

2 Brownian Motion

We begin with Brownian motion for two reasons. First, it is an essential ingredient in the definition of the Schramm-Loewner evolution. Second, it is a relatively simple example of several of the key ideas in the course - scaling limits, universality, and conformal invariance.

The article by Kager and Nienhuis has an appendix on probability and stochastic processes (Appendix B). It includes a couple of pages on Brownian motion. Lawler's book and Werner's St. Flour article assume the reader is familiar with Brownian motion. For this chapter, I am following two books: Chapter 7 of *Probability: Theory and Examples* by Richard Durrett and chapter 2 of *Brownian Motion and Stochastic Calculus* by Ioannis Karatzas and Steven Shreve.

2.1 Definition and properties

We recall a basic construction from probability theory. Let (Ω, \mathcal{F}, P) be a probability space, i.e., a measure space with $P(\Omega) = 1$. Let X_1, X_2, \dots, X_m be random variables, i.e., measurable functions. Then we can define a Borel measure μ on \mathbb{R}^m by

$$\mu(B) = P((X_1, X_2, \dots, X_m) \in B) \quad (6)$$

where B is a Borel subset of \mathbb{R}^m . One can then prove that for a function $f(x_1, x_2, \dots, x_m)$ which is integrable with respect to μ , we have

$$Ef(X_1, X_2, \dots, X_m) = \int_{\mathbb{R}^m} f(x_1, x_2, \dots, x_m) d\mu \quad (7)$$

Of course, this measure depends on the random variables; when we need to make this explicit we will write it as μ_{X_1, \dots, X_n} .

The random variables X_1, X_2, \dots, X_m are said to be independent if the measure μ_{X_1, \dots, X_n} equals the product of the measures $\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_m}$. Two collections of random variables (X_1, \dots, X_m) and (Y_1, \dots, Y_m) are said to be equal in distribution if $\mu_{X_1, \dots, X_n} = \mu_{Y_1, \dots, Y_n}$.

We now turn to Brownian motion. It is a continuous time stochastic process. This means that it is a collection of random variables X_t indexed by a real parameter t .

Definition 1 *A one-dimensional (real valued) Brownian motion is a stochastic process $B_t, t \geq 0$, with the following properties.*

(i) *If $t_0 < t_1 < t_2 < \dots < t_n$, then $B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ are independent random variables.*

(ii) *If $s, t \geq 0$, then $B_{t+s} - B_s$ has a normal distribution with mean zero and variance t .*
So

$$P(B_{t+s} - B_s \in A) = \int_A (2\pi t)^{-1/2} \exp(-x^2/2t) dx \quad (8)$$

where A is a Borel subset of the reals.

(iii) With probability one, $t \rightarrow B_t$ is continuous.

In short, Brownian motion is a stochastic process whose increments are independent, stationary and normal, and whose sample paths are continuous. Increments refer to the random variables of the form $B_{t+s} - B_s$. Stationary means that the distribution of this random variable is independent of s . Independent increments means that increments corresponding to time intervals that do not overlap are independent. Proving that such a process exists is not trivial, but we will not give the proof. The above definition makes no mention of the underlying probability space Ω . One can take it to be the set of continuous functions $\omega(t)$ from $[0, \infty)$ to \mathbb{R} with $\omega(0) = 0$. Then the random variables are given by $B_t(\omega) = \omega(t)$. Unless otherwise stated, we will take $B_0 = 0$. We list some standard consequences of the above properties.

Theorem 1 *If B_t is a Brownian motion then*

(a) B_t is a Gaussian process, i.e., for any times t_1, \dots, t_n , the distribution of B_{t_1}, \dots, B_{t_n} has a multivariate normal distribution.

(b) $EB_t = 0$ and $EB_s B_t = \min\{s, t\}$.

(c) Let $\alpha > 0$. Define a stochastic process X_t by

$$X_t = \alpha^{-2} B_{\alpha t} \tag{9}$$

for $t \geq 0$. Then X_t is a Brownian motion.

(d) Define

$$p(t, x, y) = (2\pi t)^{-1/2} \exp\left(\frac{-(x-y)^2}{2t}\right) \tag{10}$$

Then for Borel subsets A_1, A_2, \dots, A_n of \mathbb{R} ,

$$P(B_{t_1} \in A_1, B_{t_2} \in A_2, \dots, B_{t_n} \in A_n) = \int_{A_1} dx_1 \int_{A_2} dx_2 \cdots \int_{A_n} dx_n p(t_1, 0, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n)$$

Exercise: Prove parts (b) and (c) of the above. Hint for (b): If random variables X and Y are independent, then $E XY = EX EY$. For $s > t$, write B_s as $(B_s - B_t + B_t)$. The ambitious reader is welcome to prove parts (a) and (d) as well.

The definition of d -dimensional Brownian motion is easy. We take d independent copies of one-dimensional Brownian motion, and label them as $B_t^1, B_t^2, \dots, B_t^d$. Then $(B_t^1, B_t^2, \dots, B_t^d)$ is a d -dimensional Brownian motion. We can also think of the two-dimensional Brownian motion (B_t^1, B_t^2) as a complex valued Brownian motion by considering $B_t^1 + iB_t^2$.

