

Stochastic Processes notes

Kotatsu

vaffanculo

Preface

These are the notes of the Stochastic Processes course for the Academic Year 2024-2025 with Professor Sacerdote.

I took these notes personally and I integrated the (many) unclear parts and passages using online resources and occasionally the fucking shitty books that this course has.

I am not particularly enthusiast about this course but, as a person with no mathematical background from bachelor degree, I firmly believe in the necessity to understand what you are fucking doing and therefore I went out of my way to make these notes as understandable as possible for other wretched people that have not taken a single analysis exam in their whole life and now have to face integral equations. I recommend doing the same and yes, I know that “yOu CaNnOt UnDeRsTaNd EvErYtHiNg! ThIs Is A mAtHeMaThIcS dEgReE” ok shut up. I need to have at least a vague intuition of what I am doing or otherwise I WILL tune out and start thinking about murderous and mischievous tomfoolery.

There are also exercises. They should be mostly correct, but of course you don’t need to be told that these notes are not redacted by a professional and therefore are subject to errors and mistypes. But, again, this is not too far from the experience in class with Prof. Sacerdote. If I pass this exam I will throw a fucking party.

So let’s have a fucking party!

Kotatsu

Contents

1	Brownian Motion	2
1.1	Continuity of stochastic processes	2
1.2	Definitions of Brownian motion	6
1.2.1	Historical context	6
1.2.2	Random walk model	8
1.2.3	Definitions	11
Exercises		13
1.3	Brownian motion as a Gaussian processes	16
1.3.1	Properties of Gaussian Processes	16
1.3.2	Gaussianity of Brownian motion	19
1.4	Existence of Brownian motion	22
1.5	Features of Brownian motion	23
1.6	Martingales related with Brownian motion	27
1.6.1	Three golden martingales	27
1.6.2	Boundaries of BM	31
Exercises		37
1.7	Brownian motion as a Markov Process	42
1.7.1	Markov Property	42
1.7.2	First passage times and Strong Markov Property	45
1.7.3	First passage times for Brownian motion with drift	48
1.8	Useful laws about Brownian motion's path	54
1.8.1	Reflection principle	54
1.8.2	Arcsin laws	57
Exercises		61
1.8.3	Brownian bridge	62
1.9	Properties of Brownian Paths	65
1.9.1	Variation	65
1.9.2	Differentiability	68
1.9.3	Growth of Brownian motion path	69
2	Diffusion processes	72
2.1	One-dimensional diffusion processes	72
2.1.1	Definition	72
2.1.2	The Dynkin condition	74
2.1.3	Characterization of Diffusion processes	75
2.1.4	Examples of transformed diffusion processes	78
2.2	Kolmogorov backward and forward equations	81
2.2.1	Kolmogorov backward differential equation	81
2.2.2	Kolmogorov forward differential equation	84
2.2.3	Solution of the Kolmogorov Equations with transformation method	87

2.3	Differential equation for the boundary classification	90
2.3.1	Differential equations associated with certain functionals	90
2.3.2	Stationary distribution for time homogeneous diffusion	96
2.3.3	Boundary Classification	97
3	Simulation of stochastic processes	104
3.1	Generation of random variables	106
3.1.1	Uniform random variables	106
3.1.2	Classical discrete random variables	107
3.1.3	Generator of continuous random variables	108
3.2	Monte Carlo Methods	114
3.2.1	“Hit or miss” method	115
3.2.2	Sample mean method	118
3.2.3	Variance reduction: Importance Sampling	120
3.2.4	Variance reduction: Control Variable method	121
3.3	Simulation of a Brownian motion	122
3.3.1	Definition and approach	122
3.3.2	First approach	124
3.3.3	Levy’s original argument	126
3.3.4	Karhunen-Loeve theorem	128
A	Code	130
A.1	Simulation of random variables	131
A.2	Monte Carlo methods	135
A.3	Simulation of Brownian motion	137
B	Homeworks	144

Chapter 1

Brownian Motion

1.1 Continuity of stochastic processes

The sample path in a discrete scenario is given by

$$\mathbb{P}(X(t) < x | \mathcal{F}_s) = \mathbb{P}(X(t) < x | X_s).$$

This gives us a sample path like this:

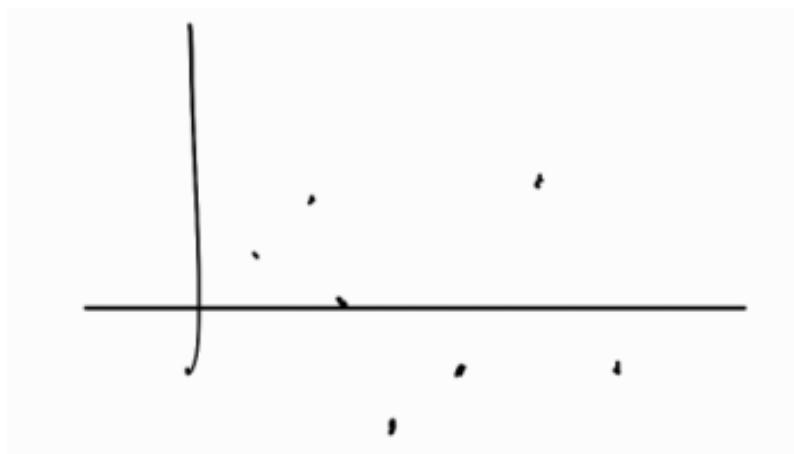


Figure 1.1: Fucking hell.

We want to think, though, about a continuous space and time process:

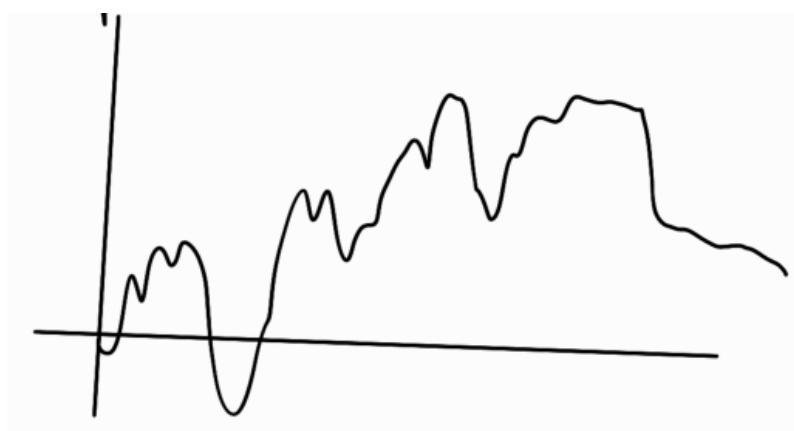


Figure 1.2: Leaning like my academic career.

How can we define the continuity of the sample paths? and how can we check this property? There are some possible definitions of continuity. To control for their robustness we check whether according to each of these definitions the Poisson process, a discrete process, is correctly classified as non continuous.

Definition 1.1.1

A stochastic process $\{M(t)\}$ is said to be a **counting process** if:

- i. $M(t) > 0$;
- ii. $M(t)$ is an integer;
- iii. $M(t)$ is increasing, meaning that $s \leq t \implies M(s) \leq M(t)$.

In general, these processes count how many times an event happens.

Definition 1.1.2

A Poisson process is a counting process such that:

- 1. $N(0) = 0$;
- 2. for $t_1 < t_2 < t_3 < t_4$ we have

$$N(t_2) - N(t_1) \perp\!\!\!\perp N(t_4) - N(t_3)$$

meaning that the increments are independent;

- 3. for $\forall h > 0$ and $t > \tau$ it holds:

$$N(t) - N(\tau) \sim N(t + h) - N(\tau + h)$$

meaning that the increments are stationary (the origin start doesn't matter);

- 4. we have

$$\mathbb{P}(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{for } k = 0, 1, \dots$$

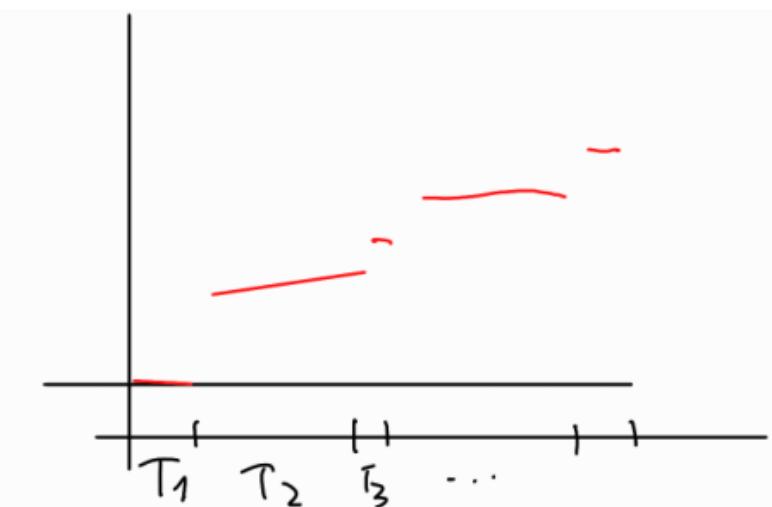


Figure 1.3: I am very good at drawing straight lines.

The inter-arrival times are i.i.d. random variables distributed as

$$T_i \sim \text{Exp}(\lambda).$$

The continuity of sample paths can be characterized in four different ways:

1. **Continuity in mean squares:** we have this if for $\forall t \geq 0$ we have

$$\lim_{s \rightarrow t} \mathbb{E} [|X(t) - X(s)|^2] = 0.$$

So according to this definition, the mean square of the distance goes to 0 if we go near s .

2. **Continuity in probability:** we have this if for $\forall t \geq 0$ and $\forall \varepsilon > 0$ we have

$$\lim_{s \rightarrow t} \mathbb{P}(|X(t) - X(s)| > \varepsilon) = 0.$$

This should be enough for all finite distributions, right?¹

Theorem 1.1.1

Let $\{X(t)\}$ be a stochastic process such that $\mathbb{E}[X^2(t)] < \infty$ for all t . Then it is continuous in mean squares if and only if:

(a) $m(t) = \mathbb{E}[X(t)]$ is continuous;

(b) the covariance function

$$\Gamma(s, t) = \mathbb{E}[(X(t) - m(t))(X(s) - m(s))]$$

is continuous on its diagonal set.

Proof

Consider the expectation

$$\mathbb{E} [|X(t) - X(s)|^2] = \mathbb{E} [|X^2(s) + X^2(t) - 2X(t)X(s)|] \quad (\bullet)$$

but

$$\mathbb{E} [X^2(s)] = \underbrace{\mathbb{E} [(X(s) - m(s))^2]}_{\Gamma(s,s)} + 2m(s)\mathbb{E} [X(s)] - m^2(s)$$

and

$$\mathbb{E} [X^2(t)] = \Gamma(t, t) + 2m(t)\mathbb{E} [X(t)] - m^2(t)$$

and, moreover,

$$\mathbb{E}[X(s)X(t)] = \Gamma(s, t) - \cancel{m(t)m(s)} + m(t)\mathbb{E}[X(s)] + \cancel{m(s)\mathbb{E}[X(t)]}.$$

So \bullet becomes

$$\begin{aligned} \bullet &= \Gamma(s, s) + 2m^2(s) - m^2(s) + \Gamma(t, t) + 2m^2(t) - m^2(t) - 2\mathbb{E}[X(s)X(t)] \\ &= \Gamma(s, s) + \Gamma(t, t) - 2\Gamma(s, t) + m^2(t) + m^2(s) - 2m(t)m(s) \\ &= \Gamma(s, s) + \Gamma(t, t) - 2\Gamma(s, t) + [m(t) - m(s)]^2. \end{aligned}$$

Hence, if $m(t)$ is continuous and $\Gamma(s, t)$ is continuous for $s = t$ then the process is continuous because we have:

¹First, but not last, question without an answer.

- $[m(t) - m(s)]^2 \rightarrow 0$ since it is a continuous function;
- $\Gamma(s, s) + \Gamma(t, t) - 2\Gamma(s, t)$ that becomes $\Gamma(t, t) + \Gamma(t, t) - 2\Gamma(t, t) = 0$.

I'd like to add that dear prof. Sacerdote didn't explain this last little point. Thank you! So now we have

$$\mathbb{E} [|X(s) - X(t)|^2] \rightarrow 0.$$

If this holds for $m(t)$ and $\Gamma(t, t)$ then it is continuous in mean squares. \square

Remark

A process continuous in mean \square (get it?) is continuous in probability (use Chebyshhev).

^aLike use him? As a person? He is dead.

Is the Poisson process continuous in mean \square (and also in probability)? We know that

$$m(t) = \lambda t$$

and

$$\begin{aligned} \Gamma(s, t) &= \mathbb{E}[(N(t) - m(t))(N(s) - m(s))] \\ &= \mathbb{E}[N(t)N(s)] - 2m(t)m(s) - m(t)m(s) \\ &\stackrel{t>s}{=} \mathbb{E}[(N(t) - 2N(s) + N(s))N(s)] - m(t)m(s) \\ &= \mathbb{E}[(N(t) - N(s))N(s)] + \mathbb{E}[N^2(s)] - m(t)m(s) \\ &= \underbrace{\mathbb{E}[N(t) - N(s)]}_{\lambda(t-s)} \underbrace{\mathbb{E}(N(s))}_{\lambda s} + \underbrace{\mathbb{E}[N^2(s)]}_{\lambda t} - \underbrace{m(t)m(s)}_{\lambda s} \\ &= \lambda^2(t - s)s - \cancel{\lambda^2ts} + \mathbb{E}[N^2(s)] \\ &= -\lambda^2s^2 + \underbrace{\mathbb{Var} N(s)}_{\lambda s} + \underbrace{[\mathbb{E}[N(s)]]^2}_{\lambda^2s^2} \\ &= \lambda s \quad \text{if } s < t. \end{aligned}$$

So $\Gamma(s, t) = \lambda \min(s, t)$ is continuous on diagonal and $m(t) = \lambda t$ is also continuous...

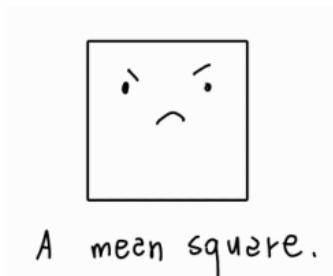


Figure 1.4: Sei un fallito.

but this would mean that Poisson processes are continuous! Which they shouldn't be!
Do I care? No!

3. **Almost sure continuity:** we can ask, as a requirement, that

$$\mathbb{P} \left(\lim_{s \rightarrow t} N(s) = N(t) \right) = 1.$$

Does this finally solve the problem with the Poisson processes? No, because it verifies the almost sure continuity (since it is discontinuous only in a countable number of instances). It is not enough to think point by point: we must think *uniformly*.

Definition 1.1.3

A stochastic process $\{X(t)\}$ has **almost sure continuous sample paths** if, with probability 1, $X(t)$ is a continuous function, that is:

$$\mathbb{P}(X(t) \text{ has continuous samples}) = 1.$$

Of course, the case in which you have exceptional points is not a problem since they have measure 0.



Figure 1.5: Me neither.

Remark

The set

$$\{\omega \in \Omega : t \rightarrow B_t(\omega) \text{ is continuous}\}$$

is not necessarily in the σ -algebra generated by the vectors

$$(B_{t_1}, B_{t_2}, \dots, B_{t_n}) \quad n \in \mathbb{N}.$$

1.2 Definitions of Brownian motion

1.2.1 Historical context

Imagine a spherical particle with a diameter of 10^{-6} m surrounded by 10^{-23} (the Avogadro number) molecules with a diameter of 10^{-10} m of diameter. How can we model the behavior of this ball in a mathematical way²?

In 1828 Robert Brown (fig. 1.7) observed the chaotic movement of pollen in the water through a microscope and noted:

- the motion was composed by translations and rotations;

²At this point of the lesson Prof. Sacerdoti started ranting about Machine Learning (seriously?) being a black box for like 10 minutes.

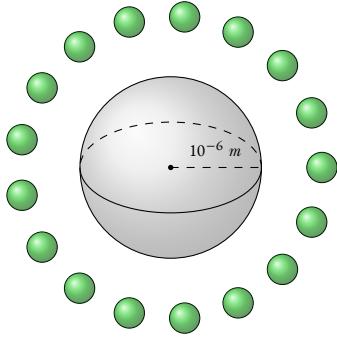


Figure 1.6: My balls.

- particles seemed to move independently one from the other;
- smallest particles moved more actively;
- less viscous fluids determined more activity movement;
- the motion never ceased;
- the motion was not determined by liquid flows or evaporation;
- particles were not animated.

In 1905 Einstein gave the correct explanation formulating a model that also proved valid for forecasting: the idea was that the atoms surrounding the particles (the atoms of the fluid) performed a temperature-dependent movement that collided with the particles, changing their direction.

In 1906 the Polish mathematician Smoluchowski independently obtained results similar to



Figure 1.7: Autechre are a IDM duo from Rochdale, England, composed by Robert Brown and Sean Booth.

Einstein's. Unrelated, but in 1909 Jean Perrin determined the size of atoms.

The Einstein approach considers the motion of a particle over a small interval T observed in a time large enough to see at least 2 intervals t independent. Consider, for the sake of simplicity, the one-dimensional case. During t the particle moves of a distance Δ , that is a random variable with density $\varphi(\Delta)$ that must be symmetric so $\varphi(\Delta) = \varphi(-\Delta)$. For example, we can imagine $\varphi(\Delta) \sim N$ with mean 0 (since in this model we do not expect a drift) and small variance. Then we can consider

$$\gamma = f(x, t) \quad \text{as the number of particles per unit of volume.}$$

Einstein proved that f verifies

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2}.$$

This was the equation of heat diffusion that was already well known at the time. And so was its solution, which is

$$f(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}.$$

In this case D had a physical meaning: Einstein proved

$D = \frac{1}{\sqrt{6\pi r\eta}} \frac{RT}{N}$	<ul style="list-style-type: none"> • η: dynamic viscosity • r: dimension of particles • R: gas constant ($8.3 \frac{\text{J}}{\text{mol K}}$) • T: absolute temperature • N: Avogadro's number
--	--

He could, moreover, calculate the average displacement of a particle

$$\lambda_x = \sqrt{\bar{x}^2} = \sqrt{2Dt}.$$

This means that the displacement (space) is proportional to the square root of time... But what is this story of space and time being related non linearly? This sounded pretty wild for that time.

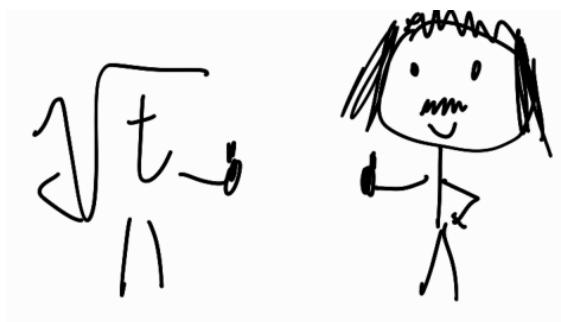


Figure 1.8: Idk I think I would have probably killed myself.

1.2.2 Random walk model

We can think that:

- each particle starts at $x = 0$;
- the particle changes position at discrete times

$$k\Delta t \quad k = 1, 2, \dots \quad \Delta t > 0;$$

- the particles move ΔX units to the right (or to the left) with probability $p = \frac{1}{2}$;
- ΔX does not depend on any past position nor on the current position.

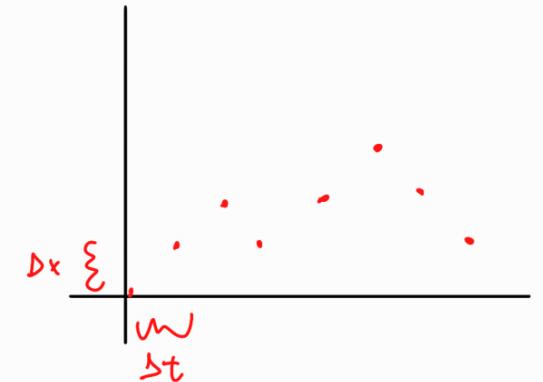


Figure 1.9: I do not know how to do curly brackets

When ΔX and Δt go to 0 in an appropriate way, we don't see "jumps" anymore but a continuous process: So we get a continuous process $\{X_t\}_{t \geq 0}$ that is the random position of the time t .

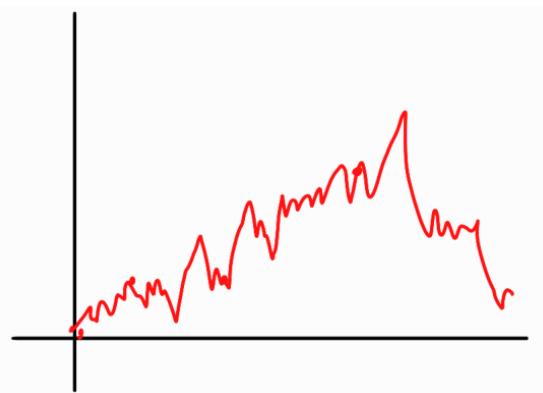


Figure 1.10: I miss tikzes but they take so much time.

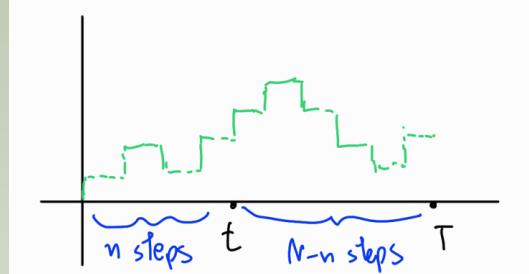
Introduce the i.i.d. random variables $\{\varepsilon_i\}$ such that $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = 0) = \frac{1}{2}$. Then $S_N = \sum_i^N \varepsilon_i$ is the number of moves to the right (since moves to the left do not contribute to this count). In the same way, I have $N - S_N$ being the number of moves to the left. Now we can write

$$\begin{aligned}
 X_T &= \text{positions of particles at } T = N\Delta t \\
 &= S_N \Delta X - (N - S_N) \Delta X \\
 &= (2S_N - N)\Delta X \\
 &= \sum_{k=1}^N (2\varepsilon_k - 1)\Delta X.
 \end{aligned}$$

Remark

For any two times $t = n\Delta t$ and $T = N\Delta t$, $0 < t < T$, we can write:

$$\begin{aligned} X_T &= (X_T - X_t) + (X_t - X_0) \\ &= \sum_{k=n+1}^N (2\varepsilon_k - 1)\Delta X + \sum_{k=1}^n (2\varepsilon_k - 1)\Delta X. \end{aligned}$$



But these quantities involve different ε , so they are independent:

$$X_T - X_t \perp X_t - X_0 \quad (\text{independent increments.})$$

But we know that

$$\mathbb{E}\varepsilon_i = \frac{1}{2} \quad \text{and} \quad \text{Var } \varepsilon_i = \frac{1}{4}$$

so

$$\begin{aligned} \text{Var } X_T &= \text{Var} \left[\sum N_k (2\varepsilon_k - 1) \Delta X \right] \\ &\stackrel{i.i.d.}{=} \sum_k^N (\Delta X)^2 \text{Var} (2\varepsilon_k - 1) \\ &= \sum_k^N (\Delta X)^2 \cancel{\text{Var } \varepsilon_k}^{\frac{1}{4}} \\ &= N(\Delta X)^2. \end{aligned}$$

But $N = \frac{T}{\Delta t}$ so

$$N(\Delta X)^2 = \frac{(\Delta X)^2}{\Delta t} T = \sigma^2 T.$$

So this should be the variance of our random walk. If we take the limit for x and t and do not take into account that space and time decrease in different ways we get a process that has mean 0 and variance 0... but that is a process that is 0 almost surely! So we get the idea that time and space must decrease at different speeds. We also said that the increments must be stationary, but that has its implications.

Remark

We know that $\text{Var } 2S_N = N$, but

$$\begin{aligned} X_T &= (2S_N - \underbrace{\mathbb{E}2S_N}_{\mathbb{E}2S_N}) \Delta X \\ &= \frac{2S_N - \mathbb{E}2S_N}{\sqrt{N}} \sqrt{N} \Delta X \\ &= \frac{2S_N - \mathbb{E}2S_N}{\sqrt{\text{Var}(2S_N)}} \underbrace{\sqrt{N}}_{\sqrt{\frac{T}{\delta t}}} \underbrace{\Delta X}_{\sigma \sqrt{\Delta t}} \\ &= \frac{2S_N - \mathbb{E}2S_N}{\sqrt{\text{Var}(2S_N)}} \sigma \sqrt{T}. \end{aligned}$$

This gets to $Z \sim \mathcal{N}(0, 1)$ as $N \rightarrow \infty$

So we get that this quantity gets to $Z\sigma\sqrt{t}$ and it is true for each t :

$$X_t \sim \mathcal{N}(0, t\sigma^2).$$

1.2.3 Definitions

Definition 1.2.1

Definition of Brownian Motion (1923, by Norbert Wiener).

A stochastic process $B = \{B_t\}_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in \mathbb{R} is called **standard Brownian motion** (or **standard Wiener process**) if:

1. the function $t \mapsto B_t$ is continuous from \mathbb{R}_+ to \mathbb{R} and $B_0 = 0$ (both almost surely);
2. B has stationary increments (they don't depend on the origin):

$$B_t - B_s \sim B_{t+h} - B_{s+h} \quad \mathbb{E}0 < s < t \quad \forall f > 0;$$

3. B has independent increments, i.e. for any $0 \leq t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ and $n \geq 1$ we have:

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}} \quad \text{independent};$$

4. $B_t \sim \mathcal{N}(0, t)$ for $\forall t \geq 0$.

Definition 1.2.2

General Brownian motion (Brownian motion with drift).

$$X_t = \underbrace{\mu t}_{\text{drift}} + \underbrace{\sigma}_{D=\frac{\sigma^2}{2}} B_t.$$

This process verifies properties 1, 2, 3 and

$$X_t \sim \mathcal{N}(\mu t, \sigma^2 t).$$

We can now proceed to define the d -dimensional Brownian motion:

Definition 1.2.3

A d -dimensional Brownian motion, $B = \{B_t\}_{t \geq 0}$ is a stochastic process indexed by $[0, \infty)$ taking values on \mathbb{R}^d such that:

1. $B_0(\omega) = 0$ a.s. $\forall \omega \in \Omega$;
2. $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ are independent for $\forall n > 1$, $0 \leq t_1, t_2 < \dots < t_n < \infty$ (**independent increments**);
3. $B_t - B_s \sim B_{t+h} - B_{s+h}$ for $\forall h > 0$, $\forall 0 \leq s < t < \infty$ (**stationary increments**);
4. $B_t - B_s \sim N(0, t-s)^{\otimes d}$ with

$$N(0, t) dx = \frac{dx}{\sqrt{2\pi t}} \exp\left\{-\frac{x^2}{2t}\right\}$$

(**Gaussian increments**);

5. $t \mapsto B_t(\omega)$ is continuous for all ω (**continuity of sample paths**).

Remember that

$$1, 2, 3, 5 \implies 4 \quad \text{and} \quad 1, 2, 3, 4 \implies 5$$

for almost all ω . A standard Brownian motion on \mathbb{R}^d is obtained by setting

$$B = (B^1, B^2, \dots, B^d)$$

where B^1, B^2, \dots are independent Brownian motion on \mathbb{R} called coordinate processes.

Definition 1.2.4

The continuous process W is called **Wiener process** with respect to \mathcal{A} if it is adapted to \mathcal{A} (check that adaptness is included in the above definitions!) and

$$\mathbb{E}_s f[W_{s+t} - W_s] = \int_{\mathbb{R}} dx \left[\frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}} f(x) \right] \quad \text{for } f \text{ positive Borel function and } \forall s, t \in \mathbb{R}_+.$$

Exercise 1

Prove that the second definition of Brownian motion implies the first for $d = 1$.

But with these properties there is the risk that no such object exists! For example, if we add differentiability to our list of requests then no object satisfies such properties. So we need to ask two questions:

- Does Brownian motion exists?
- Is it unique?

Wiener worked on the difference space, that is the space of the increments of the process, and introduced a Fourier series representation:

$$W_t = \xi_0 t + \sqrt{2} \sum_{n=1}^{\infty} \xi_n \frac{\sin(n\pi t)}{\pi n} \quad 0 \leq t \leq 1$$

where $\{\xi_i\}$ are i.i.d. random variables $\xi_i \sim N(0, 1)$, so this object is a Fourier series for which the coefficients are random variables. Thinking about difference equations makes our life a bit easier, since different times have different ξ_i but these are independent so also the increment between times are independent.

Exercises

Exercise 2

Consider a continuous process $W = (W_t)_{t \geq 0}$ adapted to \mathcal{F} and let

$$\mathbb{E}_s [f(W_{t+s} - W_s)] = \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{x^2}{2t}\right\} f(x)$$

for all s and t in \mathbb{R}_+ and all positive Borel functions f on \mathbb{R} . Show that W is a standard BM.

We need to check the five properties:

1. $B_0(\omega) = 0$ a.s. $\forall \omega \in \Omega$. We know that the increment has the same distribution as a Gaussian random variable distributed as $N(0, t)$. This means that as $t \rightarrow 0$ the distribution collapses towards a Dirac measure centered at 0:

$$W_t - W_0 \xrightarrow{t \rightarrow 0} 0 \quad \text{in distribution}$$

but since W_t is continuous in t then we know that

$$W_0 = \lim_{t \rightarrow 0} W_t.$$

So it must be the case that

$$W_0 = 0.$$

Technically, convergence in distribution + continuity should imply a.s. convergence because there are no jumps in the distribution! ✓

2. $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ are independent for $\forall n > 1$, $0 \leq t_1, t_2 < \dots < t_n < \infty$ (**independent increments**). The increment $W_{s+t} - W_s$ over the interval $(s, s+t)$ is independent of the past \mathcal{F}_s (and its distribution is $N(0, t)$). ✓

⚠️ I am losing my grip on reality ⚠️

I had to search Çinlar's horrid book for this!

3. $B_t - B_s \sim B_{t+h} - B_{s+h}$ for $\forall h > 0, \forall 0 \leq s < t < \infty$ (**stationary increments**). This works because the distribution of the interval does not depend on s but only on t . ✓

4. $B_t - B_s \sim N(0, t-s)^{\otimes d}$ with

$$N(0, t) dx = \frac{dx}{\sqrt{2\pi t}} \exp\left\{-\frac{x^2}{2t}\right\}$$

(**Gaussian increments**). This follows by definition, I guess. ✓

5. $t \mapsto B_t(\omega)$ is continuous for all ω (**continuity of sample paths**). ✓

Exercise 3

Show that there exist a random vector (U, V) such that U and V are one dimensional Gaussian random variables but (U, V) is not Gaussian. Hint: try

$$f(u, v) = g(u)g(v)(1 - \sin u \sin v)$$

with

$$g(u) = \frac{\exp\left\{-\frac{u^2}{2}\right\}}{\sqrt{2\pi}}.$$

We can compute the marginal densities:

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} g(u)g(v)(1 - \sin u \sin v) du \\ &= g(v) \int_{-\infty}^{\infty} g(u)(1 - \sin u \sin v) du \\ &= g(v) \left[\underbrace{\int_{-\infty}^{\infty} g(u) du}_{1} - \sin v \int_{-\infty}^{\infty} g(u) \sin u du \right] \\ &\quad \text{odd function, so the integral is 0} \\ &= g(v). \end{aligned}$$

So the marginal density is Gaussian! Do the same with the other marginal density to obtain the same shit.

Exercise 4

Consider a random walk with independent steps $\{Z_i\}_{i=1,\dots}$ each having the distribution:

$$\mathbb{P}(Z = \Delta) = p, \mathbb{P}(Z = -\Delta) = 1 - p = q$$

with $p > 0$. Suppose that these steps of size Δ take place at small time interval of length τ . We are interested in the limit process when Δ and τ approach zero.

1. Determine the moment generating function $\mathbb{E} [\exp \{-\vartheta Z\}]$ of Z .
2. Observe that at time t there are $n = \lfloor \frac{t}{\tau} \rfloor \sim \frac{t}{\tau}$ steps and that the total displacement at time t , $X(t)$, is the sum of n independent random variables. Determine $\mathbb{E} [\exp \{-\vartheta X(t)\}]$.
3. Determine the mean and variance of $X(t)$.
4. Propose a suitable way to let $\Delta \rightarrow 0$ and $\tau \rightarrow 0$ to obtain a meaningful result for the continuous space and time limit process. Hint: we want to obtain

$$(p - q) \frac{\Delta}{\tau} \rightarrow \mu, \quad 4pq\Delta^2 \rightarrow \sigma^2.$$

5. Is the obtained process a Brownian motion with drift?

So we know that the m.g.f of one step (and all steps are mutually independent) is

$$\mathbb{E} [e^{-\vartheta Z}] = pe^{-\vartheta Z} + (1 - p)e^{-\vartheta Z}.$$



Observe that in time t there will be $n = t/\tau$ steps. We know that the mean and variance of the total displacement $X(t)$ are

$$\begin{aligned}\mathbb{E}[X(t)] &= n \cdot \mathbb{E}[Z] = \frac{t}{\tau}(p - q)\Delta \\ \text{Var}(X(t)) &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{t}{\tau}4pq\Delta^2.\end{aligned}$$

The computations should be right... We want to let $\Delta \rightarrow 0$ and $\tau \rightarrow 0$ so that we obtain a continuous result by scaling the step size and the time step to zero. Suppose we require the limiting process to have mean μ and variance σ^2 in unit time, which means requiring Δ and τ to tend to zero in such a way that

$$(p - q)\frac{\Delta}{\tau} \rightarrow \mu, \quad 4pq\Delta^2 \rightarrow \sigma^2$$

That is, the expected displacement per step must converge to a finite value. This is like forcing the limit to make sense, otherwise just letting $\Delta \rightarrow 0$ and $\tau \rightarrow 0$ we obtain 0 mean and 0 variance. Which means 0 bitches. You want that? I don't. The main idea is that Δ must be go to zero slower than τ : by setting $\Delta = \sigma\sqrt{\tau}$ we obtain that the displacement Δ is of a considerably larger order of magnitude than the small time interval τ in which it occurs. So our request will be satisfied with

$$\begin{aligned}\Delta &= \sigma\sqrt{\tau} \\ p &= \frac{1}{2} \left(1 + \frac{\mu\sqrt{\tau}}{\sigma}\right) \\ q &= \frac{1}{2} \left(1 - \frac{\mu\sqrt{\tau}}{\sigma}\right).\end{aligned}$$

We can now substitute these mfs in  and get

$$\mathbb{E}[e^{-\vartheta X(t)}] = \left[\frac{1}{2} \left(1 + \frac{\mu\sqrt{\tau}}{\sigma}\right) e^{-\vartheta\sigma\sqrt{\tau}} + \frac{1}{2} \left(1 - \frac{\mu\sqrt{\tau}}{\sigma}\right) e^{\vartheta\sigma\sqrt{\tau}} \right]^{\frac{t}{\tau}}.$$

We already know that:

$$\mathbb{E}[X(t)] = \frac{t}{\tau}(p - q)\Delta = \frac{t}{\tau} \left(\frac{\mu\sqrt{\tau}}{\sigma}\right) \sigma\sqrt{\tau} = \mu t.$$

This shows that the expected displacement of the random walk at time t converges to a linear function of t , with slope μ , as required. This means that the limiting process has drift μ . Next, for the variance:

$$\text{Var}(X(t)) = \frac{t}{\tau} \cdot 4pq\Delta^2.$$

Substituting $\Delta = \sigma\sqrt{\tau}$, we get:

$$\text{Var}(X(t)) = \frac{t}{\tau} \cdot 4pq(\sigma\sqrt{\tau})^2 = \frac{t}{\tau} \cdot 4pq\sigma^2\tau = 4pq\sigma^2t.$$

Now, substituting $p = \frac{1}{2} \left(1 + \frac{\mu\sqrt{\tau}}{\sigma}\right)$ and $q = \frac{1}{2} \left(1 - \frac{\mu\sqrt{\tau}}{\sigma}\right)$, we get:

$$4pq = 4 \cdot \frac{1}{2} \left(1 + \frac{\mu\sqrt{\tau}}{\sigma}\right) \cdot \frac{1}{2} \left(1 - \frac{\mu\sqrt{\tau}}{\sigma}\right) = 1 - \left(\frac{\mu\sqrt{\tau}}{\sigma}\right)^2.$$

Thus, we have:

$$\text{Var}(X(t)) = \left(1 - \left(\frac{\mu\sqrt{\tau}}{\sigma}\right)^2\right)\sigma^2 t.$$

Taking the limit as $\tau \rightarrow 0$, we see that the term involving μ vanishes and the variance becomes:

$$\text{Var}(X(t)) = \sigma^2 t,$$

which is precisely the variance of a Brownian motion with variance σ^2 .

