Lecture 5: Stochastic Processes (II)

Readings

Recommended:
- Pavliotis [2014] sections 1.3-1.5
- Pavliotis [2014] sections 2.1-2.3, 2.5

Optional:
- Grimmett and Stirzaker [2001] 8.5, 8.6, 9.6, 13.1-13.3
- Koralov and Sinai [2010] Ch. 18, 19.1-19.3
- Karatzas and Shreve [1991], 2.9 (and other bits of Chapter 2), for detailed results about Brownian motion

5.1 Karhunen-Loeve decomposition

In this section we’ll consider a general spectral representation that works even for non-stationary processes.

Consider a stochastic process $X(t)$ that has mean zero, and is continuous in the $L^2$-sense:

$$\mathbb{E}X_t^2 < \infty, \quad \mathbb{E}X_t = 0, \quad \lim_{h \to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0.$$ 

Its covariance function $B(s,t) = \mathbb{E}X_sX_t$ is symmetric, positive-definite, and continuous. Suppose the process is defined on a bounded, measurable set $D \subset \mathbb{R}$, e.g. $D = [0,1], [a,b]$, etc.

Let’s define an integral operator $K : L^2(D) \to L^2(D)$ as

$$(Kf)(s) \equiv \int_D B(s,t)f(t)dt \quad (1)$$

It can be shown that this operator is:

(i) positive semi-definite: $\langle Kf, f \rangle \geq 0$ for all $f$, where $\langle u, v \rangle = \int_D uv \, dt$ is the $L^2(D)$ inner product

(ii) self-adjoint: $\langle Kf, g \rangle = \langle f, Kg \rangle$

(iii) compact: if $\{f_n\}_{n=1}^\infty$ is a bounded sequence in $L^2(D)$, then $\{Kf_n\}_{n=1}^\infty$ has a convergent subsequence.

Remark. For a finite-dimensional analogy, consider the matrix $K = (B(t_i,t_j))_{i,j=1}^n$ and let $u = (u(t_i))_{i=1}^n$ be a vector. Then $Ku$ is just a matrix (which is symmetric and nonnegative) times a vector.

Remark. By (i), (ii) it follows that all its eigenvalues are real and nonnegative. From (iii), the spectral theorem for compact operators implies that $K$ has a countable sequence of eigenvalues tending to 0.

By Mercer’s theorem, $K$ has an orthonormal basis of eigenfunctions $\{e_i\}_{i=1}^\infty$ and its eigenvalues $\{\lambda_i\}_{i=1}^\infty$ are real and nonnegative, with 0 as the only accumulation point. Furthermore, we can write the kernel as

$$B(s,t) = \sum_{i=1}^\infty \lambda_i e_i(s)e_i(t), \quad (2)$$
where the convergence of the sum is absolute and uniform on $D \times D$, and we can write every function $f \in L^2(D)$ as
\[ f = \sum_n f_n e_n(t), \]
where the series converges in $L^2$. Our goal is to write a stochastic process also in this form; basically the only difference will be that the coefficients $f_n$ will be random variables.

**Theorem** (Karhunen-Loeve). Let $X_t$ be a mean-zero stochastic process defined on $t \in D = [a, b]$ with $X_t \in L^2(\Omega \times D)$, i.e. $\int_D E X_t^2 < \infty$. Suppose it has continuous covariance function $B(s,t)$, whose corresponding integral operator has eigenvalues $\{\lambda_i\}_{i=1}^\infty$ and orthonormal eigenfunctions $\{e_i\}_{i=1}^\infty$. Then
\[ X_t = \sum_{i=1}^\infty \xi_i e_i(t), \quad \text{with} \quad \xi_i = \int_D X_t e_i(t) \, dt. \] (3)

The above series converges in mean-square ($L^2(\Omega)$) and uniformly in $t$, i.e.
\[ \lim_{n \to \infty} \sup_{t \in D} E (X_t - \sum_{i=1}^n \xi_i e_i(t))^2 \to 0. \]

The coefficients are random variables with the following properties:

(i) zero-mean: $E\xi_i = 0$

(ii) uncorrelated: $E\xi_i \xi_j = \delta_{ij}\lambda_j$

**Proof.** (Mostly from Pavliotis [2014], p.20) First, suppose that $X_t$ has representation (3). Let’s check the properties of the coefficients. We have
\[ E\xi_i = \int_D (EX_t) e_i(t) \, dt = 0 \]
which shows property (i). We used Fubini’s Theorem to interchange the expectation and the integral, which we can do because the assumption that $X_t \in L^2(\Omega \times D)$ implies that $X_t \in L^1(\Omega \times D)$. To show property (ii), calculate:
\[ E\xi_i \xi_j = \int_{D \times D} E(X_t X_s) e_i(t) e_j(t) \, dsdt = \int_{D \times D} B(s,t) e_i(s) e_j(t) \, dsdt \]
\[ = \int_D \lambda_j e_i(s) e_j(s) \, ds = \lambda_j \delta_{ij}, \]
where the second-last step follows because $e_j$ is an eigenfunction, and the last step because $e_i, e_j$ are orthogonal if $i \neq j$. Again, we used Fubini’s Theorem to interchange the expectation and the integral; this will be a common calculation in this course and we will usually assume it is possible to do so without further remarks.
To show convergence of the series to \( X_t \), consider the partial sums \( S_N = \sum_{n=1}^{N} \xi_n e_n(t) \). Then
\[
\mathbb{E}|X_t - S_N|^2 = \mathbb{E}X_t^2 + \mathbb{E}S_N^2 - 2\mathbb{E}(X_t S_N)
\]
\[
= B(t,t) + \mathbb{E} \sum_{k,l=1}^{N} \xi_k \xi_l e_k(t) e_l(t) - 2\mathbb{E} \left( X_t \sum_{n=1}^{N} \xi_n e_n(t) \right) \quad \text{(just expand)}
\]
\[
= B(t,t) + \sum_{k=1}^{N} \lambda_k e_k^2(t) - 2\mathbb{E} \sum_{k=1}^{N} \int_{\mathcal{D}} X_t X_s e_k(s) e_k(t) \, ds \quad \text{(subs. for } \xi_n \text{)}
\]
\[
= B(t,t) - \sum_{k=1}^{N} \lambda_k e_k^2(t)
\]
The final line goes to 0 uniformly in \( t \), by Mercer’s theorem. \( \square \)

**Example** (Gaussian process). If \( X_t \) is Gaussian then we can completely specify the random coefficients \( \xi_i \). Indeed, \( \xi_i \sim N(0, \lambda_i) \), and \( \xi_i, \xi_j \) are independent when \( i \neq j \).

To show that \( \xi_i \) is Gaussian, we must show \( \int X_t e_i(t) \, dt \) is Gaussian. It is clear that if we use a Riemann sum approximation for the integral this will be Gaussian, since a finite sum of Gaussian random variables is Gaussian. The limit of a sequence of Gaussian random variables which converges in probability (namely the Riemann sums) is also Gaussian; this is something that can be shown by considering the characteristic functions and using Levy’s continuity theorem.

To show that \( \xi_i, \xi_j \) are independent for \( i \neq j \), it is enough to know they are uncorrelated and Gaussian, since uncorrelated Gaussian random variables are independent. Note this will *not* be true in general.

This provides a simple way to simulate a Gaussian process, by truncating and discretizing its KL expansion.

**Example** (Brownian motion). One way to define Brownian motion (the Wiener process) \( W_t \) is by the following properties:

(i) it is Gaussian;

(ii) it has mean \( m(t) = 0 \), covariance \( B(s,t) = \min(s,t) \).

(iii) with probability 1, \( t \rightarrow W_t \) is continuous, and \( W_0 = 0 \).

Let’s calculate the KL expansion on \([0, 1]\) using these properties.

First, we find the eigenfunctions. We must solve
\[
\int_0^1 (s \wedge t) e_k(t) \, dt = \lambda_k e_k(s) \quad \implies \quad \int_0^t e_k(t) \, dt + \int_s^1 se_k(t) \, dt = \lambda_k e_k(s) \quad \text{(4)}
\]
Taking \( \frac{d}{ds}, \frac{d^2}{ds^2} \) gives
\[
\lambda_k e'_k(s) = se_k(s) + \int_s^1 e_k(t) \, dt - se_k(s), \quad \lambda_k e''_k = -e_k.
\]
Setting \( s = 0, 1 \) in (4) to find \( e_k(0) = 0, e'_k(1) = 0 \). Solving this Sturm-Liouville problem gives
\[
\lambda_k = \left( \pi k - \frac{1}{2} \right)^2, \quad e_k(s) = \sqrt{2} \sin \left( \pi \left( k - \frac{1}{2} \right)s \right), \quad k = 1, 2, \ldots
\]
Therefore

\[ W_t = \sum_{k=1}^{\infty} \xi_k \frac{\sqrt{2}}{\pi (k - \frac{1}{2})} \sin \left( \pi (k - \frac{1}{2}) t \right), \quad \xi_k \sim N(0, 1), \text{ i.i.d.} \]

Note that this is not the only way to expand \( W_t \) in a set of basis functions, for example the Haar basis is also used in some applications.

5.2 Brownian motion

Brownian motion is perhaps the most important stochastic process we will see in this course. It was first brought to popular attention in 1827 by the Scottish botanist Robert Brown, who noticed that pollen grains suspended in water moved about at random, even when the water appeared to be very still. He repeated the experiment with dust particles, and found the same behaviour, so he argued the motion was unrelated to the fact that the pollen came from living matter.