The paths of Brownian motion are continuous functions, but they are rather rough. With probability one, the Brownian path is not differentiable at any point. If $\gamma < 1/2$,

then with probability one the path is Hölder continuous with exponent γ . But if $\gamma > 1/2$, then the path is not Hölder continuous with exponent γ . For any interval (a, b) , with probability one the path is neither increasing or decreasing on (a, b) . With probability one the path does not have bounded variation. This last fact is important because it says that one cannot use the Riemann-Stieltjes integral to define integration with respect to B_t .

One of the key tools in the stochastic calculus which we will learn about later is the Ito formula. It is often summarized in the statement $(dB)^2 = dt$. The following proposition is in that spirit.

Proposition 1 *Fix $t > 0$. For each n let P_n be a partition of $[0, t]$. Let $\|P_n\|$ be the width of the largest subinterval in the partition, and suppose that $\|P_n\| \rightarrow 0$. (For example, we could let P_n consist of n intervals of width t/n .) For a partition $P = \{t_0, t_1, \dots, t_m\}$ we define*

$$\Delta_i = B_{t_i} - B_{t_{i-1}} \quad (11)$$

and then define

$$X(P) = \sum_{i=1}^m \Delta_i^2 \quad (12)$$

Then with probability one,

$$\lim_{n \rightarrow \infty} X(P_n) = t \quad (13)$$

We do not give a proof, but we note that a particular case of the theorem follows from the law of large numbers. We leave the details to the reader:

Exercise: Take P_n to be the uniform partition with n subintervals. Use the law of large numbers to prove (13) for this special case.

2.2 Brownian motion as scaling limit of random walks

A sequence of random variables X_n is independent if every finite subset is independent. It is identically distributed if each random variable has the same distribution, i.e., $P(X_n \in A) = P(X_m \in A)$ for any Borel set A and any n, m . A sequence of random variables which is both independent and identically distributed is called an i.i.d. sequence. Note that for an identically distributed sequence, the random variables all have the same mean and variance.

A one-dimensional random walk is defined as follows. Let X_n be an i.i.d. sequence of random variables. We assume that the mean, EX_n , is zero and the variance is finite. It is no loss of generality to take the variance to be 1, so $EX_i^2 = 1$. Let

$$S_m = \sum_{i=1}^m X_i \quad (14)$$

This is a rather general random walk in the sense that it allows very general steps. The simplest random walk is to just take X_i to be ± 1 with probability $1/2$. We can picture this as follows. We start at the origin and flip a fair coin. For heads we take a step forward, for tails a step backwards. We repeat this. Then S_m is our position after m steps. In this simple case S_m is always an integer, so the random walk lives on the lattice \mathbb{Z} .

The S_m form a discrete time stochastic process. We make this into a continuous time stochastic process by linear interpolation. More precisely,

$$S_t = \begin{cases} S_t & \text{if } t \text{ is an integer} \\ \text{linear on } [m, m+1] & \text{if } t \in [m, m+1] \end{cases} \quad (15)$$

Since the variance of a sum of independent random variables is the sum of their individual variances, the variance of S_m is m . So the typical size of S_t is \sqrt{t} . This motivates the following rescaling. For each positive integer n , we let

$$S_t^n = n^{-1/2} S_{nt} \quad (16)$$

If we picture a graph of S_t , then to get S_t^n we shrink the horizontal (time) axis by a factor of n and shrink the vertical (space) axis by a factor of \sqrt{n} . Note that for t which are equal to an integer divided by n , the variance of S_t^n is t .

Now consider times $0 < t_1 < t_2 < \dots < t_m$ where each time is equal to some integer divided by n . Consider the random variables $S_{t_1}, S_{t_2} - S_{t_1}, \dots, S_{t_m} - S_{t_{m-1}}$. Each of them is a sum of a subset of the X_i and no X_i appears in more than one of these sums. Thus these random variables are independent. If n is large, each of the random variables is the sum of a large number of i.i.d. random variables and so is approximately normal. So S_t^n is looking like Brownian motion, at least at the times which are multiples of $1/n$. So we can hope that as $n \rightarrow \infty$, S_t^n will converge to Brownian motion. This is indeed a theorem, proved by Donsker in 1951 and sometimes called the *invariance principle*. To state it in its strongest form requires a definition about convergence of measures. We start by stating a weaker form that is a bit easier to digest.

Theorem 2 (*invariance principle*) *Fix times $0 < t_1 < t_2 < \dots < t_m$. We use E^{rw} to denote expectation with respect to the probability measure for the original i.i.d. sequence X_i . Let X_t be a Brownian motion. We use E^{bm} to denote expectation with respect to its probability measure. Then for every bounded continuous function $f(x_1, x_2, \dots, x_m)$ on \mathbb{R}^m , we have*

$$\lim_{n \rightarrow \infty} E^{rw} f(S_{t_1}^n, S_{t_2}^n, \dots, S_{t_m}^n) = E^{bm} f(X_{t_1}, X_{t_2}, \dots, X_{t_m}) \quad (17)$$

This is already a pretty good theorem and the following somewhat technical discussion is only to get a stronger statement of the above and can be skipped without a big loss. The technical stuff ends where we consider how Brownian motion illustrates the ideas of scaling limits, critical phenomena and universality.

