Math 6810 (Probability)

Fall 2012

Lecture notes

Pieter Allaart
University of North Texas

October 26, 2012

Text: *Introduction to Stochastic Calculus with Applications*, by Fima C. Klebaner (3rd edition), Imperial College Press.

Other recommended reading: (Do not purchase these books before consulting with your instructor!)

1. *Real Analysis* by H. L. Royden (3rd edition), Prentice Hall.

2. *Probability and Measure* by Patrick Billingsley (3rd edition), Wiley.

3. *Probability with Martingales* by David Williams, Cambridge University Press.

4. *Stochastic Calculus for Finance I and II* by Steven E. Shreve, Springer.

5. *Brownian Motion and Stochastic Calculus* by Ioannis Karatzas and Steven E. Shreve, Springer. (Warning: this requires stamina, but is one of the few texts that is complete and mathematically rigorous)

Chapter 3

Stochastic Processes in continuous time

By a stochastic process in continuous time we mean a collection $(X(t))_{t\in[0,\infty)}$ of random variables on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$, or sometimes a collection $(X(t))_{t\in[0,T]}$, where $T > 0$ is a constant. As in the discrete time case, we think of t as a "time parameter". When it is necessary to indicate the dependence on ω, we write $X(t,\omega)$.

By a *filtration* we mean here an increasing collection $(\mathcal{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathcal{F}$; that is, $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$. We also define

$$\mathcal{F}_{t+} := \bigcap_{u>t} \mathcal{F}_u.$$

Note that $\mathcal{F}_{t+}$ is a σ-algebra, and $\mathcal{F}_t \subset \mathcal{F}_{t+}$. The reverse inclusion is in general false:

Example 3.1. Let Z be standard normal random variable (or any other nondegenerate r.v.), and put

$$X(t) = \begin{cases} 0, & 0 \leq t \leq 1 \\ Z, & t > 1. \end{cases}$$

Let $\mathcal{F}_t = \sigma(X(s) : 0 \leq s \leq t)$ be the smallest σ-algebra with respect to which $X(s)$ is measurable for all $s \in [0,t]$. Then

$$\mathcal{F}_t = \begin{cases} \{\emptyset, \Omega\}, & 0 \leq t \leq 1 \\ \sigma(Z), & t > 1. \end{cases}$$

Hence, $\mathcal{F}_{1+} = \sigma(Z) \neq \{\emptyset, \Omega\} = \mathcal{F}_1$.

Definition 3.2. A filtration $(\mathcal{F}_t)_t$ is *right-continuous* if $\mathcal{F}_{t+} = \mathcal{F}_t$ for all t.

We will often assume that the filtrations that we work with are right-continuous. We'll also usually assume (for reasons explained later!) that $\mathcal{F}_0$ contains the P-null sets of $\mathcal{F}$. Together, these two requirements are referred to as *the usual conditions*.

Definition 3.3. A stochastic process $(X(t))_t$ is *adapted* to the filtration $(\mathcal{F}_t)_t$ if $X(t)$ is $\mathcal{F}_t$-measurable for each t.

Definition 3.4. A *stopping time* relative to a filtration $(\mathcal{F}_t)_t$ is a $[0,\infty]$-valued r.v. τ such that for each $t > 0$,

$$\{\tau \le t\} \in \mathcal{F}_t.$$

Exercise 3.5. Show that if τ is a stopping time, then for each $t > 0$,

$$\{\tau = t\} \in \mathcal{F}_t.$$

Definition 3.6. For a stopping time τ, define the collection

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

As in the discrete-time case, $\mathcal{F}_\tau$ is a σ-algebra and τ is $\mathcal{F}_\tau$-measurable.

Definition 3.7. A process $(X(t))_t$ is called a *submartingale* relative to the filtration $(\mathcal{F}_t)_t$ if:

(i) $(X(t))_t$ is adapted to $(\mathcal{F}_t)_t$;

(ii) $\mathrm{E}|X(t)| < \infty$ for all t; and

(iii) $\mathrm{E}[X(t)|\mathcal{F}_s] \ge X(s)$ a.s. for all $0 \le s < t$.

A process $(X(t))_t$ is a *supermartingale* if $(-X(t))_t$ is a submartingale. A process that is both a submartingale and a supermartingale is called a *martingale*.

Some stochastic processes are constructed from the ground up; others are defined implicitly by a set of conditions. For this second type of process, it is necessary to prove that a stochastic process satisfying the conditions actually exists. Kolmogorov's existence theorem is an important tool for this.

Definition 3.8. Let $X = (X(t))_t$ be a stochastic process. The *finite-dimensional distributions* of X are the probability measures

$$\mu_{t_1\ldots t_k}(B) := \mathrm{P}((X(t_1),\ldots,X(t_k)) \in B), \qquad k \in \mathbb{N}, \quad 0 \le t_1 < t_2 < \ldots t_k, \quad B \in \mathcal{B}(\mathbb{R}^k).$$

Note that the family $\{\mu_{t_1\ldots t_k}\}$ satisfies the *consistency condition*

$$\begin{aligned} &\mu_{t_1\ldots t_{i-1}t_{i+1}\ldots t_k}(B_1 \times \cdots \times B_{i-1} \times B_{i+1} \times \cdots B_k) \\ &\qquad = \mu_{t_1\ldots t_k}(B_1 \times \cdots \times B_{i-1} \times \mathbb{R} \times B_{i+1} \times \cdots B_k), \end{aligned} \tag{3.1}$$

for all $i = 1,\ldots,k$, where $B_i \in \mathcal{B}(\mathbb{R})$ for each i.

Theorem 3.9 (Kolmogorov's existence theorem). *If $\{\mu_{t_1\ldots t_k}\}$ is a family of probability measures satisfying* (3.1)*, then there exists, on some probability space* $(\Omega,\mathcal{F},\mathrm{P})$*, a stochastic process $(X(t))_t$ whose finite-dimensional distributions are $\mu_{t_1\ldots t_k}$.*

Proof. See Billingsley, Section 36. ∎

3.1 The Poisson process

In this section, let $X_1, X_2, \ldots$ be i.i.d. random variables having the exponential distribution with parameter $\lambda > 0$, and let $S_n = X_1 + \cdots + X_n$. Recall from Proposition 1.48 that S_n has a gamma(n, λ) distribution. The c.d.f. of S_n is the function

$$G_n(x) = \sum_{i=n}^{\infty} e^{-\lambda x} \frac{(\lambda x)^i}{i!}, \qquad x \geq 0. \tag{3.2}$$

This can be checked by computing $G_n'(x)$ and comparing with the gamma density.

