# 5. Brownian motion

**Section 1.** The definition and some simple properties.
**Section 2.** Visualizing Brownian motion. Discussion and demystification of some strange and scary pathologies.
**Section 3.** The reflection principle.
**Section 4.** Conditional distribution of Brownian motion at some point in time, given observed values at some other times.
**Section 5.** Existence of Brownian motion. How to construct Brownian motion from familiar objects.
**Section 6.** Brownian bridge. Application to testing for unformity.
**Section 7.** A boundary crossing problem solved in two ways: differential equations and martingales.
**Section 8.** Discussion of some issues about probability spaces and modeling.

Brownian motion is one of the most famous and fundamental of stochastic processes. The formulation of this process was inspired by the physical phenomenon of Brownian motion, which is the irregular jiggling sort of movement exhibited by a small particle suspended in a fluid, named after the botanist Robert Brown who observed and studied it in 1827. A physical explanation of Brownian motion was given by Einstein, who analyzed Brownian motion as the cumulative effect of innumerable collisions of the suspended particle with the molecules of the fluid. Einstein's analysis provided historically important support for the atomic theory of matter, which was still a matter of controversy at the time—shortly after 1900. The mathematical theory of Brownian motion was given a firm foundation by Norbert Wiener in 1923; the mathematical model we will study is also known as the "Wiener process."

Admittedly, it is possible that you might not share an all-consuming fascination for the motion of tiny particles of pollen in water. However, there probably are any number of things that you do care about that jiggle about randomly. Such phenomena are candidates for modeling via Brownian motion, and the humble Brownian motion process has indeed come to occupy a central role in the theory and applications of stochastic processes. How does it fit into the big picture? We have studied Markov processes in discrete time and having a discrete state space. With continuous time and a continuous state space, the prospect arises that a process might have continuous sample paths. To speak a bit roughly for a moment, Markov processes that have continuous sample paths are called *diffusions*. Brownian motion is the simplest diffusion, and in fact other diffusions can be built up from Brownian motions in various ways. Brownian motion and diffusions are used all the time in models in all sorts of fields, such as finance ⟦in modeling the prices of stocks, for example⟧, economics, queueing theory, engineering, and biology. Just as a pollen particle is continually buffeted by collisions with water molecules, the price of a stock is buffeted by the actions of many individual investors. Brownian motion and diffusions also arise as approximations for other processes; many processes converge to diffusions when looked at in the right way. In fact, in a sense ⟦does non-sense count?⟧, Brownian motion is to stochastic processes as the standard normal distribution is to random variables: just as the normal distribution arises as a limit distribution for suitably normalized sequences of random variables, Brownian motion is a limit in distribution of certain suitably normalized sequences of stochastic processes. Roughly speaking, for many processes, if you look at them from far away and not excessively carefully, they look nearly like Brownian motion or diffusions, just as the distribution of a sum of many *iid* random variables looks nearly normal.

## 5.1 The definition

Let's scan through the definition first, and then come back to explain some of the words in it.

(5.1) DEFINITION. *A **standard Brownian motion (SBM)** $\{W(t) : t \geq 0\}$ is a stochastic process having*

(i) *continuous paths,*

(ii) *stationary, independent increments, and*

(iii) $W(t) \sim N(0, t)$ *for all* $t \geq 0$.

The letter "$W$" is often used for this process, in honor of Norbert Wiener. ⟦Then again, there is also an object called the "Wiener sausage" studied in physics.⟧

The definition contains a number of important terms. First, it is always worth pointing out that a ***stochastic process*** $W$ is really a function $W : \mathbb{R} \times \Omega \to \mathbb{R}$, and thus may be viewed in two ways: as a "collection of random variables" and as a "random function" ⟦or "random path"⟧. That is, $W = W(t, \omega)$. For each fixed $t$, $W(t, \cdot)$ is a real-valued function defined on $\Omega$, that is, a random variable. So $W$ is a collection of random variables. We will use both notations $W(t)$ and $W_t$ for the random variable $W(t, \cdot)$. For fixed $\omega$, $W(\cdot, \omega)$ is a real-valued function defined on $\mathbb{R}_+$; such a function could be viewed as a "path." Thus $W$ is a random function, or random path. Which brings us to the next item: ***continuous paths***. By this we mean that

$$P\{\omega \in \Omega : W(\cdot, \omega) \text{ is a continuous function}\} = 1.$$

Next, the ***independent increments*** requirement means that for each $n$ and for all choices of times $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, the random variables $W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$ are independent. The term ***stationary increments*** means that the distribution of the increment $W(t) - W(s)$ depends only on $t - s$. Note that from the requirement that $W(t) \sim N(0,t)$, we can see that $W(0) = 0$ with probability one. From this, using (ii) and (iii) of the definition,

$$W(t) - W(s) \sim W(t-s) - W(0) = W(t-s) \sim N(0, t-s)$$

for $s \le t$. Thus:

| |
|---|
| The increment that a standard Brownian motion makes over a time interval of length $h$ is normally distributed with mean 0 and variance $h$. |

A standard Brownian motion has a constant mean of 0, so that it has no "drift," and its variance increases at a rate of 1 per second. ⟦For now take the unit of time to be 1 second.⟧ A standard Brownian motion is a standardized version of a general Brownian motion, which need not have $W(0) = 0$, may have a nonzero "drift" $\mu$, and has a "variance parameter" $\sigma^2$ that is not necessarily 1.

(5.2) Definition. *A process $X$ is called a $(\mu, \sigma^2)$* ***Brownian motion*** *if it can be written in the form*

$$X(t) = X(0) + \mu t + \sigma W(t),$$

*where $W$ is a standard Brownian motion.*

Notice that the mean and variance of a $(\mu, \sigma^2)$ Brownian motion increase at rate $\mu$ and $\sigma^2$ per second, respectively.

This situation here is analogous to that with normal distributions, where $Z \sim N(0,1)$ is called a "standard" normal random variable, and general normal random variables are obtained by multiplying a standard normal random variable by something and adding something.

(5.3) Exercise [A sojourn time problem]. *Let $X(t) = \mu t + \sigma W(t)$, where $W$ is a standard Brownian motion and $\mu > 0$.*

(i) *For $\delta > 0$, find the expected total amount of time that the process $X$ spends in the interval $(0, \delta)$. That is, defining*

$$T = \int_0^\infty I\{X(t) \in (0, \delta)\}\, dt,$$

*what is $E(T)$? A rather involved calculus calculation should eventually arrive at a strikingly simple answer.*

(ii) *Can you give a convincing but calculation-free argument why the simple answer is correct?*

(5.4) Exercise [Another sojourn time problem]. *As in the previous problem, let $X(t) = \mu t + \sigma W(t)$, where $W$ is a standard Brownian motion and $\mu > 0$. What is the expected* *amount of time that the process $X$ spends below 0? ⟦The calculus is easier in this problem than in the previous one.⟧*

The following characterization of Brownian motion is sometimes useful.

(5.5) FACT. *If a stochastic process $X$ has continuous paths and stationary, independent increments, then $X$ is a Brownian motion.*

Thus, the assumptions of path continuity and stationary, independent increments is enough to give the normality of the increments "for free." This is not surprising, from the Central Limit Theorem.

Here is a very useful characterization of standard Brownian motion. First, $W$ is a ***Gaussian process***, which means that for all numbers $n$ and times $t_1, \ldots, t_n$ the random vector $(W(t_1), \ldots, W(t_n))$ has a joint normal distribution. An equivalent characterization of the Gaussianity of $W$ is that the sum

$$a_1 W(t_1) + \cdots + a_n W(t_n)$$

is normally distributed for each $n$, all $t_1, \ldots, t_n$, and all real numbers $a_1, \ldots, a_n$. Being a Gaussian process having mean 0, the joint distribution of all finite collections of random variables $W(t_1), \ldots, W(t_n)$ are determined by the ***covariance function***

$$r(s,t) = \mathrm{Cov}(W_s, W_t).$$

For standard Brownian motion, $\mathrm{Cov}(W_s, W_t) = s \wedge t$. To see this, suppose that $s \le t$, and observe that

$$\begin{aligned} \mathrm{Cov}(W_s, W_t) &= \mathrm{Cov}(W_s, W_s + (W_t - W_s)) \\ &= \mathrm{Var}(W_s) + \mathrm{Cov}(W_s, W_t - W_s) \\ &= s + 0 = s, \end{aligned}$$

where we have used the independence of increments to say that

$$\mathrm{Cov}(W_s, W_t - W_s) = \mathrm{Cov}(W_s - W_0, W_t - W_s) = 0.$$

It is easy to see ⟦Exercise!⟧ that a process $W$ is Gaussian with mean 0 and covariance function $r(s,t) = s \wedge t$ if and only if (ii) and (iii) of Definition (5.1) hold for $W$. Thus:

| |
|---|
| A Gaussian process having continuous paths, mean 0, and covariance function $r(s,t) = s \wedge t$ is a standard Brownian motion. |

Here is an interesting example that shows how convenient the previous characterization of Brownian motion can be.

(5.6) EXAMPLE. Suppose that $W$ is a SBM, and define a process $X$ by $X(t) = tW(1/t)$ for $t > 0$, and define $X(0) = 0$. Then we claim that $X$ is also a SBM.

