# Brownian motion

**by Peter Mörters (University of Bath)**

This is a set of lecture notes based on a graduate course given at the Taught Course Centre in Mathematics in 2011. The course is based on a selection of material from my book with Yuval Peres, entitled *Brownian motion*, which was published by Cambridge University Press in 2010.

## 1 Lévy's construction of Brownian motion and modulus of continuity

Much of probability theory is devoted to describing the *macroscopic picture* emerging in random systems defined by a host of *microscopic random effects*. Brownian motion is the macroscopic picture emerging from a particle moving randomly on a line without making very big jumps. On the microscopic level, at any time step, the particle receives a random displacement, caused for example by other particles hitting it or by an external force, so that, if its position at time zero is $S_0$, its position at time $n$ is given as $S_n = S_0 + \sum_{i=1}^n X_i$, where the displacements $X_1, X_2, X_3, \ldots$ are assumed to be independent, identically distributed random variables with mean zero. The process $\{S_n : n \geqslant 0\}$ is a random walk, the displacements represent the microscopic inputs.

It turns out that not all the features of the microscopic inputs contribute to the macroscopic picture. Indeed, all random walks whose displacements have zero mean and variance one give rise to the same macroscopic process, and even the assumption that the displacements have to be independent and identically distributed can be substantially relaxed. This effect is called *universality*, and the macroscopic process is often called a *universal object*. It is a common approach in probability to study various phenomena through the associated universal objects. If the jumps of a random walk are sufficiently tame to become negligible in the macroscopic picture, any continuous time stochastic process $\{B(t) : t \geqslant 0\}$ describing the macroscopic features of this random walk should have the following properties:

(i) for all times $0 \leqslant t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ the random variables

$$B(t_n) - B(t_{n-1}),\, B(t_{n-1}) - B(t_{n-2}),\, \ldots,\, B(t_2) - B(t_1)$$

are independent;

(ii) the distribution of the increment $B(t+h) - B(t)$ has zero mean and does not depend on $t$;

(iii) the process $\{B(t) : t \geqslant 0\}$ has almost surely continuous paths.

It follows (with some work) from the central limit theorem that these features imply that there exists $\sigma \geqslant 0$ such that

(iv) for every $t \geqslant 0$ and $h \geqslant 0$ the increment $B(t+h) - B(t)$ is normally distributed with mean zero and variance $h\sigma^2$.

The process corresponding to $\sigma = 1$ is called *Brownian motion.* If $B(0) = 0$ we say that it is a *standard Brownian motion.*

It is a substantial issue whether the conditions in the definition of Brownian motion are free of contradiction.

**Theorem 1.1** (Wiener 1923)**.** *Standard Brownian motion exists.*

**Proof.** We first construct Brownian motion on the interval $[0,1]$ as a random element on the space $\mathbf{C}[0,1]$ of continuous functions on $[0,1]$. The idea is to construct the right joint distribution of Brownian motion step by step on the finite sets

$$\mathcal{D}_n = \left\{ \tfrac{k}{2^n} \colon 0 \leqslant k \leqslant 2^n \right\}$$

of dyadic points. We then interpolate the values on $\mathcal{D}_n$ linearly and check that the uniform limit of these continuous functions exists and is a Brownian motion.

To do this let $\mathcal{D} = \bigcup_{n=0}^{\infty} \mathcal{D}_n$ and let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space on which a collection $\{Z_t \colon t \in \mathcal{D}\}$ of independent, standard normally distributed random variables can be defined. Let $B(0) := 0$ and $B(1) := Z_1$. For each $n \in \mathbb{N}$ we define the random variables $B(d)$, $d \in \mathcal{D}_n$ such that

(i) for all $r < s < t$ in $\mathcal{D}_n$ the random variable $B(t) - B(s)$ is normally distributed with mean zero and variance $t - s$, and is independent of $B(s) - B(r)$,

(ii) the vectors $(B(d) \colon d \in \mathcal{D}_n)$ and $(Z_t \colon t \in \mathcal{D} \setminus \mathcal{D}_n)$ are independent.

Note that we have already done this for $\mathcal{D}_0 = \{0,1\}$. Proceeding inductively we may assume that we have succeeded in doing it for some $n-1$. We then define $B(d)$ for $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ by

$$B(d) = \frac{B(d - 2^{-n}) + B(d + 2^{-n})}{2} + \frac{Z_d}{2^{(n+1)/2}}\,.$$

Note that the first summand is the linear interpolation of the values of $B$ at the neighbouring points of $d$ in $\mathcal{D}_{n-1}$. Therefore $B(d)$ is independent of $(Z_t \colon t \in \mathcal{D} \setminus \mathcal{D}_n)$ and the second property is fulfilled.

Moreover, as $\frac{1}{2}[B(d + 2^{-n}) - B(d - 2^{-n})]$ depends only on $(Z_t \colon t \in \mathcal{D}_{n-1})$, it is independent of $Z_d/2^{(n+1)/2}$. By our induction assumptions both terms are normally distributed with mean zero and variance $2^{-(n+1)}$. Hence their sum $B(d) - B(d - 2^{-n})$ and their difference $B(d + 2^{-n}) - B(d)$ are independent and normally distributed with mean zero and variance $2^{-n}$.

Indeed, all increments $B(d) - B(d-2^{-n})$, for $d \in \mathcal{D}_n \setminus \{0\}$, are independent. To see this it suffices to show that they are pairwise independent, as the vector of these increments is Gaussian. We have seen already that pairs $B(d)-B(d-2^{-n})$, $B(d+2^{-n})-B(d)$ with $d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ are independent. The other possibility is that the increments are over intervals separated by some $d \in \mathcal{D}_{n-1}$. Choose $d \in \mathcal{D}_j$ with this property and minimal $j$, so that the two intervals are contained in $[d-2^{-j}, d]$, respectively $[d, d+2^{-j}]$. By induction the increments over these two intervals of length $2^{-j}$ are independent, and the increments over the intervals of length $2^{-n}$ are constructed from the independent increments $B(d)-B(d-2^{-j})$, respectively $B(d+2^{-j}) - B(d)$, using a disjoint set of variables $(Z_t \colon t \in \mathcal{D}_n)$. Hence they are independent and this implies the first property, and completes the induction step.

Having thus chosen the values of the process on all dyadic points, we interpolate between them. Formally, define

$$F_0(t) = \begin{cases} Z_1 & \text{for } t = 1, \\ 0 & \text{for } t = 0, \\ \text{linear} & \text{in between,} \end{cases}$$

and, for each $n \geqslant 1$,

$$F_n(t) = \begin{cases} 2^{-(n+1)/2} Z_t & \text{for } t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1} \\ 0 & \text{for } t \in \mathcal{D}_{n-1} \\ \text{linear between} & \text{consecutive points in } \mathcal{D}_n. \end{cases}$$

These functions are continuous on $[0,1]$ and, for all $n$ and $d \in \mathcal{D}_n$,

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d). \tag{1}$$

On the other hand, we have, by definition of $Z_d$ and the from of Gaussian tails, for $c > 1$ and large $n$,

$$\mathbb{P}\{|Z_d| \geqslant c\sqrt{n}\} \leqslant \exp\left(\frac{-c^2 n}{2}\right),$$

so that the series

$$\begin{aligned} \sum_{n=0}^{\infty} \mathbb{P}\{ \text{ there exists } d \in \mathcal{D}_n \text{ with } |Z_d| \geqslant c\sqrt{n}\} &\leqslant \sum_{n=0}^{\infty} \sum_{d \in \mathcal{D}_n} \mathbb{P}\{|Z_d| \geqslant c\sqrt{n}\} \\ &\leqslant \sum_{n=0}^{\infty} (2^n + 1) \exp\left(\frac{-c^2 n}{2}\right), \end{aligned}$$

converges as soon as $c > \sqrt{2 \log 2}$. Fix such a $c$. By the Borel–Cantelli lemma there exists a random (but almost surely finite) $N$ such that for all $n \geqslant N$ and $d \in \mathcal{D}_n$ we have $|Z_d| < c\sqrt{n}$. Hence, for all $n \geqslant N$,

$$\|F_n\|_\infty < c\sqrt{n} 2^{-n/2}. \tag{2}$$

This upper bound implies that, almost surely, the series

$$B(t) = \sum_{n=0}^{\infty} F_n(t)$$

is uniformly convergent on $[0, 1]$. The increments of this process have the right finite-dimensional distributions on the dense set $\mathcal{D} \subset [0, 1]$ and therefore also in general, by approximation.

We have thus constructed a process $B \colon [0, 1] \to \mathbb{R}$ with the properties of Brownian motion. To obtain a Brownian motion on $[0, \infty)$ we pick a sequence $B_0, B_1, \ldots$ of independent $\mathbf{C}[0, 1]$-valued random variables with this distribution, and define $\{B(t) \colon t \geqslant 0\}$ by gluing together these parts to make a continuous function. ■

**Lemma 1.2** (Scaling invariance)**.** *Suppose* $\{B(t) \colon t \geqslant 0\}$ *is a standard Brownian motion and let* $a > 0$*. Then the process* $\{X(t) \colon t \geqslant 0\}$ *defined by* $X(t) = \frac{1}{a} B(a^2 t)$ *is also a standard Brownian motion.*

The definition of Brownian motion already requires that the sample functions are continuous almost surely. This implies that on the interval $[0, 1]$ (or any other compact interval) the sample functions are uniformly continuous, i.e. there exists some (random) function $\varphi$ with $\lim_{h \downarrow 0} \varphi(h) = 0$ called a **modulus of continuity** of the function $B : [0, 1] \to \mathbb{R}$, such that

$$\limsup_{h \downarrow 0} \sup_{0 \leqslant t \leqslant 1-h} \frac{|B(t+h) - B(t)|}{\varphi(h)} \leqslant 1. \tag{3}$$

Can we achieve such a bound with a deterministic function $\varphi$, i.e. is there a nonrandom modulus of continuity for the Brownian motion? The answer is yes, as the following theorem shows.

**Theorem 1.3.** *There exists a constant* $C > 0$ *such that, almost surely, for every sufficiently small* $h > 0$ *and all* $0 \leqslant t \leqslant 1 - h$*,*

$$\big|B(t+h) - B(t)\big| \leqslant C \sqrt{h \log(1/h)}.$$

**Proof.** This follows quite elegantly from Lévy's construction of Brownian motion. Recall the notation introduced there and that we have represented Brownian motion as a series $B(t) = \sum_{n=0}^{\infty} F_n(t)$, where each $F_n$ is a piecewise linear function. The derivative of $F_n$ exists almost everywhere, and by definition and (2), for any $c > \sqrt{2 \log 2}$ there exists a (random) $N \in \mathbb{N}$ such that, for all $n > N$,

$$\|F_n'\|_\infty \leqslant \frac{2\|F_n\|_\infty}{2^{-n}} \leqslant 2c\sqrt{n} 2^{n/2} \, .$$

Now for each $t, t + h \in [0, 1]$, using the mean-value theorem,

$$|B(t+h) - B(t)| \leqslant \sum_{n=0}^{\infty} |F_n(t+h) - F_n(t)| \leqslant \sum_{n=0}^{\ell} h\|F_n'\|_\infty + \sum_{n=\ell+1}^{\infty} 2\|F_n\|_\infty \, .$$

Hence, using (2) again, we get for all $\ell > N$, that this is bounded by

$$h \sum_{n=0}^{N} \|F_n'\|_\infty + 2ch \sum_{n=N}^{\ell} \sqrt{n} 2^{n/2} + 2c \sum_{n=\ell+1}^{\infty} \sqrt{n} 2^{-n/2}.$$

We now suppose that $h$ is small enough that the first summand is smaller than $\sqrt{h \log(1/h)}$ and that $\ell$ defined by $2^{-\ell} < h \leqslant 2^{-\ell+1}$ exceeds $N$. For this choice of $\ell$ the second and third summands are also bounded by constant multiples of $\sqrt{h \log(1/h)}$ as both sums are dominated by their largest element. Hence we get (3) with a deterministic function $\varphi(h) = C\sqrt{h \log(1/h)}$. ■

This upper bound is pretty close to the optimal result, as the following famous result of Lévy shows.

**Theorem 1.4** (Lévy's modulus of continuity (1937)). *Almost surely,*

$$\limsup_{h \downarrow 0} \sup_{0 \leqslant t \leqslant 1-h} \frac{|B(t+h) - B(t)|}{\sqrt{2h \log(1/h)}} = 1\,.$$

**Remark 1.5.** The limsup in Theorem 1.4 may be replaced by a limit. ◇

# 2 (Non-)differentiability of Brownian motion

While it is not so easy to construct continuous functions that are noweher differentiable, it turns out that Brownian motion has this property almost surely. For the statement of this fact define, for a function $f\colon [0,1) \to \mathbb{R}$, the **upper** and **lower right derivatives**

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

**Theorem 2.1** (Paley, Wiener and Zygmund 1933). *Almost surely, Brownian motion is nowhere differentiable. Furthermore, almost surely, for all $t$,*

$$\textit{either} \quad D^* B(t) = +\infty \quad \textit{or} \quad D_* B(t) = -\infty \quad \textit{(or both.)}$$

**Proof.** Suppose that there is a $t_0 \in [0,1]$ such that $-\infty < D_* B(t_0) \leqslant D^* B(t_0) < \infty$. Then

$$\limsup_{h \downarrow 0} \frac{|B(t_0+h) - B(t_0)|}{h} < \infty,$$

and this implies that for some finite constant $M$,

$$\sup_{h \in [0,1]} \frac{|B(t_0+h) - B(t_0)|}{h} \leqslant M.$$

It suffices to show that this event has probability zero for any $M$. From now on fix $M$. If $t_0$ is contained in the binary interval $[(k-1)/2^n, k/2^n]$ for $n > 2$, then for all $1 \leqslant j \leqslant 2^n - k$ the triangle inequality gives

$$\begin{aligned}\big|B\left((k+j)/2^n\right) - &B\left((k+j-1)/2^n\right)\big| \\ &\leqslant |B\left((k+j)/2^n\right) - B(t_0)| + |B(t_0) - B\left((k+j-1)/2^n\right)| \\ &\leqslant M(2j+1)/2^n.\end{aligned}$$

Define events

$$\Omega_{n,k} := \Big\{\big|B\left((k+j)/2^n\right) - B\left((k+j-1)/2^n\right)\big| \leqslant M(2j+1)/2^n \text{ for } j = 1,2,3\Big\}.$$

By independence of the increments and the scaling property, for $1 \leqslant k \leqslant 2^n - 3$,

$$\begin{aligned}\mathbb{P}(\Omega_{n,k}) &\leqslant \prod_{j=1}^{3} \mathbb{P}\big\{\big|B\left((k+j)/2^n\right) - B\left((k+j-1)/2^n\right)\big| \leqslant M(2j+1)/2^n\big\} \\ &\leqslant \mathbb{P}\left\{|B(1)| \leqslant 7M/\sqrt{2^n}\right\}^3,\end{aligned}$$

which is at most $(7M2^{-n/2})^3$, since the normal density is bounded by $1/2$. Hence

$$\mathbb{P}\left(\bigcup_{k=1}^{2^n-3} \Omega_{n,k}\right) \leqslant 2^n(7M2^{-n/2})^3 = (7M)^3 2^{-n/2},$$

which is summable over all $n$. Hence, by the Borel–Cantelli lemma,

$$\begin{aligned}\mathbb{P}\Big\{ &\text{ there is } t_0 \in [0,1) \text{ with } \sup_{h\in[0,1]} \frac{|B(t_0+h) - B(t_0)|}{h} \leqslant M\Big\} \\ &\leqslant \mathbb{P}\left(\bigcup_{k=1}^{2^n-3} \Omega_{n,k} \text{ for infinitely many } n\right) = 0. \qquad \square\end{aligned}$$

There is an abundance of interesting statements about the right derivatives of Brownian motion. As a taster we mention here that Lévy asked whether, almost surely, $D^*B(t) \in \{-\infty, \infty\}$ for every $t \in [0,1)$. Barlow and Perkins (1984) showed that this is not the case. We will see below that there exist points where Brownian motion has an *infinite* derivative.

