. Let W be the Wiener measure, $\mathcal{B}_C$ the Borel sigma-algebra on $C([0,\infty), \mathbb{R})$, and $\mathcal{G}$ the sigma-algebra $\sigma(\mathcal{B}_c, N)$, where N is the set of W-negligible sets. If σ, b are continuous and K-Lipschitz, then for all $x \in \mathbb{R}$ there exists a measurable function $F_x : (C([0,\infty), \mathbb{R}), \mathcal{G}) \to (C([0,\infty), \mathbb{R}), \mathcal{B}_C)$ such that the following hold:

- for all t, $F_x(w)_t$ coincides almost surely with a measurable function of $(w(s) : s \le t)$,
- for all w, the map $x \to F_x(w)$ is continuous as a map $\mathbb{R} \to C([0,\infty), \mathbb{R})$.
- for all choices of $\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B$ and for all x, $F_x(B)$ is the unique solution of $E_x(\sigma, b)$, and this is also true if we replace x with a random starting point $U \in \mathcal{F}_0$.

The second point here tells us that if we start from two points $x, y \in \mathbb{R}$ that are close to each other, and we evolve the SDE using the same Brownian motion, our paths will look similar when σ, b are bounded and Lipschitz. What we'll do is apply Theorem 178 with the filtered probability space

$$(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B) = (C([0,\infty], \mathbb{R}), \mathcal{G}, \mathcal{G}_t = \sigma(w(s) : s \le t, N), W, w),$$

where w is the canonical Brownian motion. Then the solution we get out of the theorem will be $F_x(w)$ – this is adapted to the Brownian motion, so we automatically satisfy the first points. So we just need to show that the mapping is continuous and that this works on any probability space, and we'll do this next time.

21 May 4, 2020

Recall that we've been looking at the stochastic differential equation $dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt$, where σ and b are K-Lipschitz in the x coordinate. Last time, we showed that given any $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$, we can find a strong solution X_t^x started at x and adapted to the Brownian motion, and this solution is unique by **pathwise** uniqueness. (The Yamada-Watanabe theorem, which states that pathwise uniqueness implies weak uniqueness, can be applied here, but we can also prove weak uniqueness directly in this K-Lipschitz case.) We'll now start with a proof of last time's result, which stated that we have a measurable mapping $F_x : (C([0,\infty), \mathbb{R}), \mathcal{G}) \to (C([0,\infty), \mathbb{R}), \mathcal{B})$ such that $F_x(w)_t$ is measurable of $(w_s)_{s\le t}$, the map $x \to F_x(w)$ is continuous in x, and $F_x(B)$ solves the stochastic differential equation $E_x(\sigma, b)$.

Proof of Theorem 179. We already showed there exists a strong solution X_t^x for any $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B, x)$, and we'll apply that here to the space$(C([0,\infty), \mathbb{R}), \mathcal{G}, \mathcal{G}_t, W, w, x)$. We want to show continuity, and we'll make use of the **Kolmogorov continuity lemma** – recall that if we have any stochastic process F_t which takes values in a complete separable metric space (S, d) with the bound $\boxed{\mathbb{E}\left[d(F_s, F_t)^q\right] \le C|s-t|^{1+\varepsilon}}$, then there is a modification $\tilde{F}_t$ of F_t that

is α-Hölder continuous for all $\alpha \in \left(0, \frac{\varepsilon}{q}\right)$. We applied this to Brownian motion earlier in the class, and remember that one part of the proof was to do a union bound (for the interval $[0,1]$)

$$\mathbb{P}\left(|X_{i/2^n} - X_{(i-1)/2^n}| \leq \left(\frac{1}{2^n}\right)^\alpha \text{ for all } i\right) \leq 2^{n(1+\alpha q)}/2^n(1+\varepsilon).$$

In particular, this does not rely on having independent increments! So we'll apply the Kolmogorov continuity lemma to $F_x(w)$ (where we index by **position** $x \in \mathbb{R}$ instead of by time), thinking of this process as taking values in the metric space (S, d) given by

$$S = C([0,\infty), \mathbb{R}), \quad d(f,g) = \sum_{n \geq 1} \frac{1}{2^n} \min\left(1, \sup_{t \leq n} |f(t) - g(t)|\right).$$