Exercise 5

Ignoring the requested continuity of sample paths of the Brownian motion show that the Brownian motion is continuous in mean square.

The process is continuous in mean square if the expectation of the squared difference between $B(t)$ and $B(s)$ goes to 0 as $t \rightarrow s$:

$$\lim_{s \rightarrow t} \mathbb{E}[(B(t) - B(s))^2] = 0.$$

We know that $\mathbb{E}[B(t)] = 0$, that is continuous because the mean is trivially constant over time (duh). The covariance function is, for $s \leq t$,

$$\begin{aligned} \Gamma(s, t) &= \text{Cov}(B(s), B(t)) \\ &= \mathbb{E}[B(s)B(t)] - \underbrace{\mathbb{E}[B(s)]}_{0} \underbrace{\mathbb{E}[B(t)]}_{0} \\ &= \mathbb{E}[B(s)B(t)] \\ &= \mathbb{E}[B(s)(B(s) + B(t) - B(s))] \\ &= \mathbb{E}[B(s)^2] + \mathbb{E}[B(s)(\underbrace{B(t) - B(s)}_{\perp \text{ of } B(s)})] \\ &= \mathbb{E}[B(s)^2] + \underbrace{\mathbb{E}[B(s)]}_{=0} \mathbb{E}[(B(t) - B(s))] \\ &= \mathbb{E}[B(s)^2] \\ &= s = \min\{s, t\}. \end{aligned}$$

⚠ I am losing my grip on reality ⚠

It would have been useful to know about this little property during lesson...

We know that $\min\{s, t\}|_{s=t} = t$ so our limit is:

$$\begin{aligned} \lim_{s \rightarrow t} \mathbb{E}[(B(t) - B(s))^2] &= \lim_{s \rightarrow t} \left(\mathbb{E}[B(t)^2] - 2 \underbrace{\mathbb{E}[B(t)B(s)]}_{\Gamma(s,t)} + \mathbb{E}[B(s)^2] \right) \\ &= \lim_{s \rightarrow t} (t - 2t + t) = 0. \end{aligned}$$

1.3 Brownian motion as a Gaussian processes

1.3.1 Properties of Gaussian Processes

Definition 1.3.1

A **Gaussian process** is defined as a process Γ with

$$\begin{aligned}\frac{d}{d\lambda} \mathbb{E}[e^{i\lambda\Gamma}] \Big|_{\lambda=0} &= im & m &= \mathbb{E}[\Gamma] \\ \frac{d^2}{d\lambda^2} \mathbb{E}[e^{i\lambda\Gamma}] \Big|_{\lambda=0} &= -\mathbb{E}[\Gamma^2].\end{aligned}$$

Definition 1.3.2

A random vector

$$\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_n) \in \mathbb{R}^n$$

is Gaussian if the scalar product

$$\langle \lambda, \Gamma \rangle \quad \forall \lambda \in \mathbb{R}^n$$

is a one-dimensional Gaussian random variable with:

$$\mathbb{E}[e^{i\langle \lambda, \Gamma \rangle}] = e^{i\mathbb{E}[\langle \lambda, \Gamma \rangle] - \frac{1}{2}\text{Var}(\langle \lambda, \Gamma \rangle)}.$$

So setting

$$m = (m_1, \dots, m_n) \in \mathbb{R}^n \quad m_j = \mathbb{E}[\Gamma_j]$$

$$\Sigma = (\sigma_{jk})_{jk} \in \mathbb{R}^{n \times n} \quad \begin{aligned}\sigma_{jk} &= \mathbb{E}[(\Gamma_j - m_j)(\Gamma_k - m_k)] \\ &= \text{Cov}(\Gamma_j, \Gamma_k) \\ &= \Sigma_{j,k}\end{aligned}$$

So that

$$\mathbb{E}[e^{i\langle \lambda, \Gamma \rangle}] = e^{i\langle \lambda, m \rangle - \frac{1}{2}\langle \lambda, \Sigma \lambda \rangle}.$$

The basic idea is that I can linearly transform a random vector and it still is a Gaussian random variable. Fourier transform always exists even for limits, while it is not so easy for distributions. What happens if we apply to a Gaussian vector a matrix that turns it into a subspace? Do I care? Not much! But imagine the classic Gaussian random variable:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}.$$

If I take $\sigma \rightarrow 0$ then the density degenerates into a Dirac measure. This is still technically a measure because it integrates to 1, but it is not a density anymore since we lose dimensionality on \mathbb{R} which is³ the meaning of becoming a subspace! But now consider the characteristic function

$$\mathbb{E}[e^{i\lambda x}] = e^{-1\lambda \frac{\sigma^2}{2}} \xrightarrow[\square]{\sigma \rightarrow 0} 1. \quad \text{characteristic function of degenerate random variables.}$$

The main point about this is that we could still take a limit. If we were working with density we couldn't have done that.

The alternative definition for Gaussian processes are:

³As I learn now.

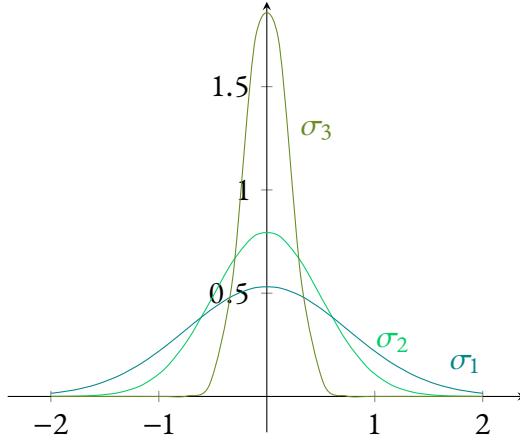


Figure 1.11: It will become a dirac measure... but that is not a fucking density.

Definition 1.3.3

A vector valued random variable X has n -dimensional standard Gaussian distribution if its n coordinates are standard Gaussian in \mathbb{R}^1 and are independent.

Definition 1.3.4

A vector valued random variable $Y : \Omega \rightarrow \mathbb{R}^n$ is Gaussian if there exists a vector valued X having n -dimensional standard Gaussian distribution and $m \times n$ matrix A and a m -dimensional vector b such that

$$Y = AX + b \quad \text{and } AA^\top = \Sigma.$$

In this case Y is a linear transformation of the Gaussian vector and A projects X in a m -dimensional subspace. This is the same procedure we do with Principal Component Analysis, where we diagonalize a matrix and other passages I don't remember and I don't care about. There are some properties:

Proposition 1.3.1

- Let $X \sim N(0, \Sigma_n)$. The distribution of $[X_{k+1}, \dots, X_n]$ conditioned to $[X_1 = x_1, \dots, X_k = x_k]$ is

$$W(0, \Sigma_{2|1})$$

where

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Note that the sub-matrices that Σ is made of have lower dimensionality.

- If $Y = \underbrace{a}_{\in \mathbb{R}^m} + \underbrace{B}_{\in \mathbb{R}^{n \times m}} X$ is an affine transformation of $X \sim N(\mu, \Sigma)$ then

$$Y \sim N(a + B\mu, B\Sigma B^\top).$$

This is like working with the affine transformation.

What about another definition? I think we need it

Definition 1.3.5

A Gaussian process is a stochastic process such that for every t_1, \dots, t_k in the index set \mathbb{T} and each $k > 1$ we have that

$$(X_{t_1}, \dots, X_{t_k}) = X_{t_1, \dots, t_k}$$

is a multivariate Gaussian random vector (possibly degenerate).

Now we have defined a class of processes:

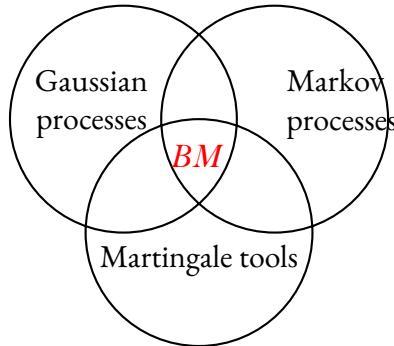


Figure 1.12: Balls.

There is another class: the Levy processes: they have independent and stationary increment (so they are Markov processes) but unlike Brownian motion we don't require continuity on the sample path, but just continuity on the right of sample paths and admit left limit.

Definition 1.3.6

We call **marginal distribution** of $\{B_t\}$ the laws of the finite dimension vectors

$$\Gamma_n = (B_{t_1}, B_{t_2}, \dots, B_{t_n})^\top \quad 0 \leq t_1 < t_2 < \dots < t_n < \infty, \quad \forall n > 0$$

There are properties about the increments

$$(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})$$

but, anyway, for the Brownian motion to be gaussian all Γ_n must be gaussian.

1.3.2 Gaussianity of Brownian motion

Proposition 1.3.2

Let $B = \{B_t\}_{t \geq 0}$ be a standard BM in $[0, \infty)$. Let s, t be given and fixed. The covariance will be

$$\mathbb{E}[B_t B_s] = s \wedge t.$$

Proof

Assume, without loss of generality, that $s < t$. Then

$$\begin{aligned}\mathbb{E}[B_t B_s] &= \mathbb{E}[(B_t - B_s + B_s)B_s] \\ &= \underbrace{\mathbb{E}[(B_t - B_s)B_s]}_{\text{independent increments}} + \underbrace{\mathbb{E}[B_s^2]}_{\text{Var}(B_s)=s} \\ &= \mathbb{E}[B_t - B_s] \underbrace{\mathbb{E}[B_s]}_{=0} + s = s.\end{aligned}$$

But, similarly, if $s > t$ then $\mathbb{E}[B_t B_s] = t$. \square

Remark

For every $0 \leq t_1 < \dots < t_n$ and $n \geq 1$, we have:

$$\begin{aligned}B_t &\sim N(0, t) \\ B_{t_2} - B_{t_1} &\sim N(0, t_2 - t_1) \\ &\vdots \\ B_{t_n} - B_{t_{n-1}} &\sim N(0, t_n - t_{n-1}).\end{aligned}$$

This implies that

$$(B_{t_1}, \dots, B_{t_n})^\top \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, [t_i \wedge t_j]_{i,j=1,\dots,n}^n\right)$$

which is a n -dimensional normal distribution.

Proof

Consider a vector

$$\Delta = \begin{bmatrix} B_{t_1} - \overbrace{B_{t_0}}^{=0} \\ \vdots \\ B_{t_n} - B_{t_0} \end{bmatrix}$$

and

$$B_{t_k} - B_{t_0} = \sum_{j=1}^k (B_{t_j} - B_{t_{j-1}}).$$

Let M be a matrix $n \times n$. We now have the increment but we want a single component. When we want to prove that a vector is Gaussian we usually find a linear transformation that generates another Gaussian.

$$M = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 1 & & & \vdots \\ \vdots & \vdots & & 1 & & \vdots \\ \vdots & \vdots & & & \ddots & 0 \\ 1 & 1 & \cdots & 1 & 1 & \end{bmatrix}.$$

Consider $M\Delta = \Gamma$. We have a linear transformation of a Gaussian vector Δ , that is therefore still Gaussian. But we notice that Γ reconstructs the values of the BM relative to B_{t_0} . Hence the BMs (that have a finite dimensional distribution) are Gaussian. But if all finite dimensional distributions of the process are Gaussian, then the BM is Gaussian. \square

But this means that we can compute the covariance matrix as a simple transformation of the covariance matrix of the original process

$$C = M\Sigma M^T$$

with Σ being the covariance matrix of Δ

$$\Sigma = \begin{bmatrix} t_1 - t_0 & 0 & \cdots & 0 & \\ 0 & t_2 - t_1 & & & \vdots \\ \vdots & & \ddots & 0 & \\ 0 & \cdots & 0 & t_n - t_{n-1} & \end{bmatrix}$$

and

$$\begin{aligned} C &= \underbrace{\begin{bmatrix} 1 & & & \\ 1 & 1 & 0 & \\ \vdots & & 1 & \\ 1 & \cdots & \ddots & \end{bmatrix}}_M \underbrace{\begin{bmatrix} t_1 - t_0 & & & \\ & t_2 - t_1 & 0 & \\ & & \ddots & \\ & & & t_n - t_{n-1} \end{bmatrix}}_\Sigma \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ & 1 & & \vdots \\ 0 & & 1 & \\ & & & \ddots \end{bmatrix}}_{M^T} \\ &= \begin{bmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & & & \vdots \\ t_1 & t_2 & \cdots & t_n \end{bmatrix} = [t_i \wedge t_j]_{n \times n}. \end{aligned}$$

Remark

The vector $\Gamma = (B_1, \dots, B_{t_n})^T$ admits a probability density function⁴

$$\begin{aligned} \mathbb{P}(\Gamma \in dx) &= \frac{dx}{(2\pi)^{\frac{n}{2}} \sqrt{\det C}} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \sqrt{\prod_{j=1}^n (t_j - t_{j-1})} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}} \right\}. \end{aligned}$$

We can see how this has independent components since the exponential depends only by $x_j - x_{j-1}$ and all the terms are separated by the summation. Since the density function factors into a product of terms for each time step, it confirms that the increments of BM are independent.

⁴ $\mathbb{P}(\Gamma \in dx)$ is just a stupid notation to say “the probability that Γ takes values in an infinitesimally small region around dx ”.

Proposition 1.3.3

Gaussian characterization. A stochastic process $B = (B_t)_{t \geq 0}$ is a BM if and only if we have:

1. $t \mapsto B_t$ is continuous from $\mathbb{R}_+ \rightarrow \mathbb{R}$ and $B_0 = 0$;
2. B is a Gaussian process (i.e. $\sum_i \alpha_i B_{t_i}$ is normally distributed for $\forall \alpha_1, \dots, \alpha_n$ in \mathbb{R} and $n \geq 1$);
3. $\mathbb{E}[B_t] = 0$ and $\mathbb{E}[B_t B_s] = t \wedge s$ for $\forall t, s \in [0, \infty)$.

We can also work with the characteristic function.

Lemma 1.3.1

Let $(X_t)_{t \geq 0}$ be a d -dimensional stochastic process and let (X_t) satisfies the properties of Brownian motion B0-B3 (property 0 is the property of independent increments). We have that this process is a BM if and only if, for all $n \geq 0$ and $0 \leq t_0 \leq \dots \leq t_n$ for $\forall n$ and $\xi_0, \xi_1, \dots, \xi_n \in \mathbb{R}^d$,

$$\mathbb{E} \left[\exp \left\{ i \sum_{j=1}^n \langle \xi_j, x_{t_j} - x_{t_{j-1}} \rangle + i \langle \xi_0, x_0 \rangle \right\} \right] = \exp \left\{ \frac{1}{2} \sum_{j=1}^n |\xi_j|^2 (t_j - t_{j-1}) \right\}$$

holds and it has continuous sample paths.

Definition 1.3.7

Let $Q \in \mathbb{R}^{d \times d}$ be a symmetric positive semi-definite $d \times d$ matrix. A **Q -Brownian motion** is a d -dimensional process $(X_t)_{t \geq 0}$ satisfying B0 to B4 and

$$X_t - X_s \sim \mathcal{N}(\mathbf{0}, (t-s)Q) \quad s < t.$$

We ask the matrix to be symmetric positive semi-definite because the covariance matrices are always positive semi-definite!

1.4 Existence of Brownian motion

There are many proofs about the existence of Brownian motion. Wiener described Brownian motion using Fourier series and so he proved its existence using that framework, but we need to use distributions. We want to use the fact that we proved it is a Gaussian process. In general, to prove that a stochastic process exists we can use Kolmogorov's consistency theorem.

The main idea is that when we have a family of distributions, the joint d -dimensional distribution must be consistent with the marginal distribution, that by definition have a lower dimensionality. The Kolmogorov theorem tells us that for every distribution that respects the consistency there exists a probability space and a stochastic process with that distribution! This is actually cool because we can move from functions (stupid & boring) to stochastic processes (cool & sexy) that would normally require every single point to be validly defined and so on. Imagine having something (say, 2-dimensional) that respects the joint probability requirements. We can ask whether it exists a 2-dimensional vector that has that distribution and this is pretty easy... But when we start having continuous time we cannot think about vectors and this could be a real pain in the ass.

Theorem 1.4.1

Kolmogorov consistency theorem (Çinlar p.437). Consider a family of probability distribution on \mathbb{R}^n with joint probability distribution

$$\{F_{t_1, t_2, \dots, t_n}\} \quad \text{for } t_1 < t_2 < \dots < t_n \text{ and } n > 1.$$

If this family verifies the consistency conditions

$$\lim_{x_k \rightarrow \infty} F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$$

for all $(x_1, \dots, x_n) \in \mathbb{R}^n$ and $1 \leq k \leq n$ then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ defined on it. The process is such that

$$\mathbb{P}(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$$

for $\forall t_1 \leq \dots \leq t_n$ in \mathbb{T} and $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$, $n \geq 1$.

So, since we saw that the BM has Gaussian distribution then it exists! But hold your horses: we know that a Gaussian process exists, but we still don't know whether it has continuous sample paths. If we add this request we still do not know if our process can exists.

Theorem 1.4.2

Existence of continuous modifications (Çinlar p. 435). Suppose that $\mathbb{E}|X_t - X_s|^p \leq c|t - s|^{1+q}$ holds for some $p, c, q \in (0, \infty)$. Then for every $\alpha \in \left[0, \frac{p}{q}\right]$ there exists a modification \tilde{X} of X such that the paths of \tilde{X} are Hölder continuous of order α on $[0, 1]$ for $\forall \omega \in \Omega$.

Revise with Kotatsu!

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is Hölder continuous of order α if there exists a constant $c > 0$ such that:

$$|f(x) - f(y)| \leq c|x - y|^\alpha, \quad \forall x, y \in \mathbb{R}^d.$$

for some $\alpha \in (0, 1]$.

Remark

For the BM, with $p = 4, q = 1, c = 3$ we have

$$\mathbb{E}|X_t - X_s|^4 = 3(t - s)^2$$

so we can apply the theorem to get a continuous modification \tilde{X} .

1.5 Features of Brownian motion

- ① **Reflection property:** if $(B_t)_{t \geq 0}$ is a BM^d then $(-B_t)_{t \geq 0}$ is a BM^d due to symmetry of Gaussian distribution.

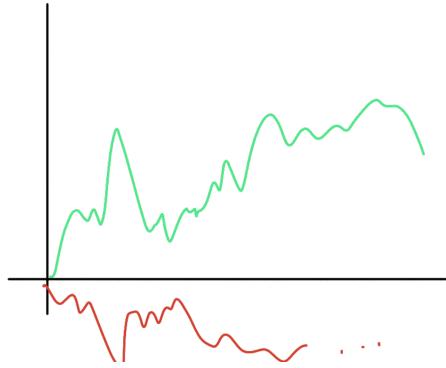


Figure 1.13: I am sorry, I cut the screenshot.

② **Renewal** (or **time deletion**): let $(B_t)_{t \geq 0}$ be a BM^d . Fix a time $a > 0$ and define the process $(W_t)_{t \geq 0}$ as

$$W_t := B_{t+a} - B_a.$$

Then $(W_t)_{t \geq 0}$ is again a BM^d .

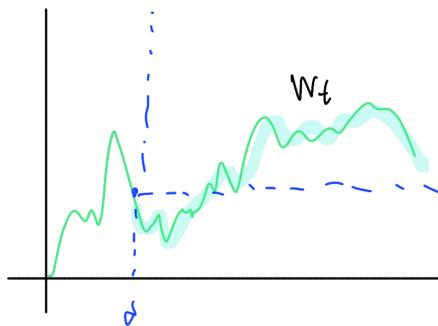


Figure 1.14: I am sorry, I didn't make the lines straight.

Proof

1. We know $W_0 = 0$. ✓

2.-3. If $t_0 = 0 < t_1 < \dots < t_n$ then

$$\begin{aligned} W_{t_h} - W_{t_{j-1}} &= B_{t_j+a} - B_a - B_{t_{j-1}+a} + B_a \\ &= B_{t_j-a} - B_{t_{j-1}+a} \sim \mathcal{N}(0, t_j - t_{j-1}) \end{aligned}$$

so the independence of the increments and their stationarity follows from those of B_t . ✓

4. The continuity is obvious since the sample paths coincide after a . ✓

□

Lemma 1.5.1

$W_t = B_{t+a} - B_a$ and B_t , $t \in (0, a)$ are independent.

③ **Time inversion**: take $W_t := B_{\tau-t} - B_t$ for $t \in [0, \tau]$.

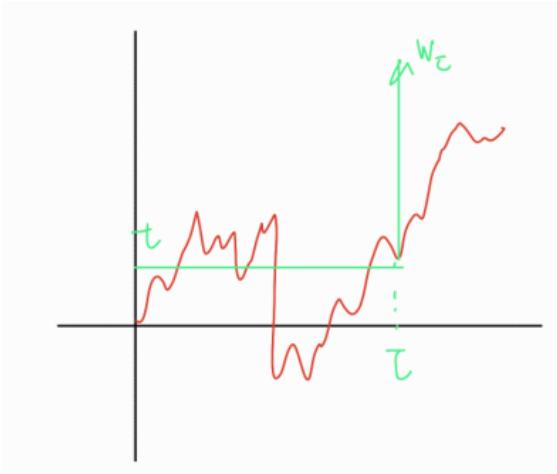


Figure 1.15: This drawing is not very clear to me either.

If B_t is a BM we get another BM inverting time.

- ④ **Scaling property:** if c is a constant then

$$B_{ct} \sim c^{1/2} B_t$$

so if I change the scale in which I measure the time I get another BM.

$$W_t = (c^{1/2} B_{ct})_{t \geq 0}$$

is a BM^d and

$$\mathcal{N}(0, ct) = \sqrt{c} \mathcal{N}(0, t).$$

- ⑤ **Projective reflection** at $t = \infty$: let B_t be a BM^d . Let

$$W_t = \begin{cases} t B_{\frac{1}{t}} & t > 0 \\ 0 & t = 0. \end{cases}$$

W_t is again a BM^d .

Proof

We must use the Gaussian definition: consider the vector

$$W = (W_{t_1}, W_{t_2}, \dots, W_{t_n}) \quad 0 \leq t_1 < \dots < t_n$$

which is a Gaussian vector. Now, we know that

(a) $\mathbb{E}[W] = 0$;

(b)

$$\begin{aligned} \text{Cov} [W_{t_j}, W_{t_k}] &= \text{Cov} \left[t_j B_{\frac{1}{t_j}}, t_k B_{\frac{1}{t_k}} \right] \\ &= t_j t_k \text{Cov} \left[B_{\frac{1}{t_j}}, B_{\frac{1}{t_k}} \right] \\ &= t_j t_k \left(\frac{1}{t_j} \wedge \frac{1}{t_k} \right) = t_j \wedge t_k. \end{aligned}$$

So this process has the mean and variance of a BM! To check the continuity of sample paths we know that for $t \geq 0$ then $t \mapsto W_t$ is a continuous function... but we must also prove the continuity at 0! (Durret p. 363). We can rewrite B_n as

$$B_n = B_1 + (B_2 - B_1) + \dots + (B_n - B_{n-1}).$$

This is a sum of i.i.d. random variables. We can use the Law of Large Numbers to consider the limit of the sum

$$\lim_{n \rightarrow \infty} \frac{B_n}{n} = \mathbb{E}[B_t] = 0$$

This means that the fluctuations of B_n become more and more irrelevant with respect to n and this means that there is a tendency over time to smooth to 0. So the limit through the integer converges to 0... But we are interested in the real numbers, not just the integers! Remember Kolmogorov's maximal inequality:

Revise with Kotatsu!

$$\mathbb{P} \left(\max_{k \leq n} |S_k| \geq x \right) \leq \frac{\text{Var } S_n}{x^2}.$$

As k we take the interval of size 1 over n :

$$\mathbb{P} \left(\sup_k \left| B_{n+\frac{k}{2^m}} - B_n \right| \geq n^{2/3} \right) \leq \frac{\sqrt{\mathbb{E}[(B_{n+1} - B_n)^2]}}{n^{2/3}}. \quad \text{if mean is 0 then variance is just second moment}$$

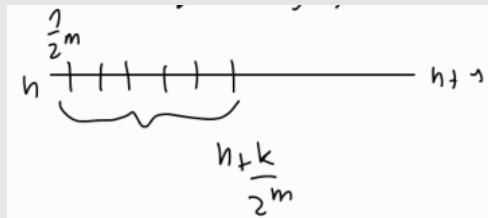


Figure 1.16: WHERE THE FUCK DID m COME FROM??

This is like breaking the interval between n and $n+1$ into finer and finer pieces. Now let $m \rightarrow \infty$:

$$\mathbb{P} \left(\sup_{u \in [n, n+1]} |B_u - B_n| > n^{2/3} \right) \leq n^{-4/3}.$$

This is like asking the probability of the event “at least once between time n and $n+1$, the Brownian motion jumps too much (more than $n^{2/3}$)”. Using Kolmogorov we can say that this probability will be at most $n^{-4/3}$. But we know that $\sum_n n^{-4/3} < \infty$ so we can apply Borel Cantelli to get that

$$\frac{B_u}{u} \rightarrow 0.$$

Now take $u = \frac{1}{t}$ and get the result. \square

1.6 Martingales related with Brownian motion

1.6.1 Three golden martingales

Definition 1.6.1

A martingale (X_t, \mathcal{F}_t) is a real or complex stochastic process

$$\begin{aligned} X_t : \Omega &\rightarrow \mathbb{R}^d \quad \text{or} \\ X_t : \Omega &\rightarrow \mathbb{C} \end{aligned}$$

satisfying

1. $\mathbb{E}|X_t| \leq \infty$ for $\forall t \in \mathbb{T}$;
2. X_t is \mathcal{F}_t -measurable for $\forall t \in \mathbb{T}$;
3. $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$ for $\forall s, t \in \mathbb{T}$ and $s < t$.

This is nothing new from what we already knew. We just added the complex definition because we may want to work with the characteristic functions.

Three golden martingales

1. $(B_t)_{t \geq 0}$;
2. $(B_t^2 - t)_{t \geq 0}$ (or $B_t^2 - dt$ if B_t is a BM^d);
3. $\left(e^{\sigma B_t - \frac{\sigma^2}{2} t}\right)_{t \geq 0}$.

Exercise 1

Check that the Brownian motion with drift is a sub-martingale if μ is positive.

Proof

1. (a) We know that $B_t \sim N(0, t)$ hence $|B_t|$ has p.d.f.

$$f_{|B_t|}(y) = \frac{\sqrt{2}}{\sqrt{\pi t}} e^{-\frac{y^2}{2t}}.$$

Hence, it has expectation

$$\begin{aligned}
 \mathbb{E}|B_t| &= \int_0^\infty f_{|B_t|}(y) dy \\
 &= \sqrt{\frac{2}{\pi t}} t \int_0^\infty y \cdot \frac{1}{t} \cdot e^{-\frac{y^2}{2t}} dy \\
 \text{change of variable: } u &= \frac{y^2}{2t} \implies du = \frac{y}{t} dy \implies dy = du \frac{t}{y} \\
 &= \sqrt{\frac{2t}{\pi}} \underbrace{\int_0^\infty e^{-u} du}_{=1} \\
 &= \sqrt{\frac{2t}{\pi}} < \infty
 \end{aligned}$$

or, if you hate yourself, note that the integrand is... kinda like the derivative of the exponential term. Indeed,

$$\begin{aligned}
 d\left(e^{-\frac{y^2}{2t}}\right) &= e^{-\frac{y^2}{2t}} d\left(-\frac{y^2}{2t}\right) \\
 &= -e^{-\frac{y^2}{2t}} \cdot \frac{y}{t} dy
 \end{aligned}$$

so we can write

$$\sqrt{\frac{2t}{\pi}} \int_0^\infty y \cdot \frac{1}{t} \cdot e^{-\frac{y^2}{2t}} dy = \sqrt{\frac{2t}{\pi}} \int_0^\infty -d\left(e^{-\frac{y^2}{2t}}\right)$$

so using the fundamental theorem of calculus $\int_0^\infty dF(y) = F(\infty) - F(0)$ we get

$$\sqrt{\frac{2t}{\pi}} \cdot - \left[e^{-\frac{y^2}{2t}} \right]_{y=0}^{y=\infty} \implies \sqrt{\frac{2t}{\pi}} \cdot -(0 - 1) = \sqrt{\frac{2t}{\pi}} < \infty$$

So it is integrable.

⚠️ I am losing my grip on reality ⚠️

On a totally unrelated tangent, I spent 2 hours over this stupid fucking equation because during the lessons *someone* decided that there was not a minus sign. Also what the fuck is this notation of $d(\text{my shit})$? Never seen this before and almost had an aneurysm. After all of this, turns out it is just a lazy way to not evaluate an integral.

- (b) It is \mathcal{F}_t^B measurable.
- (c) To check if it is a martingale:

$$\begin{aligned}
 \mathbb{E}[B_t | \mathcal{F}_s] &= \mathbb{E}[B_t - B_s + B_s | \mathcal{F}_s] \\
 &= \underbrace{\mathbb{E}[B_t - B_s | \mathcal{F}_s]}_{= 0 \text{ because they are increments of BM}} + B_s = B_s
 \end{aligned}$$

so it is a martingale.

2. Consider $M_t = |B_t|^2$ where B_t is BM^d . We need to show that M_t is a sub-martingale with respect to \mathcal{F}_t .

(a)-(b) These are easy. I think.

(c)

$$\begin{aligned}\mathbb{E}[M_t|\mathcal{F}_t] &= \sum_{j=1}^d \mathbb{E}\left[\left(B_t^j\right)^2|\mathcal{F}_s\right] \\ &\geq \sum_{j=1}^d \left(\mathbb{E}\left[B_t^j|F_s\right]\right)^2 \\ &= \sum_{j=1}^d \left(B_s^j\right)^2.\end{aligned}$$

Now let's move to the quantity of interest

$$M_t = |B_t|^2 - dt = \sum_{j=1}^d \left[\left(B_t^j\right)^2 - t\right]$$

hence it is enough to prove the result for $d = 1$. Consider $s < t$ and $\mathbb{E}[B_t^2 - t|\mathcal{F}_s]$.

$$\begin{aligned}\mathbb{E}[B_t^2 - t|\mathcal{F}_s] &= \mathbb{E}\left[\left((B_t - B_s) + B_s\right)^2 - t|\mathcal{F}_s\right] \\ &= \mathbb{E}\left[(B_t - B_s)^2 + B_s^2 + 2B_s(B_t - B_s) - t|\mathcal{F}_s\right] \\ &= \underbrace{\mathbb{E}\left[(B_t - B_s)^2|\mathcal{F}_s\right]}_{t-s} - t + B_s^2 - 2B_s \underbrace{\mathbb{E}[B_t - B_s|\mathcal{F}_s]}_0 \\ &= B_s^2 - t\end{aligned}$$

so the process is a martingale!

3. Consider $M_t := e^{i\langle \lambda, B_t \rangle + \frac{t}{2}|\lambda|^2}$ for $\lambda \in \mathbb{R}^d$. This is the same case as written above, but this time we are in the complex realm. Consider the expectation

$$\begin{aligned}\mathbb{E}[M_t^\lambda|\mathcal{F}_s] &= e^{\frac{t}{2}|\lambda|^2} \mathbb{E}\left[e^{i\langle \lambda, B_t - B_s + B_s \rangle}|\mathcal{F}_s\right] \\ &= e^{\frac{t}{2}|\lambda|^2} \underbrace{\mathbb{E}\left[e^{i\langle \lambda, B_t - B_s \rangle}|\mathcal{F}_s\right]}_{B_t - B_s \perp \mathcal{F}_s} e^{i\langle \lambda, B_s \rangle} \\ &= e^{\frac{t}{2}|\lambda|^2} e^{i\langle \lambda, B_s \rangle} \mathbb{E}\left[e^{i\langle \lambda, \overbrace{B_t - B_s}^{B_{t-2}} \rangle}\right] \\ &= e^{\frac{t}{2}|\lambda|^2} e^{i\langle \lambda, B_s \rangle} e^{-\frac{t-s}{2}\lambda^2}.\end{aligned}$$

characteristic function of $N(0, t-s)$

□

Well. But what is the point of any martingale in general? For example, consider

$$\mathbb{P} \left(\sup_{0 < t < \theta} B_t \geq b \right) \leq \exp \left\{ -\frac{b^2}{2\theta} \right\}.$$

This means that the probability that the Brownian motion exceeds a certain value b at any point in time in the interval $(0, \theta)$ decreases exponentially with $\frac{b^2}{\theta}$, which means that the probability of the Brownian motion reaching or exceeding a large value b in the interval decreases rapidly as b increases or θ (the time window) becomes large. We will be able to improve this estimation of the probability paths, but this is a simple result obtained from martingale properties.

We can use Doob's maximal inequality. Consider

$$\begin{aligned} \mathbb{P} \left(\sup_{0 < t < \theta} B_t \geq b \right) &\stackrel{\lambda \geq 0}{=} \mathbb{P} \left(e^{\lambda B_t} \geq e^{\lambda b} \right) \\ &\leq \frac{\mathbb{E}[e^{\lambda B_\theta}]}{e^{\lambda b}} \\ &= \exp \left\{ \frac{1}{2} \lambda^2 \theta \right\} e^{-\lambda b}. \end{aligned}$$

Since $\lambda > 0$ is arbitrary, we choose λ to minimize the right hand side. The result is

$$\lambda = \frac{b}{\theta}.$$

Remind that if M is a martingale in L^p , $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$ then

$$\mathbb{E} \max_{k \in (1, n)} |M_k|^p \leq q^p \mathbb{E} [|M_n|^p].$$

This inequality holds true also for continuous time. Let's apply it to BM:

$$\mathbb{E} \left[\sup_{0 < s < \theta} |B_s|^p \right] \leq \left(\frac{p}{p-1} \mathbb{E} [|B_\theta|^p] \right).$$

Of course $\frac{1}{q} = 1 - \frac{1}{p} \implies q = \frac{p}{p-1}$.