To check this, we'll check that $X$ satisfies the conditions in the last characterization. To start we ask: is $X$ a Gaussian process? Given $n$, $t_1, \ldots, t_n$, and $a_1, \ldots, a_n$, we have

$$a_1 X(t_1) + \cdots + a_n X(t_n) = a_1 t_1 W(1/t_1) + \cdots + a_n t_n W(1/t_n),$$

which, being a linear combination of $W$ evaluated at various times, has a normal distribution. Thus, the fact that $W$ is a Gaussian process implies that $X$ is also. Next, observe that the path continuity of $X$ is also a simple consequence of the path continuity of $W$: if $t \mapsto W(t)$ is continuous, then so is $t \mapsto tW(1/t)$. ⟦Well, this proves that with probability one $X = X(t)$ is continuous for all positive $t$. For $t = 0$, if you believe that $\lim_{s\to\infty} W(s)/s = 0$ with probability one—which is eminently believable by the SLLN—then making the substitution $s = 1/t$ gives $\lim_{t\to 0} tW(1/t) = 0$ with probability 1, so that $X$ is also continuous at $t = 0$. Let's leave it at this for now.⟧ The fact that $X(t)$ has mean 0 is trivial. Finally, to check the covariance function of $X$, let $s \le t$ and observe that

$$\begin{aligned}\mathrm{Cov}(X(s), X(t)) &= \mathrm{Cov}(sW(\frac{1}{s}), tW(\frac{1}{t})) = st\mathrm{Cov}(W(\frac{1}{s}), W(\frac{1}{t})) \\ &= st\left(\frac{1}{s} \wedge \frac{1}{t}\right) = st\frac{1}{t} = s.\end{aligned}$$

Thus, $X$ is a SBM. □

(5.7) Exercise ["Brownian scaling"]. *Suppose that $W$ is a standard Brownian motion, and let $c > 0$. Show that the process $X$ defined by $X(t) = c^{-1/2}W(ct)$ is also a standard Brownian motion.*

The next exercise is a typical example of the use of the Brownian scaling relationship to get substantial information about the form of a functional relationship with little effort.

(5.8) Exercise. *Imagine that you do not already know the answer to Exercise (5.4); you know only that the answer is some function of $\mu$ and $\sigma^2$. Use Brownian scaling to argue without calculus that the desired function must be of the form $a\sigma^2/\mu^2$ for some number $a$.*

(5.9) Exercise. *Suppose that $W$ is a standard Brownian motion, and let $c > 0$. Define $X(t) = W(c+t) - W(c)$. Show that $\{X(t) : t \ge 0\}$ is a standard Brownian motion that is independent of $\{W(t) : 0 \le t \le c\}$.*

The previous exercise shows that Brownian motion is continually "restarting" in a probabilistic sense. At each time $c$, the Brownian motion "forgets" its past and continues to wiggle on just as if it were a new, independent Brownian motion. That is, suppose that we know that $W(c) = w$, say. Look at the graph of the path of $W$; we are assuming the graph passes through the point $(c, w)$. Now imagine drawing a new set of coordinate axes, translating the origin to the point $(c, w)$. So the path now goes through the new origin. Exercise

(5.9) says that if we look at the path past time $c$, relative to the new coordinate axes, we see the path of a new standard Brownian motion, independent of what happened before time $c$. Brownian motion is a *Markov process*: given the current state, future behavior does not depend on past behavior.

Just as a brief note in passing, the *strong Markov property* is an extension of the restarting property of Exercise (5.9) from fixed times $c$ to random *stopping times* $\gamma$: For a stopping time $\gamma$, the process $x$ defined by $X(t) = W(\gamma + t) - W(\gamma)$ is a Brownian motion, independent of the path of $W$ up to time $\gamma$. We will discuss stopping times a bit more below, but the concept is the same as the idea from the previous chapter about martingales in discrete time: at time $t$, we know whether or not $\gamma \le t$ by looking at the path of $W$ up to time $t$. We can see the role of this sort of requirement by seeing that the restarting property can fail for a random time that isn't a stopping time. For example, let $M = \max\{B_t : 0 \le t \le 1\}$ and let $\beta = \inf\{t : B_t = M\}$; this is the first time at which $B$ achieves its maximum height over the time interval $[0,1]$. Clearly $\beta$ is not a stopping time, since we must look at the whole path $\{B_t : 0 \le t \le 1\}$ to determine when the maximum is attained. Also, the restarted process $X(t) = W(\beta + t) - W(\beta)$ is not a standard Brownian motion; in particular, by definition of $\beta$, the process $\{X(t)\}$ stays below 0 for some time after $t = 0$, which we know happens with probability 0 for Brownian motion.

(5.10) Exercise [Ornstein-Uhlenbeck process]. *Define a process $X$ by*

$$X(t) = e^{-t} W(e^{2t})$$

*for $t \ge 0$. $X$ is called an* ***Ornstein-Uhlenbeck process****.*

(i) *Find the covariance function of $X$.*

(ii) *Evaluate the functions $\mu$ and $\sigma^2$, defined by*

$$\begin{aligned} \mu(x,t) &= \lim_{h \downarrow 0} \frac{1}{h} E[X(t+h) - X(t) \mid X(t) = x] \\ \sigma^2(x,t) &= \lim_{h \downarrow 0} \frac{1}{h} \mathrm{Var}[X(t+h) - X(t) \mid X(t) = x]. \end{aligned}$$

## 5.2 Visualizing Brownian motion

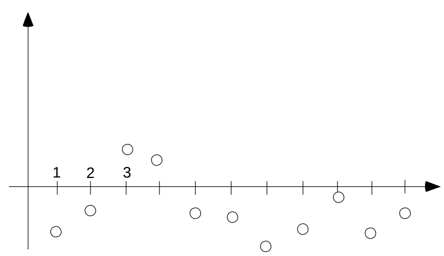
First, friends, it's time for some frank talk about Brownian motion. Brownian motion can be very difficult to visualize; in fact, in various respects it's impossible. Brownian motion has some "pathological" features that make it seem strange and somewhat intimidating. Personally, I remember that after having heard some weird things about Brownian motion, I felt rather suspicious and mistrustful of it, as if I could not use it with confidence or even speak of it without feeling apologetic somehow. We will not dwell unduly on the pathologies here, but I do not want us to completely avert our eyes, either. Let's try to take enough of a peek so that we will not be forever saddled with the feeling that we have chickened out completely. Hopefully, such a peek will result in an improved level of confidence and familiarity in working with Brownian motion.

What is all this about "pathologies"? After all, Brownian motion has continuous sample paths, and continuous functions are quite nice already, aren't they? Here's one slightly strange feature, just to get started. Recall that $W(0) = 0$. It turns out that for almost all sample paths of Brownian motion ⟦that is, for a set of sample paths having probability one⟧, for all $\epsilon > 0$, the path has infinitely many zeros in the interval $(0, \epsilon)$. That is, the path changes sign infinitely many times, cutting through the horizontal axis infinitely many times, within the interval $(0, \epsilon)$; it does more before time $\epsilon$ than most of us do all day. Another rather mind-boggling property is that with probability 1, a sample path of Brownian motion does not have a derivative *at any time*! It's easy to imagine functions—like $f(t) = |t|$, for example—that fail to be differentiable at isolated points. But try to imagine a function that *everywhere* fails to be differentiable, so that there is not even one time point at which the function has a well-defined slope.

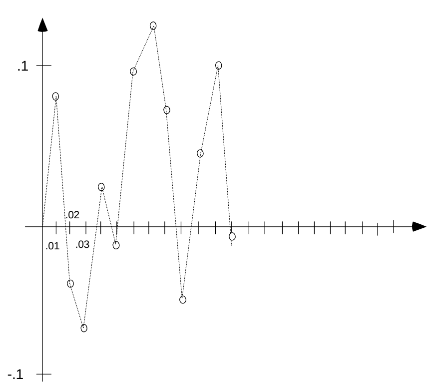
Such functions are not easy to imagine. In fact, before around the middle of the 19th century mathematicians generally believed that such functions did not exist, that is, they believed that every continuous function must be differentiable somewhere. Thus, it came as quite a shock around 1870 when Karl Weierstrass produced an example of a nowhere-differentiable function. Some in the mathematical establishment reacted negatively to this work, as if it represented an undesirable preoccupation with ugly, monstrous functions. Perhaps it was not unlike the way adults might look upon the ugly, noisy music of the next generation. It is interesting to reflect on the observation that, in a sense, the same sort of thing happened in mathematics much earlier in a different context with which we are all familiar. Pythagorus discovered that $\sqrt{2}$—which he knew to be a perfectly legitimate number, being the length of the hypotenuse of a right triangle having legs of length one—is irrational. Such numbers were initially viewed with great distrust and embarrassment. They were to be shunned; notice how even the name "irrational" still carries a negative connotation. Apparently some Pythagoreans even tried to hide their regretable discovery. Anyway, now we know that in a sense "almost all" numbers are of this "undesirable" type, in the sense that the natural measures that we like to put on the real numbers ⟦like Lebesgue measure (ordinary length)⟧ place all of their "mass" on the set of irrational numbers and no mass on the set of rational numbers. Thus, the proof of existence of irrational numbers by producing an example of a particular irrational number was dwarfed by the realization that if one chooses a real number at random under the most natural probability measures, the result will be an irrational number with probability 1. The same sort of turnabout has occurred in connection with these horrible nowhere-differentiable functions. Weierstrass constructed a particular function and showed that it was nowhere differentiable. The strange nature of this discovery was transformed in the same sense by Brownian motion, which puts probability 0 on "nice" functions and probability 1 on nowhere differentiable functions.

Having presented two scary pathologies, let us now argue that they are not really all that strange or unexpected. We'll start with the fact that a Brownian motion $W$ has infinitely many zeros in the interval $(0, \epsilon)$. Let $b$ be an arbitrary positive number, perhaps very large. Do you believe that a Brownian motion will necessarily hit 0 infinitely many times in the time interval $(b, \infty)$? This proposition seems to me to be quite believable and not at all scary ⟦for example, by recurrence considerations we know that a simple random walk will behave this way⟧. Well, recall that $X(s) = sW(1/s)$ is a Brownian motion. So you believe that $X(s) = sW(1/s) = 0$ for infinitely many values of $s$ in $(b, \infty)$. But this implies that $W(t) = 0$ for infinitely many values of $t$ in the interval $(0, 1/b)$. Making the identification of $1/b$ with $\epsilon$ shows that the scary pathology and the believable proposition are the same. Now for the nondifferentiability of the Brownian paths. This should not be very surprising, by the assumption of independent increments. Indeed, for each $t$ and each $\delta > 0$, the increment $W(t+\delta) - W(t)$ is independent of the increment $W(t) - W(t-\delta)$, so that it would just be the wildest stroke of luck if the increments on both sides of $t$ "matched up" well enough for $W$ to be differentiable at $t$!