Before we approach this, we establish the strong Markov property, which is an essential tool in the study of Brownian motion. It states that Brownian motion is started afresh at certain (possibly random) times called stopping times.

For $t \geqslant 0$ we define a $\sigma$-algebra

$$\mathcal{F}^+(t) := \bigcap_{\varepsilon>0} \sigma(B_s\colon 0 \leqslant s < t+\varepsilon).$$

The following result is easy to show.

**Theorem 2.2** (Markov property)**.** *For every $s \geqslant 0$ the process $\{B(t+s) - B(s)\colon t \geqslant 0\}$ is a standard Brownian motion and independent of $\mathcal{F}^+(s)$.*

A random variable $T$ with values in $[0,\infty]$ is called a **stopping time** if $\{T \leqslant t\} \in \mathcal{F}^+(t)$, for every $t \geqslant 0$. In particular the times of first entry or exit from an open or closed set are stopping times. We define, for every stopping time $T$, the $\sigma$-algebra

$$\mathcal{F}^+(T) = \{A \in \mathcal{A}\colon A \cap \{T \leqslant t\} \in \mathcal{F}^+(t) \text{ for all } t \geqslant 0\}$$

which is the collection of events that depend only on $\{B(t)\colon 0 \leqslant t \leqslant T\}$ . We now state the *strong Markov property* for Brownian motion, which was rigorously established by Hunt and Dynkin in the 1950s.

**Theorem 2.3** (Strong Markov property)**.** *For every almost surely finite stopping time $T$, the process $\{B(T+t) - B(T)\colon t \geqslant 0\}$ is a standard Brownian motion independent of $\mathcal{F}^+(T)$.*

The proof follows by approximation from Theorem 2.2 and will be skipped here. To formulate an important consequence of the Markov property, we introduce the convention that $\{B(t)\colon t \geqslant 0\}$ under $\mathbb{P}_x$ is a Brownian motion started in $x$.

**Theorem 2.4** (Blumenthal's 0-1 law)**.** *Every $A \in \mathcal{F}^+(0)$ has $\mathbb{P}_x(A) \in \{0,1\}$.*

**Proof.** Using Theorem 2.3 for $s = 0$ we see that any $A \in \sigma(B(t)\colon t \geqslant 0)$ is independent of $\mathcal{F}^+(0)$. This applies in particular to $A \in \mathcal{F}^+(0)$, which therefore is independent of itself, hence has probability zero or one. ■

As a first application we show that a standard Brownian motion has positive and negative values and zeros in every small interval to the right of 0.

**Theorem 2.5.** *Let $\tau = \inf\{t > 0\colon B(t) > 0\}$ and $\sigma = \inf\{t > 0\colon B(t) = 0\}$. Then*

$$\mathbb{P}_0\{\tau = 0\} = \mathbb{P}_0\{\sigma = 0\} = 1\,.$$

**Proof.** The event

$$\{\tau = 0\} = \bigcap_{n=1}^{\infty} \Big\{ \text{ there is } 0 < \varepsilon < 1/n \text{ such that } B(\varepsilon) > 0 \Big\}$$

is clearly in $\mathcal{F}^+(0)$. Hence we just have to show that this event has positive probability. This follows, as $\mathbb{P}_0\{\tau \leqslant t\} \geqslant \mathbb{P}_0\{B(t) > 0\} = 1/2$ for $t > 0$. Hence $\mathbb{P}_0\{\tau = 0\} \geqslant 1/2$ and we have shown the first part. The same argument works replacing $B(t) > 0$ by $B(t) < 0$ and from these two facts $\mathbb{P}_0\{\sigma = 0\} = 1$ follows, using the intermediate value property of continuous functions. ■

We now exploit the Markov property to study the local and global extrema of Brownian motion.

**Theorem 2.6.** *For a Brownian motion $\{B(t)\colon 0 \leqslant t \leqslant 1\}$, almost surely,*

**(a)** *every local maximum is a strict local maximum;*

**(b)** *the set of times where the local maxima are attained is countable and dense;*

**(c)** *the global maximum is attained at a unique time.*

**Proof.** We first show that, given two nonoverlapping closed time intervals the maxima of Brownian motion on them are different almost surely. Let $[a_1, b_1]$ and $[a_2, b_2]$ be two fixed intervals with $b_1 \leqslant a_2$. Denote by $m_1$ and $m_2$, the maxima of Brownian motion on these two intervals. Note first that, by the Markov property together with Theorem 2.5, almost surely $B(a_2) < m_2$. Hence this maximum agrees with maximum in the interval $[a_2 - \frac{1}{n}, b_2]$, for some $n \in \mathbb{N}$, and we may therefore assume in the proof that $b_1 < a_2$.

Applying the Markov property at time $b_1$ we see that the random variable $B(a_2) - B(b_1)$ is independent of $m_1 - B(b_1)$. Using the Markov property at time $a_2$ we see that $m_2 - B(a_2)$ is also independent of both these variables. The event $m_1 = m_2$ can be written as

$$B(a_2) - B(b_1) = (m_1 - B(b_1)) - (m_2 - B(a_2)).$$

Conditioning on the values of the random variables $m_1 - B(b_1)$ and $m_2 - B(a_2)$, the left hand side is a continuous random variable and the right hand side a constant, hence this event has probability 0.

**(a)** By the statement just proved, almost surely, all nonoverlapping pairs of nondegenerate compact intervals with rational endpoints have different maxima. If Brownian motion however has a non-strict local maximum, there are two such intervals where Brownian motion has the same maximum.

**(b)** In particular, almost surely, the maximum over any nondegenerate compact interval with rational endpoints is not attained at an endpoint. Hence every such interval contains a local maximum, and the set of times where local maxima are attained is dense. As every local maximum is strict, this set has at most the cardinality of the collection of these intervals.

**(c)** Almost surely, for any rational number $q \in [0, 1]$ the maximum in $[0, q]$ and in $[q, 1]$ are different. Note that, if the global maximum is attained for two points $t_1 < t_2$ there exists a rational number $t_1 < q < t_2$ for which the maximum in $[0, q]$ and in $[q, 1]$ agree. ■

We will see many applications of the strong Markov property later, however, the next result, the reflection principle, is particularly interesting.

**Theorem 2.7** (Reflection principle)**.** *If $T$ is a stopping time and $\{B(t)\colon t \geqslant 0\}$ is a standard Brownian motion, then the process $\{B^*(t)\colon t \geqslant 0\}$ defined by*

$$B^*(t) = B(t)1_{\{t \leqslant T\}} + (2B(T) - B(t))1_{\{t > T\}}$$

*is also a standard Brownian motion.*

**Proof.** If $T$ is finite, by the strong Markov property both paths

$$\{B(t+T)-B(T)\colon t\geqslant 0\} \text{ and } \{-(B(t+T)-B(T))\colon t\geqslant 0\} \tag{4}$$

are Brownian motions and independent of the beginning $\{B(t)\colon 0\leqslant t\leqslant T\}$. The process arising from glueing the first path in (4) to $\{B(t)\colon 0\leqslant t\leqslant T\}$ and the process arising from glueing the second path in (4) to $\{B(t)\colon 0\leqslant t\leqslant T\}$ have the same distribution. The first is just $\{B(t)\colon t\geqslant 0\}$, the second is $\{B^*(t)\colon t\geqslant 0\}$, as introduced in the statement. ∎

Now we apply the reflection principle. Let $M(t)=\max_{0\leqslant s\leqslant t} B(s)$. A priori it is not at all clear what the distribution of this random variable is, but we can determine it as a consequence of the reflection principle.

**Lemma 2.8.** $\mathbb{P}_0\{M(t)>a\}=2\mathbb{P}_0\{B(t)>a\}=\mathbb{P}_0\{|B(t)|>a\}$ *for all* $a>0$.

**Proof.** Let $T=\inf\{t\geqslant 0\colon B(t)=a\}$ and let $\{B^*(t)\colon t\geqslant 0\}$ be Brownian motion reflected at the stopping time $T$. Then

$$\{M(t)>a\}=\{B(t)>a\}\cup\{M(t)>a,\ B(t)\leqslant a\}.$$

This is a disjoint union and the second summand coincides with event $\{B^*(t)\geqslant a\}$. Hence the statement follows from the reflection principle. ∎

**Theorem 2.9.** *Almost surely,*

$$D_*B(t_0)=D^*B(t_0)=-\infty,$$

*where* $t_0$ *is the uniquely determined maximum of Brownian motion on* $[0,1]$.

**Proof.** We first fix $\varepsilon, a>0$. We denote by $B$ the event that $\varepsilon$ is small enough in the sense of Theorem 1.3. Given an interval $I\subset[\varepsilon,1-\varepsilon]$ with length $0<h<(\varepsilon/6)^4$, we consider the event $A$ that $t_0\in I$ and we have

$$B(t_0+\tilde h)-B(t_0)>-2ah^{1/4} \qquad \text{for some } h^{1/4}<\tilde h\leqslant 2h^{1/4}.$$

We now denote by $t_L$ the left endpoint of $I$. By Theorem 1.3, on the event $B$,

$$B(t_0)-B(t_L)\leqslant C\sqrt{h\log(1/h)}.$$

Hence the event $A\cap B$ implies the following events

$$A_1=\big\{B(t_L-s)-B(t_L)\leqslant C\sqrt{h\log(1/h)} \text{ for all } s\in[0,\varepsilon]\big\},$$

$$A_2=\big\{B(t_L+s)-B(t_L)\leqslant C\sqrt{h\log(1/h)} \text{ for all } s\in[0,h^{1/4}]\big\}.$$

We now define the stopping time

$$T:=\inf\{s>t_L+h^{1/4} : B(s)>B(t_L)-2ah^{1/4}\}.$$

Then the event $A \cap B$ implies that $T \leqslant t_L + 3h^{1/4}$ and this implies the event

$$A_3 = \big\{ B(T+s) - B(T) \leqslant 2ah^{1/4} + C\sqrt{h \log(1/h)} \text{ for all } s \in [0, \varepsilon/2] \big\}.$$

Now by the strong Markov property, these three events are independent and we obtain

$$\mathbb{P}(A \cap B) \leqslant \mathbb{P}(A_1)\, \mathbb{P}(A_2)\, \mathbb{P}(A_3).$$

Using Lemma 2.8 we obtain

$$\begin{aligned}
\mathbb{P}(A_1) &= \mathbb{P}\big\{ |B(\varepsilon)| \leqslant C\sqrt{h \log(1/h)} \big\} \leqslant 2 \tfrac{1}{\sqrt{2\pi\varepsilon}}\, C\sqrt{h \log(1/h)},\\
\mathbb{P}(A_2) &= \mathbb{P}\big\{ |B(h^{1/4})| \leqslant C\sqrt{h \log(1/h)} \big\} \leqslant 2 \tfrac{1}{\sqrt{2\pi h^{1/4}}}\, C\sqrt{h \log(1/h)},\\
\mathbb{P}(A_3) &= \mathbb{P}\big\{ |B(\varepsilon/2)| \leqslant 2ah^{1/4} + C\sqrt{h \log(1/h)} \big\} \leqslant 2 \tfrac{1}{\sqrt{\pi\varepsilon}} \big( C\, h^{1/4} + 2ah^{1/4} \big).
\end{aligned}$$

Hence we obtain, for a suitable constant $K > 0$, depending on $a$ and $\varepsilon$, that

$$\mathbb{P}(A \cap B) \leqslant K\, h^{9/8} \log(1/h).$$

Summing first over a covering collection of $1/h$ intervals of length $h$ that cover $[\varepsilon, 1-\varepsilon]$ and then taking $h = 2^{-4n-4}$ and summing over $n$, we see that

$$\sum_{n=1}^{\infty} \mathbb{P}\Big\{ \varepsilon \leqslant t_0 \leqslant 1-\varepsilon, \varepsilon \text{ small, and } \sup_{2^{-n-1} < h \leqslant 2^{-n}} \frac{B(t_0+h) - B(t_0)}{h} > -a \Big\} < \infty,$$

and from the Borel–Cantelli lemma we obtain that, almost surely, either $t_0 \notin [\varepsilon, 1-\varepsilon]$, or $\varepsilon$ is too large in the sense of Theorem 1.3 or

$$\limsup_{h \downarrow 0} \frac{B(t_0+h) - B(t_0)}{h} \leqslant -a.$$

Taking $a \uparrow \infty$ and $\varepsilon \downarrow 0$ gives that, almost surely, $D^* B(t_0) = -\infty$, as required. ∎

# 3 Lévy's theorem on the maximum process

We have seen from the reflection principle that the maximum of a Brownian motion at time $t$ has the same law as the absolute value as the same time $|B(t)|$. Obviously, this does not extend to the maximum process $\{M(t) \colon t \geqslant 0\}$ , defined by $M(t) = \max_{0 \leqslant s \leqslant t} B(s)$, and the reflected Brownian motion $\{|B(t)| \colon t \geqslant 0\}$. The relationship between these processes is giben by a famous theorem of Lévy.

**Theorem 3.1** (Lévy 1948)**.** *Let $\{M(t) \colon t \geqslant 0\}$ be the maximum process of a standard Brownian motion, then, the process $\{Y(t) \colon t \geqslant 0\}$ defined by $Y(t) = M(t) - B(t)$ is a reflected Brownian motion.*

**Proof.** Fix $s > 0$, and consider the two processes $\{\hat{B}(t)\colon t \geqslant 0\}$ defined by

$$\hat{B}(t) = B(s+t) - B(s) \text{ for } t \geqslant 0,$$

and $\{\hat{M}(t)\colon t \geqslant 0\}$ defined by

$$\hat{M}(t) = \max_{0 \leqslant u \leqslant t} \hat{B}(u) \text{ for } t \geqslant 0.$$

Because $Y(s)$ is $\mathcal{F}^+(s)$-measurable, it suffices to check that conditional on $\mathcal{F}^+(s)$, for every $t \geqslant 0$, the random variable $Y(s+t)$ has the same distribution as $|Y(s) + \hat{B}(t)|$. Indeed, this directly implies that $\{Y(t)\colon t \geqslant 0\}$ is a Markov process with the same transition kernel as the reflected Brownian motion, and the result follows by continuity of paths. To prove the claim fix $s, t \geqslant 0$ and observe that $M(s+t) = M(s) \vee (B(s) + \hat{M}(t))$, and hence

$$Y(s+t) = (M(s) \vee (B(s) + \hat{M}(t))) - (B(s) + \hat{B}(t)).$$

Using the fact that $(a \vee b) - c = (a-c) \vee (b-c)$, we have

$$Y(s+t) = \big(Y(s) \vee \hat{M}(t)\big) - \hat{B}(t).$$

To finish, it suffices to check, for every $y \geqslant 0$, that $y \vee \hat{M}(t) - \hat{B}(t)$ has the same distribution as $|y + \hat{B}(t)|$. For any $a \geqslant 0$ write

$$P_1 = \mathbb{P}\{y - \hat{B}(t) > a\}, \qquad P_2 = \mathbb{P}\big\{y - \hat{B}(t) \leqslant a \text{ and } \hat{M}(t) - \hat{B}(t) > a\big\}.$$

Then $\mathbb{P}\{y \vee \hat{M}(t) - \hat{B}(t) > a\} = P_1 + P_2$. Since $\{\hat{B}(t)\colon t \geqslant 0\}$ has the same distribution as $\{-\hat{B}(t)\colon t \geqslant 0\}$ we have $P_1 = \mathbb{P}\{y + \hat{B}(t) > a\}$. To study the second term it is useful to define the time reversed Brownian motion $\{W(u)\colon 0 \leqslant u \leqslant t\}$ by $W(u) := \hat{B}(t-u) - \hat{B}(t)$. Note that this process is also a Brownian motion on $[0,t]$. Let $M_W(t) = \max_{0 \leqslant u \leqslant t} W(u)$. Then $M_W(t) = \hat{M}(t) - \hat{B}(t)$. Since $W(t) = -\hat{B}(t)$, we have

$$P_2 = \mathbb{P}\{y + W(t) \leqslant a \text{ and } M_W(t) > a\}.$$

Using the reflection principle by reflecting $\{W(u)\colon 0 \leqslant u \leqslant t\}$ at the first time it hits $a$, we get another Brownian motion $\{W^*(u)\colon 0 \leqslant u \leqslant t\}$. In terms of this Brownian motion we have $P_2 = \mathbb{P}\{W^*(t) \geqslant a + y\}$. Since it has the same distribution as $\{-\hat{B}(t)\colon t \geqslant 0\}$, it follows that $P_2 = \mathbb{P}\{y + \hat{B}(t) \leqslant -a\}$. The Brownian motion $\{\hat{B}(t)\colon t \geqslant 0\}$ has continuous distribution, and so, by adding $P_1$ and $P_2$, we get $\mathbb{P}\{y \vee \hat{M}(t) - \hat{B}(t) > a\} = \mathbb{P}\{|y + \hat{B}(t)| > a\}$. This proves the main step and, consequently, the theorem. ■

While, as seen above, $\{M(t) - B(t)\colon t \geqslant 0\}$ is a Markov process, it is important to note that the maximum process $\{M(t)\colon t \geqslant 0\}$ itself is not a Markov process. However the times when new maxima are achieved form a Markov process, as the following theorem shows.