We now think of the X_i's as waiting times between consecutive events of some kind, like arrivals of customers at a post office or tasks at a computer network. The total number of events that have happened on or before time t (where $t \geq 0$) is then

$$N(t) := \max\{n \geq 0 : S_n \leq t\}. \tag{3.3}$$

Clearly, $N(t)$ is piecewise constant, changing only in positive jumps of size 1. The jump times are the random times S_n. Furthermore, $N(t, \omega)$ is right-continuous as a function of t for each ω. We say $N(t)$ has *right-continuous sample paths.* Note that

$$\{N(t) = n\} = \{S_n \leq t < S_{n+1}).$$

By (3.2), we have

$$\mathrm{P}(N(t) \geq n) = \mathrm{P}(S_n \leq t) = G_n(t),$$

and so

$$\mathrm{P}(N(t) = n) = G_n(t) - G_{n+1}(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \qquad n = 0, 1, 2, \ldots.$$

Thus, for each t, $N(t)$ has a Poisson (λt) distribution. From this it follows that the expected number of events (say arrivals) in a time interval $(s, t]$ is proportional to $t - s$. We call $(N(t))_t$ a *Poisson process* with *intensity* λ. In fact, we can say more:

Theorem 3.10 (Independent and stationary increments of the Poisson process)**.**

(i) For $0 < t_1 < t_2 < \cdots < t_k$, the increments $N(t_1), N(t_2) - N(t_1), \ldots, N(t_k) - N(t_{k-1})$ are independent.

(ii) For all $s < t$, the increment $N(t) - N(s)$ has a Poisson$(\lambda(t - s))$ distribution.

Proof. Fix $t > 0$, and define new random variables

$$X_1^{(t)} = S_{N(t)+1} - t, \qquad X_2^{(t)} = X_{N(t)+2}, \qquad X_3^{(t)} = X_{N(t)+3}, \qquad \ldots$$

These are the waiting times after time t, because

$$S_{N(t)} \leq t < S_{N(t)+1}.$$

Verify that

$$N(t+s) - N(t) = \max\{m \geq 0 : X_1^{(t)} + \cdots + X_m^{(t)} \leq s\}. \tag{3.4}$$

We first show that

$$\mathrm{P}\left(N(t) = n, \left(X_1^{(t)}, \ldots, X_j^{(t)}\right) \in H\right) = \mathrm{P}(N(t) = n)\, \mathrm{P}((X_1, \ldots, X_j) \in H), \tag{3.5}$$

for all $H \in \mathcal{B}(\mathbb{R}^j)$. By Carathéodory's extension theorem, it is enough to show this for all H of the form

$$H = (y_1, \infty) \times (y_2, \infty) \times \cdots \times (y_j, \infty). \tag{3.6}$$

In order to do so, we calculate

$$\begin{aligned}
\mathrm{P}(S_n \leq t < S_{n+1}, S_{n+1} - t > y) &= \mathrm{P}(S_n \leq t, X_{n+1} > t + y - S_n) \\
&= \int_{x \leq t} \mathrm{P}(X_{n+1} > t + y - x)\, dG_n(x) \\
&= e^{-\lambda y} \int_{x \leq t} \mathrm{P}(X_{n+1} > t - x)\, dG_n(x) \\
&= e^{-\lambda y}\, \mathrm{P}(S_n \leq t, X_{n+1} > t - S_n),
\end{aligned}$$

where the third equality follows by the memoryless property of the exponential distribution. Using this and the independence of the X_n, we obtain for H of the form (3.6),

$$\begin{aligned}
&\mathrm{P}\left(N(t) = n, \left(X_1^{(t)}, \ldots, X_j^{(t)}\right) \in H\right) \\
&\quad = \mathrm{P}(S_{n+1} - t > y_1, X_{n+2} > y_2, \ldots, X_{n+j} > y_j, S_n \leq t < S_{n+1}) \\
&\quad = \mathrm{P}(S_{n+1} - t > y_1, S_n \leq t < S_{n+1}) e^{-\lambda y_2} \cdots e^{-\lambda y_j} \\
&\quad = \mathrm{P}(S_n \leq t < S_{n+1}) e^{-\lambda y_1} \cdots e^{-\lambda y_j} \\
&\quad = \mathrm{P}(N_t = n)\, \mathrm{P}((X_1, \ldots, X_j) \in H).
\end{aligned}$$

Thus, we have (3.5).

Now let $0 < s_1 < s_2 < \cdots < s_k$ and $0 \leq m_1 \leq m_2 \leq \cdots \leq m_k$. Then

$$\{N(s_i) = m_i, 1 \leq i \leq k\} = \{(X_1, \ldots, X_j) \in H\},$$

where $j = m_k + 1$ and

$$H = \{(x_1, \ldots, x_j) \in \mathbb{R}^j : x_1 + \cdots + x_{m_i} \leq s_i < x_1 + \cdots + x_{m_i+1}, 1 \leq i \leq k\}.$$

In the same way, by (3.4),

$$\{N(t+s_i) - N(t) = m_i, 1 \leq i \leq k\} = \left\{\left(X_1^{(t)}, \ldots, X_j^{(t)}\right) \in H\right\}.$$

Hence, (3.5) gives

$$\mathrm{P}\big(N(t) = n, N(t+s_i) - N(t) = m_i, 1 \leq i \leq k\big) = \mathrm{P}(N(t) = n)\, \mathrm{P}\left(N(s_i) = m_i, 1 \leq i \leq k\right).$$

Applying this repeatedly, we obtain for $0 = t_0 < t_1 < \cdots < t_k$,

$$\mathrm{P}\big(N(t_i) - N(t_{i-1}) = n_i, 1 \le i \le k\big) = \prod_{i=1}^{k} \mathrm{P}\left(N(t_i - t_{i-1}) = n_i\right).$$

From this, the assertions of the theorem follow. □

Exercise 3.11. Show that the following are martingales:

(i) $N(t) - \lambda t$

(ii) $(N(t) - \lambda t)^2 - \lambda t$

3.2 Brownian motion

Botanist Robert Brown described the highly irregular motion of a pollen particle suspended in liquid in 1828. Albert Einstein gave a physical explanation for this motion and derived mathematical equations for it in 1905. The process that we now call *Brownian motion* was formulated rigorously (from a mathematical point of view) by Norbert Wiener in the 1920s, and because of his work, the process is also frequently called the *Wiener process*. In the 1940s, Paul Lévy analyzed Brownian motion more deeply and introduced the notion of local time, important for the theory of stochastic calculus. Around the same time, Kiyoshi Itô laid the groundwork for this new kind of calculus, publishing what is now known as Itô's rule, which replaces the chain rule from ordinary differential calculus. Brownian motion is now used in many areas, including physics, engineering and mathematical finance.

Definition 3.12. A (one-dimensional) *Brownian motion* is a stochastic process $\{W(t) : t \ge 0\} = \{W(t, \omega) : t \ge 0\}$ on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with the following properties:

(i) $W(0) \equiv 0$.