Definition 1.6.2

Take a filtration $\mathcal{F} = (\mathcal{F}_t)$. We say that it is **augmented** if $(\Omega, \mathcal{H}, \mathbb{P})$ is complete (i.e. each negligible set is measurable) and all negligible events in \mathcal{H} are also in \mathcal{F}_0 (and hence in \mathcal{F}_t).

Definition 1.6.3

Let \mathcal{F} be a filtration. Define

$$\mathcal{F}_{t^+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon} \quad t \in \mathbb{R}^+.$$

In this case \mathcal{F}_{t^+} is again a filtration and it is finer than \mathcal{F}_t . Imagine we want to compute

$$v_t = \lim_{\varepsilon \rightarrow 0} \frac{x_{t+\varepsilon} - x_t}{\varepsilon}$$

which is like calculating the velocity of X (given the position), where we need an infinitesimal quantity of information beyond t . This is what the filtration \mathcal{F}_{t^+} tells us. v_t is defined on \mathcal{F}_{t^+} but not on \mathcal{F}_t .

Definition 1.6.4

If $\mathcal{F}_t = \mathcal{F}_{t^+}$, $t \in \mathbb{R}^+$, the filtration is said to be **right continuous**.

For example, the Levy process does not have a right continuous filtration.

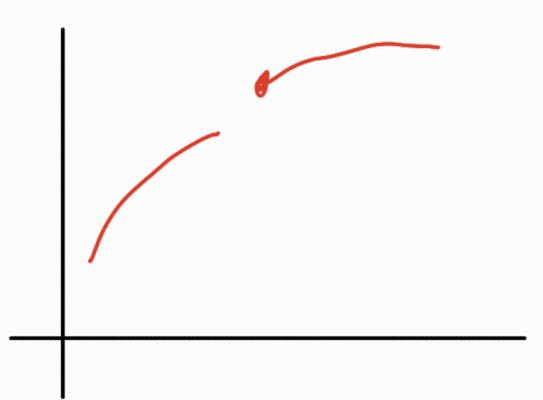


Figure 1.17: It's broken!

On the other hand, BM has right-continuous natural filtration.

Consider two continuous positive random variables T, V . Define

$$X_t(\omega) = \begin{cases} V(\omega)t & t < T(\omega) \\ V(\omega)T(\omega) & t \geq T(\omega). \end{cases}$$

We are describing, for example, the motion of a particle that moves from the origin with speed V until T and then stops after T . Consider the filtration generated by X . Is T a stopping time of \mathcal{F} ? When $X_s(\omega) = V(\omega)s$ we still don't know if $T(\omega) = s$ or $T(\omega) > s$... but

$$\{T \leq t\} \in \mathcal{F}_{t+\varepsilon} \forall \varepsilon > 0 \implies \{T \leq t\} \in \mathcal{F}_{t^+}.$$

We know that with continuous distributions, $T = t$ has probability 0... but this doesn't mean that the object doesn't exist, just that it is negligible.

1.6.2 Boundaries of BM

Consider

$$\begin{aligned} \tau_A^0 &:= \{t \geq 0 : X_t \in A\} && \text{first entry time in } A. \\ \tau_A &:= \inf \{t > 0 : X_t \in A\} && \text{first hitting time in } A. \\ \tau_{A^c} &:= \inf \{t \geq 0 : X_t \notin A\} && \text{first exit time from } A. \end{aligned}$$

If the inferior is the null set we say $\inf \{\emptyset\} = \infty$ and

$$\tau_A^0 \leq \tau_A.$$

Lemma 1.6.1

Consider a d -dimensional stochastic process $\{X_t\}$ with continuous sample paths and $A \subset \mathbb{R}^d$ closed. Then τ_A^0 is an \mathcal{F}_t^X stopping time and τ_A is an $\mathcal{F}_{t^+}^X$ stopping time.

The difference is that τ_A^0 includes the 0, while τ_A doesn't. We must have a peek of what comes next.

Let's take into consideration the one dimensional case. Consider $\tau_b := \tau_{\{b\}}^0$, the first entry time into the closed set $\{b\}$. Our problem consists in looking when the BM crosses the boundary b . Will it even do it? We are asking

$$\mathbb{P}(\tau_b < \infty) = ?$$

In other words, is it a sure event?

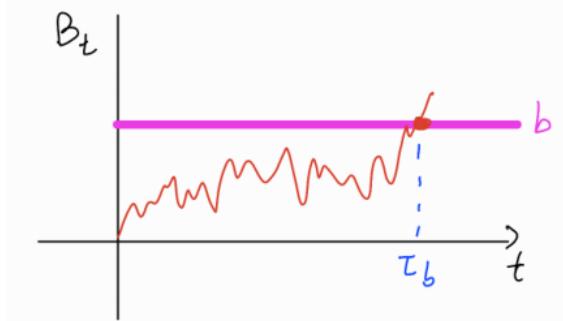


Figure 1.18: Wasn't b a set? I don't even care anymore.

In general, X_t can be a continuous stochastic process and b can depend on time (being a function $b(t)$).

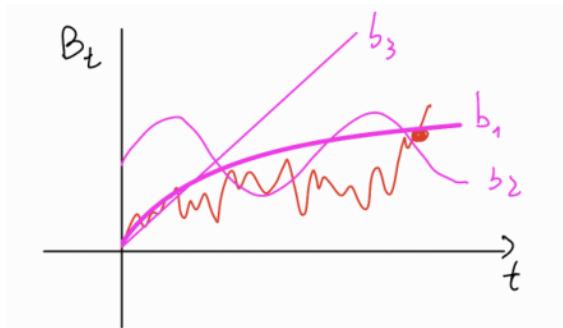


Figure 1.19: The point of entry shouldn't be the same for all three functions but you hopefully get the point.

Sometimes we may be able to prove analitically how $\tau_{b(t)}$ is distributed, sometimes not. What we care about now is to understand whether it even passes the threshold. For example. GPS satellites works using atomic clocks to triangulate the position and a small error in time can completely fuck your position up. Atomic clocks must always be aligned and one of the models used to compute the error of the clocks is the integral of Brownian motion, so yeah. Is pretty important to know how this bitch behaves.

Remark

Of course,

$$\sup_{t \geq 0} B_t \geq \sup_{n \geq 1} B_n$$

since if I start to consider the process earlier I will get at least the maximum I would obtain if I started later. But

$$\sup_{n \geq 1} B_n = \sup_{n \geq 1} (\xi_1 + \xi_2 + \dots + \xi_n)$$

with $\xi_i = (B_i - B_{i-1})$ i.i.d. by definition of Brownian motion. Now define

$$S := \sup_{n \geq 1} \sum_{i=1}^n \xi_i = \xi_1 + \sup_{n \geq 1} \sum_{i=2}^n \xi_i = \xi_1 + S'.$$

Note that

- ξ_i 's are i.i.d. so $S \sim S'$;
- ξ_1 is independent from S' ;
- ξ_1 is not trivial, so it is not 0.

But this means that $S = \infty$ a.s. because this would be the only case in which the sum stays the same after I add a non-trivial quantity!

We can apply the same idea to the infimum and say that

$$\mathbb{P}(\sup_{t \geq 0} B_t = \infty, \inf_{t \geq 0} B_t = -\infty) = 1.$$

This remark means that

$$\mathbb{P}(\tau_b < \infty) = 1$$

because if I can go from $-\infty$ to ∞ then I cross every value at least twice! This means that I can

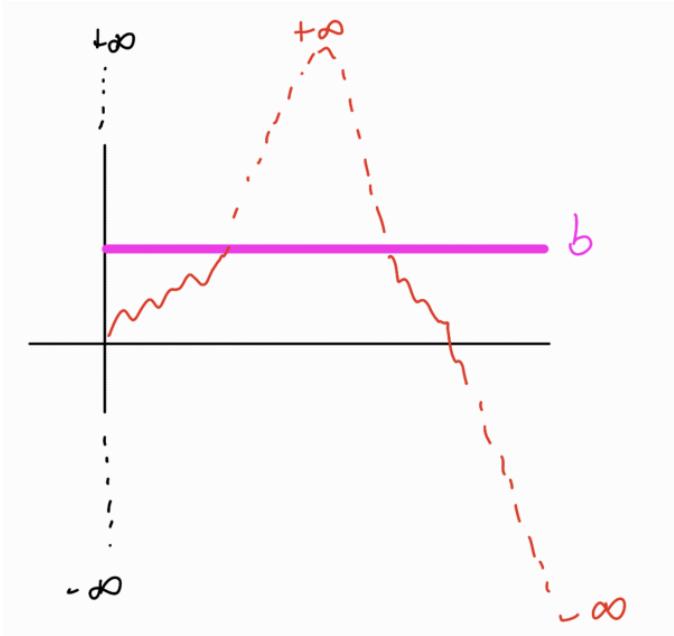


Figure 1.20: Now the lines are more straight, c'mon.

“split” the process at every level b . Not only that, but it will cross the threshold of b infinitely often since it has to go to and fro infinity infinite times! Like Markov chains, Brownian motion is **point recurrent**: each time the process leaves a state b it will return there sooner or later.

We proved that Brownian motion has the martingale property. We will now use the Doob's optional stopping theorem.

Theorem 1.6.1

Doob's optional stopping theorem. Let M be a stochastic process adapted to \mathcal{F} . The following are equivalent:

1. M is a \mathcal{F} -submartingale.
2. For every bounded stopping times S, T with $S \leq T$ the random variables M_S and M_T are integrable and

$$\mathbb{E}_S [M_T - M_S] \geq 0.$$

This means that the conditional expectation satisfies the submartingale property even after stopping.

Theorem 1.6.2

Wald's identity. Let (B_t, \mathcal{F}_t) be a BM¹ and assume that τ is a \mathcal{F}_t stopping time. If $\mathbb{E}\tau < \infty$ then $B_\tau \in L^2$ and

$$\begin{aligned} \mathbb{E}B_\tau &= 0 \\ \mathbb{E}B_\tau^2 &= \mathbb{E}\tau. \end{aligned}$$

Here we are relating a second moment with a first moment expectation. This tells us that the expectation of the Brownian motion is 0 also on stopping times (we already saw this with deterministic times) and its variance is equal to the mean value of τ .

Proof

We need Doob's stopping theorem. Introduce $(\tau \wedge t)_{t \geq 0}$: this gives us a bounded stopping time for Doob's Stopping theorem.

1. We know that $(B_{\tau \wedge t}, \mathcal{F}_{\tau \wedge t})$ is a martingale and

$$\mathbb{E}B_{\tau \wedge t} = \mathbb{E}B_0 = 0.$$

2. We must prove that $B_\tau \in L^2$. Use the inequality of the maximum:

$$\mathbb{E} \left[\sup_{s \leq t} |B_s|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}|B_t|^p.$$

Choose $p = 2$. We have

$$\mathbb{E} [B_{\tau \wedge t}^2] \leq \mathbb{E} \left[\sup_{s \leq t} B_s^2 \right] \leq 4 \mathbb{E} [B_t^2] = 4t$$

hence

$$B_{\tau \wedge t} \in L^2.$$

3. Use Doob's optional stopping theorem applies to the $\mathcal{F}_{\tau \wedge t}$ martingale: define

$$M_t^2 := B_{\tau \wedge t}^2 - \tau \wedge t$$

so that

$$\mathbb{E}M_t^2 = \mathbb{E}M_0^2 = 0 \implies \mathbb{E} [B_{\tau \wedge t}^2] = \mathbb{E} [\tau \wedge t].$$

We now need to remove t . We have to find a L^2 Cauchy sequence to prove that there is a limit that we can take.

Remark

For all $s \leq t$ we have that

$$\mathbb{E}[B_{\tau \wedge t} B_{\tau \wedge s}] = \mathbb{E}[B_{\tau \wedge s}^2].$$

So

$$\begin{aligned}\mathbb{E}[(B_{\tau \wedge t} - B_{\tau \wedge s})^2] &= \mathbb{E}[B_{\tau \wedge t}^2 + B_{\tau \wedge s}^2 - 2B_{\tau \wedge s}^2] \\ &= \mathbb{E}[B_{\tau \wedge t}^2 - B_{\tau \wedge s}^2] \\ &= \mathbb{E}[\tau \wedge t - \tau \wedge s] \xrightarrow{s,t \rightarrow \infty} 0.\end{aligned}$$

So $(B_{\tau \wedge t})$ is an L^2 Cauchy sequence. Furthermore, we know that $\tau < \infty$ a.s. by hypothesis and we know that B_t has continuous samples, so we can get almost sure convergence:

$$L^2 - \lim_{t \rightarrow \infty} B_{\tau \wedge t} = B_\tau.$$

Using the fact that $B_\tau \in L^2$ and using the monotone convergence theorem we get that

$$\mathbb{E}[B_\tau^2] = \lim_{t \rightarrow \infty} \mathbb{E}[B_{\tau \wedge t}^2] = \lim_{t \rightarrow \infty} \mathbb{E}[\tau \wedge t] = \mathbb{E}[\tau].$$

Finally, L^2 convergence implies L^1 convergence:

$$\mathbb{E}B_\tau = \lim_{t \rightarrow \infty} B_{\tau \wedge t} = 0.$$

□

Corollary

Let $(B_t)_{t \geq 0}$ be a one-dimensional Brownian motion and let

$$\tau := \inf_{t \geq 0} \{t \geq 0 : B_t \notin (-a, b)\}$$

be the first entry time into the set $(-a, b)^C$ with $a, b \geq 0$.

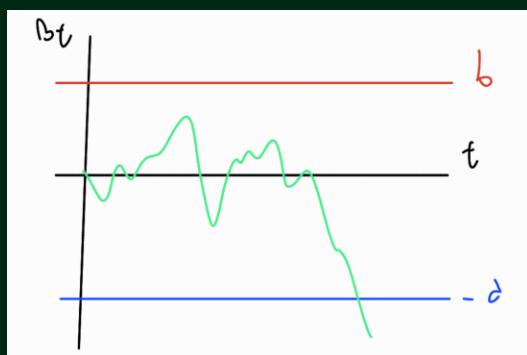


Figure 1.21: I completely tuned out from this part of the lecture on.

Then

$$\begin{aligned}\mathbb{P}(B_\tau = -a) &= \frac{b}{a+b} \\ \mathbb{P}(B_\tau = b) &= \frac{a}{a+b} \\ \mathbb{E}\tau &= ab.\end{aligned}$$

Proof

We know that $\tau \wedge t$ is a bounded stopping time and that $B_{\tau \wedge t} \in [a, b]$: hence

$$|B_{\tau \wedge t}| \leq a \vee b.$$

But by the previous theorem we know that

$$\mathbb{E}[\tau \wedge t] = \mathbb{E}[B_{\tau \wedge t}^2] \leq a^2 \vee b^2.$$

Now use the monotone convergence theorem to get that

$$\mathbb{E}\tau \leq a^2 \vee b^2 \leq \infty.$$

From Wald's identities we have that

$$\begin{aligned}\mathbb{E}B_t &= 0 \\ -a\mathbb{P}(B_\tau = -1) + b\mathbb{P}(B_\tau = b) &= 0.\end{aligned}$$

Furthermore, we know

$$\mathbb{P}(B_\tau = -a) + \mathbb{P}(B_\tau = b) = 1.$$

So we have a nice little system that we can solve to get

$$\begin{aligned}\mathbb{P}(B_\tau = -a) &= \frac{b}{a+b} \\ \mathbb{P}(B_\tau = b) &= \frac{a}{a+b}.\end{aligned}$$

Now use the second Wald's identity to get that

$$\begin{aligned}\mathbb{E}_\tau &= \mathbb{E}B_\tau^2 \\ &= a^2\mathbb{P}(B_\tau = -a) + b^2\mathbb{P}(B_\tau = b) \\ &= a^2 \frac{b}{a+b} + b^2 \frac{a}{a+b} \\ &= ab.\end{aligned}$$

□

Remark

If $a \rightarrow \infty$ (which means we are in the one boundary case, or the *absorbing barrier* case):

1. $\mathbb{P}(B_\tau = b) = \lim_{a \rightarrow \infty} \frac{a}{a+b} = 1;$

2. if we want

$$\tau_1 = \inf_{t \geq 0} \{t \geq 0 : B_t = 1\}$$

then we have

$$\mathbb{E}\tau_1 = \lim_{a \rightarrow \infty} ab = \infty.$$

This means heavy tails for the distribution of this stopping time.

Remember that the one-dimensional Brownian motion is recurrent, meaning that it visits 1 infinitely often... but the time to attain 1 diverges.

Proof

We know that

$$\tau < \infty \quad \text{a.s.}$$

and

$$\mathcal{B}_{\tau_1} = 1 \text{ and } \mathbb{E}B_{\tau_1} = \mathbb{E}B_{\tau_1}^2 = 1.$$

This means that

$$\mathbb{E}\tau = \infty.$$

□

Exercises

Exercise 1

Let $(B_t)_{t \geq 0}$ be a Brownian motion in dimension 1 on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $(\mathcal{R}_t)_{t \geq 0}$ be its natural filtration. Define

$$X_t := \int_0^t B_s \, ds$$

and

$$M_t := X_t - \frac{1}{3}B_t^3.$$

Prove that:

1. $\mathbb{E}[B_t^3 | \mathcal{F}_s] = B_s^3 + 3B_s(t-s)$;
2. $\mathbb{E}\left[\int_s^t B_u \, du | \mathcal{F}_s\right] = (t-s)B_s$. Hint: use the following result

$$\mathbb{E}\left[\int_s^t Y_u \, du | \mathcal{G}\right] = \int_s^t \mathbb{E}[Y_u | \mathcal{G}] \, du$$

that holds for each $0 \leq s \leq t$ and each σ -algebra $\mathcal{G} \subset \mathcal{F}$ and each process $Y = (Y_u)_{u \in (0, \infty)}$ continuous and with $Y_u^2 \in L^2$;

3. M is a martingale.

1. We know that

$$\begin{aligned}
\mathbb{E}[B_t^3 | \mathcal{F}_s] &= \mathbb{E}[((B_t - B_s) + B_s)^3 | \mathcal{F}_s] \\
&= \mathbb{E}\left[B_s^3 + \underbrace{3B_s(B_t - B_s)^2}_{\mathcal{F}_s\text{-meas}} + \underbrace{3B_s^2(B_t - B_s)}_{\mathcal{F}_s\text{-meas}} + (B_t - B_s)^3 | \mathcal{F}_s\right] \\
&= B_s^3 + 3B_s \mathbb{E}\left[\underbrace{(B_t - B_s)^2}_{\perp \text{ from } \mathcal{F}_s} | \mathcal{F}_s\right] + 3B_s^2 \mathbb{E}\left[\underbrace{B_t - B_s}_{\perp \text{ from } \mathcal{F}_s} | \mathcal{F}_s\right] + \mathbb{E}\left[\underbrace{(B_t - B_s)^3}_{\perp \text{ from } \mathcal{F}_s} | \mathcal{F}_s\right] \\
&= B_s^3 + 3B_s \underbrace{\mathbb{E}[(B_t - B_s)^2]}_{t-s} + 3B_s^2 \underbrace{\mathbb{E}[B_t - B_s]}_0 + 3 \underbrace{\mathbb{E}[B_t - B_s]^2}_0 \\
&= B_s^3 + 3B_s(t - s).
\end{aligned}$$

2. Using the suggested hint we get:

$$\begin{aligned}
\mathbb{E}\left[\int_s^t B_u du | \mathcal{F}_s\right] &= \int_s^t \mathbb{E}[B_u | \mathcal{F}_s] du \\
&= \int_s^t \underbrace{B_s du}_{B_s \text{ does not depend on } u} \\
&= B_s \int_s^t du \\
&= B_s(t - s).
\end{aligned}$$

3. (a) The process $(M_t)_{t \geq 0}$ is \mathcal{F}_t -adapted by definition. ✓

(b) $M_t \in L^1$ for $\forall t \geq 0$. We want to show that $\mathbb{E}[|M_t|] < \infty$. We know that

$$M_t = X_t - \frac{1}{3}B_t^3 \implies \mathbb{E}[M_t] = \mathbb{E}[X_t] - \frac{1}{3}\mathbb{E}[B_t^3].$$

We know that $\mathbb{E}[B_t^3] = 0$ because the odd moments of a Gaussian are zero (really?) and we know that

$$\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}\left[\int_0^t B_s ds\right] \\
&= \int_0^t \mathbb{E}[B_s] ds \quad \leftarrow \text{Fubini's theorem} \\
&= 0
\end{aligned}$$

so

$$\mathbb{E}[M_t] = 0 - \frac{1}{3} \cdot 0 = 0 < \infty. \quad \checkmark$$

(c) For $0 < s < t$ we have

$$\begin{aligned}
\mathbb{E}[M_t | \mathcal{F}_s] &= \mathbb{E}[X_t | \mathcal{F}_s] - \frac{1}{3}\mathbb{E}[B_t^3 | \mathcal{F}_s] \\
&= \mathbb{E}[X_t - X_s + X_s | \mathcal{F}_s] - \frac{1}{3}\mathbb{E}[B_t^3 | \mathcal{F}_s] \\
&= X_s + \mathbb{E}[X_t - X_s | \mathcal{F}_s] - \frac{1}{3}\mathbb{E}[B_t^3 | \mathcal{F}_s] \\
&= X_s + \mathbb{E}\left[\int_s^t B_u du | \mathcal{F}_s\right] - \frac{1}{3}\mathbb{E}[B_t^3 | \mathcal{F}_s] \\
&= X_s + B_s(t - s) - \frac{1}{3}(B_s^3 + 3B_s(t - s)) \\
&= M_s. \quad \checkmark
\end{aligned}$$

Exercise 2

Is

$$Y_t = \frac{1}{3}B_t^3 - tB_t$$

a martingale with respect to the natural filtration of B_t ?

1. We know that Y_t is adapted to \mathcal{F}_t . ✓
2. We need to show that Y_t is integrable, that is

$$\mathbb{E} \left[\left| \frac{1}{3}B_t^3 - tB_t \right| \right] < \infty.$$

So

$$\begin{aligned} \mathbb{E} [|Y_t|] &= \mathbb{E} \left[\left| \frac{1}{3}B_t^3 - tB_t \right| \right] \\ &\stackrel{\triangle \text{ ineq}}{\leq} \frac{1}{3}\mathbb{E} [|B_t^3|] + t\mathbb{E} [|B_t|] \\ &= \underbrace{\frac{1}{3}\mathbb{E} [|B_t|^3]}_{<\infty} + t\underbrace{\mathbb{E} [|B_t|]}_{<\infty}. \end{aligned} \quad \checkmark$$

3. We know that

$$\begin{aligned} \mathbb{E} [Y_t | \mathcal{F}_s] &= \mathbb{E} \left[\frac{1}{3}(B_t - B_s + B_s)^3 - tB_t | \mathcal{F}_s \right] \\ &= \frac{1}{3}\mathbb{E} [(B_t - B_s)^3 | \mathcal{F}_s] + \frac{1}{3}\mathbb{E} [B_s^3 | \mathcal{F}_s] + \mathbb{E} [(B_t - B_s)^2 B_s | \mathcal{F}_s] + \\ &\quad + \mathbb{E} [(B_t - B_s)B_s^2 | \mathcal{F}_s] - t\mathbb{E} [B_t - B_s | \mathcal{F}_s] - tB_s \\ &= \underbrace{\frac{1}{3}\mathbb{E} [(B_t - B_s)^3]}_{=0} + \frac{1}{3}B_s^3 + B_s(\cancel{t} - s) + B^2 \underbrace{\mathbb{E} [B_t - B_s]}_{=0} - \cancel{t}B_s \\ &= \frac{1}{3}B_s^3 + sB_s. \end{aligned} \quad \checkmark$$

Exercise 3

Is there any implication between Markov and Martingale properties? Justify the answer with examples.

In practice: no. A process can be Markov without having martingale property. For example, think about the Brownian motion with drift μ :

$$X_t = B_t + \mu t.$$

This is Markov (because it has independent increments) but it is not a martingale (it is a sub-martingale, though):

$$\begin{aligned} \mathbb{E} [X_t | \mathcal{F}_s] &= \mathbb{E} [B_t + \mu t | \mathcal{F}_s] \\ &= \underbrace{\mathbb{E} [B_t | \mathcal{F}_s]}_{B_s} + \underbrace{\mathbb{E} [\mu t | \mathcal{F}_s]}_{\mu t} > X_s. \end{aligned}$$

We can also find a process that is martingale but not Markov: consider $(\xi_n)_{n \geq 0}$ i.i.d. random variables with $\mathbb{E}[\xi_i] = 0$ for $\forall i$. Consider the random variable

$$X_n = \begin{cases} X_0 : \text{random variable independent from } (\xi_n) \\ X_{n+1} := X_n + \xi_{n+1} X_0. \end{cases}$$

In this case X_n is \mathcal{F}_n -martingale:

$$\begin{aligned} \mathbb{E}[X_{n+1} | \mathcal{F}_n] &= \underbrace{\mathbb{E}[X_n | \mathcal{F}_n]}_{X_n} + \underbrace{\mathbb{E}[\xi_{n+1} X_0 | \mathcal{F}_n]}_{\xi \perp\!\!\!\perp X_0} \\ &= X_n + \underbrace{\mathbb{E}[\xi_{n+1} | \mathcal{F}_n]}_{\perp\!\!\!\perp \mathcal{F}_n} \cdot \underbrace{\mathbb{E}[X_0 | \mathcal{F}_n]}_{\mathcal{F}_n-\text{meas.}} \\ &= X_n + \underbrace{\mathbb{E}[\xi_{n+1}]}_{=0} X_0 = X_n \end{aligned}$$

but it is not Markov:

$$\mathbb{P}(\underbrace{X_n + \xi_{n+1} X_0 \leq x}_{X_{n+1}} | \mathcal{F}_n) \neq \mathbb{P}(X_n + \xi_{n+1} X_0 \leq x | X_n).$$

This is because in this case X_{n+1} depends also on X_0 , not only on X_n . In other words, the filtration generated by X_n is “not enough” to explain the probability of X_{n+1} .

Exercise 4

Let

$$p(x, t) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp\left\{-\frac{|x|^2}{2t}\right\}$$

be the transition probability density function of a d -dimensional Brownian motion.

1. Show that $p(x, t)$ is a solution of the heat equation

$$\frac{\partial p(x, t)}{\partial t} = \frac{1}{2} \Delta_x p(x, t) = \frac{1}{2} \sum_1^d \frac{\partial^2 p(x, t)}{\partial x^2}.$$

2. Is the proposed solution unique?

1. Thank you! This was just what I needed! Some fucking partial differential equations!

Yeah! Cool! Anyway, we can check this by substituting $p(x, t)$ in the partial differential equation.

$$\begin{aligned} \frac{\partial p(x, t)}{\partial t} &= \frac{\partial}{\partial t} \left(\frac{1}{(2\pi t)^{\frac{d}{2}}} \exp\left\{-\frac{|x|^2}{2t}\right\} \right) \\ &= \frac{\partial}{\partial t} \left(\frac{1}{(2\pi t)^{\frac{d}{2}}} \right) \cdot \exp\left\{-\frac{|x|^2}{2t}\right\} + \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \frac{\partial}{\partial t} \left(\exp\left\{-\frac{|x|^2}{2t}\right\} \right) \\ &= -\frac{d}{2} \frac{1}{(2\pi t)^{\frac{d}{2}+1}} \cdot \exp\left\{-\frac{|x|^2}{2t}\right\} + \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \exp\left\{-\frac{|x|^2}{2t}\right\} \cdot \frac{|x|^2}{2t^2} \\ &= \exp\left\{-\frac{|x|^2}{2t}\right\} \left(-\frac{d}{2} \cdot \frac{1}{(2\pi t)^{\frac{d}{2}+1}} + \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \frac{|x|^2}{2t^2} \right). \end{aligned} \tag{△}$$

We know that the Laplacian is the sum of second derivatives with respect to the spatial dimensions

$$\Delta_x p = \nabla \cdot \nabla p = \nabla^2 p = \sum_{i=1}^d \frac{\partial^2 p}{\partial x_i^2}.$$

Cool! You just learned about the divergence operator ∇^2 , that is the dot product (= derivative, in this case) of the gradient with itself! We know that

$$\nabla^2 \exp \left\{ -\frac{|x|^2}{2t} \right\} = \left(\frac{|x|^2}{t^2} - \frac{d}{t} \right) \exp \left\{ -\frac{|x|^2}{2t} \right\}$$

So we can add the constant to the equation and get that

$$\begin{aligned} \Delta_x p(x, t) &= \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \left(\frac{|x|^2}{t^2} - \frac{d}{t} \right) \exp \left\{ -\frac{|x|^2}{2t} \right\} \\ &= \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \left(\frac{|x|^2 - dt}{t^2} \right) \exp \left\{ -\frac{|x|^2}{2t} \right\}. \end{aligned}$$

We can further simplify Δ to get our result:

$$\begin{aligned} \exp \left\{ -\frac{|x|^2}{2t} \right\} \left(-\frac{d}{2} \cdot \frac{1}{(2\pi t)^{\frac{d}{2}+1}} + \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \frac{|x|^2}{2t^2} \right) &= \exp \left\{ -\frac{|x|^2}{2t} \right\} \cdot \frac{1}{(2\pi t)^{\frac{d}{2}}} \left(-\frac{d}{2} \cdot \frac{1}{t} + \frac{|x|^2}{2t^2} \right) \\ &= \exp \left\{ -\frac{|x|^2}{2t} \right\} \cdot \frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \left(\frac{|x|^2 - dt}{2t^2} \right) \\ &= \frac{1}{2} \exp \left\{ -\frac{|x|^2}{2t} \right\} \cdot \underbrace{\frac{1}{(2\pi t)^{\frac{d}{2}}} \cdot \left(\frac{|x|^2 - dt}{t^2} \right)}_{\Delta_x p(x, t)}. \end{aligned}$$

2. The answer is no: this solution verifies the equation for $x \in -\infty, +\infty$. Other functions can verify the same equation but with different boundary conditions.

Wait, so what did we learn? The heat equation describes the evolution of a temperature distribution. A BM^d can be thought of as the random motion of a particle in space, where the position $X(t)$ of the particle at time t follows a probability distribution governed by the heat equation. So heat equation successfully model how the random motion of a particle that we expect moving with a Brownian motion!

Exercise 5

Consider a BM¹, $B = (B_t)_{t \geq 0}$ and define the BM with drift as

$$X(t) = \sigma B(t) + \mu t$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

1. Is it a sub/super martingale?
2. If the answer to the first question was affirmative, write the Doob's decomposition of $X(t)$.

1. If $\mu > 0$ then $X(t)$ is a sub-martingale, if $\mu < 0$ then $X(t)$ is a super-martingale. Indeed, we know that $\mathbb{E}[B(t)|\mathcal{F}_s] = B(s)$ for $s \leq t$, so

$$\begin{aligned}\mathbb{E}[X(t)|\mathcal{F}_s] &= \mathbb{E}[\sigma B(t) + \mu t|\mathcal{F}_s] \\ &= \sigma \mathbb{E}[B(t)|\mathcal{F}_s] + \mu t \\ &= \sigma B(s) + \mu t.\end{aligned}$$

We know that

$$X(s) = \sigma B(s) + \mu s$$

Just close... For it to be a martingale we should have $\mu t = \mu s$, but according to the sign of μ we can tell whether the quantity that we got is greater or smaller than $X(s)$ and is therefore a sub or a super martingale.

2. $X(t) = \sigma B(t) + \mu t$ is already the unique decomposition! $\sigma B(t)$ is a martingale and μt is a predictable process.

1.7 Brownian motion as a Markov Process

1.7.1 Markov Property

Definition 1.7.1

A stochastic process $(X_t)_{t \geq 0}$ verifies the **Markov property** with respect to a filtration $\mathcal{F}_{tt \geq 0}$ if it holds

$$\mathbb{P}(X_t \in B | \mathcal{F}_s) = \mathbb{P}(X_t \in B | X_s) \quad \forall s, t \in \mathbb{R}, \quad 0 \leq s < t, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Theorem 1.7.1

The Brownian motion is a **Markov process with respect to its natural filtration**.

Proof

We want to prove that

$$\mathbb{P}(B_t \in A | \mathcal{F}_s) = \mathbb{P}(B_t \in A | B_s).$$

Consider

$$\begin{aligned}\mathbb{P}(B_t \in A | \mathcal{F}_s) &= \mathbb{E}[\mathbb{1}_A(B_t) | \mathcal{F}_s] \\ &= \mathbb{E}\left[\mathbb{1}_A(\underbrace{B_t - B_s}_{\perp \mathcal{F}_s} + \underbrace{B_s}_{\mathcal{F}_s\text{-meas, so we can write } B_s = X}) | \mathcal{F}_s\right] \\ &= \mathbb{E}[\mathbb{1}_A(B_t - B_s + X)] \\ &= \mathbb{P}((B_t - B_s + X) \in A).\end{aligned}\tag{\(\cup\)}$$

Here the X means that B_s is already known at time s . Now consider the right hand side of \cup . With analogous steps you get

$$\mathbb{P}(B_t \in A | B_s) = \mathbb{P}((B_t - B_s + X) \in A)\tag{\(\cup\)}$$

So combining  and  we get

$$\mathbb{P}(B_t \in A | B_s) = \mathbb{P}(B_t \in A | \mathcal{F}_s).$$

□

Can we “split” the sample path at stopping (that is, random) times?

Definition 1.7.2

Define the process $X_T(\omega)$ as:

$$X_T(\omega) = \begin{cases} X_T(\omega) & \text{if } T(\omega) < \infty \\ \lim_{t \rightarrow \infty} X_t(\omega) & \text{if the limit exists} \\ 0 & \text{otherwise.} \end{cases}$$

We know the Markov property:

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s).$$

Here we are working with deterministic times, but does this hold also when T is a stopping time?

Definition 1.7.3

We say that a stochastic process has **strong Markov property** if

$$\mathbb{P}(X_{T+h} \in A | \mathcal{F}_T) = \mathbb{P}(X_{T+h} \in A | X_T)$$

holds for all stopping times $T, h > 0$ and $\forall A \in \mathcal{B}(\mathbb{R})$.

Remark

For discrete processes, Markov and strong Markov properties coincide.

The problem lies with continuity. Take, for example, a Brownian motion BT that not necessarily starts from 0. Define

$$X_t = B_t \mathbb{1}_{\{B_0 \neq 0\}} = \begin{cases} B_t & \text{if } B_0 \neq 0 \\ 0 & \text{if } B_0 = 0. \end{cases}$$

Show that X_t is Markov but it is not strong Markov. The heuristic intuition is that this is not strong Markov because to know the next step we must know whether we started from 0 or not. It is, though, Markov because the next move depends from a certain previous state. We cannot “forget” the origin.

Definition 1.7.4

Let

$$\psi(x) = \mathbb{E}[f(X_t) | X_0 = x] \quad x \in \mathbb{R}$$

with f being a function and $t > 0$ fixed. We say that the process $X(t)$ has the **Feller property** if

$$f \in C_b \implies \psi \in C_b \quad (\boxed{\text{Feller}})$$

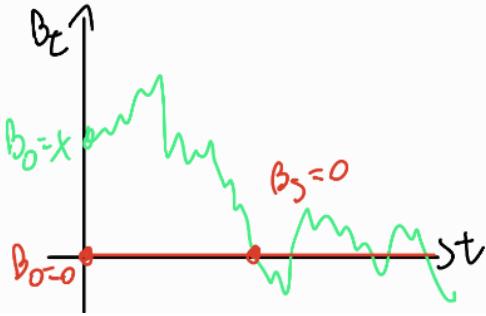


Figure 1.22: Consider $B_s = 0$, when the two paths coincide

where $C(b)$ is the space of continuous and bounded functions of \mathbb{R} .