Several mathematicians tried to explain this behaviour (Theile, 1880, Bachelier, 1900). However the phenomenon really took off as a model in physics after a paper by Einstein in 1905, which showed how the random motion could arise if water were made of many discrete components, rather than forming a continuum. He argued that this indirectly confirmed that matter was made of atoms.

Here is a movie showing 1\( \mu \)m glass beads sitting in water under a microscope: [https://www.youtube.com/watch?v=Xscn-QSmFo4](https://www.youtube.com/watch?v=Xscn-QSmFo4)

Here is another one, with actual pollen grains (but you have to put up with the soundtrack): [https://www.youtube.com/watch?v=jLQ66ytMa9I](https://www.youtube.com/watch?v=jLQ66ytMa9I)

![Figure 1: Some approximate realizations of Brownian motion. These were constructed by simulating a random walk with i.i.d. steps with distribution \( N(0, \sqrt{\Delta t}) \), at times \( \Delta t = 0.01 \). The total time of each realization is 10 units.](image)

The most common way to define a Brownian Motion is by the following properties:

**Definition.** A Brownian motion or Wiener process \( W_t, t \geq 0 \) is a real-valued stochastic process such that
Definition. Consider a sequence of random variables $X_1, X_2, \ldots$ defined on a sequence of probability spaces $(\Omega_n, \mathcal{F}_n, P_n)_{n=1}^{\infty}$ and taking values in some metric space $(S, \rho)$. Let $(\Omega, \mathcal{F}, P)$ be another probability space, on which another random variable $X$ is defined, which takes values in $(S, \rho)$. Then $\{X_n\}_{n=1}^{\infty}$ converges in distribution, or converges weakly to $X$, written $X_n \overset{d}{\rightarrow} X$, if $\mathbb{E}_n f(X_n) \rightarrow \mathbb{E} f(X)$ for all bounded, continuous, real-valued functions $f$, where $\mathbb{E}_n, \mathbb{E}$ denote expectations with respect to the measures associated with $X_n, X$ respectively.

Remark. It is possible to replace (ii) with the condition that the increments $W(s+t) - W(s)$ do not depend on $t$, plus a continuity condition that $\lim_{s \rightarrow 0} \mathbb{P}(|W(s+t) - W(t)| \geq \delta) = 0$ for all $\delta > 0$. One can then show the increments must be Gaussian. See [Brieman 1992], Ch. 12, p.248.

Property (i) says the increments are independent. Property (ii) says the increments are stationary and Gaussian. See Brieman [1992], Ch. 12, p.248.

An important question is whether such a process exists. The answer is yes, and is a subtle, technical question that is dealt with in many probability books, e.g. Durrett [2005], p.373, Karatzas and Shreve [1991], Brieman [1992]. The major difficulty is in showing property (iii): that there exists a version of Brownian motion that is continuous everywhere, almost surely.

5.2.1 Brownian motion as a limit of random walks

An intuitive way of thinking about Brownian motion is as a limit of random walks. Let $X_1, X_2, \ldots$ be iid random variables with mean 0 and variance 1. Consider the sum $S_n = \sum_{j=1}^{n} X_j$, with $S_0 = 0$. This is a discrete-time process, but we can make a continuous-time process by linearly interpolating between values of $S_n$.

We want the interpolated process to approach Brownian motion as the number of steps goes to infinity, so we will need to scale time and space appropriately. To figure out how, suppose time of the approximate Brownian motion is $t = n\Delta t$ for some $\Delta t$, and suppose we scale the jumps to have size $\Delta x$. We need $\Delta t, \Delta x \rightarrow 0$, but how should they be related?

Note that $\mathbb{E}S_n^2 = n$, so $\mathbb{E}(\Delta x S_n/\Delta t)^2 \approx (\Delta x)^2 t/\Delta t$. Since $\mathbb{E}W_t^2 = t$, we should choose $\Delta x = \sqrt{\Delta t}$. Let $\Delta t = 1/n$, and $\Delta x = 1/\sqrt{n}$. Let the interpolated, rescaled process be

$$W^n_t = \frac{S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)S_{\lfloor nt \rfloor} + 1}{\sqrt{n}}, \quad (5)$$

where $\lfloor nt \rfloor$ means the largest integer less than or equal to $nt$.

We would like to show that $W^n_t$ converges in some way to Brownian motion. For this, we need a definition of convergence.