(ii) (Independence of increments) For $0 \le t_1 < t_2 < \cdots < t_n < \infty$, the increments $W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_n) - W(t_{n-1})$ are independent.

(iii) (Stationarity and normality of increments) For $0 < s < t$, the increment $W(t) - W(s)$ has a normal distribution with mean 0 and variance $t - s$.

(iv) (Continuity of sample paths) For each $\omega \in \Omega$, the function $t \mapsto W(t, \omega)$ is continuous.

The first question that needs to be addressed is whether such a process actually exists. It is not difficult to check that the stipulations (ii) and (iii) satisfy the consistency condition (3.1), so Kolmogorov's existence theorem implies the existence of a process $\{W(t)\}$ satisfying (i)-(iii). However, there is no guarantee that this process will have continuous sample paths. One way around this is to begin with a process $\{W(t)\}$ as given by Kolmogorov's theorem, and prove that the restriction of this process to dyadic rational time points is with probability one uniformly continuous on compact intervals. One can then redefine $W(t)$ at nondyadic t by taking limits over dyadic rationals approaching t, which will guarantee continuity of $W(t)$. This approach, which can be found in Billingsley, section 37, is rather technical. Instead, we will construct Brownian motion from the ground up. But first, some preliminaries.

Definition 3.13. A process $\{X(t) : t \geq 0\}$ is *Gaussian* if every finite linear combination $a_1X(t_1) + \cdots + a_nX(t_n)$ has a normal distribution, where $a_i \in \mathbb{R}$ and $t_i \geq 0$ for all i.

Lemma 3.14. *Brownian motion is a Gaussian process. It has mean and covariance functions given by*

$$\mu(t) := \mathrm{E}(W(t)) = 0,$$

and

$$r(s,t) := \mathrm{Cov}(W(s), W(t)) = \min\{s,t\}.$$

Proof. Let $0 \leq t_1 < t_2 < \cdots < t_n$ and $a_1, \ldots, a_n \in \mathbb{R}$. Put $t_0 = 0$. Note that for each i, we can write $W(t_i)$ as a linear combination of the increments $W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$. But then $a_1W(t_1) + \cdots + a_nW(t_n)$ is also a linear combination of these increments. Hence, $\{W(t)\}$ is Gaussian.

That the mean of $W(t)$ is zero is obvious. To compute the covariance, assume WLOG that $s < t$. Then $W(s)$ and $W(t) - W(s)$ are independent and have mean zero, so

$$\begin{aligned} \mathrm{E}[W(s)W(t)] &= \mathrm{E}\left[W(s)(W(t) - W(s))\right] + \mathrm{E}\left[(W(s))^2\right] \\ &= \mathrm{E}(W(s))\,\mathrm{E}(W(t) - W(s)) + \mathrm{Var}(W(s)) = s. \end{aligned}$$

Hence,

$$\mathrm{Cov}(W(s), W(t)) = \mathrm{E}(W(s)W(t)) - \mathrm{E}(W(s))\,\mathrm{E}(W(t)) = s,$$

as required. □

It can be shown that the finite-dimensional distributions of a Gaussian process are completely determined by its mean and covariance functions. Thus, if we can construct a Gaussian process with continuous sample paths and with the correct mean and covariance functions, then this process must be Brownian motion.

3.2.1 Construction of Brownian motion

To begin, note that it is sufficient to construct Brownian motion on $0 \leq t \leq 1$: we can then construct infinitely many independent copies of this Brownian motion and paste them together to obtain Brownian motion on $[0, \infty)$. Precisely, let $W_j(t)$, $j \in \mathbb{N}$ be independent Brownian motions on $[0, 1]$ and put

$$W(t) = \begin{cases} W_1(t), & 0 \leq t < 1 \\ \sum_{j=1}^{n-1} W_j(1) + W_n(t-n), & n \leq t < n+1, \quad n \in \mathbb{N}. \end{cases}$$

Then one checks easily that $W(t)$ is a Brownian motion on $[0,\infty)$.

Step 1. Define the *Haar functions*

$$H_1(t) = 1, \qquad 0 \le t \le 1$$
$$H_{2^n+1}(t) = \begin{cases} 2^{n/2}, & 0 \le t < 2^{-(n+1)} \\ -2^{n/2}, & 2^{-(n+1)} \le t \le 2^{-n} \\ 0, & \text{elsewhere} \end{cases} \qquad (n = 0,1,2,\dots)$$
$$H_{2^n+j}(t) = H_{2^n+1}\left(t - \frac{j-1}{2^n}\right), \qquad j = 1,\dots,2^n, \qquad n \ge 0.$$

Define the *Schauder functions* by

$$S_k(t) = \int_0^t H_k(u)\,du, \qquad k \in \mathbb{N}, \quad 0 \le t \le 1.$$

Then $S_1(t) = t$, and for $n \ge 0$ and $1 \le j \le 2^n$, the graph of S_{2^n+j} is a "tent" of height $2^{-(n+2)/2}$ over the interval $[(j-1)/2^n, j/2^n]$. Note that

$$S_{2^n+j}(t)S_{2^n+k}(t) = 0 \qquad \forall\, t \qquad \text{if } 1 \le k < j \le 2^n. \tag{3.7}$$

Step 2. Let $Z_k, k \in \mathbb{N}$ be independent standard normal r.v.'s. Put

$$W^{(n)}(t) = \sum_{k=1}^{2^n} Z_k S_k(t), \qquad 0 \le t \le 1, \quad n \ge 0.$$

Lemma 3.15. *As $n \to \infty$, $W^{(n)}$ converges uniformly on $[0,1]$ to a continuous function $W(t)$ with probability one.*

Proof. Let

$$M_n := \max_{1 \le j \le 2^{n-1}} |Z_{2^{n-1}+j}|, \qquad n \in \mathbb{N}.$$

For $x > 0$ and $k \in \mathbb{N}$, we have

$$\begin{aligned} \mathrm{P}(|Z_k| > x) = 2\,\mathrm{P}(Z_k > x) &= \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-u^2/2}\,du \\ &\le \sqrt{\frac{2}{\pi}} \int_x^\infty \frac{u}{x} e^{-u^2/2}\,du = \sqrt{\frac{2}{\pi}}\,\frac{e^{-x^2/2}}{x}. \end{aligned}$$

Thus, for $n \in \mathbb{N}$,

$$\begin{aligned}
\mathrm{P}(M_n > n) &= \mathrm{P}\left(\bigcup_{1 \le j \le 2^{n-1}} \{|Z_{2^{n-1}+j}| > n\}\right) \\
&\le \sum_{j=1}^{2^{n-1}} \mathrm{P}\left(|Z_{2^{n-1}+j}| > n\right) \\
&= 2^n\,\mathrm{P}(|Z_1| > n) \le \sqrt{\frac{2}{\pi}} \cdot \frac{2^n e^{-n^2/2}}{n}.
\end{aligned}$$