Remark

The Brownian motion has the feller property:

$$\begin{aligned}\psi(x) &= \mathbb{E}[f(B_t) | B_0 = x] \\ &= \mathbb{E}\left[f\left(\frac{B_t}{\sqrt{t}} + x\right)\right] \quad \text{starting from } x \text{ means shifting} \\ &= \int_{\mathbb{R}} f(x+y) \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} dy\end{aligned}$$

and $\boxed{\psi}$ holds because since f is continuous and bounded also the shifted function $f(x+y)$ is. The Gaussian density is smooth and integrates to 1.

Theorem 1.7.2

The Brownian motion is a strong Markov process with respect to its natural filtration.

Remark

Strong Markov property can be rewritten as

$$\mathbb{E}[f(X_{T+h}) | \mathcal{F}_T] = \mathbb{E}[f(X_h) | X_0 = x] \Big|_{x=X_T}$$

for all stopping times T , all $h > 0$ and all continuous bounded functions $f : \mathbb{R} \rightarrow \mathbb{R}$.

Proof

1. Consider the discrete stopping time T_n :

$$\begin{aligned}\mathbb{E}\left[f(B_{T_n+h}) | \mathcal{F}_{T_n}^B\right] &= \sum_n \mathbb{1}_{\{T_n=n\}} \mathbb{E}\left[f(B_{T_n+h}) | \mathcal{F}_n^B\right] \quad (\text{conditioning all possibilities}) \\ &\stackrel{\text{markov}}{=} \sum_n \mathbb{1}_{\{T_n=n\}} \mathbb{E}\left[f(B_{T_n+h}) | B_n\right].\end{aligned}$$

But this means that, since there is only one term with $T_n = n$, at time n the process

restarts

$$\sum_n \mathbb{1}_{\{T_n=n\}} \mathbb{E} [f(B_h)|B_0 = x] \Big|_{B_{T_n}=B_n=X} = \mathbb{E} [f(B_h)|B_0 = x] \Big|_{B_n=x}.$$

The result works for T_n , so for discrete times we have the result.

2. Now move to arbitrary stopping times. When we want to prove something about markoviality, we always start with discrete times and then move to continuous times. We know that for any stopping time T there exists a sequence of decreasing stopping times $\{T_n\}$ converging to T :

$$\mathbb{E} \left[f(B_{T_n+h}) | \mathcal{F}_{T_n}^B \right] = \mathbb{E} [f(B_h)|B_0 = x] \Big|_{x=B_{T_n}} \quad \text{for } \forall n \geq 1. \quad (\text{III})$$

Revise with Kotatsu!

Remember **Hunt's lemma**: if $X_n \xrightarrow{n \rightarrow \infty} X$ a.s., $|X_n| \rightarrow Z, Z \in L^1$ and $\{\mathcal{F}_n\}$ is a filtration then

$$\mathbb{E} [X_n | \mathcal{F}_n] \rightarrow \mathbb{E} [X | \mathcal{F}_\infty] \quad \text{a.s.}$$

where $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$.

We take the limit of III for $n \rightarrow \infty$, then we apply Hunt's lemma to the left hand side and the Feller property to the right hand side:

$$\mathbb{E} [f(B_{T+h}) | \mathcal{F}_T^B] = \mathbb{E} [f(B_h)|B_0 = x] \Big|_{x=B_T}.$$

This allows us to say that every Feller process is strong Markov!

□

1.7.2 First passage times and Strong Markov Property

As we have seen before, we can split the sample paths of a BM^d at a stopping time. We are

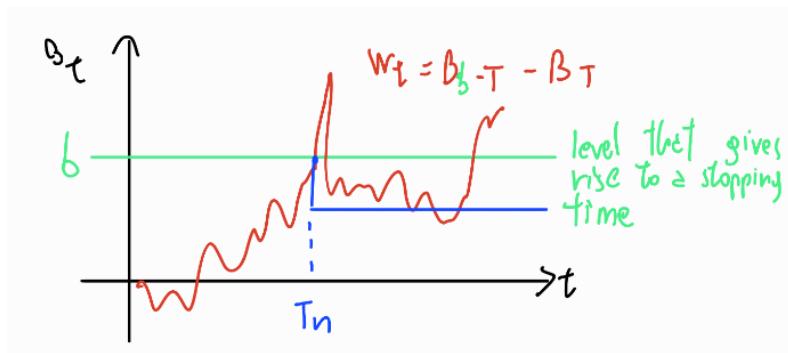


Figure 1.23: More diagrams... are these even useful?

interested in knowing the distribution of the first passage time through b.

Theorem 1.7.3

It holds

$$\mathbb{P}(\tau_b < t) = 2\mathbb{P}(B_t \geq b)$$

Proof

Assume B_t attained b at time τ_b . Stop at τ_b and start a new Brownian motion W_s from $B_{\tau_b} = b$

$$\begin{aligned} W_s + b &= \left(B_{\tau_b+s} - \underbrace{B_{\tau_b}}_b \right) + b \\ &= B_{\tau_b+s} \end{aligned}$$

and this is again a BM but started at b . The probability of being above or below b is the same, so

$$\begin{aligned} \mathbb{P}(\tau_b \leq t, B_t < b) &= \mathbb{P}\left(\underbrace{\{\tau_b \leq t\}}_{\in \mathcal{F}_{\tau_b}^B} \cap \underbrace{\{B_{\tau_b+(t-\tau_b)} - B_{\tau_b} < 0\}}_{\in \mathcal{F}_{\infty}^W \perp \mathcal{F}_{\tau_b}^B}\right) \\ &= \mathbb{P}(\tau_b \leq t) \underbrace{\mathbb{P}(B_{\tau_b+(t-\tau_b)} - B_{\tau_b} < 0)}_{\frac{1}{2} \text{ because of symmetry around } 0} \\ &= \frac{1}{2} \mathbb{P}(\tau_b \leq t). \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathbb{P}(\tau_b \leq t) &= \mathbb{P}(\tau_b \leq t, B_t \geq b) + \underbrace{\mathbb{P}(\tau_b \leq t, B_t < b)}_{\frac{1}{2} \mathbb{P}(\tau_b \leq t)} \\ &\implies \frac{1}{2} \mathbb{P}(\tau_b \leq t) = \mathbb{P}(B_t \geq b) \end{aligned}$$

□

We can compute the probability density of τ_b . We know

$$\mathbb{P}(\tau_b \leq t) = 2 \int_b^\infty \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{y^2}{2t}\right\} dy$$

Hence if I take the derivative

$$\begin{aligned} \frac{d}{dt} &= \frac{b}{2\pi t^3} \exp\left\{-\frac{b^2}{2t}\right\} \\ &= g(t) \end{aligned}$$

which is the probability distribution function of our first passage time. Now we can understand why the expectation of t diverges. Consider

$$\begin{aligned} \mathbb{E}[\tau_b] &= \int_0^\infty t \frac{b}{\sqrt{2\pi t^3}} \exp\left\{-\frac{b^2}{2t}\right\} dt \\ &= \int_0^\infty \frac{b}{\sqrt{2\pi t}} \exp\left\{-\frac{b^2}{2t}\right\} dt \\ &= \infty \end{aligned} \tag{1.1}$$

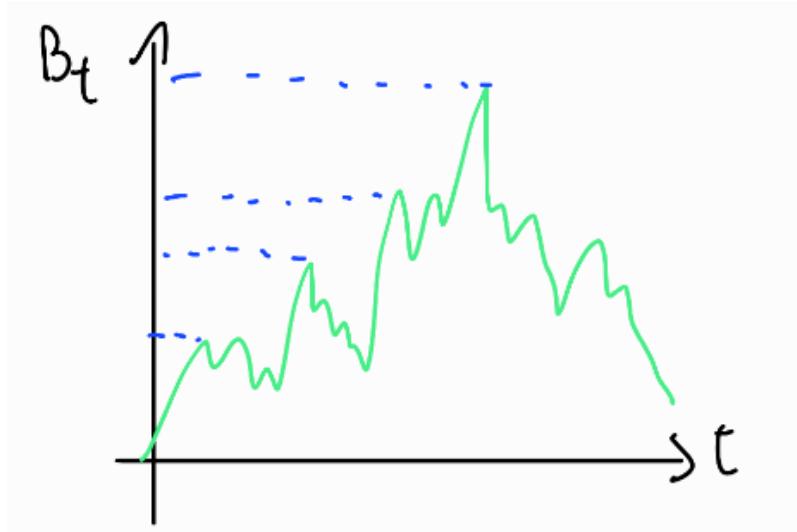


Figure 1.25: As time t evolves, its superior changes

so we have heavy tails!

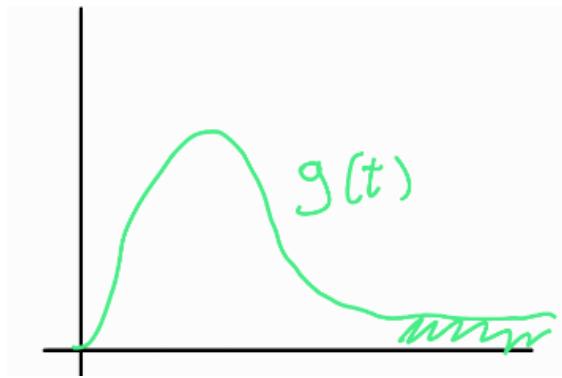


Figure 1.24: Metal

There is a related property about this: let

$$M_t := \sup_{\theta \leq t} B_\theta$$

It holds

$$\mathbb{P}(M_t \geq b) = 2\mathbb{P}(B_t \geq b)$$

Proof

Observe that

$$\{\tau_b \leq t\} = \{M_t \geq b\}.$$

□

M_t and $|B_t|$ have the same distribution but what about their sample paths? We know that:

- M_t is non decreasing;
- $|B_t|$ is not non decreasing.

Their sample paths is very different!

Remark

We have

$$M_t \sim M_t - B_t \quad \forall y \geq 0 \quad \text{given and fixed.}$$

Proof

By symmetry,

$$\begin{aligned} M_t &\sim \sup_{s \leq t} (-B_s) = \sup_{s \leq t} ((-B_s + B_t) - B_t) \\ &= \sup_{s \leq t} \left[\frac{(B_t - B_{t-s}) - B_s}{\tilde{B}_s} \right] \\ &= \sup_{s \leq t} \tilde{B}_s - B_t \sim M_t - B_t. \end{aligned}$$

□

1.7.3 First passage times for Brownian motion with drift

Remember that for a Brownian motion the first passage time probability is

$$\mathbb{P}(\tau_b < t) = \frac{b}{\sqrt{2\pi t^3}} \exp\left\{-\frac{b^2}{2t}\right\}$$

And that $\mathbb{P}(\tau_b < \infty) = 1$. For a process with strong Markov property (but not necessarily symmetry, like in the case of Brownian motion with drift) can we determine the distribution of the first passage time? This means looking for

$$\mathbb{P}(\tau_b < t) = \mathbb{P}(\tau_b < t | X(t_0) = x_0).$$

Taking the derivative we get the probability density function of this distribution:

$$\frac{d}{dt} \mathbb{P}(\tau_b < t) = g_{\tau_b}(x_0) = \dots?$$

Aw hell nah. This is something we want to know for sure. This can apply to processes with continuous sample paths with strong Markov property (that is, diffusion processes). Think about Brownian motion with drift $X(t) = \sigma B_t + \mu t$, $\mu > 0$, and its sample path:

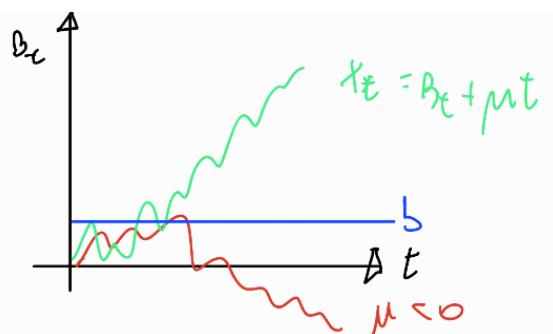


Figure 1.26: This is the start of a downward spiral, believe me.

If the drift is positive and $b > 0$ then we know that $\mathbb{P}(\tau_b^\mu < \infty) = 1$ because the process will incur in b sooner or later. The focal point is that $X_t \geq B_t$ a.s., that is $X(t)$ dominates B_t .

On the other hand, if $\mu' < 0$ then $\mathbb{P}(\tau_b^{\mu'} < \infty) < 1$ because the process will tend to escape from the bounds. It may happen that b is reached, but in general it doesn't happen. Consider the probability that our process is above b in figure 1.27. Here we look for $\mathbb{P}(\tau_b^{x_0} < t)$, that is

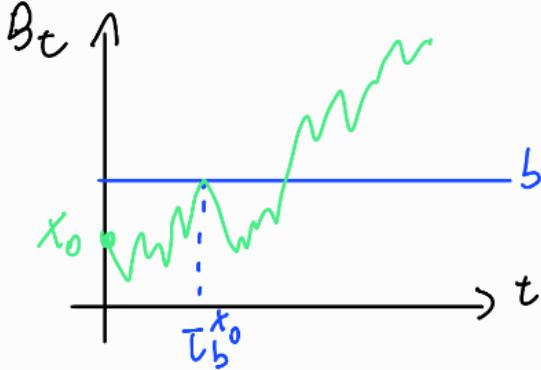


Figure 1.27: b is attained at $\tau_b^{x_0}$ after starting at x_0

the probability density function of $\tau_b^{x_0}$, *the first passage time in b with starting point $X(t_0) = x_0$* . This is like asking for the probability of $X(t)$ being greater than x , with $x > b$ and the process starting from $X(t_0) = x_0$:

$$\mathbb{P}(X(t) > x | X(t_0) = x_0) \quad x > b.$$



Inverting the signs we can introduce the notation for a cumulative distribution function

$$\mathbb{P}(X(t) < x | X(t_0) = x_0) = F(x, t | x_0, t_0)$$

and therefore define the survival function

$$\mathbb{P}(X(t) > x | X(t_0) = x_0) = 1 - F(x, t | x_0, t_0) = \bar{F}(x, t | x_0, t_0).$$

Furthermore, consider $g_b^{x_0}(\tau)$, the probability of attaining b at time τ with $X(t_0) = x_0$:

$$g_b^{x_0}(\tau) = g(b, \tau | X(t_0) = x_0).$$

Note that this is a function of τ but it is characterized by the parameters x_0 (the starting point) and b (the level we want to attain). Now let's go back to the probability of $X(t)$ being above level b (the one in equation 😊). We can rewrite it as

$$\underbrace{\mathbb{P}(X(t) > x | X(t_0) = x_0)}_{\bar{F}(x, t | x_0, t_0)} = \underbrace{\int_0^t g_b^{x_0}(\tau) \mathbb{P}(X(t) > x | X(\tau) = b) d\tau}_{\text{note the reversed integral}}.$$

What we did here was using strong Markov property to “split” the probability between the **probability of attaining b starting from x_0** and **a new process that starts from b** , seeking the probability of transitioning from b to x in the remaining time! But this also means that

$$\bar{F}(x, t | x_0, t_0) = \int_0^t g_b^{x_0}(\tau) \bar{F}(x, t | b, \tau) d\tau.$$

Now derive with respect to x :

$$f(x, t | x_0, t_0) = \int_0^t d\tau g_b^{x_0}(\tau) f(x, t | b, \tau) \quad \text{for } x > b.$$

here $f(x, t|x_0, t_0)$ is simply the **probability distribution function of the transition from x_0 to x at time t** .

In 1943 Fortet⁵ proved that this equation holds also when $x = b$, introducing the so-called Fortet equation:

$$f(b, t|x_0, t_0) = \int_0^t d\tau g_b^{x_0}(\tau) f(b, t|b, \tau). \quad \blacksquare$$

In the Brownian motion with drift, we have that the process $X(t) = \sigma B(t) + \mu t$ is distributed as $N(\mu(t - t_0), \sigma^2(t - t_0))$ and therefore our transition function is simply the normal distribution.

$$f(x, t|x_0, t_0) = \frac{\exp\left\{-\frac{(x-x_0-\mu(t-t_0))^2}{2\sigma^2(t-t_0)}\right\}}{\sqrt{2\pi\sigma^2(t-t_0)}}. \quad \clubsuit$$

What we don't know is the function $g_b^{x_0}(t)$. Our problem now is solving a particular kind of equation called **integral equations**. They are like differential equation, but with integrals. Pretty rad, if you ask me.



Figure 1.28: Manco fa ridere 'sta troiata, ma l'ho fatta mentre aspettavo che la prof. caricasse gli appunti su Moodle. Cosa che non ha ancora fatto, tra l'altro.

We want to solve an equation of this form:

$$f(t) = \int_0^t d\tau K(t, \tau)y(\tau).$$

We call these “Volterra integral equation of the first type”. In our case

$$f(t) = f(x, t|x_0, t_0) \quad \text{known } \checkmark$$

$$K(t, \tau) = f(b, t|b, \tau) \quad \text{known } \checkmark$$

$$y(\tau) = g_b^{x_0}(\tau) \quad \text{unknown } \clubsuit.$$



This is an equation of the first type because the unknown is only in the integral. There exist also the second type Volterra equations:

$$f(t) = \varphi(y(t)) + \int_0^t d\tau K(t, \tau)y(\tau)$$

but we leave these to more unhappy people than we are. For time-homogeneous processes, we have

$$f(b, t|b, \tau) = f(b, t - \tau|b, 0).$$

This is a simple consequence of stationarity of increments. Consider

$$f(b, t|x_0, t_0) = \int_0^t d\tau g_b^{x_0}(\tau)f(b, t - \tau|b, 0).$$



We can see, though, that the integral is a convolution.

Revise with Kotatsu!

You don't know remember what a convolution is, you poor wretched bastard who studied economics as undergrad? The convolution of two functions $f(t)$ and $g(t)$ is given by

$$(f * g)(t) = \int_0^t f(\tau)g(t - \tau) d\tau.$$

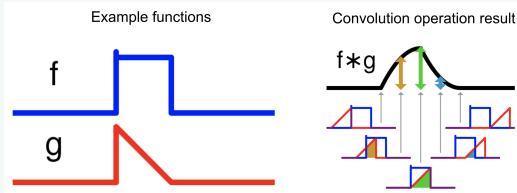


Figure 1.29: This image has always been helpful to me. I just love it.

Convolutions have nice properties. For example, we can easily turn them into simple multiplications using Laplace transforms. The Laplace transform of a function is defined as

$$\mathcal{L}\{f(t)\}(\lambda) = \mathcal{L}_\lambda\{f(t)\} = \int_0^\infty e^{-\lambda t} f(t) dt \quad \text{with } \lambda = \sigma + i\omega.$$

Here σ introduces a decaying/growing component while $i\omega$ introduces the oscillatory component, just like Fourier transforms! Indeed, Fourier transforms lack the “ σ ” part, only relying on the “spinning” part of the signal. While Fourier transform lets us go from the time domain to the frequency domain, Laplace transform lets us go to the *complex frequency domain*. This is why this bad boy Laplace transform is also well suited to represent signals/functions that are not periodic. This is bananas, very cool!

Anyway, back to our matter: Laplace transforms can make our life easier with convolutions. This is because

$$\mathcal{L}_\lambda\{(f * g)(t)\} = \mathcal{L}_\lambda\{f(t)\} \cdot \mathcal{L}_\lambda\{g(t)\}.$$

This is much more manageable! Just remember to do the inverse transform after to go back to your original domain. This can be tricky, because the inverse Laplace transform depends on the form of the transformed function. We often have to use tables of known Laplace transforms or other extravagant techniques that I will not delve into.

As we said, it's pretty standard to use Laplace transforms to deal with convolutions. We could also use Fourier transforms, but we are cooler than that. So transform  into our new equation:

$$f(b, t|x_0, t_0) = \int_0^t d\tau g_b^{x_0}(\tau) f(b, t - \tau|b, 0) = \int_0^t d\tau g_b^{x_0}(\tau) f(b, t|b, \tau)$$

$$\downarrow$$

$$\mathcal{L}_\lambda \{f(b, t|x_0, t_0)\} = \mathcal{L}_\lambda \{g_b^{x_0}(\tau)\} \cdot \mathcal{L}_\lambda \{f(b, t|b, \tau)\}.$$

We know that

$$\begin{aligned}\mathcal{L}_\lambda \{f(b, t|x_0, t_0)\} &= f_\lambda(b|x_0) \\ \mathcal{L}_\lambda \{f(b, t|b, \tau)\} &= f_\lambda(b|b) \\ \mathcal{L}_\lambda \{g_b^{x_0}(\tau)\} &= g_\lambda(b|x_0).\end{aligned}$$

Note that these equations do not show t anymore because they don't live in the time domain anymore, but in the complex frequency domain. t "becomes" λ (intuitively). Same goes for $g_b^{x_0}(\tau)$, where τ gets "integrated out" leaving behind the parameters b and x_0 . So we get

$$f_\lambda(b|x_0) = g_\lambda(b|x_0) \cdot f_\lambda(b|b) \implies g_\lambda(b|x_0) = \frac{f_\lambda(b|x_0)}{f_\lambda(b|b)}.$$

Let's take the only example we can apply in this stage of our life: the Brownian motion with drift (and $\sigma^2 = 1$). Apply the Laplace transform to  (setting $t_0 = 0$) and get:

$$\begin{aligned}f_\lambda(b|x_0) &= \int_0^\infty e^{-\lambda t} \frac{1}{\sqrt{2\pi t}} \exp \left\{ -\frac{(b - \mu t - x_0)^2}{2t} \right\} dt \\ &= \frac{e^{\mu(b-x_0)} e^{-\sqrt{\mu^2+2\lambda}(b-x_0)}}{\sqrt{2\lambda + \mu^2}}\end{aligned}$$

and

$$\begin{aligned}f_\lambda(b|b) &= \int_0^\infty e^{-\lambda t} \frac{1}{\sqrt{2\pi t}} \exp \left\{ -\frac{\mu^2 t}{2} \right\} dt \\ &= \frac{1}{\sqrt{2\lambda + \mu^2}}.\end{aligned}$$

It becomes now quite easy to compute $g_\lambda(b|x_0)$:

$$\begin{aligned}g_\lambda(b|x_0) &= \frac{e^{\mu(b-x_0)} e^{-\sqrt{\mu^2+2\lambda}(b-x_0)}}{\sqrt{2\lambda + \mu^2}} \cdot \cancel{\sqrt{2\lambda + \mu^2}} \\ &= e^{\mu(b-x_0)} e^{-\sqrt{\mu^2+2\lambda}(b-x_0)}.\end{aligned}$$

We are lucky: for this function we can analytically compute the inverse Laplace transform.

$$g(b, t|x_0) = \frac{|b - x_0|}{\sqrt{2\pi t^3}} \exp \left\{ -\frac{(b - x_0 - \mu t)^2}{2t} \right\}.$$

This is our function $g_b^{x_0}(\tau)$ expressed in the time domain! This is a good first result, but there is more to it. Consider the Laplace transform evaluated at $\lambda = 0$: this is like summing over all time without the decay, since when $\lambda = 0$ we are essentially removing the exponential decay factor

from the integral. In this case the fact that $e^{-\lambda t} = 1$ gives us the *total probability* of the event occurring with finite time.

$$\begin{aligned} g_\lambda(b|x_0)|_{\lambda=0} &= \int_0^\infty e^{-\lambda t} g(b, t|x_0, t_0) dt|_{\lambda=0} \\ &= \int_0^\infty g(b, t|x_0, t_0) dt \\ &= \mathbb{P}(\tau_b^{x_0} < \infty). \end{aligned}$$

Since

$$g_\lambda(b|x_0) = e^{\mu(b-x_0)} e^{-\sqrt{\mu^2+2\lambda}(b-x_0)}$$

we have that

$$g_\lambda(b|x_0)|_{\lambda=0} \begin{cases} 1 & \mu \geq 0 \\ e^{-2|\mu|(b-x_0)} & \mu < 0. \end{cases}$$

This is our second result and it tells us that, as a function of μ , the distribution of the first hitting time in a finite time has a decaying probability of reaching b (as b gets away from the starting point x_0) and then it degenerates to 1. But hold tight, there is MORE! Consider

$$\begin{aligned} \frac{dg_\lambda}{d\lambda}|_{\lambda=0} &= \frac{d}{d\lambda} \int_0^\infty dt e^{-\lambda t} g_b^{x_0}(t)|_{\lambda=0} \\ &= \int_0^\infty (-t) e^{-\lambda t} g_b^{x_0}(t) dt|_{\lambda=0} \\ &= -\mathbb{E}[\tau_b^{x_0}] \end{aligned}$$

and

$$\frac{d^n g_\lambda}{d\lambda^n}|_{\lambda=0} = (-1)^n \mathbb{E}\left[\left(\tau_b^{x_0}\right)^n\right]$$

So we can compute moments from the Laplace transform! For the Brownian motion with drift $\mu > 0$ (for $\mu < 0$ it doesn't make sense to compute the moments) we have

$$\begin{aligned} -\frac{d}{d\lambda} e^{\left(\mu-\sqrt{\mu^2+2\lambda}\right)(b-x_0)}|_{\lambda=0} &= \frac{b-x_0}{\sqrt{\mu^2+2\lambda}} e^{\left(\mu-\sqrt{\mu^2+2\lambda}\right)(b-x_0)}|_{\lambda=0} \\ &= \frac{b-x_0}{\mu} = \mathbb{E}[\tau_b^{x_0}] \end{aligned}$$

So the expectation is not only finite, but it also tends to be small since it depends inversely on the drift and directly on the distance between b and x_0 . The crossing time exists, but remember how in normal Brownian motion this time was infinite (see 1.1)? This is the proof that the drift helps attaining b !

Consider the case $b < x_0$:

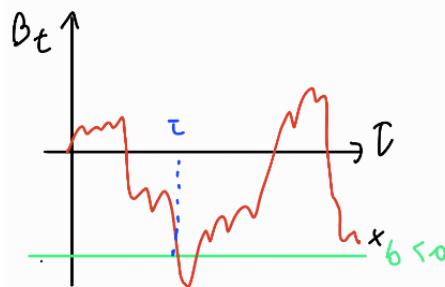


Figure 1.30: This hellish part is almost over... but you gotta admit it was pretty cool, in a sense. Maybe I am just losing my mind.

We know that

$$\begin{aligned}\mathbb{P}(X(t) < x | X(t_0) = x_0) &= \int_0^t g_b^{x_0}(\tau) \mathbb{P}(X(t) < x | X(t) = b) d\tau \\ &\Downarrow \\ F(x, t | x_0) &= \int_0^t g_b^{x_0}(\tau) F(x, t | b, \tau) d\tau\end{aligned}$$

Which is exactly the Fortet equation. So everything we said before applies to this case as well.

And what's the deal with time-dependent boundaries?

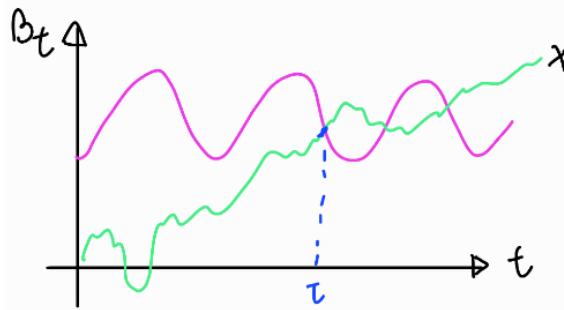


Figure 1.31: And what's the deal with airplane food?

We should understand that the probability of τ is not affected.

$$\begin{aligned}\mathbb{P}(X(t) > x | X(t_0) = x_0) &= \int_0^t d\tau g(b(\tau), t | x_0) \cdot \mathbb{P}(X(t) > x | b(\tau), \tau) \\ &\Downarrow \\ f(b(t), t | x_0) &= \int_0^t d\tau g(b(\tau), \tau | x_0) f(b(t), t | b(\tau), \tau).\end{aligned}$$

This is still the Fortet equation. I would have never guessed that that guy was not only capable of doing terrible Skrillex and Fred Again featurings but also well versed in stochastic processes.

Up to now, we worked having the process (meaning its probability density function) and the shape of the boundary. But what if we are in the inverse situation? What if we have the process and the first passage time distribution and *we want to determine the boundary shape*? Heh.

1.8 Useful laws about Brownian motion's path

1.8.1 Reflection principle

The **reflection principle** is the property according to which, if a path of a Brownian motion attains a level b at time s , the subsequent path has the same probability as its reflected path. There

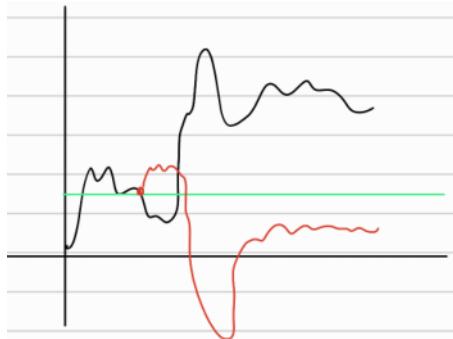


Figure 1.32: When I exported my notes I forgot to remove the page lines...

is an alternative definition.

Definition 1.8.1

The reflected sample path of a Brownian motion after a stopping time $\tau(\omega)$ is defined as

$$W(t, \omega) := \begin{cases} B(t, \omega) & 0 < t < \tau(\omega) < \infty \\ 2B(\tau(\omega), \omega) - B(t, \omega) & \tau(\omega) < t < \infty. \end{cases}$$

We can see how $2B(\tau(\omega), \omega) - B(t, \omega)$ is simply the mirrored sample path obtained by $B_\tau - (B_t - B_\tau)$: after τ we have the same increment but with opposite sign.

Theorem 1.8.1

Levy's triple law 1948: Let $B = (B_t)_{t \geq 0}$ be a standard BM. Then, if

$$m_t = \inf_{s \leq t} B_s \quad M_t = \sup_{s \leq t} B_s,$$

the probability

$$\mathbb{P}(m_t > a, M_t < b, B_t \in dx)$$



is equal to

$$\frac{dx}{\sqrt{2\pi t}} \sum_{n=-\infty}^{+\infty} \left[\exp \left\{ -\frac{x + 2n(b-a)^2}{2t} \right\} - \exp \left\{ -\frac{x - 2a - 2n(b-a)^2}{2t} \right\} \right].$$

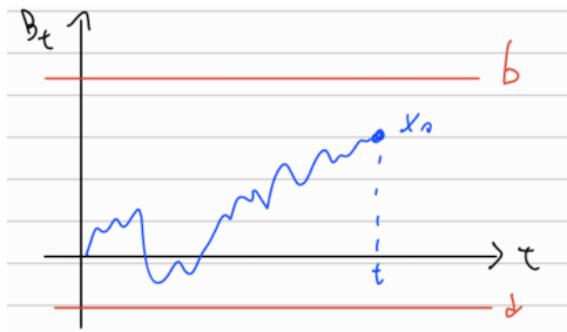


Figure 1.33: We are looking for the probability of the Brownian motion to be between a and b .

We get an infinite summation with two terms: the first one represent the “normal” path, while the other one represents the mirrored path reflected at a . We don’t do the whole the proof⁶ because that would take like two pages of useless calculations and we don’t want that. We just give a sketch of the proof:

Proof

Let

$$\tau = \inf \{t \geq 0 : B_t \notin (a, b)\}$$

⁶Yippie!!

be the first exit time from a, b . We have that

$$\begin{aligned}\mathbb{P}(m_t > a, M_t, B_t \in K) &= \mathbb{P}(\tau > t, B_t \in K) \\ &= \mathbb{P}(B_t \in K) - \mathbb{P}(\tau \leq t, B_t \in K) \\ &= \mathbb{P}(B_t \in K) - \mathbb{P}(B_\tau = a, \tau \leq t, B_t \in K) - \\ &\quad - \mathbb{P}(B_\tau = b, \tau \leq t, B_t \in K).\end{aligned}$$

Introduce the reflection operator:

$$\begin{aligned}\mathfrak{R}_a x &= 2a - x \\ \mathfrak{R}_b x &= 2b - x.\end{aligned}$$

Note that, using the reflection principle,

$$\begin{aligned}\mathbb{P}(B_\tau = a, \tau \leq t, B_t \in K) &= \mathbb{P}(B_\tau = a, \tau \leq t, B_t \in \mathfrak{R}_a K) \\ &= \mathbb{P}(B_\tau = a, B_t \in \mathfrak{R}_a K) \\ &= \mathbb{P}(B_t \in \mathfrak{R}_a K) - \mathbb{P}(B_\tau = b, B_t \in \mathfrak{R}_a K) \\ &= \mathbb{P}(B_t \in \mathfrak{R}_a K) - \mathbb{P}(B_\tau = b, B_t \in \mathfrak{R}_b \mathfrak{R}_a K) \\ &= \mathbb{P}(B_t \in \mathfrak{R}_a K) - \mathbb{P}(B_t \in \mathfrak{R}_b \mathfrak{R}_a K) + \\ &\quad + \mathbb{P}(B_\tau = a, B_t \in \mathfrak{R}_n \mathfrak{R}_a K) \\ &\quad \dots\end{aligned}$$

The idea is that we keep on rewriting the probabilities as their marginals minus (or plus) something until we arrive to our summation. \square

We know that

$$M_t = \sup_{\vartheta \leq t} B_\vartheta = \sup_{\vartheta \leq t} (-B_\vartheta) = \sup_{\vartheta \leq t} \left[\underbrace{-B_\vartheta + B_t}_{B_t - B_\vartheta \sim \tilde{B}_\vartheta} - B_t \right] \sim \sup_{\vartheta \leq t} [\tilde{B}_\vartheta - B_t] \sim M_t - B_t.$$

So... $M_t \sim M_t - B_t$? Are we going crazy yet?

Theorem 1.8.2

Consider a BM^d with $b \in \mathbb{R}$. Let

$$m_t = \inf_{s \leq t} B_s \quad M_t = \sup_{s \leq t} B_s$$

and

$$\tau_b := \inf \{t \geq 0 : B_t = b\}.$$

Then

$$\begin{aligned}M_t &\sim |B_t| \sim -m_t \sim M_t - B_t \sim B_t - m_t \\ &\stackrel{(1)}{\sim} \stackrel{(2)}{\sim} \stackrel{(3)}{\sim} \stackrel{(4)}{\sim} \\ &\sim \frac{\sqrt{2}}{\sqrt{\pi t}} e^{-\frac{x^2}{2t}} dx \\ &\stackrel{(5)}{\sim}\end{aligned}$$

and

$$\tau_b \sim \frac{|b|}{\sqrt{2\pi t^3}} \exp \left\{ -\frac{b^2}{2t} \right\}.$$

Proof

- (1) Already proven.
- (2) This is left as an exercise: we need to prove that

$$m_t = \inf_{s < t} B_s = -\sup_{s \leq t} (-B_s) = -t.$$

- (3) Rethink this as

$$M_t - B_t = \sup_{s \leq t} (B_s - B_t) = \sup_{s \leq t} (B_{t-s} - B_t) \sim \sup_{s \leq t} B_s = M_t.$$

Remark

we know that $M_t \sim |B_t|$ (we proved it), so

$$M_t - B_t \sim |B_t| - B_t \sim |B_t| \sum M_t.$$

- (4) This follows from combining (3) with symmetry. □

1.8.2 Arcsin laws

Consider a Brownian motion in the interval $(0, t)$. We want to study the last 0 before t , that is

$$\xi_t := \sup \{s \leq t : B_s = 0\}.$$

Of course this is not a stopping time because to assess whether a 0 is the last 0 before t we need

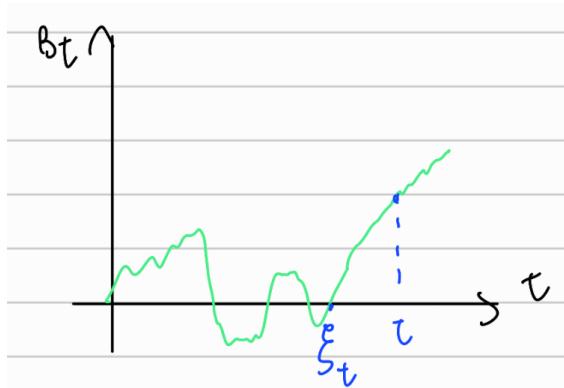


Figure 1.34: I am not sure the sup gets defined in this way...

to peek into the future.