Enough of that for a while. How does Brownian motion look and behave? We can get a good idea of the behavior on a rough scale by sampling the process at every integer, say. If we are looking at the process over a large time interval and are not concerned about little fluctuations over short time intervals, then this sort of view may be entirely adequate. It is also very easy to understand, since $W(0), W(1), W(2), \ldots$ is just a random walk with *iid* standard normal increments. This a very familiar, non-mysterious sort of process.




What if we want to get a more detailed picture? Let's zoom in on the first tenth of a second, sampling the process in time intervals of length 0.01 instead of length 1. Then we might get a picture that looks like this.



We get another normal random walk, this time with the increments having variance 0.01 instead of variance 1. Notice that the standard deviation of the increments is 0.1, which is 10 times bigger than the time interval 0.01 over which the increments take place! That is, the random walk changes by amounts of order of magnitude 0.1 over intervals of length 0.01, so that we get a random walk that has "steep" increments having "slope" on the order of magnitude of 10.

We could continue to focus in on smaller and smaller time scales, until we are satisfied that we have seen enough detail. For example, if we sampled 10,000 times every second instead of 100 times, so that the sampling interval is .0001, the standard deviation of the increments would be $\sqrt{.0001} = .01$, so that the random walk would have even "steeper" increments whose slope is now measured in the hundreds rather than in tens. Notice again how we should not be surprised by Brownian motion's catastrophic failure of differentiability.

It is reassuring to know that in a sense we can get as accurate and detailed a picture of Brownian motion as we like by sampling in this way, and that when we do so, we simply get a random walk with normally distributed increments.

## 5.3 A simple calculation with Brownian motion: the reflection principle

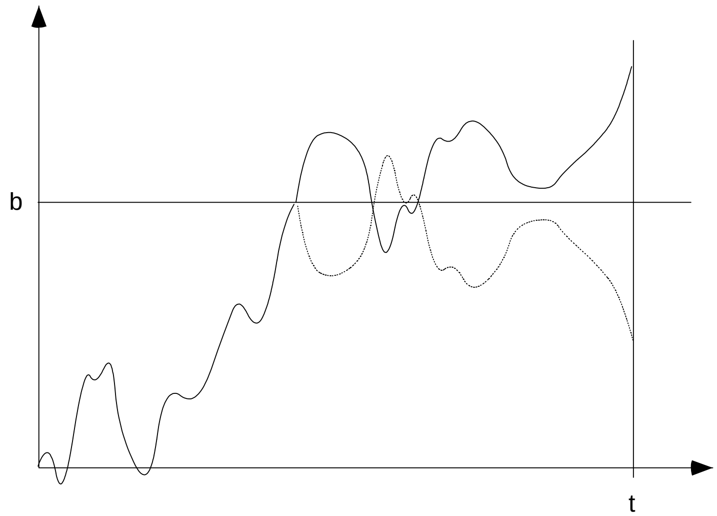
Let $\{W_t\}$ be a standard Brownian motion. For $b > 0$, define the first passage time

$$\tau_b = \inf\{t : W_t \geq b\};$$

by path continuity, an equivalent definition would be $\inf\{t : W_t = b\}$. Here is a problem: what is $P\{\tau_b \leq t\}$? Here is how to do it. First note that $\{\tau_b \leq t, W_t > b\} = \{W_t > b\}$, since by path continuity and the assumption that $W(0) = 0$, the statement that $W(t) > b$ *implies* that $\tau_b \leq t$. Using this,

$$\begin{aligned} P\{\tau_b \leq t\} &= P\{\tau_b \leq t, W_t < b\} + P\{\tau_b \leq t, W_t > b\} \\ &= P\{W_t < b \mid \tau_b \leq t\} P\{\tau_b \leq t\} + P\{W_t > b\}. \end{aligned}$$

The term $P\{W_t > b\}$ is easy: since $W_t \sim N(0,t)$, the probability is $1 - \Phi(\frac{b}{\sqrt{t}})$. Next, a little thought will convince you that $P\{W_t < b \mid \tau_b \leq t\} = \frac{1}{2}$. Indeed, path continuity guarantees that $W_{\tau_b} = b$, so that, knowing that $\tau_b \leq t$, the process is equally likely to continue on to be above $b$ or below $b$ at time $t$. [[A rigorous justification involves the "strong Markov property," but let's not get into that now. Also note that path continuity is important here. We could not make a statement like this about a discrete-time random walk having $N(0,1)$ increments, for example, since there will always be an "overshoot" when such a process first jumps above $b$.]]




Making the above substitutions and solving for $P\{\tau_b \le t\}$ gives

$$P\{\tau_b \le t\} = 2P\{W_t > b\} = 2\left[1 - \Phi\left(\frac{b}{\sqrt{t}}\right)\right],$$

a nice explicit result! This is one reason why people like to use Brownian motion in models—it sometimes allows explicit, tidy results to be obtained.

Here's another example of a nice, explicit formula. For now let's just present it "for enrichment"; we'll come back to derive it later. For Brownian motion $X_t = W_t + \mu t$ with drift $\mu$, defining $\tau_b = \inf\{t : X_t \ge b\}$, we have

$$P_\mu\{\tau_b \le t\} = 1 - \Phi\left(\frac{b - \mu t}{\sqrt{t}}\right) + e^{2\mu b}\Phi\left(\frac{-b - \mu t}{\sqrt{t}}\right).$$

Isn't that neat?

(5.11) EXERCISE. *Throughout let $W$ be a standard Brownian motion.*

(i) *Defining $\tau_b = \inf\{t : W(t) = b\}$ for $b > 0$ as above, show that $\tau_b$ has probability density function*

$$f_{\tau_b}(t) = \frac{b}{\sqrt{2\pi}} t^{-3/2} e^{-b^2/(2t)}$$

*for $t > 0$.*

(ii) *Show that for $0 < t_0 < t_1$,*

$$P\{W(t) = 0 \text{ for some } t \in (t_0, t_1)\} = \frac{2}{\pi}\tan^{-1}\left(\sqrt{\frac{t_1}{t_0} - 1}\right) = \frac{2}{\pi}\cos^{-1}\left(\sqrt{\frac{t_0}{t_1}}\right).$$

*⟦Hint: The last equality is simple trigonometry. For the previous equality, condition on the value of $W(t_0)$, use part (i), and Fubini (or perhaps integration by parts).⟧*

(iii) *Let* $L = \sup\{t \in [0,1] : W_t = 0\}$ *be the last zero of* $W$ *on [0,1]. Find and plot the probability density function of* $L$. *Rather peculiar, wouldn't you say?*

(5.12) EXERCISE. *Let* $X_t = W_t + \mu t$ *be a Brownian motion with drift* $\mu$, *and let* $\epsilon > 0$. *Show that*

$$P\{\max_{0\le t\le h} |X_t| > \epsilon\} = o(h) \text{ as } h \downarrow 0.$$

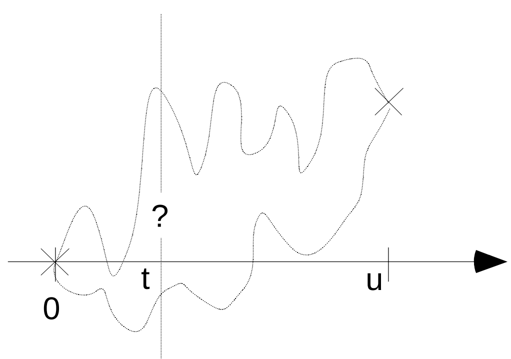
*⟦That is,*

$$\frac{1}{h} P\{\max_{0\le t\le h} |X_t| > \epsilon\} \to 0 \text{ as } h \downarrow 0.$$

*Hint: You might want to do the special case* $\mu = 0$ *first. For the general case, you can transform the problem about* $X$ *into an equivalent problem about* $W$, *and use the special case to do an easy bound. This shouldn't be a very involved calculation.⟧ This result is useful when we are calculating something and keeping terms up to order* $h$, *and we want to show that the probability of escaping from a strip can be neglected.*

## 5.4 Conditional distributions for Brownian motion

What happens between sampled points of a Brownian motion? That is, if we are given the value of Brownian motion at two time points, what is the conditional distribution of the process between those times?




We will examine this question in more detail when we discuss the Brownian bridge in Section 5.6. For now we'll just do what we need for the construction of Brownian motion in the next section. So here is a problem. Let $W$ be a standard Brownian motion (so we know that $W_0 = 0$) and let $0 < t < u$. What is the conditional distribution of $W_t$ given $W_u$?

The next result gives a nice way of working with the relevant normal distributions.

(5.13) CLAIM. $W_t - (t/u)W_u$ *is independent of* $W_u$.

To verify this, just note that

$$\mathrm{Cov}(W_t - (t/u)W_u, W_u) = (t \wedge u) - (t/u)(u \wedge u) = t - (t/u)u = 0,$$

and recall that jointly Gaussian variables are independent if they have zero covariance.

This simple claim makes everything easy. To get the conditional mean $E(W_t \mid W_u)$, observe that

$$\begin{aligned} 0 &= E(W_t - (t/u)W_u) \\ &\stackrel{(a)}{=} E(W_t - (t/u)W_u \mid W_u) \\ &= E(W_t \mid W_u) - (t/u)W_u, \end{aligned}$$

where $(a)$ follows from the claim. Thus,

$$E(W_t \mid W_u) = (t/u)W_u.$$

This relation makes sense; undoubtedly you would have guessed it from the picture! The conditional mean of Brownian motion is obtained by linearly interpolating between the points we are given.