Since

$$\sum_{n=1}^{\infty} \frac{2^n e^{-n^2/2}}{n} < \infty$$

(check!!), the first Borel-Cantelli lemma implies that

$$\mathrm{P}(M_n > n \text{ for infinitely many } n) = 0.$$

Hence, with probability one, there is an index N such that $M_n \le n$ whenever $n \ge N$. But then,

$$\sum_{k=2^N+1}^{\infty} |Z_k S_k(t)| \le \sum_{n=N}^{\infty} M_{n+1} 2^{-(n+2)/2} \le \sum_{n=N}^{\infty} (n+1) 2^{-(n+2)/2} < \infty$$

for all $0 \le t \le 1$. Thus, by the Cauchy criterion, $W^{(n)}(t)$ converges uniformly on $[0,1]$ to a function $W(t)$, which is continuous as the uniform limit of continous functions on $[0,1]$. □

Step 3. Note that $W^{(n)}$ is a mean-zero Gaussian process for each n. It can be shown (for instance using the method of characteristic functions) that the almost-sure limit of a sequence of Gaussian processes is Gaussian. Hence W is Gaussian, and it remains to check that it has the correct mean and covariance functions.

Exercise 3.16. Show that

$$\sum_{k=1}^{\infty} S_k(t) < \infty \qquad \forall\, t \tag{3.8}$$

and

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} S_j(s) S_k(t) < \infty \qquad \forall\, s, t. \tag{3.9}$$

By (3.8) and Fubini's theorem,

$$\mathrm{E}\left(\sum_{k=1}^{\infty} |Z_k S_k(t)|\right) = \mathrm{E}\,|Z_1| \sum_{k=1}^{\infty} S_k(t) < \infty,$$

and so

$$\mathrm{E}(W(t)) = E\left(\sum_{k=1}^{\infty} Z_k S_k(t)\right) = \sum_{k=1}^{\infty} \mathrm{E}(Z_k) S_k(t) = 0.$$

To compute the covariance function, we need a little Fourier analysis. Verify that the functions H_k, $k \in \mathbb{N}$ form a complete orthonormal system; that is,

$$\int_0^1 H_j(t) H_k(t)\, dt = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Therefore, Parseval's identity implies that for any two bounded functions f, g on $[0,1]$,

$$\int_0^1 f(u) g(u)\, du = \sum_{k=1}^{\infty} a_k b_k,$$

where

$$a_k = \int_0^1 f(t) H_k(t)\, dt, \qquad b_k = \int_0^1 g(t) H_k(t)\, dt.$$

Apply this to $f = \chi_{[0,s]}$ and $g = \chi_{[0,t]}$. Then $a_k = S_k(s)$ and $b_k = S_k(t)$. Now

$$\begin{aligned}
\mathrm{E}\left(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} |Z_j Z_k S_j(s) S_k(t)|\right) &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \mathrm{E}\,|Z_j Z_k| S_j(s) S_k(t) \\
&\leq \max\{\mathrm{E}\,|Z_1^2|, (\mathrm{E}\,|Z_1|)^2\} \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} S_j(S) S_k(t) \\
&< \infty
\end{aligned}$$

by (3.9) and Fubini's theorem. Hence,

$$\begin{aligned}
r(s,t) = \mathrm{E}[W(s)W(t)] &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \mathrm{E}(Z_j Z_k) S_j(s) S_k(t) \\
&= \sum_{j=1}^{\infty} \mathrm{E}(Z_j^2) S_j(s) S_j(t) \qquad \text{(by independence of the } Z_j\text{)} \\
&= \sum_{j=1}^{\infty} S_j(s) S_j(t) \qquad \text{(since } \mathrm{E}(Z_j^2) = 1\text{)} \\
&= \int_0^1 \chi_{[0,s]}(u) \chi_{[0,t]}(u)\, du \qquad \text{(by Parseval's identity)} \\
&= \min\{s,t\}.
\end{aligned}$$

This shows that $\{W(t), 0 \leq t \leq 1\}$ is a Brownian motion.

3.2.2 Various properties of Brownian motion

Definition 3.17. A process $(X(t))_{t\in[0,\infty)}$ is *measurable* if it is measurable viewed as a function $X : [0,\infty)\times\Omega\to\mathbb{R}$. That is, if

$$\{(t,\omega) : X(t,\omega)\in B\}\in\mathcal{B}([0,\infty))\times\mathcal{F}, \qquad \forall\, B\in\mathcal{B}(\mathbb{R}).$$

Proposition 3.18. *Brownian motion is measurable.*

Proof. Define the piecewise constant approximands

$$W_n(t,\omega) = W(k2^{-n},\omega) \qquad \text{for} \qquad k2^{-n}\le t<(k+1)2^{-n}, \quad k=0,1,2,\ldots.$$

It is easy to see that the mapping $(t,\omega)\mapsto W_n(t,\omega)$ is $\mathcal{B}([0,\infty))\times\mathcal{F}$-measurable for each n. By continuity of sample paths, $W_n(t,\omega)\to W(t,\omega)$ for each (t,ω). Hence, the mapping $(t,\omega)\mapsto W(t,\omega)$ is $\mathcal{B}([0,\infty))\times\mathcal{F}$-measurable. □

This last result is useful because random integrals of the form $\int_a^b \phi(W(t))\,dt$, where ϕ is a Borel function, are now well defined. Provided the interchange of expectation and integral can be justified, we have

$$\mathrm{E}\left(\int_a^b \phi(W(t))\,dt\right) = \int_a^b \mathrm{E}\left(\phi(W(t))\right)dt.$$

Exercise 3.19. Compute

$$\mathrm{E}\left(\int_0^T W(t)\,dt\right) \qquad \text{and} \qquad \mathrm{E}\left(\int_0^T W^2(t)\,dt\right)$$

Exercise 3.20 (Brownian motion martingales)**.** Show that the following are martingales:

(i) $W(t)$

(ii) $W(t)^2-t$

(iii) $e^{\lambda W(t)-\lambda^2 t/2}$, where $\lambda\in\mathbb{R}$.

