Aspects of Brownian Motion

Michelle Yeo and Đorđe Žikelić

1 Introduction

Brownian motion was firstly introduced by a Scottish botanist Robert Brown in 1827, who observed pollen grains moving randomly in water. Later it was introduced as a mathematical model that describes random movement of particles suspended in a fluid. This is caused by the particles bumping into the fast moving molecules in the liquid.

Let $\{\Omega, \mathcal{F}, \mathbb{P}\}$ be a probability space and $\{E, \mathcal{E}\}$ be a measurable space. Then an **E-valued random variable** is a measurable map $R : \Omega \to E$. By default, if $(E, \mathcal{E})$ is not specified, **random variable** refers to the case when $(E, \mathcal{E}) = (\mathbb{R}, \lambda)$. A **stochastic process** is a family of random variables $\{X_t\}_{t \in \mathcal{T}}$. Stochastic processes are functions of 2 variables: t and ω. If we fix t, we get a random variable $\omega \to X_t(\omega)$. If we fix ω, we get a **trajectory** of the process which is a random function $t \to X_t(\omega)$.

2 Existence of Brownian motion

In order to define and reason about Brownian motion, we firstly recall the definition and standard properties of Gaussian random variables. A real valued random variable X is said to have normal distribution with mean μ and variance σ^2 if for all $x \in \mathbb{R}$

$$\mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du.$$

Defintion 2.1 (Gaussian random variable)**.** *A random vector $X = (X_1, \ldots, X_n)$ is said to be a Gaussian random variable if there exists an $n \times m$ matrix A and $b \in \mathbb{R}^n$ such that $X^T = AY + b$, where Y is a random column vector with i.i.d. $N(0, 1)$ entries.*

Lemma 2.2. *(Properties of Gaussian random variables).*
Suppose X is a Gaussian random variable with mean μ and variance Σ. Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then

(i) $AX + b$ is a Gaussian random variable in $\mathbb{R}^m$ with mean $A\mu + b$ and variance $A\Sigma A^T$.

(ii) The density function on $\mathbb{R}^n$ given by

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} (\det \Sigma)^{-1/2} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

(iii) Suppose $X = (X_1, X_2)$ with $Cov(X_1, X_2) = 0$, then X_1 and X_2 are independent.

Lemma 2.3. *Suppose $X_1, X_2 \sim N(0, \sigma^2)$ are independent. Then $X_1 + X_2$ and $X_1 - X_2$ are independent and have distributions $N(0, 2\sigma^2)$.*

Proofs of these properties are standard and do not illustrate essential concepts for this short course, so we omit them. We are now ready to define the Brownian motion.

Defintion 2.4 (Brownian Motion). *Stochastic process $\{B_t | t \geq 0\}$ in $\mathbb{R}^d$ is called a* ***Brownian Motion*** *with start in $x \in \mathbb{R}^d$ if*

(i) $B_0 = x$

(ii) the process has independent increments, that is, $\forall 0 \leq t_1 \leq ... \leq t_n$, $B_{t_n} - B_{t_{n-1}}, ..., B_{t_2} - B_{t_1}$ are independent

(iii) $\forall h > 0$, $B_{t+h} - B_t$ is a Gaussian with mean 0 and variance hI

(iv) $\forall \omega$, the trajectory $t \to B_t(\omega)$ is almost surely continuous.

If $x = 0$, we say the Brownian motion is ***standard****.*

It is a subtle but important point that Brownian Motion has almost surely continuous trajectories but not necessarily continuous.

The existence of Brownian Motion is non trivial since it is not clear whether the finite-dimensional distributions given in the first three items above allow for the process to be a.s. continuous. The following construction shows that such a process indeed exists.

Theorem 2.5 (Wiener, 1923). *For any $x \in \mathbb{R}^d$, the d-dimensional Brownian motion starting in x exists.*

Proof. Note that, having constructed Brownian motion $\{B_t\}_{t \geq 0}$ starting in 0, for any x we have that $\{B_t + x\}_{t \geq 0}$ is a Brownian motion starting in x. Hence, we may w.l.o.g. assume that $x = 0$.
We start by proving existence of standard 1-dimensional Brownian motion. We do this by constructing Brownian motion on the interval $[0, 1]$ and extending it to the whole non-negative line by gluing.
For each $n \in \mathbb{N} \cup \{0\}$, define $\mathcal{D}_n = \{\frac{k}{2^n} \mid 0 \leq k \leq 2^n\}$, and let $\mathcal{D} = \cup_{n=0}^{\infty} \mathcal{D}_n$. Intuitively, our aim is to iteratively construct the right joint distribution for Brownian motion on each $\mathcal{D}_n$, and then obtain $\{B_t\}_{t \geq 0}$ by linearly interpolating the values on each $\mathcal{D}_n$ and check that the uniform limit of these continuous functions exists and is a Brownian motion.
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which $\{Z_t\}_{t \in \mathcal{D}}$ of i.i.d. $N(0, 1)$ random variables can be defined. Define $B_0 = 0$ and $B_1 = Z_1$. For each $d \in \mathcal{D}$, we define random variable B_d such that, for each $n \in \mathbb{N} \cup \{0\}$,

1. for all $r < s < t$ in $\mathcal{D}_n$, $B_t - B_s \sim N(0, t - s)$ and $B_t - B_s$ and $B_s - B_r$ are independent;

2. families of random variables $\{B_d \mid d \in \mathcal{D}_n\}$ and $\{Z_d \mid d \in \mathcal{D} \backslash \mathcal{D}_n\}$ are independent.

To see that such random variables exist, define B_t inductively on each $\mathcal{D}_n$ with B_0 and B_1 given above, and for each $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ define

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

Properties 2 easily follows from induction hypothesis and the fact that Z_d were chosen to be independent. Property 1 is straightforward to verify using induction hypothesis and Lemmas 2.2 and 2.3.
Having defined our stochastic process on $\mathcal{D}$, consider

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

and for each $n \geq 1$

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

These functions are clearly continuous.

<u>Claim 1:</u> for all $n \in \mathbb{N} \cup \{0\}$ and $d \in \mathcal{D}_n$,

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

Second equality is clear from definition of F_i's. We prove first equality by induction on n, with basis case following since $B_0 = 0$, $B_1 = Z_1$. Suppose it is true for $\mathcal{D}_{n-1}$, and let $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$. Then

$$\begin{aligned} B_d &= \frac{B_{d-2^{-n}} + B_{d+2^{-n}}}{2} + \frac{Z_d}{2^{(n+1)/2}} \\ &= \frac{1}{2} \sum_{i=0}^{n-1} F_i(d - 2^{-n}) + \frac{1}{2} \sum_{i=0}^{n-1} F_i(d + 2^{-n}) + \frac{Z_d}{2^{(n+1)/2}} \\ &= \sum_{i=0}^{n-1} \frac{F_i(d - 2^{-n}) + F_i(d + 2^{-n})}{2} + \frac{Z_d}{2^{(n+1)/2}} \\ &= \sum_{i=0}^{n-1} F_i(d) + F_n(d) = \sum_{i=0}^{n} F_i(d) \end{aligned}$$

with last row following by definition of F_i's, thus finishing the proof of Claim 1 by induction.