For the conditional variance,

$$\begin{aligned} \mathrm{Var}(W_t \mid W_u) &= E[(W_t - E(W_t \mid W_u))^2 \mid W_u] \\ &= E[(W_t - (t/u)W_u)^2 \mid W_u] \\ &\stackrel{(a)}{=} E[(W_t - (t/u)W_u)^2] \\ &= (t \wedge t) - 2(t/u)(t \wedge u) + (t/u)^2(u \wedge u) \\ &= t - 2(t^2/u) + (t^2/u) = t\left(1 - \frac{t}{u}\right) = \frac{t(u-t)}{u}, \end{aligned}$$

where we have again used the claim at $(a)$. The functional form $t(u-t)/u$ for the conditional variance makes some qualitative sense at least: notice that it approaches 0 as $t$ approaches either of the two points 0 or $u$, which makes sense. Also, for fixed $u$, the conditional variance is maximized by taking $t$ in the middle: $t = u/2$.

In summary, we have found that for $0 < t < u$

$$\mathcal{L}(W_t \mid W_u) = N\left(\frac{t}{u}W_u, \frac{t(u-t)}{u}\right).$$

Observe that the conditional variance does not depend on $W_u$! Does this surprise you? For example, to take an extreme example, we have found that the conditional distribution of $W(1/2)$ given that $W(1) = 10$ billion is normal with mean 5 billion and variance 1/4! How do you feel about that? In this example, should we really claim that $W(1/2)$ is within 1.5 (that's 3 standard deviations) of 5 billion with probability 0.997? Well, that is what is implied by the Brownian motion model. Here is one say to conceptualize what is going on. It's *extremely* painful for a Brownian motion to get to 10 billion at time 1 (that's why it is so extremely rare). Among all of the painful ways the Brownian motion can do this, by far the least painful is for it to spread the pain equally over the two subintervals [0,1/2] and [1/2,1], making an increment of very nearly 5 billion over each. This last property of Brownian motion, which basically stems from the small tail of the normal distribution, could be viewed as a defect of the model. In real life, if one observes a value that seems outlandish according to our model, such as the value $W(1)$ =10 billion as discussed above, it does not seem sensible to be pig-headedly sure about the value of $W(1/2)$. In fact, an outlandish observation should be an occasion for healthy respect for the limitations of models in general and for skepticism about the suitability of this model in particular, which should lead to a humbly large amount of uncertainty about the value of $W(1/2)$.

(5.14) Exercise. *Show that for* $0 \le s < t < u$,

$$\mathcal{L}(W_t \mid W_s, W_u) = N\left(W_s + \frac{t-s}{u-s}(W_u - W_s), \frac{(t-s)(u-t)}{u-s}\right).$$

*[[Hint: This may be obtained by applying the previous result to the Brownian motion* $\tilde{W}$ *defined by* $\tilde{W}(v) = W(s+v) - W(s)$ *for* $v \ge 0$.*]]*

(5.15) Exercise. *Let* $0 < s < t < u$.

1. *Show that* $E(W_s W_t \mid W_u) = \frac{s}{t} E(W_t^2 \mid W_u)$.
2. *Find* $E(W_t^2 \mid W_u)$ *[[you know* $\operatorname{Var}(W_t \mid W_u)$ *and* $E(W_t \mid W_u)$*!]] and use this to show that*
$$\operatorname{Cov}(W_s, W_t \mid W_u) = \frac{s(u-t)}{u}.$$

(5.16) Exercise. *What is the conditional distribution of the vector* $(W_5, W_9, W_{12})$ *given that* $W_{10} = 3$*? [[It is a joint normal distribution; you should give the* $(3 \times 1)$ *mean vector and* $(3 \times 3)$ *covariance matrix.]]*

## 5.5 Existence and construction of Brownian motion (Or: Let's Play Connect-the-Dots)

I like the way David Freedman puts it: "One of the leading results on Brownian motion is that it exists." It is indeed comforting to know that we have not been talking about nothing. We will show that Brownian motion exists by "constructing" it. This means that we will show how we can obtain Brownian motion by somehow putting together other simpler, more familiar things whose existence we are not worried about. Why do I want to do this? It is not because I think that a mathematical proof of the existence of Brownian motion is somehow legally necessary before we can do anything else (although I suppose it could be a bit embarrassing to be caught analyzing things that do not exist). Rather, it is because I think that seeing how a construction of Brownian motion works gives one a much better, more "familiar" feeling for Brownian motion. Personally, after having heard some weird things about Brownian motion, I felt much less queasy about it after seeing how it could be constructed rather simply from familiar objects. For gaining familiarity and understanding, there's nothing like taking something apart and putting it together again.

We will construct Brownian motion on the time interval [0,1]; having done that, it will be easy to construct Brownian motion on $[0, \infty)$. We'll do it by an intuitive connect-the-dots approach, in which at each stage of the construction we obtain a more and more detailed picture of a sample path. We know $W(0) = 0$. At the initial stage, we start with the modest goal of simulating the value of the Brownian motion at time 1. Since we know that $W(1) \sim N(0,1)$, we can do this by going to our computer and generating a $N(0,1)$ random variable $Z_1$; take $Z_1$ to be $W(1)$. Given just the information that the path passes through the two points $(0,0)$ and $(1, Z_1)$, the conditional expectation is the linear interpolation $X^{(0)}$ shown in Figure 1, that is, $X^{(0)}(t) = Z_1 t$. This will be our first crude approximation to a sample path.

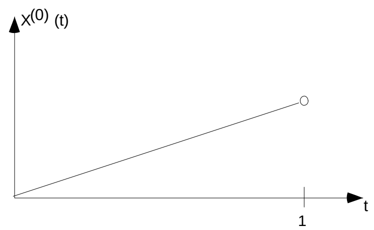



Figure 1

Next let's simulate a value for $W(1/2)$. Given the values we have already generated for $W(0)$ and $W(1)$, we know that $W(1/2)$ is normally distributed with mean $Z_1/2$ and variance $(1/2)(1/2) = 1/4$. Since $X^{(0)}(1/2)$ is already $Z_1/2$, we need only add a normal random variable with mean 0 and variance 1/4 to $X^{(0)}(1/2)$ to get the right distribution. Accordingly, generate another independent $N(0,1)$ random variable $Z_2$ and take $W(1/2)$ to be $X^{(0)}(1/2)+(1/2)Z_2$. Having done this, define the approximation $X^{(1)}$ to be the piecewise linear path joining the three points (0,0), $(1/2, W(1/2))$, and $(1, W(1))$ as in Figure 2.

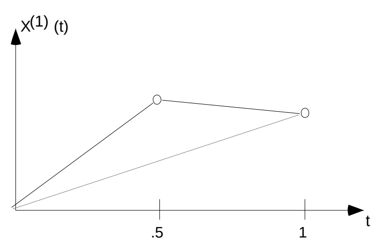



Figure 2

Now let's simulate $W(1/4)$ and $W(3/4)$. Notice how the correct conditional means are already given by the piecewise linear path $X^{(1)}$; that is, $E(W(t) \mid W(0), W(1/2), W(1)) = X^{(1)}(t)$; this holds for all $t$, and in particular for $t = 1/4$ and $t = 3/4$. The conditional variance of $W(1/4)$ given $W(0)$, $W(1/2)$, and $W(1)$ is $(1/4)(1/4)/(1/2) = 1/8$. Similarly, the conditional variance of $W(3/4)$ is 1/8. Thus, to simulate these points we generate two more independent standard normal random variables $Z_3$ and $Z_4$, and define

$$W(1/4) = X^{(1)}(1/4) + \frac{1}{\sqrt{8}} Z_3,$$
$$W(3/4) = X^{(1)}(3/4) + \frac{1}{\sqrt{8}} Z_4.$$

The approximation $X^{(2)}$ is then defined to be the piecwise linear interpolation of the simulated values we have obtained for the times 0, 1/4, 1/2, 3/4, and 1, as in Figure 3.

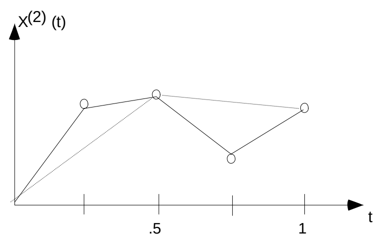



Figure 3

One more time: Given the values simulated so far, each of $W(1/8), W(3/8), W(5/8)$, and $W(7/8)$ has conditional variance (1/8)(1/8)/(1/4)=1/16. So we can simulate $W(1/8), W(3/8), W(5/8)$, and $W(7/8)$ by multiplying some more standard normal random variables $Z_5, Z_6, Z_7, Z_8$ by $\sqrt{1/16} = 1/4$ and adding these to the values $X^{(2)}(1/8), X^{(2)}(3/8), X^{(2)}(5/8)$, and $X^{(2)}(7/8)$ given by the previous approximation. The piecewise linear interpolation gives $X^{(3)}$.

And so on. In general, to get from $X^{(n)}$ to $X^{(n+1)}$, we generate $2^n$ new standard normal random variables $Z_{2^n+1}, Z_{2^n+2}, \ldots, Z_{2^{n+1}}$, multiply these by the appropriate conditional standard deviation $\sqrt{2^{-n-2}} = 2^{-(n/2)-1}$, and add to the values $X^{(n)}(1/2^{n+1}), X^{(n)}(3/2^{n+1}), \ldots, X^{(n)}(1 - 1/2^{n+1})$ to get the new values $X^{(n+1)}(1/2^{n+1}), X^{(n+1)}(3/2^{n+1}), \ldots, X^{(n+1)}(1 - 1/2^{n+1})$.