Definition 2 *Suppose that the sample space Ω is a metric space. Suppose that P_n is a sequence of probability measures on Ω defined on the Borel subsets. Let P be another such probability measure. We say that P_n converges weakly to P if*

$$\lim_{n \rightarrow \infty} \int f dP_n = \int f dP \quad (18)$$

for every bounded, continuous real-valued function f on Ω .

Now look at the conclusion of the theorem. For each n let μ_n be the probability measure on \mathbb{R}^m that comes from the random variables $S_{t_1}^n, S_{t_2}^n, \dots, S_{t_m}^n$. Let μ be the probability measure on \mathbb{R}^m that comes from $X_{t_1}, X_{t_2}, \dots, X_{t_m}$. Then the conclusion of the above theorem is that μ_n converges weakly to μ . A probabilist says that the sequence of random vectors $(S_{t_1}^n, S_{t_2}^n, \dots, S_{t_m}^n)$ *converges in distribution* to $(X_{t_1}, X_{t_2}, \dots, X_{t_m})$. And the conclusion of the above theorem is that the finite dimensional distributions of S_t^n converge in distribution to those of Brownian motion.

The stronger form of the theorem does not just look at the process at a finite set of times. Let $C[0, \infty)$ be the space of continuous functions on $[0, \infty)$. We let P denote the probability measure on this space for Brownian motion. For each n , S_t^n is a continuous function of t . So S_t^n also defines a probability measure on $C[0, \infty)$. We denote it by P_n . It is supported on piecewise linear functions.

Theorem 3 *(Invariance principle of Donsker) Let X_i be an i.i.d. sequence of random variables defined on the probability space (Ω, \mathcal{F}, P) . Suppose that they have mean zero and variance 1. Define S_t^n by the linear interpolation and scaling defined above, and let P_n be the probability measure on $C[0, \infty)$ induced by the process S_t^n . Then P_n converges weakly to a probability measure P for which $B_t(\omega) = \omega(t)$ is standard one-dimensional Brownian motion.*

We now consider how Brownian motion illustrates the ideas of scaling limits, critical phenomena and universality. We start with the scaling limit. Usually in statistical physics one starts with a model defined on a lattice and then tries to understand what the scaling limit is. If we take $X_i = \pm 1$ with equal probability, then the random walk stays on the lattice \mathbb{Z} . The scaling limit is what we did above when we shrunk time by a factor of n and space by a factor of \sqrt{n} . For this model we have a candidate for the scaling limit (Brownian motion) and a theorem that says the scaling limit is indeed equal to Brownian motion. This is not the typical situation in statistical physics. There we are lucky if we have an explicit candidate for the scaling limit and extremely lucky if we have a theorem that says the scaling limit does converge to the candidate. What is exciting about SLE is that it defines in a fairly explicit way candidates for scaling limits, and for some models we even have a theorem.

Now consider universality. The invariance principle is a very strong form of universality. It says that we can start with any random walk, subject only to the conditions that

the steps have mean zero and variance 1, and the scaling limit will converge to the same stochastic process, i.e., Brownian motion. We have stated the invariance principle only for one dimension. But it is true in any number of dimensions. For example, we can take a random walk on the lattice \mathbb{Z}^d which at each step moves by $\pm e_i$ with probability $1/2d$ where e_i is the unit vector in the i th coordinate direction. We then take a scaling limit as we did above. This will converge to a d -dimensional Brownian motion. (I am ignoring a slight rescaling that needs to be done here.)

Finally we consider criticality. In the scaling limit the steps of the random walk are of size $1/\sqrt{n}$. So the random walk is formed by combining infinitely many microscopic random inputs. The result, Brownian motion, is clearly random. So it appears that Brownian motion is a critical phenomena. This is a bit confusing from the viewpoint of statistical physics. Usually in a statistical physics model one must adjust a parameter, e.g., the temperature, to a particular value to make the model have critical behavior. There appears to be no such parameter in the random walk model. In some sense the condition that the mean of the step X_i must be zero plays the role of adjusting a parameter to make the model critical. Consider a one-dimensional random walk with steps of ± 1 , but now take $X_i = 1$ with probability p and $X_i = -1$ with probability $1 - p$ with $p \neq 1/2$. Now the typical size of S_n is n , not \sqrt{n} as before. So to construct a scaling limit we must define

$$S_t^n = n^{-1} S_{nt} \quad (19)$$

Now in the scaling limit, S_t^n will converge to a straight line with a slope which depends only on p . So the scaling limit has no randomness at all.

Exercise: For $p \neq 1/2$, find the slope m of the line to which (19) converges. Prove that for $t > 0$,

$$\lim_{n \rightarrow \infty} S_t^n = mt \quad (20)$$

with probability one. Hint: law of large numbers.