Claim 2: A.s. the series $\sum_{i=0}^{\infty} F_i$ converges uniformly on $[0,1]$, therefore is a.s. continuous as a uniform limit of continuous functions.
By the Weierstrass M-test, it suffices to show that $\sum_{n=0}^{\infty} ||F_n||_\infty < \infty$. Note that $||F_n||_\infty \leq 2^{-(n+1)/2} \max_{d\in\mathcal{D}_n} |Z_d|$, so we try to bound this maximum. Fix $c > 0$. Since $Z_d \sim N(0,1)$, we have

$$\mathbb{P}[|Z_d| \geq cn] = \frac{1}{\sqrt{2\pi}} \int_{c\sqrt{n}}^{+\infty} e^{-u^2/2} du \leq \frac{1}{\sqrt{2\pi}} \int_{c\sqrt{n}}^{+\infty} \frac{u}{c\sqrt{n}} e^{-u^2/2} du = e^{-c^2 n/2}.$$

Thus, by the union bound

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbb{P}[\text{there exists } d \in \mathcal{D}_n \text{ s.t. } |Z_d| \geq c\sqrt{n}] &\leq \sum_{n=0}^{\infty} \sum_{d\in\mathcal{D}_n} \mathbb{P}[|Z_d| \geq c\sqrt{n}] \\ &\leq \sum_{n=0}^{\infty} (2^n + 1) e^{-cn^2/2},\end{aligned}$$

which converges for $c > \sqrt{2\log 2}$. Picking any such c, by the Borel - Cantelli lemma there is a.s. finite $N \in \mathbb{N}$ such that for any $n > N$ and $d \in \mathcal{D}_n$, we have $|Z_d| < c\sqrt{n}$, implying that $||F_n||_\infty < c\sqrt{n} 2^{-n/2}$. So almost surely

$$\sum_{n=0}^{\infty} ||F_n||_\infty \leq \sum_{n=0}^{N} ||F_n||_\infty + \sum_{n=N+1}^{\infty} c\sqrt{n} 2^{-n/2} < \infty,$$

thus proving the claim.

These two claims imply that a stochastic process on $[0,1]$ defined via $B_t = \sum_{n=0}^{\infty} F_n(t)$ starts at 0, has a.s. continuous trajectories, and satisfies properties 1. and 2. for points in $\mathcal{D}$. In order to conclude that it is a 1-dimensional Brownian motion on $[0,1]$, we need to check that increments have right finite-dimensional distributions.
To verify point (ii) in the definition, pick $0 \leq t_1 < \cdots < t_n \leq 1$. We can pick points $0 \leq t_{1,k} \leq \cdots \leq t_{n,k} \leq 1$ in $\mathcal{D}$ with $t_{i,k} \overset{k\to\infty}{\to} t_i$ for each $1 \leq i \leq n$. By a.s. continuity of trajectories of B, we have that for each $1 \leq i \leq n-1$ almost surely

$$B_{t_{i+1}} - B_i = \lim_{k\to\infty} (B_{t_{i+1,k}} - B_{t_{i,k}})$$

As a.s. convergence of random variables implies convergence in distribution, by looking at limits of mean and variance we conclude that $B_{t_{i+1}} - B_{t_i}$ is a Gaussian and that (argument for taking the limit out of covariance slightly more involved and omitted)

$$\mathbb{E}[B_{t_{i+1}} - B_{t_i}] = \lim_{k\to\infty} (B_{t_{i+1,k}} - B_{t_{i,k}}) = 0,$$

$$\begin{aligned}Cov(B_{t_{i+1}} - B_{t_i}, B_{t_{j+1}} - B_{t_j}) &= \lim_{k\to\infty} Cov(B_{t_{i+1,k}} - B_{t_{i,k}}, B_{t_{j+1,k}} - B_{t_{j,k}}) \\ &= \lim_{k\to\infty} 1_{i=j}(t_{i+1,k} - t_{i,k}) I.\end{aligned}$$

This implies that each $B_{t_{i+1}} - B_{t_i}$ is a Gaussian with mean 0 and variance matrix $(t_{i+1,k} - t_{i,k})I$. Furthermore, as by Lemma 2.2 two Gaussians with covariance matrix 0 are independent, we conclude that all increments are independent. Thus, $\{B_t \mid t \in [0,1]\}$ is indeed a 1-dimensional Brownian motion on $[0,1]$.
We obtain 1-dimensional Brownian motion by gluing together the parts of the above form. Namely, take a sequence $B^0, B^1, B^2, \ldots$ of independent Brownian motions on $[0,1]$ as constructed above. Define a stochastic process B_t via

$$B_t = B^{\lfloor t \rfloor}_{t - \lfloor t - \rfloor} + \sum_{n=0}^{\lfloor t \rfloor - 1} B^n_1.$$

This process is easily seen to have continuous trajectories and to satisfy all finite distribution properties.
Finally, to obtain standard d-dimensional Brownian motion just consider a random vector $(B^1, B^2, \ldots, B^d)$ where each coordinate is a standard 1-dimensional Brownian motion.

□

We note some simple invariance properties of Brownian Motion.

Lemma 2.6 (Scaling invariance)**.** *Let $\{B_t \mid t \geq 0\}$ be a standard d-dimensional Brownian motion. Let $a > 0$, and define $\{X_t \mid t \geq 0\}$ via $X_t = \frac{1}{a} X_{a^2 t}$. Then X is also a d-dimensional standard Brownian motion.*

Proof. Continuity of the paths, independence and stationarity at origin are not influenced by scaling. By Lemma 2.2(ii) we obtain the right distributions of increments, thus concluding the proof. □

Lemma 2.7 (Time inversion)**.** *Let $\{B_t \mid t \geq 0\}$ be a standard d-dimensional Brownian motion. The process $\{X_t \mid t \geq 0\}$ defined by*

$$X_t = \begin{cases} tB_{\frac{1}{t}}, & \textit{if } t > 0 \\ 0, & \textit{otherwise} \end{cases}$$

is also a standard d-dimensional Brownian motion.