Checking the assumptions of the continuity lemma requires the boxed estimate above, but we can read that on our own, and once we verify this we do indeed get continuity. The last thing we must check is that $F_x(B)$ solves our SDE given any $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$; we know it's true on the Wiener space, but we need to check that it's true for any Brownian motion B. If we let w be our Brownian motion on the Wiener space, then $F_x(w)$ solves $E_x(\sigma, b)$, meaning

$$F_x(w)_t - \left(x + \int_0^t \sigma(s, w(s)) dw(s) + \int_0^t b(s, w(s)) ds\right) = 0.$$

where w is our Brownian motion. Letting $\Psi(w)_t$ denote the left-hand side, we have $\int_w |\Psi(w)| dW(w) = 0$. But then any Brownian motion B has the same law as W, so we must also have $\int |\Psi(B)| d\mathbb{P}(B) = 0$, which implies that $F_x(B)$ is a valid solution to our SDE. □

Remark 180. *As a technical sidenote, we do need to make sure Ψ is measurable as a function of w – the main difficulty here is showing that the stochastic integral $\int_0^t \sigma(s, w(s)) dw(s)$ is a measurable function of w, but this follows from the approximation*

$$\int_0^t \sigma(s, w(s)) dw(s) = \lim_{n \to \infty} \sum_{i=1}^{2^n} \sigma\left(\frac{(i-1)t}{2^n}, w\left(\frac{(i-1)t}{2^n}\right)\right)\left(w\left(\frac{it}{2^n} - w\left(\frac{(i-1)t}{2^n}\right)\right)\right).$$

Note that this theorem we've just proved implies weak uniqueness without needing the Yamada-Watanabe theorem – the solution comes from applying the same map F_x to our Brownian motion, no matter what probability space we're on, so the law of X^x is determined: for any event A we have

$$\mathbb{P}(X^x \in A) = \mathbb{P}(F_x(B) \in A) = \mathbb{P}(B \in (F_x)^{-1} A),$$

and then because the law of Brownian motion is given by the Wiener measure, this is just $W((F_x)^{-1}A)$. So the law is just $(F_x)_\# W$.

Remark 181. *As stated last lecture, if we replace our starting point $X_0 = X$ with a random variable $X_0 = U \in \mathcal{F}_0$, we can still get a solution to our SDE with $F_U(B)$. But we can read the book for more details.*

We will now make a connection to Markov processes: suppose for this part of the lecture that σ and b **don't depend on time** and that they are still K-Lipschitz in x.

Theorem 182

Suppose $\sigma(t,x) = \sigma(x)$ and $b(t,x) = b(x)$ are K-Lipschitz, and let X be a solution of the SDE $E(\sigma, b)$ on any probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. Then X is a Markov process, and the semigroup can be described via

$$Q_t f(x) = \mathbb{E}_x[f(X_t)] = \int f(F_x(w)_t) dW(w).$$

(Remember that we showed the existence and uniqueness of a strong solution X already.)

Proof. We'll first show that $\mathbb{E}[f(X_{s+t})|\mathcal{F}_s] = Q_t f(X_s)$, where f must be a bounded and measurable function. Define the shifted process $\hat{X}_t = X_{s+t}$, and write it out as

$$\hat{X}_t = X_s + \int_s^{s+t} \sigma(X_r) dB_r + \int_s^{s+t} b(X_r) dr = \hat{X}_0 + \int_0^t \sigma(\hat{X}_r) d\hat{B}_r + \int_0^t b(\hat{X}_r) dr.$$