2.3 Conditional expectation

This subsection is not primarily about Brownian motion. We are going to define the conditional expectation of a random variable with respect to a σ -field. This is needed in the next section on the Markov property of Brownian motion. It is an essential part of the definition of a martingale, a concept that we will use throughout the course.

The material in this subsection is covered in Math 563a. See Prof. Watkins notes or any graduate level probability book for proofs. The student who has not seen conditional probability or conditional expectation at an undergraduate level should take a look at an undergrad probability book, e.g., *The Essentials of Probability* by Richard Durrett, *Elementary Probability* by David Stirzaker or *Probability: An introduction* by Geoffrey Grimmett and Dominic Welsh.

To develop some intuition we will first look at the conditional expectation of one random variable given another. We start with the notion of conditional probability. Let A be an event with $P(A) > 0$. The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (21)$$

What does $P(A|B)$ mean? This is the probability of A if we know for certain that the outcome is in B . Put another way, if we do the experiment N times, keep only the outcomes that belong to B , then the fraction of these outcomes that belong to A converges to $P(A|B)$ as $N \rightarrow \infty$.

If we fix the given event B and think of this conditional probability as a function of A , $A \rightarrow P(A|B)$, then this defines a new probability measure. So given a random variable X , we can compute the integral of X with respect to this new probability measure. This is called the conditional expectation of X given B and written as $E(X|B)$. For example, consider a discrete random variable X , i.e., a random variable that only takes on a countable set of values. Let x_1, x_2, \dots be the values that X takes on. Then the integral of X with respect to a general measure μ is

$$\int X d\mu = \sum_i x_i \mu(E_i) \quad (22)$$

where $E_i = X^{-1}(x_i)$. So

$$E(X|B) = \sum_i x_i P(E_i|B) \quad (23)$$

Probabilists usually write the event $X^{-1}(x_i)$ as just $X = x_i$. So

$$E(X|B) = \sum_i x_i P(X = x_i|B) \quad (24)$$

If we do the experiment N times, keep only the outcomes that belong to B , and average the value of X we get for these outcomes, then this average will converge to $E(X|B)$ as $N \rightarrow \infty$.

Now let Y be another discrete random variable whose possible values are y_1, y_2, \dots . Then for each i , $Y = y_i$ is an event. We can use the above to define $E(X|Y = y_i)$. Now think of this as a function of y_i . More precisely, let $\psi(y_i) = E(X|Y = y_i)$ and define $\psi(y) = 0$ when y is not one of the values y_i . Now we define a random variable by $\omega \rightarrow \psi(Y(\omega))$. This random variable is called the conditional expectation of X given Y and denoted $E(X|Y)$. Note that for an event B , $E(X|B)$ is a number, but for a random variable Y , $E(X|Y)$ is another random variable.

This is a rather convoluted definition, so some intuition is in order. Let B_i denote the event $Y = y_i$. Note that the B_i are disjoint and their union is the full sample space Ω .

So they form a partition of Ω . We can write the conditional expectation of X given Y as

$$E(X|Y) = \sum_i E(X|B_i)1_{B_i} \quad (25)$$

Here 1_B denotes the random variable which equals 1 when the outcome is in B and equal 0 when it is not in B . Note that with $\psi : \mathbb{R} \rightarrow \mathbb{R}$ defined as above, then $E(X|Y) = \psi(Y)$. In other words, the random variable $E(X|Y)$ is a function of Y . Of course, there are a zillion functions of Y . What is special about the function $E(X|Y)$? It is the function of Y which best approximates X in the following sense.

Proposition 2 For any Borel function $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$E[E(X|Y) - X]^2 \leq E[f(Y) - X]^2 \quad (26)$$

Exercise: Prove the above proposition.

Let (Ω, \mathcal{F}, P) be a probability space. Recall that a random variable X is a measurable function, i.e., for every Borel subset B of the reals, $X^{-1}(B)$ must belong to \mathcal{F} . In analysis one typically has a single σ field. In probability there is often more than one σ field. So one cannot just say measurable, one must specify the σ -field. We will abbreviate “ X is measurable with respect to \mathcal{F} ” by just $X \in \mathcal{F}$.

Given a random variable Y , the collection of events $\{Y \in A\}$, for Borel subsets A of \mathbb{R} is a σ -field which we denote by $\sigma(Y)$ and call “the σ -field generated by Y .” Obviously, Y is measurable with respect to the σ -field it generates, i.e., $Y \in \sigma(Y)$.

Exercise: Let Y be a discrete random variable with values y_1, y_2, \dots . Let B_i be the event $Y = y_i$. Show that the following are equivalent for a random variable X .

- (i) X is measurable with respect to $\sigma(Y)$
- (ii) X is constant on each B_i , i.e., it is of the form

$$X = \sum_i c_i 1_{B_i} \quad (27)$$

for some real numbers c_i .

- (iii) There is a Borel function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ such that $X = \psi(Y)$.

Thus for two discrete random variables X and Y , the conditional expectation of X given Y is the best approximation to X (in the above L^2 sense) using random variables that are measurable with respect to $\sigma(Y)$. The following exercise gives another way to think about how $E(X|Y)$ approximates X .