Proof. Exercise. □

Corollary 2.7.1 (Strong Law of Large Numbers for BM)**.** $\frac{B_t}{t} \to 0$ *as* $t \to \infty$ *almost surely.*

Proof. Define $\{X_t \mid t \geq 0\}$ as in Time Inversion lemma. Then

$$\lim_{t \to \infty} \frac{B_t}{t} = \lim_{t \to \infty} X_{\frac{1}{t}} = X_0 = 0$$

almost surely (since trajectories are a.s. continuous). □

3 Properties of 1-dimensional Brownian motion

By definition of Brownian motion, trajectories are almost surely continuous, hence almost surely uniformly continuous on $[0,1]$ (or any other compact interval). Fix $\omega \in \Omega$ defining a continuous trajectory, i.e. for which $B_t(\omega)$ is continuous. Then there exists a function $\varphi(\omega) : [0,1] \to \mathbb{R}$ called a **modulus of continuity**, such that $\lim_{h \searrow 0} \varphi(\omega)(h) = 0$ and

$$\limsup_{h \searrow 0} \sup_{0 \le t \le 1-h} \frac{B_{t+h}(\omega) - B_t(\omega)}{\varphi(\omega)(h)} \le 1.$$

However, modulus on continuity depends on the choice of a trajectory (that is, ω). A natural question to ask is if there exists a deterministic choice of φ. The answer to this question is positive, and is given by the following theorem which we do not prove.

Theorem 3.1 (Levy's modulus of continuity)**.** *Almost surely,*

$$\limsup_{h \searrow 0} \sup_{0 \le t \le 1-h} \frac{B_{t+h}(\omega) - B_t(\omega)}{\sqrt{2h \log(1/h)}} = 1.$$

Maybe suprisingly, for every t Brownian motion turns out to be non differentiable almost surely.

Theorem 3.2 (Non-differentiability of Brownian motion)**.** *Let $\{B_t \mid t \ge 0\}$ be a Brownian motion and fix $t \ge 0$. Then a.s. B_t is not differentiable at t.*

For a small flavour of the intuition behind this phenomenon, note that Brownian motion is a.s. non monotone on any interval $[a,b]$, a, $b \in \mathbb{R}^+$.

Proof. Fix any non degenerate interval $[a,b]$. If the interval is an interval of monotonicity, split the interval into n parts $[a_i, a_i+1], i = 0, ...n+1, a = a_0, b = a_{n+1}$. As increments are independent, the probability of the whole interval being monotone is $\frac{2}{2^n}$. Send $n \to \infty$ to make this probability go to 0. $\square$

4 Strong Markov Property

4.1 Markov Property

An interesting property of Brownian Motion is that it is time-shift invariant. Intuitively, this means that we do not need to know about the process for t in an interval of s if we're just interested in the process after a given time $t > s$. We can just pretend the process started from B_s. Stochastic processes which have this quality are sometimes called **memoryless**. Formally, we call this the **Markov Property**.

Theorem 4.1 (Markov Property)**.** *Suppose that $\{B_t\}$ is a Brownian motion started at $x \in \mathbb{R}^d$. Let $s > 0$, then the process $\{B_{t+s} - B_s \mid t \ge 0\}$ is a standard Brownian motion independent of the process $\{B_t \mid 0 \le t \le s\}$.*

Proof. At $t = 0$, $B_{0+s} - B_s = 0$ so this new process clearly starts at 0. Independence of increments and their distributions follow from those of $\{B_t \mid t \geq 0\}$ since it is a Brownian motion. Finally, a.s. continuity of $\{B_t - B_s \mid t \geq s\}$ follows since $\{B_t \mid t \geq 0\}$ is a.s. continuous. $\square$

4.2 Filtration

Defintion 4.2 (Filtration)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A* ***filtration*** *is a family $\{\mathcal{F}_t \mid t \geq 0\}$ of σ-algebras such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ whenever $s \leq t$. We call a probability space with a filtration a* ***filtered probability space****. A stochastic process $\{X_t \mid t \geq 0\}$ on a filtered probability space with filtration $\{\mathcal{F}_t \mid t \geq 0\}$ is called* ***adapted*** *if each X_t is measurable with respect to $\mathcal{F}_t$.*

Example 4.3 (Canonical Filtration)**.** *Let $\{X_t\}$ be a stochastic process on a given probability space. We can define a filtration $\{\mathcal{F}_t \mid t \geq 0\}$ by letting $\mathcal{F}_t = \sigma(X_s | 0 \leq s \leq t)$. This is the σ-algebra generated by the random variables X_s up to some time t. Intuitively, this σ-algebra contains all information from the process up to time t.*

Markov property states that $\{B_{t+s} - B_s \mid t \geq 0\}$ is independent of the process $\{B_t \mid 0 \leq t \leq s\}$. We can use the concept of filtrations to improve this statement and show that $\{B_{t+s} - B_s\}$ is in fact independent of the slightly larger σ-algebra:

$$\mathcal{F}_s^+ = \cap_{t>s} \mathcal{F}_t$$

This σ-algebra is larger than the previous one due to the strict inequality. This gives an infinitesimal glimpse into the future.

Proof. Take a strictly decreasing sequence $\{s_n\}$ converging to s. Then almost surely $\{B_{t+s} - B_s\} = \lim_{n\to\infty}\{B_{t+s_n} - B_{s_n}\}$.

Now take any $t_1, ..., t_m \geq 0$. Almost surely $(B_{t_1+s} - B_s, ..., B_{t_m+s} - B_s) = \lim_{n\to\infty}(B_{t_1+s_n} - B_{s_n}, ..., B_{t_m+s_n} - B_{s_n})$ and this is independent of $\mathcal{F}_s^+$ from the Markov property. So the process $\{B_{t+s} - B_s\}$ is independent of $\mathcal{F}_s^+$ $\square$

A crucial difference between $\{\mathcal{F}^+\}$ and $\{\mathcal{F}\}$ is **right continuity**, which gives

$$\cap_{\epsilon>0}\mathcal{F}_{t+\epsilon}^+ = \cap_{n=1}^{\infty} \cap_{k=1}^{\infty} \mathcal{F}_{t+1/n+1/k} = \mathcal{F}_t^+ = \cap_{\epsilon>0}\mathcal{F}_{t+\epsilon}$$

The σ-algebra $\mathcal{F}_0^+$ has a special name called **germ σ-algebra**. This comprises all events defined in terms of Brownian motion on an infinitesimally small interval to the right of the origin. The concept of germ σ-algebra is used to prove 2 important 0-1 laws in probability theory

4.3 0-1 laws

Theorem 4.4 (Blumenthal 0-1 law)**.** *Let $x \in \mathbb{R}^d$ be the starting point of some Brownian Motion. Consider $\mathcal{A} \in \mathcal{F}_0^+$. Then $\mathbb{P}(A) \in \{0, 1\}$.*

Proof. Since $A \in \sigma(B_t \mid t \geq 0)$, by the above extension of Markov property we know that $\mathcal{A}$ is independent of $\mathcal{F}_0^+$. In particular, $\mathcal{A}$ is also $\in \mathcal{F}_0^+$ so it is independent of itself. Thus $\mathbb{P}(A) = \mathbb{P}(A) \cdot \mathbb{P}(A)$ and so $\mathbb{P}(A) \in \{0, 1\}$. □

Let $\mathcal{G}_t := \sigma(B_s \mid s \geq t)$ and let $\mathcal{T} = \cap_{t \geq 0} \mathcal{G}_t$ be the **tail σ-algebra**, that is, the σ-algebra of all tail events. We have the following theorem for tail events:

Theorem 4.5 (Tail 0-1 law)**.** *Let $x \in \mathbb{R}^d$ be the starting point of some Brownian Motion. Consider $\mathcal{A} \in \mathcal{T}$. Then $\mathbb{P}(A) \in \{0, 1\}$.*

Proof. By time inversion, we can map the tail σ-algebra to the germ σ-algebra of time inversed Brownian motion. This is done by letting B_s be Brownian motion and defining a new Brownian motion $sW_{\frac{1}{s}} = B_s$. Then

$$\mathcal{G}_t = \sigma(sW_{\frac{1}{s}} \mid s \geq t) = \sigma(W_{\frac{1}{s}} \mid s \geq t) = \sigma(W_s \mid s \leq \frac{1}{t})$$

But the intersection of these becomes the germ of W_s. Applying Blumenthal 0-1 Law to it, we conclude the proof. □

4.4 Stopping time

Defintion 4.6 (Stopping time)**.** *A random variable T with values in $[0, \infty]$ on a probability space with filtration $\{\mathcal{F}_t \mid t \geq 0\}$ is called a* ***stopping time*** *if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.*

Intuitively, we call a random time T a stopping time if we can know whether $T \leq t$ occurs only if we know the stochastic process up to point t.

Example 4.7 (Trivial stopping time)**.** *Every deterministic time t is a stopping time w.r.t. every filtration.*

Example 4.8 (First hitting time of closed set)**.** *Suppose H is closed and define $T := inf\{t \geq 0 \mid B_t \in H\}$. WLOG we can consider H to be a singleton. Then T is a stopping time w.r.t. $\{\mathcal{F}_t\}$.*

Indeed we note that

$$\{T \leq t\} = \cap_{n=1}^{\infty} \cup_{s \in \mathbb{Q} \cap (0,t)} \cup_{x \in \mathbb{Q}^d \cap H} \{B_s \in B(x, \frac{1}{n})\} \in \mathcal{F}_t$$

It is common to work with $\mathcal{F}_t^+$ instead of $\mathcal{F}_t$ because because $\mathcal{F}_t^+$ is bigger than $\mathcal{F}_t$ and we get more stopping times. The example below highlights this:

Example 4.9 (First hitting time of open set)**.** *Suppose $G \in \mathbb{R}^d$ open and define $T := inf\{t \geq 0 \mid B_t \in G\}$. T is a stopping time w.r.t. $\{\mathcal{F}_t^+\}$ but not $\{\mathcal{F}_t\}$.*

Observe that by right continuity,

$$\{T \leq t\} = \cap_{s > t} \{T \leq s\} = \cap_{s > t} \cup_{r \in \mathbb{Q} \cap (0,s)} \{B_r \in G\} \in \mathcal{F}_t^+$$

so T is a stopping time w.r.t. $\{\mathcal{F}^+\}$. To see why T is not a stopping time w.r.t $\{\mathcal{F}\}$, suppose G is bounded and B_0 is not in $\overline{G}$. So fix a path $\gamma : [0, t] \to \mathbb{R}^d$ s.t. $\gamma(0) \cap \overline{G} = \emptyset$ and $\gamma(t) \in \partial G$. At t, the σ-algebra $\mathcal{F}_t$ contains only 2 subsets of $\{B_s = \gamma(s) \ \forall 0 \leq s \leq t\}$ which is the empty set and the set itself. But if $\{T \leq t\} \in \mathcal{F}_t$, then the set $\{B_s = \gamma(s) \ \forall 0 \leq s \leq t \, T = t\}$ will also be in $\mathcal{F}_t$. As this is also another subset of $\{B_s = \gamma(s) \ \forall 0 \leq s \leq t\}$, we get a contradiction.

From now on all stopping times will by assumption refer to the probability space with filtration $\{\mathcal{F}^+\}$, unless otherwise stated.

We will state one more useful fact of stopping times as a lemma.

Lemma 4.10. *Consider $\{T_n\}$ an increasing sequence of stopping times defined on a probability space with filtration $\mathcal{F}_t^+$ and converging to T. Then $\{T \leq t\} \in \mathcal{F}_t^+$.*

Proof. $\{T \leq t\} = \cap_{n=1}^{\infty} \{T_n \leq t\} \in \mathcal{F}_t^+$ □

4.5 Strong Markov Property

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{B_t \mid t \geq 0\}$ be a Brownian motion and define the filtration $\mathcal{F}^+$ as before. Let T be a stopping time w.r.t. filtration $\mathcal{F}^+$. Recall from last lecture, whenever we talk about Brownian motion and stopping times, unless otherwise stated we always mean that stopping time is defined w.r.t. this filtration. Define $\mathcal{F}_T^+$ the following σ-algebra

$$\mathcal{F}_T^+ = \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_t^+ \ \forall t \geq 0\}.$$

Intuitively, this means that for every A, the part of A contained in $T \leq t$ is measurable with respect to information available to our Brownian motion right after time t. Thus we can think of it as a set of events generated by this Brownian motion before reaching stopping time T.
We now state and prove the Strong Markov Property, which has many beautiful and insightful application of which unfortunately we can mention only a few. Note that up until now, we have proved that for a fixed time $s \geq 0$, the process $\{B_{t+s} - B_s \mid t \geq 0\}$ is a standard Brownian motion independent of $\mathcal{F}_s^+$. Strong Markov Property generalizes this result, in the sense that it is just a special case when the stopping time is deterministic.

Theorem 4.11 (Strong Markov Property)**.** *For every a.s. finite stopping time T with respect to some filtration $\{\mathcal{F}_t \mid t \geq 0\}$, the process $\{B_{T+t} - B_T \mid t \geq 0\}$ is a standard Brownian motion independent of $\mathcal{F}_T^+$.*

Proof. The proof strategy is as follows: we approximate T from above with a sequence of discrete stopping times T_n for which we prove the claim, and then conclude it for T.
For each $n \in \mathbb{N}$, define a stopping time T_n

$$T_n = (m+1)2^{-n} \text{ whenever } m2^{-n} \leq T < (m+1)2^{-n}.$$

We clearly have $T_n > T$. Also, each T_n is a stopping time w.r.t. $\{\mathcal{F}_t \mid t \geq 0\}$ since for each $t \geq 0$, if $m2^{-n} \leq t < (m+1)2^{-n}$ then

$$\{T_n \leq t\} = \{T_n \leq m2^{-n}\} = \{T < m2^{-n}\} \subseteq \mathcal{F}_{m2^{-n}} \subseteq \mathcal{F}_t.$$