Theorem 1.8.3

Arcsin law I. Let $(B_t)_{t \geq 0}$ be a one-dimensional Brownian motion and let ξ_t be the largest zero of B_s in $[0, t]$. Then for $\forall 0 < s < t$ we have

$$\mathbb{P}(\xi_t < s) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}}.$$

Proof

Take

$$h(s) = \mathbb{P}(B_u \neq 0, \forall u \in [s, t])$$

conditioning the \mathbb{P} to the value of B at $s \rightarrow \infty$ = $\int_{\mathbb{R}} \mathbb{P}^{B_s(\omega)}(B_{u-s}, \forall u \in [s, t]) \cdot \mathbb{P}(B_{s \in db})$

So my process at time s will be in $B_s(\omega) = b$. I can write the probability conditioned to being in b as \mathbb{P}^b . What the fuck is this horrid notation now?

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{P}^{B_s(\omega)}(B_{u-s}, \forall u \in [s, t]) \cdot \mathbb{P}(B_{s \in db}) &= \int_{\mathbb{R}} \mathbb{P}^b(B_{u-s} \neq 0, \forall u \in [s, t]) f_{B_s}(b) db \\ &= \int_{\mathbb{R}} \mathbb{P}^0(B_{u-s} \neq -b, \forall u \in [s, t]) f_{B_s}(b) db \quad (\text{by definition of } f_{B_s}) \\ &= \int_{\mathbb{R}} \mathbb{P}^0(B_v \neq -b, \forall v \in [0, t-s]) f_{B_s}(b) db \\ &= \int_{\mathbb{R}} \mathbb{P}^0(\tau_{-b} > t-s) \underbrace{f_{B_s}(b)}_{\frac{1}{\sqrt{2\pi s}} e^{-\frac{b^2}{2s}}} db. \end{aligned}$$

Note that

$$\mathbb{P}^0(\tau_b \geq t-s) = 2 \sqrt{\frac{1}{2\pi(t-s)}} \int_0^b e^{-\frac{x^2}{2(t-s)}} dx$$

so

$$(\text{by definition of } f_{B_s}) = 2 \int_0^\infty \left(\sqrt{\frac{1}{2\pi(t-s)}} \int_0^b e^{-\frac{x^2}{2(t-s)}} dx \right) \cdot \frac{1}{\sqrt{2\pi s}} \cdot e^{-\frac{b^2}{2s}} db.$$

So, calling $\beta = \frac{b}{\sqrt{s}}$ and $\mu = \frac{x}{\sqrt{t-s}}$, we get

$$h(s) = \frac{2}{\pi} \int_0^\infty \left(\int_0^{\beta \sqrt{\frac{s}{t-s}}} e^{-\frac{u^2}{2}} du \right) e^{-\frac{\beta^2}{2}} d\beta.$$

Why do we have to endure suffering? Consider

$$h'(s) = \frac{1}{\pi} \frac{1}{\sqrt{s(t-s)}} \quad 0 \leq s \leq t.$$

Revise with Kotatsu!

Remember that

$$\frac{d}{ds} \arcsin \sqrt{\frac{s}{t}} = \frac{\pi}{2} h'(s).$$

So

$$h(s) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}} = 1 - \frac{2}{\pi} \arccos \sqrt{\frac{s}{t}} \quad s \leq t.$$

□

⚠ I am losing my grip on reality ⚠

Trigonometric equations?? In MY fucking measure theory notes?

Remark

The density is

$$\mathbb{P}(\xi \in ds) = \frac{1}{\pi} \frac{1}{\sqrt{s(t-s)}}.$$

If we choose $t = 1$ we get

$$\mathbb{P}(\xi \in ds) = \frac{1}{\pi} s^{-1/2} (1-s)^{1/2}.$$

This is a Beta distributions with parameters $(1/2, 1/2)$.

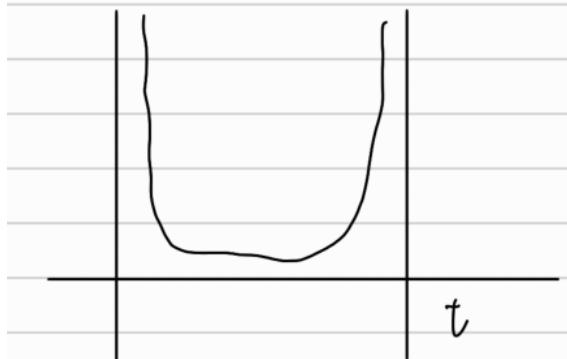


Figure 1.35: You will take this distribution and you WILL like it.

The shape is symmetric: there is a high probability of values at the extreme of the interval, so it is highly probable that all the values ar either at 0 or at t .

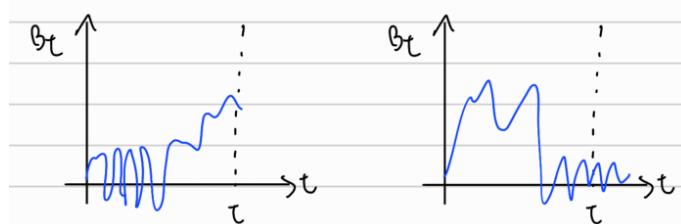


Figure 1.36: I am not entirely sure bout there graphs, but I think they are correct.

Remark

We can read the law in an alternative way. Recall that

$$X_t = M_t - B_t$$

has the same law as B_t . Then observe that:

- zeroes of B_t coincide with zeroes of $|B_t|$;
- the last zero of X_t has the same distribution as the last zero of B_t ;

- the last zero of X_t happens when B_t attains its supremum.

This means that (and this is **arcsin law III**)

$$\mathbb{P}(\underbrace{M}_{\text{time of last maximum}} < X) = \frac{2}{\pi} \arcsin \sqrt{x}.$$

What about the **arcsin law II**? Consider a Brownian motion (B_t) and

$$G_t := \int_0^t \mathbf{1}_{(0,\infty)}(B_s) ds.$$

G_t is the time that the process spends above the level 0. We have that

$$\mathbb{P}(G_t \leq x) = \frac{2}{\pi} \arcsin \sqrt{\frac{x}{t}} \quad x \in [0, t].$$

Again, small and large times are more probable. There is another arcsin law (I call it **arcsin law IV**). We want the probability α that B_t has at least one 0 in $[t_0, t]$.

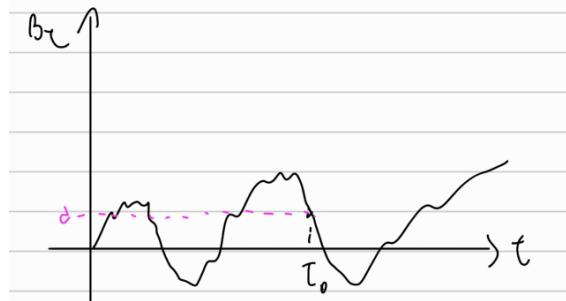


Figure 1.37: What is a ? We will never know!

$$\alpha = \frac{2}{\pi} \arccos \sqrt{\frac{t_0}{t}}.$$

So we are learning about how the process behaves either at the origin or far away from it.

Proof

$$\begin{aligned} \beta &= \mathbb{P}(B_t \text{ has at least one zero in } [t_0, t_1] | B_{t_0} = a) \\ &= \mathbb{P}(\tau_a \leq t_1 - t_0) \\ &= \frac{|a|}{\sqrt{2\pi}} \int_0^{t_1-t_0} \frac{e^{-\frac{a^2}{2u}}}{\sqrt{u^3}} du \end{aligned}$$

and

$$\begin{aligned}
\alpha &= \int_0^\infty da \beta \mathbb{P}(|B_{t_0}| \in (a, a+da) | B_0 = a) \\
&= \int_0^\infty \left(\frac{s}{\sqrt{2\pi}} \int_0^{t_1-t_0} \frac{e^{-\frac{u^2}{2s}}}{\sqrt{u^3}} du \right) \sqrt{\frac{2}{\pi t_0}} e^{-\frac{a^2}{2t}} da \\
&= \dots = \frac{2}{\pi} \arccos \sqrt{\frac{t_0}{t}}.
\end{aligned}$$

□

Truly horrible. Here is a cheatsheet:

Arcsin law I	Arcsin law II
$\mathbb{P}(\xi_t < s) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}}$	$\mathbb{P}(G_t \leq x) = \frac{2}{\pi} \arcsin \sqrt{\frac{x}{t}}$
Arcsin law III	Arcsin law IV
$\mathbb{P}(M < X) = \frac{2}{\pi} \arcsin \sqrt{x}$	$\alpha = \frac{2}{\pi} \arccos \sqrt{\frac{t_0}{t}}$

Exercises

Exercise 1

Let $B(t)$ be a Brownian motion and let $0 < s \leq t \leq u \leq v$.

1. Show that the random variables $\frac{1}{t}B(t) - \frac{1}{s}B(s)$ and $aB(u) + bB(v)$ are independent for any $a, b \in \mathbb{R}$.
2. Show that the random variables $aB(s) + bB(t)$ and $\frac{1}{v}B(v) - \frac{1}{u}B(u)$ are independent for any $a, b \in \mathbb{R}$ satisfying the condition $as + bt = 0$.

Exercise 2

Let $X(t)$ be a one-dimensional Brownian motion with drift μ and let $T = \inf\{t \geq 0 : X(t) > b\}$ be its first passage time from a boundary $b > 0$.

1. Using the Fortet equation determine when

$$\mathbb{P}(T < \infty) = 1.$$

2. Determine the distribution of T . Hint: make use of Laplace transforms, their expression is reported in the additional material uploaded on Moodle.
3. Determine $\mathbb{E}[T]$.

Exercise 3

Let $B(t)$ be a one dimensional Brownian motion and let $T_i = \inf\{t \geq 0 : B(t) > i\}$, $i = -1, 1, 2$. Determine

$$\mathbb{P}(T_1 < T_{-1} < T_2).$$

Exercise 4

Consider a one dimensional Brownian motion. Let

- $\tau_b := \inf\{t > 0 : B(t) \geq b\};$
- $\tau_{-b} := \inf\{t > 0 : B(t) \leq -b\};$
- $\tau_{cb} := \inf\{t > 0 : B(t) \geq cb\}.$

Prove that

1. $\tau_b \sim \tau_{-b};$
2. $\tau_{cb} \sim c^2 \tau_b.$

1.8.3 Brownian bridge

Definition 1.8.2

Let $(B_t)_{t \geq 0}$ be a one-dimensional Brownian motion. Then B conditioned on $B_0 = B_1 = 0$ (or $B_0 = B_t = 0$ with t fixed).

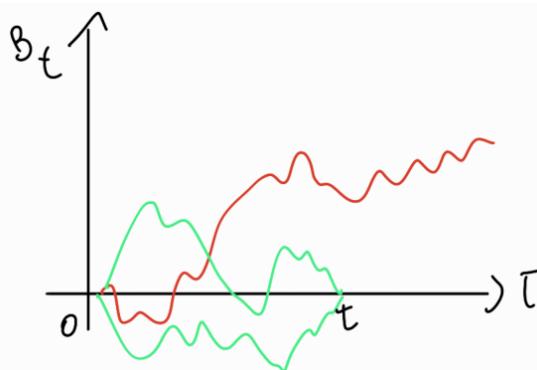


Figure 1.38: This time I remembered to turn off the lines.

Theorem 1.8.4

The **Brownian bridge** \tilde{B}_t is a Gaussian process with covariance function

$$\begin{aligned}\mathbb{E} [\tilde{B}_s \tilde{B}_\omega] &= \mathbb{E} [B_s B_\omega | B_0 = B_t = 0] \\ &= \frac{1}{t} [(s \wedge \omega)((t-s) \wedge (t-\omega))] \quad 0 < s, \omega \leq t\end{aligned}$$

and mean

$$\mathbb{E} [\tilde{B}_s] = 0.$$

So the Gaussian bridge is basically a Brownian motion to which we impose the condition that it must start at 0 at time 0 and end at 0 at time t . Is a “selection” of the infinitely possible sample paths (among which there are some that start at 0 and end at 0 at t) and as we know, “selecting” means conditioning.

Proof

Consider $(B_s B_\omega | B_t)$ for $\omega < s < t$. The vector (B_t, B_s, B_ω) has covariance matrix

$$\Sigma = \begin{bmatrix} t & s & \omega \\ s & s & \omega \\ \omega & \omega & \omega \end{bmatrix}. \quad \text{⑩}$$

Note that ⑩ is divided in a 1×1 matrix Σ_{11} , a 1×2 matrix Σ_{12} , a 2×1 matrix Σ_{21} and 2×2 matrix Σ_{22} . The covariance matrix $\Sigma_{2|1}$ (that is the conditional covariance of one subset of random variables “2”, in this case the Gaussian sub-vector (B_s, B_ω) to another subset of random variables “1”, in this case the sub-vector (B_t)) of (B_s, B_ω) conditioned on B_t is therefore

$$\begin{aligned} \Sigma_{2|1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &= \begin{bmatrix} s & \omega \\ \omega & \omega \end{bmatrix} - \frac{1}{t} \begin{bmatrix} s \\ \omega \end{bmatrix} \begin{bmatrix} s & \omega \end{bmatrix} \\ &= \begin{bmatrix} s & \omega \\ \omega & \omega \end{bmatrix} - \frac{1}{t} \begin{bmatrix} s^2 & s\omega \\ s\omega & \omega^2 \end{bmatrix} \\ &= \begin{bmatrix} s - \frac{1}{t}s^2 & \omega - \frac{1}{t}s\omega \\ \omega - \frac{1}{t}s\omega & \omega - \frac{1}{t}\omega^2 \end{bmatrix}. \end{aligned}$$

This means that

$$\begin{aligned} \mathbb{E}[B_s B_\omega | B_0 = B_t = 0] &= \begin{cases} \omega - \frac{1}{t}s\omega & \omega < s \\ s - \frac{1}{t}s\omega & \omega > s \end{cases} \\ &= \frac{1}{t}(s \wedge \omega) [(t - s) \wedge (t - \omega)]. \end{aligned}$$

So the variance of \tilde{B} is

$$\text{Var } \tilde{B}_s = \frac{1}{t}s(t-s).$$

□

Question: are the increments of \tilde{B} independent? Actually not... Because we have the conditions that the process must be at 0 at time t . So the increments, having to reach a certain point, are dependent on their position respect to t ! Also, it doesn't have to be 0: I can also set a value A at time ϑ and a value B at time t and only consider the sample paths that go through these points.

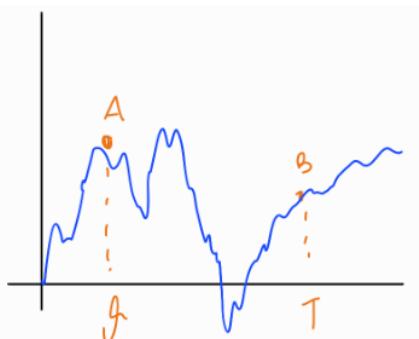


Figure 1.39: This is... cool I guess?

Remark

The Brownian bridge can be represented as

$$X_s = B_s - \frac{s}{t} B_t \quad 0 \leq s \leq t.$$

If $t = 1$ then

$$X_s = B_s - sB_1.$$

At time 1 this is equal to 0 and it has the same distribution as the equation above. Another representation is

$$Y_s = (1-s) \frac{B_s}{1-t} \quad 0 \leq s \leq t.$$

These representations are equal to the Brownian bridge in distribution. The shape of the sample path, of course, can be different.

Why do we care? There are many reasons:

1. Simulations: imagine that we want to simulate a Brownian motion. We know that we can obtain a simulated Brownian motion by partitioning our times and then extracting a Gaussian random variable to get the increment between times. Imagine we want to

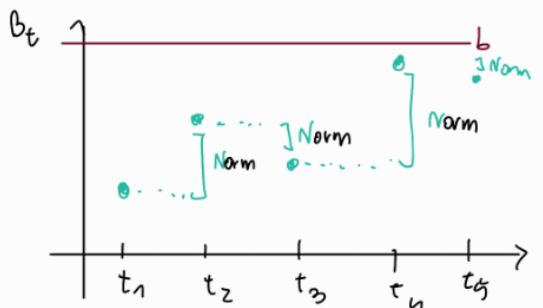


Figure 1.40: Many little normal increments form a big Brownian noise.

check whether the process crosses a threshold b . What if it does, but it happens in between two of the discrete values we computed? For example, in figure 1.40, what if it crosses b between the last two points? I could either recalculate everything with a finer mesh (huge waste of time) or... just study the Brownian bridge between those two points!

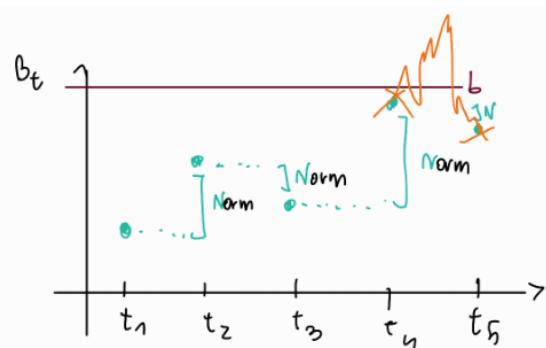


Figure 1.41: I am going to do this motherfucker on Max/Msp.

2. The list could go on.

Remark

The Brownian bridge transition probability density function verifies the heat equation. It is the same as Brownian motion but with different boundary conditions. I don't really care.

1.9 Properties of Brownian Paths



Figure 1.42: Many paths but nowhere to go!

1.9.1 Variation

Subdivide an interval $[a, b]$ through a finite collection of disjoint intervals of the form (s, t) whose union is $[a, b]$. Then we take

$$\mathcal{A} = \text{subdivision with mesh } \|\mathcal{A}\| = \sup \{t - s : (s, t] \in \mathcal{A}\}.$$

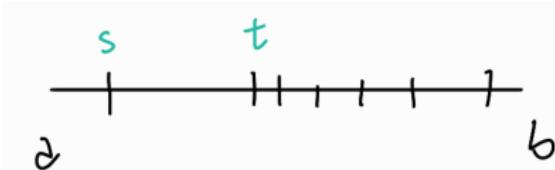


Figure 1.43: In practice, we take the biggest subdivision. This helps when extending results to arbitrary subdivisions.

Definition 1.9.1

Consider a right-continuous function $f : \mathbb{R} \rightarrow \mathbb{R}_+$. Fix an interval $[a, b] \in \mathbb{R}_+$. In general, for $p > 0$, consider the sum

$$\sum_{(s,t] \in \mathcal{A}} |f(t) - f(s)|^p.$$

The supremum of this sum over $[a, b]$

$$\sup_{[a,b]} \sum_{(s,t] \in \mathcal{A}} |f(t) - f(s)|^p$$

is called **p-variation**.

Setting $p = 2$ gives us the **quadratic variation**. Setting $p = 1$ gives us the **total variation**.

Theorem 1.9.1

Let $(a, b]$ be fixed and let \mathcal{A}_n be a sequence of subdivisions of it with the mesh going to 0:

$$\|\mathcal{A}_n\| \xrightarrow{n \rightarrow \infty} 0.$$

The sequence of random variables

$$V_b = \sum_{(s,t] \in \mathcal{A}_n} |B_t - B_s|^2$$

converges in L^2 and in probability to the length $b - a$.

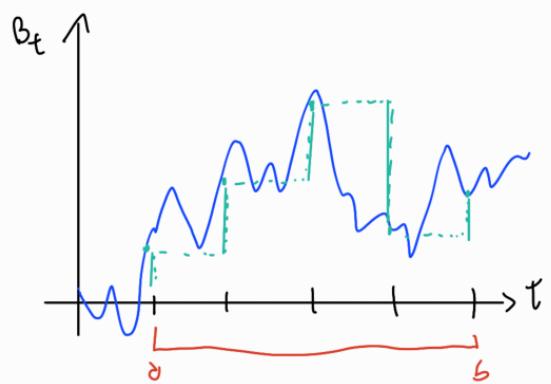


Figure 1.44: Imagine that the grid becomes smaller and smaller.

Proof

We know that

$$|B_t - B_s|^2 \sim (t - s)Z^2 \quad Z \sim N(0, 1).$$

The intervals of the mesh are disjoint and their corresponding increments are therefore independent. Consider

$$\begin{aligned} \mathbb{E}[V_n] &= \sum_{(s,t] \in \mathcal{A}_n} (t - s) \\ &= b - a \end{aligned}$$

and

$$\text{Var } V_n = \sum_{(s,t] \in \mathcal{A}_n} (t - s)^2 \leq (b - a) \|\mathcal{A}_n\|.$$

This means that

$$\mathbb{E}[(V_n - (b - a))^2] = \text{Var } V_n \xrightarrow{\text{since } \|\mathcal{A}_n\| \rightarrow 0} 0.$$

So the sequence converges in L^2 and this implies convergence in probability. The limit

$$\lim_{n \rightarrow \infty} V_n$$

is the quadratic variation of B over $(a, b]$. □

Theorem 1.9.2

For each $n \in \mathbb{N}$ let \mathcal{A}_n be a subdivision of $(a, b]$ that consists of 2^n intervals of the same length. Then V_n converges to $(b - a)$ almost surely.

Proof

For each $(s, t]$ we have length $(b - a)2^{-n}$ and this means that

$$\begin{aligned}\mathbb{E}[V_n] &= b - a \\ \text{Var } V_n &= 2^n(b - a)^2 2^{-n}.\end{aligned}$$

If we use Chebyshev's inequality we get

$$\mathbb{P}(|V_n - (b - a)| > \varepsilon) \leq \frac{1}{\varepsilon^2} (b - a)^2 2^{-n}$$

this is summable and finite

so using the Borel-Cantelli lemma 1 we get

$$V_n \rightarrow (b - a) \text{ a.s.}$$

□

What about the total variation, tough?

Proposition 1.9.1

For almost every ω the path $B(\omega)$ has infinite total variation over every interval $(a, b]$, $a < b$.

Revise with Kotatsu!

A function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is said to be of **bounded variation** on $[0, t]$ if

$$V(f : [0, t]) = \sup \left\{ \sum |f(t_i) - f(t_{i-1})| \right\}$$

is finite for $0 \leq t_0 < t_1 < \dots < t_n < t$, $n \geq 1$.

Here is an example of infinite variation function:

$$f(t) = \begin{cases} t \sin \frac{1}{t} & t > 0 \\ 0 & t = 0. \end{cases}$$

As we can see in figure 1.45 the function has an extremely oscillatory nature. This is what happens with out Brownian motion, with the difference that the variations are very small. So much that taking their square makes them vanish.

Proof

Let Ω_{ab} be the set of ω such that $V_n \rightarrow (b - a)$. Pick $\omega \in \Omega_{ab}$ and let v^\star be the total

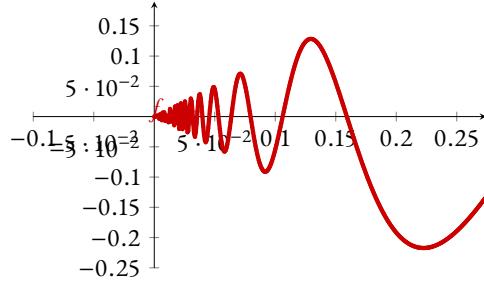


Figure 1.45: First time using the geogebra export function. I hate it.

variation of $B(\omega)$. By contradiction, let $v^* < \infty$. We have

$$\begin{aligned} \sum_{(s,t] \in \mathcal{A}_n} |B_t - B_s|^2 &\leq \sup_{(s,t] \in \mathcal{A}_n} |B_t - B_s| \sum_{(s,t] \in \mathcal{A}_n} |B_t - B_s| \\ &\leq \left(\sup_{(s,t] \in \mathcal{A}_n} |B_t - B_s| \right) v^* \end{aligned}$$

but

$$\sum_{(s,t] \in \mathcal{A}_n} |B_t - B_s|^2 \xrightarrow{\|\mathcal{A}_n\| \rightarrow 0} (b-a)$$

while

$$\sup_{(s,t] \in \mathcal{A}_n} |B_t - B_s| \rightarrow 0$$

so v^* cannot be finite. Let Ω_0 be the intersection of Ω_{ab} over the rationals a, b such that $0 \leq a < b$. The thesis holds for each ω in the a.s. event Ω_0 . \square

1.9.2 Differentiability

We saw how the sample path is highly oscillatory with extremely small oscillation. We will now discover that they are also **nowhere differentiable**. To show this we need the notion of Hölder continuity.

Definition 1.9.2

Hölder continuity. A function f is said to be Hölder continuous of order α on $S \subset \mathbb{R}_+$ if there exists a constant k such that

$$|f(t) - f(s)| \leq k|t - s|^\alpha \quad s, t \in S.$$

We can see how differentiable functions are always Hölder continuous due to the very definition of derivative. So if we show that the Brownian motion is not Hölder continuous, we will know it is now differentiable.

Proposition 1.9.2

For almost every ω , the Brownian motion path is Hölder continuous on **no interval** for order $\alpha > \frac{1}{2}$. In particular, for almost every ω , the paths are nowhere differentiable.

Proof

Pick an ω . Suppose

$$|B_t(\omega) - B_s(\omega)| \leq k|t - s|^\alpha \quad \forall s, t \in [a, b] \text{ and } \alpha > \frac{1}{2} \text{ and some } k.$$

Now we square everything and sum over the partitions of $[a, b]$:

$$\begin{aligned} \sum_{(s,t] \in \mathcal{A}_n} |B_t(\omega) - B_s(\omega)|^2 &\leq k^2 \sum_{(s,t] \in \mathcal{A}_n} |t - s|^{2\alpha} \\ &\leq k^2 \sum_{(s,t] \in \mathcal{A}_n} (t - s)|t - s|^{2\alpha-1} \\ &\leq k^2 \sup_{(s,t] \in \mathcal{A}_n} (t - s)^{2\alpha-1} \underbrace{\sum_{(s,t] \in \mathcal{A}_n} (t - s)}_{b-a, \text{ our original interval}}. \end{aligned}$$

As $n \rightarrow \infty$ we have

$$\sup_{(s,t] \in \mathcal{A}_n} |t - s|^{2\alpha-1} \rightarrow 0$$

if $\alpha > \frac{1}{2}$, so

$$\sum_{(s,t] \in \mathcal{A}_n} |B_t(\omega) - B_s(\omega)|^2 \rightarrow 0 \implies \omega \notin \Omega_{ab}$$

by convergence of quadratic variation. This holds of $\forall \omega$ in the intersection of Ω_{ab} , \forall rational $a, b | a < b$. \square

Remark

The result holds for $\alpha = \frac{1}{2}$ too:

$$\frac{B(t) - B(s)}{(t - s)} = \frac{B(t) - B(s)}{(t - s)^{1/2}(t - s)^{1/2}}$$

and taking the limit

$$\lim_{s \rightarrow t} \frac{B(t) - B(s)}{(t - s)^{1/2}} \frac{1}{(t - s)^{1/2}} = \infty$$

k (it is Hölder continuous)

1.9.3 Growth of Brownian motion path

We proved that

$$tB\left(\frac{1}{t}\right)$$

is again a Brownian motion. We can think of B_t as the sum of increments of the sample path and t as the sum of the times (that is, occurrences). This lets us apply the Law of Large Numbers

$$\lim_{t \rightarrow \infty} \frac{B_t}{t} \stackrel{\text{LLN}}{=} \mathbb{E}[B_t] = 0.$$

This in turn means that

$$B(t, \omega) \leq \varepsilon t \quad \forall \varepsilon > 0.$$

The central limit theorem holds as well:

$$\lim_{t \rightarrow \infty} \frac{B_t - \mathbb{E}[B_t]}{\sqrt{\text{Var } B_t}} = \lim_{t \rightarrow \infty} \frac{B_t}{\sqrt{t}} \sim Z \sim N(0, 1).$$

This is like saying

$$\limsup \frac{B_t}{\sqrt{t}} = +\infty \quad \liminf \frac{B_t}{\sqrt{t}} = -\infty \quad \frac{B_t}{\sqrt{t}} = -\infty$$

What we have discovered is that our sample path will always grow slower than the bisector!



Figure 1.46: Yes I know... The lines.

Theorem 1.9.3

Le'vy 1937. Let B_t be a one-dimensional Brownian motion. Then

$$\mathbb{P}\left(\limsup_{h \rightarrow 0} \frac{|B_{t+h} - B_t|}{\sqrt{2h \ln \frac{1}{h}}} = 1\right) = 1.$$

Note that $\sqrt{2h \ln \frac{1}{h}}$ is increasing in $(0, \frac{1}{e})$.

Theorem 1.9.4

Law of iterated logarithm (Khinchin 1933). Let B_t be a one-dimensional Brownian motion. Then

$$\mathbb{P}\left(\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \ln \ln t}} = 1\right) = 1.$$

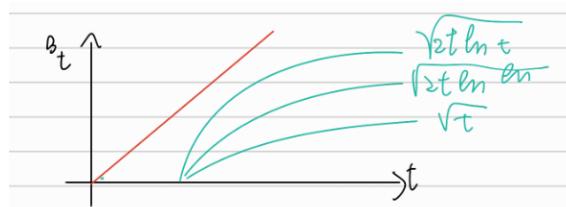


Figure 1.47: I am sorry. The fucking homeworks are due soon.

Corollary

Let B_t be a one-dimensional Brownian motion. Then, a.s.,

$$\begin{aligned}\overline{\lim}_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \ln \ln t}} &= 1 & \overline{\lim}_{t \rightarrow 0} \frac{B_t}{\sqrt{2t \ln \ln \frac{1}{t}}} &= 1 \\ \underline{\lim}_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \ln \ln t}} &= -1 & \underline{\lim}_{t \rightarrow 0} \frac{B_t}{\sqrt{2t \ln \ln \frac{1}{t}}} &= -1.\end{aligned}$$

Define

$$h(t) = \sqrt{2t \ln \ln t}.$$

For every $\varepsilon > 0$, $\delta > 0$, the Brownian motion B_t is infinitely often in the region

$$\{(t, y) : 0 \leq t \leq \delta \quad (1 - \varepsilon)h(t) \leq y \leq (1 + \varepsilon)h(t)\}.$$

And also, infinitely often,

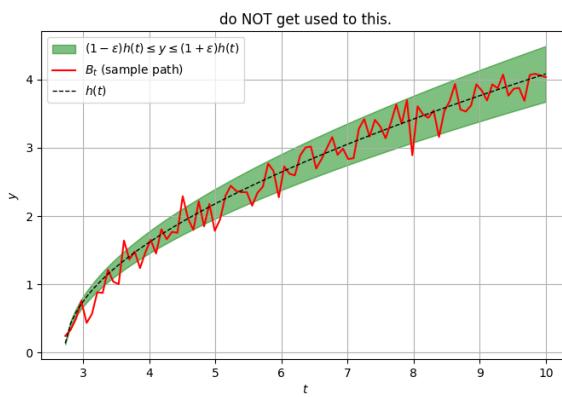


Figure 1.48: FIRST and LAST time I use Matplotlib.

$$\{(t, y) : 0 \leq t \leq \delta \quad -(1 + \varepsilon)n(t) < y < -(1 - \varepsilon)h(t)\}.$$

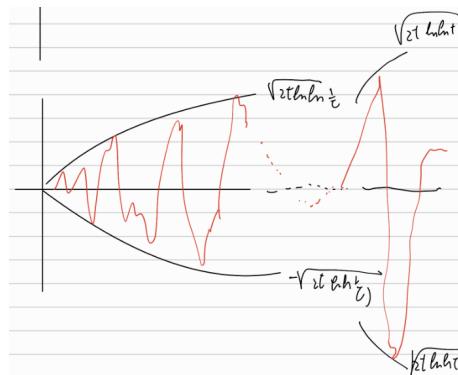


Figure 1.49: Idk bruv she drew this thing but I don't even know what this is supposed to be in the first place.

Chapter 2

Diffusion processes

2.1 One-dimensional diffusion processes

2.1.1 Definition

Definition 2.1.1

Let $\{X(t)\}$ be a continuous time stochastic process such that:

1. its sample paths are continuous functions of t , a.s.;
2. it has strong Markov property.

Then $\{X(t)\}$ is called **diffusion process**.

Here we incur in a problem: while before we just took for granted that Brownian motion was continuous, now we need to *actually define* what “continuous” means. But why should we care about diffusion processes?

1. Many real world phenomena can be modeled through diffusion. We saw how Brownian motion can model the motion of particles, but if we want to take into account friction as well we need to use a diffusion process to model that behavior.
2. For diffusions we are able to compute many useful functionals. So this is motivation to use them!
3. Suitable rescaling allows us to approximate discrete time or space processes through diffusion processes. This is useful because for many discrete processes, as soon as combinatorics enter the stage the computations become a nightmare (think about a random walk with random step size... and think about considering all the various combinations of steps to reach a certain state).

Definition 2.1.2

The interval $I = (l, r)$ (or $[l, r)$ or $(l, r]$ or $[l, r]$) where we allow $l = -\infty$ and $r = +\infty$, is called **diffusion interval**.

Definition 2.1.3

A stochastic process is said to be **regular** if, starting from any point x_0 in the interval I , any other point z interior to I may be attained with positive probability. This means, if

$$T_z = \inf\{t > 0 : X(t) > z | X(0) = x_0 < z\},$$

then the process is regular if

$$\mathbb{P}(T_z < \infty | X(0) = x_0) > 0 \quad \forall l < x_0 < z < r.$$

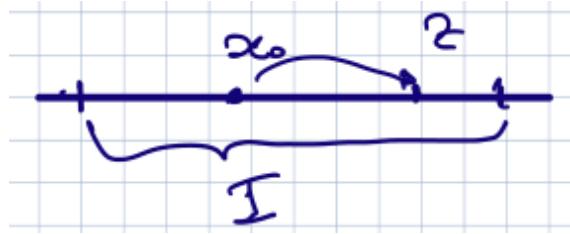


Figure 2.1: Drawing from the one and only Professor herself.

We will consider only regular processes. For Markov processes we have

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(|X(t+h) - X(t)| > \varepsilon | X(t) = x) = \lambda(x, \varepsilon) \geq 0.$$

Remark

For diffusion processes we have

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(|X(t+h) - X(t)| > \varepsilon | X(t) = x) = 0.$$

This means that large displacements over small intervals are very unlikely.