**Theorem 3.2.** *For any $a \geqslant 0$ define the stopping times*

$$T_a = \inf\{t \geqslant 0 \colon B(t) = a\}.$$

*Then $\{T_a \colon a \geqslant 0\}$ is an increasing Markov process with transition densities*

$$p(a,t,s) = \tfrac{a}{\sqrt{2\pi(s-t)^3}} \exp\big(-\tfrac{a^2}{2(s-t)}\big)\, 1\{s > t\}, \qquad \text{for } a > 0.$$

*This process is called the* **stable subordinator** *of index $\frac{1}{2}$.*

**Remark 3.3.** As the transition densities satisfy the *shift-invariance property*

$$p(a,t,s) = p(a,0,s-t) \quad \text{ for all } a \geqslant 0 \text{ and } s,t \geqslant 0,$$

the stable subordinator has stationary, independent increments. ◇

**Proof.** Fix $a \geqslant b \geqslant 0$ and note that for all $t \geqslant 0$ we have

$$\begin{aligned} &\big\{T_a - T_b = t\big\} \\ &\quad = \big\{B(T_b+s) - B(T_b) < a-b, 0 < s < t, \text{ and } B(T_b+t) - B(T_b) = a-b\big\}. \end{aligned}$$

By the strong Markov property of Brownian motion this event is independent of $\mathcal{F}^+(T_b)$ and therefore in particular of $\{T_d \colon d \leqslant b\}$. This proves the Markov property of $\{T_a \colon a \geqslant 0\}$. The form of the transition kernel follows from the reflection principle,

$$\begin{aligned} \mathbb{P}\{T_a - T_b \leqslant t\} &= \mathbb{P}\{T_{a-b} \leqslant t\} = \mathbb{P}\big\{\max_{0 \leqslant s \leqslant t} B(s) \geqslant a-b\big\} \\ &= 2\mathbb{P}\big\{B(t) \geqslant a-b\big\} = 2\int_{a-b}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\big(-\tfrac{x^2}{2t}\big)\, dx \\ &= \int_0^t \frac{1}{\sqrt{2\pi s^3}}\, (a-b) \exp\big(-\tfrac{(a-b)^2}{2s}\big)\, ds, \end{aligned}$$

where we used the substitution $x = \sqrt{t/s}\,(a-b)$ in the last step. ■

# 4 Exit and hitting times, and Wald's lemmas

Given a stopping time $T$, what can we say about $E[B(T)]$? Note that if $T$ is a fixed time, we always have $E[B(T)] = 0$, but this does not extend to all stopping times, as the example of $T = \inf\{t > 0 \colon B(t) = 1\}$ shows. To decide which stopping times have this property, we develop a bit of martingale theory in continuous time.

A real-valued stochastic process $\{X(t) \colon t \geqslant 0\}$ is a **martingale** with respect to a filtration $(\mathcal{F}(t) \colon t \geqslant 0)$ if it is adapted to the filtration, $\mathbb{E}|X(t)| < \infty$ for all $t \geqslant 0$ and, for any pair of times $0 \leqslant s \leqslant t$,

$$\mathbb{E}\big[X(t)\,\big|\,\mathcal{F}(s)\big] = X(s) \text{ almost surely.}$$

The process is called a **submartingale** if $\geqslant$ holds, and a **supermartingale** if $\leqslant$ holds in the display above. We observe that Brownian motion is a martingale, and reflected Brownian motion is a submartingale, but not a martingale.

We now state a useful fact about martingales, which we will exploit extensively, the *optional stopping theorem.* This results is well-known in the discrete time setting and adapt easily to the continuous world.

**Proposition 4.1** (Optional stopping theorem)**.** *Suppose $\{X(t) \colon t \geqslant 0\}$ is a continuous martingale, and $0 \leqslant S \leqslant T$ are stopping times. If the process $\{X(t \wedge T) \colon t \geqslant 0\}$ is dominated by an integrable random variable $X$, i.e. $|X(t \wedge T)| \leqslant X$ almost surely, for all $t \geqslant 0$, then*

$$\mathbb{E}\big[X(T)\,\big|\mathcal{F}(S)\big] = X(S), \text{ almost surely.}$$

We now use the martingale property and the optional stopping theorem to prove Wald's lemmas for Brownian motion. These results identify the first and second moments of the value of Brownian motion at well-behaved stopping times.

**Theorem 4.2** (Wald's lemma)**.** *Let $\{B(t) \colon t \geqslant 0\}$ be a standard Brownian motion, and $T$ be a stopping time such that either*

**(i)** $\mathbb{E}[T] < \infty$*, or*

**(ii)** $\big\{B(t \wedge T) \colon t \geqslant 0\big\}$ *is dominated by an integrable random variable.*

*Then we have $\mathbb{E}[B(T)] = 0$.*

**Proof.** We first show that a stopping time satisfying condition (i), also satisfies condition (ii). So suppose $\mathbb{E}[T] < \infty$, and define

$$M_k = \max_{0 \leqslant t \leqslant 1} |B(t+k) - B(k)| \quad \text{and } M = \sum_{k=1}^{\lceil T \rceil} M_k.$$

Then

$$\begin{aligned}\mathbb{E}[M] = \mathbb{E}\Big[\sum_{k=1}^{\lceil T \rceil} M_k\Big] &= \sum_{k=1}^{\infty} \mathbb{E}\big[1\{T > k-1\}\, M_k\big] = \sum_{k=1}^{\infty} \mathbb{P}\{T > k-1\}\, \mathbb{E}[M_k] \\ &= \mathbb{E}[M_0]\, \mathbb{E}[T+1] < \infty\,,\end{aligned}$$

where, using Fubini's theorem and Lemma 2.8,

$$\mathbb{E}[M_0] = \int_0^\infty \mathbb{P}\big\{ \max_{0 \leqslant t \leqslant 1} |B(t)| > x \big\}\, dx \leqslant 1 + \int_1^\infty \tfrac{2\sqrt{2}}{x\sqrt{\pi}} \exp\big\{ -\tfrac{x^2}{2} \big\}\, dx < \infty\,.$$

Now note that $|B(t \wedge T)| \leqslant M$, so that (ii) holds. It remains to observe that under condition (ii) we can apply the optional stopping theorem with $S = 0$, which yields that $\mathbb{E}[B(T)] = 0$. ∎

**Corollary 4.3.** *Let $S \leqslant T$ be stopping times and $\mathbb{E}[T] < \infty$. Then*

$$\mathbb{E}\big[(B(T))^2\big] = \mathbb{E}\big[(B(S))^2\big] + \mathbb{E}\big[\big(B(T) - B(S)\big)^2\big].$$

**Proof.** The tower property of conditional expectation gives

$$\mathbb{E}\big[\big(B(T)\big)^2\big] = \mathbb{E}\big[\big(B(S)\big)^2\big] + 2\mathbb{E}\Big[B(S)\mathbb{E}\big[B(T) - B(S) \mid \mathcal{F}(S)\big]\Big] + \mathbb{E}\big[\big(B(T) - B(S)\big)^2\big].$$

As $\mathbb{E}[T] < \infty$ implies that $\{B(t \wedge T) \colon t \geqslant 0\}$ is dominated by an integrable random variable, the optional stopping theorem implies that the conditional expectation on the right hand side and the middle term hence vanishes. ■

To find the second moment of $B(T)$ and thus prove Wald's second lemma, we identify a further martingale derived from Brownian motion.

**Lemma 4.4.** *Suppose $\{B(t) \colon t \geqslant 0\}$ is a Brownian motion. Then the process*

$$\big\{B(t)^2 - t \colon t \geqslant 0\big\}$$

*is a martingale.*

**Proof.** The process is adapted to $(\mathcal{F}^+(s) \colon s \geqslant 0)$ and

$$\begin{aligned}
&\mathbb{E}\big[B(t)^2 - t \big| \mathcal{F}^+(s)\big] \\
&\quad = \mathbb{E}\big[\big(B(t) - B(s)\big)^2 \,\big|\, \mathcal{F}^+(s)\big] + 2\,\mathbb{E}\big[B(t)B(s) \,\big|\, \mathcal{F}^+(s)\big] - B(s)^2 - t \\
&\quad = (t - s) + 2B(s)^2 - B(s)^2 - t = B(s)^2 - s\,,
\end{aligned}$$

which completes the proof. ■

**Theorem 4.5** (Wald's second lemma)**.** *Let $T$ be a stopping time for standard Brownian motion such that $\mathbb{E}[T] < \infty$. Then*

$$\mathbb{E}\big[B(T)^2\big] = \mathbb{E}[T].$$

**Proof.** Look at the martingale $\{B(t)^2 - t \colon t \geqslant 0\}$ and define stopping times

$$T_n = \inf\{t \geqslant 0 \colon |B(t)| = n\}$$

so that $\{B(t \wedge T \wedge T_n)^2 - t \wedge T \wedge T_n \colon t \geqslant 0\}$ is dominated by the integrable random variable $n^2 + T$. By the optional stopping theorem we get $\mathbb{E}[B(T \wedge T_n)^2] = \mathbb{E}[T \wedge T_n]$. By Corollary 4.3 we have $\mathbb{E}[B(T)^2] \geqslant \mathbb{E}[B(T \wedge T_n)^2]$. Hence, by monotone convergence,

$$\mathbb{E}\big[B(T)^2\big] \geqslant \lim_{n \to \infty} \mathbb{E}\big[B(T \wedge T_n)^2\big] = \lim_{n \to \infty} \mathbb{E}\big[T \wedge T_n\big] = \mathbb{E}[T]\,.$$

Conversely, now using Fatou's lemma in the first step,

$$\mathbb{E}\big[B(T)^2\big] \leqslant \liminf_{n\to\infty} \mathbb{E}\big[B(T\wedge T_n)^2\big] = \liminf_{n\to\infty} \mathbb{E}\big[T\wedge T_n\big] \leqslant \mathbb{E}[T]\,. \qquad \blacksquare$$

Wald's lemmas suffice to obtain exit probabilities and expected exit times for Brownian motion.

**Theorem 4.6.** *Let $a < 0 < b$ and, for a standard Brownian motion $\{B(t) : t \geqslant 0\}$, define $T = \min\{t \geqslant 0 \colon B(t) \in \{a, b\}\}$. Then*

- $\mathbb{P}\{B(T) = a\} = \dfrac{b}{|a| + b}$ *and* $\mathbb{P}\{B(T) = b\} = \dfrac{|a|}{|a| + b}$.
- $\mathbb{E}[T] = |a|b$.

**Proof.** Let $T = \tau(\{a, b\})$ be the first exit time from the interval $[a, b]$. This stopping time satisfies the condition of the optional stopping theorem, as $|B(t \wedge T)| \leqslant |a| \vee b$. Hence, by Wald's first lemma,

$$0 = \mathbb{E}[B(T)] = a\mathbb{P}\{B(T) = a\} + b\mathbb{P}\{B(T) = b\}.$$

Together with the easy equation $\mathbb{P}\{B(T) = a\} + \mathbb{P}\{B(T) = b\} = 1$ one can solve this, and obtain $\mathbb{P}\{B(T) = a\} = b/(|a| + b)$, and $\mathbb{P}\{B(T) = b\} = |a|/(|a| + b)$. To use Wald's second lemma, we check that $\mathbb{E}[T] < \infty$. For this purpose note

$$\mathbb{E}[T] = \int_0^\infty \mathbb{P}\{T > t\}\, dt = \int_0^\infty \mathbb{P}\{B(s) \in (a, b) \text{ for all } s \in [0, t]\}\, dt,$$

and that, for $t \geqslant k \in \mathbb{N}$ the integrand is bounded by the $k^{\text{th}}$ power of

$$\max_{x\in(a,b)} \mathbb{P}_x\{B(1) \in (a, b)\},$$

i.e. decreases exponentially. Hence the integral is finite.

Now, by Wald's second lemma and the exit probabilities, we obtain

$$\mathbb{E}[T] = \mathbb{E}[B(T)^2] = \frac{a^2 b}{|a| + b} + \frac{b^2 |a|}{|a| + b} = |a|b. \qquad \blacksquare$$

We now discuss a strengthening of Theorem 4.2, which works with a weaker (essentially optimal) moment condition. The lemma is a weak version of the famous Burkholder-Davis-Gundy inequality.

**Theorem 4.7.** *Let $\{B(t) \colon t \geqslant 0\}$ be a standard Brownian motion and $T$ a stopping time with $\mathbb{E}[T^{1/2}] < \infty$. Then $\mathbb{E}[B(T)] = 0$.*

**Proof.** Let $\{M(t)\colon t \geqslant 0\}$ be the maximum process of $\{B(t)\colon t \geqslant 0\}$ and $T$ a stopping time with $\mathbb{E}[T^{1/2}] < \infty$. Let $\tau = \lceil \log_4 T \rceil$, so that $B(t \wedge T) \leqslant M(4^\tau)$. In order to get $\mathbb{E}[B(T)] = 0$ from the optional stopping theorem it suffices to show that the majorant is integrable, i.e. that

$$\mathbb{E}M(4^\tau) < \infty.$$

Define a discrete time stochastic process $\{X_k\colon k \in \mathbb{N}\}$ by $X_k = M(4^k) - 2^{k+1}$, and observe that $\tau$ is a stopping time with respect to the filtration $(\mathcal{F}^+(4^k)\colon k \in \mathbb{N})$. Moreover, the process $\{X_k\colon k \in \mathbb{N}\}$ is a supermartingale. Indeed,

$$\mathbb{E}\big[X_k \,\big|\, \mathcal{F}_{k-1}\big] \leqslant M(4^{k-1}) + \mathbb{E}\Big[\max_{0 \leqslant t \leqslant 4^k - 4^{k-1}} B(t)\Big] - 2^{k+1}\,,$$

and the supermartingale property follows as

$$\mathbb{E}\Big[\max_{0 \leqslant t \leqslant 4^k - 4^{k-1}} B(t)\Big] = \sqrt{4^k - 4^{k-1}}\,\mathbb{E}\Big[\max_{0 \leqslant t \leqslant 1} B(t)\Big] \leqslant 2^k\,,$$

using that, by the reflection principle, Theorem 2.8, and the Cauchy–Schwarz inequality,

$$\mathbb{E}\big[\max_{0 \leqslant t \leqslant 1} B(t)\big] = \mathbb{E}|B(1)| \leqslant \big(\mathbb{E}[B(1)^2]\big)^{\frac{1}{2}} = 1.$$

Now let $t = 4^\ell$ and use the supermartingale property for $\tau \wedge \ell$ to get

$$\mathbb{E}\big[M(4^\tau \wedge t)\big] = \mathbb{E}\big[X_{\tau\wedge\ell}\big] + \mathbb{E}\big[2^{\tau\wedge\ell+1}\big] \leqslant \mathbb{E}[X_0] + 2\,\mathbb{E}\big[2^\tau\big]\,.$$

Note that $X_0 = M(1) - 2$, which has finite expectation and, by our assumption on the moments of $T$, we have $\mathbb{E}[2^\tau] < \infty$. Thus, by monotone convergence,

$$\mathbb{E}\big[M(4^\tau)\big] = \lim_{t \uparrow \infty} \big[M(4^\tau \wedge t)\big] < \infty\,,$$

which completes the proof of the theorem. ∎

# 5 How often does Brownian motion visit zero?

The set $\{t \geqslant 0\colon B_t = 0\}$ of zeros of a Brownian motion is almost surely

- uncountable (as it is a perfect set: closed with no isolated points),
- and has zero Lebesgue measure (easily seen from Fubini's theorem).