This means that $\hat{X}_t$ solves the SDE $E_{\hat{X}_0}(\sigma, b)$ on the probability space $(\Omega, \mathcal{F}, \hat{\mathcal{F}}_t = \mathcal{F}_{s+t}, \mathbb{P})$ with the shifted Brownian motion $\hat{B}_t = B_{s+t} - B_s$. Therefore, Theorem 179 tells us that $F_{X_s}(\hat{B})$ must solve the SDE, meaning we have

$$\hat{X}_t = F_{X_s}(\hat{B}) \implies \mathbb{E}\left[f(X_{s+t})|\mathcal{F}_s\right] = \mathbb{E}\left[f(F_{X_s}(\hat{B})_t)|\mathcal{F}_s\right].$$

And now $\hat{B}$ on the right-hand side is independent of $\mathcal{F}_s$, so we can write this out as

$$\hat{X}_t = \int f(F_{X_s}(w)_t) dW(w) = Q_t f(X_s),$$

as desired. To finish, we need to show that Q_t is a valid semigroup, which means that (1) $Q_0(x, \cdot)$ is the Dirac measure δx concentrated at x, (2) the Chapman-Kolmogorov equations $Q_{s+t} = Q_s Q_t$ are satisfied, and (3) the map $(t, x) \to Q_t(x, A)$ must be measurable. The last point follows from Q_t being continuous in t and x, and everything else is straightforward. □

Theorem 183

In the same setting as the theorem above, Q_t is a Feller semigroup. The space of functions $A = C^{2,\text{cpt}}(\mathbb{R})$ that are compactly supported and twice differentiable is contained in the domain $D(L)$, and for any $f \in A$ we have

$$Lf(x) = b(x)f'(x) + \sigma^2 \frac{f''(x)}{2}.$$

Proof sketch. We'll omit the proof of the Feller property, which is showing that (1) whenever $f \in C_0(\mathbb{R})$, we have $Q_t f \in C_0(\mathbb{R})$, and (2) $Q_t f \to f$ converges in the sup-norm as $t \downarrow 0$. For the remaining claims, we apply Itô's formula to $f \in A$ and then plug in the SDE to find

$$\begin{aligned} df(X_t) &= f'(X_t) dX_t + \frac{f''(X_t)}{2} \sigma(X_t)^2 dt \\ &= f'(X_t)\left(\sigma(X_t) dB_t + b(X_t) dt\right) + \frac{f''(X_t)}{2} \sigma(X_t)^2 dt. \end{aligned}$$

Subtracting off the drift term, we can define

$$\boxed{M_t = f(X_t) - f(x) - \int_0^t \left(f'(X_s) b(X_s) + \frac{f''(X_s)\sigma(X_s)^2}{2}\right) ds},$$

so that $M_t = \int_0^t \sigma(X_s) dB_s$ is a local martingale. Call the integrand in the boxed expression $Gf(X_s)$ – we want to show that G is the generator we're looking for. For simplicity, let's assume σ, b are bounded, so that M is a true martingale. We then have

$$0 = \mathbb{E}[M_t] = \mathbb{E}_x[f(X_t)] - f(x) - \int_0^t \mathbb{E}_x[Gf(X_s)]\, ds.$$

Dividing through by t and taking the limit $t \downarrow 0$, we find that

$$\begin{aligned} Lf(x) = \lim_{t\downarrow 0} \frac{Q_t f(x) - f(x)}{t} &= \lim_{t\downarrow 0} \frac{1}{t} \int_0^t \mathbb{E}_x[Gf(X_s)] ds \\ &= \lim_{t\downarrow 0} \frac{1}{t} \int_0^t Q_s Gf(x) ds. \end{aligned}$$

But we know that $Q_s(Gf)$ converges to Gf by the Feller property as $s \downarrow 0$, so this last expression is indeed $Gf(x)$, showing that G is our generator. □

In short, we can phrase the results above as saying that **time-independent SDEs with Lipschitz coefficients correspond to Feller processes**.

Example 184 (Ornstein-Uhlenbeck process)
We'll apply our theory to the SDE $dX_t = dB_t - \lambda X_t dt$.