(5.17) CLAIM. *With probability 1, the sequence of functions $X^{(1)}, X^{(2)}, \ldots$ converges uniformly over the interval [0,1].*

The importance of the uniformity of the convergence stems from the following fact from analysis:

| |
|---|
| The limit of a uniformly convergent sequence of continuous functions is a continuous function. |

⟦To appreciate the need for uniformity of convergence in order to be guaranteed that the limit function is continuous, recall the following standard example. For $n = 1, 2, \ldots$ consider the function $t \mapsto t^n$ for $t \in [0, 1]$. Then as $n \to \infty$, this converges to 0 for all $t < 1$ whereas it converges to 1 for $t = 1$, so that the limit is not a continuous function.⟧ Since each of the functions $X^{(n)}$ is clearly continuous, the claim then implies that with probability 1, the sequence $X^{(1)}, X^{(2)}, \ldots$ converges to a limit function $X$ that is continuous.

PROOF: Define the maximum difference $M_n$ between $X^{(n+1)}$ and $X^{(n)}$ by

$$M_n = \max_{t \in [0,1]} |X^{(n+1)}(t) - X^{(n)}(t)|.$$

It is clear that if $\sum M_n < \infty$, then the sequence of function $X^{(1)}, X^{(2)}, \ldots$ converges uniformly over [0,1]. Thus, it is sufficient to show that $P\{\sum M_n < \infty\} = 1$. Observe that

$$M_n = 2^{-(n/2)-1} \max\{|Z_{2^n+1}|, |Z_{2^n+2}|, \ldots, |Z_{2^{n+1}}|\}.$$

We will use the following result about normal random variables.

(5.18) FACT. *Let $G_1, G_2, \ldots$ be iid standard normal random variables, and let $c$ be a number greater than 2. Then*

$$P\{|G_n| \le \sqrt{c \log n} \text{ for all sufficiently large } n\} = 1.$$

PROOF: Remember the tail probability bound

$$P\{G > x\} \le \frac{\varphi(x)}{x} = \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x}$$

for a standard normal random variable $G$ and for $x > 0$. From this,

$$\sum_{n=1}^{\infty} P\{|G_n| > \sqrt{c \log n}\} \le 2 \frac{1}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{e^{-(1/2)c \log n}}{\sqrt{c \log n}} = \sqrt{\frac{2}{\pi}} \sum_{n=1}^{\infty} \frac{n^{-(1/2)c}}{\sqrt{c \log n}},$$

which is finite for $c > 2$. Thus, by the Borel-Cantelli lemma,

$$P\{|G_n| > \sqrt{c \log n} \text{ infinitely often}\} = 0,$$

which is equivalent to the desired statement. □

Taking $c > 2$, the fact implies that with probability 1,

$$M_n \le 2^{-(n/2)-1} \sqrt{c \log(2^{n+1})}$$

holds for all sufficiently large $n$. That is,

$$P\{M_n \le 2^{-(n/2)-1} \sqrt{n+1} \sqrt{c \log 2} \text{ eventually}\} = 1.$$

Thus, since $\sum 2^{-(n/2)} \sqrt{n+1} < \infty$, we have $\sum M_n < \infty$ with probability 1, which completes the proof of Claim (5.17). □

So we know that with probability 1, the limit $X = \lim_{n \to \infty} X^{(n)}$ is a well-defined, continuous function. It remains to check that $X(t) \sim N(0,t)$ and the process $X$ has stationary independent increments. By construction, our process $X$ satisfies these properties at least at all *dyadic rational* $t$, that is, values of $t$ of the form $t = k/(2^n)$. From here, one uses the fact that any $t \in [0,1]$ is the limit of some sequence of dyadic rationals $\{t_n\}$, say; that is, $t = \lim t_n$. So, by path continuity of the $X$ process, $X(t) = \lim X(t_n)$. Thus, the conclusion $X(t) \sim N(0,t)$ follows from the following fact, whose proof is an exercise.

(5.19) FACT. *Suppose that $\{Y_n\}$ is a sequence of random variables that converges with probability 1 to a limit $Y$, with $P\{|Y| < \infty\} = 1$. Suppose that each $Y_n$ has a normal distribution: $Y_n \sim N(\mu_n, \sigma_n^2)$. Show that as $n \to \infty$, $\mu_n$ and $\sigma_n$ converge to finite limits, and $Y \sim N(\mu, \sigma^2)$, where $\mu_n \to \mu$ and $\sigma_n \to \sigma$.*

(5.20) EXERCISE. *Prove the fact. ⟦Hint: This is not hard using characteristic functions or moment generating functions.⟧*

The stationary and independent increments properties are proved by similar arguments.

(5.21) EXERCISE. *Complete the proof that $X$ is a standard Brownian motion by showing that the increments of $X$ have the right joint distribution. For example, show that for $0 < t_1 < t_2$, we have $X(t_1) \sim N(0,t_1)$, $X(t_2) - X(t_1) \sim N(0, t_2 - t_1)$, and $X(t_1)$ is independent of $X(t_2) - X(t_1)$.*
*⟦Hint: Joint characteristic functions. For example, consider $E(e^{i[\lambda_1 X(t_1) + \lambda_2(X(t_2)-X(t_1))]})$.⟧*

## 5.6 The Brownian bridge

A standard Brownian bridge over the interval [0,1] is a standard Brownian motion $W(\cdot)$ conditioned to have $W(1) = 0$. People say the Brownian motion is "tied down" at time 1 to have the value 0. By Exercise (5.15), we know that $E(W(t) \mid W(1) = 0) = 0$ and $\mathrm{Cov}(W(s), W(t) \mid W(1) = 0) = s(1-t)$ for $0 \le s \le t \le 1$.

(5.22) DEFINITION. *A standard Brownian bridge is a Gaussian process $X$ with continuous paths, mean 0, and covariance function $\mathrm{Cov}(X(s), X(t)) = s(1-t)$ for $0 \le s \le t \le 1$.*

Here is an easy way to manufacture a Brownian bridge from a standard Brownian motion: define

$$X(t) = W(t) - tW(1) \quad \text{for } 0 \le t \le 1. \tag{5.23}$$

It is easy and pleasant to verify that the process $X$ defined this way satisfies the definition of a Brownian bridge; I wouldn't dream of denying you the pleasure of checking it for yourself! Notice that, given the construction of standard Brownian motion $W$, now we do not have to worry about the existence or construction of the Brownian bridge. Another curious but sometimes useful fact is that the definition

$$Y(t) = (1-t)W\left(\frac{t}{1-t}\right) \quad \text{for } 0 \le t < 1, \quad Y(1) = 0 \tag{5.24}$$

also gives a Brownian bridge.

(5.25) EXERCISE. *Verify that the definitions (5.23) and (5.24) give Brownian bridges.*

(5.26) EXERCISE. *We defined the Brownian bridge by conditioning a standard Brownian motion to be 0 at time 1. Show that we obtain the same Brownian bridge process if we start with a* $(\mu, 1)$ *Brownian motion and condition it to be 0 at time 1.*

### 5.6.1 A boundary crossing probability

Earlier, using the reflection principle, we found the probability that a Brownian motion reaches a certain height by a certain time. What is this probability for a Brownian bridge? Letting $W$ be a standard Brownian motion, recall the definition of the first hitting time of the positive level $b$:

$$\tau_b = \inf\{t : W(t) = b\}.$$

In the standard Brownian bridge, we considered the particular condition $W(1) = 0$. Instead of the particular time 1 and the particular value 0, let us consider the general condition where we tie the Brownian motion down at an arbitrary time $t$ to an arbitrary value $x$, so that we are interested in the probability $P\{\tau_b \le t \mid W(t) = x\}$. Clearly by path continuity the answer is 1 if $x \ge b$, so let us assume that $x < b$. Adopting rather informal notation, we have

$$P\{\tau_b \le t \mid W(t) = x\} = \frac{P\{\tau_b \le t, W(t) \in dx\}}{P\{W(t) \in dx\}}. \tag{5.27}$$

Heuristically, $dx$ is a tiny interval around the point $x$. Or, somewhat more formally, think of the $dx$ as shorthand for a limiting statement—the usual limiting idea of conditioning on a random variable taking on a particular value that has probability 0. We can calculate the right side of (5.27) explicitly as follows:

$$\begin{aligned}
\text{numerator} &= P\{\tau_b \le t\}P\{W(t) \in dx \mid \tau_b < t\} \\
&= P\{\tau_b \le t\}P\{W(t) \in 2b - dx \mid \tau_b < t\} \\
&= P\{W(t) \in 2b - dx, \tau_b < t\} \\
&= P\{W(t) \in 2b - dx\} \\
&= \frac{1}{\sqrt{t}}\varphi\left(\frac{2b - x}{\sqrt{t}}\right)\, dx
\end{aligned}$$

and of course

$$\text{denominator} = \frac{1}{\sqrt{t}}\varphi\left(\frac{x}{\sqrt{t}}\right)\, dx,$$

so that

$$P\{\tau_b < t \mid W(t) = x\} = \frac{\varphi\left(\frac{2b-x}{\sqrt{t}}\right)}{\varphi\left(\frac{x}{\sqrt{t}}\right)}$$

$$= \exp\Big[-\frac{1}{2}\frac{(2b-x)^2}{t} + \frac{1}{2}\frac{x^2}{t}\Big]$$
$$= \exp\Big[\frac{-2b(b-x)}{t}\Big],$$

a handy formula! Of course, it makes qualitative sense: the probability goes to 0 ⟦very fast!⟧ as $b \to \infty$, and the probability is nearly 1 if $b$ is small or if $b - x$ is small or if $t$ is large.