To measure its size on a crude scale we use the notion of *Hausdorff dimension*, which we now introduce. For every $\alpha > 0$ the $\alpha$-**value** of a sequence $E_1, E_2, \ldots$ of sets in a metric space is (with $|E_i|$ denoting the diameter of $E_i$)

$$\sum_{i=1}^{\infty} |E_i|^\alpha\,.$$

For every $\alpha > 0$ denote

$$\mathcal{H}^\alpha_\delta(E) := \inf\Big\{ \sum_{i=1}^\infty |E_i|^\alpha \colon E_1, E_2, \ldots \text{ is a covering of } E \text{ with } |E_i| \leqslant \delta \Big\}.$$

The $\alpha$**-Hausdorff measure** of $E$ is defined as

$$\mathcal{H}^\alpha(E) = \lim_{\delta \downarrow 0} \mathcal{H}^\alpha_\delta(E),$$

informally speaking the $\alpha$-value of the most efficient covering by small sets. If $0 < \alpha < \beta$, and $\mathcal{H}^\alpha(E) < \infty$, then $\mathcal{H}^\beta(E) = 0$. If $0 < \alpha < \beta$, and $\mathcal{H}^\beta(E) > 0$, then $\mathcal{H}^\alpha(E) = \infty$. Thus we can define

$$\dim E = \inf\Big\{\alpha > 0 \colon \mathcal{H}^\alpha(E) < \infty\Big\} = \sup\Big\{\alpha > 0 \colon \mathcal{H}^\alpha(E) > 0\Big\},$$

the **Hausdorff dimension** of the set $E$.

**Theorem 5.1.** *For every $\varepsilon > 0$ we have, almost surely,*

$$\mathcal{H}^{1/2}\big\{\varepsilon < t < 1 \colon B(t) = 0\big\} < \infty.$$

We use a covering consisting of intervals. Define the collection $\mathfrak{D}_k$ of intervals $[j2^{-k}, (j+1)2^{-k})$ for $j = 0, \ldots, 2^k - 1$, and let $Z(I) = 1$ if there exists $t \in I$ with $B(t) = 0$, and $Z(I) = 0$ otherwise. To estimate the dimension of the zero set we need an estimate for the probability that $Z(I) = 1$, i.e. for the probability that a given interval contains a zero of Brownian motion.

**Lemma 5.2.** *There is an absolute constant $C$ such that, for any $a, \delta > 0$,*

$$\mathbb{P}\big\{ \text{ there exists } t \in [a, a+\delta] \text{ with } B(t) = 0\big\} \leqslant C\sqrt{\tfrac{\delta}{a+\delta}}.$$

**Proof.** Consider the event $A = \{|B(a+\delta)| \leqslant \sqrt{\delta}\}$. By the scaling property of Brownian motion, we can give the upper bound

$$\mathbb{P}(A) \leqslant \mathbb{P}\left\{ |B(1)| \leqslant \sqrt{\tfrac{\delta}{a+\delta}} \right\} \leqslant 2\sqrt{\tfrac{\delta}{a+\delta}}.$$

Applying the strong Markov property at $T = \inf\{t \geqslant a \colon B(t) = 0\}$, we have

$$\begin{aligned} \mathbb{P}(A) &\geqslant \mathbb{P}\big(A \cap \{0 \in B[a, a+\delta]\}\big) \\ &\geqslant \mathbb{P}\{T \leqslant a+\delta\} \min_{a \leqslant t \leqslant a+\delta} \mathbb{P}\{|B(a+\delta)| \leqslant \sqrt{\delta} \,|\, B(t) = 0\}. \end{aligned}$$

Clearly the minimum is achieved at $t = a$ and, using the scaling property of Brownian motion, we have $\mathbb{P}\{|B(a+\delta)| \leqslant \sqrt{\delta} \,|\, B(a) = 0\} = \mathbb{P}\{|B(1)| \leqslant 1\} =: c > 0$. Hence,

$$\mathbb{P}\{T \leqslant a+\delta\} \leqslant \tfrac{2}{c}\sqrt{\tfrac{\delta}{a+\delta}}.$$

■

We have thus shown that, for any $\varepsilon > 0$ and sufficiently large integer $k$, we have

$$\mathbb{E}[Z(I)] \leqslant c_1\, 2^{-k/2}, \qquad \text{for all } I \in \mathfrak{D}_k \text{ with } I \cap (\varepsilon, 1] \neq \emptyset\,,$$

for some constant $c_1 > 0$. Hence the covering of the set $\{t \in (\varepsilon, 1]\colon B(t) = 0\}$ by all $I \in \mathfrak{D}_k$ with $I \cap (\varepsilon, 1] \neq \emptyset$ and $Z(I) = 1$ has an expected $\frac{1}{2}$-value of

$$\mathbb{E}\Big[\sum_{\substack{I \in \mathfrak{D}_k \\ I \cap (\varepsilon,1] \neq \emptyset}} Z(I)\, 2^{-k/2}\Big] = \sum_{\substack{I \in \mathfrak{D}_k \\ I \cap (\varepsilon,1] \neq \emptyset}} \mathbb{E}[Z(I)]\, 2^{-k/2} \leqslant c_1\, 2^k\, 2^{-k/2}\, 2^{-k/2} = c_1.$$

We thus get, from Fatou's lemma,

$$\mathbb{E}\Big[\liminf_{k \to \infty} \sum_{\substack{I \in \mathfrak{D}_k \\ I \cap (\varepsilon,1] \neq \emptyset}} Z(I)\, 2^{-k/2}\Big] \leqslant \liminf_{k \to \infty} \mathbb{E}\Big[\sum_{\substack{I \in \mathfrak{D}_k \\ I \cap (\varepsilon,1] \neq \emptyset}} Z(I)\, 2^{-k/2}\Big] \leqslant c_1.$$

Hence the liminf is almost surely finite, which means that there exists a family of coverings with maximal diameter going to zero and bounded $\frac{1}{2}$-value. This implies the statement.

With some (considerable) extra effort it can even be shown that

$$\mathcal{H}^{1/2}\{0 < t < 1\colon B(t) = 0\} = 0.$$

But even from the result for fixed $\varepsilon > 0$ we can infer that

$$\dim\ \big\{0 < t \leqslant 1\colon B(t) = 0\big\} \leqslant \frac{1}{2} \text{ almost surely,}$$

and the stability of dimension under countable unions gives the same bound for the unbounded zero set.

From the definition of the Hausdorff dimension it is plausible that in many cases it is relatively easy to give an upper bound on the dimension: just find an efficient cover of the set and find an upper bound to its $\alpha$-value. However it looks more difficult to give lower bounds, as we must obtain a lower bound on $\alpha$-values of *all* covers of the set. The energy method is a way around this problem, which is based on the existence of a nonzero measure on the set. The basic idea is that, if this measure distributes a positive amount of mass on a set $E$ in such a manner that its local concentration is bounded from above, then the set must be large in a suitable sense. Suppose $\mu$ is a measure on a metric space $(E, \rho)$ and $\alpha \geqslant 0$. The $\alpha$**-potential** of a point $x \in E$ with respect to $\mu$ is defined as

$$\phi_\alpha(x) = \int \frac{d\mu(y)}{\rho(x,y)^\alpha}\,.$$

In the case $E = \mathbb{R}^3$ and $\alpha = 1$, this is the Newton gravitational potential of the mass $\mu$. The $\alpha$**-energy** of $\mu$ is

$$I_\alpha(\mu) = \int \phi_\alpha(x)\, d\mu(x) = \iint \frac{d\mu(x)\, d\mu(y)}{\rho(x,y)^\alpha}\,.$$

Measures with $I_\alpha(\mu) < \infty$ spread the mass so that at each place the concentration is sufficiently small to overcome the singularity of the integrand. This is only possible on sets which are large in a suitable sense.

**Theorem 5.3** (Energy method)**.** *Let $\alpha \geqslant 0$ and $\mu$ be a nonzero measure on a metric space $E$. Then, for every $\varepsilon > 0$, we have*

$$\mathcal{H}^\alpha_\varepsilon(E) \geqslant \frac{\mu(E)^2}{\iint_{\rho(x,y)<\varepsilon} \frac{d\mu(x)\,d\mu(y)}{\rho(x,y)^\alpha}}\,.$$

*Hence, if $I_\alpha(\mu) < \infty$ then $\mathcal{H}^\alpha(E) = \infty$ and, in particular,* $\dim E \geqslant \alpha$.

**Proof.** (Due to O. Schramm) If $\{A_n : n = 1, 2, \ldots\}$ is any pairwise disjoint covering of $E$ consisting of Borel sets of diameter $< \varepsilon$, then

$$\iint_{\rho(x,y)<\varepsilon} \frac{d\mu(x)\,d\mu(y)}{\rho(x,y)^\alpha} \geqslant \sum_{n=1}^\infty \iint_{A_n\times A_n} \frac{d\mu(x)\,d\mu(y)}{\rho(x,y)^\alpha} \geqslant \sum_{n=1}^\infty \frac{\mu(A_n)^2}{|A_n|^\alpha},$$

and moreover,

$$\mu(E) \leqslant \sum_{n=1}^\infty \mu(A_n) = \sum_{n=1}^\infty |A_n|^{\frac{\alpha}{2}} \frac{\mu(A_n)}{|A_n|^{\frac{\alpha}{2}}}$$

Given $\delta > 0$ choose a covering as above such that additionally

$$\sum_{n=1}^\infty |A_n|^\alpha \leqslant \mathcal{H}^\alpha_\varepsilon(E) + \delta.$$

Using now the Cauchy–Schwarz inequality, we get

$$\mu(E)^2 \leqslant \sum_{n=1}^\infty |A_n|^\alpha \sum_{n=1}^\infty \frac{\mu(A_n)^2}{|A_n|^\alpha} \leqslant \Big(\mathcal{H}^\alpha_\varepsilon(E) + \delta\Big) \iint_{\rho(x,y)<\varepsilon} \frac{d\mu(x)\,d\mu(y)}{\rho(x,y)^\alpha}.$$

Letting $\delta \downarrow 0$ gives the stated inequality. Further, letting $\varepsilon \downarrow 0$, if $I_\alpha(\mu) < \infty$ the integral converges to zero, so that $\mathcal{H}^\alpha_\varepsilon(E)$ diverges to infinity. ■

**Remark 5.4.** To get a lower bound on the dimension from this method it suffices to show finiteness of a single integral. In particular, in order to show for a random set $E$ that $\dim E \geqslant \alpha$ almost surely, it suffices to show that $\mathbb{E} I_\alpha(\mu) < \infty$ for a (random) measure on $E$. ◇

**Theorem 5.5** (Taylor 1955)**.**

$$\dim\ \big\{t \geqslant 0 \colon B(t) = 0\big\} = \frac{1}{2} \textit{ almost surely.}$$

**Proof.** Recall that we already know the upper bound. We now look at the lower bound using the energy method, for which we require a suitable positive finite measure on the zero set. We use Lévy's theorem, which implies that the random variables

$$\dim\big\{t>0\colon |B(t)|=0\big\} \text{ and } \dim\big\{t>0\colon M(t)=B(t)\big\}$$

have the same law. On the set on the right we have a measure $\mu$ given by

$$\int f\,d\mu = \int_0^\infty f(T_a)\,da,$$

where $T_a$ is the first hitting time of level $a>0$. It thus suffices to show, for any fixed $l>0$ and $\alpha<\frac{1}{2}$, the finiteness of the integral

$$\begin{aligned}\mathbb{E}\int_0^{T_l}\int_0^{T_l}\frac{d\mu(s)\,d\mu(t)}{|s-t|^\alpha} &= \int_0^l da\int_0^l db\,\mathbb{E}|T_a-T_b|^{-\alpha}\\ &= 2\int_0^l da\int_0^a db\int_0^\infty ds\,s^{-\alpha}\frac{1}{\sqrt{2\pi s^3}}\,(a-b)\,\exp\big(-\tfrac{(a-b)^2}{2s}\big)\\ &= \sqrt{\frac{2}{\pi}}\int_0^\infty ds\,s^{-\alpha-\frac{1}{2}}\int_0^l da\,\big(1-e^{-a^2/2s}\big),\end{aligned}$$

where we used the transition density of the $\frac{1}{2}$-stable subordinator. The inner integral converges to a constant when $s\downarrow 0$ and decays like $\mathrm{O}(1/s)$ when $s\uparrow\infty$. Hence the expression is finite when $(0\leqslant)\alpha<\frac{1}{2}$. ∎

We define the $\alpha$**-capacity** of a metric space $(E,\rho)$ as

$$\mathrm{Cap}_\alpha(E) := \sup\Big\{I_\alpha(\mu)^{-1}\colon \mu \text{ a probability measure on } E\Big\}.$$

In the case of the Euclidean space $E=\mathbb{R}^d$ with $d\geqslant 3$ and $\alpha=d-2$ the $\alpha$-capacity is also known as the **Newtonian capacity**. Theorem 5.3 states that a set of positive $\alpha$-capacity has dimension at least $\alpha$. The famous Frostman's lemma states that in Euclidean spaces this method is sharp, i.e., for any closed (or, more generally, analytic) set $A\subset\mathbb{R}^d$,

$$\dim A = \sup\big\{\alpha:\ \mathrm{Cap}_\alpha(A)>0\big\}.$$

We omit the (nontrivial) proof.

# 6 Donsker's invariance principle

Given a random variable $X$ can we find a stopping time $T$ with $\mathbb{E}[T]<\infty$, such that $B(T)$ has the law of $X$? This is called the *Skorokhod embedding problem.* By Wald's lemmas we have $\mathbb{E}[B(T)]=0$ and $\mathbb{E}[B(T)^2]=\mathbb{E}[T]<\infty$, so that the Skorokhod embedding problem can only be solved for random variables $X$ with mean zero and finite second moment. However, under these assumptions there are several solutions.

**Theorem 6.1** (Azéma–Yor embedding theorem)**.** *Let $X$ be a real valued random variable with $E[X] = 0$ and $E[X^2] < \infty$. Define*

$$\Psi(x) = \begin{cases} E\big[X \,\big|\, X \geqslant x\big] & \text{if } P\{X \geqslant x\} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

*and a stopping time $\tau$ by*

$$\tau = \inf\{t \geqslant 0 \colon M_t \geqslant \Psi(B_t)\}.$$

*Then $\mathbb{E}[\tau] = E[X^2]$ and $B(\tau)$ has the same law as $X$.*

We prove this for random variables taking only finitely many values, the general case follows by an approximation from this case.