Exercise: Show that for any random variable $Z \in \sigma(Y)$, we have

$$EXZ = E[E(X|Y)Z] \quad (28)$$

This way of looking at the conditional expectation now generalizes easily.

Definition 3 Let $(\Omega, \mathcal{F}_0, P)$ be a probability space. Let \mathcal{F} be a sub σ -field of \mathcal{F}_0 , i.e., a subset of \mathcal{F}_0 that is itself a σ -field. Let $X \in \mathcal{F}_0$ with $E|X| < \infty$. The conditional expectation of X given \mathcal{F} , denoted $E(X|\mathcal{F})$, is the unique random variable Y which satisfies

- (i) $Y \in \mathcal{F}$
- (ii) $EXZ = EYZ$ for any random variable $Z \in \mathcal{F}$.

The definition asserts both that such a Y exists and is unique. Of course, uniqueness means that any other random variable with these two properties must equal Y with probability one. Uniqueness is not too hard to prove. Existence is harder to prove and requires the Radon-Nikodym theorem.

Theorem 4

- (i) $E(E(X|\mathcal{F})) = EX$
- (ii) If $X \in \mathcal{F}$ then $E(X|\mathcal{F}) = X$.
- (iii) If X is independent of \mathcal{F} , then $E(X|\mathcal{F}) = EX$.
- (iv) If $Y \in \mathcal{F}$, $E|XY| < \infty$, then $E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$.
- (v) $E(aX + BY|\mathcal{F}) = aE(X|\mathcal{F}) + bE(Y|\mathcal{F})$
- (vi) If $X \leq Y$ a.s., then $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$ a.s.
- (vii) If $\mathcal{F}_1 \subset \mathcal{F}_2$ then $E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$

Here is some intuition for these properties. The conditional expectation of X given \mathcal{F} is the random variable which best approximates X within the set of \mathcal{F} measurable functions. (“Best” is in the sense of minimizing the L^2 distance.) Property (i) says that the mean of the approximation is the same as the mean of the random variable we are approximating. Property (ii) says that if X is already in the set of functions you are using to approximate X , then the best approximation is X itself. Property (iii) says that if X is independent of the set of random variables you are using to approximate, then the best you can do is to take the approximating function to just be a constant. Property (iv) says that if Y is in the space of functions we are using to approximate with, then we get the approximation of XY by just multiplying the approximation of X by Y . Properties (v) and (vi) say that if we fix \mathcal{F} , then $X \rightarrow E(X|\mathcal{F})$ acts like the usual expectation, $X \rightarrow EX$. I don’t know a simple interpretation of property (vii). To remember it think “the smaller σ -field wins.” Note that if \mathcal{F}_1 and \mathcal{F}_2 are not related by inclusion, then $E(E(X|\mathcal{F}_1)|\mathcal{F}_2)$ and $E(E(X|\mathcal{F}_2)|\mathcal{F}_1)$ need not be equal. (This is very similar to projections in a Hilbert space commuting if their ranges are related by inclusion, but typically not commuting if their ranges are not so related.)

Property (iii) refers to a random variable being independent of a σ -field. What does this mean? One way to define it is that the random variable X and the σ -field \mathcal{F} are independent if for every Borel set $B \subset \mathbb{R}$ and every $A \in \mathcal{F}$, we have $P(A \cap \{X \in B\}) = P(A)P(X \in B)$. The notation is being tortured in the usual way: $A \cap \{X \in B\}$ means the intersection of A and the event $\{\omega : X(\omega) \in B\}$.

2.4 Markov properties

Filtrations

The above definition of Brownian motion is missing a key component - a filtration. A σ -field is a collection of subsets of the sample space satisfying some axioms. If X is a random variable, then the collection of subsets of Ω of the form $X \in B$ where B is a Borel subset of \mathbb{R} is a σ -field. We will denote it by $\sigma(X)$. Given a collection of random variables X_α , $\alpha \in \mathcal{A}$, we let $\sigma(X_\alpha, \alpha \in \mathcal{A})$ denote the σ -field generated by the X_α , i.e., the smallest σ -field which contains all sets of the form $X_\alpha \in B$ where B is a Borel subset of \mathbb{R} .

Definition 4 *A filtration is a family of σ -fields \mathcal{F}_t indexed by $t \geq 0$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ for $t < s$. It is right continuous if*

$$\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t \quad (29)$$

A stochastic process X_t is said to be adapted to the filtration \mathcal{F}_t if for every $t \geq 0$, X_t is measurable with respect to \mathcal{F}_t .

Given a stochastic process X_t we can construct a filtration by simply defining $\mathcal{F}_t = \sigma(X_s, s \leq t)$. Obviously, X_t is adapted to this filtration. Unfortunately if we do this with Brownian motion the resulting filtration is not right continuous. This can be fixed. One can construct a filtration which is right continuous and is the same as $\sigma(X_s, s \leq t)$ up to sets of measure zero. See Durrett for details. We will denote this right continuous filtration by \mathcal{F}_t . I will fudge and just pretend that \mathcal{F}_t is $\sigma(X_s, s \leq t)$.