Definition 2.1.4

A strong Markov process $X(t)$ is called **standard prcoess** if its sample path verifies:

1. $X_\omega(t)$ is right continuous;

$$\lim_{t \rightarrow s} X(t) = X(s);$$

2. the left limit exists;

3. $X(t)$ is continuous from the left through Markov (stopping) times.

The third characteristic means that if $T_1 < T_2 < \dots$ are stopping times then

$$\lim_{n \rightarrow \infty} X(T_n) = X(T)$$

whenever $T < \infty$ a.s. So it is continuous on the left just by stopping times: this means, ultimately, that the only kind of discontinuity allowed in a standard process are (unpredictable) jumps.

For example, the Poisson process is standard.

Proposition 2.1.1

Every strong Markov process continuous in probability and subject to “mild conditions” admits a standard version.

2.1.2 The Dynkin condition

We may now ask whether there are conditions to recognize if a standard process is also a diffusion. Clearly Poisson processes are not diffusions so we need a tool to restrict the class. This is the **Dynkin condition**:

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{P}(|X(t+h) - X(t)| > \varepsilon | X(t) = x) = 0$$

where the convergence prevails uniformly for x restricted to any compact sub-interval of (l, r) and t on any finite interval on $[0, N]$.

Since continuity of a process is tied to the increments, we should analyze them.

Definition 2.1.5

Let $\Delta_h X(t)$ be the increment of the process $X(t)$ in an interval of size h .

$$\Delta_h X(t) = (X(t-h)) - X(t).$$

Then for $l < xL < r$ define

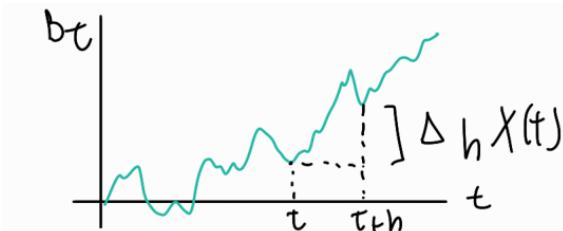


Figure 2.2: It is hard to think about the easy way out.

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [\Delta_h X(t) | X(t) = x] = \mu(x, t)$$

as the **drift** or **infinitesimal mean** and

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [(\Delta_h X(t))^2 | X(t) = x] = \sigma^2(x, t)$$

as the **infinitesimal variance** of $X(t)$ or **diffusion parameter**. Keep in mind that the infinitesimal mean and variance are the mean and variance of the infinitesimal increment, not of the whole process!

Remark

A diffusion process is characterized as having the following conditions:

1. it has strong Markov property;
2. $\lim_{h \searrow 0} \mathbb{P}(|\Delta_h X(t)| > \varepsilon | X(t) = x) = 0, x \in I$;
3. $\mu(x, t)$ and $\sigma^2(x, t)$ are continuous functions of x and t .

This is equivalent to the Dynkin condition, but it is much easier to check: we are just asking that the probability of large increments goes to zero with some conditions on the speed at which the process leaves the current state.

Definition 2.1.6

If

$$\begin{aligned}\mu(x, t) &= \mu(x) \\ \sigma^2(x, t) &= \sigma^2(x)\end{aligned}\forall x, t$$

the process is said to be **time homogeneous**.

Remark

Think about

$$\mathbb{E} [\Delta_h X(t) | X(t) = x] = \mu(x)h + o(h).$$

Moreover, the variance can be rewritten

$$\begin{aligned}\text{Var} [\Delta_h X(t) | X(t) = x] &= \mathbb{E} [(\Delta_h X(t))^2 | X(t) = x] - (\mathbb{E} [\Delta_h X(t) | X(t) = x])^2 \\ &= \sigma^2(x)h + o(h) - (\mu(x)h + o(h))^2 \\ &= \sigma^2(X)h + o(h)\end{aligned}$$

The different numbers under o just means that those are different errors.

Definition 2.1.7

We define the infinitesimal moment of order r as

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [(\Delta_h X(t))^r | X(t) = x] \quad r = 3, 4, \dots$$

For diffusion processes these infinitesimal moments are zero for $r > 2$.

Lemma 2.1.1

If a standard process satisfies

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [(\Delta_h X(t))^p | X(t) = x] = 0$$

for some $p > 2$ uniformly for x in a compact subinterval of (l, r) and t in any finite interval in $[0, N]$ then the Dynkin condition is verified.

Proof

Use Chebyshev:

$$\frac{1}{h} \mathbb{P}(|X(t+h) - X(t)| > \varepsilon | X(t) = x) \leq \frac{\mathbb{E} [|\Delta_h X(t)|^p | X(t) = x]}{h \varepsilon^p}.$$

□

2.1.3 Characterization of Diffusion processes

We can define a diffusion process:

1. In terms of infinitesimal coefficients:

$$\begin{aligned} & \mu(x, t) \\ & \sigma^2(x, t) \\ & I \text{ (diffusion interval)} \end{aligned}$$

We should also specify, for example, the behavior at the boundaries of I : it could be an absorbing boundary or some other shit. Think about a Brownian motion on $I(-\infty, \infty]$: depending on whether 0 is absorbing we have two different sample paths.

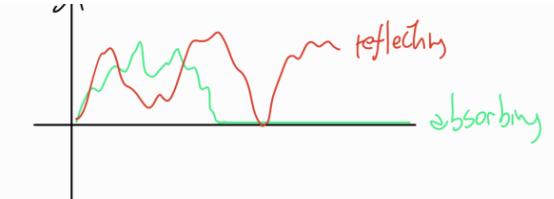


Figure 2.3: The “internal” law of movement remains the same and so do the diffusion parameters. Of course, this is until the process attains 0.

2. We have an alternative definition using stochastic differential equations:

$$\begin{cases} dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_s \\ X_0 = c \end{cases}$$

This is of course just notation. dB_s doesn't make sense as a concept because the Brownian motion doesn't admit derivative! This notation actually means

$$X_t = c + \underbrace{\int_{t_0}^t \mu(X_s, s) ds}_{\text{normal Riemann}} + \underbrace{\int_{t_0}^t \sigma(X_s, s) dB_s}_{\text{Ito integral...}}$$

We will need a new type of integral to deal with this shit: the Ito integral. Which is bad and stupid.

3. We can use martingales and, in particular, the **Stroock-Varadhan** martingale. We do not care.

A diffusion process which is often used is the **Orstein-Uhlenbeck** process:

$$I = (-\infty, +\infty), \quad \begin{cases} \mu(x) = -\alpha(x) \\ \sigma^2(x) = \sigma^2 \end{cases}$$

Where $\alpha > 0, \sigma > 0$ are arbitrary. This gives us the following stochastic differential equation:

$$dX_t = -\alpha X_t dt + \sigma dB_t.$$

Not considering the noise (that in this model is represented by the Brownian motion component) we describe the phenomenon as a simple first order differential equation that we can quite easily solve. Then we add pack the noise (that contains all stochastic elements) and so we have our perturbed model.

There are other methods to “build” new diffusions.

Theorem 2.1.1

Let $X(t)$ be a regular diffusion process on $I = (\lambda, r)$ (it has no difference open or closed). Let $\mu(x)$ and $\sigma^2(x)$ be its infinitesimal moments. Let $g(x)$ be a strictly monotone function on I with continuous first and second derivatives $g'(x)$ and $g''(x)$ for $x \in (l, r)$. Then the transformed process

$$Y(t) = g(X(t))$$

is a regular diffusion process on $(g(l), g(r))$ and

$$\begin{aligned}\mu_Y(y) &= \frac{1}{2}\sigma^2(x)g''(x) + \mu(x)g'(x)\Big|_{x=g^{-1}(y)} \\ \sigma_Y^2(y) &= \sigma^2(x)g'(x)\Big|_{x=g^{-1}(y)}\end{aligned}$$

Where $y = g(x)$.

Proof

Let g be strictly increasing and twice continuously differentiable. Take the Taylor expansion (this is a classic way of dealing with these things)

$$g(x + \Delta X) = g(x) + \Delta X g'(x) = \frac{(\Delta X)^2}{2} g''(x) + \frac{1}{2}(\Delta X)^2 [g''(\xi) - g''(x)]$$

for $x \leq \xi < x + \Delta X$. Substitute

$$\begin{aligned}x &\rightarrow X(t) \\ \Delta x &\rightarrow \Delta X_h(t)\end{aligned}$$

So we get

$$\begin{aligned}g(X(t+h)) &= g(X(t)) + \Delta X_h g'(X(t)) + \frac{(\Delta X_h)^2}{2} g''(X(t)) + \\ &+ \frac{(\Delta X_h)^2}{2} [g''(\xi(\omega)) - g''(X(t))].\end{aligned}$$

Here $X(t) \leq \xi(\omega) \leq X(t+h)$: ξ has an ω as argument because it is now a random variable. Substitute $Y(t)$:

$$\begin{aligned}Y(t+h) - Y(t) &= \Delta X_h g'(X(t)) + \frac{1}{2}(\Delta X_h)^2 g''(X(t)) + \\ &+ \frac{(\Delta X_h)^2}{2} [g''(\xi(\omega)) - g''(X(t))] \quad \text{for } g(X(t)) = Y(t).\end{aligned}$$

Now:

1. condition on $Y(t) = y$ (which means $X(t) = x, g(X(t)) = Y(t) = y$);
2. take the expectation;
3. divide by h and take the limit:

$$\begin{aligned}\lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [Y(t-h) - Y(t) | Y(t) = y] &= \mu(x)g'(X(t)) + \frac{\sigma^2(x)}{2} g''(X(t)) + \\ &+ \frac{1}{2} \lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [g''(\xi(\omega)) - g''(X(t)) | Y(t) = y] \quad \text{with } X(t) = g^{-1}(Y(t)).\end{aligned}$$

We know that $g''(x)$ is continuous, so

$$g''(\xi(\omega)) \rightarrow g''(X(t))$$

and therefore

$$\mu_Y(y) = \lim_{h \searrow 0} \mathbb{E}[Y(t+h) - Y(t) | Y(t) = y] = \mu(x)g'(x) + \frac{\sigma^2(x)}{2} g''(x)|_{x=g^{-1}(y)}.$$

We still need to prove that the diffusion parameter is equal to the one in the hypothesis.
We know that

$$Y(t+h) - Y(t) = \Delta X g'(X(t)) + \frac{1}{2}(\Delta X)^2 g''(X(t)) + \frac{1}{2}(\Delta X)^2 \cdot [g''(\xi(\omega)) + g''(X(t))]$$

and this means that

$$[Y(t+h) - Y(t)]^2 = (\Delta X)^2(g(X(t)))^2 + R_h$$

where R_h contains only $(\Delta X)^3$ and higher order terms: this makes them vanish faster than $(\Delta X)^2$, making them negligible as $h \searrow 0$:

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E}[(Y(t+h) - Y(t))^2 | Y(t) = y] = \sigma^2(y) (g'(y))^2$$

because we know that

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E}[|\Delta X|^r | X(t) = x] = 0 \quad \forall r \geq 3.$$

The condition

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{P}(|X(t+h) - x| > \varepsilon | X(t) = x) = 0$$

can be verified also for $Y(t) = g(X(t))$. This means (???) Thank you Zucca) that $Y(t)$ is a diffusion process with infinitesimal parameters as the one stated in the thesis. \square

2.1.4 Examples of transformed diffusion processes

An example of transformed process is the **Lamperti transform**:

$$g(x) = \int^x \frac{1}{\sigma_X(z)} dz.$$

This is very useful¹ because it allows us to transform any diffusion into a diffusion with constant diffusion parameter, which is much easier to handle. Indeed, the parameters are

$$\begin{aligned} \sigma_Y^2(y) &= \sigma_X^2(x) \cdot \frac{1}{\sigma_X^2(x)} = 1 \\ \mu_Y(y) &= \frac{1}{2} \sigma_X^2(x) \frac{d}{dx} \left(\frac{1}{\sigma_X(x)} \right) + \frac{\mu_X(x)}{\sigma_X(x)} \Big|_{y=g(x)}. \end{aligned}$$

So all the dependency is conveniently contained in the infinitesimal mean.

¹Of course, not in an *actual* sense.

Remark

The meaning of $\mu_Y(y)$ is to capture the drift, or deterministic behavior, of the process while $\sigma_Y^2(y)$ is a noise component. We can also use this fact inside the stochastic differential equation:

$$\begin{cases} dX_t = \underbrace{\mu(t, X_t) dt}_{\text{deterministic function}} + \underbrace{\sigma(t, X_t) dB_t}_{\text{noise that could involve the process itself}} \\ X_0 = x_0. \end{cases}$$

Never forget that this is just ugly notation for

$$X_t = X_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s$$

(yes I know, it's a little bit different than above but it is because we switched professor, be a big boy about it) because technically it doesn't make sense to say dB_s since you can't differentiate Brownian motion. The integral notation holds because we define a new object, the Ito integral, just to make sense for this setting. If σ^2 is continuous is like a Riemann integral, otherwise shit happens.

Another example of transformed process is the **geometric Brownian motion**:

Definition 2.1.8

Let $(X_t)_{t \geq 0}$ be a Brownian motion with drift μ and diffusion coefficient σ^2 (remember that σ is present in the definition of Brownian motion with drift?). Then the process

$$Y(t) = e^{X(t)}$$

is called **geometric Brownian motion** with $I = (0, \infty)$.

This is not anymore a Gaussian process, but it is enough to take its log to go back to the normal Brownian motion. This is often used to model prices of assets traded in a perfect market because it is non negative and it exhibits a long-term exponential growth or decay. We know that

$$\begin{aligned} y &= g(x) = e^x \\ g'(x) &= g''(x) = e^x = y \end{aligned}$$

so using the theorem above we know that

$$\begin{aligned} \mu_Y(y) &= \left(\mu + \frac{1}{2} \sigma^2 \right) y \\ \sigma^2(y) &= \sigma^2 y^2. \end{aligned}$$

We can plug this into our SDE:

$$dY_t = \left(\mu + \frac{1}{2} \sigma^2 \right) Y_t dt + \sigma Y_t dB_t.$$

The interesting thing to note here is that while the process is exponential, the coefficients of the SDE are linear. This tells us how batshit insane the variability of this thing is. If the coefficient was squared we would already be having a very unpleasant time trying to solve this SDE in closed form.

Another useful example is the **Bessel process**. It is the euclidean distance from the origin of an n-dimensional Brownian motion.

$$Y(t) = \sqrt{\underbrace{X_1(t)^2 + \dots + X_n(t)^2}_{\text{components of } n\text{-dim. BM}}}$$

We want to compute the infinitesimal parameters. We need two steps.

1. Define

$$Z(t) = X_1(t)^2 + \dots + X_n(t)^2$$

where $\{X_i(t)\}_{t \geq 0}$ are independent Brownian motion processes. We condition on $X_i(t) = x_i$ for $i = 1, \dots, n$ and we write

$$\begin{aligned} X_i(t + \Delta t) &= x_i + \Delta X_i \\ Z_i(t + \Delta t) &= z + \Delta Z \end{aligned}$$

where $z = x_1^2 + \dots + x_n^2$. We know that

$$\begin{aligned} \Delta Z &= Z(t + \Delta t) - Z(t) \\ &= \underbrace{[X_1(t + \Delta t)]^2 - x_1^2 + \dots + [X_n(t + \Delta t)]^2 - x_n^2}_{(x_1 + \Delta X_1)^2} \\ &= 2(x_1 \Delta X_1 + \dots + x_n \Delta X_n) + [(\Delta X_1)^2 + \dots + (\Delta X_n)^2]. \end{aligned}$$

We know that $\Delta X_1, \dots, \Delta X_n$ are i.i.d. $\sim N(0, \Delta t)$. So we know that

$$\mathbb{E} [\Delta Z | Z(t) = z] = n \Delta t$$

and

$$\mathbb{E} [(\Delta Z)^2 | Z(t) = z] = 4 \left(\sum_{i=1}^n x_i^2 \right) \Delta t + o(\Delta t)$$

where $o(\Delta t)$ contains terms of order $(\Delta X)^4$ or higher. Moreover the fourth moment is

$$\mathbb{E} [(\Delta Z)^4 | Z(t) = z] = o((\Delta t)^2)$$

So the lemma condition holds for $p = 4$. This means that $\{Z(t)\}_{t \geq 0}$ is a diffusion with

$$\begin{aligned} \mu(z) &= n \\ \sigma^2(z) &= 4z. \end{aligned}$$

2. The Bessel process is:

$$Y(t) = \sqrt{Z(t)}$$

so we can set

$$\begin{aligned} g(z) &= \sqrt{z} \\ g'(z) &= \frac{1}{2\sqrt{z}} = \frac{1}{2y} \\ g''(z) &= -\frac{1}{4z^{3/2}} = -\frac{1}{4y^3}. \end{aligned}$$

We can now apply the theorem to get the infinitesimal moments:

$$\begin{aligned} \mu_Y(y) &= \frac{n-1}{2y} \\ \sigma_Y^2(y) &= 1. \end{aligned}$$

Plugging this into the SDE we get

$$dY_t = \frac{n-1}{2Y_t} dt + dB_t.$$

If $n = 1$ we get

$$\begin{cases} \mu_Y(y) = 0 & \text{reflecting Brownian motion} \\ \sigma_Y^2 = 1 & I = [0, \infty). \end{cases}$$

2.2 Kolmogorov backward and forward equations

Let $\{X_t, t \geq 0\}$ be a regular, time homogeneous diffusion process on $I = (l, r)$. Let

$$F(y, t|x, 0) = \mathbb{P}(X(t) < y | X(0) = x)$$

be the transition *distribution* function of $X(t)$ with initial condition

$$F(y, 0|x, 0) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{if } x > y. \end{cases}$$

Let us assume that $F(y, t|x, 0)$ has density on I :

$$f(y, t|x, 0) = \frac{d}{dt} F(y, t|x, 0)$$

which is the transition *density* function.

2.2.1 Kolmogorov backward differential equation

We want to derive a partial differential equation for

$$u(t, x) = \mathbb{E}[g(X(t)) | X(0) = x]$$

where $g(x)$ is a bounded and piecewise continuous function on I . We will see that $u(t, x)$ satisfies

$$\frac{\partial u}{\partial t} = \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} + \mu(x)\frac{\partial u}{\partial x} \quad \text{frog icon}$$

with $u(0^+, x) = \lim_{h \searrow 0} u(h, x) = g(x)$. The particular case

$$g(\eta) = \begin{cases} 1 & \text{if } \eta \leq y \\ 0 & \text{if } \eta > y \end{cases}$$

gives

$$u(t, x) = \mathbb{E}[\mathbb{1}_{\{X(t) \leq y\}} | X(0) = x] = F(y, t|x, 0)$$

and becomes the **Kolmogorov backward equation**:

$$\frac{\partial F(y, t|x, 0)}{\partial t} = \frac{1}{2}\sigma^2(x)\frac{\partial^2 F(y, t|x, 0)}{\partial x^2} + \mu(x)\frac{\partial F(y, t|x, 0)}{\partial x} \quad \text{double arrow icon}$$

for $t > 0, l < x, y > r$ and with initial conditions

$$F(y, 0^+|x, 0) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{if } x > y. \end{cases}$$

Also, the transition density function $f(y, t|x, 0)$ satisfies the Kolmogorov backward equation

$$\frac{\partial f(y, t|x, 0)}{\partial t} = \mu(x)\frac{\partial f(y, t|x, 0)}{\partial x} = \frac{1}{2}\sigma^2(x)\frac{\partial^2 f(y, t|x, 0)}{\partial x^2} \quad \text{cross icon}$$

for $t > 0, l < x, y < r$.

Remark

The solution of \Leftarrow and \rightarrow in general is not unique.

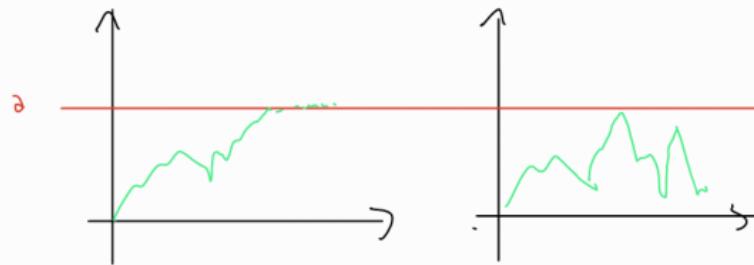


Figure 2.4: On the left, Brownian motion with absorbing boundary; on the right, with reflecting boundary.

Both of the Brownian motions in figure 2.4 satisfy \Leftarrow and \rightarrow but they are clearly different things. We need to add information on the behavior of the process around the boundaries. This is because to find the solution we should add the boundary conditions.

Proof

It is difficult to prove that F, f, u are differentiable in t and twice differentiable in x so we start from these hypotheses and we prove that $f(y, t|x, 0)$ satisfies \rightarrow .

We know that the transition density function f satisfies the **Chapman-Kolmogorov/Smoluchowski** equation

$$f(x, t|x_0, t_0) = \int_I f(x, t|y, \tau) f(y, \tau|x_0|t_0) dy. \quad \text{🏃}$$

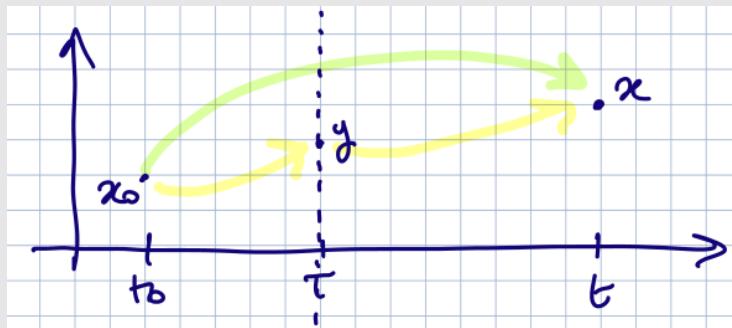


Figure 2.5: I didn't manage to draw this myself because at this point I had largely fallen asleep in the classroom.

We consider the 🏃 equation with

$$\begin{aligned} t_0 &\leq \tau < t \\ x &= X(t) \\ y &= X(\tau) \\ x_0 &= X(t_0). \end{aligned}$$

After substituting t_0 with $t_0 - \Delta t$ and τ with t_0 the equation 🏃 becomes

$$f(x, t|x_0, t_0 - \Delta t) = \int_I f(x, t|y, t_0) f(y, t_0|x_0, t_0 - \Delta t) dy$$

where

$$\begin{aligned}x &= X(t) \\x_0 &= X(t_0 - \Delta t) \\y &= X(t_0) \\ \Delta t &\text{ is a small increment.}\end{aligned}$$

We subtract

$$f(x, t|x_0, t_0) = \int_I f(x, t|x_0, t_0) f(y, t|x_0, t_0 - \Delta t) dy$$

where the yellow part integrates to 1.

$$\begin{aligned}f(x, t|x_0, t_0 - \Delta t) - f(x, t|x_0, t_0) &= \int_I f(y, t_0|x_0, t_0 - \Delta t) \cdot \\&\quad \cdot [f(x, t|y, t_0) - f(x, t|x_0, t_0)] dy \\&= \sum_{n=1}^{\infty} \int_I f(y, t_0|x_0, t_0 - \Delta t) \cdot \\&\quad \cdot \underbrace{\frac{\partial^n f(x, t|x_0, t_0)}{\partial x_0^n} \frac{(y - x_0)^n}{n!}}_{\text{Taylor series in } x_0} dy.\end{aligned}$$

Now divide by $(-\Delta t)$ and let $\Delta t \rightarrow 0$ so that

$$\frac{\partial f(x, t|x_0, t_0)}{\partial t_0} = - \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial^n f(x, t|x_0, t_0)}{\partial x_0^n} \underbrace{\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \int_I (y - x_0)^n f(y, t_0|x_0, t_0 - \Delta t) dy}_{A_n(x_0, t_0)}$$

where $A_n(x_0, t_0)$ are the infinitesimal moments of a diffusion process

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E} [|\Delta_h X(t)|^n | X(t) = x].$$

We know that

1. $A_1(x_0, t_0) = \mu(x_0, t_0);$
2. $A_2(x_0, t_0) = \sigma^2(x_0, t_0);$
3. $A_k(x_0, t_0) = 0 \text{ for } \forall k \geq 3.$

Then

$$-\frac{\partial f}{\partial t_0} = \mu(x_0, t_0) \frac{\partial f}{\partial x_0} + \frac{\sigma^2(x_0, t_0)}{2} \frac{\partial^2 f}{\partial x_0^2}.$$

If the process is time-homogeneous then we have that

$$\begin{aligned}f(x, t|x_0, t_0) &= f(x, t - t_0|x_0, 0) \\A_n(x_0, t_0) &= A_n(x_0) \quad \forall n.\end{aligned} \implies \frac{\partial f}{\partial t} = -\frac{\partial f}{\partial t_0}$$

and therefore

$$\frac{\partial f}{\partial t} = \mu(x_0) \frac{\partial f}{\partial x_0} + \frac{\sigma^2(x_0)}{2} \frac{\partial^2 f}{\partial x_0^2}.$$

□

2.2.2 Kolmogorov forward differential equation

This is also called **Fokker-Plank** equation². This is the dual of the \blacktriangleleft equation. It is the following:

$$\frac{\partial f}{\partial x} = -\frac{\partial}{\partial x} [\mu(x, t)f] + \frac{\partial^2}{\partial x^2} \left[\frac{\sigma^2(x, t)}{2} f \right] \quad \blacktriangleright$$

where f is $f(x, t|x_0, t_0)$.

This equation seems more complex because μ multiplies f and because we have second order derivatives. We have to note that here we are differentiating with respect to *forward space* x while before we were taking into account the starting point at time 0. We won't prove this result, but we need to use the Chapman-Kolmogorov equation with

$$\begin{aligned} t &\rightsquigarrow t + \Delta t \\ \tau &\rightsquigarrow t. \end{aligned}$$

We see some examples of application.

1. *General Brownian motion.* We know that the Brownian motion is time-homogeneous. Recall the \blacktriangleleft equation that in this case is:

$$\frac{\partial f}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x_0^2} \quad I = (-\infty, +\infty).$$

This is the classical heat equation. We can prove that the result is the following:

$$f(x, t|x_0, t_0) = \frac{1}{\sqrt{2\pi\sigma^2(t-t_0)}} \exp \left\{ -\frac{(x-x_0)^2}{2\sigma^2(t-t_0)} \right\}$$

and this is none other than the solution to the \blacktriangleleft equation. The system that gives us this solution is:

Cauchy Problem

$$\begin{cases} \frac{\partial f}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x_0^2} \\ \lim_{t \rightarrow t_0} f(x, t|x_0, t_0) = \underbrace{\delta(x-x_0)}_{\text{Dirac funct.}} & \text{initial condition} \\ \lim_{|x| \rightarrow \infty} f(x, t|x_0, t_0) = \lim_{|x| \rightarrow \infty} \frac{\partial f}{\partial x} = 0 & \text{boundary condition} \end{cases}$$

The boundary condition specifies what happens at the boundaries of the diffusion interval, that in this case are $-\infty$ and $+\infty$. This is just an analytical problem: this has the advantage that we can basically forget about probability and other similar things.

2. *Reflected Brownian motion with boundary a .* We said that the Kolmogorov equation is the same for different processes, if we don't specify the boundaries. But how can we modify the system of equations to have a reflecting boundary? Recall \blacktriangleright equation. The transition probability density function $f^+(x, t|x_0, t_0)$ satisfies

$$\frac{\partial f^+}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 f^+}{\partial x^2} \quad I = (-\infty, a).$$

This seems equal to the \blacktriangleleft equation but just because the coefficients are constants.

²Fokker? I hardly know her.

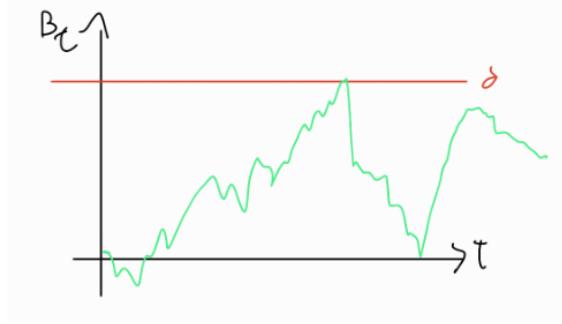


Figure 2.6: Do we really need this?

We know that

$$\int_{-\infty}^a f^+(x, t|x_0, t_0) dx = 1$$

because all the mass of the motion is below a . We take the derivative with respect to time:

$$\int_{-\infty}^a \frac{\partial f^+}{\partial t} dx = 0$$

but I know from the \blacktriangleright equation that $\frac{\partial f^+}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 f^+}{\partial x^2}$ so we can substitute and get

$$\int_{-\infty}^a \frac{\sigma^2}{2} \frac{\partial^2 f^+}{\partial x^2} dx = \left. \frac{\sigma^2}{2} \frac{\partial^2 f^+}{\partial x^2} \right|_{-\infty}^a = 0.$$

So the boundary condition for a reflecting boundary is

$$\left. \frac{\partial f^+}{\partial x} \right|_{x=a} = 0.$$

The system becomes

$$\begin{cases} \frac{\partial f^+}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 f^+}{\partial x^2} & x < a \\ \lim_{t \rightarrow t_0} f^+(x, t|x_0, t_0) = \underbrace{\delta(x - x_0)}_{\text{Dirac funct.}} & t > 0 \\ \left. \frac{\partial f^+}{\partial x} \right|_{x=a} = 0 & \text{upper bound} \\ \lim_{x \rightarrow \infty} f^+(x, t|x_0, t_0) = \lim_{x \rightarrow \infty} \frac{\partial^+ f}{\partial x} = 0 & \text{lower bound} \end{cases} \quad \text{initial condition}$$

3. *Brownian motion with absorbing boundary a .* The transition probability density function $\bar{f}(x, t|x_0, t_0)$ satisfies \blacktriangleright :

$$\frac{\partial \bar{f}}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 \bar{f}}{\partial x^2}.$$

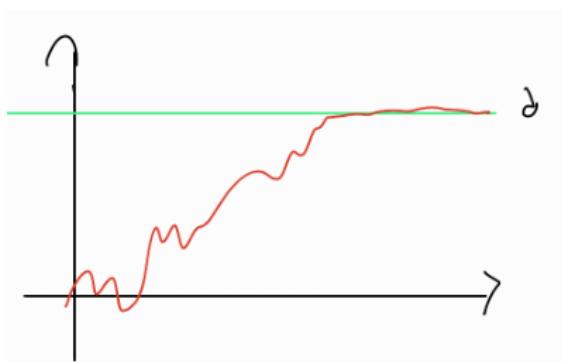


Figure 2.7: No, we don't.

The integral of the density will be lower than 1! This is because everything is concentrated in a which being a point has probability 0:

$$\int_{-\infty}^a \bar{f}(x, t|x_0, t_0) dx < 1.$$

The system becomes

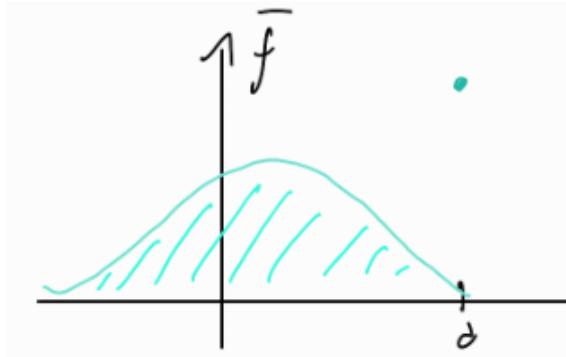


Figure 2.8: This is a mixed distribution of continuous and discrete and it is “incomplete”!

$$\begin{cases} \frac{\partial \bar{f}}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 \bar{f}}{\partial x_0^2} & x < a \\ \lim_{t \rightarrow t_0} \bar{f}(x, t|x_0, t_0) = \underbrace{\delta(x - x_0)}_{\text{Dirac funct.}} & t > 0 \\ \bar{f}(x, t|x_0, t_0) = 0 & \text{initial condition} \\ \lim_{x \rightarrow \infty} \bar{f}(x, t|x_0, t_0) = \lim_{|x| \rightarrow \infty} \frac{\partial \bar{f}}{\partial x} = 0 & \text{lower bound} \\ & \text{upper bound} \end{cases}$$

4. *Brownian motion with absorbing boundary $a > 0$.* Let us denote $\{\bar{B}(t), t \geq 0\}$ as the Brownian motion with absorbing boundary a . Let

$$T_a := \inf\{t > B(t)\}$$

be the first passage time through a . We want to define

$$\bar{B}(t) = \begin{cases} B(t) & \text{if } t < T_a \\ a & \text{if } t \geq T_a. \end{cases}$$

We want to find the transition density function of $\bar{B}(t)$, that is

$$\mathbb{P}(\bar{B}(t) \leq y) \quad \text{which is short for } \mathbb{P}(\bar{B}(t) \leq y | B(0) = 0).$$

The aim here is to express it in terms of the normal Brownian motion $B(t)$. We know that

$$\begin{aligned} \mathbb{P}(\bar{B}(t) \leq y) &= \mathbb{P}(B(t) \leq y, T_a > t) \quad y < a \\ &= \mathbb{P}(B(t) \leq y) - \mathbb{P}(B(t) \leq y, T_a < t) \quad \text{because } \mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cup B^c) \\ &= \mathbb{P}(B(t) \leq y) - \mathbb{P}\left(B(t) \leq y \max_{0 \leq s \leq t} B(s) \geq a\right) \\ &= \mathbb{P}(B(t) \leq y) - \mathbb{P}(B(t) \geq 2a - y) \quad \text{refl. principle} \\ &= \mathbb{P}(B(t) \leq y) - [1 - \mathbb{P}(B(t) < 2a - y)] \end{aligned}$$

so then transition density function is

$$\begin{aligned}\bar{f}_a(t, y|0, 0) &= f(t, y|0, 0) - f(t, y|0, 2a) \\ &= \underbrace{\frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} - \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-2a)^2}{2t}}}_{\leq 1 \text{ as expected!}}.\end{aligned}$$

I basically delete from the normal transition density function $f(t, y|0, 0)$ all trajectories that are “not acceptable” and for the reflection principle they happen to equivalent to all the trajectories that start from $2a$! This is very cool and it is a neat trick to solve the system . We could do this only because the Brownian motion has the reflection property... and this is not a common thing.

5. *Brownian motion with reflecting boundary $a > 0$.* Define $\{B^+(t), t \geq 0\}$ as the Brownian motion with reflecting boundary $a > 0$:

$$B^+(t) = \begin{cases} B(t) & \text{if } B(t) \leq a \\ 2a - B(t) & \text{if } B(t) > a. \end{cases}$$

As before, remember that $\mathbb{P}(B^+(t) \leq y)$ should actually be $\mathbb{P}(B^+(t) \leq y|0, 0)$.

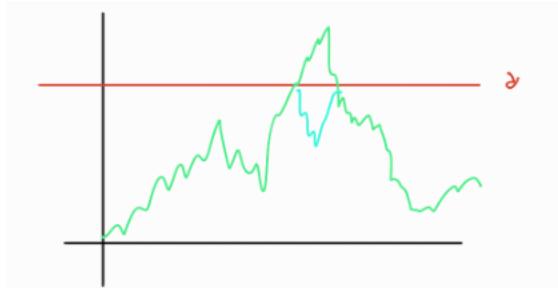


Figure 2.9: At this point you should know how a fucking Brownian motion looks like, why do you keep on drawing it?

$$\begin{aligned}\mathbb{P}(B^+(t) \leq y) &= \mathbb{P}(\{B(t) \leq y\} \cup \{B(t) > 2a - y\}) \\ &= \mathbb{P}(B(t) \leq y) + \mathbb{P}(B(t) > 2a - y) \\ &= \mathbb{P}(B(t) \leq y) + [1 - \mathbb{P}(B(t) \leq 2a - y)]\end{aligned}$$

so the transition density function is

$$\begin{aligned}f^+(t, y|0, 0) &= f(t, y|0, 0) + f(t, y|0, 2a) \\ &= \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} + \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-2a)^2}{2t}}.\end{aligned}$$

The sign $+$ makes sense since now we are *adding* new trajectories.