**Lemma 6.2.** *Suppose the random variable $X$ with $\mathbb{E}X = 0$ takes only finitely many values $x_1 < x_2 < \cdots < x_n$. Define $y_1 < y_2 < \cdots < y_{n-1}$ by $y_i = \Psi(x_{i+1})$, and define stopping times $T_0 = 0$ and*

$$T_i = \inf\big\{t \geqslant T_{i-1} \colon B(t) \notin (x_i, y_i)\big\} \qquad \text{for } i \leqslant n-1.$$

*Then $T = T_{n-1}$ satisfies $\mathbb{E}[T] = E[X^2]$ and $B(T)$ has the same law as $X$.*

**Proof.** First observe that $y_i \geqslant x_{i+1}$ and equality holds if and only if $i = n-1$. We have $\mathbb{E}[T_{n-1}] < \infty$, by Theorem 4.6, and $\mathbb{E}[T_{n-1}] = \mathbb{E}[B(T_{n-1})^2]$, from Theorem 4.5. For $i = 1, \ldots, n-1$ define random variables

$$Y_i = \begin{cases} E[X \,|\, X \geqslant x_{i+1}] & \text{if } X \geqslant x_{i+1}, \\ X & \text{if } X \leqslant x_i. \end{cases}$$

Note that $Y_1$ has expectation zero and takes on the two values $x_1, y_1$. For $i \geqslant 2$, given $Y_{i-1} = y_{i-1}$, the random variable $Y_i$ takes the values $x_i, y_i$ and has expectation $y_{i-1}$. Given $Y_{i-1} = x_j$, $j \leqslant i-1$ we have $Y_i = x_j$. Note that $Y_{n-1} = X$. We now argue that

$$(B(T_1), \ldots, B(T_{n-1})) \stackrel{d}{=} (Y_1, \ldots, Y_{n-1}).$$

Clearly, $B(T_1)$ can take only the values $x_1, y_1$ and has expectation zero, hence the law of $B(T_1)$ agrees with the law of $Y_1$. For $i \geqslant 2$, given $B(T_{i-1}) = y_{i-1}$, the random variable $B(T_i)$ takes the values $x_i, y_i$ and has expectation $y_{i-1}$. Given $B(T_{i-1}) = x_j$ where $j \leqslant i-1$, we have $B(T_i) = x_j$. Hence the two tuples have the same law and, in particular, $B(T_{n-1})$ has the same law as $X$. ■

It remains to show that the stopping time we have constructed in Lemma 6.2 agrees with the stopping time $\tau$ in the Azéma–Yor embedding theorem. Indeed, suppose that $B(T_{n-1}) = x_i$, and hence $\Psi(B(T_{n-1})) = y_{i-1}$. If $i \leqslant n-1$, then $i$ is minimal with the property that $B(T_i) = \cdots = B(T_{n-1})$, and thus $B(T_{i-1}) \neq B(T_i)$. Hence $M(T_{n-1}) \geqslant y_{i-1}$. If $i = n$ we also have $M(T_{n-1}) =$

$x_n \geqslant y_{i-1}$, which implies in any case that $\tau \leqslant T$. Conversely, if $T_{i-1} \leqslant t < T_i$ then $B(t) \in (x_i, y_i)$ and this implies $M(t) < y_i \leqslant \Psi(B(t))$. Hence $\tau \geqslant T$, and altogether we have seen that $T = \tau$.

The main application of Skorokhod embedding is to *Donsker's invariance principle*, or *functional central limit theorem*, which offers a direct relation between random walks and Brownian motion. Let $\{X_n : n \geqslant 0\}$ be a sequence of independent and identically distributed random variables and assume that they are normalised, so that $\mathbb{E}[X_n] = 0$ and $\mathrm{Var}(X_n) = 1$. We look at the *random walk* generated by the sequence

$$S_n = \sum_{k=1}^{n} X_k \,,$$

and interpolate linearly between the integer points, i.e.

$$S(t) = S_{[t]} + (t - [t])(S_{[t]+1} - S_{[t]}) \,.$$

This defines a random function $S \in \mathbf{C}[0,\infty)$. We now define a sequence $\{S_n^* : n \geqslant 1\}$ of random functions in $\mathbf{C}[0,1]$ by

$$S_n^*(t) = \frac{S(nt)}{\sqrt{n}} \text{ for all } t \in [0,1].$$

**Theorem 6.3** (Donsker's invariance principle)**.** *On the space* $\mathbf{C}[0,1]$ *of continuous functions on the unit interval with the metric induced by the sup-norm, the sequence* $\{S_n^* : n \geqslant 1\}$ *converges in distribution to a standard Brownian motion* $\{B(t) : t \in [0,1]\}$*.*

By the Skorohod embedding theorem there exists a sequence of stopping times

$$0 = T_0 \leqslant T_1 \leqslant T_2 \leqslant T_3 \leqslant \, \ldots$$

with respect to the Brownian motion, such that the sequence $\{B(T_n) : n \geqslant 0\}$ has the distribution of the random walk with increments given by the law of $X$. Moreover, using the finiteness of the expectations it is not too hard to show that the sequence of functions $\{S_n^* : n \geqslant 0\}$ constructed from this random walk satisfies

$$\lim_{n\to\infty} \mathbb{P}\Big\{ \sup_{0 \leqslant t \leqslant 1} \Big| \frac{B(nt)}{\sqrt{n}} - S_n^*(t) \Big| > \varepsilon \Big\} = 0 \,.$$

To complete the proof of Donsker's invariance principle recall from the scaling property of Brownian motion that the random functions $\{W_n(t) : 0 \leqslant t \leqslant 1\}$ given by $W_n(t) = B(nt)/\sqrt{n}$ are standard Brownian motions. Suppose that $K \subset \mathbf{C}[0,1]$ is closed and define

$$K[\varepsilon] = \{ f \in \mathbf{C}[0,1] : \|f - g\|_{\sup} \leqslant \varepsilon \text{ for some } g \in K \}.$$

Then $\mathbb{P}\{S_n^* \in K\} \leqslant \mathbb{P}\{W_n \in K[\varepsilon]\} + \mathbb{P}\{\|S_n^* - W_n\|_{\sup} > \varepsilon\}$. As $n \to \infty$, the second term goes to 0, whereas the first term does not depend on $n$ and is equal to $\mathbb{P}\{B \in K[\varepsilon]\}$ for a Brownian motion $B$. Letting $\varepsilon \downarrow 0$ we obtain $\limsup_{n\to\infty} \mathbb{P}\{S_n^* \in K\} \leqslant \mathbb{P}\{B \in K\}$, which implies weak convergence and completes the proof of Donsker's invariance principle.

# 7 Applications: Arcsine laws and Pitman's $2M - X$ Theorem

We now show by the example of the *arc-sine laws* how one can transfer results between Brownian motion and random walks by means of Donsker's invariance principle. The name comes from the **arcsine distribution**, which is the distribution on $(0,1)$ which has the density

$$\frac{1}{\pi\sqrt{x(1-x)}} \qquad \text{for } x \in (0,1).$$

The distribution function of an arcsine distributed random variable $X$ is therefore given by $\mathbb{P}\{X \leqslant x\} = \frac{2}{\pi}\arcsin(\sqrt{x})$.

**Theorem 7.1** (First arcsine law for Brownian motion)**.** *The random variables*

- *$L$, the last zero of Brownian motion in $[0,1]$, and*
- *$M^*$, the maximiser of Brownian motion in $[0,1]$,*

*are both arcsine distributed.*

**Proof.** Lévy's theorem shows that $M^*$, which is the last zero of the process $\{M(t) - B(t) \colon t \geqslant 0\}$ has the same law as $L$. Hence it suffices to prove the second statement. Note that

$$\begin{aligned}
\mathbb{P}\{M^* < s\} &= \mathbb{P}\big\{\max_{0\leqslant u\leqslant s} B(u) > \max_{s\leqslant v\leqslant 1} B(v)\big\} \\
&= \mathbb{P}\big\{\max_{0\leqslant u\leqslant s} B(u) - B(s) > \max_{s\leqslant v\leqslant 1} B(v) - B(s)\big\} \\
&= \mathbb{P}\big\{M_1(s) > M_2(1-s)\big\},
\end{aligned}$$

where $\{M_1(t) \colon 0 \leqslant t \leqslant s\}$ is the maximum process of the Brownian motion $\{B_1(t) \colon 0 \leqslant t \leqslant s\}$, which is given by $B_1(t) = B(s-t) - B(s)$, and $\{M_2(t) \colon 0 \leqslant t \leqslant 1\}$ is the maximum process of the independent Brownian motion $\{B_2(t) \colon 0 \leqslant t \leqslant 1-s\}$, which is given by $B_2(t) = B(s+t) - B(s)$. Since, for any fixed $t$, the random variable $M(t)$ has the same law as $|B(t)|$, we have

$$\mathbb{P}\big\{M_1(s) > M_2(1-s)\big\} = \mathbb{P}\big\{|B_1(s)| > |B_2(1-s)|\big\}.$$

Using the scaling invariance of Brownian motion we can express this in terms of a pair of two independent standard normal random variables $Z_1$ and $Z_2$ , by

$$\mathbb{P}\big\{|B_1(s)| > |B_2(1-s)|\big\} = \mathbb{P}\big\{\sqrt{s}\,|Z_1| > \sqrt{1-s}\,|Z_2|\big\} = \mathbb{P}\Big\{\frac{|Z_2|}{\sqrt{Z_1^2+Z_2^2}} < \sqrt{s}\Big\}.$$

In polar coordinates, $(Z_1, Z_2) = (R\cos\theta, R\sin\theta)$ pointwise. The random variable $\theta$ is uniformly distributed on $[0,2\pi]$ and hence the last quantity becomes

$$\begin{aligned}
\mathbb{P}\Big\{\frac{|Z_2|}{\sqrt{Z_1^2+Z_2^2}} < \sqrt{s}\Big\} &= \mathbb{P}\big\{|\sin(\theta)| < \sqrt{s}\big\} = 4\mathbb{P}\big\{\theta < \arcsin(\sqrt{s})\big\} \\
&= 4\left(\frac{\arcsin(\sqrt{s})}{2\pi}\right) = \frac{2}{\pi}\arcsin(\sqrt{s}).
\end{aligned}$$

For random walks the first arcsine law takes the form of a limit theorem, as the length of the walk tends to infinity.

**Theorem 7.2** (Arcsine law for the last sign-change). *Suppose that $\{X_k : k \geqslant 1\}$ is a sequence of independent, identically distributed random variables with $\mathbb{E}[X_1] = 0$ and $0 < \mathbb{E}[X_1^2] = \sigma^2 < \infty$. Let $\{S_n : n \geqslant 0\}$ be the associated random walk and*

$$N_n = \max\{1 \leqslant k \leqslant n \,:\, S_k S_{k-1} \leqslant 0\}$$

*the last time the random walk changes its sign before time $n$. Then, for all $x \in (0,1)$,*

$$\lim_{n\to\infty} \mathbb{P}\{N_n \leqslant xn\} = \frac{2}{\pi} \arcsin(\sqrt{x})\,.$$

**Proof.** The strategy of proof is to use Theorem 7.1, and apply Donsker's invariance principle to extend the result to random walks. As $N_n$ is unchanged under scaling of the random walk we may assume that $\sigma^2 = 1$. Define a bounded function $g$ on $\mathbf{C}[0,1]$ by

$$g(f) = \max\{t \leqslant 1 : f(t) = 0\}.$$

It is clear that $g(S_n^*)$ differs from $N_n/n$ by a term, which is bounded by $1/n$ and therefore vanishes asymptotically. Hence Donsker's invariance principle would imply convergence of $N_n/n$ in distribution to $g(B) = \sup\{t \leqslant 1 : B(t) = 0\}$ — if $g$ was continuous. $g$ is *not* continuous, but we claim that $g$ is continuous on the set $\mathcal{C}$ of all $f \in \mathbf{C}[0,1]$ such that $f$ takes positive and negative values in every neighbourhood of every zero and $f(1) \neq 0$. As Brownian motion is almost surely in $\mathcal{C}$, we get from the Portmanteau theorem and Donsker's invariance principle, that, for every continuous bounded $h : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\Big[h\Big(\frac{N_n}{n}\Big)\Big] &= \lim_{n\to\infty} \mathbb{E}\big[h \circ g(S_n^*)\big] = \mathbb{E}\big[h \circ g(B)\big] \\ &= \mathbb{E}\big[h(\sup\{t \leqslant 1 : B(t) = 0\})\big],\end{aligned}$$

which completes the proof subject to the claim. To see that $g$ is continuous on $\mathcal{C}$, let $\varepsilon > 0$ be given and $f \in \mathcal{C}$. Let

$$\delta_0 = \min_{t \in [g(f)+\varepsilon, 1]} |f(t)|\,,$$

and choose $\delta_1$ such that $(-\delta_1, \delta_1) \subset f(g(f) - \varepsilon, g(f) + \varepsilon)$. Let $0 < \delta < \delta_0 \wedge \delta_1$. If now $\|h - f\|_\infty < \delta$, then $h$ has no zero in $(g(f) + \varepsilon, 1]$, but has a zero in $(g(f) - \varepsilon, g(f) + \varepsilon)$, because there are $s, t \in (g(f) - \varepsilon, g(f) + \varepsilon)$ with $h(t) < 0$ and $h(s) > 0$. Thus $|g(h) - g(f)| < \varepsilon$. This shows that $g$ is continuous on $\mathcal{C}$. ■

There is a second arcsine law for Brownian motion, which describes the law of the random variable $\mathcal{L}\big\{t \in [0,1] : B(t) > 0\big\}$, the time spent by Brownian motion above the $x$-axis. This statement is much harder to derive directly for Brownian motion. At this stage we can use random walks to derive the result for Brownian motion.

**Theorem 7.3** (Second arcsine law for Brownian motion)**.** *If* $\{B_t : t \geqslant 0\}$ *is a standard Brownian motion, then* $\mathcal{L}\{t \in [0,1] : B_t > 0\}$ *is arcsine distributed.*

The idea is to prove a direct relationship between the first maximum and the number of positive terms for a *simple* random walk by a combinatorial argument, and then transfer this to Brownian motion using Donsker's invariance principle.

**Lemma 7.4.** *Let* $\{S_k : k = 1, \ldots, n\}$ *be a simple, symmetric random walk on the integers. Then*

$$\#\big\{k \in \{1, \ldots, n\} : S_k > 0\big\} \stackrel{\mathrm{d}}{=} \min\big\{k \in \{0, \ldots, n\} \,:\, S_k = \max_{0 \leqslant j \leqslant n} S_j\big\}. \tag{5}$$

For the proof of the lemma let $X_k = S_k - S_{k-1}$ for each $k \in \{1, \ldots, n\}$, with $S_0 := 0$, and rearrange the tuple $(X_1, \ldots, X_n)$ by

- placing first in *decreasing* order of $k$ the terms $X_k$ for which $S_k > 0$,
- and then in *increasing* order of $k$ the $X_k$ for which $S_k \leqslant 0$.

Denote the new tuple by $(Y_1, \ldots, Y_n) := T_n(X_1, \ldots, X_n)$. One first shows by induction that $T_n$ is a bijection for every $n \in \mathbb{N}$. Then one defines $\{S_k(Y) : k = 1, \ldots, n\}$ by $S_k(Y) = \sum_{j=1}^{k} Y_j$ and checks by induction on $n$ that

$$\begin{aligned}\#\big\{k \in \{1, \ldots, n\} &: S_k(X) > 0\big\} \\ &= \min\{k \in \{0, \ldots, n\} : S_k(Y) = \max_{0 \leqslant j \leqslant n} S_j(Y)\big\}.\end{aligned}$$

To prove Theorem 7.3 we look at the right hand side of the equation (5), which divided by $n$ can be written as $g(S_n^*)$ for the function $g : \mathbf{C}[0,1] \to [0,1]$ defined by

$$g(f) = \inf\big\{t \in [0,1] : f(t) = \sup_{s \in [0,1]} f(s)\big\}\,.$$

The function $g$ is continuous in every $f \in \mathbf{C}[0,1]$ which has a unique maximum, hence almost everywhere with respect to the distribution of Brownian motion. By Donsker's invariance principle and the Portmanteau theorem the right hand side in (5) divided by $n$ converges to the distribution of $g(B)$, which by Theorem 7.1 is the arcsine distribution.