We now prove the claim for each T_n. Fix $n \in \mathbb{N}$. Write $B^k = \{B_t^k \mid t \geq 0\}$ for the Brownian motion defined by $B_t^k = B_{t+k/2^n} - B_{k/2^n}$, and $B^* = \{B_t^* \mid t \geq 0\}$ for the process $B_t^* = B_{t+T_n} - B_{T_n}$. Suppose $E \in \mathcal{F}_{T_n}^+$. Then for any event $A \in \mathcal{F}$, due to discreteness of T_n and finiteness of T, hence T_n,

$$\begin{aligned}\mathbb{P}[\{B^* \in A\} \cap E] &= \sum_{k=0}^{\infty} \mathbb{P}[\{B^k \in A\} \cap E \cap \{T_n = k2^{-n}\}] \\ &= \sum_{k=0}^{\infty} \mathbb{P}[B^k \in A]\mathbb{P}[E \cap \{T_n = k2^{-n}\}],\end{aligned}$$

using that the Brownian motion B^k is independent of $\mathcal{F}_{k2^{-n}}^+$ for each k. By Markov property, $\mathbb{P}[\{B^k \in A\}]$ does not depend on k so we may just write $\mathbb{P}[\{B \in A\}]$, and get

$$\begin{aligned}\mathbb{P}[\{B^* \in A\} \cap E] &= \sum_{k=0}^{\infty} \mathbb{P}[B^k \in A]\mathbb{P}[E \cap \{T_n = k2^{-n}\}] \\ &= \mathbb{P}[\{B \in A\}] \sum_{k=0}^{\infty} \mathbb{P}[E \cap \{T_n = k2^{-n}\}] \\ &= \mathbb{P}[\{B \in A\}]\mathbb{P}[E].\end{aligned}$$

Taking E to be the whole space, we conclude that all finite distributions of B^* agree with those of standard Brownian motion. Since a.s. continuity of trajectories follows from those of the original Brownian motion, we conclude that B^* is a standard Brownian motion. Since $E \in \mathcal{F}_{T_n}^+$ was arbitrary, we conclude that B^* is independent of $\mathcal{F}_{T_n}^+$.
It remains to prove the claim for T. Since $T_n \searrow T$, $\{B_{T_n+t} - B_{T_n} \mid t \geq 0\}$ is a Brownian motion independent of $\mathcal{F}_T^+ \subseteq \mathcal{F}_{T_n}^+$. Since a.s. convergence of random variables implies convergence in distribution, we conclude that the increments $B_{s+t+T} - B_{t+T}$ have the right distributions and that they are independent. Also, a.s. continuity of trajectories follows from those of the original Brownian motion, therefore we conclude that $\{B_{t+T} - B_T \mid t \geq 0\}$ is indeed a standard Brownian motion. Finally, as a limit of random variables independent of $\mathcal{F}_T^+$ we conclude that this process is also independent of $\mathcal{F}_T^+$, thus finishing the proof. □

Next, we present one nice application of the Strong Markov Property.

Theorem 4.12 (Reflection principle)**.** *Let $\{B_t \mid t \geq 0\}$ be a standard Brownian motion and T be an a.s. finite stopping time. Then the process $\{B_t^* \mid t \geq 0\}$ defined via*

$$B_t^* = B_t 1_{t \leq T} + (2B_T - B_t)1_{t > T}$$

is again a standard Brownian motion and is called the ***Brownian motion reflected at*** T.

Proof. Suppose w.l.o.g. T is finite (redefining it on a set of probability zero does not influence finite distributions or a.s. continuity of trajectories). By Strong Markov Property, both processes $\{B_{t+T} - B_T \mid t \geq 0\}$ and $\{-(B_{t+T} - B_T) \mid t \geq 0\}$ are standard Brownian motions and independent of $\{B_t \mid t \leq T\}$. Hence process obtained by gluing each of these to the end of $\{B_t \mid t \leq T\}$ will have same distributions (it is straightforward to check processes obtained by gluing are indeed measurable, moreover they are clearly both a.s. continuous at each point). But the first process is just our original standard Brownian motion, whereas the second obtained process is

$$B_t^* = B_t 1_{t \leq T} + (2B_T - B_t) 1_{t > T},$$

thus this proves $\{B_t^* \mid t \geq 0\}$ is also standard Brownian motion. □

5 Skorokhod Embedding Problem

In 1961, Ukrainian mathematician Anatoliy Skorokhod formulated and solved an extremely interesting problem concerned with embedding an increment of some 1-dimensional random walk into a given Brownian motion.
Consider a random walk on the real line (w.l.o.g. starting in 0) such that at each point the increment has distribution of a given real valued random variable X. We would like to embed this random walk into standard linear Brownian motion, so we could use general results on Brownian motion to analyze 1-dimensional random walks. Given a standard linear Brownian motion $\{B_t \mid t \geq 0\}$, our aim is to find a stopping time T such that B_T (that is, our Brownian motion stopped at T) has the distribution of X. Therefore, positions in random walk would reduce to the sequence $0, B_T, B_{2T}, \ldots$.

Regarding the class of (from now on always 1-dimensional) random walks we could potentially analyze this way, we immediately run into a necessary condition imposed just by properties of linear Bronian motion. For any integrable stopping time T, by Wald's lemma we have that

$$\mathbb{E}[B_T] = 0 \qquad \text{and} \qquad \mathbb{E}[B_T^2] = \mathbb{E}[T] < \infty,$$

so if we restrict to considering integrable stopping time (which intuitively makes sense), we limit ourselves to random walks with increment X having mean 0 and finite variance. It turns out that this is also sufficient, and is reflected in the following theorem which is the main result of this section.

Theorem 5.1 (Skorokhod Embedding Theorem)**.** *Suppose that $\{B_t \mid t \geq 0\}$ is a standard linear Brownian motion, and let X be a real valued random variable with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$. Then there exists a stopping time T (w.r.t. the natural filtration of the Brownian motion) such that B_T has the law of X and $\mathbb{E}[T] = \mathbb{E}[X^2]$.*

Since Wald's lemmas essentially rely on Optional Stopping Theorem and martingale properties of Brownian motion, due to time constraints we omit their proofs. In what follows, we give a constructive proof of the Skorokhod Embedding Theorem through the following result, yielding a specific construction for T.

Theorem 5.2 (Azema-Yor Embedding Theorem)**.** *Suppose that X is a real valued random variable with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$. Let*

$$\Psi(x) = \begin{cases} \mathbb{E}[X \mid X \geq x], & \textit{if } \mathbb{P}[X \geq x] > 0 \\ 0, & \textit{otherwise.} \end{cases}$$

Suppose $\{B_t \mid t \geq 0\}$ is a standard linear Brownian motion and define a process $\{M_t \mid t \geq 0\}$ via $M_t = \sup_{0 \leq s \leq t} B_s$. Define a stopping time τ via

$$\tau = \inf\{t \geq 0 \mid M_t \geq \Phi(B_t)\}.$$

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

To prove Azema-Yor Embedding Theorem, we proceed in three steps (pictures provided in actual lectures):

5.1 Step 1: Skorokhod Embedding Theorem for discrete random variables

We firstly prove the Skorokhod Embedding Theorem for discrete real valued random variables with a stopping time which is not yet clear to coincide with the one from Azema-Yor Embedding Theorem.