Definition. Consider a sequence of random variables $X_1, X_2, \ldots$ defined on a sequence of probability spaces $(\Omega_n, \mathcal{F}_n, P_n)_{n=1}^{\infty}$ and taking values in some metric space $(S, \rho)$. Let $(\Omega, \mathcal{F}, P)$ be another probability space, on which another random variable $X$ is defined, which takes values in $(S, \rho)$. Then $\{X_n\}_{n=1}^{\infty}$ converges in distribution, or converges weakly to $X$, written $X_n \overset{d}{\rightarrow} X$, if $\mathbb{E}_n f(X_n) \rightarrow \mathbb{E} f(X)$ for all bounded, continuous, real-valued functions $f$, where $\mathbb{E}_n, \mathbb{E}$ denote expectations with respect to the measures associated with $X_n, X$ respectively.

(i) if $t_0 < t_1 < \cdots < t_n$, then $W(t_0), W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$ are independent.

(ii) For all $s, t \geq 0$, the random variable $W(s+t) - W(s)$ is Gaussian, with mean zero and variance $s$.

(iii) With probability 1, $t \rightarrow W_t$ is continuous, and $W(0) = 0$. 

\[ E(W_{nt}) = nt, \quad \text{Var}(W_{nt}) = nt \]

\[ \sqrt{n} (W_{nt} - nt) \rightarrow N(0, t) \text{ in distribution} \]

\[ \sqrt{n} (W_{nt} - nt - t) \rightarrow Z_n \text{ in distribution} \]

\[ \mathbb{P}(\lim_{n \rightarrow \infty} \frac{\sqrt{n} (W_{nt} - nt - t)}{Z_n} = 1) = 1 \]

\[ \mathbb{E}(\sqrt{n} (W_{nt} - nt - t)) = 0, \quad \text{Var}(\sqrt{n} (W_{nt} - nt - t)) = t \]

\[ \mathbb{E}(\sqrt{n} (W_{nt} - nt - t)^2) = t \]

\[ \sqrt{n} (W_{nt} - nt - t - t) \rightarrow Z_n \text{ in distribution} \]

\[ \mathbb{P}(\lim_{n \rightarrow \infty} \frac{\sqrt{n} (W_{nt} - nt - t - t)}{Z_n} = 1) = 1 \]

\[ \mathbb{E}(\sqrt{n} (W_{nt} - nt - t - t)) = 0, \quad \text{Var}(\sqrt{n} (W_{nt} - nt - t - t)) = t \]

\[ \mathbb{E}(\sqrt{n} (W_{nt} - nt - t - t)^2) = t \]
If \( X_n, X \) are real-valued, then an equivalent definition is that \( F_n(x) \to F(x) \) at each point of continuity of \( F(x) \), where \( F_n, F \) are the cumulative distribution functions of the random variables. (The former definition works for random variables taking values in a path space, while the latter works also for real-valued random variables.)

We know by the Central Limit Theorem that \( W^n_d \overset{d}{\to} N(0,t) \) as \( n \to \infty \). Therefore the one-point distributions of \( W^n_d \) converge to those of Brownian motion.

To show that the \( d \)-point distributions converge uses a similar technique as in the proof of the CLT; one considers the characteristic functions and shows these converge. See [Karatzas and Shreve (1991), p. 67].

Showing the entire process \( W^n_t \) converges in distribution to \( W_t \) is significantly harder and is the statement of Donker’s Invariance Principle; see [Karatzas and Shreve (1991), p. 70]. This is a generalization of the CLT to path space.

5.2.2 Properties of Brownian motion

Scaling properties

(i) \( -W_t \) is a Brownian motion (symmetry)
(ii) \( W_{s+t} - W_s \) for fixed \( s \) is a Brownian motion (time-homogeneity)
(iii) \( \frac{1}{\sqrt{c}} W_{ct} \), with \( c > 0 \) is a fixed constant, is a Brownian motion (scaling)
(iv) \( t W_{1/t} \) is a Brownian motion (time-inversion)

Proof. (i), (ii), (iii) follow straightforwardly from the definition.

To check (iv), we use the definition of BM as a Gaussian process. We have that \( tW_t \) is Gaussian, with mean 0. It has covariance function \( \text{Cov}(W_s, W_t) = st \left( \frac{1}{2} \wedge \frac{1}{2} \right) = s \wedge t \). It is continuous for \( t > 0 \). It remains to check that it is continuous at 0. But \( \lim_{t \to 0} t W_{1/t} = \lim_{s \to \infty} \frac{W_s}{s} \to 0 \) a.s., by the result below. □

Behaviour as \( t \to \infty \) There are several ways to characterize BM in the limit as \( t \to \infty \):

(i) \( \lim_{t \to \infty} \frac{W_t}{\sqrt{t}} = 0 \) a.s..
(ii) \( \limsup_{t \to \infty} \frac{W_t}{\sqrt{t}} = \infty \), \( \liminf_{t \to \infty} \frac{W_t}{\sqrt{t}} = -\infty \) (both a.s.).
(iii) (Law of the Iterated Logarithm)

\[
\limsup_{t \to \infty} \frac{W_t}{\sqrt{2t \log \log t}} = 1 \quad \text{a.s.,} \quad \limsup_{t \to 0^+} \frac{W_t}{\sqrt{2t \log \log 1/t}} = 1 \quad \text{a.s.}
\]

If limsup is replaced by liminf in either of the above, the limits are \( -1 \).