Similarly, the left hand side of (5) divided by $n$ can be approximated in probability by $h(S_n^*)$ for the function $h : \mathbf{C}[0,1] \to [0,1]$ defined by

$$h(f) = \mathcal{L}\{t \in [0,1] : f(t) > 0\big\}\,.$$

It is not hard to see that the function $h$ is continuous in every $f \in \mathbf{C}[0,1]$ with $\mathcal{L}\{t \in [0,1] : f(t) = 0\big\} = 0$, a property which Brownian motion has almost surely. Hence, again by Donsker's invariance principle and the Portmanteau theorem the left hand side in (5) divided by $n$ converges to the distribution of $h(B) = \mathcal{L}\{t \in [0,1] : B(t) > 0\}$, and this completes the argument.

Pitman's $2M - X$ theorem describes an interesting relationship between the process $\{2M(t) - B(t) \colon t \geqslant 0\}$ and the 3-dimensional Bessel process, which, loosely speaking, can be considered as a Brownian motion conditioned to avoid zero. We will obtain this result from a random walk analogue, using Donsker's invariance principle to pass to the Brownian motion case.

We start by discussing simple random walk conditioned to avoid zero. Consider a simple random walk on $\{0, 1, 2, \dots, n\}$ conditioned to reach $n$ before 0. This conditioned process is a Markov chain with the following transition probabilities: $\widehat{p}(0,1) = 1$ and for $1 \leqslant k < n$,

$$\widehat{p}(k, k+1) = (k+1)/2k \quad ; \quad \widehat{p}(k, k-1) = (k-1)/2k \,. \tag{6}$$

Taking $n \to \infty$, this leads us to *define* the **simple random walk on** $\mathbb{N} = \{1, 2, \dots\}$ **conditioned to avoid zero** (forever) as a Markov chain on $\mathbb{N}$ with transition probabilities as in (6) for all $k \geqslant 1$.

**Lemma 7.5.** *Let $\{S(j) \colon j = 0, 1, \dots\}$ be a simple random walk on $\mathbb{Z}$ and let $\{\tilde{\rho}(j) \colon j = 0, 1, \dots\}$ be a simple random walk on $\mathbb{N}$ conditioned to avoid zero. Then for $\ell \geqslant 1$ and any sequence $(x_0, \dots, x_\ell)$ of positive integers, we have*

$$\mathbb{P}\big\{\tilde{\rho}(1) = x_1, \dots, \tilde{\rho}(\ell) = x_\ell \,\big|\, \tilde{\rho}(0) = x_0\big\} = \frac{x_\ell}{x_0}\, \mathbb{P}\big\{S(1) = x_1, \dots, S(\ell) = x_\ell \,\big|\, S(0) = x_0\big\} \,.$$

**Proof.** We prove the result by induction on $\ell$. The case $\ell = 1$ is just (6). Assume the lemma holds for $\ell - 1$ and let $(x_0, \dots, x_\ell)$ be a sequence of positive integers such that $|x_j - x_{j-1}| = 1$ for $j = 1, \dots, \ell$. Clearly, the probability on the right hand side of the equation is just $2^{-\ell}$. Moreover, using the induction hypothesis and the Markov property,

$$\begin{aligned} \mathbb{P}\big\{\tilde{\rho}(1) = x_1, \dots, \tilde{\rho}(\ell) = x_\ell \,\big|\, \tilde{\rho}(0) = x_0\big\} &= \frac{x_{\ell-1}}{x_0}\, 2^{1-\ell}\, \mathbb{P}\big\{\tilde{\rho}(\ell) = x_\ell \,\big|\, \tilde{\rho}(\ell-1) = x_{\ell-1}\big\} \\ &= \frac{x_{\ell-1}}{x_0}\, 2^{1-\ell} \frac{x_\ell}{2x_{\ell-1}} \; = \; \frac{x_\ell}{x_0}\, 2^{-\ell}, \end{aligned}$$

as required to complete the proof. ■

Define the **three-dimensional Bessel process** $\{\rho(t) \colon t \geqslant 0\}$ by taking three independent Brownian motions and putting

$$\rho(t) = \sqrt{B_1(t)^2 + B_2(t)^2 + B_3(t)^2} \,.$$

The only nontrivial fact we need about this process is that, for $0 < r < a < R$,

$$\mathbb{P}\{\text{exit } (r, R) \text{ at } R \,|\, \rho(0) = a\} = \frac{\frac{1}{r} - \frac{1}{a}}{\frac{1}{r} - \frac{1}{R}},$$

see Theorem 3.18 in the Brownian motion book.

Fix $h > 0$ and assume $\rho(0) = h$. Define the stopping times $\{\tau_j^{(h)} : j = 0, 1, \ldots\}$ by $\tau_0^{(h)} = 0$ and, for $j \geqslant 0$,

$$\tau_{j+1}^{(h)} = \min\left\{t > \tau_j^{(h)} : |\rho(t) - \rho(\tau_j^{(h)})| = h\right\}.$$

Given that $\rho(\tau_j^{(h)}) = kh$ for some $k > 0$, by our hitting estimate, we have that

$$\rho(\tau_{j+1}^{(h)}) = \begin{cases} (k+1)h, & \text{with probability } \frac{k+1}{2k}\,, \\ (k-1)h, & \text{with probability } \frac{k-1}{2k}\,. \end{cases} \tag{7}$$

We abbreviate $\tau_j = \tau_j^{(1)}$. By (6) and (7), the sequence $\{\rho(\tau_j) : j = 0, 1, \ldots\}$ has the same distribution as the simple random walk on $\mathbb{N}$ conditioned to avoid zero, with the initial condition $\tilde{\rho}(0) = 1$.

**Lemma 7.6.** *The sequence $\{\tau_n - n : n \geqslant 0\}$ is a martingale and there exists $C > 0$ with*

$$\mathrm{Var}(\tau_n - n) \leqslant C\, n\,.$$

**Proof.** If $\{B(t) : t \geqslant 0\}$ is standard Brownian motion, then we know from Lemma 4.4 that $\{B(t)^2 - t : t \geqslant 0\}$ is a martingale. As $\{\rho(t)^2 - 3t : t \geqslant 0\}$ is the sum of three independent copies of this martingale, it is also a martingale. Given that $\rho(\tau_{n-1}) = k$, optional sampling for this martingale at times $\tau_{n-1}$ and $\tau_n$ yields

$$k^2 - 3\tau_{n-1} = \frac{(k+1)^3}{2k} + \frac{(k-1)^3}{2k} - 3\mathbb{E}[\tau_n \mid \mathcal{F}^+(\tau_{n-1})]\,,$$

hence $\mathbb{E}[\tau_n - \tau_{n-1} \mid \mathcal{F}^+(\tau_{n-1})] = 1$, so that $\{\tau_n - n : n \geqslant 0\}$ is a martingale. To bound its variance, consider the scalar product

$$Z := \left\langle W(t+1) - W(t)\,,\, \tfrac{W(t)}{|W(t)|} \right\rangle,$$

where $W(t) = (B_1(t), B_2(t), B_3(t))$. Given $\mathcal{F}^+(t)$ the distribution of $Z$ is standard normal. Moreover,

$$Z = \left\langle W(t+1)\,,\, \tfrac{W(t)}{|W(t)|} \right\rangle - |W(t)| \leqslant |W(t+1)| - |W(t)|\,.$$

Therefore $\mathbb{P}\{|W(t+1)| - |W(t)| > 2 \,|\, \mathcal{F}(t)\} \geqslant \mathbb{P}\{Z > 2\}$. For any $n$,

$$\bigcup_{j=1}^{k} \left\{|W(\tau_{n-1} + j)| - |W(\tau_{n-1} + j - 1)| > 2\right\} \subset \{\tau_n - \tau_{n-1} \leqslant k\},$$

so that, given $\tau_{n-1}$, the difference $\tau_n - \tau_{n-1}$ is stochastically bounded from above by a geometric random variable with parameter $p := \mathbb{P}\{Z > 2\}$. Hence,

$$\mathrm{Var}(\tau_n - \tau_{n-1} - 1) \leqslant \mathbb{E}\big[(\tau_n - \tau_{n-1})^2\big] \leqslant \frac{2}{p}\,.$$

By orthogonality of martingale differences, we conclude that $\mathrm{Var}(\tau_n - n) \leqslant 2n/p$, which completes the proof. ■

We use the following notation,

- $\{S(j)\colon j = 0, 1, \ldots\}$ is a simple random walk in $\mathbb{Z}$,
- $\{\tilde{M}(j)\colon j = 0, 1, \ldots\}$ defined by $\tilde{M}(j) = \max_{0\leqslant a\leqslant j} S(a)$;
- $\{\tilde{\rho}(j)\colon j = 0, 1, \ldots\}$ is simple random walk on $\mathbb{N}$ conditioned to avoid 0,
- $\{\tilde{I}(j)\colon j = 0, 1, \ldots\}$ defined by $\tilde{I}(j) = \min_{k\geqslant j} \tilde{\rho}(k)$.

Let $\{I(t)\colon t \geqslant 0\}$ defined by $I(t) = \min_{s\geqslant t} \rho(s)$ be the future minimum process of the process $\{\rho(t)\colon t \geqslant 0\}$.

**Proposition 7.7.** *Let $\tilde{I}(0) = \tilde{\rho}(0) = 0$, and extend the processes $\{\tilde{\rho}(j)\colon j = 0, 1, \ldots\}$ and $\{\tilde{I}(j)\colon j = 0, 1, \ldots\}$ to $[0, \infty)$ by linear interpolation. Then*

$$\big\{h\tilde{\rho}(t/h^2)\colon 0 \leqslant t \leqslant 1\big\} \stackrel{\mathrm{d}}{\to} \big\{\rho(t)\colon 0 \leqslant t \leqslant 1\big\} \quad \text{as } h \downarrow 0\,, \tag{8}$$

*and*

$$\big\{h\tilde{I}(t/h^2)\colon 0 \leqslant t \leqslant 1\big\} \stackrel{\mathrm{d}}{\to} \big\{I(t)\colon 0 \leqslant t \leqslant 1\big\} \quad \text{as } h \downarrow 0\,, \tag{9}$$

*where $\stackrel{\mathrm{d}}{\to}$ indicates convergence in law as random elements of $\mathbf{C}[0, 1]$.*

**Proof.** For any $h > 0$, Brownian scaling implies that the process $\{\tau_n^{(h)}\colon n = 0, 1, \ldots\}$ has the same law as the process $\{h^2\tau_n\colon n = 0, 1, \ldots\}$. Doob's $\mathbf{L}^2$ maximal inequality, $\mathbb{E}[\max_{1\leqslant k\leqslant n} X_k^2] \leqslant 4\,\mathbb{E}[|X_n|^2]$ and Lemma 7.6 yield

$$\mathbb{E}\big[\max_{0\leqslant j\leqslant n} (\tau_j - j)^2\big] \leqslant C\,n,$$

for a suitable constant $C > 0$. Therefore, taking $n = \lfloor h^{-2}t\rfloor$,

$$\mathbb{E}\Big[\max_{0\leqslant t\leqslant 1} (\tau^{(h)}_{\lfloor h^{-2}t\rfloor} - h^2\lfloor h^{-2}t\rfloor)^2\Big] = h^4\,\mathbb{E}\Big[\max_{0\leqslant t\leqslant 1} (\tau_{\lfloor h^{-2}t\rfloor} - \lfloor h^{-2}t\rfloor)^2\Big] \leqslant C\,h^2\,,$$

whence also (for a slightly larger constant)

$$\mathbb{E}\Big[\max_{0\leqslant t\leqslant 1} (\tau^{(h)}_{\lfloor h^{-2}t\rfloor} - t)^2\Big] \leqslant C\,h^2\,. \tag{10}$$

Since $\{\rho(t)\colon 0 \leqslant t \leqslant 1\}$ is uniformly continuous almost surely, we infer that

$$\max_{0\leqslant t\leqslant 1} |\rho(\tau^{(h)}_{\lfloor h^{-2}t\rfloor}) - \rho(t)| \to 0 \quad \text{in probability as } h \downarrow 0,$$

and similar reasoning gives the analogous result when $\lfloor\cdot\rfloor$ is replaced by $\lceil\cdot\rceil$. Since $\tilde{\rho}(t/h^2)$ is, by definition, a weighted average of $\tilde{\rho}(\lfloor h^{-2}t\rfloor)$ and $\tilde{\rho}(\lceil h^{-2}t\rceil)$, the proof of (8) is now concluded by recalling that $\{\rho(\tau_j^{(h)})\colon j = 0, 1, \ldots\}$ has the same distribution as $\{h\tilde{\rho}(j)\colon j = 0, 1, \ldots\}$. Similarly, $\{I(\tau_j^{(h)})\colon j = 0, 1, \ldots\}$ has the same distribution as $\{h\tilde{I}(j)\colon j = 0, 1, \ldots\}$, so (9) follows from (10) and the continuity of $I$. ■

**Theorem 7.8. (Pitman's $2M - X$ theorem)** *Let $\{B(t)\colon t \geqslant 0\}$ be a standard Brownian motion and let $M(t) = \max_{0\leqslant s\leqslant t} B(s)$ denote its maximum up to time $t$. Also let $\{\rho(t)\colon t \geqslant 0\}$ be a three-dimensional Bessel process and let $\{I(t)\colon t \geqslant 0\}$ be the corresponding future infimum process given by $I(t) = \inf_{s\geqslant t} \rho(s)$. Then*

$$\big\{(2M(t) - B(t), M(t))\colon t \geqslant 0\big\} \stackrel{\mathrm{d}}{=} \big\{(\rho(t), I(t))\colon t \geqslant 0\big\}\,.$$

*In particular, $\{2M(t) - B(t)\colon t \geqslant 0\}$ is a three-dimensional Bessel process.*

**Proof.** Following Pitman's original paper, we prove the theorem in the discrete setting, i.e. we show that, for $S(0) = \tilde\rho(0) = 0$,

$$\big\{(2\tilde M(j) - S(j), \tilde M(j))\colon j = 0, 1, \dots\big\} \stackrel{\mathrm{d}}{=} \big\{(\tilde\rho(j), \tilde I(j))\colon j = 0, 1, \dots\big\}\,. \tag{11}$$

The theorem then follows directly by invoking Donsker's invariance principle and Proposition 7.7. First note that (11) is equivalent to

$$\big\{(S(j), \tilde M(j))\colon j = 0, 1, \dots\big\} \stackrel{\mathrm{d}}{=} \big\{(2\tilde I(j) - \tilde\rho(j), \tilde I(j))\colon j = 0, 1, \dots\big\}\,,$$

which we establish by computing the transition probabilities. If $S(j) < \tilde M(j)$, then clearly

$$(S(j+1), \tilde M(j+1)) = \begin{cases} (S(j)+1, \tilde M(j)), & \text{with probability } \frac12\,,\\ (S(j)-1, \tilde M(j)), & \text{with probability } \frac12\,. \end{cases} \tag{12}$$

If $S(j) = \tilde M(j)$, then

$$(S(j+1), \tilde M(j+1)) = \begin{cases} (S(j)+1, \tilde M(j)+1), & \text{with probability } \frac12\,,\\ (S(j)-1, \tilde M(j)), & \text{with probability } \frac12\,. \end{cases} \tag{13}$$

We now compute the transition probabilities of $\{(2\tilde I(j) - \tilde\rho(j), \tilde I(j))\colon j = 0, 1, \dots\}$. To this end, we first show that $\{\tilde I(j)\colon j = 0, 1, \dots\}$ is the maximum process of $\{2\tilde I(j) - \tilde\rho(j)\colon j = 0, 1, \dots\}$. Indeed, for all $j \leqslant k$, since $(\tilde I - \tilde\rho)(j) \leqslant 0$, we have

$$2\tilde I(j) - \tilde\rho(j) = \tilde I(j) + (\tilde I - \tilde\rho)(j) \leqslant \tilde I(k)\,.$$

On the other hand, let $j_*$ be the minimal $j_* \leqslant k$ such that $\tilde I(j_*) = \tilde I(k)$. Then $\tilde\rho(j_*) = \tilde I(j_*)$ and we infer that $(2\tilde I - \tilde\rho)(j_*) = \tilde I(j_*) = I(k)$.