Proposition 3.21. *On a set of probability one,*

$$\lim_{t\to\infty}\frac{W(t)}{t}=0. \tag{3.10}$$

Proof. Writing $W(n)=W(0)+[W(1)-W(0)]+\cdots+[W(n)-W(n-1)]$, we see by the SLLN that

$$\lim_{n\to\infty}\frac{W(n)}{n}=0 \qquad \text{a.s.}$$

It is now not hard to believe that (3.10) should hold, because for arbitrary $t>0$, we can write

$$\frac{W(t)}{t}=\frac{[t]}{t}\cdot\left(\frac{W(t)-W([t])}{[t]}+\frac{W([t])}{[t]}\right),$$

where $[t]$ denotes the greatest integer in t. Intuitively, the first term in parentheses should approach zero as $t \to \infty$. But proving this requires some care; see Karatzas and Shreve, Problem 2.9.3. $\square$

Exercise 3.22. Show that each of the following processes is a Brownian motion:

(i) (Reflection) $W_1(t) = -W(t)$

(ii) (Time scaling) $W_2(t) = cW(t/c^2)$, for fixed $c > 0$

(iii) (Time shift) $W_3(t) = W(t_0 + t) - W(t_0)$, for fixed $t_0 > 0$

(iv) (Time inversion)

$$W_4(t) = \begin{cases} tW(1/t), & t > 0 \\ 0, & t = 0 \end{cases}$$

(Hint: check that each process is a mean zero Gaussian process with continuous sample paths and the correct covariance function. For W_4, continuity at $t = 0$ follows from (3.10) and the substitution $u = 1/t$.)

Proposition 3.21 has the following strengthening, which specifies the maximum growth rate of Brownian paths. Its proof is beyond the scope of this course.

Theorem 3.23 (Law of the iterated logarithm)**.** *We have*

$$\limsup_{t\to\infty} \frac{W(t)}{\sqrt{2t \log\log t}} = 1 \qquad a.s.$$

and

$$\liminf_{t\to\infty} \frac{W(t)}{\sqrt{2t \log\log t}} = -1 \qquad a.s.$$

(Check that these statements together imply (3.10)!)

Exercise 3.24. Use Theorem 3.23 to show that

$$\limsup_{t\downarrow 0} \frac{W(t)}{\sqrt{2t \log\log(1/t)}} = 1 \qquad \text{a.s.}$$

and the corresponding $\liminf$ equals -1, almost surely.

The last exercise has the remarkable consequence that, on any time interval $[0, \varepsilon]$, $W(t)$ changes sign infinitely many times with probability one. Since $W(t)$ is continuous in t, this means that $W(t) = 0$ for infinitely many t in $[0, \varepsilon]$ with probability one. In fact, more can be proved:

Theorem 3.25 (Zero set of Brownian motion)**.** *The set $\{t \geq 0 : W(t) = 0\}$ is with probability one an uncountable closed set without isolated points and of Lebesgue measure zero.*

We will not use this theorem. For a proof, see Klebaner, Theorem 3.28.

Theorem 3.26 (Nowhere differentiability of sample paths)**.** *For all ω outside a set of probability* 0*,* $W(\cdot, \omega)$ *is nowhere differentiable.*

Note that we avoid saying something like

$$\mathrm{P}(\omega : W(\cdot, \omega) \text{ is nowhere differentiable}) = 1.$$

The reason is, that it is not at all clear whether the set in question is measurable. Proving the theorem entails finding a measurable subset of this set which has probability 1.

Proof. (We follow Billingsley, Theorem 37.3.) Put

$$\Delta_{n,k} = W\left(\frac{k+1}{2^n}\right) - W\left(\frac{k}{2^n}\right),$$

and let

$$X_{n,k} = \max\{|\Delta_{n,k}|, |\Delta_{n,k+1}|, |\Delta_{n,k+2}|\}.$$

Note that each $\Delta_{n,k}$ has the same distribution as $2^{-n/2}W(1)$, namely Normal$(0, 2^{-n})$. Furthermore, for fixed n, the $\Delta_{n,k}$ are independent. Thus, given $\varepsilon > 0$,

$$\mathrm{P}(X_{n,k} \le \varepsilon) = \left[\mathrm{P}(|W(1)| \le 2^{n/2}\varepsilon)\right]^3.$$

Now for $\alpha > 0$,

$$\mathrm{P}(|W(1)| \le \alpha) = \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \le 2\alpha \frac{1}{\sqrt{2\pi}} < \alpha.$$

Hence,

$$\mathrm{P}(X_{n,k} \le \varepsilon) \le (2^{n/2}\varepsilon)^3.$$

Put

$$Y_n = \min_{k \le n2^n} X_{n,k}.$$

Then

$$\mathrm{P}(Y_n \le \varepsilon) \le n2^n(2^{n/2}\varepsilon)^3 = n2^{5n/2}\varepsilon^3. \tag{3.11}$$

We consider the upper and lower right-hand derivatives

$$D^+(t, \omega) = \limsup_{h \downarrow 0} \frac{W(t+h, \omega) - W(t, \omega)}{h},$$
$$D_+(t, \omega) = \liminf_{h \downarrow 0} \frac{W(t+h, \omega) - W(t, \omega)}{h}.$$

Define the set

$$E = \{\omega : \text{ there is } t \ge 0 \text{ such that } D^+(t, \omega) \text{ and } D_+(t, \omega) \text{ are both finite}\}.$$

Suppose $\omega \in E$; then we can find $t \geq 0$ and $K > 0$ such that

$$-K < D_+(t,\omega) \leq D^+(t,\omega) < K.$$

This implies that there is $\delta > 0$ (depending on t, K and ω) such that

$$t \leq s \leq t+\delta \qquad \Rightarrow \qquad |W(s,\omega) - W(t,\omega)| \leq K|s-t|. \tag{3.12}$$

Choose n large enough so that $2^{-n} < \delta/4$, $8K < n$, and $n > t$. Choose k such that $(k-1)/2^n \leq t < k/2^n$. Then $|t - i/2^n| < \delta$ for $i = k, k+1, k+2, k+3$, and hence, by (3.12) and the triangle inequality,

$$X_{n,k}(\omega) \leq 2K(4/2^n) < n2^{-n}.$$

Since also $k-1 \leq t2^n < n2^n$, it follows that $Y_n(\omega) \leq n2^{-n}$.

Define the set $A_n = \{Y_n \leq n2^{-n}\}$. The above argument shows that $E \subset \liminf A_n$. Note that each A_n is measurable, and so $\liminf A_n$ is measurable. By (3.11),

$$\mathrm{P}(A_n) \leq n2^{5n/2}(n2^{-n})^3 = n^4 2^{-n/2} \to 0.$$

It follows (check!) that $\mathrm{P}(\liminf A_n) = 0$. And outside the set $\liminf A_n$, $W(\cdot,\omega)$ is nowhere differentiable (in fact, it does not have finite upper and lower right-hand derivatives anywhere). □

3.2.3 Quadratic variation

Definition 3.27. The *quadratic variation* of a function $f : [0,\infty) \to \mathbb{R}$ is defined by

$$[f,f](t) = \lim_{\delta_n \to 0} \sum_{i=1}^{n} \left(f(t_i^n) - f(t_{i-1}^n) \right)^2$$

(provided the limit exists), where $\{0 = t_0^n < t_1^n < \cdots < t_{n-1}^n < t_n^n = t\}$ is a partition of $[0,t]$, and $\delta_n = \max_{1\leq i\leq n}(t_i^n - t_{i-1}^n)$. (Compare the definition of Riemann integral!)