Proof. (i) (from [Brieman (1992), p. 265]) This follows from the Strong Law of Large Numbers. For \( n \in \mathbb{N} \) we can write \( W_n = (W_1 - W_0) + (W_2 - W_1) + \ldots + (W_n - W_{n-1}) \), which is a sum of iid random variables. By the SLLN, \( W_n/n \to 0 \) a.s. To obtain behaviour at non-integer \( t \), let

\[
Z_k = \max_{0 \leq t \leq 1} |B(k + t) - B(k)|.
\]
For $t \in [k, k+1]$, 
\[
\left| \frac{W_t}{t} - \frac{W_k}{k} \right| \leq \frac{1}{k(k+1)} |W_k| + \frac{1}{k+1} Z_k.
\]

The first term on the RHS $\to 0$ a.s., and $Z_k$ has the same distribution as $\max_{0 \leq t \leq 1} |W_t|$. It can be shown that $\mathbb{E}Z_k < \infty$, and that this implies $Z_k/k \to 0$ a.s.

(iii) For the second part, see e.g. [Brieman 1992], p. 263, or [Karatzas and Shreve 1991], p. 112. The first part follows from the second using the time inversion property (iv) of BM.

(ii) This follows from the Law of the Iterated Logarithm.

**Differentiability**

**Theorem.** With probability one, Brownian paths are not Lipschitz continuous (and hence not differentiable) at any point.

**Remark.** We can see this heuristically as follows. First let’s argue why it should not be differentiable at $t = 0$. This implies it’s not differentiable anywhere, by the time-homogeneity property (ii) (though be careful: we should really show it’s not differentiable all points at once.) If it were differentiable at 0, then we would have $W'_t = \lim_{t \to 0} \frac{W_t}{t} = \lim_{s \to \infty} sW_{1/s} = \lim_{t \to \infty} \tilde{W}_t$, where $\tilde{W}_t$ is another BM, by the scaling property (iv). But this limit doesn’t exist.

**Remark.** Here is another heuristic argument. We have $\mathbb{E}W_t^2 = t$, so $\text{Var}(W_t) = \frac{t}{n} = \frac{1}{t}$. This $\to \infty$ as $t \to 0$, so we would not expect the derivative to exist at 0.

**Proof.** (From [Brieman 1992], p. 261, in turn from Dvoretsky, Erdős, and Kakutani (1961). The same proof is presented in [Durrett 2005], p 377.)

Notice that if a function $x(t)$ has a derivative $x'(s)$, with $|x'(s)| < \beta$ at some point $s \in [0, 1]$, then there is an $n_0$ such that for $n > n_0$,
\[
|x(t) - x(s)| \leq 2\beta |t - s|, \quad \text{if } |t - s| \leq 2/n. \tag{6}
\]

Let 
\[A_n = \{ \omega : \text{there is an } s \in [0, 1] \text{ s.t. } |W_t - W_s| \leq 2\beta |t - s| \text{ when } |t - s| \leq 2/n \}.\]

The $A_n$ increase with $n$, and the limit set $A$ includes the set of all sample paths on $[0, 1]$ having a derivative at any point which is less than $\beta$ in absolute value. If (6) holds, then let $k$ be the largest integer such that $k/n \leq s$, so that
\[
y_k = \max \left\{ \left| W_{k+1/n} - W_{k/n} \right|, \left| W_{k+1/n} - W_k \right|, \left| W_k - W_{k-1/n} \right| \right\} \leq \frac{6\beta}{n}.
\]

Therefore, if we let
\[
C_n = \left\{ B(\cdot) : \text{at least one } y_k \leq \frac{6\beta}{n} \right\},
\]
then $A_n \subset C_n$. To show $P(A) = 0$, which implies the theorem, it is sufficient to get $\lim_n P(C_n) = 0$. But
\[
C_n = \bigcup_{k=1}^{n-2} \left\{ B(\cdot) : y_k \leq \frac{6\beta}{n} \right\}.
\]
so

\[ P(C_n) \leq \sum_{k=1}^{n-2} P \left( \max \left\{ |W_{k+2} - W_{k+1}|, |W_{k+1} - W_k|, |W_k - W_{k-1}| \right\} \leq \frac{6\beta}{n} \right) \]

\[ \leq nP \left( \max \left\{ |W_{3/n} - W_{2/n}|, |W_{2/n} - W_{1/n}|, |W_{1/n}| \right\} \leq \frac{6\beta}{n} \right) \]

\[ = nP \left( W_{1/n} \leq \frac{6\beta}{n} \right)^3 \]

\[ = n \left( \frac{n}{2\pi} \int_{-6\beta/n}^{6\beta/n} e^{-x^2/2} \, dx \right)^3 \]

\[ = n \left( \frac{1}{\sqrt{2\pi n}} \int_{-6\beta}^{6\beta} e^{-x^2/2n} \, dx \right)^3 \]

The final integral converges to 0 as \( n \to \infty \), so \( P(W_n) \to 0 \).

\[ \square \]

**Corollary.** Almost surely, every sample path of \( W_t \) has infinite variation on every finite interval.

**Proof.** If a function has bounded variation on an interval \( I \), then it has a derivative existing almost everywhere on \( I \).