2.2.3 Solution of the Kolmogorov Equations with transformation method

Other than using the Fourier transform (characteristic method) we can solve the Kolmogorov equations with the **transformation** method. What the fuck?

Take the Brownian motion with drift with

$$\begin{aligned}\mu(X) &= \mu \\ \sigma^2(X) &= \sigma^2.\end{aligned}$$

We observe that the transformation

$$\begin{cases} x'_0 = \frac{\sqrt{2}}{\sigma}(x_0, \mu_t) \\ t'_0 = t_0 \end{cases}$$

changes space but not time. The conservation of probability mass allows us to say

$$f'(x', t' | x'_0, t'_0) dx' = f(x, t | x_0, t_0) dx$$

and therefore the Komlogorov equation becomes

$$\frac{\partial f}{\partial t_0} = \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x_0^2} + \mu \frac{\partial f}{\partial x_0} \implies \frac{\partial f'}{\partial t'_0} = \frac{\partial^2 f'}{\partial x'_0^2}.$$

On a closer look we see that this is not the standard Brownian motion, because that would have

$$\frac{\partial f'}{\partial t'_0} = \frac{\sigma^2}{2} \frac{\partial^2 f'}{\partial x'_0^2}.$$

So this is actually a Brownian motion with $\sigma^2 = 2$. The initial condition becomes:

$$\begin{aligned} \lim_{t \rightarrow t_0} f(x, t | x_0, t_0) &= \lim_{t' \rightarrow t'_0} \frac{\sqrt{2}}{\sigma} f'(x', t' | x'_0, t'_0) \\ &= \lim_{t' \rightarrow t'_0} \frac{\sqrt{2}}{\sigma} f'\left(\frac{\sqrt{2}}{\sigma}(x - \mu t), t \middle| \frac{\sqrt{2}}{\sigma}(x_0 - \mu t_0)\right) \\ &= \frac{\sqrt{2}}{\sigma} \delta\left(\frac{\sqrt{2}}{\sigma}(x - x_0)\right) \\ &= \delta(x - x_0). \end{aligned}$$

What does this mean? This transformation lets us move from a process to another one that is more manageable. We know $f'(x', t' | x'_0, t'_0)$ so we get

$$\begin{aligned} f(x, t | x_0, t_0) &= \frac{\sqrt{2}}{\sigma} \frac{1}{2\pi(t' - t'_0)} e^{-\frac{(x' - x'_0)^2}{4(t' - t'_0)}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2(t - t_0)}} e^{-\frac{(x - x_0 - \mu(t - t_0))^2}{2\sigma^2(t - t_0)}}. \end{aligned}$$

Remark

Using the transition density function I can prove

$$\lim_{t \rightarrow \infty} F(y, t | x_0, t_0) = \begin{cases} \frac{1}{2} & \mu = 0 \\ 1 & \mu < 0 \\ 0 & \mu > 0. \end{cases}$$

As an example, we see the Ornstein-Uhlenbeck process (also known as Vasicek process). We have

$$\begin{cases} \mu(x) = -\frac{x}{\theta} \\ \sigma^2(x) = \sigma^2. \end{cases} \quad (\text{for Vasicek process } \mu(x) = \alpha x)$$

This is the easier diffusion process after Brownian motion. It is Gaussian so we can write

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x} \left(\frac{x}{\vartheta} f \right) + \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}.$$

We can use Fourier transforms to solve it:

$$f(x, t|x_0, t_0) = \frac{1}{2\pi\sigma^2 \frac{\vartheta}{2} \left(1 - e^{-\frac{2(t-t_0)}{\vartheta}}\right)} \exp \left\{ -\frac{\left(x - x_0 e^{-\frac{(t-t_0)}{\vartheta}}\right)^2}{2\sigma^2 \vartheta \left(1 - e^{-\frac{2(t-t_0)}{\vartheta}}\right)} \right\}.$$

Even if this looks strange, it is still a Gaussian distribution. It makes sense: it has constant diffusion and the drift is a linear term. If we want to use the transformation method we do

$$\begin{cases} x'_0 = \sqrt{\frac{2}{\sigma^2}} e^{\frac{t_0}{\vartheta}} x_0 \\ t'_0 = \frac{\vartheta}{2} e^{\frac{2t_0}{\vartheta}} \end{cases} \quad \begin{cases} x' = \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} x \\ t' = \frac{\vartheta}{2} e^{\frac{2t}{\vartheta}} \end{cases}$$

so we get

$$f(x, t|x_0, t_0) = \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} f'(x', t'|x'_0, t'_0)$$

What we basically did was to go from an Ornstein-Uhlenbeck (x, t) process to a Brownian motion (x', t') .

Proof

We know three things:

1.

$$\begin{aligned} \frac{\partial f}{\partial t_0} &= \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \left[\frac{\partial f'}{\partial x'_0} \frac{\partial x'_0}{\partial t_0} + \frac{\partial f'}{\partial t'_0} \frac{\partial t'_0}{\partial t_0} \right] \\ &= \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \left[\frac{\partial f'}{\partial x'_0} \frac{\partial x'_0}{\vartheta} + \frac{\partial f'}{\partial t'_0} \frac{\partial t'_0}{\vartheta} \right] \end{aligned}$$

2.

$$\begin{aligned} \frac{\partial f}{\partial x_0} &= \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \frac{\partial f'}{\partial x'_0} \frac{\partial x'_0}{\partial x_0} \\ &= \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \frac{\partial f'}{\partial x'_0} \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \end{aligned}$$

3.

$$\frac{\partial^2 f}{\partial x_0^2} = \dots = \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \frac{\partial^2 f'}{\partial x'^2} e^{\frac{2t_0}{\vartheta}} \frac{2}{\sigma^2}.$$

Substitute the last equation in the Kolmogorov equation and get

$$\begin{aligned} \sqrt{\frac{2}{\sigma^2}} e^{\frac{t}{\vartheta}} \left[\frac{\partial f'}{\partial x'_0} \frac{x'_0}{\vartheta} + \frac{\partial f'}{\partial t'_0} \frac{2t'_0}{\vartheta} - \cancel{\frac{x'_0}{\vartheta} \frac{\partial f'}{\partial x'_0}} + e^{\frac{2t_0}{\vartheta}} \frac{\partial^2 f'}{\partial x'^2} \right] &= 0 \\ \Downarrow \\ \cancel{\frac{\partial t'_0}{\vartheta} \frac{\partial f'}{\partial t'_0}} + \cancel{\frac{2t'_0}{\vartheta} \frac{\partial^2 t'_0}{\partial x'^2}} &= 0 \end{aligned}$$

That is the Kolmogorov equation for Brownian motion. We still need to prove that

$$\lim_{t \rightarrow t_0} f(x, t | x_0, t_0) = \delta(x - x_0) \implies \lim_{t' \rightarrow t'_0} f(x', t' | x'_0, t'_0) = \delta(x' - x'_0)$$

but you can do this as an exercise. 

□

Now we want to transform arbitrary processes in Brownian motions so that we can enjoy the “nice” properties of Brownian motion also for those processes.

Theorem 2.2.1

Let $c_1(t)$ and $c_2(t)$ be two arbitrary functions of t . The process with transition density function f , that is the solution of

$$\frac{\partial f}{\partial t_0} + b(x_0, t_0) \frac{\partial f}{\partial x_0} + a(x_0, t_0) \frac{\partial^2 f}{\partial x_0^2} = 0$$

can be transformed into a Brownian motion with $\sigma^2 = 2$ and transition density function

$$\frac{\partial f'}{\partial t'_0} + \frac{\partial^2 f'}{\partial x'^2_0} = 0$$

if and only if

$$b(x_0, t_0) = \frac{1}{2} \frac{\partial a(x_0, t_0)}{\partial x_0} + \frac{\sqrt{a(x_0, t_0)}}{2} \left[c_1(t_0) + \int^{x_0} \frac{c_2(t_0)a(z, t_0) + \frac{\partial a(z, t_0)}{\partial t_0}}{[a(z, t_0)]^{3/2}} dz \right].$$

*

This transformation (apart from constants) is

$$\begin{cases} x'_0 = \psi(x_0, t_0) = e^{-\frac{1}{2} \int^{t_0} c_s(s) ds} \left[\int^{x_0} (a(z, t_0))^{-\frac{1}{2}} dz - \frac{1}{2} \int^{t_0} c_1(\vartheta) e^{-\frac{1}{2} \int^\vartheta c_1(s) ds} d\vartheta \right] \\ t'_0 = \varphi(t_0) = \int^{t_0} e^{-\int^\vartheta c_2(s) d\vartheta} d\vartheta \end{cases}$$

with

$$f(x, t | x_0, t_0) dx = f'(x', t' | x'_0, t'_0) dx'$$

so the probability mass is conserved.

So we move from two diffusion processes, with diffusion coefficient a and drift coefficient b , to a Brownian motion. This is possible only if we can write b as in *, that is if we can find c_1 and c_2 ! The space transformation x'_0 depends on space and time, while t'_0 depends only on time.

2.3 Differential equation for the boundary classification

Never has a title sounded less appealing.

2.3.1 Differential equations associated with certain functionals

Let us fix a and b such that $l < a < b < r$. Let $T(y) = T_y$ be the first hitting time of a random process through a constant level y . Let us consider $\{X(t), t \geq 0\}$ as a time-homogeneous diffusion process such that:

1. $I = (l, r)$ or $[l, r]$ or $(l, r]$ or $[l, r)$ with $\infty \leq l < r \leq \infty$;

2. the process is regular in the interior of I :

$$\mathbb{P}(T(y) < \infty | X(0) = x) > 0 \quad l < x, t < r;$$

3. The process has infinitesimal parameters $\mu(x)$ and $\sigma^2(x)$;

4. $\mu(x)$ and $\sigma^2(x)$ are continuous functions of $\sigma^2(x) > 0$ and $l < x < r$.

We define

$$T^* = T_{a,b} = \min\{T_a, T_b\} = T_a \wedge T_b$$

as the first hitting time between barrier a and barrier b .

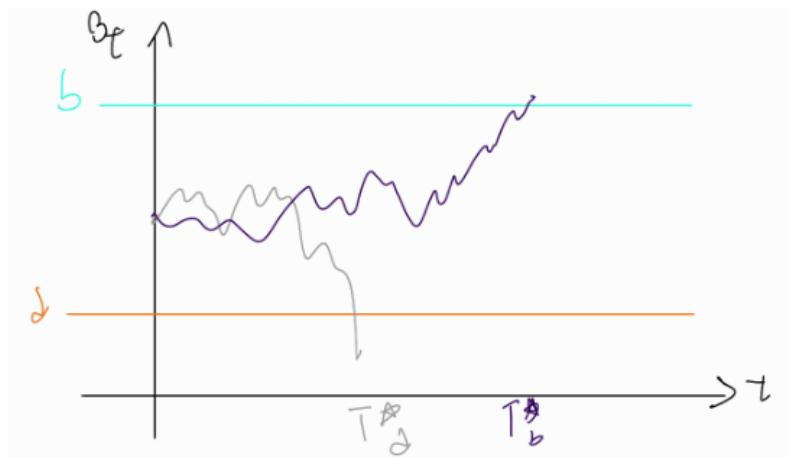


Figure 2.10: T^* is the time in which whichever boundary gets attained.

Then we consider three problems:

① Find

$$u(x) = \mathbb{P}(T_b < T_a | X(0) = x) \quad a < x < b$$

that is the probability of crossing b earlier than a .

② Find

$$v(x) = \mathbb{E}[T^* | X(x) = x] \quad a < x < b.$$

③ For a bounded continuous function g find

$$w(x) = \mathbb{E}\left[\int_0^{T^*} g(X(s)) ds | X(0) = x\right] \quad a < x < b.$$

Remark

Is $\int_0^{T^*} g(X(s)) ds$ well defined? Yes because $X(t)$ is a diffusion process with continuous trajectory. Moreover, if $g(x) = 1$ we have

$$\int_0^{T^*} g(X_s) ds = T^*$$

so problem ③ is a special case of problem ②.

In a standard Brownian motion the mean is 0. The probability of reaching a level y is 1 but as we saw the expected time to reach it is infinite! We are trying to look whether the mean time behaves like that also in other diffusion processes. Under the four hypotheses stated above it can be proved that

- $v(x)$ and $w(x)$ are finite;
- $u(x), v(x)$ and $w(x)$ have two bounded derivatives for $a < x < b$;
- they satisfy the following differential equations:

A:

$$0 = \mu(x) \frac{du}{dx} + \frac{\sigma^2(x)}{2} \frac{d^2u}{dx^2} \quad a < x < b \\ u(a) = 0 \\ u(b) = 1.$$

This one looks similar to the Kolmogorov equation...

B:

$$-1 = \mu(x) \frac{dv}{dx} + \frac{\sigma^2(x)}{2} \frac{d^2v}{dx^2} \quad a < x < b \\ v(a) = v(b) = 0.$$

C:

$$-g(x) = \mu(x) \frac{dw}{dx} + \frac{\sigma^2(x)}{2} \frac{d^2w}{dx^2} \quad a < x < b \\ w(a) = w(b) = 0.$$

We are going to see an heuristic derivation of equation A. We want to find u , that is the probability of reaching b before a . Consider $x \in (a, b)$.

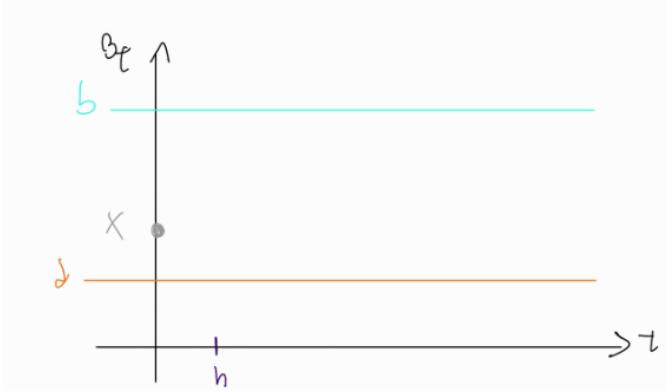


Figure 2.11: Very informative.

We choose h such that the probability of reaching a or b is negligible (h is small so the process doesn't "have time" to reach either a or b). So the probability of reaching b before a conditioning on the position $X(h)$ is $u(X(h))$ so we get

$$\begin{aligned} u(x) &= \mathbb{P}(T_b < T_a | X(0) = x) \\ &= \int_y \mathbb{P}(T_b < T_a | \underbrace{X(0) = x}_{\text{Markov p.ty}}, X(h) = y) \underbrace{\mathbb{P}(X(h) \in dy | X(0) = x)}_{\text{negligible by hypothesis}} \\ &= \mathbb{E}[u(X(h)) | X(0) = x] + o(h). \end{aligned}$$

Then consider $\Delta X = X(h) - x$ and expand $u(X(h))$ in the Taylor series:

$$\begin{aligned} u(X(h)) &= u(\Delta X + x) \\ &= u(x) + \Delta X u'(x) + \frac{1}{2} (\Delta X)^2 u''(x) + \dots \end{aligned}$$

so

$$\begin{aligned}
u(x) &= \mathbb{E}[u(X(h))|X(0)=x] + o(h) \\
&= \mathbb{E}[u(x+\Delta X)|X(0)=x] + o(h) \\
&= u(x) + \underbrace{\mathbb{E}[\Delta X|X(0)=x]}_{\text{drift: } \mu(x)h+o(h)} u'(x) + \frac{1}{2} \underbrace{\mathbb{E}[(\delta X)^2|X(0)=x]}_{\sigma^2(x)h+o(h)} u''(h) + o(h) \\
&= u(x) + \mu(x)hu'(x) + \frac{1}{2}\sigma^2(x)hu''(x) + o(h).
\end{aligned}$$

Divide by h and let $h \rightarrow 0$:

$$0 = \mu(x)u'(x) + \frac{1}{2}\sigma^2(x)u''(x) \quad x \in (a, b).$$

Why should we care about all this? Because we exited the realm of probability (ugly, uncertain, retarded) by transforming the probability in a problem of solving an ordinary differential equation (deterministic, analytic, sexy). Now we look for the solutions of the problems A, B and C. We need some definitions:

Definition 2.3.1

Let $\sigma^2(x) > 0$ for $x \in (l, r)$. The **scale density** is defined as

$$s(x) = \exp \left\{ - \int^x \frac{2\mu(\xi)}{\sigma^2(\xi)} d\xi \right\} \quad x \in (l, r).$$

Definition 2.3.2

The **scale function** or **scale measure** is defined as

$$S(x) = \int^x s(\eta) d\eta = \int^x \exp \left\{ - \int^\eta \frac{2\mu(\xi)}{\sigma^2(\xi)} d\xi \right\} d\eta$$

Definition 2.3.3

The **speed density** is defined as

$$m(x) = \frac{1}{\sigma^2(x)s(x)} \quad x \in (l, r).$$

In our process σ^2 tells us how “noisy” the process is. If $m(x)$ is high, the diffusion is low. This is like asking «How fast can I go away from a certain level?»

Definition 2.3.4

The **speed measure** is defined as

$$M(x) = \int^x m(y) dy.$$

Our differential equations A, B and C involve the differential operator \mathcal{L} :

$$\mathcal{L}f(x) := \mu(x)f'(x) + \frac{\sigma^2(x)}{2}f''(x)$$



where f is twice differentiable with continuity on (a, b) . We express the operator \mathfrak{L} in terms of scale and speed. Observe that

$$\frac{s'(x)}{s(x)} = -\frac{2\mu(x)}{\sigma^2(x)}$$

so  becomes

$$\mathfrak{L}f(x) = \frac{1}{2} \left(\frac{1}{m(x)} \right) \frac{d}{dx} \left[\frac{1}{s(x)} \frac{df(x)}{dx} \right].$$

We write

- $s(x) = \frac{ds(x)}{dx} \implies dS(x) = s(x) dx;$
- $m(x) = \frac{dM(x)}{dx} \implies dM(x) = m(x) dx.$

So

$$\mathfrak{L}f(x) = \frac{1}{2} \frac{d}{dM} \left[\frac{df(x)}{dS} \right].$$

This is called **canonical representation of the differential infinitesimal operator**. It is now easy to write the solutions:

A: canonical representation of the problem A:

$$\frac{1}{2} \frac{d}{dM} \left[\frac{du(x)}{dS} \right] = 0 \quad \begin{aligned} x &\in (a, b) \\ u(a) &= 0 \\ u(b) &= 1. \end{aligned}$$

Integrate twice:

$$\frac{du(x)}{dS(x)} = \beta \implies u(x) = a + \beta S(x) \quad x \in (a, b).$$

We use the boundary conditions:

$$\begin{aligned} u(a) = 0 &\implies a = -\beta S(a) \\ u(b) = 1 &\implies \beta = \frac{1}{S(b)} \\ &\Downarrow \\ u(x) &= \frac{S(x) - S(a)}{S(b) - S(a)} \quad x \in [a, b]. \end{aligned}$$

Remark

If we substitute $S(x)$ with $S^\star(x) = \alpha + \beta S(x)$ with constants $\alpha, \beta \neq 0$ then $u(x)$ does not change. Our solution is invariant to linear transformations of the scale function. The scale function can be used to rescale the state space (l, r) in terms of the probability of reaching all levels.

So what we need to do is:

- fix the origin x_0 ;
- determine the scale function in such a way that the translation yields $S(x_0) = 0$;
- consider $Y(t) = S(X(t))$. This process is just a function applied to X so I can use the transformation theorem. Analyze $S(X(t))$:

- ◇ it lives in $I = (S(l), S(r))$;
- ◇ it is strictly monotone;
- ◇ it is twice continuously differentiable.

So we can use the transformation theorem!

The infinitesimal parameters of $\{Y(t), t \geq 0\}$ are:

$$\begin{aligned}\mu_Y(y) &= \frac{1}{2}\sigma^2(x)S''(x) + \mu(X)S'(x) \\ &= \frac{1}{2}\sigma^2(x)s'(x) + \mu(x)s(x) \\ &= \frac{1}{2}\cancel{\sigma^2(x)} \left[-\cancel{\frac{\mu(x)}{\sigma^2(x)}} \right] s(x) + \mu(x)s(x) \\ &= 0 \\ \sigma_Y^2(y) &= \sigma^2(x) [S'(x)]^2 \\ &= \sigma^2 s(x)^2 \Big|_{y=S(x)}.\end{aligned}$$

The scale measure of $\{Y(t), t \geq 0\}$ is

$$S_Y(y) = y \text{ or, equivalently, } S_Y(y) = \alpha + \beta y.$$

Definition 2.3.5

A process $\{Y(t), t \geq 0\}$ whose scaling function is linear is said to be in **natural/canonical scale** since hitting probabilities are proportional to actual distances:

$$\mathbb{P}(T_a < T_b | Y(0) = y) = \frac{b-y}{b-a} \quad y \in (a, b).$$

We see two examples of this thing.

(a) Standard Brownian motion: we have

$$\begin{aligned}\mu(x) &= 0 \\ \sigma^2 &= 1.\end{aligned}$$

then

$$\begin{aligned}s(x) &= \exp \left\{ -2 \int^x \frac{\mu(\xi)}{\sigma^2(\xi)} d\xi \right\} = 1 \\ m(\xi) &= \frac{1}{s(\xi)} = 1.\end{aligned}$$

(b) Brownian motion with drift: we have

$$\begin{aligned}\mu(x) &= \mu \\ \sigma^2 &= \sigma^2.\end{aligned}$$

so we have

$$\begin{aligned}s(x) &= e^{\frac{2\mu}{\sigma^2}} \\ S(x) &= Ae^{-\frac{2\mu}{\sigma^2}x} + B \quad \text{for any } A, B \\ u(x) &= \frac{e^{-\frac{2\mu}{\sigma^2}x} - e^{-2\frac{\mu a}{\sigma^2}}}{e^{-\frac{2\mu b}{\sigma^2}} - e^{-2\frac{\mu a}{\sigma^2}}}.\end{aligned}$$

C:

$$w(x) = 2 \left[u(x) \int_x^b [S(b) - S(\xi)] m(\xi) g(\xi) d\xi + (1 - u(x)) \int_a^x [S(\xi) - S(a)] m(\xi) g(\xi) d\xi \right].$$

B:

$$v(x) = 2 \left[u(x) \int_x^b [S(b) - S(\xi)] m(\xi) g(\xi) d\xi + (1 - u(x)) \int_a^x [S(\xi) - S(a)] m(\xi) g(\xi) d\xi \right].$$

So it is always dependent on u .

2.3.2 Stationary distribution for time homogeneous diffusion

Does it exist

$$\lim_{t \rightarrow \infty} f(y, t|x, 0) = \psi(y)?$$

Is it possible that the limit does not depend on the initial position anymore? If such $\psi(y)$ exists, it satisfies

$$\psi(y) = \int \psi(z) f(y, s|z, 0) dz \quad \forall t > 0.$$

We start from the Chapman-Kolmogorov equation

$$f(y, t+s|x, t) = \int f(x, t|x, 0) f(y, s|z, 0) dz$$

and take the limit for $t \rightarrow \infty$.

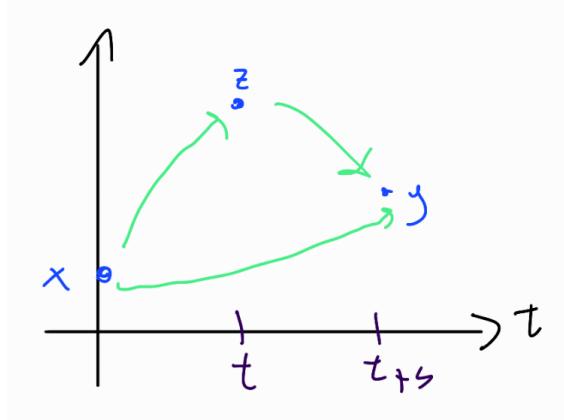


Figure 2.12: Classic Chapman-Kolmogorov action here. Going to y is like going to y and doing a pit-stop at z .

It is possible to prove that $\psi(x)$ satisfies the Fokker-Plank equation

$$0 = \frac{1}{2} \frac{\partial^2}{\partial y^2} [\sigma^2(y) \psi(y)] - \frac{\partial}{\partial y} [\mu(y) \psi(y)].$$

⋮

Here ψ does not depend on t because

$$\lim_{t \rightarrow \infty} \frac{\partial f(y, t|x, 0)}{\partial t} = 0.$$

We now compute the stationary distribution: we integrate ψ :

$$\frac{d}{dy} \left[\frac{\sigma^2(y)}{2} \psi(t) \right] - \mu(y) \psi(y) = \frac{1}{2} c_1$$

where c_1 is a constant. Multiply by the scale density

$$s(y) = e^{-\int^y \frac{\mu(\xi)}{\sigma^2(\xi)} d\xi}$$

so to get

$$\frac{d}{dy} [s(y) \sigma^2(y) \psi(y)] = c_1 s(y).$$

Not gonna make the computation explicitly of course because I don't hate myself. Integrating we get

$$\begin{aligned} \psi(x) &= c_1 \frac{S(x)}{s(x) \sigma^2(x)} + c_2 \frac{1}{s(x) \sigma^2(x)} \\ &= m(x) [c_1 S(x) + c_2]. \end{aligned}$$

If I cannot find the right constants c_1 and c_2 the stationary density function doesn't exist. The constants are determined by the condition that $\psi(x)$ is a density:

$$\psi(x) = 0 \quad \int_I \psi(x) dx = 1.$$

For example, the Brownian motion has no stationary distribution! Consider the Ornstein-Uhlenbeck process. Its scale density is

$$s(x) = e^{\mu \frac{x^2}{\sigma^2}}$$

so we get

$$\psi(x) = c_1 \left(\int_0^x e^{\frac{\mu}{\sigma^2} y} dy \right) e^{-\frac{\mu}{\sigma^2} x^2} + c_2 e^{-\frac{\mu}{\sigma^2} x^2}.$$

We add conditions to get a density: $\psi(x)$ should be positive, we set $c_1 = 0$. This gives us

$$\psi(x) = \underbrace{c e^{-\frac{\mu}{\sigma^2} x}}_{\text{integrates to 1}}.$$

2.3.3 Boundary Classification

The idea is that we want to understand the behavior of the process near the boundaries a and b . Let us consider $\{X(t), t \geq 0\}$ as a regular diffusion process on $I = (l, r)$. For $x \in (l, r)$ we postulate that $\mu(x)$ and $\sigma^2(x)$ are continuous. Consider the lower boundary (for the upper boundary things don't change). The idea is that we want to consider these two probabilities

$$u(x) = u_{a,b}(x) = \underbrace{\mathbb{P}(T_b < T_a | X(0) = x)}_{\text{what we called "problem A"}} \quad l < a < x < b < r$$

$$v(x) = v_{a,b}(x) = \underbrace{\mathbb{E}[T_{ab} | X(0) = x]}_{\text{what we called "problem C}}} \quad l < a < x < b < r$$

and let $a \rightarrow l$.

Definition 2.3.6

The scale measure of a closed interval

$$J = [c, d] \subset (l, r)$$

is

$$S(J) = S[c, d] = S(d) - S(c).$$

Definition 2.3.7

The scale measure of an infinitesimal interval interval $[x, x + dx]$ is

$$\begin{aligned} dS(x) &= S[dx] = S(x + dx) - S(x) \\ &= s(x) dx. \end{aligned}$$

We can extract three properties:

1. $0 < S[c, d] < \infty$ for $l < c, d < r$;
2. $S[c, d] = S[c, x] + S[x, d]$ for $l < c < x < d < r$;
3. $S[a, d]$ is monotone in a for a fixed d .

Definition 2.3.8

The speed measure of the interval $J = [c, d] \subset (l, r)$ is

$$M(J) = M[c, d] = \int_c^d m(x) dx = \int_c^d \frac{1}{\sigma^2(x)s(x)} dx.$$

$M[J]$ is positive and finite for $J = [c, d] \subset (l, r)$. Let us recall that $u(x)$ and $v(x)$ can be rewritten in terms of scale and speed measure:

$$\begin{aligned} u(x) &= \frac{S[a, x]}{S[a, b]} = \frac{S(x) - S(a)}{S(b) - S(a)} \\ v(x) &= 2 \left[u(x) \int_x^b S[\eta, b] dM(\eta) + (1 - u(x)) \int_a^x S[a, \eta] dM(\eta) \right]. \end{aligned}$$

Since property 3) holds, we can define

$$S(l, b] = \lim_{a \searrow l} S[a, b] \leq \infty.$$

We know that if $[a, b] \subset (l, r)$ then $0 \leq S[a, b] \leq \infty$. For this reason and for property 2 we have

$$S(l, b] = \infty \quad \text{for some } b \in (l, r) \iff S(l, b] = \infty \quad \text{for all } b \in (l, r)$$

Remark

Note that $S(l, b]$ depends only on the process parameters in the interior of the state space.

This is because at the boundary the process is not explained by the parameters alone (what if the boundary is absorbing or reflecting?). Let us consider the hitting time problem for l . We are basically asking two questions:

1. can we reach l ?
2. If we can reach l , is the mean time for reaching l finite?

Let's start with the first problem. For each sample path starting at $X(0) = x \in (a, b)$ the hitting time T_a is a monotonically non-increasing function of a so we can define

$$T_{l^+} = \lim_{a \searrow l} T_a < \infty.$$

Proposition 2.3.1

We have that

$$T_{l^+} = T_l.$$

Proof

- When $X(0) = x \in (a, b)$ then $T_a \leq T_l$. This means that

$$T_{l^+} = \lim_{a \searrow l} T_a < T_l \implies T_{l^+} \leq T_l.$$

- If $T_{l^+} = \infty$ then $T_l = T_{l^+} = \infty$.

- If $T_{l^+} < \infty$ then

$$\begin{aligned} X(T_{l^+}) &= \lim_{a \searrow l} X(T_a) \\ &= \lim_{a \searrow l} a = l > -\infty \end{aligned}$$

so

$$T_{l^+} \geq T_l = \inf\{t \geq 0 | X(t) = l\} \implies T_{l^+} \geq T_l.$$

So $T_{l^+} \geq T_l$ and $T_{l^+} \leq T_l$. This means that

$$T_{l^+} = T_l.$$

□

Lemma 2.3.1

1. If $S(l, x_0] < \infty$ for some $x_0 \in (l, r)$ then

$$\mathbb{P}(T_{l^+} \leq T_b | X(0) = x) > 0 \quad \forall l < x < b < r.$$

2. If $S(l, x_0] = \infty$ for some $x_0 \in (l, r)$ then

$$\mathbb{P}(T_{l^+} \leq T_b | X(0) = x) = 0 \quad \forall l < x < b < r.$$

Definition 2.3.9

A boundary is **attracting** if $S(l, x] < \infty$ independently of $x \in (l, r)$.

This is our first kind of identification of the boundary. For example, consider:

1. a standard Brownian motion: here $l = -\infty$. This boundary is not attracting because

$$S[a, x] = x - a \implies S(-\infty, x] = \infty;$$

2. Brownian motion with drift $\mu = -\alpha < 0$:

$$S[a, x] = e^{2\alpha x} - e^{2\alpha a} \xrightarrow[a \rightarrow -\infty]{} e^{2\alpha x} < \infty$$

so $l = -\infty$ is attracting for Brownian motion with negative drift... no shit.

Let's now turn to the second problem. When is a boundary attainable in a finite expected time? Let us consider an attracting boundary l . We know that

$$\begin{aligned} &\rightarrow S(l, x) < \infty \quad \forall x \in (l, r) \\ &\rightarrow \mathbb{P}(T_{l^+} \leq T_b | X(0) = x) > 0 \quad \forall l < x < b < r. \end{aligned}$$

We want to study

$$\lim_{a \searrow l} \mathbb{E}_x [T_a \wedge T_b] \quad x < b < r, \quad b \text{ fixed.}$$

Here \mathbb{E}_x only means that we are starting to evaluate the expectation from x . Recall that

$$v_{a,b}(x) = \mathbb{E}_x [T_a \wedge T_b].$$

We are not interested in the value of the limit but we only want to know whether it is finite or not.

$$\lim_{a \searrow l} \mathbb{E}_x [T_a \wedge T_b] = \lim_{a \searrow l} 2 \underbrace{\frac{S[a, x]}{S[a, b]}}_{\substack{\text{finite}}} \int_x^b \underbrace{S[\xi, b] dM(\xi)}_{\substack{\text{finite}}} + \quad (a)$$

$$+ \lim_{a \searrow l} 2 \underbrace{\frac{S[x, b]}{S[a, b]}}_{\substack{\text{finite}}} \int_a^x S[a, \xi] dM(\xi). \quad (b)$$

We know that l is attracting:

$$\begin{cases} \lim_{a \searrow l} \frac{S[a, x]}{S[a, b]} & \text{finite positive} \implies (a) < \infty \\ \lim_{a \searrow l} \frac{S[x, b]}{S[a, b]} & \text{finite positive.} \end{cases}$$

So everything depends from $\int_a^x S[a, \xi] dM(\xi)$! We must check whether this quantity is finite or infinite. We know that

$$\lim_{a \searrow l} \mathbb{E}_x [T_a \wedge T_b] < \infty \iff \lim_{a \searrow l} \underbrace{\int_a^x S[a, \xi] dM(\xi)}_{:= \Sigma(l)}.$$

We define $\Sigma(l)$ as:

$$\begin{aligned} \Sigma(l) &:= \lim_{a \searrow l} \int_a^x S[a, \xi] dM(\xi) \\ &= \int_l^x S[l, \xi] dM(\xi) \\ &= \int_l^x \left(\int_l^\xi s(\eta) d\eta \right) m(\xi) d\xi \quad \text{change integration order} \\ &= \int_l^x \left(\int_\eta^x m(\xi) d\xi \right) s(\eta) d\eta \\ &= \int_l^x M[\eta, x] dS(\eta). \end{aligned}$$

Definition 2.3.10

The boundary l is:

- **attainable** if $\Sigma(l) < \infty$;
- **unattainable** if $\Sigma(l) = \infty$.

This is our second classification. We can prove that $\Sigma(l) < \infty \implies S[l, x] < \infty$ so if l is attainable then it is also attracting.

Consider a Brownian motion with drift $\mu < 0$. In this case $l = -\infty$ is attracting but unattainable. On the other hand, if the Brownian motion has absorbing boundary in 0 we have that $l = 0$ is attracting and attainable:

$$\mathbb{E}[T_a \wedge T_b] = x(b-x) < \infty \quad 0 < x < b.$$

Remember that $\mathbb{E}[T_a | X(0) = x] = \infty$!

Lemma 2.3.2

Let l be an attracting boundary and $l < x < b < r$. The following are equivalent:

- i) $\mathbb{P}(T_l < \infty | X(0)) > 0$;
- ii) $\mathbb{E}[T_l \wedge T_b | X(0) = x] < \infty$;
- iii) $\Sigma(l) = \int_l^x S(l, \eta) dM(\eta) < \infty$.