Proposition 3.28. *If f is continuous and of bounded variation in $[0,t]$, then $[f,f](t) = 0$.*

Proof. See Klebaner, Theorem 1.10. □

Because of the last proposition, quadratic variation is not a useful concept in ordinary differential calculus. But it is of critical importance in stochastic calculus, because the processes encountered there are typically not of bounded variation. This includes Brownian motion.

Theorem 3.29 (Quadratic variation of Brownian motion)**.** *Brownian motion accumulates quadratic variation at unit rate. That is,*

$$[W,W](t) = t.$$

More precisely, for each fixed $t > 0$*,*

$$\sum_{i=1}^{n} \big(W(t_i^n) - W(t_{i-1}^n)\big)^2 \to t \qquad \textit{in } L^2,$$

and under the extra condition $\sum_{n=1}^{\infty} \delta_n < \infty$*, the convergence above also takes place almost surely.*

Proof. See Klebaner, Theorem 3.4 for a proof of the almost-sure statement. The L^2 statement can be gleaned from the proof without difficulty (exercise!). □

Observe the surprising fact that, although the paths of Brownian motion are random, its quadratic variation (which is a path property!) is deterministic. We will later see that Brownian motion is the unique mean-zero martingale whose quadratic variation on $[0, t]$ is t (Lévy's characterization), and this fact offers a practical way to recognize Brownian motions "in disguise".

3.2.4 The strong Markov property of Brownian motion

In this subsection, let $(\mathcal{F}_t)_t$ denote the natural filtration of $(W(t))_t$. That is, $\mathcal{F}_t = \sigma(W(s) : 0 \leq s \leq t)$. Fix $t_0 \geq 0$, and put

$$W'(t) = W(t_0 + t) - W(t_0), \qquad t \geq 0.$$

We have seen before that $W'(t)$ is a Brownian motion, and in view of the independent increments of $W(t)$, it is independent of $\mathcal{F}_t$. In particular, we have (check!)

$$\mathrm{P}(W(t_0 + t) \leq y | \mathcal{F}_{t_0}) = \mathrm{P}(W(t_0 + t) \leq y | W(t_0)) \qquad \text{a.s.} \tag{3.13}$$

We say that Brownian motion possesses the *Markov property.* More fully, we have, for $0 \leq t_1 < t_2 < \cdots < t_k$ and $H \in \mathcal{B}(\mathbb{R}^k)$,

$$\begin{aligned} \mathrm{P}\big((W'(t_1), \ldots, W'(t_k)) \in H \cap A\big) &= \mathrm{P}\left((W'(t_1), \ldots, W'(t_k)) \in H\right) \mathrm{P}(A) \\ &= \mathrm{P}\left((W(t_1), \ldots, W(t_k)) \in H\right) \mathrm{P}(A), \qquad A \in \mathcal{F}_{t_0}. \end{aligned} \tag{3.14}$$

We now want to show that the above identities remain true when t_0 is replaced by a stopping time τ. A process $X = (X(t))_t$ satisfying

$$\mathrm{P}(X(\tau + t) \leq y | \mathcal{F}_\tau) = \mathrm{P}(X(\tau + t) \leq y | X(\tau)) \qquad \text{a.s.} \tag{3.15}$$

for every finite stopping time τ is said to possess the *strong Markov property.* Note that the strong Markov property implies the Markov property, because any constant $t_0 > 0$ is a stopping time.

Some care is needed here; for instance, it is not a priory clear that $X(\tau + t)$ is even a random variable (i.e. is $\mathcal{F}$-measurable). We also must be careful with the use of conditional probabilities. We ignore these subtleties here; the details can be found for instance in Karatzas and Shreve, Chapter 2.

Theorem 3.30 (Strong Markov property of Brownian motion)**.** *Let τ be a finite stopping time, and put*

$$W^*(t,\omega) := W(\tau(\omega)+t,\omega) - W(\tau(\omega),\omega), \qquad t \geq 0.$$

Then $\{W^*(t) : t \geq 0\}$ is a Brownian motion independent of $\mathcal{F}_\tau$. That is, for $0 \leq t_1 < t_2 < \cdots < t_k$, $H \in \mathcal{B}(\mathbb{R}^k)$ and $A \in \mathcal{F}_\tau$,

$$\begin{aligned} \mathrm{P}\big(\{(W^*(t_1),\dots,W^*(t_k)) \in H\} \cap A\big) &= \mathrm{P}\left((W^*(t_1),\dots,W^*(t_k)) \in H\right)\mathrm{P}(A) \\ &= \mathrm{P}\left((W(t_1),\dots,W(t_k)) \in H\right)\mathrm{P}(A). \end{aligned} \tag{3.16}$$

The proof of the theorem follows Billingsley, section 37, and uses the following lemma.

Lemma 3.31. *Let $X_n = (X_n^1,\dots,X_n^k)$ and $X = (X^1,\dots,X^k)$ be random vectors in $\mathbb{R}^k$, and let $F_n(x)$ be the distribution function of X_n; that is,*

$$F_n(x) = \mathrm{P}(X_n^1 \leq x_1,\dots,X_n^k \leq x_k), \qquad x = (x_1,\dots,x_k) \in \mathbb{R}^k.$$

If $X_n \to X$ a.s. and $F_n(x) \to F(x)$ for all $x \in \mathbb{R}^k$, then $F(x)$ is the distribution function of X.

Proof. Let H denote the distribution function of X. Then for all $h > 0$,

$$\begin{aligned} F(x_1,\dots,x_k) &= \limsup_{n\to\infty} F_n(x_1,\dots,x_k) && \text{(by hypothesis)} \\ &= \limsup_{n\to\infty} \mathrm{P}(X_n^1 \leq x_1,\dots,X_n^k \leq x_k) \\ &\leq \mathrm{P}(\limsup\{X_n^1 \leq x_1,\dots,X_n^k \leq x_k\}) \\ &\leq \mathrm{P}(X^1 \leq x_1,\dots,X^k \leq x_k) && \text{(since } X_n \to X \text{ a.s.)} \\ &= H(x_1,\dots,x_k) \\ &\leq \mathrm{P}(\liminf\{X_n^1 \leq x_1+h,\dots,X_n^k \leq x_k+h\}) && \text{(again since } X_n \to X \text{ a.s.)} \\ &\leq \liminf_{n\to\infty} \mathrm{P}(X_n^1 \leq x_1+h,\dots,X_n^k \leq x_k+h) \\ &= F(x_1+h,\dots,x_k+h). \end{aligned}$$

But $F(x_1+h,\dots,x_k+h) \to F(x_1,\dots,x_k)$ as $h \downarrow 0$, so it follows that $F = H$. $\square$

Exercise 3.32. We have used in the above proof that

$$\mathrm{P}(\liminf A_n) \leq \liminf \mathrm{P}(A_n) \qquad \text{and} \qquad \limsup \mathrm{P}(A_n) \leq \mathrm{P}(\limsup A_n).$$

Check these inequalities!