Define the process $M_t = e^{\lambda t} X_t$ – by Itô's formula, we have

$$d(e^{\lambda t} X_t) = e^{\lambda t}(dB_t - \lambda X_t dt) + \lambda e^{\lambda t} X_t dt.$$

Since the drift terms cancel out, we have $dM_t = e^{\lambda t} dB_t$, meaning that M_t is a local martingale with increments given by integrating a deterministic function $e^{\lambda t}$ against a Brownian motion. So if we take M_0 to be some integrable function, M_t will be a true martingale, and

$$e^{\lambda t} X_t - X_0 = M_t - M_0 = \int_0^t e^{\lambda s} dB_s,$$

and rearranging yields our solution $\boxed{X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dB_s}$. In particular, if X_0 is deterministic or normal and independent of B, then X_t is actually a **Gaussian process**, where we can calculate

$$\mathrm{Var}(X_t) = \frac{\mathrm{Var}(X_0)}{e^{2\lambda t}} + \int_0^t \frac{ds}{e^{2\lambda(t-s)}} = \frac{\mathrm{Var}(X_0)}{e^{2\lambda t}} + \frac{1}{2\lambda}\left(1 - \frac{1}{e^{2\lambda t}}\right).$$

If we now choose X_0 so that $\mathrm{Var}(X_0) = \frac{1}{2\lambda}$, then we have $\mathrm{Var}(X_t) = \frac{1}{2\lambda}$ for all t. We can also check that $\mathrm{Cov}(X_s, X_t) = \frac{1}{2\lambda \exp(\lambda|t-s|)}$, which only depends on the difference between s and t. This means the centered Gaussian process X is also **stationary**, and such a process is called the **Ornstein-Uhlenbeck process**.

Example 185 (Geometric Brownian motion)
Next, consider the SDE $dX_t = \sigma X_t dB_t + r X_t dt$.

This is the crudest possible model we could have for the stock market (we have some rate of appreciation, as well as some volatility). The idea here is to apply Itô's formula to $\log X_t$ (as long as X stays positive), which yields

$$d(\log X_t) = \frac{1}{X_t}(\sigma X_t dB_t + r X_t dt) - \frac{1}{2X_t^2}\sigma^2 X_t^2 dt.$$

Again the drift terms cancel, so we end up with the equation $d(\log X_t) = \sigma dB_t + \left(r - \frac{\sigma^2}{2}\right) dt$. Since there is no X-dependence on the right side, we can now integrate to get our solution $\boxed{X_t = X_0 \exp\left(\sigma B_t - \left(r - \frac{\sigma^2}{2}\right) t\right)}$.

22 May 11, 2020

Today, we'll discuss a few results that are applications of what we've learned in this class, centered around the **Dyson Brownian motion in random matrices**.

Definition 186

Let $(B_{ij})_{i \leq j}$ be a family of iid Brownian motions. Then the **symmetric matrix Brownian motion** is the symmetric matrix H such that

$$H(t)_{ij} = \begin{cases} \sqrt{2} B_{ij} & i = j, \\ B_{ij} & i < j, \\ B_{ji} & j < i. \end{cases}$$

In other words, all entries evolve according to independent Brownian motions, except that we want the matrix to be symmetric. We include the $\sqrt{2}$ factor here because another way that we can obtain this matrix is via

$$H(t) = \frac{X(t) + X(t)^*}{\sqrt{2}},$$

where X is a standard Brownian motion in $\mathbb{R}^{n \times n}$ (meaning **all** entries are independent) which we symmetrize and rescale. So all off-diagonal entries evolve like standard Brownian motions, but the diagonal terms will have a larger variance.

Definition 187

The **Hermitian matrix Brownian motion** is similarly defined as

$$H(t) = \frac{X(t) + X(t)^*}{\sqrt{2}},$$

where X is a standard Brownian motion in $\mathbb{C}^{n \times n}$.

We say that $\frac{H(t)}{\sqrt{t}}$ for a symmetric BM H is a sample from the **Gaussian orthogonal ensemble** or **GOE**, and similarly $\frac{H(t)}{\sqrt{t}}$ for a Hermitian BM H is a sample from the **Gaussian unitary ensemble** or **GUE**.

