Lecture 6: Brownian motion

Readings

Recommended:

- Pavliotis (2014) section 1.3, 2.1-2.3

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

6.1 Introduction

Brownian motion is perhaps the most important stochastic process we will see in this course. It is named after Scottish botanist Robert Brown, who in 1827 noticed that pollen grains suspended in water move about at random even when the water is still. Some people thought this motion was because grains come from living matter and are moving about of their own accord. To rule out this explanation, Brown showed that inanimate objects like dust particles more in the same erratic fashion. Brown was not the first person to notice this peculiar motion – for example, it is described by a Roman Lucretius in a poem in 60 BC, and pointed out by Dutch scientist Jan Ingenhousz in 1785 – but he was the first to investigate it systematically.

Brown's work inspired both physicists to investigate the phenomenon. It was finally understood by Einstein in 1905 (Einstein, 1905), who 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. Smoluchowski constructed a related model in 1906 (Von Smoluchowski, 1906), which was experimentally verifieid by Jean Baptiste Perrin in 1908, who received the Nobel prize for his work.

Meanwhile, mathematicians realized that a function describing a particle moving according to Brownian motion would have some bizarre properties that meant it couldn't be constructed as a function using classical mathematical techniques. Several mathematicians tried to invent new mathematics to construct a Brownian motion (e.g. Theile, 1880, Bachelier, 1900), and a rigorous construction was finally given by Wiener in 1923. For this reason in mathematics Brownian motion is often called a *Wiener process*.

6.2 Definitions

We'll start by asking how to construct a stochastic process that could model the erratic motion of a dust particle or other processes that are "very random." We'd like the process to be "as random as possible", because if there is any deterministic or predictable part of the process, then we could model that part in a more conventional way. Therefore we'd like to ask for a continuous process whose derivative is independent at each point in time – the idea being that if the derivative has some correlations, then we could predict the value of the process at least some time into the future, and then the process wouldn't be as random as possible. It turns out that the best approximation for such a process is a Brownian motion.

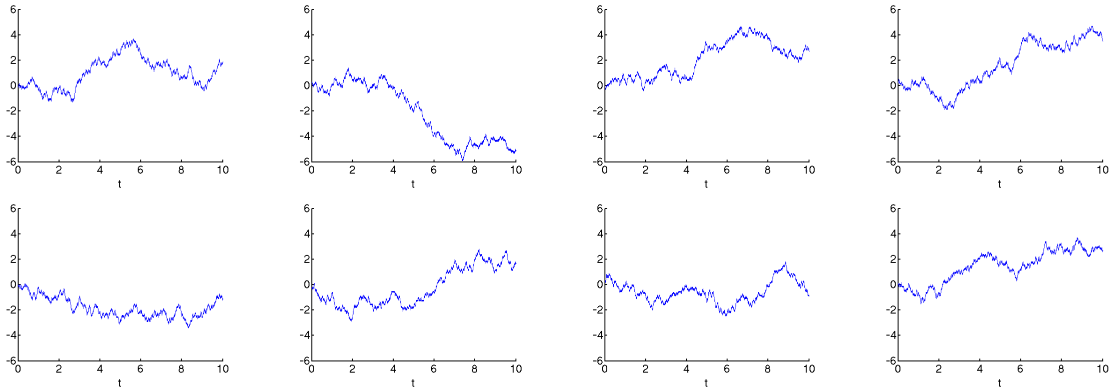

Figure 1: Some approximate realizations of Brownian motion. These were constructed by simulating a random walk with i.i.d. steps with distribution $N(0,\sqrt{\Delta t})$, at times $\Delta t = 0.01$. The total time of each realization is 10 units.

We'll first study the path properties of Brownian motion, and then we'll look at what we can say about its statistics. Brownian motion is our first example of a diffusion process, which we'll study a lot in the coming lectures, so we'll use this lecture as an opportunity for introducing some of the tools to think about more general Markov processes.

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

Definition (#1.)**.** A *Brownian motion* or *Wiener process* $W = (W_t)_{t\geq 0}$ is a real-valued stochastic process that has

(i) $W_0 = 0$;
(ii) Independent increments: the random variables $W_v - W_u$, $W_t - W_s$ are independent whenever $u \leq v \leq s \leq t$ (so the intervals (u,v), (s,t) are disjoint.)
(iii) Normal increments:[1] $W_{s+t} - W_s \sim N(0,t)$ for all $s,t \geq 0$.
(iv) Continuous sample paths: with probability 1, the function $t \to W_t$ is continuous.

That such a process exists, and that its probability law is uniquely determined by the above properties, is a result shown in many probability texts (e.g. Durrett (2005), p.373, Karatzas and Shreve (1991), Breiman (1992)). The major difficulty is in showing property (iv): that there exists a version of Brownian motion that is continuous everywhere, almost surely.

The properties of Brownian motion are a lot like those of the Poisson process. Property (iii) implies the increments are stationary, so a Brownian motion has stationary, independent increments, just like the Poisson process. The differences from the Poisson process is that the increments of Brownian motion have a normal distribution, not a Poisson distribution, and that it is a continuous process.

[1] Somewhat remarkably, it is possible to replace this condition with the condition that the increments $W_{s+t} - W_t$ do not depend on t, plus a continuity condition $\lim_{s\to 0} \frac{P(|W_{s+t}-W_t|\geq\delta)}{s} = 0$ for all $\delta > 0$. That the increments are normal comes from the Central Limit Theorem, by breaking up an increment over a finite interval into a sum of increments over smaller intervals. See Breiman (1992), Ch. 12, p.248.