Assume now that $2\tilde I(j) - \tilde\rho(j) < \tilde I(j)$, i.e., $\tilde\rho(j) > \tilde I(j)$. Lemma 7.5 and the fact that $\{S(j)\colon j = 0, 1, \dots\}$ is recurrent imply that, for integers $k \geqslant i > 0$,

$$\mathbb{P}\big\{\exists j \text{ with } \tilde\rho(j) = i \,\big|\, \tilde\rho(0) = k\big\} = \frac{i}{k}\,\mathbb{P}\big\{\exists j \text{ with } S(j) = i \big|\, S(0) = k\big\} = \frac{i}{k}.$$

Thus, for $k \geqslant i > 0$,

$$\begin{aligned}&\mathbb{P}\big\{\tilde I(j) = i \,\big|\, \tilde\rho(j) = k\big\}\\ &\quad= \mathbb{P}\big\{\exists j \text{ with } \tilde\rho(j) = i \,\big|\, \tilde\rho(0) = k\big\} - \mathbb{P}\big\{\exists j \text{ with } \tilde\rho(j) = i - 1 \,\big|\, \tilde\rho(0) = k\big\} = \frac{1}{k}\,.\end{aligned}$$

Therefore,

$$\mathbb{P}\big\{\tilde\rho(j+1)=k-1 \mid \tilde\rho(j)=k, \tilde I(j)=i\big\}$$
$$= \frac{\mathbb{P}\{\tilde\rho(j+1)=k-1, \tilde I(j)=i \mid \tilde\rho(j)=k\}}{\mathbb{P}\{\tilde I(j)=i \mid \tilde\rho(j)=k\}} \tag{14}$$
$$= \frac{\frac{k-1}{2k}\frac{1}{k-1}}{\frac{1}{k}} = \frac{1}{2}\,.$$

We conclude that if $2\tilde I(j)-\tilde\rho(j)<\tilde I(j)$, then

$$(2\tilde I(j+1)-\tilde\rho(j+1), \tilde I(j+1))$$
$$= \begin{cases} (2\tilde I(j)-\tilde\rho(j)+1, \tilde I(j)), & \text{with probability } \frac12, \\ (2\tilde I(j)-\tilde\rho(j)-1, \tilde I(j)), & \text{with probability } \frac12. \end{cases} \tag{15}$$

Assume now that $\tilde\rho(j)=\tilde I(j)=k$. Then $\tilde\rho(j+1)=k+1$, and

$$\mathbb{P}\{\tilde I(j+1)=k+1 \mid \tilde I(j)=\tilde\rho(j)=k\}$$
$$= \frac{\mathbb{P}\{\tilde\rho(j+1)=k+1 \mid \tilde\rho(j)=k\}\mathbb{P}\{\tilde I(j+1)=k+1 \mid \tilde\rho(j+1)=k+1\}}{\mathbb{P}\{\tilde I(j)=k \mid \tilde\rho(j)=k\}}$$
$$= \frac{\frac{k+1}{2k}\frac{1}{k+1}}{\frac{1}{k}} = \frac{1}{2}\,.$$

Hence, if $\tilde\rho(j)=\tilde I(j)=k$, then we have

$$(2\tilde I(j+1)-\tilde\rho(j+1), \tilde I(j+1))$$
$$= \begin{cases} (2\tilde I(j)-\tilde\rho(j)+1, \tilde I(j)+1), & \text{with probability } \frac12\,, \\ (2\tilde I(j)-\tilde\rho(j)-1, \tilde I(j)), & \text{with probability } \frac12\,. \end{cases} \tag{16}$$

Finally, comparing (12) and (13) to (15) and (16) completes the proof. ■

# 8 Local times of Brownian motion

How can we measure the amount of time spent by a standard Brownian motion $\{B_t\colon t\geqslant 0\}$ at zero? Recall that, almost surely, the zero set has Hausdorff dimension 1/2, so its Lebesgue measure is zero. We approach this problem by counting the number of downcrossings of a nested sequence of intervals decreasing to zero. Given $a<b$, we define stopping times $\tau_0=0$ and, for $j\geqslant 1$,

$$\sigma_j = \inf\big\{t>\tau_{j-1}\colon B(t)=b\big\}, \qquad \tau_j=\inf\big\{t>\sigma_j\colon B(t)=a\big\}. \tag{17}$$

For every $t>0$ we denote by

$$D(a,b,t)=\max\big\{j\in\mathbb{N}\colon \tau_j\leqslant t\big\}$$

the number of *downcrossings* of the interval $[a, b]$ before time $t$. Note that $D(a, b, t)$ is almost surely finite by the uniform continuity of Brownian motion on the compact interval $[0, t]$.

**Theorem 8.1** (Downcrossing representation of the local time at zero)**.** *There exists a stochastic process* $\{L_t : t \geqslant 0\}$ *called the* **local time at zero** *such that for all sequences* $a_n \uparrow 0$ *and* $b_n \downarrow 0$ *with* $a_n < b_n$*, almost surely,*

$$\lim_{n\to\infty} 2(b_n - a_n)\, D(a_n, b_n, t) = L_t \qquad \textit{for every } t > 0$$

*and this process is almost surely locally* $\gamma$*-Hölder continuous for any* $\gamma < 1/2$*.*

We first prove the convergence for the case when the Brownian motion is stopped at the time $T = T_b$ when it first reaches some level $b > b_1$. This has the advantage that there cannot be any uncompleted upcrossings.

**Lemma 8.2.** *For any two sequences* $a_n \uparrow 0$ *and* $b_n \downarrow 0$ *with* $a_n < b_n$*, the discrete time stochastic process* $\{2\frac{b_n - a_n}{b - a_n}\, D(a_n, b_n, T) : n \in \mathbb{N}\}$ *is a submartingale.*

**Proof.** Without loss of generality we may assume that, for each $n$, we have either (1) $a_n = a_{n+1}$ or (2) $b_n = b_{n+1}$. In case (1), we observe that the total number $D(a_n, b_{n+1}, T)$ of downcrossings of $[a_n, b_{n+1}]$ given $D(a_n, b_n, T)$ is the sum of $D(a_n, b_n, T)$ independent geometric random variables with success parameter $p = \frac{b_{n+1} - a_n}{b_n - a_n}$ plus a nonnegative contribution. Hence,

$$\mathbb{E}\big[\tfrac{b_{n+1} - a_n}{b - a_n}\, D(a_n, b_{n+1}, T) \,\big|\, \mathcal{F}_n\big] \geqslant \tfrac{b_n - a_n}{b - a_n}\, D(a_n, b_n, T),$$

which is the submartingale property. Case (2) follows similarly. ∎

**Lemma 8.3.** *For any two sequences* $a_n \uparrow 0$ *and* $b_n \downarrow 0$ *with* $a_n < b_n$ *the limit*

$$L(T_b) := \lim_{n\to\infty} 2(b_n - a_n)\, D(a_n, b_n, T_b) \tag{18}$$

*exists almost surely. It is not zero and does not depend on the choices.*

**Proof.** Observe that $D(a_n, b_n, T_b)$ is a geometric random variable on $\{0, 1, \ldots\}$ with parameter $(b_n - a_n)/(b - a_n)$. Recall that the variance of a geometric random variable on $\{0, 1, \ldots\}$ with parameter $p$ is $(1 - p)/p^2$, and so its second moment is bounded by $2/p^2$. Hence

$$\mathbb{E}\big[4(b_n - a_n)^2\, D(a_n, b_n, T_b)^2\,\big] \leqslant 8\,(b - a_n)^2,$$

and thus the submartingale in Lemma 8.2 is $\mathbf{L}^2$-bounded. By the submartingale convergence theorem its limit exists almost surely and in $\mathbf{L}^2$ ensuring that the limit is nonzero. Finally, note that the limit does not depend on the choice of the sequence $a_n \uparrow 0$ and $b_n \downarrow 0$ because if it did, then given two sequences with different limits in (18) we could construct a sequence of intervals alternating between the sequences, for which the limit in (18) would not exist. ∎

**Lemma 8.4.** *For any fixed time $t > 0$, almost surely, the limit*

$$L(t) := \lim_{n\to\infty} 2(b_n - a_n)\, D(a_n, b_n, t) \qquad \textit{exists.}$$

**Proof.** We define an auxiliary Brownian motion $\{B_t(s)\colon s \geqslant 0\}$ by $B_t(s) = B(t + s)$. For any integer $b > b_1$ we denote by $D_t(a_n, b_n, T_b)$ the number of downcrossings of the interval $[a_n, b_n]$ by the auxiliary Brownian motion before it hits $b$. Then, almost surely,

$$L_t(T_b) := \lim_{n\uparrow\infty} 2(b_n - a_n)\, D_t(a_n, b_n, T_b),$$

exists by the previous lemma. Given $t > 0$ we fix a Brownian path such that this limit and the limit in Lemma 8.3 exist for all integers $b > b_1$. Pick $b$ so large that $T_b > t$. Define

$$L(t) := L(T_b) - L_t(T_b)\,.$$

To show that this is the required limit, observe that

$$D(a_n, b_n, T_b) - D_t(a_n, b_n, T_b) - 1 \leqslant D(a_n, b_n, t) \leqslant D(a_n, b_n, T_b) - D_t(a_n, b_n, T_b),$$

where the correction $-1$ on the left hand side arises from the possibility that $t$ interrupts a downcrossing. Multiplying by $2(b_n - a_n)$ and taking a limit gives $L(T_b) - L_t(T_b)$ for both bounds, proving convergence. ■

We now have to study the dependence of $L(t)$ on the time $t$ in more detail. To simplify the notation we write

$$I_n(s, t) = 2(b_n - a_n)\big(D(a_n, b_n, t) - D(a_n, b_n, s)\big) \qquad \text{for all } 0 \leqslant s < t\,.$$

The following lemma contains a probability estimate, which is sufficient to get the convergence of the downcrossing numbers jointly for all times and to establish Hölder continuity.

**Lemma 8.5.** *Let $\gamma < 1/2$ and $0 < \varepsilon < (1 - 2\gamma)/3$. Then, for all $t \geqslant 0$ and $0 < h < 1$, we have*

$$\mathbb{P}\big\{L(t + h) - L(t) > h^\gamma\big\} \leqslant 2\, \exp\{-\tfrac{1}{2}\, h^{-\varepsilon}\}\,.$$

**Proof.** As, by Fatou's lemma,

$$\begin{aligned}\mathbb{P}\big\{L(t + h) - L(t) > h^\gamma\big\} &= \mathbb{P}\big\{\liminf_{n\to\infty} I_n(t, t + h) > h^\gamma\big\} \\ &\leqslant \liminf_{n\to\infty} \mathbb{P}\big\{I_n(t, t + h) > h^\gamma\big\}\end{aligned}$$

we can focus on estimating $\mathbb{P}\{I_n(t, t + h) > h^\gamma\}$ for fixed large $n$. It clearly suffices to estimate $\mathbb{P}_{b_n}\{I_n(0, h) > h^\gamma\}$. Let $T_h = \inf\{s > 0\colon B(s) = b_n + h^{(1-\varepsilon)/2}\}$ and observe that

$$\big\{I_n(0, h) > h^\gamma\big\} \subset \big\{I_n(0, T_h) > h^\gamma\big\} \cup \big\{T_h < h\big\}.$$

The number of downcrossings of $[a_n, b_n]$ during the period before $T_h$ is geometrically distributed on $\{0, 1, \ldots\}$ with failure parameter $(b_n - a_n + h^{(1-\varepsilon)/2})^{-1} h^{(1-\varepsilon)/2}$ and thus

$$\mathbb{P}_{b_n}\big\{I_n(0, T_h) > h^\gamma\big\} \leqslant \Big(\frac{h^{(1-\varepsilon)/2}}{b_n - a_n + h^{(1-\varepsilon)/2}}\Big)^{\lfloor \frac{1}{2(b_n - a_n)}\, h^\gamma \rfloor}$$
$$\stackrel{n\to\infty}{\longrightarrow} \exp\big\{-\tfrac{1}{2}\, h^{\gamma - \frac{1}{2} + \frac{\varepsilon}{2}}\big\} \leqslant \exp\big\{-\tfrac{1}{2}\, h^{-\varepsilon}\big\}.$$

With $\{W_s : s \geqslant 0\}$ denoting a standard Brownian motion,

$$\mathbb{P}_{b_n}\big\{T_h < h\big\} = \mathbb{P}\Big\{\max_{0 \leqslant s \leqslant h} W_s \geqslant h^{(1-\varepsilon)/2}\Big\} \leqslant \sqrt{\tfrac{2}{\pi h^{-\varepsilon}}}\, \exp\big\{-\tfrac{1}{2}\, h^{-\varepsilon}\big\}.$$

The result follows by adding the last two displayed formulas. ■

**Lemma 8.6.** *Almost surely,*

$$L(t) := \lim_{n\to\infty} 2(b_n - a_n)\, D(a_n, b_n, t)$$

*exists for every* $t \geqslant 0$*.*

**Proof.** It suffices to prove the simultaneous convergence for all $0 \leqslant t \leqslant 1$. We define a countable set of gridpoints

$$\mathcal{G} = \bigcup_{m\in\mathbb{N}} \mathcal{G}_m \cup \{1\}, \qquad \text{for } \mathcal{G}_m = \big\{\tfrac{k}{m} : k \in \{0, \ldots, m-1\}\big\}$$

and show that the stated convergence holds on the event

$$E_M = \bigcap_{t\in\mathcal{G}} \big\{L(t) = \lim_{n\to\infty} 2(b_n - a_n)\, D(a_n, b_n, t) \text{ exists }\big\}$$
$$\cap \bigcap_{m>M} \bigcap_{t\in\mathcal{G}_m} \big\{L(t + \tfrac{1}{m}) - L(t) \leqslant (1/m)^\gamma\big\}.$$

which, by choosing $M$ suitably, has probability arbitrarily close to one by the previous two lemmas. Given any $t \in [0, 1)$ and a large $m$ we find $t_1, t_2 \in \mathcal{G}_m$ with $t_2 - t_1 = \frac{1}{m}$ and $t \in [t_1, t_2]$. We obviously have

$$2(b_n - a_n)\, D(a_n, b_n, t_1) \leqslant 2(b_n - a_n)\, D(a_n, b_n, t) \leqslant 2(b_n - a_n)\, D(a_n, b_n, t_2).$$

Both bounds converge on $E_M$, and the difference of the limits is $L(t_2) - L(t_1)$, which is bounded by $m^{-\gamma}$ and thus can be made arbitrarily small by choosing a large $m$. ■

**Lemma 8.7.** *For* $\gamma < \frac{1}{2}$*, almost surely, the process* $\{L(t) : t \geqslant 0\}$ *is locally* $\gamma$*-Hölder continuous.*

**Proof.** It suffices to look at $0 \leqslant t < 1$. We use the notation of the proof of the previous lemma and show that $\gamma$-Hölder continuity holds on the set $E_M$ constructed there. Indeed, whenever $0 \leqslant s < t < 1$ and $t - s < 1/M$ we pick $m \geqslant M$ such that

$$\tfrac{1}{m+1} \leqslant t - s < \tfrac{1}{m}\,.$$

We take $t_1 \leqslant s$ with $t_1 \in \mathcal{G}_m$ and $s - t_1 < 1/m$, and $t_2 \geqslant t$ with $t_2 \in \mathcal{G}_m$ and $t_2 - t < 1/m$. Note that $t_2 - t_1 \leqslant 2/m$ by construction and hence,

$$L(t) - L(s) \leqslant L(t_2) - L(t_1) \leqslant 2(1/m)^\gamma \leqslant 2\big(\tfrac{m+1}{m}\big)^\gamma\,(t-s)^\gamma.$$

The result follows as the fraction on the right is bounded by 2. ■

This completes the proof of Theorem 8.1. It is easy to see from this representation that, almost surely, the local time at zero increases only on the zero set of the Brownian motion. The following theorem is a substantial refinement of the downcrossing representation, the proof of which we omit.