**Theorem.** (a) With probability 1, a Brownian sample path is locally Hölder continuous with exponent \( \gamma \) for every \( \gamma \in (0, \frac{1}{2}) \).

(b) With probability 1, Brownian paths are nowhere locally Hölder continuous for any exponent \( \gamma > \frac{1}{2} \).

**Proof.** See [Karatzas and Shreve 1991], [Durrett 2005].

\[ \square \]

### 5.2.3 Quadratic variation

Even though Brownian motion is not differentiable, hence has infinite variation, it actually has finite quadratic variation. This will be a very important property that we will use when constructing the stochastic integral.

**Definition.** The \( p \)th-variation of a function \( f \) on \([a, b]\) given partition \( \sigma = \{t_0, t_1, \ldots, t_n\} \), with \( a = t_0 < t_1 < \cdots < t_n = b \) is

\[ V_p^{f|[a,b]}(\sigma) = \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p. \]

**Definition.** The quadratic variation of a function \( f \) on \([0, t]\) is often written as \( Q^\sigma_t(f) \equiv V^2_{[0,t]}(f, \sigma) \), or just \( Q^\sigma_t \) if the function is clear. Explicitly, we have

\[ Q^\sigma_t(f) = \sum_{i=0}^{n} |f(t_{i+1}) - f(t_i)|^2. \]
Definition. A sequence of random variables \(X_1, X_2, \ldots\) converges in mean-square to another random variable \(X\), written \(X_n \xrightarrow{m.s.} X\), if \(E[|X_n - X|^2] \to 0\) as \(n \to \infty\).

Lemma. The quadratic variation of Brownian motion \(Q_t(W_i)\) converges in mean-square to \(t\) as \(|\sigma| \to 0\), where \(|\sigma| = \max_i |t_{i+1} - t_i|\), i.e. \(\sum_{i=1}^{n} |W_{i+1} - W_i|^2 \xrightarrow{m.s.} t\).

Remark. This means that increments of \(W_i\) behave as \(\Delta t\), i.e. formally we can write \((\Delta W)^2 = \Delta t\), or \((dW_t)^2 = dt\).

Proof. (From Koralov and Sinai [2010], p. 269.)

\[
\mathbb{E}
\left(
Q_t(W_i) - t
\right)^2
= \mathbb{E}
\left(
\sum_{i=1}^{n} [(W_{i} - W_{i-1})^2 - (t_i - t_{i-1})]
\right)^2
\geq\]
\[
\sum_{i=1}^{n} \mathbb{E}
\left([W_{i} - W_{i-1}]^2 - (t_i - t_{i-1})
\right)^2
\geq\]
\[
\sum_{i=1}^{n} \mathbb{E}(W_{i} - W_{i-1})^4 + \sum_{i=1}^{n} (t_i - t_{i-1})^2
\geq\]
\[
4 \sum_{i=1}^{n} (t_i - t_{i-1})^2
\geq\]
\[
4 \max_{1 \leq i \leq n} (t_i - t_{i-1}) \sum_{i=1}^{n} (t_i - t_{i-1})
\geq 4t(\sigma)
\]

The second equality uses the fact that \(\mathbb{E}((W_{i} - W_{i-1})^2 - (t_i - t_{i-1}))((W_{j} - W_{j-1})^2 - (t_j - t_{j-1})) = 0\) if \(i \neq j\). \(\square\)

5.2.4 Brownian motion as a Markov process

Suppose we know the value of Brownian motion for all times up to some time \(s\). What can we say about \(W_t\) for \(t > s\)? Since \(W_t = W_s + (W_t - W_s)\), and the increment \(W_t - W_s\) is independent of all observations up to time \(s\), we can obtain the distribution of \(W_t\) using only our knowledge of \(W_s\), and not any earlier observations. In other words, we can write

\[
P(W_{t_n} \in F_n | W_{t_{n-1}} \in F_{n-1}, \ldots, W_0 \in F_0) = P(W_{t_n} \in F_n | W_{t_{n-1}} \in F_{n-1}),
\]

where \(t_0 < t_1 < \cdots < t_n\), and \(F_i\) are Borel sets.