With these properties we can say a lot about the trajectories and statistics of the process. For example, we can calculate all finite dimensional distributions.

Example 6.1 Let's calculate the two-point distributions $P(W_s \in A, W_t \in B)$, with $A, B \subset \mathbb{R}$ and $s < t$. We have

$$\begin{aligned}
P(W_s \in A, W_t \in B) &= P(W_t - W_s + W_s \in B, W_s \in A) \\
&= \int_{x\in A} P(W_t - W_s \in B - x, W_s \in [x, x+dx)) \\
&= \int_{x\in A} P(W_t - W_s \in B - x) P(W_s \in [x, x+dx)) && \text{increments are independent} \\
&= \int_{x\in A}\int_{y\in B} \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(y-x)^2}{2(t-s)}} \frac{1}{\sqrt{2\pi s}} e^{-\frac{x^2}{2s}}\, dy dx && \text{increments are normal}
\end{aligned}$$

This calculation also shows that the joint density for (W_s, W_t) is

$$p_{s,t}(x,y) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(y-x)^2}{2(t-s)}} \frac{1}{\sqrt{2\pi s}} e^{-\frac{x^2}{2s}}\,. \tag{1}$$

This density is Gaussian, so (W_s, W_t) is a two-dimensional Gaussian vector. If we define the two-point transition density to be

$$p(y,t|x,s)dx = P(W_t \in [y, y+dy] | W_s = x) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(y-x)^2}{2(t-s)}} dx \tag{2}$$

to be the probability density for $W_t = y$, given $W_s = x$, then we can write the two-point density as

$$p_{s,t}(x,y) = p(y,t|x,s)p(x,s|0,0)\,.$$

Example 6.2 Let's calculate the n-point fdds. Let $t_1 < \cdots < t_n$, and let $p_{t_1,t_2,\ldots,t_n}(x_1,x_2,\ldots,x_n)$ be the joint density of $(W_{t_1},\cdots,W_{t_n})$. Similar calculations to our earlier example show that

$$p_{t_1,t_2,\ldots,t_n}(x_1,x_2,\ldots,x_n) = p(x_1,t_1|x_1,0)\; p(x_2,t_2|x_1,t_1)\cdots p(x_n,t_n|x_{n-1},t_{n-1})\,.$$

That is, the joint density of BM at n different timepoints, is obtained by multiplying the two-point transition densities together. Since these are each Gaussian, the whole product is Gaussian, and we find the n-point fdd is a multivariate Gaussian.

Recall the definition of Brownian motion from Lecture 5:

Definition (#2.)**.** A *Brownian motion* or *Wiener process* is a stochastic process $W = (W_t)_{t\ge 0}$ with the following properties:

(i) $W_0 = 0$;
(ii) It is a Gaussian process;
(iii) It has mean $m(t) = 0$ and covariance $B(s,t) = \min(s,t)$.
(iv) It has continuous sample paths: with probability 1, the function $t \to W_t$ is continuous.

Proposition. *Definitions 1&2 are equivalent.*

Therefore you can work with whichever one is more convenient for the problem at hand.

Proof. (For a full proof, see Durrett (2005), p.373.) Let's show that Definition #1 $\Rightarrow$ Definition #2. Given a Brownian motion satisfying Definition #1, we need to show that it satisfies properties (ii),(iii) of Definition # 2. Properties (i),(iv) are included in Definition #1. Property (ii), that BM is a Gaussian process, follows from our examples above.

It remains to check property (iii) of Definition #2. Since $W_t \sim N(0,t)$ by property (iii) of Definition #1, we have $\mathbb{E}W_t = 0$. Let's compute its covariance. For $s < t$, we have

$$B(s,t) = \mathbb{E}W_sW_t - (\mathbb{E}W_s)(\mathbb{E}W_t) = \mathbb{E}W_s(W_t - W_s + W_s) = (\mathbb{E}W_s)\mathbb{E}(W_t - W_s) + \mathbb{E}W_s^2 = s.$$

The second-last step follows since $W_t - W_s$ is independent of W_s (property (ii)), and the third by the distribution of the increments (property (iii).) Therefore the BM satisfies property (iii) of Definition #2. $\square$

***Exercise* 6.1.** Show that Definition #2 $\Rightarrow$ Definition #1.

6.3 Brownian motion as a limit of random walks

One way to construct a Brownian motion is as a limit of random walks. Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean 0 and variance 1. For the sake of illustration let's suppose that $X_i = \pm 1$ with equal probability; the argument below will hold for more general step distributions. Consider the sum

$$S_n = \sum_{j=1}^{n} X_j \qquad \text{with } S_0 = 0.$$

This is a simple symmetric random walk on the integers. It is a discrete-time process, but we can make a continuous-time process by linearly interpolating between values of S_n.

Consider the properties of $(S_t)_{t\in\mathbb{N}}$:

(i) $\mathbb{E}S_t = 0$
(ii) $\mathrm{Var}(S_t) = t$
(iii) $(S_t)_{t\in\mathbb{N}}$ has stationary increments.
To see why, note that

$$S_t - S_s = X_{s+1} + \cdots X_t\,, \qquad S_{t-s} = X_1 + \cdots + X_{t-s}\,.$$

Each of $S_t - S_s, S_{t-s}$ is a sum of $t-s$ i.i.d random variables, so $S_t - S_s \sim S_{t-s}$.
(iv) $(S_t)_{t\in\mathbb{N}}$ has independent increments.
To see why, let $0 < q < r \le s < t$, and write

$$S_t - S_s = X_{s+1} + \cdots + X_t\,, \quad S_r - S_q = X_{q+1} + \cdots + X_r\,.$$

Each of $S_t - S_s$, $S_r - S_q$ is a sum of distinct, independent random variables, so they are independent.
(v) For t large, $S_t \approx N(0,t)$. This follows from the Central Limit Theorem.