**Theorem 8.8** (Trotter's theorem)**.** *Let $D^{(n)}(a,t)$ be the number of downcrossings before time $t$ of the $n^{\text{th}}$ stage dyadic interval containing $a$. Then, almost surely,*

$$L^a(t) = \lim_{n\to\infty} 2^{-n+1}\, D^{(n)}(a,t) \qquad \textit{exists for all } a \in \mathbb{R} \textit{ and } t \geqslant 0.$$

*The process $\{L^a(t)\colon t \geqslant 0\}$ is called the* **local time at level** *$a$. Moreover, for every $\gamma < \frac{1}{2}$, the random field*

$$\{L^a(t)\colon a \in \mathbb{R}, t \geqslant 0\}$$

*is almost surely locally $\gamma$-Hölder continuous.*

# 9 The Ray–Knight theorem

We now have a closer look at the distributions of local times $L^x(T)$ as a function of the level $x$ in the case that Brownian motion is started at an arbitrary point and stopped at the time $T$ when it first hits level zero. The following remarkable distributional identity goes back to the work of Ray and Knight.

**Theorem 9.1** (Ray–Knight theorem)**.** *Suppose $a > 0$ and $\{B_t\colon 0 \leqslant t \leqslant T\}$ is a Brownian motion started at $a$ and stopped at time $T = \inf\{t \geqslant 0\colon B_t = 0\}$, when it reaches level zero for the first time. Then*

$$\{L^x(T)\colon 0 \leqslant x \leqslant a\} \stackrel{\mathrm{d}}{=} \{B_1(x)^2 + B_2(x)^2\colon 0 \leqslant x \leqslant a\}\,,$$

*where $B_1$, $B_2$ are independent Brownian motions.*

**Remark 9.2.** The process $(|W(x)|\colon x \geqslant 0)$ given by $|W(x)| = \sqrt{B_1(x)^2 + B_2(x)^2}$ is the **two-dimensional Bessel process**. ◇

As a warm-up, we look at one point $0 < x \leqslant a$. Recall that

$$\lim_{n\to\infty} \tfrac{2}{n}\, D_n(x) = L^x(T) \qquad \text{almost surely,}$$

where $D_n(x)$ denotes the number of downcrossings of $[x-1/n, x]$ before time $T$.

**Lemma 9.3.** *For any $0 < x \leqslant a$, we have $\frac{2}{n}\, D_n(x) \xrightarrow{\mathrm{d}} |W(x)|^2$ as $n \uparrow \infty$.*

**Proof.** By the strong Markov property and the exit probabilities from an interval it is clear that, provided $n > 1/x$, the random variable $D_n(x)$ is geometrically distributed with parameter $1/(nx)$. Hence, as $n \to \infty$,

$$\mathbb{P}\{D_n(x) > ny/2\} = \big(1 - \tfrac{1}{nx}\big)^{\lfloor ny/2 \rfloor} \longrightarrow e^{-y/(2x)}\,,$$

and the result follows, as $|W(x)|^2$ is exponentially distributed with mean $2x$. ■

The essence of the Ray–Knight theorem is captured in the following 'two-point version', which we will prove here instead of the full result. We fix two points $x$ and $x+h$ with $0 < x < x+h < a$.

**Lemma 9.4.** *Suppose $nu_n$ are nonnegative, even integers and $u_n \to u$. For any $\lambda \geqslant 0$,*

$$\lim_{n\to\infty} \mathbb{E}\big[\exp\big\{-\lambda \tfrac{2}{n} D_n(x+h)\big\} \,\big|\, \tfrac{2}{n} D_n(x) = u_n\big] = \mathbb{E}\big[\exp\big\{-\lambda |W(h)|^2\big\}\big],$$

*where $\{|W(x)| \colon x \geqslant 0\}$ is a two-dimensional Bessel process started in $\sqrt{u}$.*

The next three lemmas are the crucial ingredients for the proof of Theorem 9.1.

**Lemma 9.5.** *Let $0 < x < x+h < a$. Then, for all $n > h$, we have*

$$D_n(x+h) = D + \sum_{j=1}^{D_n(x)} I_j\, N_j\,,$$

*where*

- *$D = D^{(n)}$ is the number of downcrossings of the interval $[x+h-\frac{1}{n}, x+h]$ before the Brownian motion hits level $x$,*
- *for any $j \in \mathbb{N}$ the random variable $I_j = I_j^{(n)}$ is Bernoulli distributed with mean $\frac{1}{nh+1}$,*
- *for any $j \in \mathbb{N}$ the random variable $N_j = N_j^{(n)}$ is geometrically distributed with mean $nh+1$,*

*and all these random variables are independent of each other and of $D_n(x)$.*

**Proof.** The decomposition of $D_n(x+h)$ is based on counting the number of downcrossings of the interval $[x+h-1/n, x+h]$ that have taken place between the stopping times in the sequence

$$\begin{aligned} \tau_0 &= \inf\big\{t > 0\colon B(t) = x\big\}, & \tau_1 &= \inf\big\{t > \tau_0\colon B(t) = x - \tfrac{1}{n}\big\}, \\ \tau_{2j} &= \inf\big\{t > \tau_{2j-1}\colon B(t) = x\big\}, & \tau_{2j+1} &= \inf\big\{t > \tau_{2j}\colon B(t) = x - \tfrac{1}{n}\big\}, \end{aligned}$$

for $j \geqslant 1$. By the strong Markov property the pieces

$$\begin{aligned} B^{(0)}\colon [0, \tau_0] &\to \mathbb{R}, & B^{(0)}(s) &= B(s) \\ B^{(j)}\colon [0, \tau_j - \tau_{j-1}] &\to \mathbb{R}, & B^{(j)}(s) &= B(\tau_{j-1} + s),\ j \geqslant 1, \end{aligned}$$

are all independent. The crucial observation of the proof is that the vector $D_n(x)$ is a function of the pieces $B^{(2j)}$ for $j \geqslant 1$, whereas we shall define the random variables $D, I_1, I_2, \ldots$ and $N_1, N_2 \ldots$ depending only on the other pieces $B^{(0)}$ and $B^{(2j-1)}$ for $j \geqslant 1$.
First, let $D$ be the number of downcrossings of $[x+h-1/n, x+h]$ during the time interval $[0, \tau_0]$. Then fix $j \geqslant 1$ and hence a piece $B^{(2j-1)}$. Define $I_j$ to be the indicator of the event that $B^{(2j-1)}$ reaches level $x+h$ during its lifetime. By Theorem 4.6 this event has probability $1/(nh+1)$. Observe that the number of downcrossings by $B^{(2j-1)}$ is zero if the event fails. If the event holds, we define $N_j$ as the number of downcrossings of $[x+h-1/n, x+h]$ by $B^{(2j-1)}$, which is a geometric random variable with mean $nh+1$ by the strong Markov property and Theorem 4.6.

The claimed decomposition follows now from the fact that the pieces $B^{(2j)}$ for $j \geqslant 1$ do not upcross the interval $[x+h-1/n, x+h]$ by definition and that $B^{(2j-1)}$ for $j = 1, \ldots, D_n(x)$ are exactly the pieces that take place before the Brownian motion reaches level zero. ∎

**Lemma 9.6.** *Suppose $nu_n$ are nonnegative, even integers and $u_n \to u$. Then*

$$\frac{2}{n}\, D^{(n)} + \frac{2}{n} \sum_{j=1}^{\frac{nu_n}{2}} I_j^{(n)}\, N_j^{(n)} \xrightarrow{\ \mathrm{d}\ } \tilde{X}^2 + \tilde{Y}^2 + 2 \sum_{j=1}^{M} \tilde{Z}_j \qquad \text{as } n \uparrow \infty,$$

*where $\tilde{X}$, $\tilde{Y}$ are normally distributed with mean zero and variance $h$, the random variable $M$ is Poisson distributed with parameter $u/(2h)$ and $\tilde{Z}_1, \tilde{Z}_2, \ldots$ are exponentially distributed with mean $h$, and all these random variables are independent.*

**Proof.** By Lemma 9.3, we have, for $\tilde{X}$, $\tilde{Y}$ as defined in the lemma,

$$\frac{2}{n}\, D^{(n)} \xrightarrow{\ \mathrm{d}\ } |W(h)|^2 \stackrel{\mathrm{d}}{=} \tilde{X}^2 + \tilde{Y}^2 \qquad \text{as } n \uparrow \infty.$$

Moreover, we observe that

$$\frac{2}{n}\sum_{j=1}^{\frac{nu_n}{2}} I_j^{(n)}\, N_j^{(n)} \stackrel{\mathrm{d}}{=} \frac{2}{n}\sum_{j=1}^{B_n} N_j^{(n)},$$

where $B_n$ is binomial with parameters $nu_n/2 \in \{0,1,\ldots\}$ and $1/(nh+1) \in (0,1)$ and independent of $N_1^{(n)}, N_2^{(n)}, \ldots$. We now show that, when $n \uparrow \infty$, the random variables $B_n$ converge in distribution to $M$ and the random variables $\frac{1}{n} N_j^{(n)}$ converge to $\tilde{Z}_j$, as defined in the lemma. First note that, for $\lambda, \theta > 0$, we have

$$\mathbb{E}\exp\big\{-\lambda \tilde{Z}_j\big\} = \tfrac{1}{\lambda h+1}, \quad \mathbb{E}\big[\theta^M\big] = \exp\big\{-\tfrac{u(1-\theta)}{2h}\big\},$$

and hence

$$\mathbb{E}\exp\Big\{-\lambda\sum_{j=1}^{M}\tilde{Z}_j\Big\} = \mathbb{E}\big(\tfrac{1}{\lambda h+1}\big)^M = \exp\big\{-\tfrac{u}{2h}\,\tfrac{\lambda h}{\lambda h+1}\big\} = \exp\big\{-\tfrac{u\lambda}{2\lambda h+2}\big\}.$$

Convergence of $\frac{1}{n} N_j^{(n)}$ is best seen using tail probabilities

$$\mathbb{P}\big\{\tfrac{1}{n} N_j^{(n)} > a\big\} = \big(1-\tfrac{1}{nh+1}\big)^{\lfloor na\rfloor} \longrightarrow \exp\big\{-\tfrac{a}{h}\big\} = \mathbb{P}\{\tilde{Z}_j > a\}.$$

Hence, for a suitable sequence $\delta_n \to 0$,

$$\mathbb{E}\exp\big\{-\lambda\tfrac{1}{n} N_j^{(n)}\big\} = \frac{1+\delta_n}{\lambda h+1}.$$

For the binomial distributions we have

$$\mathbb{E}\big[\theta^{B_n}\big] = \Big(\tfrac{\theta}{nh+1} + \big(1-\tfrac{1}{nh+1}\big)\Big)^{nu_n/2} \longrightarrow \exp\big\{-\tfrac{u(1-\theta)}{2h}\big\},$$

and thus

$$\begin{aligned}\lim_{n\uparrow\infty}\mathbb{E}\exp\Big\{-\lambda\frac{1}{n}\sum_{j=1}^{B_n} N_j^{(n)}\Big\} &= \lim_{n\uparrow\infty}\mathbb{E}\Big[\big(\tfrac{1+\delta_n}{\lambda h+1}\big)^{B_n}\Big] = \lim_{n\uparrow\infty}\exp\big\{-\tfrac{u}{2h}\,\tfrac{\lambda h-\delta_n}{\lambda h+1}\big\}\\ &= \exp\big\{-\tfrac{u\lambda}{2\lambda h+2}\big\} = \mathbb{E}\exp\Big\{-\lambda\sum_{j=1}^{M}\tilde{Z}_j\Big\}. \qquad \blacksquare\end{aligned}$$

**Lemma 9.7.** *Suppose $X$ is standard normally distributed, $Z_1, Z_2, \ldots$ standard exponentially distributed and $N$ Poisson distributed with parameter $\ell^2/2$ for some $\ell > 0$. If all these random variables are independent, then*

$$(X+\ell)^2 \stackrel{\mathrm{d}}{=} X^2 + 2\sum_{j=1}^{N} Z_j.$$

**Proof.** It suffices to show that the Laplace transforms of the random variables on the two sides of the equation agree. Let $\lambda > 0$. Completing the square, we find

$$\begin{aligned}
\mathbb{E}\exp\{-\lambda\,(X+\ell)^2\} &= \frac{1}{\sqrt{2\pi}}\int \exp\{-\lambda\,(x+\ell)^2 - x^2/2\}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int \exp\Big\{-\tfrac{1}{2}\big(\sqrt{2\lambda+1}\,x + \tfrac{2\lambda\ell}{\sqrt{2\lambda+1}}\big)^2 - \lambda\ell^2 + \tfrac{2\lambda^2\ell^2}{2\lambda+1}\Big\}\,dx \\
&= \frac{1}{\sqrt{2\lambda+1}}\,\exp\Big\{-\tfrac{\lambda\ell^2}{2\lambda+1}\Big\}.
\end{aligned}$$

From the special case $\ell = 0$ we get $\mathbb{E}\exp\{-\lambda\,X^2\} = \frac{1}{\sqrt{2\lambda+1}}$. For any $\theta > 0$,

$$\mathbb{E}[\theta^N] = \exp\{-\ell^2/2\}\sum_{k=0}^{\infty}\tfrac{(\ell^2\theta/2)^k}{k!} = \exp\{(\theta-1)\ell^2/2\}\,.$$

Using this and that $\mathbb{E}\exp\{-2\lambda\,Z_j\} = \frac{1}{2\lambda+1}$ we get

$$\mathbb{E}\exp\Big\{-\lambda\big(X^2 + 2\sum_{j=1}^{N} Z_j\big)\Big\} = \frac{1}{\sqrt{2\lambda+1}}\,\mathbb{E}\Big(\frac{1}{2\lambda+1}\Big)^N = \frac{1}{\sqrt{2\lambda+1}}\,\exp\Big\{-\tfrac{\lambda\ell^2}{2\lambda+1}\Big\},$$

which completes the proof. ∎

Combining Lemmas 9.5 and 9.6 we get

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{E}\big[\exp\big\{-\lambda\tfrac{2}{n}D_n(x+h)\big\}\,\big|\,\tfrac{2}{n}D_n(x) = u_n\big] \\
&= \mathbb{E}\Big[\exp\Big\{-\lambda\Big(\tilde{X}^2 + \tilde{Y}^2 + 2\sum_{j=1}^{M}\tilde{Z}_j\Big)\Big\}\Big] \\
&= \mathbb{E}\Big[\exp\Big\{-\lambda h\Big(X^2 + Y^2 + 2\sum_{j=1}^{M} Z_j\Big)\Big\}\Big],
\end{aligned}$$

where $X$, $Y$ are standard normally distributed, $Z_1, Z_2, \dots$ are standard exponentially distributed and $M$ is Poisson distributed with parameter $\ell^2/2$, for $\ell = \sqrt{u/h}$. By Lemma 9.7 the right hand side can thus be rewritten as

$$\mathbb{E}\big[\exp\big\{-\lambda h\big((X+\sqrt{u/h})^2 + Y^2\big)\big\}\big] = \mathbb{E}_{(0,\sqrt{u})}\big[\exp\big\{-\lambda|W(h)|^2\big\}\big],$$

which proves the two-point version of the Ray-Knight theorem.