We can show that H sampled from either GOE or GUE will always have n distinct eigenvalues almost surely, which are real because we have a Hermitian matrix – we'll order them as $\lambda_1 < \cdots < \lambda_n$. In fact, the ordered eigenvalue process $\lambda(t) = (\lambda_1(t), \cdots, \lambda_n(t))$ is such that the eigenvalues never collide almost surely, so this eigenvalue process does not leave the **Weyl chamber**

$$W_n = \{z \in \mathbb{R}^n : z_1 < \cdots < z_n\}.$$

There are two main results we'll be covering today:

Theorem 188

If H is a symmetric or Hermitian matrix Brownian motion, then the eigenvalue process solves the **β-Dyson SDE**

$$d\lambda_i(t) = \beta \sum_{\ell \neq i} \frac{1}{\lambda_i(t) - \lambda_\ell(t)} dt + \sqrt{2} dB_i(t),$$

where B_i are independent Brownian motions, where we have $\beta = 1$ in the symmetric case and $\beta = 2$ in the Hermitian case.

Notice that all eigenvalues $\lambda_\ell < \lambda_i$ give a positive contribution to the drift term, and all eigenvalues $\lambda_\ell > \lambda_i$ give a negative contribution. So the eigenvalues will "repel" each other more strongly as the eigenvalues grow closer, and this is related to why the eigenvalues are distinct almost surely for the GOE and GUE.

We won't focus too much on the proof of existence and uniqueness for now – it turns out that for any $x \in W_n$ in the Weyl chamber, there is a unique strong solution of the β-Dyson SDE started from x for all $\beta \geq 1$ which stays inside the chamber for all time. After that, we'll also cover the following result:

Theorem 189

For any $x \in W_n$, the $\beta = 2$ Dyson process started from x is equidistributed as an n-dimensional Brownian motion started from x, conditioned to stay inside the Weyl chamber.

Note that n-dimensional Brownian motion will almost surely exit the Weyl chamber, because even two Brownian motions will intersect almost surely. So we'll need to be more precise about this statement to work with it.

We'll first discuss the main ideas of the first result – most of the work is calculation:

Proof sketch of Theorem 188. We'll just do the symmetric case – the method of proof is the same for the Hermitian case. Call the entries of our symmetric matrix H_{jk}. We'll calculate the first and second derivative of the ith eigenvalue with respect to each matrix entry H_{jk} (for all $j \leq k$) and then apply Itô's formula, but we must break into separate cases because the Brownian motions are different on the diagonals. We'll find that

$$d\lambda_i(t) = \left[\sum_{k=1}^{n} \frac{\partial \lambda_i}{\partial H_{kk}} \sqrt{2} dB_{kk} + \sum_{j<k} \frac{\partial \lambda_i}{\partial H_{jk}} dB_{jk}(t) \right] + \left[\frac{1}{2} \sum_{k=1}^{n} 2 \frac{\partial^2 \lambda_i}{\partial H_{kk}^2} + \frac{1}{2} \sum_{j<k} \frac{\partial^2 \lambda_i}{\partial H_{kk}^2} \right] dt.$$

The first bracketed term is the continuous local martingale term, and if we set it equal to $\sqrt{2} dB_i(t)$, the Lévy characterization verifies that B_i is indeed a Brownian motion. Similarly, evaluating the partial derivatives on the second bracketed term will show that it matches up with the finite variation term in the theorem statement. So it just remains to explain how we actually calculate the needed derivatives: we write

$$\frac{\partial \lambda_i}{\partial H_{jk}} = \frac{d}{dt} \lambda_i(H(t)),$$

where $H(t) = H + t(E_{jk} + E_{kj})$ and E_{jk} and E_{kj} are the "matrix units" which are 0 everywhere except with a 1 in the (j,k) and (k,j) entries, respectively. This can then be computed using implicit differentiation, using the fact that $Hu_i = \lambda u_i$. □

We'll need a technical lemma for the second result:

Lemma 190 (Andréif integration formula)