Therefore $(S_t)_{t\in\mathbb{N}}$ has many of the properties of a Brownian motion. We might wonder if there is a way to scale it so it approaches a Brownian motion in some limit.

We will construct such a limit by scaling space and time in a particular way. Suppose we scale spatial steps by Δx, and time steps by Δt. The rescaled process is

$$S_t^{\Delta t,\Delta x} = \Delta x\, S_{t/\Delta t} = \Delta x\left(X_1 + \cdots + X_{t/\Delta t}\right) . \tag{3}$$

We want to consider the limit of the process $(S_t^{\Delta t,\Delta x})_{t\in\mathbb{N}/\Delta t}$ as $\Delta t, \Delta x \to 0$. How should these parameters be related? If the limit is to approach something finite, then the variance should be finite too. Since

$$\mathrm{Var}(S_t^{\Delta t,\Delta x}) = \frac{(\Delta x)^2}{\Delta t}t \qquad \Rightarrow \qquad \text{we should choose } \frac{(\Delta x)^2}{\Delta t} = constant\ . \tag{4}$$

This is an important point – *for a diffusion process, space scales as the square root of time*. We will call this *diffusive scaling*. It will come up again and again throughout the course.

Let's suppose the constant equals 1 so the limiting process has the same variance at a point as a Brownian motion. We write $\Delta t = 1/n$, $\Delta x = 1/\sqrt{n}$, and define a sequence of processes in terms of parameter n. Since our original process was a discrete-time process, it is convenient to make a continuous-time process by linearly interpolating between the discrete values of t. The interpolated, rescaled process is

$$S_t^{(n)} = \underbrace{\frac{S_{[nt]}}{\sqrt{n}}}_{\substack{\text{rescaled}\\ \text{random walk}}} + \underbrace{\frac{(nt-[nt])S_{[nt]+1}}{\sqrt{n}}}_{\text{needed for interpolation}}, \tag{5}$$

where $[nt]$ means the largest integer less than or equal to nt.

Then *Donsker's Theorem* or *Donsker's Invariance Principle* says that $S^{(n)} \equiv (S_t^{(n)})_{t\in[0,1]}$ converges in distribution[2] to a Brownian motion W on $[0,1]$ (see Durrett (1996) Section 8.5, or Durrett (2005) Section 7.6), where the convergence is in the space of continuous functions. (This result also holds on any finite interval of time, by rescaling time appropriately.) That is, given the space $\mathscr{C}([0,1])$ of continuous functions on $[0,1]$ equipped with the sup-norm $\|f\| = \sup\{f(t) : t\in[0,1]\}$, we have that $S^{(n)} \Rightarrow W^{(1)} \equiv (W_t)_{t\in[0,1]}$, i.e. the associated measures on $\mathscr{C}([0,1])$ converge weakly.[3]

Donsker's Theorem is a powerful theorem, which is a generalization of the Central Limit Theorem to path space. It implies that all the finite-dimensional distributions of $S^{(n)}$ converge to the finite-dimensional distributions of a Brownian motion. It also goes further, and implies that $S^{(n)}$ converges weakly as an entire path. That is, given a functional $\psi : \mathscr{C}([0,1]) \to \mathbb{R}$ that is a.s. continuous (with respect to the measure of $W^{(1)}$), we have that

$$\psi(S^{(n)}) \Rightarrow \psi(W^{(1)}).$$

Example 6.3 Let $\psi(f) = f(1)$. Then $\psi : \mathscr{C}([0,1]) \to \mathbb{R}$ is continuous, and Donsker's theorem implies that $S_1^{(n)} \Rightarrow W_1$. This is the Central Limit Theorem.

[2] See the Appendix for a definition of convergence in distribution.

[3] Specifically: let ϕ_n be the measures on $\mathscr{C}([0,1])$ associated with $S^{(n)}$ and let ϕ be the measure associated with $W^{(1)}$. Then for all bounded continuous functions $\phi : \mathscr{C}([0,1]) \to \mathbb{R}$, we have that $\int \phi\, d\mu_n \to \int \phi\, d\mu$.

Example 6.4 Let $\psi(f) = \max\{f(t) : t \in [0,1]\}$. Again ψ is continuous, and we have that

$$\max_{t\in[0,1]} S_t^{(n)} \Rightarrow M_1 \equiv \max_{t\in[0,1]} W_t\,.$$

Proving Donsker's Theorem is a fair amount of work; see e.g. Durrett (1996) p.287 or Karatzas and Shreve (1991) p.70. However, it is not hard to see why the finite-dimensional distributions should converge, at least heuristically. For the one-point distributions, the Central Limit theorem gives that $S_t^{(n)} \Rightarrow N(0,t)$ as $n \to \infty$ (with a little bit of care with the interpolated parts.) Therefore the one-point distributions of $S^{(n)}$ converge to the one-point distributions of Brownian motion.

For the two-point distributions, we need to consider the joint distribution of pairs of random variables of the form $\left(\frac{S_{nt}}{\sqrt{n}}, \frac{S_{ns}}{\sqrt{n}}\right)$ (again ignoring the interpolated parts here.) Using a similar technique as in the proof of the CLT (i.e. considering the characteristic functions) shows that such a random vector has a distribution which converges to $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\min(s,t) & 0\\ 0 & \min(s,t)\end{pmatrix}\right)$.