Proof of Theorem 3.30. Assume first that τ has a countable range V. Because

$$\{\omega : W^*(t,\omega) \in H\} = \bigcup_{t_0 \in V} \{\omega : W(t_0+t,\omega) - W(t_0,\omega) \in H, \tau(\omega) = t_0\},$$

it follows that $W^*(t)$ is a random variable. Furthermore,

$$\mathrm{P}\big(\{(W^*(t_1),\dots,W^*(t_k)) \in H\} \cap A\big) = \sum_{t_0 \in V} \mathrm{P}\left(\{(W^*(t_1),\dots,W^*(t_k)) \in H\} \cap A \cap \{\tau = t_0\}\right).$$

Now $A \in \mathcal{F}_\tau$ implies $A \cap \{\tau = t_0\} \in \mathcal{F}_{t_0}$ (why?), and if $\tau = t_0$ then $W^*(t) = W'(t)$. Thus, (3.14) reduces the last sum above to

$$\sum_{t_0 \in V} \mathrm{P}\left((W(t_1),\dots,W(t_k)) \in H\right) \mathrm{P}(A \cap \{\tau = t_0\}) = \mathrm{P}\left((W(t_1),\dots,W(t_k)) \in H\right) \mathrm{P}(A).$$

This shows that the first and third terms in (3.16) are equal. That the second term is equal to the third follows by taking $A = \Omega$. Thus, the theorem holds when τ has countable range.

For general τ, we approximate τ as follows. Let

$$\tau_n = \begin{cases} k2^{-n}, & \text{if } (k-1)2^{-n} < \tau \le k2^{-n}, k \in \mathbb{N} \\ 0, & \text{if } \tau = 0. \end{cases}$$

Check that each τ_n is a stopping time and $\tau \le \tau_n$, so that $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$. Define

$$W_n(t,\omega) = W(\tau_n(\omega) + t, \omega) - W(\tau_n(\omega), \omega), \qquad t \ge 0.$$

Since τ_n has countable range, we can apply (3.16) to obtain, for all $A \in \mathcal{F}_\tau$ (!),

$$\mathrm{P}\left(\{(W_n(t_1),\dots,W_n(t_k)) \in H\} \cap A\right) = \mathrm{P}((W(t_1),\dots,W(t_k)) \in H)\,\mathrm{P}(A). \tag{3.17}$$

Now $\tau_n(\omega) \to \tau(\omega)$ for all ω, and by continuity of sample paths, $W_n(t,\omega) \to W^*(t,\omega)$ for each ω. Hence,

$$(W_n(t_1),\dots,W_n(t_k)) \to (W^*(t_1),\dots,W^*(t_k)) \qquad \text{a.s.}$$

Apply Lemma 3.31 with $F_n = F =$ the conditional distribution function of $(W_n(t_1),\dots,W_n(t_k))$ given A. Then (3.16) follows from (3.17). □

3.2.5 The reflection principle, hitting times and the maximum of BM

The strong Markov property of Brownian motion says roughly speaking that at a finite stopping τ, Brownian motion "starts afresh" from the (random) level $W(\tau)$. One consequence of this is the so-called reflection principle. Let τ be a finite stopping time, and define a new process

$$W'(t) = \begin{cases} W(t), & t \le \tau \\ W(\tau) - [W(t) - W(\tau)], & t \ge \tau. \end{cases} \tag{3.18}$$

Thus, the sample path of $W'(t)$ is the same as the sample path of $W(t)$ up to time τ, and after that, it is the reflection of this path in the line $y = W(\tau)$.

Theorem 3.33 (Reflection principle)**.** *The process $\{W'(t) : t \geq 0\}$ is a Brownian motion.*

This theorem is intuitively obvious from the strong Markov property: $W'(t)$ is a Brownian motion before time τ, and it is a reflection of a Brownian motion, hence a Brownian motion, after time τ, so it ought to be a Brownian motion "everywhere". However, since τ is random, some care is needed to check that the finite-dimensional distributions of $W'(t)$ are really the same as those of $W(t)$. (Continuity of sample paths is obvious from the definition of $W'(t)$.) See Billingsley, p. 512 for a precise proof.

We now define the *hitting time*

$$\tau_a := \inf\{t > 0 : W(t) \geq a\}, \qquad a > 0$$

and the maximum of Brownian motion up to time t,

$$M(t) := \max\{W(s) : 0 \leq s \leq t\}.$$

(By continuity of sample paths, this maximum is well defined.) Also by continuity of paths, $W(\tau_a) = a$ on the event $\{\tau_a < \infty\}$.

Proposition 3.34. *The stopping time τ_a is finite with probability one:* $\mathrm{P}(\tau_a < \infty) = 1$.

Proof. See Klebaner, Theorem 3.14 for a proof based on the strong Markov property; or Example 7.8 for an alternative proof using the exponential martingale $e^{\sigma W(t) - \sigma^2 t/2}$. □

As we did for random walks, we will use the reflection principle to derive the distributions of τ_a and $M(t)$, as well as the joint distribution of $M(t)$ and $W(t)$. First, define $W'(t)$ by (3.18) with $\tau = \tau_a$, and let $\tau_a' = \inf\{t > 0 : W'(t) \geq a\}$. Of course $\tau' = \tau$, and for any $t > 0$, the pair $(\tau, W(t))$ has the same joint distribution as the pair $(\tau', W'(t))$ by the reflection principle. Hence, for $x \leq a$,

$$\begin{aligned}
\mathrm{P}(\tau_a \leq t, W(t) \leq x) &= \mathrm{P}(\tau_a' \leq t, W'(t) \leq x) \\
&= \mathrm{P}(\tau_a \leq t, W(\tau_a) - [W(t) - W(\tau_a)] \leq x) \\
&= \mathrm{P}(\tau_a \leq t, 2a - W(t) \leq x) \\
&= \mathrm{P}(\tau_a \leq t, W(t) \geq 2a - x) \\
&= \mathrm{P}(W(t) \geq 2a - x).
\end{aligned}$$

The second equality follows since $W'(t) = W(\tau_a) - [W(t) - W(\tau_a)]$ when $\tau_a \leq t$, and the third since $W(\tau_a) = a$. The last equality follows since $2a - x \geq a$, so if $W(t) \geq 2a - x$, then certainly $\tau_a \leq t$. In summary,

$$\mathrm{P}(\tau_a \leq t, W(t) \leq x) = \mathrm{P}(W(t) \geq 2a - x), \qquad a > 0, \quad x \leq a. \tag{3.19}$$

Setting $x = a$ in this equation, we obtain

$$\begin{aligned}
\mathrm{P}(\tau_a \leq t) &= \mathrm{P}(\tau_a \leq t, W(t) \leq a) + \mathrm{P}(\tau_a \leq t, W(t) \geq a) \\
&= 2\,\mathrm{P}(W(t) \geq a) = \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-u^2/2t}\, du,
\end{aligned}$$

since $\{\tau_a \leq t, W(t) \geq a\} = \{W(t) \geq a\}$ and $W(t) \sim \text{Normal}(0, t)$. Differentiating, we find (eventually - integrate by parts!) that τ_a has density function

$$f_{\tau_a}(t) = \frac{d}{dt}\,\mathrm{P}(\tau_a \leq t) = \frac{a}{t^{3/2}\sqrt{2\pi}} e^{-a^2/2t}, \qquad t \geq 0.$$

(How does this change when $a < 0$?)