Markov property

In general “Markov” means *the future depends on the past only through the present*. We make this precise. Start with a random walk $S_n = \sum_{i=1}^n X_i$ (without the interpolation and rescaling). We fix a time m which we think of as the present and assume that we know S_m , the location of the walk at time m , as well as how the walk got there, i.e., S_j for $j < m$. Consider what the walk does after m . Let $n > m$ and write S_n as $S_m + (S_n - S_m)$. The random variable $S_n - S_m$ is independent of the steps the walk took to get to S_m . So what the walk does after time m only depends on S_m , not on how the walk got to S_m . We can give a mathematical formulation using conditional expectation and σ -fields. We start with a special case in which we can prove the statement. Let $n > m$. Then

$$\begin{aligned} E(S_n | \sigma(S_1, \dots, S_m)) &= E(S_m + (S_n - S_m) | \sigma(S_1, \dots, S_m)) \\ &= E(S_m | \sigma(S_1, \dots, S_m)) + E(S_n - S_m | \sigma(S_1, \dots, S_m)) = S_m + E(S_n - S_m) \\ &= E(S_m | \sigma(S_m)) + E(S_n - S_m | \sigma(S_m)) = E(S_n | \sigma(S_m)) \end{aligned} \quad (30)$$

In fact a stronger statement is true. For any Borel function $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$E(f(S_n)|\sigma(S_1, \dots, S_m)) = E(f(S_n)|\sigma(S_m)) \quad (31)$$

In words, the distribution of $f(S_n)$ given S_1, S_2, \dots, S_m is the same as the distribution of $f(S_n)$ given only S_m .

Brownian motion should have a similar property. For $t > s$, $B_t - B_s$ is independent of what happened before time s , i.e., independent of B_u for $u \leq s$. So B_t will depend on $\{B_u : 0 \leq u \leq s\}$ only through B_s . Thus we should have for any Borel function f and $t > s$,

$$E(f(B_t)|\mathcal{F}_s) = E(f(B_t)|\sigma(B_s)) \quad (32)$$

This is true, but it is only a statement about the Brownian motion at a single time t in the future. The entire future should depend on the past only through the present. To state a stronger form of the Markov property we need to revisit the definition of Brownian motion.

Until now we have only considered Brownian motions that start at 0. We now need Brownian motions that start at other points. Let $C[0, \infty)$ be the set of continuous functions on $[0, \infty)$. Let $B_t(\omega) = \omega(t)$. There is a probability measure P which makes these random variables a Brownian motion and for which $B_0 = 0$ with P probability 1. One can prove that there is a family of probability measures P_x for $x \in \mathbb{R}$ on $C[0, \infty)$ such that B_t and P_x form a Brownian motion, and $P_x(\{\omega : \omega(0) = x\}) = 1$. We let E_x denote the expectation with respect to P_x .

Define $\theta_s : C[0, \infty) \rightarrow C[0, \infty)$ by

$$(\theta_s \omega)(t) = \omega(t + s), \quad t \geq 0 \quad (33)$$

So θ_s removes the part of the path for the time interval $[0, s]$ and then shifts time so that the path begins at time 0.

Theorem 5 (*Markov property*) *Let Y be a bounded random variable on $C[0, \infty)$ and $s \geq 0$. Define $\phi(z) = E_z Y$, and let $E_{B_s} Y$ denote $\phi(B_s)$. Then for any x*

$$E_x(Y \circ \theta_s | \mathcal{F}_s) = E_{B_s} Y \quad (34)$$

Note that the right side does not depend on x , so one of the assertions of the theorem is that the left side does not depend on x . The left side of the above equation is a random variable since it is a conditional expectation. By properties of conditional expectation it is measurable with respect to \mathcal{F}_s . Intuitively, this means it is a function of the B_t with $t \leq s$. $E_z Y$ is a number. When we replace z by B_s , we get a function of B_s , and so a random variable which is measurable with respect to $\sigma(B_s)$. There is an even stronger Markov property for Brownian motion involving stopping times. We will return to it at the end of this section.

Martingales

We first define a martingale in the discrete time case. Suppose that our stochastic process comes from some form of gambling. X_m represents the amount of money we have at time m . This is often called our “stake.” Heuristically, a martingale is a *fair game*. Think of m as the present. We know what has happened up to time m . Let $n > m$ be a time in the future. Then the expected value of X_n given our knowledge of the present and the past should be the current value of the stake, i.e., X_m . More formally,

$$E(X_n | \sigma(X_1, X_2, \dots, X_m)) = X_m \quad (35)$$

For example, consider a random walk S_m . This can be thought of as a simple gambling game where we play the same game at each time step. X_i represents the amount of money we win (or lose if $X_i < 0$) on the i th play, and S_m is our stake at time m . Then as we saw when we considered the Markov property of random walks,

$$E(S_n | \sigma(S_1, S_2, \dots, S_m)) = S_m + E(S_n - S_m) = S_m + (n - m)EX_1 \quad (36)$$

So the random walk is a martingale if and only if $EX_1 = 0$. In words, it is a fair game if the average amount we win on a single play is zero. We now turn to a continuous time process X_t . We assume that $E|X_t| < \infty$ for all $t \geq 0$ so that $E(X_t | \mathcal{F}_t)$ is defined.