One can treat the k-point distributions similarly, see e.g. Karatzas and Shreve (1991), p. 67. Showing that the distribution of the entire process converges in some sense, for all values of t at once, requires tools beyond the Central limit theorem.

6.4 Properties of Brownian motion

Brownian motion has a number of useful and sometimes surprising properties, surveyed in this section.

6.4.1 Scaling properties

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

Property (iii) shows that Brownian motion is like a fractal: it looks statistically "the same" at all scales, no matter how much you zoom in, provided that space and time are scaled in the right way (again, we see the diffusive scaling space $\propto \sqrt{\text{time}}$.) This property follows naturally from the construction of Brownian motion as a limit of random walks.

Proof. (i), (ii), (iii) follow straightforwardly from Definition #1, by checking the required conditions are satisfied. For example, for (iii): let $X_t = c^{-1/2} W_{ct}$. Then

(i) $X_0 = c^{-1/2} W_0 = 0$.
(ii) X_t has independent increments – this is straightforward to check.
(iii) Normal increments: for $t \geq s$, $X_t - X_s = c^{-1/2}(W_{ct} - W_{cs}) \sim c^{-1/2} N(0, c(t-s)) \sim N(0, t-s)$.
(iv) Continuity – this follows from continuity of W_t.

To check (iv), we use Definition #2 of BM as a Gaussian process. We have that tW_t is Gaussian, with mean 0. It has covariance function $\mathbb{E} st W_{1/s} W_{1/t} = st\left(\frac{1}{s} \wedge \frac{1}{t}\right) = s \wedge t$. It is continuous for $t > 0$. It remains to check that it is continuous at 0. But $\lim_{t\to 0} tW_{1/t} = \lim_{s\to\infty} \frac{W_s}{s} \to 0$ a.s., by a result in the next section. $\square$

6.4.2 Behaviour as $t \to \infty$

There are several ways to characterize BM in the limit as $t \to \infty$:

Proposition. *(i)* $\lim_{t\to\infty} \frac{W_t}{t} = 0$ *a.s..*

(ii) $\limsup_{t\to\infty} \frac{W_t}{\sqrt{t}} = \infty, \quad \liminf_{t\to\infty} \frac{W_t}{\sqrt{t}} = -\infty$ *(both a.s.).*

(iii) (Law of the Iterated Logarithm)

$$\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = 1 \quad a.s., \qquad \limsup_{t\to 0+} \frac{W_t}{\sqrt{2t\log\log 1/t}} = 1 \quad a.s..$$

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

Proof. (i) (from Breiman (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 i.i.d. random variables. By the SLLN, $W_n/n \to 0$ a.s.. To obtain behaviour at non-integer t, let

$$Z_k = \max_{0\le t\le 1} |B(k+t) - B(k)|.$$

For $t \in [k, k+1]$,

$$\left|\frac{W_t}{t} - \frac{W_k}{k}\right| \le \frac{1}{k(k+1)}|W_k| + \frac{1}{k+1} Z_k.$$

The first term on the RHS $\to 0$ a.s., and Z_k has the same distribution as $\max_{0\le t\le 1}|W_t|$. It can be shown that $\mathbb{E}Z_k < \infty$ (see this week's homework!), and that this implies $Z_k/k \to 0$ a.s..

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

(iii) One only needs to show one of these limits, since they are related to each other by the time inversion property (iv) of BM. The proof is long; see e.g. Durrett (2005) section 7.9 p.431, or Breiman (1992), p. 263, or Karatzas and Shreve (1991), p. 112.

$\square$

6.4.3 Differentiability

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

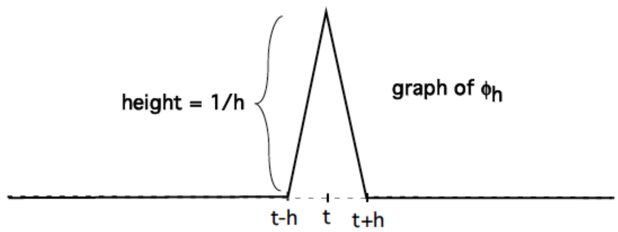

Figure 2: A sketch of $\phi_h(t) = \phi_h(s, s+t)$. From (Evans, 2013).

A rigorous proof is in the Appendix. Here is a heuristic explanation for why the derivative doesn't exist, at least at a single point. Suppose we try to calculate the derivative as

$$\xi_t = \frac{dW_t}{dt} = \lim_{h\to 0} \frac{W_{t+h} - W_t}{h} \ . \tag{6}$$

But $W_{t+h} - W_t \sim N(0, h)$ so $\frac{W_{t+h}-W_t}{h} \sim N(0, \frac{1}{h})$. This random variable doesn't converge to anything as $h \to 0$, since it is a Gaussian with a variance that blows up to infinity.

Another way to see that Brownian motion is not differentiable, is to argue that $X_h = W_{t+h} - W_t$ is a Brownian motion, by the translation property, and then that $Y_s = sX_{1/s}$ for $s = 1/h$ is a Brownian motion, by the time-inversion property. So then $\frac{dW_t}{dt} = \lim_{s\to\infty} Y_s$, which doesn't exist.

Of course, the theorem makes a much stronger argument, which is that the derivative doesn't exist *anywhere*, with probability 1. That is, for any given path, there is not even a single point at which a sample path of Brownian motion is differentiable!