For "nice" functions f_i and g_i, we have

$$\frac{1}{n!}\int_{\mathbb{R}^n} \det\begin{bmatrix} f_1(x_1) & \cdots & f_1(x_n) \\ \vdots & \ddots & \vdots \\ f_n(x_1) & \cdots & f_n(x_n) \end{bmatrix}\begin{bmatrix} g_1(x_1) & \cdots & g_1(x_n) \\ \vdots & \ddots & \vdots \\ g_n(x_1) & \cdots & g_n(x_n) \end{bmatrix} dx = \det\begin{bmatrix} \int_{\mathbb{R}} f_1 g_1 & \cdots & \int_{\mathbb{R}} f_1 g_n \\ \vdots & \ddots & \vdots \\ \int_{\mathbb{R}} f_n g_1 & \cdots & \int_{\mathbb{R}} f_n g_n \end{bmatrix}$$

Proof. Writing out the definition of the determinant on the left-hand side, we have

$$\frac{1}{n!}\int_{\mathbb{R}^n}\left(\sum_\sigma \operatorname{sgn}(\sigma)\prod_{i=1}^n f_{\sigma(i)}(x_i)\right)\left(\sum_\tau \operatorname{sgn}(\tau)\prod_{j=1}^n g_{\tau(j)}(x_j)\right),$$

which can also be rewritten as

$$= \frac{1}{n!}\sum_{\sigma,\tau}\operatorname{sgn}(\sigma\tau)\prod_{i=1}^n\left(\int f_{\sigma(i)}(x_i)g_{\tau(i)}(x_i)dx_i\right)$$

(where we've used that the integral over the product is the product over the integral of the individual independent x_is). But now summing over all permutations $\rho = \sigma\tau$ will count each one $n!$ times, so those factors cancel out and we get the formula of the determinant on the right side. □

Lemma 191 (Karlin-McGregor formula)

Let $B(t)$ be a Brownian motion in $\mathbb{R}^n$, and let T be the first exit time of B from the Weyl chamber. Then for any $x \in W_n$ and measurable subset $A \subseteq W_n$,

$$\mathbb{P}_x(B(t)\in A; T > t) = \int_A \det\begin{bmatrix} p_t(x_1,y_1) & \cdots & p_t(x_1,y_n) \\ \vdots & \ddots & \vdots \\ p_t(x_n,y_1) & \cdots & p_t(x_n,y_n) \end{bmatrix} dy,$$

where p_t is the transition density of a usual one-dimensional Brownian motion.

If the BMs were all independent and we didn't care whether they collided or not, the transition kernel would just be $p_t(x_1,y_1)\cdots p_t(x_n,y_n)$. So this result is saying that we need to use the determinant of p_ts instead if we require our BMs not to collide. (And here, we're **restricting** on the left side to the event that we haven't left by time t, not conditioning.)

Proof. Let T_i be the collision time $\inf\{t : B_i(t) = B_{i+1}(t)\}$, so that $T = \min T_i$. Again expanding out the determinant and writing the transition p_t functions in terms of a Brownian motion yields (letting $q_t(x,y)$ be the determinant)

$$q_t(x,y)dy = \sum_\sigma \operatorname{sgn}(\sigma)\mathbb{E}_x\left(\prod_{i=1}^n 1\{B_i(t)\in dy_{\sigma(i)}\}\left[1\{T\geq t\} + \sum_{j=1}^{n-1} 1\{T = T_j < t\}\right]\right),$$

where bracketed term we've inserted is just 1. But when we look at the contribution from the $1\{T \geq t\}$ term (meaning we haven't left the chamber) the only permutation that is relevant is that where the y_is haven't gone out of order from the x_is, so this contributes $\boxed{\mathbb{E}_x\left[\prod_{i=1}^n 1\{B_i(t)\in dy_i\}; T\geq t\right]}$. And we can make a swapping argument to show that the total contribution from the other part is zero (if T_j happens before t, we can take the two Brownian motions

that cross and switch them, changing the sign of the difference). So the boxed term is the only contribution from the integral overall, and integrating it over A yields the left-hand side as desired. □

In order to condition on "not colliding," we'll construct a martingale:

Corollary 192

For any $x \in \mathbb{R}^n$, define the classic **Vandermonde determinant**

$$v(x) = \prod_{i<j}(x_j - x_i) = \det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}.$$

Then if B is a Brownian motion in $\mathbb{R}^n$ and T is the exit time from the Weyl chamber, then $M_t = v(B_{t\wedge T})$ is a nonnegative martingale.