(5.28) EXERCISE. *Let $X(t) = x_0 + \mu t + \sigma W(t)$ be a $(\mu, \sigma^2)$ Brownian motion starting from $x_0$ at time 0. What is the probability*

$$P\{X(s) \ge b + cs \textit{ for some } s \le t \mid X(t) = x_t\}?$$

*⟦Your answer should be a function of $\mu$, $\sigma$, $x_0$, $x_t$, $b$, $c$, and $t$ (maybe not depending on all of these). This should not require significant calculation, but rather a reduction to something we have done.⟧*

### 5.6.2 Application to testing for uniformity

Suppose $U_1, \ldots U_n$ are *iid* having a distribution $F$ on [0,1], and we are interested in testing the hypothesis that $F$ is the uniform distribution $F(t) = t$. The empirical distribution function $F_n$ is defined by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{U_i \le t\} \quad \text{for } 0 \le t \le 1.$$

Thus, $F_n(t)$ is the fraction of the sample that falls in the interval $[0, t]$; this is a natural estimator of $F(t)$, the probability that one random observation from $F$ falls in $[0, t]$. By the law of large numbers, $F_n(t) \to F(t)$ for all $t$ as $n \to \infty$. So if $F$ is Unif[0,1], we have $F_n(t) \to t$. The idea of the ***Kolmogorov-Smirnov test*** is to look at the difference $F_n(t) - t$, and reject the uniformity hypothesis if the difference gets large enough at any $t \in [0, 1]$. The question is: how large is large enough? For example, we might want to find a rejection threshold that gives a probability of false rejection of .05; that is, find $b$ so that $P\{\max(F_n(t) - t) : t \in [0, 1]\} = .05$.

Again, the Strong Law of Large Numbers says that for all $t$, the difference $F_n(t) - t$ approaches 0 as $n \to \infty$. A limit distribution is obtained by multiplying the difference by $\sqrt{n}$: since

$$\mathrm{Var}(I\{U_1 \le t\}) = P\{U_1 \le t\} - (P\{U_1 \le t\})^2 = t(1-t)$$

the Central Limit Theorem tells us that

$$\sqrt{n}(F_n(t) - t) \xrightarrow{\mathcal{D}} N(0, t(1-t)).$$

So define $X_n(t) = \sqrt{n}(F_n(t) - t)$. Then, similarly to the above, since

$$\mathrm{Cov}(1_{\{U_1 \le s\}}, 1_{\{U_1 \le t\}}) = P\{U_1 \le s, U_1 \le t\} - (P\{U_1 \le s\})(P\{U_1 \le t\}) = s - st = s(1-t)$$

for $s \le t$, the vector Central Limit Theorem tells us that

$$\begin{pmatrix} X_n(s) \\ X_n(t) \end{pmatrix} = \begin{pmatrix} \sqrt{n}(F_n(s) - s) \\ \sqrt{n}(F_n(t) - t) \end{pmatrix} \xrightarrow{\mathcal{D}} N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s & s \\ s & t \end{pmatrix} \right) \sim \begin{pmatrix} X(s) \\ X(t) \end{pmatrix},$$

where $X$ is a Brownian bridge, and, in general,

$$\begin{pmatrix} X_n(t_1) \\ \vdots \\ X_n(t_k) \end{pmatrix} \xrightarrow{\mathcal{D}} N\left( \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} t_1 & \cdots & t_1 \\ \vdots & & \\ t_1 & & t_k \end{pmatrix} \right) \sim \begin{pmatrix} X(t_1) \\ \vdots \\ X(t_k) \end{pmatrix}.$$

Thus, as $n \to \infty$, the joint distribution of process $X_n$ sampled at any finite number of time points converges to the joint distribution to the Brownian bridge $X$ sampled at those same times. Therefore, for any finite collection of times $T = \{t_1, \ldots, t_k\} \subset [0,1]$,

$$\lim_{n\to\infty} P\{\max\{X_n(t) : t \in T\} \ge b\} = P\{\max\{X(t) : t \in T\} \ge b\}$$

This leads one to suspect that we should also have

$$\lim_{n\to\infty} P\{\max\{X_n(t) : t \in [0,1]\} \ge b\} = P\{\max(X(t) : t \in [0,1]) \ge b\}$$

In fact, this last convergence can be rigorously shown; the proof is a bit too involved for us to get into now. ⟦For the general subject of weak convergence of stochastic processes see the books *Convergence of Stochastic Processes* by David Pollard and *Convergence of Probability Measures* by P. Billingsley.⟧ Since we know the exact expression for the last probability, we can say that

$$\lim_{n\to\infty} P\{\max\{X_n(t) : t \in [0,1]\} \ge b\} = e^{-2b^2}.$$

Thus, for example, since $e^{-2b^2} = 0.05$ for $b = 1.22$, then if $n$ is large we have

$$P\{\max\{X_n(t) : t \in [0,1]\} \ge 1.22\} \approx 0.05.$$

So we have found an approximate answer to our question of setting a rejection threshold in the test for uniformity.

## 5.7 Two Approaches to a Simple Boundary Crossing Problem

Let $W$ be standard Brownian motion as usual, let $b > 0$, and define $\tau_b = \inf\{t : W_t = b\}$. Recall that the reflection principle gives

$$P\{\tau_b \le t\} = 2P\{W_t > b\} = 2\left[1 - \Phi\left(\frac{b}{\sqrt{t}}\right)\right],$$

from which, letting $t \to \infty$ with $b$ fixed, it follows that $P\{\tau_b < \infty\} = 1$. That is, $W$ is sure to cross the horizontal level $b$ eventually.

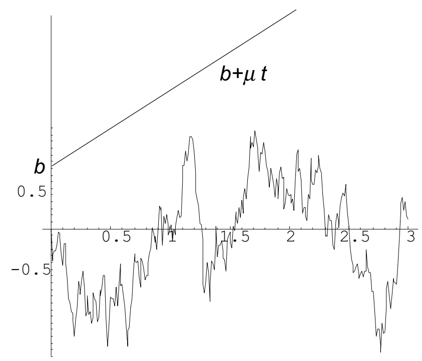
More generally, we could ask:

**Problem A:** *What is the probability*

$$P\{W(t) = b + \mu t \quad \text{for some } t \geq 0\}$$

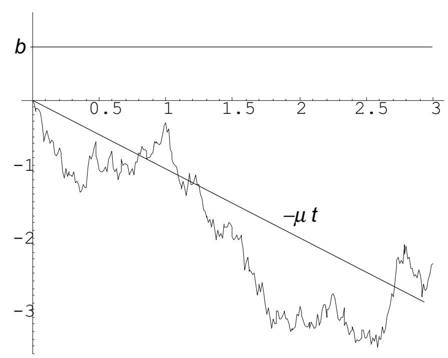
*that the process $W$ ever crosses the linear boundary $b + \mu t$?*

Clearly the answer is 1 if $\mu < 0$, and we just observed in the previous paragraph that the answer is also 1 when $\mu = 0$. Accordingly, let's look at linear boundaries having positive slope $\mu > 0$.




Note that if we subtract $\mu t$ from $W$, we get Brownian motion with drift $-\mu$, and if we subtract $\mu t$ from the line $b+\mu t$, we get the horizontal level $b$. Thus, letting $X(t) := W(t) - \mu t$ and defining the stopping time $\tau_b$ by $\tau_b = \inf\{t : X_t = b\}$, it is clear that our problem is equivalent to the following problem.

**Problem A$'$:** *Find $P_{-\mu}\{\tau_b < \infty\}$. Here the subscript "$-\mu$" is attached to the $P$ in order to remind us that the Brownian motion that we are considering has drift $-\mu$.*

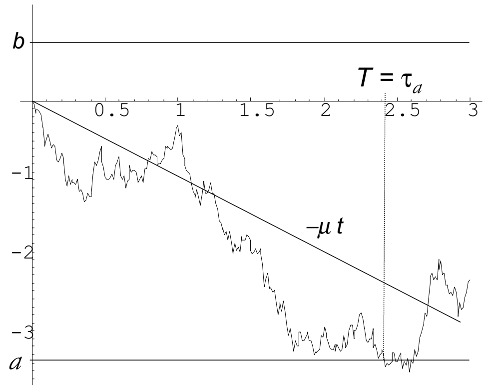



To solve problems A and A$'$, we will in fact solve the following "better" problem.

**Problem B:** *Let $a < 0 < b$ and $\mu$ be given, and define $T = \min\{\tau_a, \tau_b\}$. What is the probability*

$$P_{-\mu}\{\tau_b < \tau_a\} = P_{-\mu}\{X_T = b\}$$

*that the Brownian motion hits the level $b$ before the level $a$?*




Solving Problem B will enable us to solve the other two, since

$$P_{-\mu}\{\tau_b < \infty\} = \lim_{a \to -\infty} P_{-\mu}\{\tau_b < \tau_a\}$$

.

### 5.7.1 Differential Equations

We could formulate Problem B in terms of an absorbing Markov process. Consider a $(-\mu,1)$-Brownian motion $\{X_t\}$ having two absorbing states $a$ and $b$. Then our problem is to find the probability that the process gets absorbed in $b$ rather than $a$. This should have a familiar feel to it: we solved this sort of problem for finite-state Markov chains earlier. In fact, we did this by two different methods, both of which solved the given problem by simultaneously solving the whole family of analogous problems starting from all of the possible starting states of the chain. The first method used the "fundamental matrix." The second method involved conditioning on what the chain did at the first step.

Here we will do the same sort of thing, using the continuous-time analog of the "conditioning on what happened at the first step" method. We won't try to be rigorous here. Let $P^x$ and $E^x$ denote probability and expectation when the $(-\mu,1)$-Brownian motion $\{X_t\}$ starts in the state $X_0 = x$. Then, defining the function $u(x) = P^x\{X_T = b\}$, Problem B asks for the value of $u(0)$, which we will find by in fact solving for the whole function $u(x)$ for $x \in [a, b]$. Clearly $u(a) = 0$ and $u(b) = 1$, so let $x \in (a, b)$. In continuous time there is no "first step" of the process, but we can think of conditioning on the value of $X(h)$ where $h$ is a tiny positive number. This gives

$$u(x) = E^x P^x\{X_T = b | X(h)\} = E^x u[X(h)] + o(h),$$

where in the last equality we have used the Markov "restarting" property to say that $P^x\{X_T = b | X(h), T > h\} = u[X(h)]$, and Exercise (5.12) to say that $P^x\{T \le h\} = o(h)$.