Nevertheless, in the Physics literature it is common to speak of the derivative of Brownian motion, where it is called *white noise*. It turns out that even though white noise doesn't exist in a classical sense, it is possible to define it in a weak sense, which we'll do when we construct stochastic integrals. Let's pretend for a moment that the derivative of Brownian motion exists, and see what we can learn about it. If we calculate its mean and covariance function from (6) (e.g. Evans, 2013, p.41), we obtain:

$$\mathbb{E}\xi_t = \mathbb{E} \lim_{h\to 0} \frac{W_{t+h} - W_t}{h} = \lim_{h\to 0} \frac{\mathbb{E}W_{t+h} - \mathbb{E}W_t}{h} = 0\,,$$

and

$$\mathrm{Cov}(\xi_s, \xi_t) = \lim_{h\to 0} \mathbb{E}\left(\frac{W_{t+h} - W_t}{h}\right)\left(\frac{W_{s+h} - W_s}{h}\right) = \lim_{h\to 0} \phi_h(s,t)\,,$$

where

$$\phi_h(s,t) = \frac{1}{h^2}\left[(t+h)\wedge(s+h) - (t+h)\wedge s - t\wedge(s+h) + s\wedge t\right] = \begin{cases} 0 & \text{if } |s-t| > h \\ \frac{1}{h^2}(h - |s-t|) & \text{if } |s-t| \le h \end{cases} . \tag{7}$$

The function $\phi_h(s,t)$ is only a function of $s-t$, so we can also write it (with a slight abuse of notation) as $\phi_h(t)$. (See Figure 2 for a sketch of $\phi_h(t)$.) As $h \to 0$, the function becomes narrower and taller, but the area under the function remains constant, $\int \phi_h(t)dt = 1$. Therefore, we expect $\phi_h(t) \to \delta(t)$ as $h \to 0$, so, formally at least, we should have

$$\mathrm{Cov}(\xi_s, \xi_t) = \mathbb{E}\xi_s\xi_t = \delta(s-t)\,. \tag{8}$$

Therefore, ξ_t is weakly stationary, with covariance function $C(t) = \delta(t)$. Calculating the spectral density of the covariance function gives

$$f(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda t}\delta(t)\,dt \quad = \frac{1}{2\pi} \quad \text{for all } \lambda \in \mathbb{R}.$$

The spectral density is flat: all frequencies contribute equally. This is why ξ_t is called "white noise": just as white light is a superposition of light from all wavelengths, white noise is a superposition of random oscillations of all frequencies.

Here are some other facts about the sample path properties of Brownian motion.

Theorem. *(a) With probability 1, a Brownian sample path is locally Hölder continuous with exponent γ 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). □

6.5 Quadratic variation

Recall that the concept of total variation from analysis:

Definition. The *total variation* of a function $f(t)$ on an interval $[a,b]$ is defined by

$$V_{[a,b]}(f) = \sup_{\sigma}\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|\,, \tag{9}$$

where the supremum is over all partitions $\sigma = \{t_0, t_1, \ldots, t_n\}$ of $[a,b]$ with $a = t_0 < t_1 < \cdots < t_n = b$. If $V_{[a,b]}(f) < \infty$ then f is said to be of *bounded variation*, and if $V_{[a,b]}(f) = \infty$ then f is said to be of *infinite variation*.

If a function is of bounded variation on $[a,b]$, then it has a derivative almost everywhere on $[a,b]$ (i.e. except for a set of measure zero.) Conversely, if a function is nowhere differentiable, then it must have infinite variation on any interval.

Since Brownian motion is nowhere differentiable, it has infinite variation on any interval. However, it has finite *quadratic* variation (in a mean-square sense). This will turn out to be an important property when we construct the stochastic integral.

Definition. The *quadratic variation* of a function f on $[0,t]$ with respect to a partition σ is

$$Q_t^{\sigma}(f) = \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|^2. \tag{10}$$

We would like to define the quadratic variation of a Brownian motion W on $[0,t]$ as $\sup_\sigma Q_t^\sigma(W)$. However, for a stochastic processes we need to be careful with how the supremum over partitions is calculated, since not all types of stochastic convergence will give a finite result. We will use the following notion of convergence.

Definition. A sequence of random variables $X_1, X_2, \ldots$ *converges in mean-square* to another random variable X, written $X_n \xrightarrow{m.s.} X$ or $\text{m.s.}\lim_{n\to\infty} X_n = X$, if $\mathbb{E}|X_n - X|^2 \to 0$ as $n \to \infty$.

Remark. Convergence in mean-square implies convergence in probability, which in turn implies convergence in distribution. Convergence in mean-square does *not* imply almost sure convergence nor vice versa. (Almost sure convergence does imply convergence in probability and hence convergence in distribution.)

Lemma. *Let* $|\sigma| = \max_{0\le i\le n-1} |t_{i+1} - t_i|$. *The quadratic variation of Brownian motion* $Q_t^\sigma(W)$ *converges in mean-square to* t *as* $|\sigma| \to 0$*:*

$$\sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|^2 \xrightarrow{m.s.} t\,.$$

We can write this result formally as $(\Delta W)^2 = \Delta t$. We see the diffusive scaling yet again.