This seems a lot like the Markov property that we studied in Lectures 2&3 – the difference is that previously, we considered Markov chains that could take on a discrete set of values, whereas Brownian motion can take on a continuum. It turns out that \((7)\) is actually the definition of a general Markov process\(^1\) and that Brownian motion satisfies this condition (Koralov and Sinai [2010], p.278.) Therefore Brownian motion is our first example of a continuous-time, continuous-space Markov process; there are many more to come.

\(^1\)For a formulation of the Markov property in terms of \(\sigma\)-algebras, see Lecture 3 notes.
Since Brownian motion is Markov, it makes sense to study its transition function $P(\Gamma, t|y, s) = P(W_t \in \Gamma|W_s = y)$, where $s < t$ and $\Gamma$ is a Borel set. This can be written in terms of a transition density $p(x, t|y, s)$ as

$$P(\Gamma, t|y, s) = \int_{\Gamma} \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(x-y)^2}{2(t-s)}} dx = \int_{\Gamma} p(x, t|y, s) dx \quad (8)$$

The transition density $p(x, t|y, s)$ gives the probability density to be at $x$ at time $t$, given the process was at $y$ at time $s$. It is the continuous analog of the transition matrix that we introduced for Markov chains. We were able to write it down explicitly in (8), because we know the increments of Brownian Motion are Gaussian.

An approach to studying Markov processes that is common in the Physics literature is to start with a set of transition functions or densities with known infinitesimal properties, and to show they satisfy certain evolution equations (and/or analyze the consistency of the initial assumptions.) We will explore such techniques on the homework. For now, let’s consider how the transition probabilities for Brownian motion evolve, both forward and backward in time.

The transition probability density for Brownian motion is stationary in both time and space:

$$p(x, t|y, s) = p(x-y, t-s|0, 0) \quad (9)$$

This is a version of the Kolmogorov forward equation.

It also satisfies a PDE in $s, y$:

$$\frac{\partial p}{\partial s} = -\frac{1}{2} \frac{\partial^2 p}{\partial y^2}, \quad p(x, t|y, t) = \delta(x-y) \quad (10)$$

This is a version of the Kolmogorov backward equation. It should be solved backward in time in $s$, starting at $s = t$.

For a continuous-time Markov chain, we calculated the infinitesimal generator and derived a number of properties from this. It turns out this concept is also very useful when studying more general Markov processes. The definition of the generator is very similar:

**Definition.** The generator of a Markov process is the operator $\mathcal{A}$ defined on a subset $D(\mathcal{A})$ of the set of bounded, measurable functions $f$ by

$$\mathcal{A} f \equiv \lim_{t \to 0} \frac{T_t f - f}{t} \implies (\mathcal{A} f)(x) = \lim_{t \to 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t} \quad (11)$$

The set $D(\mathcal{A})$ on which this limit exists is the domain of $\mathcal{A}$. We have introduced the operator

$$T_t f(x) \equiv \mathbb{E}_x f(X_t) = \int_{\mathbb{R}} f(y) P(dy, t|x, 0) \quad (12)$$

which acts on the set of bounded, measurable functions $f$. Here $\mathbb{E}_x$ means the expected value, given $X_0 = x$. The convergence is understood as the norm convergence, i.e. we mean that $\lim_{t \to 0} \|T_t f - f\| = 0$, for an appropriate norm $\| \cdot \|$ on the function space.
Remark. The set of operators \( \{ T_t \}_{t \geq 0} \) forms an operator semigroup, i.e. \( T_t \circ T_s = T_{t+s}, f \). Note the relation to the Chapman-Kolmogorov equations \( P_t ^t P_s = P_{t+s}, f \). You can show this by considering the general version of the Chapman-Kolmogorov equations: \( P(B, s|x, t) = \int _\mathbb{R} P(B, x|y, u)P(dy, u|x, t), t \leq u \leq s \).

Let’s calculate the generator of Brownian motion. Suppose \( f \in C_0^\infty (\mathbb{R}) \).

\[
(\mathcal{A} f)(x) = \lim _{t \to 0} \frac{E_x f(X_t) - f(x)}{t} = \lim _{t \to 0} \frac{1}{t} \int _\mathbb{R} (f(y) - f(x))p(y, t|x, 0)dy
\]

\[
= \lim _{t \to 0} \frac{1}{t} \int _0 ^t \partial _t \left[ \int _\mathbb{R} (f(y) - f(x))p(y, s|x, 0)dy \right] ds \quad \quad \text{(since can’t put } \partial _t \text{ under integral)}
\]

\[
= \lim _{t \to 0} \frac{1}{t} \int _0 ^t \int _\mathbb{R} (f(y) - f(x)) \frac{\partial ^2 p}{\partial y^2} (y, s|x, 0)dyds \quad \quad \text{(using the forward equation)}
\]

\[
= \lim _{t \to 0} \frac{1}{t} \int _0 ^t \frac{\partial ^2 f}{\partial x^2} (x) dyds \quad \quad \text{(after integrating by parts twice)}
\]

\[
= \frac{1}{2} \frac{\partial ^2 f}{\partial x^2} (x)
\]

Note that in the second step, it would be nice to differentiate under the integral and replace \( \partial _t p \) with \( \frac{1}{2} \partial _{yy} p \), directly, but we don’t know that we can since \( p|_{t=0} \) is unbounded, so we have to get around that by first writing \( p \) as a time integral of its time derivative.