Now since $h$ is tiny, $X(h)$ will be very close to $x$ with high probability under $P^x$. So $u[X(h)]$ can be closely approximated by the first few terms of its Taylor expansion

$$u[X(h)] = u(x) + u'(x)[X(h) - x] + (1/2)u''(x)[X(h) - x]^2 + \cdots.$$

Combining this with the previous equation gives

$$u(x) = u(x) + u'(x)E^x[X(h) - x] + (1/2)u''(x)E^x\{[X(h) - x]^2\} + \cdots.$$

However, the $(-\mu,1)$-Brownian motion $\{X_t\}$ satisfies

$$E^x[X(h) - x] = -\mu h + o(h),$$
$$E^x\{[X(h) - x]^2\} = h + o(h),$$

and

$$E^x\{[X(h) - x]^k\} = o(h) \quad \text{for } k > 2.$$

Thus,

$$0 = u'(x)(-\mu h) + (1/2)u''(x)h + o(h),$$

so that, dividing through by $h$ and letting $h \downarrow 0$, we see that $u$ satisfies the differential equation

(5.29) $$(1/2)u''(x) - \mu u'(x) = 0.$$

The boundary conditions are $u(a) = 0$ and $u(b) = 1$. This differential equation is very easy to solve: the general solution is $u(x) = Ce^{2\mu x} + D$ where $C$ and $D$ are constants, so the solution satifying the boundary conditions is

(5.30) $$u(x) = \frac{e^{2\mu x} - e^{2\mu a}}{e^{2\mu b} - e^{2\mu a}}.$$

This is a handy, explicit result, from which the solution to our original problems follow easily. In particular, since we wanted our Brownian motion $X$ to start at 0, we are interested in

$$P_{-\mu}\{\tau_b < \tau_a\} = u(0) = \frac{1 - e^{2\mu a}}{e^{2\mu b} - e^{2\mu a}}.$$

Since $\mu > 0$, by letting $a \to -\infty$ we obtain

$$P_{-\mu}\{\tau_b < \infty\} = \lim_{a \to -\infty} P_{-\mu}\{\tau_b < \tau_a\} = e^{-2\mu b}.$$

Let's pause to make sure that (5.30) is consistent with what we know about the cases where $\mu$ is not positive. If $\mu$ were negative, so that $-\mu$ were positive, then we would have obtained the limit $\lim_{a\to-\infty} u(0) = 1$, which makes sense, as we said before: Brownian motion with positive drift is sure to pass through the level $b$ eventually. What if $\mu = 0$? In that case the solution (5.30) breaks down, reducing to the form "0/0". What happens is that the differential equation (5.29) becomes $u''(x) = 0$, whose solutions are linear functions of $x$. The solution satisfying the boundary conditions $u(a) = 0$ and $u(b) = 1$ is

$$u(x) = \frac{x - a}{b - a},$$

so that $u(0) = a/(a - b)$, which approaches 1 as $a \to -\infty$, again as expected.

It is interesting to contemplate our solution still a bit more. Let $M$ denote the maximum height ever attained by the Brownian motion, that is, $M = \max\{X_t : t \geq 0\}$. Then what we have found is that $P_{-\mu}\{M \geq b\} = e^{-2\mu b}$, or, in other words, $M \sim \text{Exp}(2\mu)$. Now this result also makes a good deal of sense, at least qualitatively: we could have guessed that $M$ should be exponentially distributed, since it seems intuitively clear that $M$ should have the memoryless property. To see this, suppose I tell you that I observed $M$ to be greater than $b$, and I ask you for your conditional probability that $M > b + y$. Then you should think to yourself: "He just told me that $\tau_b < \infty$. The portion of the BM path after time $\tau_b$ should look just like ordinary BM with drift $-\mu$, except started at the level $b$. ⟦Then you should mutter something about the strong Markov property: $\tau_b$ is a stopping time.⟧ So I should say $P_{-\mu}\{M > b + y \mid M > b\} = P_{-\mu}\{M > y\}$." This is the memoryless property.

The fact that the parameter of the exponential distribution should be $2\mu$ does not seem obvious, so our calculation has given us some substantial information. Again, you should check that $2\mu$ at least makes some crude, qualitative sense; for example, it is monotone in the right direction and that sort of thing.

### 5.7.2 Martingales

The results we discussed in the previous chapter for martingales in discrete time have counterparts for continuous-time martingales. Here is a definition. Suppose that $W$ is a standard Brownian motion.

(5.31) DEFINITION. *A stochastic process $M$ is a martingale with respect to $W$ if it satisfies the following two conditions:*

1. *$M$ is adapted to $W$; that is, for every $t$, $M(t)$ is some deterministic function of the portion $\langle W\rangle_0^t := \{W(s) : 0 \leq s \leq t\}$ of the path of $W$ up to time $t$.*
2. *For all $0 < s < t$ we have $E\{M(t)|\langle W\rangle_0^s\} = M(s)$.*

The second condition is a "fair game" sort of requirement. If we are playing a fair game, then we expect to neither win nor lose money on the average. Given the history of our fortunes up to time $s$, our expected fortune $M(t)$ at a future time $t > s$ should just be the fortune $M(s)$ that we have at time $s$.

The important *optional sampling* ("conservation of fairness") property of martingales extends to continuous time. Let $M$ be a martingale with respect to $W$, and suppose that we know that $M(0)$ is just the constant $m_0$. By the "fair game" property, $EM(t) = m_0$ for all times $t \geq 0$. ⟦Exercise: check this!⟧ That is, I can say "stop" at any predetermined time $t$, like $t = 8$, say, and my winnings will be "fair": $EM(8) = m_0$. As before, the issue of optional sampling is this: If $\tau$ is a *random* time, that is, $\tau$ is a nonnegative random variable, does the equality $EM(\tau) = m_0$ still hold? As before, we can be assured that this holds if we rule out two sorts of obnoxious behaviors: "taking too long" and "taking back moves." That is, optional sampling holds for bounded stopping times.

(5.32) DEFINITION. *We say that a nonnegative random variable $\tau$ is a stopping time (with respect to $W$) if for each $t$ it is possible to determine whether or not $\tau \leq t$ just by looking at $\langle W\rangle_0^t$. That is, the indicator random variable $I\{\tau \leq t\}$ is a function of $\langle W\rangle_0^t$.*

⟦We could also perversely express this in the language introduced above by saying that $\tau$ is a stopping time if the process $X$ defined by $X(t) = \{\tau \le t\}$ is adapted to $W$.⟧ We say that a random time $\tau$ is bounded if there is a number $c$ such that $P\{\tau \le c\} = 1$.

(5.33) Theorem [Optional Sampling Theorem]. *Suppose that $M$ is a martingale with respect to $W$ starting at the value $M(0) = m_0$, and let $\tau$ be a bounded stopping time with respect to $W$. Then we have $EM(\tau) = m_0$.*

You may be disappointed by the boundedness restriction on the stopping time $\tau$. However, often we can prove optional sampling for unbounded stopping times by combining the above optional sampling theorem with a result like the bounded convergence theorem, for example. We will see this in our application, to which I think it is high time we got back now. If in the force Yoda's so strong, construct a sentence with the words in the proper order then why can't he?

The next result introduces a martingale called "Wald's martingale" or "the exponential martingale." Any martingale with more than one name must be important!

(5.34) Claim. *For any real $\lambda$, $M(t) := \exp\{\lambda W(t) - \frac{1}{2}\lambda^2 t\}$ is a martingale.*

Proof: Easy exercise. Use the fact that if $Z \sim N(0,1)$, then $E\{e^{\theta Z}\} = e^{\theta^2/2}$ for real $\theta$; this is the moment generating function of the $N(0,1)$ distribution. □

We can use this to form a martingale out of our process $X(t) = W(t) - \mu t$, which has drift $-\mu$: since $X(t) + \mu t$ is a standard Brownian motion, for every $\lambda$

$$M(t) := \exp\{\lambda[X(t) + \mu t] - \frac{1}{2}\lambda^2 t\} = \exp\{\lambda X(t) + \lambda(\mu - \frac{1}{2}\lambda)t\} \tag{5.35}$$

is a martingale. The nice thing is that since this holds for every $\lambda$, we are free to choose any $\lambda$ that we like. There is a clear choice here that appears to simplify things: if we take $\lambda = 2\mu$, we see that

$$M(t) := e^{2\mu X(t)} \text{ is a martingale.}$$

Retaining the notation $T = \min\{\tau_a, \tau_b\}$ from before, in accordance with the optional sampling ideas we discussed above, we would like to say that

$$E\{e^{2\mu X(T)}\} = EM(T) = M(0) = 1.$$

Is this right? Well, clearly $T$ is a stopping time; that's good. However, $T$ is not bounded; that might be bad. Here's the trick. For any number $n$, the random time $T \wedge n$ is a bounded stopping time, so that we can say that $E\{e^{2\mu X(T\wedge n)}\} = 1$. So clearly we would like to take a limit as $n \to \infty$ as follows

$$\begin{aligned} E\{e^{2\mu X(T)}\} &= E\left\{\lim_{n\to\infty} e^{2\mu X(T\wedge n)}\right\} \\ &= \lim_{n\to\infty} E\left\{e^{2\mu X(T\wedge n)}\right\} \\ &= \lim_{n\to\infty} 1 = 1. \end{aligned}$$

This interchange of limit and expectation is permitted by the bounded convergence theorem in our case, since the fact that $X(T \wedge n)$ must be between $a$ and $b$ implies that the random variables $e^{2\mu X(T \wedge n)}$ are bounded [[between $e^{2\mu a}$ and $e^{2\mu b}$]].