Proof. (From Koralov and Sinai (2010), p. 269.) Write $\Delta W_i = W_{t_{i+1}} - W_{t_i}$, and $\Delta t_i = t_{i+1} - t_i$. Then

$$\mathbb{E}\Big(Q_t^\sigma(W) - t\Big)^2 = \mathbb{E}\left(\sum_{i=0}^{n-1} \Delta W_i^2 - \Delta t_i\right)^2 = \sum_{i=0}^{n-1} \mathbb{E}(\Delta W_i^2 - \Delta t_i)^2 + \sum_{i,j=0,i\neq j}^{n-1} \mathbb{E}(\Delta W_i^2 - \Delta t_i)(\Delta W_j^2 - \Delta t_j)\,.$$

Now, we have

$$\mathbb{E}(\Delta W_i^2 - \Delta t_i)(\Delta W_j^2 - \Delta t_j) = \mathbb{E}\Delta t_i \Delta t_j - \Delta t_i \Delta t_j - \Delta t_i \Delta t_j + \Delta t_i \Delta t_j\,,$$

since $\Delta W_i, \Delta W_j$ are independent for $i \neq j$. Therefore

$$\begin{aligned}
\mathbb{E}\Big(Q_t^\sigma(W) - t\Big)^2 &= \sum_{i=0}^{n-1} \mathbb{E}(\Delta W_i^2 - \Delta t_i)^2 \\
&\le \sum_{i=0}^{n-1} \mathbb{E}\Delta W_i^4 + \Delta t_i^2 && \text{since } (a-b)^2 \le a^2 + b^2 \text{ for } a,b \ge 0 \\
&= 4\sum_{i=0}^{n-1} \Delta t_i^2 && \text{since } \mathbb{E}(W_t - W_s)^4 = 3|t-s|^2 \\
&\le 4 \max_{0\le i\le n-1} (t_{i+1} - t_i) \sum_{i=0}^{n-1} (t_{i+1} - t_i) \\
&= 4t|\sigma| \quad \to 0 \ \text{ as } |\sigma| \to 0.
\end{aligned}$$

□

6.6 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

$$P(W_t \in F|W_{t_{n-1}}{=}x_{n-1},\ldots,W_{t_0}{=}x_0) = P(W_{t_n} \in F_n|W_{t_{n-1}}{=}x_{n-1}), \tag{11}$$

where $t_0 < t_1 < \cdots < t_n$, and conditioning on points is defined as $P(W_t \in F|W_s = y) = \lim_{\varepsilon\searrow 0} P(W_t \in F|W_s \in [y, y+\varepsilon))/\varepsilon$, and similarly for conditioning on multiple points.

Transition density To describe the transition probabilities for a Markov process with a discrete state space, we had a transition matrix $P(t)$ or $P(s,t)$. For a process with a continuous state space, we need a *transition density* $p(y,t|x,s)$, which is the function such that

$$P(X_t \in A|X_s = x) = \int_{x\in A} p(y,t|x,s)dx\,.$$

We calculated earlier (see (2)) that the transition density for Brownian motion is

$$p(y,t|x,s) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(y-x)^2}{2(t-s)}}\,.$$

This function is *time-homogeneous*: $p(y,t|x,s) = p(y,t-s|x,0)$ for all t,s. It satisfies the *Chapman-Kolmogorov equations*:

$$p(y,t|x,s) = \int_{z\in\mathbb{R}} p(y,t|z,u)p(z,u|x,s)dz\,, \qquad s < u < t\,. \tag{12}$$

You can check this relation by direct calculation using the transition densities for Brownian motion. It also holds for a general Markov process with transition density $p(y,t|x,s)$ (see Pavliotis (2014), p.35 for a formal proof).

Let's consider how probabilities and expectations for a continuous process, are related to those of a discrete process. Consider expectations first. For a discrete Markov process, we calculated expectations as $\mathbb{E}_x f(X_t) = P(0,t)u^{(0)}$, where $u^{(0)} = (f(1), f(2), \ldots)^T$. That is, we multiplied the transition matrix by a vector on the right, summing over the rows. For a continuous Markov process, we calculate expectations as

$$\mathbb{E}_x f(X_t) = \int_y f(y)p(y,t|x,0)dy\,.$$

That is, we integrate a function f, against the first variable of the transition density.

Now consider probabilities. For a discrete Markov process, we calculated the probability distribution at time t as $\mu(t) = \mu(t)P(t)$: we multiplied the transition matrix on the left by a row vector, summing over the columns. For a continuous process, the probability density at time t, $\rho(y,t)dy = P(X_t \in [y, y+dy))$, is obtained from the initial density ρ_0 as

$$\rho(y,t) = \int_x p(y,t|x,0)\rho_0(x)dx\,.$$

Now we integrate against the second variable in the transition density. So in the density $p(y,t|x,s)$, the variable x is analogous to the columns of the transition matrix, and y is analogous to the rows of the transition matrix.

Infinitesimal generator Recall that the fundamental quantity that characterized a continuous-time Markov chain was its generator. From the generator we obtained the forward and backward equations, describing how probability and statistics respectively evolve. Let's look at how these ideas generalize to a continuous Markov process, focusing on the specific case of Brownian motion.

Definition. The *infinitesimal generator* of a Markov process is the operator $\mathscr{L}$ acting on functions in $L^\infty = \{f : \|f\| < \infty\}$ with norm $\|f\| = \sup_x |f(x)|$, is defined by

$$(\mathscr{L}f)(x) = \lim_{t\to 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t}, \tag{13}$$

The set $D(\mathscr{L}) \subset L^\infty$ on which this limit exists is the *domain* of $\mathscr{L}$.

Remark. The convergence above is understood as the norm convergence, i.e.

$$\lim_{t\to 0} \Big\| \frac{\mathbb{E}_x f(X_t) - f}{t} - g(x) \Big\| = 0$$

for some function g, which is identified as $\mathscr{L}f$.