Lemma 5.3. *Suppose that X is a real valued random variable with $\mathbb{E}[X] = 0$ and support in a finite set $x_1 < x_2 < \cdots < x_n$. Define $y_i = \Phi(x_{i+1})$, thus $y_1 < y_2 < \cdots < y_{n-1}$. Furthermore, define stopping times $T_0 = 0$ and*

$$T_i = \inf\{t \geq T_{i-1} \mid B(t) \notin (x_i, y_i)\}$$

for $i \leq n-1$. Then $T = T_{n-1}$ satisfies $\mathbb{E}[T] = \mathbb{E}[X^2]$ and B_T has the same law as X.

In order to prove this, we will actually need the general versions of two Wald's lemmas.

Lemma 5.4 (Wald's first lemma)**.** *Let $\{B_t \mid t \geq 0\}$ be a standard linear Brownian motion, and T a stopping time such that either $\mathbb{E}[T] < \infty$ or the set of random variables $\{B_{t \wedge T} \mid t \geq 0\}$ is dominated by an integrable function. Then $\mathbb{E}[B_T] = 0$.*

Lemma 5.5 (Wald's second lemmma)**.** *Let $\{B_t \mid t \geq 0\}$ be a standard linear Brownian motion, and T a stopping time such that $\mathbb{E}[T] < \infty$. Then $\mathbb{E}[B_T^2] = \mathbb{E}[T]$.*

Lemma 5.6. *Let $a < 0 < b$ and, for a standard linear Brownian motion $\{B_t \mid t \geq 0\}$, define $T = \min\{t \geq 0 \mid B_t \in \{a, b\}\}$. Then T is a stopping time with*

- $\mathbb{P}[B_T = a] = \frac{b}{|a|+b}$ *and* $\mathbb{P}[B_T = b] = \frac{|a|}{|a|+b}$;
- $\mathbb{E}[T] = |a|b$.

Proof. First part is an exercise in the example sheet. For the second part, we wish to apply second Wald's lemma. The claim easily follows and is left as an exercise once we prove that $\mathbb{E}[T] < \infty$:

$$\begin{aligned}
\mathbb{E}[T] = \int_0^\infty \mathbb{P}[T > t] dt &= \int_0^\infty \mathbb{P}\Big[B_s \in (a, b)\ \forall s \in [0, t]\Big] dt \\
&\leq \sum_{k=0}^\infty \mathbb{P}\Big[B_s \in (a, b)\ \forall s \in [0, k]\Big] \\
&\leq \sum_{k=0}^\infty \mathbb{P}\Big[B_j \in (a, b)\ \forall j \in \{0, 1, \dots, k\}\Big] \\
&\leq \sum_{k=0}^\infty \prod_{j=1}^k \mathbb{P}\Big[B_j \in (a, b) \mid B_{j-1} \in (a, b)\Big] \\
&\leq \sum_{k=0}^\infty \mathbb{P}\Big[B_1 \in (2a, 2b)\Big]^k < \infty.
\end{aligned}$$

□

Proof of Lemma 5.3. Note that for every i $y_i \geq x_{i+1}$ and equality holds iff $i = n-1$. By Lemma 5.6, $\mathbb{E}[T_{n-1}] < \infty$, so by Second Wald's lemma $\mathbb{E}[B_{T_{n-1}}^2] = \mathbb{E}[T_{n-1}]$. For $i = 1, \dots, n-1$, define random variables

$$Y_i = \begin{cases} \mathbb{E}[X \mid X \geq x_{i+1}], & \text{if } X \geq x_{i+1}, \\ X, & \text{if } X \leq x_i. \end{cases}$$

We firstly observe some properties of Y_i's conditioned on X taking some values. Firstly, Y_1 has expectation 0 and takes on two values x_1 and y_1 with non-zero probability. For $i \geq 2$, given $Y_{i-1} = y_{i-1}$ (i.e. $X \geq x_i$), Y_i takes values x_i and y_i with non-zero probability and has expectation y_{i-1}. Given $Y_{i-1} = x_j$ for $j \leq i-1$ (thus $X = x_j$), we have $Y_j = x_j$. Finally, $Y_{n-1} = X$.
Claim.

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

To prove the claim, observe that it suffices to show that these two probability measures agree on products of measurable sets in $\mathbb{R}$ (as they form a π-system generating the product σ-algebra). For a set $A_1 \times \cdots \times A_{n-1}$ in the product

σ-algebra,

$$\mathbb{P}[(B_{T_1},\dots,B_{T_{n-1}}) \in A_1 \times \cdots \times A_{n-1}] = \prod_{i=1}^{n-1} \mathbb{P}[B_{T_i} \in A_i \mid B_{T_{i-1}} \in A_{i-1}] \cdot \mathbb{P}[B_{T_1} \in A_1]. \tag{1}$$

Hence, it suffices to check that B_{T_1} has the same distribution as Y_1 and that each B_{T_i} conditioned on $B_{T_{i-1}}$ has the same distribution as Y_i conditioned on Y_{i-1}.

By a.s. continuity of trajectories, B_{T_1} can only take values x_1 and y_1 with non-zero probability. By Lemma 5.6 $\mathbb{E}[T_1] < \infty$ hence by Wald's first lemma $\mathbb{E}[B_{T_1}] = 0$. From this we can compute probabilities of B_{T_1} being equal to x_1 and y_1, and thus conclude that it has the same distribution as Y_1.

For $i \geq 2$, proceed by induction with basis given above (for T_1 and T_0). Then by induction hypothesis and the same argument as in equation (1), we have that $B_{T_{i-1}}$ has the same distribution as Y_{i-1}, and can thus take values $x_1, x_2, \dots, x_{i-1}$ and y_{i-1}. By writing

$$\begin{aligned}\mathbb{P}[B_{T_i} \in A_i \mid B_{T_{i-1}} \in A_{i-1}] &= \frac{\mathbb{P}[B_{T_i} \in A_i, B_{T_{i-1}} \in A_{i-1}]}{\mathbb{P}[B_{T_{i-1}} \in A_{i-1}]} \\ &= \sum_{a \in A_{i-1}} \frac{\mathbb{P}[B_{T_i} \in A_i, B_{T_{i-1}} = a]}{\mathbb{P}[B_{T_{i-1}} \in A_{i-1}]},\end{aligned}$$

since $B_{T_{i-1}}$ has the same distribution as Y_{i-1} to prove (1) we may w.l.o.g. assume A_{i-1} is a singleton and do analogous analysis as for Y_i's. Consider the following two cases:

- $B_{T_{i-1}} = y_{i-1}$. Then B_{T_i} can take both values x_i and y_i and by Lemma 5.6 has expectation y_{i-1}. Therefore it has the same distribution as Y_i conditioned on $Y_{i-1} = y_{i-1}$.
- $B_{T_{i-1}} = x_j$ for $j \leq i-1$. Then as we start from a point smaller than x_i so we are already outside the interval (x_i, y_i), have $B_{T_i} = x_j$ concluding it again has the same distribution as Y_i conditioned on $Y_{i-1} = x_j$.