Definition 5 *A stochastic process X_t is a martingale with respect to the filtration \mathcal{F}_t if for $t > s$,*

$$E(X_t | \mathcal{F}_s) = X_s \quad (37)$$

We construct three examples of martingales involving Brownian motion. As always, B_t denotes a Brownian motion. For the first example, we compute $E(B_t | \mathcal{F}_s)$ for $t > s$. Since $B_t - B_s$ is independent of \mathcal{F}_s , $E(B_t - B_s | \mathcal{F}_s) = E(B_t - B_s) = 0$. So

$$E(B_t | \mathcal{F}_s) = E((B_t - B_s) + B_s | \mathcal{F}_s) = E(B_t - B_s | \mathcal{F}_s) + E(B_s | \mathcal{F}_s) = B_s \quad (38)$$

Thus Brownian motion itself is a martingale.

For the second example we compute $E(B_t^2 | \mathcal{F}_s)$ for $t > s$.

$$E(B_t^2 | \mathcal{F}_s) = E((B_t - B_s + B_s)^2 | \mathcal{F}_s) = E((B_t - B_s)^2 | \mathcal{F}_s) + E(B_s^2 | \mathcal{F}_s) + 2E((B_t - B_s)B_s | \mathcal{F}_s) \quad (39)$$

Using independence, $E((B_t - B_s)^2 | \mathcal{F}_s) = E(B_t - B_s)^2 = t - s$. Of course, $E(B_s^2 | \mathcal{F}_s) = B_s^2$. In the last term we use property (iv) to get

$$E((B_t - B_s)B_s | \mathcal{F}_s) = B_s E((B_t - B_s) | \mathcal{F}_s) = B_s E(B_t - B_s) = 0 \quad (40)$$

Thus

$$E(B_t^2 | \mathcal{F}_s) = B_s^2 + t - s \quad (41)$$

So B_t^2 is not a martingale, but we can rewrite the above as

$$E(B_t^2 - t | \mathcal{F}_s) = B_s^2 - s \quad (42)$$

Thus $X_t = B_t^2 - t$ is a martingale.

We leave the third example for the reader.

Exercise: (a) For $t > s$, compute $E(\exp(B_t) | \mathcal{F}_s)$. Use the result to show that $\exp(B_t - t/2)$ is a martingale.

(b) Find $f(t)$ so that $\exp(iB_t + f(t))$ is a martingale.

Stopping times

Consider a two-dimensional Brownian motion starting from the origin. Let T be the first time it hits the unit circle. This is a random variable and it is a special kind of random variable called a stopping time. Let $t > 0$. We can tell if $T \leq t$ if we know B_s for $0 \leq s \leq t$. We do not need to know B_s for $s > t$. This motivates the following definition

Definition 6 *A stopping time is a random variable T taking values in $[0, \infty]$ such that for all $t \geq 0$, the event $\{T < t\}$ belongs to \mathcal{F}_t .*

It is not hard to show that if the filtration is right continuous then the above definition is equivalent to the definition with $\{T < t\}$ replaced by $\{T \leq t\}$. We make the trivial observation that a nonnegative constant is a stopping time. One of the most common ways stopping times arise is by looking at when the Brownian motion enters or leaves some set.

Proposition 3 *If A is an open or closed subset of \mathbb{R}^d and B_t is a d -dimensional Brownian motion, then*

$$T = \inf\{t : B_t \in A\} \quad (43)$$

is a stopping time.

The stopping times from the above proposition are often called the *hitting time* for A . There are many ways to combine stopping times to get new stopping times.

Proposition 4 (i) *If S and T are stopping times, then $\min\{S, T\}$, $\max\{S, T\}$ and $S + T$ are all stopping times.*

(ii) *If T_n is a sequence of stopping times, then $\sup_n T_n$, $\inf_n T_n$, $\liminf_n T_n$, and $\limsup_n T_n$ are all stopping times. In particular, if $\lim_n T_n$ exists a.s., then the limit is a stopping time.*

Caution: $S - T$ need not be a stopping time.

Strong Markov property

We now return to the Markov property of Brownian motion. Consider a two-dimensional Brownian motion and let S be the first time it hits the unit circle. If we look at the Brownian motion from time S onwards, it should look the same as a 2d Brownian motion started at the point on the circle where the Brownian motion first hit the circle. In other words, when we say the future depends on the past only through the present, we can take the present to be a stopping time.

To make this precise we need to define \mathcal{F}_S for a stopping time S .

Definition 7 *Let S be a stopping time. \mathcal{F}_S is the set of events A such that for all $t \geq 0$, $A \cap \{S \leq t\}$ is in \mathcal{F}_t .*

In words, the part of A that lies in $S \leq t$ should be measurable with respect to the information available at time t . We also need to define the time shift θ_S . We let $(\theta_S \omega)(t) = \omega(S(\omega) + t)$. So the path ω is shifted backwards in time by $S(\omega)$ and the part of the part between times 0 and $S(\omega)$ is discarded.