Let's calculate the generator of Brownian motion (following Varadhan, 2007). We have that

$$\mathbb{E}_x f(X_t) = \int_{-\infty}^{\infty} f(y) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} dy = \int_{-\infty}^{\infty} f(x+z\sqrt{t}) \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\,.$$

The infinitesimal generator is

$$(\mathscr{L}f)(x) = \lim_{t\to 0} \int_{-\infty}^{\infty} \frac{f(x+z\sqrt{t}) - f(x)}{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\,. \tag{14}$$

Suppose f is bounded and has three bounded derivatives, so we can expand it using Taylor's formula as

$$f(x+z\sqrt{t}) - f(x) = z\sqrt{t} f'(x) + \frac{tz^2}{2} f''(x) + t^{3/2} R(t,z)$$

where the remainder term satisfies $|R(t,z)| \le C|z|^3$ for some constant C. Substituting this expansion into (14) and calculating the integral directly shows that

$$\mathscr{L}f = \frac{1}{2} f'' = \frac{1}{2} \frac{d^2 f}{dx^2}\,. \tag{15}$$

The generator of Brownian motion is the Laplacian operator.

Remark. Note the analogy to the generator for a continuous-time Markov chain: recall the generator was a matrix, acting on vectors. Here, the generator is a functional, but it is still 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. For diffusion processes the functional is a partial differential operator; for processes containing jumps the generator is an integral operator. The *Hille-Yosida theorem* provides the conditions for a closed linear operator $\mathscr{L}$ on a Banach space to be the infinitesimal generator of a Markov process.

Forward and backward equations Now let's use the generator to obtain the backward equation. Let $u(x,t) = \mathbb{E}_x f(X_t)$. We seek an evolution equation for u. It will be convenient to introduce the operator

$$T_t f(x) = \mathbb{E}_x f(X_t) = \int_y f(y) p(y,t|x,0) dy \,, \tag{16}$$

so that $T_t f(x) = u(x,t)$. The domain of T_t is the set of bounded, measurable functions $f : \mathbb{R} \to \mathbb{R}$. We claim that

$$T_{s+t} f = (T_s \circ T_t) f \tag{17}$$

Remark. This is the *semigroup property* for the set of operators $\{T_t\}_{t\geq 0}$, and hence the set is called the *transition semigroup* (Durrett, 1996, Chapter 7), (Varadhan, 2007). The transition semigroup is obtained from the generator as $T_t = e^{\mathscr{L} t} \equiv I + t\mathscr{L} + \frac{t^2}{2}\mathscr{L}^2 + \cdots$.

Property (17) is a succinct form of the Chapman-Kolmogorov equations (12); notice the relationship to the Chapman-Kolmogorov equations $P(s+t) = P(s)P(t)$ for Markov chains. To show it, we calculate

$$\begin{aligned}
T_{s+t} f = \mathbb{E}_x f(X_{t+s}) &= \int_y f(y) p(y,t+s|x,0) dy \\
&= \int_y \int_z f(y) p(y,t+s|z,s) p(z,s|x,0) dz dy && \text{Chapman-Kolmogorov equations (12)} \\
&= \int_y \int_z f(y) p(y,t|z,0) p(z,s|x,0) dz dy && \text{time-homogeneity} \\
&= \int_z p(z,s|x,0) \left[\int f(y) p(y,t|z,0) dy \right] dz && \text{Fubini's theorem} \\
&= T_s(T_t f).
\end{aligned}$$

We use this property to derive the backward equation:

$$\begin{aligned}
\frac{\partial u(x,t)}{\partial t} &= \lim_{h\to 0} \frac{\mathbb{E}_x f(X_{t+h}) - \mathbb{E}_x f(X_t)}{h} = \lim_{h\to 0} \frac{T_{t+h} f - T_t f}{h} \\
&= \lim_{h\to 0} \frac{(T_h \circ T_t) f - T_t f}{h} = \lim_{h\to 0} \frac{T_h u - u}{h} \\
&= \mathscr{L} u \,.
\end{aligned}$$

We obtain the *Kolmogorov backward equation* for a continuous Markov process, $\frac{\partial u(x,t)}{\partial t} = \mathscr{L} u$. Therefore, the backward equation for Brownian motion is the heat equation,

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \,, \qquad u(x,0) = f(x) \,. \tag{18}$$

You can also verify this directly for Brownian motion, by computing derivatives of its transition density, since (formally at least),

$$\frac{\partial u}{\partial t} = \int_y f(y) \frac{\partial}{\partial t} p(y,t|x,0) dy = \int_y f(y) \frac{1}{2} \frac{\partial^2}{\partial x^2} p(y,t|x,0) dy = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \,.$$

Now let's consider how the probability density $\rho(x,t)$ evolves, assuming some initial density $\rho(x,0) = \rho_0(x)$. Our argument will be purely formal; it needs assumptions on each of the functions involved to be made rigorous.

First, notice that by time-homogeneity, $p(y,t|x,s) = p(y,t-s|x,0)$, and therefore $\partial_t p(y,t|x,s) = -\partial_s p(y,t|x,s)$.

Next, let $T > 0$, let $f \in D(\mathscr{L})$, and let

$$v(x,s) = \mathbb{E}_{X_s=x} f(X_T)\,.$$

One can verify that v satisfies the backward equation $\partial_s v = -\mathscr{L} v$ with $v(x,T) = f(x)$.

Consider the inner product $\langle \rho(x,s) v(x,s) \rangle$. Our strategy will be to show this inner product is constant in s, and therefore since v satisfies the backward equation, ρ must satisfy its adjoint, the forward equation.