(5.36) EXERCISE. *Strictly speaking, in order to apply bounded convergence this way, we should show that $T$ is finite with probability one. Can you do that? How about if $\mu = 0$?*

We are almost done. Write

$$1 = E\{e^{2\mu X(T)}\} = e^{2\mu a} P\{X(T) = a\} + e^{2\mu b} P\{X(T) = b\}.$$

Noting that $P\{X(T) = a\} = 1 - P\{X(T) = b\}$, this becomes a simple equation for the desired probability $P\{X(T) = b\}$, and the answer is the same as before.

(5.37) EXERCISE. *We can do more with this approach; the martingale (5.35) is quite powerful. For example, let $\mu > 0$ and $b > 0$ and consider $X(t)$ to be BM with drift $\mu$. Find the "Laplace transform" (or moment generating function) $E(e^{\theta \tau_b})$ as a function of $\theta$. Note that the answer is finite for some values of $\theta$ and infinite for others. Can you solve the same problem for $T = \tau_a \wedge \tau_b$?*

Let's end this section with something really strange. If you really think about the optional sampling result, from a certain point of view it is exceedingly strange. Here is what I mean. We know that if $T$ is any bounded stopping time and $W$ is a standard Brownian motion, then $E[W(T)]=0$. In gambling terms, you cannot stop a Brownian motion before time 1, for example, and make a profit — i.e. end up with a positive expected amount of money. However, we also know that with probability 1, a Brownian motion *must* become positive before time 1. In fact, in each interval of the form $(0, \epsilon)$, $W$ hits 0 infinitely often and is positive infinitely often and is negative infinitely often. Not only that, but since $W$ has continuous paths, whenever $W$ is positive at any point in time, in fact there is a whole *interval* of times over which $W$ is positive. So, with probability 1, if we ride along the Brownian path, before time 1 we will experience many intervals of times on which the Brownian motion is positive. Can't we just look down, see the time axis below us, think "Good, I'm positive now," and say "Stop"? It doesn't seem hard, does it? If you are a miscroscopic rider on the graph of a Brownian path, there will be all these stretches of time over which the Brownian motion is positive. You don't have to be greedy; just choose any of those times and say "Stop." Isn't it clear that you can do this with probability 1? If so, your winnings will be positive with probability 1, and so obviously your expected winnings are positive. But the optional sampling theorem implies that there is no such stopping time!

## 5.8 Some confusing questions (or answers)

I suspect the effect of this section may be to toggle your state of confusion about some issues — if you were not already confused and searching for answers, this section may confuse you, whereas if you have been wondering about these issues yourself, I hope this section will help you sort them out. So, have you ever wondered about questions of this sort?

(1) What is the $\Omega$ for a Brownian motion?
(2) What does it mean to "fix $\omega$" for a pollen particle in water?

I've been asked questions like these a number of times by a number of students, and they are good questions (perhaps a little too good... grumble grumble...).

The answer to (1) is: the set $\Omega$ is unimaginably complicated. Or not. Actually, there are many ways to answer this sort of question. This shouldn't be surprising. For example, there are many different random variables that have (or are commonly modeled as having) a $N(0,1)$ distribution; for example, (women's height $-$ 65 inches)/3, (IQ $-$ 100)/10, and so on. Just as there are many different $N(0,1)$ random variables, there are many different Brownian motions.

Let's start by reviewing these ideas in a context much simpler than Brownian motion: tossing a coin, just once. We want to define a random variable $X$ that models a coin toss, according to the mathematical framework of probability. That is, we need a probability space $(\Omega, \mathcal{F}, P)$ and then we define $X$ as a function $X : \Omega \to \mathbb{R}$. ⟦Recall $\Omega$ is called the sample space, and $P$ is a probability measure on $\Omega$. $\mathcal{F}$ is a collection of subsets of $\Omega$ called events.⟧

A standard description of the concept of a sample space $\Omega$ is that "$\Omega$ is the set of all possible outcomes of the experiment under consideration." Here that would be $\Omega = \{H, T\}$. So defining the probability measure $P$ by $P\{H\} = P\{T\} = 1/2$ and the random variable $X$ by $X(\omega) = \omega$ for $\omega \in \Omega$, we have a model. Notice that the random variable $X$ here is rather trivial—the identity function. Given this generic choice for $X$, we have customized the probability $P$ to model the phenomenon.

Here is another way to model a coin toss. Imagine simulating the toss using a uniformly distributed random number: take $\Omega = [0,1]$, the unit interval, and $P$=Lebesgue measure (ordinary length) on $\Omega$. Then we could define $X$ by $X(\omega) = H$ if $\omega \le 1/2$ and $X(\omega) = T$ if $\omega > 1/2$. Notice that this is an entirely different random variable than the $X$ defined in the previous paragraph: they are different functions, with different domains! However, the two random variables have the same probability distribution: each satisfies $P\{\omega \in \Omega : X(\omega) = H\} = P\{\omega \in \Omega : X(\omega) = T\} = 1/2$. In this second setup we have used a generic sort of source of randomness: a uniformly distributed $\omega$. So $P$ was not tailored to our application here; it is the random variable $X$ that was carefully defined to give the desired distribution. Notice the contrast with the last sentences of the previous paragraph.

Now for a more physical picture. Imagine a person actually tossing a coin and letting it fall to the floor. What is it about the randomness in this situation that we would need to specify in order to know the outcome? This motivates a description like $\Omega = \{$all possible initial conditions, i.e., all possible values for (initial position, initial velocity, initial angular velocity)$\}$. Here $P$ would be the probability distribution over $\Omega$ that describes the way that our flipper will "choose" initial conditions. And $X$ is a complicated function that you could in principle write down from the laws of physics ⟦good luck⟧ that tells us, for each possible initial condition, whether the toss will be heads or tails.

What does it mean to fix $\omega$? For each fixed $\omega$, the outcome $X(\omega)$ is determined—recall that $X$ is simply a function of $\omega$. So having fixed $\omega$, there is no randomness left in $X(\omega)$. The "randomness" is all in the choice of $\omega$; this applies to each of the three descriptions above.

Let's stop modeling a coin toss now; you can think of other probability spaces and random variables that we could use to do it. Now to Brownian motion: each of the 3 ways of thinking described above has a natural analog here. The first approach had $\Omega$ as the set of all possible outcomes. We know that the "outcome" of Brownian motion is a continuous path. So we could take $\Omega$ to be the set of all continuous functions

$$\Omega = C[0,\infty) = \{\omega : \omega(\cdot) \text{ is a continuous function on } [0,\infty)\}.$$

Then $P$ would be the probability measure on $C[0,\infty)$ that corresponds to Brownian motion; this is called *Wiener measure*. It is rather complicated conceptually; an example of a simple statement that one could make about Wiener measure $P$ is

$$P\{\omega \in C[0,\infty) : \omega(1) < 2\} = \Phi(2)$$

where $\Phi$ is the standard normal cumulative distribution function. The definition of Brownian motion $W$ in terms of this sample space is trivially simple, just as the definition of $X$ was trivial in the first example above; just define $W(\omega) = \omega$. That is, for each $t$, we define $W_t(\omega) = \omega(t)$. So simple it looks like gibberish. ⟦Remember, $\omega$ is a function in its own right; it is an element of $C[0,\infty)$.⟧ The function $W$ is trivial; the interesting part is the Wiener measure.

A second approach uses a simpler measure and a more complicated function. The question comes down to: how could you simulate a realization (that is, a path) of Brownian motion? We have seen (from our "construction" of Brownian motion) that it can be done from an independent sequence of uniformly distributed random variables. But in fact a whole sequence of such random variables can be produced from a single uniformly distributed random variable. ⟦How?⟧ So Brownian motion can be defined on the nice friendly probability space $[0,1]$ with Lebesgue measure. The tradeoff for the simple probability space is that the function $W$ must then be more complicated — we can't get by with the trivial identity function any more! This function must perform the tasks of producing a sequence of *iid* uniforms, transforming them to a sequence of *iid* $N(0,1)$'s, then combining them into the series of functions discussed in the construction of Brownian motion, all starting from a single uniformly distributed random variable.

Finally, one could imagine modeling an actual pollen particle in a drop of water. We could define $\Omega$ to be the set of all possible initial conditions of the positions and velocities of the pollen particle and of all of the molecules of water in the drop. Wow. Then $P$ would be our probability distribution for such initial conditions. The function $W$ would again be determined ⟦in principle!⟧ from the laws of physics, with $W_t(\omega)$ giving the position of the pollen at time $t$ if the initial conditions were $\omega$.

Here are some words. A function $X : \Omega \to \mathbb{R}$ is usually called a random variable. We can consider functions from $\Omega$ to a more general space $\mathcal{X}$, say. Such a function $X : \Omega \to \mathcal{X}$ would be called a *random variable taking values in* $\mathcal{X}$ or a *random element of* $\mathcal{X}$. We can consider a stochastic process such as $W$ to be a random element of a function space such as $C[0,\infty)$, since, for each $\omega$, we get a whole path (i.e. continuous function) $W(\omega)$.

I hope you are not put off by all these different modeling approaches. Thankfully, as we have seen, most of the time, we don't have to agonize over just what $\Omega$ we have in our minds and how complicated or trivial the corresponding random element is. This works because usually we are interested in probabilities that the random element is in various sets, and these depend only on the probability distribution of the random element. It is well to keep in mind that this is all a game going on inside our heads. Of course, the same statement applies to all probability and statistics, and indeed to all of mathematics (and perhaps much more...). There are no points or lines in the physical universe, yet the mental game of geometry can be useful.