Proof sketch. M_t is indeed nonnegative because each term $(x_j - x_i)$ in the product is nonnegative while we're in the Weyl chamber. Consider the expectation $\mathbb{E}_x[v(B_{t\wedge T})]$ – if $T \le t$, then $M_t = 0$ (because v is zero when it hits the boundary of the chamber) and there is no contribution to the expectation. So taking $(y_1, \cdots, y_n)$ to be the location of the Brownian motion at time t, we can just calculate

$$\mathbb{E}_x[v(B_{t\wedge t}); T > t] = \int_{W_n} \det \begin{bmatrix} 1 & y_1 & \cdots & y_1^{n-1} \\ 1 & y_2 & \cdots & y_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_n & \cdots & y_n^{n-1} \end{bmatrix} \det \begin{bmatrix} p_t(x_1, y_1) & \cdots & p_t(x_1, y_n) \\ \vdots & \ddots & \vdots \\ p_t(x_n, y_1) & \cdots & p_t(x_n, y_n) \end{bmatrix} dy$$

where we've used the the Karlin-MacGregor formula. But symmetry in the variables here means we can integrate over all of $\mathbb{R}^n$ instead of W_n by adding a factor of $\frac{1}{n!}$, and now the Andréif integration formula yields

$$\mathbb{E}_x[v(B_{t\wedge t})] = \det \begin{bmatrix} 1 & x_1 & \cdots & \mathbb{E}[(x_1 + \sqrt{t}Z)^{n-1}] \\ 1 & x_2 & \cdots & \mathbb{E}[(x_2 + \sqrt{t}Z)^{n-1}] \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \cdots & \mathbb{E}[(x_n + \sqrt{t}Z)^{n-1}] \end{bmatrix}$$

(because by definition of the transition kernel, integrating y_1 against $p_t(x_1, y_1)$ will just yield x_1). And now this is just equal to the simpler Vandermonde matrix

$$\det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} = v(x),$$

because we can expand out the rightmost columns by the binomial theorem and use row operations to subtract off lower powers. And this is basically the martingale identity we want once we use the Markov property. □

Proposition 193

Let $B(t)$ be a Brownian motion started from $x \in W_n$. Then there exists a unique measure $\mathbb{Q}$ such that for all stopping times $S < \infty$,

$$\frac{d\mathbb{Q}|_{\mathcal{F}_S}}{d\mathbb{P}|_{\mathcal{F}_S}} = \frac{M_S}{M_0} = \frac{v(B_{S\wedge T})}{v(x)}.$$

Then the law of B under $\mathbb{Q}$ can be thought of as Brownian motion conditioned not to exit W_n – it is a Feller process with generator given by $Lf(x) = \langle \nabla \log v(x), \nabla f(x)\rangle + \frac{1}{2}\Delta f(x)$.

Proof. Let R_i be the first time that $M_t \geq i$. Then M^{R_i} is a bounded martingale, so the optional stopping theorem tells us that the exit time satisfies $\mathbb{P}(T > R_i) = \frac{v(x)}{i}$. Define the measure

$$\left.\frac{d\mathbb{Q}_i}{d\mathbb{P}_i}\right|_{\mathcal{F}_{R_i}} = \lim_{t\to\infty} \frac{M_{t\wedge R_i}}{v(x)} = \frac{i1\{T > R_i\}}{v(x)}$$