For the inner product, we have

$$\langle \rho(x,s) v(x,s) \rangle = \iiint f(y) p(y,T|x,s) p(x,s|z,0) \rho_0(z) dz dy dx = \iint f(y) p(y,T|z,0) \rho_0(z) dz dy\,.$$

We moved x to the inner integral and used the Chapman-Kolmogorov equations. Therefore $\langle \rho(x,s) v(x,s) \rangle$ is independent of s. Taking $\frac{d}{ds}$ we have

$$\int \partial_s v(x,s) \rho(x,s) dx = -\int v(x,s) \partial_s \rho(x,s) dx\,.$$

By the backward equation, we have

$$\int \partial_s v(x,s) \rho(x,s) dx = \int \mathscr{L} v(x,s) \rho(x,s) dx = \int v(x,s) \mathscr{L}^* \rho(x,s) dx\,,$$

where $\mathscr{L}^*$ is the formal adjoint of $\mathscr{L}$, i.e. the operator such that $\langle \mathscr{L} f, g \rangle = \langle f, \mathscr{L}^*, g \rangle$, for all for all functions $f \in D(\mathscr{L})$ and $g \in L^1$, and where s$\langle f, g \rangle = \int_{\mathbb{R}} f(x) g(x) dx$ is the L^2 inner product. Therefore

$$\int v(x,s) \mathscr{L}^* \rho(x,s) dx = \int v(x,s) \partial_s \rho(x,s) dx\,.$$

Choosing $T = s$ (which affects the definition of v but not ρ), so that $v(x,s) = v(x,T) = f(x)$, and then we have that $\int f(x) \mathscr{L}^* \rho(x,s) dx = \int f(x) \partial_s \rho(x,s) dx$. By choosing $f(x)$ to approximate a δ-function, we have that

$$\partial_s \rho(x,s) = \mathscr{L}^* \rho(x,s)\,.$$

This is the *Kolmogorov forward equation* for a continuous Markov process. For Brownian motion, $\mathscr{L} = \mathscr{L}^* = \frac{1}{2}\frac{\partial^2}{\partial x^2}$, so the forward equation is

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2}\,, \qquad \rho(x,0) = \rho_0(x)\,.$$

In terms of the transition probabilities $p(x,t|y,s)$ the forward equation is

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \qquad p(x,0|y,0) = \delta(x-y).$$

You can check that the latter equation holds directly from the explicit form of $p(x,t|y,s)$. In general, you won't have an explicit formula for the transition probabilities, so you will obtain them by solving the corresponding PDEs.

References

Breiman, L. (1992). *Probability*. SIAM.

Durrett, R. (1996). *Stochastic Calculus: A practical introduction*. CRC Press, Taylo & Francis Group.

Durrett, R. (2005). *Probability: Theory and Examples*. Thomson, 3rd edition.

Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 322(8):549–560.

Evans, L. C. (2013). *An Introduction to Stochastic Differential Equations*. American Mathematical Society.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer.

Koralov, L. B. and Sinai, Y. G. (2010). *Theory of Probability and Random Processes*. Springer.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications*. Springer.

Varadhan, S. R. S. (2007). *Stochastic Processes*, volume 16 of *Courant Lecture Notes*. American Mathematical Society.

Von Smoluchowski, M. (1906). Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik*, 326(14):756–780.

6.7 Appendix

6.7.1 Convergence in distribution

Here is the definition of convergence in distribution (see e.g. Durrett (1996), Chapter 8, Theorem 1.1.)

Definition. Consider a sequence of random variables $X_1, X_2, \dots$ defined on a sequence of probability spaces $\{(\Omega_n, \mathscr{F}_n, P_n)\}_{n=1}^{\infty}$ and taking values in some metric space (S, ρ). Let $(\Omega, \mathscr{F}, P)$ be another probability space, on which another random variable X is defined, which takes values in (S, ρ). Then $\{X_n\}_{n=1}^{\infty}$ *converges in distribution*, or *converges weakly* to X, written $X_n \xrightarrow{d} X$, if $\mathbb{E}_n f(X_n) \to \mathbb{E} f(X)$ 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.

If X_n, X are real-valued (i.e. not path-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.

6.7.2 Brownian motion is not differentiable

Proof of the non-differentiability of Brownian motion. This proof is from Breiman (1992), p. 261, in turn from Dvoretsky, Erdös, 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 $n > n_0$,

$$|x(t) - x(s)| \leq 2\beta|t-s|, \quad \text{if } ||t-s| \leq 2/n. \tag{19}$$

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 β in absolute value. If (19) holds, then let k be the largest integer such that $k/n \leq s$, so that

$$y_k = \max\left\{\left|W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}}\right|, \left|W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right|, \left|W_{\frac{k}{n}} - W_{\frac{k-1}{n}}\right|\right\} \leq \frac{6\beta}{n}.$$

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

$$\begin{aligned}
P(C_n) &\leq \sum_{k=1}^{n-2} P\left(\max\left\{\left|W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}}\right|, \left|W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right|, \left|W_{\frac{k}{n}} - W_{\frac{k-1}{n}}\right|\right\} \leq \frac{6\beta}{n}\right) \\
&\leq nP\left(\max\left\{\left|W_{3/n} - W_{2/n}\right|, \left|W_{2/n} - W_{1/n}\right|, \left|W_{1/n}\right|\right\} \leq \frac{6\beta}{n}\right) \\
&= nP\left(\left|W_{1/n}\right| \leq \frac{6\beta}{n}\right)^3 \\
&= n\left(\sqrt{\frac{n}{2\pi}}\int_{-6\beta/n}^{6\beta/n} e^{-nx^2/2}dx\right)^3 \\
&= n\left(\frac{1}{\sqrt{2\pi n}}\int_{-6\beta}^{6\beta} e^{-x^2/2n}dx\right)^3
\end{aligned}$$

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

$\square$