Theorem 6 *(Strong Markov property) Let Y be a bounded random variable on $C[0, \infty)$. Let S be a stopping time. Define $\phi(z) = E_z Y$, and let $E_{B_S} Y = \phi(B_S)$. Then for any x*

$$E_x(Y \circ \theta_S | \mathcal{F}_S) = E_{B_S} Y \quad (44)$$

on the event $S < \infty$.

To help digest the theorem, consider a Y which only depends on the sample path at one time. So let $t > 0$ and let $f : \mathbb{R} \rightarrow \mathbb{R}$. Define $Y(\omega) = f(\omega(t))$. Then the theorem says

$$E_x(f(\omega(t + S(\omega))) | \mathcal{F}_S) = E_{B_S} f(\omega(t)) \quad (45)$$

We start the Brownian motion at some point x and run the Brownian motion up to the stopping time S . Knowing what the path is up to this time, we look at the distribution of B_{t+S} . The theorem says it has the same distribution as B_t if we start the Brownian motion at $\omega(S)$.

Optional sampling theorem

If we think of our stochastic process as a gambling game, then a stopping time is a simple kind of system. It is a rule for when to quit playing given complete knowledge of what has happened up to the present, but no knowledge about the future. Suppose we quit playing at the stopping time T and ask what is the average of X_T . For a fair game we expect that it is the same as our initial stake, i.e., X_0 . So we expect $EX_T = EX_0$.

The equality is not true without some further conditions. For example, consider a 1d Brownian motion which starts at 0, and let T be the first time it hits +1. Then $B_T = 1$. So $EB_T = 1$, but $EB_0 = 0$.

Theorem 7 (*Optional sampling theorem*) Let X_t be a martingale with respect to the filtration \mathcal{F}_t , and let T be a stopping time. Assume that $P(T < \infty) = 1$, $E|X_T| < \infty$ and

$$\lim_{s \rightarrow \infty} E[|X_s| | T > s] P(T > s) = 0 \quad (46)$$

Then

$$EX_T = EX_0 \quad (47)$$

In the example before the theorem, $P(T < \infty) = 1$ and $E|X_T| = 1$. So it must be that (46) is not satisfied.

Exercise: Let $a < 0 < b$. Consider a 1d Brownian motion started at 0. Let T be the first time it reaches a or b , i.e., the hitting time for $\{a, b\}$. The hypotheses of the optional sampling theorem are true in this case. Use the conclusion of the theorem to compute the probability that the Brownian motion reaches b before it reaches a .

There are deep connections between Brownian motions and PDE's. Here is a nice example. We consider Brownian motion and think of it as a complex valued Brownian motion. Let $D \subset \mathbb{C}$ be a simply connected domain with continuous boundary. Let $f(z)$ be a bounded harmonic function on D that extends continuously to ∂D . (f is harmonic if $\Delta f = 0$.) Let B_t be a complex Brownian motion which starts at $z \in D$. We will show later that $X_t = f(B_t)$ is a martingale. (This requires the stochastic calculus and Ito's formula.) Now let T be the hitting time for ∂D . Then the optional sampling theorem says

$$E_z f(B_T) = E_z f(B_0) = f(z) \quad (48)$$

By the definition of the stopping time, $B_T \in \partial D$. Thus the left side of the above depends only on the boundary values of f . So the above tells you how to solve Laplace's equation with given boundary values. To find the value of the solution at z , you start a Brownian motion at z and run it until you hit the boundary. Then you average the value of f at this exit point.

Borel-Cantelli

Definition 8 Let E_n be a sequence of events. The event $\{E_n \text{ i.o.}\}$ is the event that E_n occurs infinitely often, i.e., the outcome belongs to infinitely many of the E_n .

A little thought convinces you that

$$\{E_n \text{ i.o.}\} = \bigcap_{m=1}^{\infty} \bigcup_{n:n>m} E_n \quad (49)$$

Theorem 8 (*Borel-Cantelli Lemma*) If E_n is a sequence of events with $\sum_n P(E_n) < \infty$, then $P(E_n \text{ i.o.}) = 0$. If E_n is a sequence of **independent** events with $\sum_n P(E_n) = \infty$, then $P(E_n \text{ i.o.}) = 1$.

Note that for a sequence of independent events, the probability that E_n occurs i.o. can only be 0 or 1; which value depends only on whether the sum converges or diverges.

Exercise: The first half of the Borel-Cantelli lemma is an easy exercise in measure theory using (49). Prove it. Prove the second half. This is harder, so here are some hints. Prove that it suffices to show $P(\cup_{n:n>m} E_n) = 1$ for all m . Show that $\sum_n P(E_n) = \infty$ implies $\prod_n (1 - P(E_n)) = 0$. Now use $1 - P(E_n) = P(E_n^c)$ and the independence of the E_n^c to see what this says.