This proves the claim. But then $B_T = B_{T_{n-1}}$ has the same distribution as $Y_{n-1} = X$. Moreover, by Lemma 5.6 T is integrable hence by Wald's second lemma we have $\mathbb{E}[T] = \mathbb{E}[B_T^2] = \mathbb{E}[X^2]$. This finishes the proof. $\square$

5.2 Step 2: Identifying stopping time with that in Azema-Yor Embedding Theorem

We show that the stopping time T constructed in the previous step agrees with the stopping time τ in the Azema-Yor embedding. This would complete the proof of the Azema-Yor Embedding Theorem for random variables with finite support.

Lemma 5.7. *The stopping time τ in the Azema-Yor Embedding Theorem and the stopping time T in Lemma 5.3 are equal.*

Proof. We prove that $\tau \leq T$ and $\tau \geq T$, hence conclude the claim. Suppose that $B_{T_{n-1}} = x_i$ for some $1 \leq i \leq n$, hence $\Phi(B_{T_{n-1}}) = y_i$.

- $\tau \leq T$: Suppose first that $i \leq n-1$. Then as $x_i \leq x_{n-1} < y_{n-1} = x_n$, have that i is the smallest index s.t. $B_{T_{n-1}} = \cdots = B_{T_i}$. So $B_{T_{i-1}} \neq B_{T_i}$. Hence cannot have $B_{T_{i-1}} \leq x_{i-1}$, thus $B_{T_{i-1}} \geq y_{i-1}$ and so $M(T_{n-1}) \geq y_{i-1} = \Phi(x_i) = \Phi(B_{T_{n-1}})$. Thus $\tau \leq T$.
  If $i = n$, we have $M(T_{n-1}) = x_n \geq \Phi(B_{T_{n-1}})$ by maximality of x_n, hence $\tau \leq T$.

- $\tau \geq T$: Suppose $t < T_{n-1}$, we show that $t < \tau$. As $T_0 = 0$ there is $1 \leq i \leq n-1$ such that $T_{i-1} \leq t < T_i$. In particular, this means that the Brownian motion has not exceeded y_i up to time t as otherwise we would have $t \geq T_i$, therefore $M(t) < y_i$. Also, as $t < T_i$ have $B_{T_i} \in (x_i, y_i)$, thus $\Phi(B_t) \geq y_i > M(t)$. Thus $\tau > t$, as wanted.

□

5.3 Step 3: Deducing Skorokhod Embedding Theorem for the general case

Given a real valued random variable X with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$, we aim to construcy a sequence of discrete real random variables with finite support $\{X_n\}_{n\in\mathbb{N}}$ such that:

1. $\forall n \in \mathbb{N}$, $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] < \infty$;
2. $X_n \to X$ in distribution;
3. for $\Phi_n(x) = \mathbb{E}[X_n \mid X_n \geq x]$, the embedding stopping times
$$\tau_n = \inf\{t \geq 0 \mid M(t) \geq \Phi_n(B_t)\}$$
almost surely converge to τ.

If we do so, since trajectories of B are a.s. continuous and $\tau_n \to \tau$ a.s. we conclude that $B_{\tau_n} \to B_\tau$ a.s. and thus in distribution as well. But by Steps 1 and 2 we know that B_{τ_n} has the same distribution as X_n, and limit of convergence in distribution is unique, hence B_τ has the same distribution as X. Moreover, as $\tau_n \to \tau$ a.s. by Fatou's lemma (and dominated convergence with dominating function $(X+1)^2$ in the last inequality)

$$\mathbb{E}[\tau] \leq \liminf_{n\to\infty} \mathbb{E}[\tau_n] = \liminf_{n\to\infty} \mathbb{E}[X_n^2] < \infty.$$

Hence, by Wald's second lemma $\mathbb{E}[\tau] = \mathbb{E}[B_\tau^2] = \mathbb{E}[X^2]$, thus finishing the proof of the Skorokhod Embedding Theorem.

Thus it remains to construct such sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables. For each $n \in \mathbb{N}$, define

$$\mathcal{D}_n = \{\frac{k}{2^n} \mid k \in \{-n2^n, -n2^n + 1, \dots, n2^n\}\}.$$

Denote $x_k = \frac{k-n2^n-1}{2^n}$ for each $1 \le k \le n2^{n+1}+1$, and let $x_{n2^n+2} = \infty$. For each $1 \le i \le n2^{n+1} + 1$, let $y_i = \mathbb{E}[X \mid x_i \le X < x_{i+1}]$ and $p_i = \mathbb{P}[x_i \le X < x_{i+1}]$. Finally, define a random variable X_n to take values in the set $\{y_i \mid 1 \le i \le 2^{n+1} + 1\}$ with $\mathbb{P}[X_n = y_i] = p_i$ for each i.
Simple summation gives $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] < \infty$ for each $n \in \mathbb{N}$. Moreover, since mesh of $\mathcal{D}_n$ converges to 0 we have $X_n \to X$ in distribution. Finally, as for each x $\Phi_n(x) \to \Phi(x)$, from a.s. continuity of B it follows that $\tau_n \to \tau$ almost surely, finishing our construction.

6 Donsker invariance principle

Donsker invariance principle is named after American mathematician Monroe Donsker and is commonly referred to as the "functional Central Limit Theorem". This is because the convergence in distribution we get does not depend on the distribution of the random variables.

Let $\{X_i\} \overset{iid}{\sim} N(0,1)$. Define the random walk $S_n := \sum_{i=1}^n X_i$. We can interpolate linearly between the S_n by doing the following:

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)(S_{\lfloor t+1 \rfloor} - S_{\lfloor t \rfloor})$$

$S(t)$ is now a random function taking values in $\mathbf{C}[0,\infty]$. Recall that for any random variable Y with finite first and second moment, one can **normalise** the random variable to have mean 0 and variance 1 by doing:

$$\frac{Y - \mathbb{E}[Y]}{\sqrt{Var[Y]}}$$

Since a sum of n independent Gaussians with mean 0 and unit variance is also Gaussian with mean 0 and variance n, we can obtain another random function in $\mathbf{C}[0,1]$ from this function in $\mathbf{C}[0,\infty]$ by looking at the **normalised random walks** and interpolating them

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

Theorem 6.1 (Donsker's invariance principle)**.** *On the space of* $\mathbf{C}[0,1]$ *continuous functions on the unit interval with the sup-norm metric, the sequence* $\{S_n^* \mid n \ge 1\}$ *converges in distribution to a standard Brownian motion* $\{B_t \mid t \in [0,1]\}$.

Intuitively, this means that the S_n^* has the same trajectories as standard Brownian motion on $[0,1]$. The idea of the proof is to construct some variables

X_i on the same probability space as the Brownian motion such that when we take S_n^* from these variables, S_n^* is extremely close (in probability) to a scaled Brownian motion.