This shows that the generator of Brownian motion is the Laplacian: \( \mathcal{A} f = \frac{1}{2} \frac{\partial ^2 f}{\partial x^2} \). Note the analogy to Lecture 3: there, the generator was a matrix applied to a vector (or, in the case of the Poisson process, a linear function applied to a countable sequence.) Here, the generator is still linear, but now it acts on functions, not vectors: it is a linear operator. This will be true for Markov processes in general: their evolution can be described by a generator, which is always a linear functional. The functional won’t always be a partial differential operator – though for a wide class of processes, namely diffusion processes, it is. Another common possibility is an integral operator, which occurs for jump processes. The Hille-Yosida theorem provides the conditions for a closed linear operator \( \mathcal{A} \) on a Banach space to be the infinitesimal generator of a Markov process (alternatively, a strongly continuous one-parameter semigroup.)

Why are we interested in the generator? This is the fundamental object from which we can describe the evolution of probability and measurable statistics. Let’s start with the latter. Let’s define \( u(x, t) \equiv E_x f(X_t) = T_t f(x) \). To find out how \( u \) evolves in time, we calculate:

\[ \frac{\partial u}{\partial t} = \lim _{h \to 0} \frac{T_{t+h} f - T_t f}{h} = \lim _{h \to 0} \frac{T_h(T_t f) - T_0(T_t f)}{h} = \mathcal{A} T_t f = \mathcal{A} u. \] (13)

We obtain the

**Backward Kolmogorov Equation.**

\[ \frac{\partial u}{\partial t} = \mathcal{A} u, \quad u(x, 0) = f(x). \] (14)
Note that we could have factored the other way in (13), and found that
\[
\frac{\partial u}{\partial t} = \lim_{h \to 0} \frac{1}{h} \left( T_h f - T_0 f \right) = T_t A f,
\]
assuming that we can interchange the limit and expectation. (See the handout for a discussion of conditions under which this is possible; basically we need to adapt the Bounded/Monotone/Dominated Convergence theorems to the case of random variables.) This shows that for a certain class of processes, \( T_t A = A T_t \), i.e. the generator and transition operator commute.

Now consider how the probability density evolves. This derivation will be formal and heuristic, but you can find more rigorous derivations elsewhere. The probability density at time \( t \) is \( \rho(y,t) = \int \rho_0(x) p(y,t|x,0) \, dx \), where \( \rho_0(y) \) is the initial density. Therefore to compute \( \frac{\partial \rho}{\partial t} \), we need an equation for \( \frac{\partial}{\partial t} p(y,t|x,0) \). We will do this indirectly, working with our previous calculations for the backward equation. Let \( f \) be a suitable test function, e.g. a bounded, measurable function with compact support. If we let \( u = T_t f \) as before, we know that
\[
\frac{\partial u}{\partial t} = T_t A f = \int p(y,t|x,0) A f(y) \, dy = \int (A^* y p) f(y) \, dy.
\]
Here \( A^* \) is the adjoint of \( A \) with respect to the \( L^2 \) inner product, i.e. \( \int u A v = \int v A^* u \), and the subscript is a reminder that it acts on the \( y \)-variables of \( p \).

We also know that \( \frac{\partial u}{\partial t} = \int \frac{\partial p}{\partial t}(x,t,y,0) f(y) \, dy \). (We can differentiate under the integral provided \( p_t \) exists and is locally in \( L^1 \). As in the case for Brownian motion above, this won’t hold at \( t = 0 \), but we can use the same trick to get around it so we just assume it holds here.) Since the two calculations hold for all test functions \( f \), we have that \( \frac{\partial u}{\partial t} = A^* y p(y,t|x,0) \) (weakly). Integrating over the initial density \( \rho_0(x) \) gives

**Forward Kolmogorov Equation.**
\[
\frac{\partial \rho}{\partial t} = A^* \rho, \quad \rho(0,y) = \rho_0(y).
\]

**Remark.** The forward equation exists for all the processes we will encounter, though in general its mathematical existence is trickier to establish than the backward equation. Sometimes, (15) must be interpreted in weak form, for example when the probability measure is restricted to a submanifold so it contains \( \delta \)-functions.

**References**