The distribution of $M(t)$ for fixed t is derived similarly, by noting that

$$\mathrm{P}(M(t) \geq a) = \mathrm{P}(\tau_a \leq t),$$

so that

$$f_{M(t)}(a) = -\frac{d}{da}\,\mathrm{P}(M(t) \geq a) = \frac{2}{\sqrt{2\pi t}} e^{-a^2/2t}, \qquad a \geq 0.$$

Exercise 3.35. Show that $M(t)$ and $W(t)$ have the joint density function

$$f_{M(t),W(t)}(a, x) = \frac{2(2a - x)}{t^{3/2}\sqrt{2\pi}} e^{-(2a-x)^2/2t}, \qquad a > 0, \quad x \leq a.$$

(*Hint*: rewrite (3.19) as

$$\mathrm{P}(M(t) \geq a, W(t) \leq x) = \mathrm{P}(W(t) \geq 2a - x).$$

Differentiate both sides of this equation, first to a and then to x.)

3.2.6 The invariance principle and Wiener measure

Recall the central limit theorem: If $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with mean zero and variance σ^2, and if $S_n = \xi_1 + \cdots + \xi_n$, then

$$\frac{S_n}{\sigma\sqrt{n}} \Rightarrow N,$$

where $N \sim \text{Normal}(0, 1)$. Note that this is also the distribution of $W(1)$, where $W(t)$ is a Brownian motion. Since the process $\{S_n\}$ and Brownian motion both have independent and stationary increments, it should not be a surprise that, in a sense to be made precise below, the *entire process* $\{S_n\}$, suitably scaled, converges to a Brownian motion. We begin with a definition.

Definition 3.36. Let (S, ρ) be a metric space, let $\{\mathrm{P}_n\}$ be a sequence of Borel probability measures on S, and let P be a Borel measure on S. We say P_n *converges weakly* to P if

$$\int_S f(s)\, d\,\mathrm{P}_n(s) \to \int_S f(s)\, d\,\mathrm{P}(s)$$

for every bounded, continuous function $f : S \to \mathbb{R}$.

(If $S = \mathbb{R}$, this corresponds to weak convergence of random variables; compare Theorem **??**.) It follows from the definition that if P_n converges weakly to P, then P is a probability measure. (Take $f \equiv 1$.)

Next, let $\xi_1, \xi_2, \dots$ be i.i.d. r.v.'s on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with mean zero and variance σ^2, and put $S_0 \equiv 0$, and $S_n = \xi_1 + \cdots + \xi_n$ for $n \in \mathbb{N}$. We construct a continuous-time process $\{Y(t) : t \geq 0\}$ by linear interpolation of the sequence $\{S_n\}$: $Y(n) = S_n$, and for $0 \leq \lambda \leq 1$,

$$Y\big(\lambda n + (1-\lambda)(n+1)\big) = \lambda S_n + (1-\lambda) S_{n+1}, \qquad n = 0, 1, 2, \dots.$$

Finally, define

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} Y(nt), \qquad n \in \mathbb{N}, \quad t \geq 0.$$

Thus, X_n is a linear interpolation of $\{S_n\}$, scaled in both time and space. As such, X_n has continuous sample paths. Note that $Y(t)$ has mean zero, and $\mathrm{Var}(Y(n)) = \mathrm{Var}(S_n) = \sigma^2 n$ for $n \in \mathbb{N}$. Thus, $X_n(t)$ has mean zero for every t, and if nt is an integer, $\mathrm{Var}(X_n(t)) = (\sigma^2 n)^{-1} \mathrm{Var}(Y(nt)) = t$. Hence $X_n(t)$ has the same mean and variance as $W(t)$, at least when nt is integer.

In the following theorem, $C[0,\infty)$ denotes the space of all continuous functions $f : [0,\infty) \to \mathbb{R}$. We equip it with the metric

$$\rho(x_1, x_2) := \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{0 \leq t \leq n} \left(|x_1(t) - x_2(t)| \wedge 1\right).$$

It can be shown that ρ is a metric, and $(C[0,\infty), \rho)$ is a complete, separable metric space. Furthermore, the Borel σ-algebra $\mathcal{B}(C[0,\infty))$ generated by the open sets of $(C[0,\infty), \rho)$ is the smallest σ-algebra containing all the finite-dimensional cylinder sets of the form

$$C = \{x \in C[0,\infty) : (x(t_1), \dots, x(t_n)) \in A\}, \qquad n \geq 1, \quad A \in \mathcal{B}(\mathbb{R}^n), \quad t_i \geq 0.$$

Theorem 3.37 (Donsker's invariance principle)**.** *Let* P_n *be the measure induced by the process* X_n *on* $(C[0,\infty)), \mathcal{B}(C[0,\infty)))$. *That is,*

$$\mathrm{P}_n(H) = \mathrm{P}(\omega \in \Omega : X_n(\cdot, \omega) \in H), \qquad H \in \mathcal{B}(C[0,\infty)).$$

Then P_n *converges weakly to a probability measure* P_* *on* $(C[0,\infty)), \mathcal{B}(C[0,\infty)))$, *under which the coordinate mapping process* $W(t,x) := x(t)$ *on* $C[0,\infty)$ *is a Brownian motion. Equivalently, the sequence of processes* $\{X_n\}_n$ *converges weakly to a Brownian motion.*

This theorem is intuitively quite plausible, but its proof is long and technical - see Karatzas and Shreve, pp. 66-71. The probability measure P_* in the above theorem is called *Wiener measure.* We can now think of Brownian motion as a (function-valued!) random variable on the probability space $(C[0,\infty)), \mathcal{B}(C[0,\infty)), \mathrm{P}_*)$, defined by

$$W(t,x) := x(t), \qquad x \in C[0,\infty), \quad t \geq 0.$$

This space is called the *canonical probability space* for Brownian motion, and $W(t,\cdot)$ as defined above is called *canonical Brownian motion.*