The proof of Donsker's invariance principle is straightforward once we prove the following lemma:

Lemma 6.2. *Suppose $\{B_t \mid t \geq 0\}$ is a 1 dimensional Brownian motion. Then for any random variable X with $\mathbb{E}[X] = 0$ and $Var[X] = 1$, there exists a sequence of stopping times*

$$0 = T_0 \leq T_1 \leq T_2 \leq ...$$

w.r.t. the Brownian motion s.t.

a) the sequence $\{B_{T_n} \mid n \geq 0\}$ has the distribution of the random walk with increments given by the law of X.

b) the sequence $\{S_n^* \mid n \geq 0\}$ constructed from this random walk satisfies

$$\lim_{n\to\infty} \mathbb{P}\{\sup_t \mid \frac{B_{nt}}{\sqrt{n}} - S_n^*(t)| \ > \epsilon\} = 0$$

Proof. Since $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$, we can use Skorokhod embedding theorem to show that there exists a stopping time T s.t. $\mathbb{E}[T] = \mathbb{E}[X^2] = 1$ and $\text{law}(B_T) = \text{law}(X)$. We can just define $T_1 = T$.

By strong Markov property, $\{B_t^2 \mid t \geq 0\} := \{B_{T_1+t} - B_{T_1} \mid t \geq 0\}$ is a Brownian motion independent of $\mathcal{F}^+_{T_1}$. Again using Skorokhod on B_t^2, we obtain another stopping time T' s.t. $\mathbb{E}[T'] = 1$ and $\text{law}(B^2_{T'}) = \text{law}(X)$. Then $T_2 = T_1 + T'$ becomes another stopping time for the original Brownian motion with $\mathbb{E}[T_2] = 2$ and B_{T_2} is the second value in the random walk with increments given by X.

Proceed analogously to get a sequence $0 = T_0 \leq T_1 \leq T_2 \leq ...$ s.t. $S_n = B_{T_n}$ and $\mathbb{E}[T_n] = n$. This proves part a).

Now for b), denote $W_n(t) := \frac{B_{nt}}{\sqrt{n}}$. From scaling invariance property, this is also standard Brownian motion. Let $A_n = \{\exists t \in [0,1) \mid |S_n^*(t) - W_n(t)| \geq \epsilon\}$. Need to show: $\mathbb{P}(A_n) \to 0$.

Let k be the integer s.t. $(k-1)/n \leq t < k/n$. Since S_n^* was constructed by linear interpolation, we have

$$A_n \subset \{\exists t \in [0,1) \mid |S_k/\sqrt{n} - W_n(t)| > \epsilon\} \cup \{\exists t \in [0,1) \mid |S_{k-1}/\sqrt{n} - W_n(t)| > \epsilon\}$$

But since $S_k = B_{T_k} = \sqrt{n} W_n(T_k/n)$, we can rewrite the above as

$$A_n \subset \{\exists t \in [0,1) \mid |W_n(T_k/n) - W_n(t)| > \epsilon\} \\ \cup \{\exists t \in [0,1) \mid |W_n(T_{k-1}/n) - W_n(t)| > \epsilon\} =: A_n^*$$

Since for each $k \in \mathbb{N}$ $\{T_k - T_{k-1}\}$ is a random variable with mean 1, Strong law of large numbers implies that

$$\lim_{n\to\infty} \frac{T_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (T_k - T_{k-1}) = 1$$

And using a known fact,

$$\lim_{n\to\infty} \frac{T_n}{n} = 1 \implies \lim_{n\to\infty} \sup_{0\le k\le n} \frac{|T_k - k|}{n} = 0$$

Therefore we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\sup_{0\le k\le n} \frac{|T_k - k|}{n} \ge \delta \Big) = 0 \tag{2}$$

We want to make use of (2) to show $\mathbb{P}(A_n^*) \to 0$. By inspection, one can see that A_n^* is contained within

$$\{\exists s, t \in [0,2] \mid |s-t| < \delta, W_n(s) - W_n(t) > \epsilon\} \cup \{\exists t \in [0,1) \mid |\frac{T_k}{n} - t| \vee |\frac{T_{k-1}}{n} - t| \ge \delta\}.$$

Thus, fixing $\delta > 0$, it suffices to show that the probability of the second term $\mathbb{P}\{\exists t \in [0,1) \mid |\frac{T_k}{n} - t| \vee |\frac{T_{k-1}}{n} - t| \ge \delta\}$ tends to 0 as $n \to \infty$. If we do this, letting $\delta \to 0$ would imply the claim.

Let $n \ge 2/\delta$. Then we have

$$\begin{aligned}
&\mathbb{P}\{\exists t \in [0,1) \mid |\frac{T_k}{n} - t| \vee |\frac{T_{k-1}}{n} - t| \ge \delta\} \\
&\qquad \le \mathbb{P}\{ \sup_{1\le k\le n} \frac{(T_k - (k-1)) \vee (k - T_{k-1})}{n} \ge \delta\} \\
&\qquad \le \mathbb{P}\{ \sup_{1\le k\le n} \frac{T_k - k}{n} \ge \frac{\delta}{2}\} + \mathbb{P}\{ \sup_{1\le k\le n} \frac{k - 1 - T_{k-1}}{n} \ge \frac{\delta}{2}\}
\end{aligned}$$

By equation (2), letting $n \to \infty$ both of these tend to 0.

□

Finally, we are ready to deduce the Donsker's invariance principle.

Proof of Donsker's Invariance Principle. Choose stopping times as in the above lemma. Suppose $K \in \mathbf{C}[0,1]$ is closed. Define an ϵ-fattening of the set K as

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

Then $\mathbb{P}\{S_n^* \in K\} \le \mathbb{P}\{W_n \in K_\epsilon\} + \mathbb{P}\{\|S_n^* - W_n\|_\infty > \epsilon\}$. From lemma, we know the second term goes to 0 as $n \to \infty$. As K is closed,

$$\lim_{\epsilon \searrow 0} \mathbb{P}\{W_n \in K_\epsilon\} = \mathbb{P}\{W_n \in \cap_{\epsilon>0} K_\epsilon\} = \mathbb{P}\{W_n \in K\}$$

One of the conditions in the Portmanteau theorem states that $X_n \xrightarrow{d} X \iff \limsup_{n\to\infty} \mathbb{P}\{X_n \in K\} \leq \mathbb{P}\{X \in K\}$ for all closed sets K. Since we get $\limsup_{n\to\infty} \mathbb{P}\{S_n^* \in K\} \leq \mathbb{P}\{W_n \in K\}$, we can use Portmanteau and complete the proof. $\square$

Acknowledgement

The authors would like to thank Jan Maas and Uli Wagner for help in the preparation of this manuscript.

References

[1] Peter Morters and Yuval Peres. Brownian Motion. Cambridge: Cambridge University Press, 2010