(notice that this is zero if we exit the chamber before hitting i). Because M is a martingale, the $\mathbb{Q}_i$ are consistent with each other, so there is a measure $\mathbb{Q}$ whose restrictions to $\mathcal{F}_{R_i}$ are consistent with the Q_is, and it will have the correct values of $\frac{d\mathbb{Q}|_{\mathcal{F}_S}}{d\mathbb{P}|_{\mathcal{F}_S}}$. But now we have

$$\mathbb{P}(A|T > R_i) = \frac{\mathbb{P}(A \cdot 1\{T > R_i\})}{v(x)/i} = \mathbb{E}_{\mathbb{P}}\left(1_A \frac{i1\{T > R_i\}}{v(x)}\right) = \mathbb{Q}(A)$$

by definition of the Radon–Nikodym derivative. So as we take $i \to \infty$, we can think of this as conditioning on the Brownian motion never exiting the chamber (since it takes arbitrarily long time to travel arbitrarily long distances). We can now find the generator by noting that

$$Lf(x) = \lim_{t\downarrow 0} \frac{1}{t}\left[\mathbb{E}_{\mathbb{Q}}\left[f(B(t))|B(0) = x\right] - f(x)\right].$$

Applying the change of measure, this can be written in terms of $\mathbb{P}$ as

$$\lim_{t\downarrow 0} \frac{1}{t}\left[\mathbb{E}_{\mathbb{P}}\left(f(B(t))\frac{v(B_{t\wedge T})}{v(x)}\middle| B(0) = x\right) - f(x)\right].$$

Now $v(x)$ is a constant, and Itô's formula tells us that

$$d(f(B(t))M(t)) = f(B(t))dM_t + f'(B(t))M_t dB_t + \frac{1}{2}M_t f''(B(t))dt + f'(B(t))d\langle B, M\rangle_t.$$

We can ignore the martingale term because we're taking expectations; M_t has mean $v(x)$, and $d\langle B, M\rangle = \nabla v$. So plugging this back in yields

$$Lf(x) = \frac{1}{2}\Delta f(x) + \left\langle \nabla f(x), \frac{\nabla v(x)}{v(x)}\right\rangle,$$

and we're done because $\frac{\nabla v(x)}{v(x)} = \nabla \log v(x)$. □

Proof of Theorem 189. Recall that the $\beta = 2$ Dyson SDE is

$$d\lambda_i(t) = 2\sum_{\ell\neq i} \frac{1}{\lambda_i(t) - \lambda_\ell(t)}dt + \sqrt{2}dB_i(t).$$

Apply a rescaling $\theta = \frac{\lambda}{\sqrt{2}}$, so that we have the equation

$$d\theta_i(t) = \sum_{\ell\neq i} \frac{dt}{\theta_i - \theta_\ell} + dB_i(t).$$

This SDE gives us a generator

$$Lf = \sum_i \sum_{\ell \neq i} \frac{1}{x_i - x_\ell} \frac{\partial f}{\partial x_i} + \frac{1}{2} \Delta f,$$

and we can check this is indeed $\langle \nabla \log v(x), \nabla f(x) \rangle + \frac{1}{2}\Delta f$, which is the generator for the non-colliding Brownian motion. Since the generators are the same, the two processes are equal in law, as desired. □

As a final note, the reason we care about this identification is that when we consider some symmetric random matrix X, we may want it to have spectral statistics like those of the GOE (this has to do with **universality theorems** for random matrices). To show this, we construct a flow on the matrices, where $H(0) = X$ and $H(t)$ evolves via an Ornstein-Uhlenbeck process (it's like a Brownian motion, but we want it to stay stationary). Then $H(\infty)$ looks like the stationary distribution for Ornstein-Uhlenbeck, which is the GOE, and we can bound

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(H)]| \leq |\mathbb{E}[f(H(0))] - \mathbb{E}[f(H(t))]| + |\mathbb{E}[f(H(t))] - \mathbb{E}[f(H(\infty))]|$$

for a GOE matrix H. If we take t very small, the first term is small by perturbation theory, but the surprising fact is that the Dyson process mixes very quickly, so we can also control the second term (this is called the **fast mixing of the Dyson process**). And many references for further study can be found on the official course website!

References

[1] N. Sun. *18.676. Stochastic Calculus.* URL: https://math.mit.edu/~nsun/s20-18676.html.

[2] J.F.L. Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Graduate Texts in Mathematics. Springer International Publishing, 2013.

[3] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 2019.

[4] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer New York, 2002.