Remember that

$$\Sigma(l) = \int_l^x S(l, \xi) dM(\xi) = \int_l^x M[\eta, x] d(\eta)$$

measures the time that the process takes to reach l or another state b starting from an interior point $x < b$. Really?

Definition 2.3.11

Recall that $M(l, x) = \lim_{a \searrow l} M[a, x]$. We define

$$N(l) = \int_l^x S[\eta, x] dM(x) = \int_l^x M(l, \xi) dS(\xi).$$

Here:

- $M(l, x]$ measures the speed of the process near l ;
- $N(l)$ measures the time the process takes to reach an interior point $X \in (l, r)$ starting from the boundary l .

What we get from this is that the boundary classification depends entirely from $S(l, x]$, $M(l, x]$, $\Sigma(l)$ and $N(l)$.

Criteria				Terminology				
$S(l, x]$	$M(l, x]$	$\Sigma(l)$	$N(l)$	Feller	Gikhman and Skorokhod			
< ∞	< ∞	< ∞	< ∞	Regular				
< ∞	= ∞	< ∞	= ∞	Exit-Trap-Absorbing				
< ∞	= ∞	= ∞	= ∞	Natural				
= ∞	< ∞	= ∞	= ∞	$(\Sigma(l) = \infty, N(l) = \infty)$				
= ∞	= ∞	= ∞	= ∞	Natural				
= ∞	< ∞	= ∞	< ∞	$(S(l, x] = \infty$				
				Entrance				
					Attracting			
					Attainable			
					Unattainable			

Table 2.1: Do you know how much time it took me to make this?

We have the following cases.

a) *Regular boundary*: the process can enter and leave from the boundary. We can add the behavior at the boundary:

- $M[\{l\}] = \infty$: absorbing.
- $M[\{l\}] = 0$: reflecting.
- $0 < M[\{l\}] < \infty$: sticky boundary.

Proposition 2.3.2

If $S(l, x] < \infty, M(l, x] < \infty \implies l$ is regular.

b) *Exit boundary*: if the process starts in l then it cannot reach any interior state b . We can reach l but we cannot leave it.

Proposition 2.3.3

If $\Sigma(l) < \infty$ (which means l is attainable), $M(l, x] = \infty \implies l$ is exit.

c) *Entrance boundary*: we cannot enter l , but if we start from it we can exit.

Proposition 2.3.4

If $S(l, x] = \infty$ (which means l is not attractive), $N(l) < \infty \implies l$ is entrance.

d) *Natural/Feller boundary*: the process can neither reach the boundary in finite mean time nor can be started from the boundary.

Proposition 2.3.5

If $\Sigma(l) = \infty, N(l) = \infty \implies l$ is natural.

Remark

If both boundaries are exit then the process is transient and there not exists a stationary distribution. If both boundaries are entrance then there exists a stationary distribution.

We consider some examples.

- **Standard Brownian motion.** Here

$$\begin{aligned} I &= (l, r) = (-\infty, +\infty) \\ s(\xi) &= 1 \\ m(\xi) &= 1 \\ \Sigma(l) &= \int_l^x \left(\int_l^\xi s(\eta) d\eta \right) m(\xi) d\xi = \int_l^x (\xi - l) d\xi = \frac{\xi^2}{2} - l\xi \Big|_l^x. \end{aligned}$$

So we have

$$\begin{aligned} \Sigma(-\infty) &= \infty \\ N(l) &= \int_l^x M(l, \xi) dS(\xi) = \infty \text{ (if } l = \infty) \implies l = \text{natural boundary} \\ \Sigma(r) &= N(r) = \infty \text{ if } r = \infty \implies r = \text{natural boundary}. \end{aligned}$$

- **Ornstein-Uhlenbeck process.** Here $I = (l, r) = (-\infty, +\infty)$.

$$\begin{aligned} \mu(x) = -x &\implies s(x) = e^{x^2} \\ \sigma^2(x) = 1 &\implies m(x) = e^{-x^2} \implies \Sigma(\infty) = \dots = \infty \\ &\quad N(\infty) = \dots = \infty. \end{aligned}$$

So $l = \infty$ is natural. It is also possible to prove that $l = -\infty$ is natural.

- **Bessel process.** Here $I = [0, \infty)$.

$$\begin{aligned} \mu(x) = \frac{\alpha-1}{2^x} &\implies s(x) = x^{1-\alpha} \\ \sigma^2(x) = 1 &\implies m(x) = x^{\alpha-1} \implies \Sigma(\infty) = \begin{cases} < \infty & \text{if } \alpha < 2 \\ \infty & \text{if } \alpha \geq 2 \end{cases} \\ &\quad N(\infty) = \begin{cases} < \infty & \text{if } \alpha < 2 \\ \infty & \text{if } \alpha = 0. \end{cases} \end{aligned}$$

The boundary 0 is:

$$\begin{cases} \text{entrance} & \text{if } \alpha \geq 2 \\ \text{regular} & \text{if } 0 < \alpha < 2 \\ \text{exit} & \text{if } \alpha = 0. \end{cases}$$

The boundary ∞ is always regular.

Chapter 3

Simulation of stochastic processes

This part is kept by the visiting professor Samuel Hermann. Imagine we have the following problems:

- estimating the average profit of an investment strategy;
- assessing the viability of a complex system of productions;
- calculating when an insurance company will (rightfully) go bankrupt;
- calculating the average (= integrated) value of a function.

In general, we want to estimate the average of a probability. We only care about that because by computing averages we can also compute probabilities! Remember that

$$\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A].$$

To estimate the average be basically make use of the law of large numbers:

$$\mathbb{E}[A] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n A_k \quad \text{for a sequence of i.i.d. } A_k \sim A.$$

To control the error, instead, we will use the central limit theorem. The main problem, for now, is generating a sequence of independent random variables with a given distribution. Consider the following examples:

- If we look at a system of production and we know that every small component of an object has a lifetime then we can represent the object as an electrical circuit.

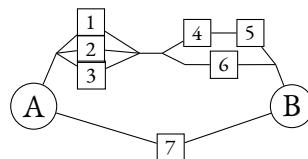


Figure 3.1: Yeah I spent a non negligible amount of time on this and it also looks like shit.

If component 7 is dead then the object is still working, but if 7, 4 and 6 are dead then the object is still not working. To estimate the lifetime of the object, we need to evaluate all the cases in which A and B are still connected. If T_n is the lifetime of component n , we have

$$\text{Lifetime: } \mathbb{E}[\max(T_7), \min(\max(T_1, T_2, T_3), \max(T_6, \min(T_4, T_5)))] .$$

This is not very fun to compute: it is a 7-dimensional integral with min and max functions: it is way better to simulate it. Here A_k is linked to a random vector of finite dimension

$$(T_1, T_2, \dots, T_7) \in \mathbb{R}^7$$

- Probability of default of an insurance company. The surplus of a company is

$$X_t = x + ct - \sum_{k=1}^{N_t} \xi_k$$

where x is the value of the company at time 0, ct is the premium income (which is a linear quantity that grows with the time), ξ_k is the value of the k -th claim (they are all i.i.d.) and N_t is the number of shocks at time t (we can model it as a counting Poisson Process).

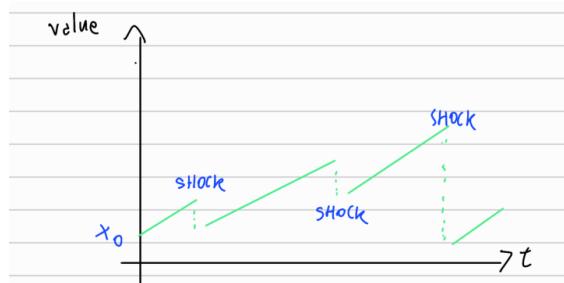


Figure 3.2: I know. Lines.

N_t is Poisson so the time between shocks is exponentially distributed.

Probability of failure: $\mathbb{P}(\inf\{t \geq 0 : X_t \leq 0\} \leq T)$.

To evaluate this probability we need

$$\left(\begin{array}{c} T_1 \\ \text{Time of 1st shock} \end{array}, \begin{array}{c} \xi_1 \\ \text{Value of claim of 1st shock} \end{array} \right), (T_2, \xi_2), \dots$$

Here the number of shocks is random, so the dimension of the vector is random here!

- Calculate the price of a financial product.

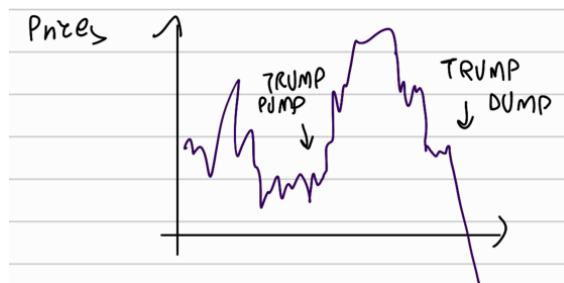


Figure 3.3: This is getting too political.

Here the price is

$$\mathbb{E}[F(X_t : t \in [0, T])]$$

where

\mathbb{E} : average gain of the contract

F : some functional

$[0, T]$: time left from today to the expiry of the contract.

Here we need to generate full trajectories of infinite point, which means that our random vector is infinite dimensional! Of course we won't be able to simulate this but we will need to approximate.

3.1 Generation of random variables

3.1.1 Uniform random variables

We want to generate a sequence of values corresponding to the value of i.i.d. random variables with uniform distribution. We want this generation to be reproducible.

When we ask a computer to generate a random number of course that number is not “really random” but it is pseudo-random: the sequence of numbers in output is deterministic, but it is “hard enough” to tell apart from an actually random stream of numbers. The most basic random number generator is the **congruence generator**. It is obtained as follows:

Congruence generator

Let (y_n) be a sequence $\in \{0, 1, \dots, N - 1\}^{\mathbb{N}}$ defined by

$$y_{n+1} = a \cdot y_n + b \pmod{N}$$

where a is called **multiplier** and b is called increment. The starting value y_0 is called the **seed**.

If I want to have numbers between 0 and 1, I take the sequence and divide it by N :

$$x_n = \frac{y_n}{N} \in [0, 1].$$

What is the probability distribution for a, b, N ? We want to choose these parameters so that the sequence looks random. There are tests for checking whether the sequence is uniformly distributed and whether it is random (like the DieHard test).

Definition 3.1.1

The period of the sequence is defined as

$$T : x_{n+T} = x_n$$

and it is $\leq N$.

Ideally we want the period to be as large as possible. We can now see some examples of congruence generators.

1

$$\left. \begin{array}{l} a = 7^5 = 16807 \\ b = 0 \\ N = 2^{31} - 1 = 2147483641 \end{array} \right\} \text{period: } 2^{31} - 2.$$

② Fortran IBM:

$$\left. \begin{array}{l} a = 2^{16} + 3 \\ N = 2^{32} \end{array} \right\} \text{period: } 2^{30}.$$

③ Scilab:

$$\left. \begin{array}{l} a = 2843314861 \\ N = 2^{31} \end{array} \right\} \text{period: } 2^{30}.$$

④ Mersenne Twister. This was made by Matsumoto and Nishimura in 1997 and it is the Phyton  random generator. It has a period of $2^{19937} - 1$ but it is not a simple congruence generator.

It is hard to tell whether a generator is good or not. The Fortran one was used for a long time before discovering that in 3 dimension, if you line up your data in a certain way, there is clearly a correlation between the random instances. We can use the χ^2 test to check whether the instances are actually independent.

3.1.2 Classical discrete random variables

This is the **general setting**. Let X be a random variable with values in

$$\{X_1, \dots, X_n, \dots\} = \{X : i \in I\} \text{ and } \mathbb{P}(X = x_i) = p_i.$$

The idea is to divide the interval $[0, 1]$ in many sub intervals accordingly to the various probabilities:

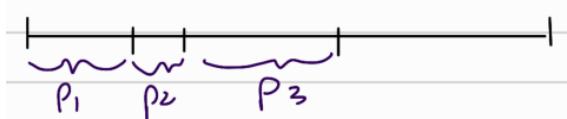


Figure 3.4: They can have different lengths.

The extract a uniform random variable. If it falls in the first interval (which happens with probability p_1 since we defined the interval accordingly) then give to that uniform random variable the value of X_1 .

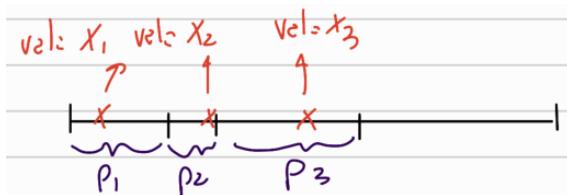


Figure 3.5: Imagine that if we had only 2 probabilities $p_1 = 0.7$, $p_2 = 0.3$ and we splitted the interval so that p_1 covers 70% of the length and p_1 the remainig 30% we will fall in p_1 70% of the times and in p_2 30% of the times.

So our after extracting $U \sim U(0, 1)$ our actual extracted value X' will be

$$X' = x_1 \mathbb{1}_{\{U \leq p_1\}} + x_2 \mathbb{1}_{\{p_1 \leq U \leq p_2\}} + \dots + x_n \mathbb{1}_{\{p_1 + \dots + p_{n-1} \leq U \leq p_1 + \dots + p_n\}}$$

Proof

$$\mathbb{P}(X' = x_n) = \mathbb{P}(p_1 + \dots + p_{n-1} \leq U < p_1 + \dots + p_n).$$

□

Remark

We need only one random variable U to generate one value of X .

In particular, we can exploit this to generate random variables from many discrete distributions.

① **Bernoulli distribution $\text{Be}(p)$** . We take

$$X' = \mathbb{1}_{\{U \leq p\}}.$$

② **Binomial distribution $\text{Bin}(n, p)$** . We use the general setting and consider X' as the sum of independent Bernoulli random variables.

$$X' = \sum_{k=1}^n \mathbb{1}_{\{U_k \leq p\}} \sim \text{Bin}(n, p).$$

③ **Geometric distribution $\text{Geom}(p)$** . We use the general setting and we take

$$\begin{aligned} X' &= \mathbb{1}_{\{U \leq p\}} + \sum_{i \geq 1}^n i \mathbb{1}_{\{(1-p)^i \leq 1-U < (1-p)^{i-1}\}} \\ &\Downarrow \\ X' &= 1 + \lfloor -\log\left(\frac{\text{rand}()}{-\log(1-p)}\right) \rfloor. \end{aligned}$$

3.1.3 Generator of continuous random variables

① **Gaussian random variables**. We use the Box-Müller method.

Box-Müller transform

Let $U \sim \text{U}(0, 1)$ and $V \sim \text{Exp}(1)$ with $U \perp\!\!\!\perp V$. We set

$$\begin{cases} X = \sqrt{2V} \cos(2\pi U) \\ Y = \sqrt{2V} \sin(2\pi U). \end{cases}$$

Then $X \perp\!\!\!\perp Y$ and $X \sim Y \sim \mathcal{N}(0, 1)$.

Proof

Let f be a Borel non-negative function. Then

$$\begin{aligned} \mathbb{E}[f(X, Y)] &= \mathbb{E}\left[f\left(\sqrt{2V} \cos(2\pi U), \sqrt{2V} \sin(2\pi U)\right)\right] \\ &= \int_0^\infty \int_0^1 f\left(\sqrt{2v} \cos(2\pi u), \sqrt{2v} \sin(2\pi u)\right) e^{-v} dv du. \end{aligned}$$

exponential density

To perform a change of variable form (U, V) to (X, Y) we need the determinant of the Jacobian matrix

$$\begin{bmatrix} \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \end{bmatrix} = \begin{bmatrix} \frac{x}{2v} & \frac{y}{2v} \\ -2\pi y & 2\pi x \end{bmatrix}$$

that gives us

$$\det = \frac{\pi(x^2 + y^2)}{v} = 2\pi$$

so we have

$$\mathbb{E}[f(X, Y)] = \int \int_{\mathbb{R}^2} \frac{1}{2\pi} f(x, y) e^{-\frac{x^2+y^2}{2}} dx dy$$

which is the density function of a bivariate normal $\mathcal{N}(0, 1)$. □

If $U' \sim U(0, 1)$ then $-\frac{1}{\lambda} \ln(U') \sim \text{Exp}(\lambda)$ and

$$\mathbb{P}\left(-\frac{1}{\lambda} \ln(U') > t\right) = \mathbb{P}(U' \leq e^{-\lambda t}) = e^{-\lambda t} \quad \forall t > 0.$$

Then we can use Box-Müller and get

$$\begin{cases} X = \sqrt{-2 \ln(U')} \cos(2\pi U) \\ Y = \sqrt{-2 \ln(U')} \sin(2\pi U). \end{cases}$$

In this way the whole transform is expressed in terms of uniform random variables, without needing to sample from an exponential distribution. We can also construct n -dimensional Gaussian vectors (X_1, \dots, X_d) by 2.

② Inverse transformation sampling. We aim to generate a random variable X with probability density function f and cumulative distribution function F with:

- F is increasing and right-continuous;
- $F(-\infty) = 0$ if X is finite;
- $F(x) - F(x^-) = \mathbb{P}(X = x)$;
- F has only a countable number of discontinuities.

For $u \in (0, 1)$ we can construct the inverse function (quantile function):

$$F^{-1}(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\}$$

Proposition 3.1.1

F^{-1} is increasing and left continuous. it satisfies:

$$\forall x \in \mathbb{R}, \forall u \in (0, 1) : F(x) \geq u \iff x \geq F^{-1}.$$

Proof

By definition of $F(x) \geq u$ we know that $x \geq F^{-1}(u)$ so

$$x + \varepsilon > F^{-1}(u) \implies F(x + \varepsilon) > u \quad \forall \varepsilon > 0.$$

Then we take the limit and we use right continuity:

$$\lim_{\varepsilon \rightarrow 0} F(x + \varepsilon) = F(x) \geq u.$$

□

Proposition 3.1.2

X and $F^{-1}(U)$ with $U \sim U([0, 1])$ are identically distributed.

Proof

$$\begin{aligned} \mathbb{P}(F^{-1}(U) \geq X) &= \mathbb{P}(U \leq F(x)) \\ &= F(x). \end{aligned}$$

□

Let's see how to apply this method in practice.

- $X \sim \text{Exp}(\lambda) \rightarrow F(x) = 1 - e^{-\lambda x}$. We get that

$$F(x) = u = 1 - e^{-\lambda x} \iff x = -\frac{1}{\lambda} \ln(1 - u) = F^{-1}(u)$$

so

$$-\frac{1}{\lambda} \left(\underbrace{1 - U}_{\sim U([0,1])} \right) \sim \text{Exp}(\lambda).$$

- $X \sim \text{Cauchy}(a)$. The probability density function is $f_X(x) = \frac{a}{\pi(x^2 + a^2)}$. This means

$$F(x) = \frac{1}{\pi} \arctan\left(\frac{x}{a}\right) + \frac{1}{2} = u \iff a \cdot \tan\left(\pi\left(u - \frac{1}{2}\right)\right) = F^{-1}(u)$$

so

$$a \cdot \tan\left(\pi\left(U - \frac{1}{2}\right)\right) \sim \text{Cauchy}(a).$$

- As an exercise, apply this to the arcsin law:

$$\begin{aligned} f_x(t) &= \frac{1}{\pi\sqrt{t(1-t)}} \\ F(t) &= \frac{2}{\pi} \arcsin\left(\sqrt{t}\right). \end{aligned}$$

In many cases, though, we don't have F^{-1} explicitly. We need a way to extend this theorem.

- ③ **Acceptance-Rejection Method (“exact method”).** We aim to generate a random variable X with density f . We assume we are able to generate Y with density g such that g dominates f up to a constant c :

$$\exists c > 0 | f(x) \leq cg(x) \forall x \in \mathbb{R}.$$

So we use Y to generate X .

Acceptance-Rejection

- (a) Generate Y with probability density function g and $U \sim U([0, 1])$ so that $U \perp\!\!\!\perp Y$.
- (b) Test whether

$$U \leq \frac{f(y)}{cg(y)} = h(y).$$

If yes, set $X = y$; if not, go to step 1.

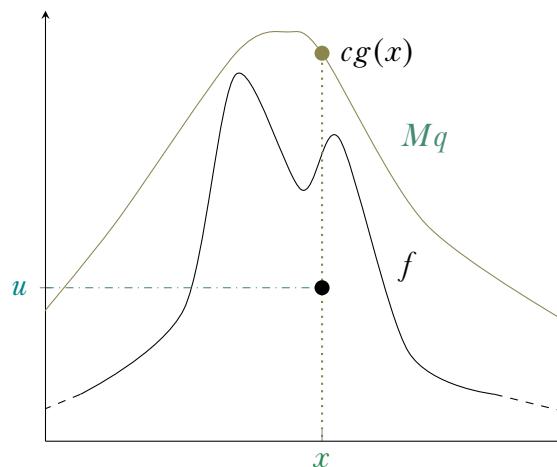


Figure 3.6: I had made this drawing for a dead course in statistics for stochastic processes

How many loops should we observe?

Remark

The number of iterations N is geometrically distributed.

For each random variable X we need 2 random variables, Y and U . This may create problems with large computations.

Proof

We need k iterations:

$$\begin{aligned} (Y_1, U_1) &\quad \text{first iteration} \\ &\vdots \\ (Y_k, U_k) &\quad k - \text{th iteration.} \end{aligned}$$

Let φ be a Borel non-negative function. We have that

$$\begin{aligned}\mathbb{E} [\varphi(X)] &= \mathbb{E} [\varphi(Y_1) \mathbb{1}_{\{U_1 \leq h(y)\}}] + \\ &\quad + \sum_{k=1}^{\infty} \mathbb{E} [\varphi(Y_{k+1}) \mathbb{1}_{\{U_1 > h(Y_1), \dots, U_k > h(Y_k), U_{k+1} \leq h(Y_{k+1})\}}]\end{aligned}$$

Every time we start a new loop we get independent random variables so we can factorize

$$\begin{aligned}\mathbb{E} [\varphi(X)] &= \mathbb{E} [\varphi(Y_1) \mathbb{1}_{\{U_1 \leq h(y)\}}] + \\ &\quad + \underbrace{\sum_{k=1}^{\infty} \mathbb{E} [\varphi(Y_{k+1}) \mathbb{1}_{\{U_{k+1} \leq h(Y_{k+1})\}}]}_{\text{Geometric series}} \mathbb{P}(U_1 > h(Y_1))^k\end{aligned}$$

$(U_1, Y_1) \sim (U_2, Y_2) \sim \dots$ so the expectation becomes

$$\begin{aligned}\mathbb{E} [\varphi(Y_1) \mathbb{1}_{\{U_1 \leq h(Y_1)\}}] \frac{1}{\mathbb{P}(U_1 \leq h(Y_1))} &= \mathbb{E} [\varphi(Y_1) | U_1 \leq h(Y_1)] \\ &= \frac{1}{c_0} \int_{\mathbb{R}} \int_0^1 \varphi(y) \mathbb{1}_{\{u \leq h(y)\}} g(u) du dy \\ &\stackrel{\text{use Fubini}}{=} \frac{1}{c_0} \int_{\mathbb{R}} \varphi(y) \frac{f(y)}{cg(y)} g(y) dy \\ &= \frac{1}{cc_0} \int_{\mathbb{R}} \varphi(y) f(y) dy\end{aligned}$$

Taking $\varphi = 1$ we get $cc_0 = 1 \implies c_0 = \frac{1}{c}$. So the probability distribution function of X is

$$N \sim \text{Geom} \left(\frac{1}{c} \right) \implies \mathbb{E}[N] = c.$$

□

Notice that c should be as small as possible. For example, consider

$$f(x) = \frac{2}{\sqrt{\pi}} \sqrt{x} e^{-x} \mathbb{1}_{\{x \geq 0\}} \sim \Gamma(3/2, 1).$$

We can set

$$g(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}.$$

We need to find λ and $c > 0$ such that

$$f(x) \leq cg(x) \forall x \in \mathbb{R}.$$

Do it as an exercise.

④ **Convergent series method.** We assume that

$$f(x) = \sum_{k \geq 0} f_k(x), \quad \forall x \in \mathbb{R}$$

and define

$$S_n(x) = \sum_{k=0}^n f_k(x)$$

and the remainder R_{n+1} (the tail of the series) as

$$R_{n+1} = \sum_{k=n+1}^{\infty} f_k(x).$$

If the remainder tends to 0 and it is known (or an upper bound of the remainder is known) we can use the algorithm.

Convergent series method

(a) Generate Y with density g and $U \sim U([0, 1]), Y \perp U$.

(b) Choose the smallest n such that

$$|Ucg(Y) - S_n(Y)| \leq R_{n+1}(Y).$$

(c) If $Ucg(Y) \leq S_n$ then $X = Y$ otherwise go to step 1.

This makes sense if we think of S_n like an approximation of f . The algorithm stops when $f(y) = Ucg(y)$ but this happens with probability 0.

We are very interested in knowing how to approximate integrals. After all expectations are integrals! Imagine we are working with an integral

$$I = \int_U g(x) dx \quad U \subset \mathbb{R}^d.$$

For example imagine calculating the expected price of the S&P 500: that would be an integral of dimension 500... is better to approximate this than to compute explicitly. Think about this: we have to approximate

$$\int_a^b g(x) dx.$$

If we know that $c \geq g(x) \geq 0$ then we have a “rectangle” that we can throw points into and see how many of these fall below the function. The proportion between those points and the total number of points should give us the integral.

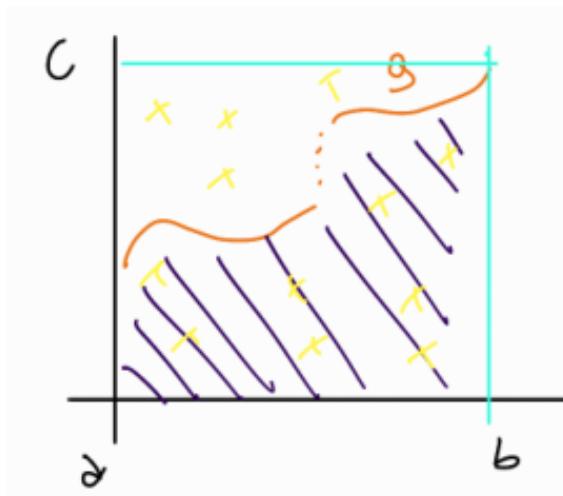


Figure 3.7: Under a starry sky.

This is called the “**hit or miss**” method and we will analyze it in the next section.

3.2 Monte Carlo Methods

Randomized algorithms are basically divided in two families: **Monte Carlo** and **Las Vegas** algorithms.

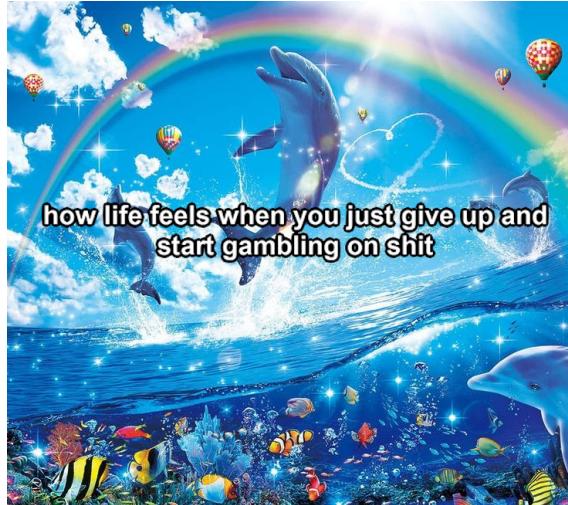


Figure 3.8: Please bro just one more simulation please bro now the ergodic average will be good please I swear just one more run it's only 3000 nested loops to generate a gazillion-dimensional vector please bro

Las Vegas algorithms always give the true result, without approximation. An example is the so called “ n -queens problem” where we have to place n queens in a $n \times n$ checkerboard so that no two queens attack each other. So we place the first queen, we choose a *random* position that does not conflict with the previous queens and if no legal moves are available then backtrack or restart. It may take a lot of time, but the results will always be correct.

X	X	X	M	X	X	X	X
			X	T	T		
	X		X		T		
		X				T	
			X				T
				X			
					X		
						X	

Figure 3.9: Now we playing games cuh.

Monte Carlo methods, instead, give a random value close to a deterministic unknown. If

the aim is computing the value of an integral

$$I = \int_U g(x) dx \quad \text{for } U \subseteq \mathbb{R}^d$$

we can approximate them with deterministic methods: for example consider the “Riemann rectangle” method or “trapezoidal method”. This has some problems:

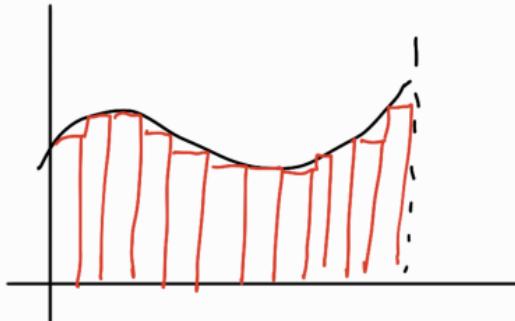


Figure 3.10: Riemann “rettangoli balenghi” method.

- it requires some sort of regularity of the function;
- for high dimension it takes a shitload of time.

3.2.1 “Hit or miss” method

This is easy to generalize for more than one dimension, but we analyze the case for $d = 1$. We have

$$I = \int_a^b g(x) dx \quad \text{with } -\infty < a < b < +\infty.$$

This is valid if $\exists c > 0$ s.t. $0 \leq g(x) \leq c$ for $\forall x \in [a, b]$. We do not need regularity! Introduce

$$\begin{aligned} \Omega &= [a, b] \times [0, c] \\ S &= \{(x, y) \in \Omega : y \leq g(x)\}. \end{aligned}$$

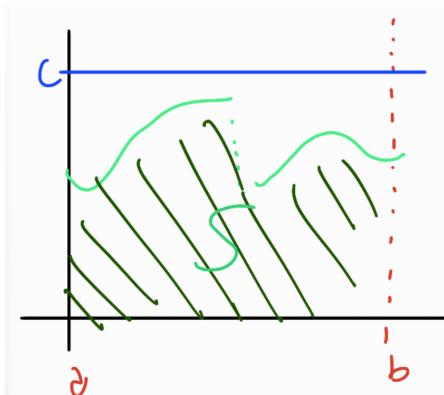


Figure 3.11: S is a subspace of Ω .

Let (X, Y) be a random couple uniformly distributed on Ω :

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{c(b-a)} & \text{if } (x, y) \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

Define p as

$$\begin{aligned} p &= \mathbb{P}((X, Y) \in S) \\ &= \int \int \mathbb{1}_S(x, y) f_{(X,Y)}(x, y) dx dy \\ &= \frac{1}{c(b-a)} \int_a^b \left(\int_0^{g(x)} dy \right) dx \\ &= \frac{1}{c(b-a)} \underbrace{\int_a^b g(x) dx}_I \end{aligned}$$

so

$$I = \underbrace{c(b-a)}_{\text{area of } \Omega} \cdot \underbrace{p}_{\frac{\text{area of } S}{\text{area of } \Omega}}.$$

We can approximate the value of I using asymptotic empirical frequency. Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be n independent copies of (X, Y) . We have that N_n^S , the number of couples belonging to S , is a sum of Bernoulli random variables:

$$N_n^S = \sum_n^{k=1} \mathbb{1}_{\{(X_k, Y_k) \in S\}} \sim \text{Bin}(n, p)$$

so by the LLN we have

$$\frac{N_n^S}{n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} p.$$

We want to get

$$\vartheta_1 := \frac{c(b-a)N_n^S}{n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} I$$

so that ϑ_1 is an estimator of I . It is a non biased estimator because

$$\mathbb{E}[\vartheta_1] = \frac{c(b-a)\cancel{n}p}{\cancel{n}} = I.$$

We are basically counting the number of points under the function.

We now want to compute the confidence interval.

Theorem 3.2.1

Moire-Laplace theorem. If $X \sim \text{Bin}(n, p)$ then

$$\mathbb{P}\left(\lambda_1 \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \lambda_2\right) \xrightarrow{n \rightarrow \infty} \int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

This is the baby-ahh version of the central limit theorem. Let us consider a confidence level $1 - \alpha$. Then we denote z_α the quantile of the Gaussian distribution such that

$$\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$$

where $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$. We have that with probability $1 - \alpha$

$$\frac{X - np}{\sqrt{np(1-p)}} \in [-z_\alpha, z_\alpha].$$

Applying this to ϑ_1 we have

$$I \in \left[\vartheta_1 - \frac{z_\alpha}{\sqrt{n}} \sqrt{p(1-p)} c(b-a); \vartheta_1 + \frac{z_\alpha}{\sqrt{n}} \sqrt{p(1-p)} c(b-a) \right].$$

In practice we replace p with $\frac{N_n^S}{n}$ since p is unknown. The rate of convergence is $\frac{1}{\sqrt{n}} \cdot \text{a constant}$ while the deterministic method has a convergence rate of $\mathcal{O}\left(\frac{\varepsilon''}{n^2}\right)$. The hit or miss method doesn't explode with $n!$ Now we want to compute the variance of this estimator:

$$\begin{aligned} \text{Var}(\vartheta_1) &= \frac{c^2(b-a)^2}{n^2} np(1-p) \\ &= \frac{I(c(b-a) - I)}{n}. \end{aligned}$$

In order to increase the precision we need to minimize the numerator. This is not always explicitly

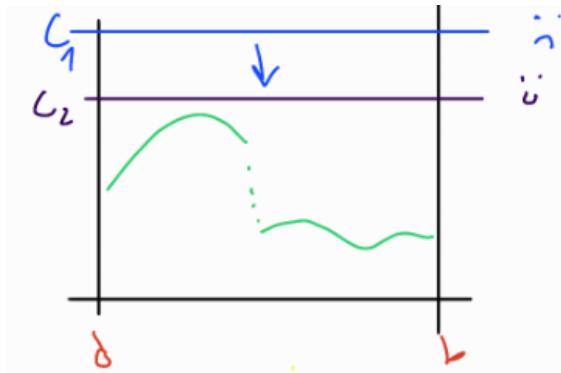


Figure 3.12: Of course...

computable. We could also think about splitting the function so that we can use different more fitting values of c . In general, the hit or miss method is based on a sequence of n experiments

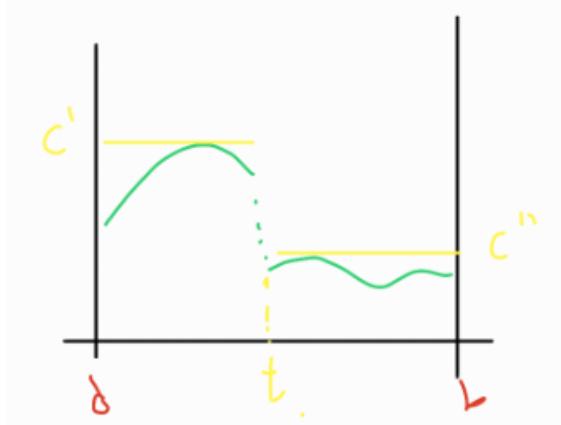


Figure 3.13: Just like this!

and each experiment concerns 1 point but there are variations to this:

- if one experiment is to consider k points which should be under the curve we have

$$\begin{aligned} p &= \mathbb{P}((X_1, Y_1) \in S, (X_2, Y_2) \in S, \dots, (X_k, Y_k) \in S) \\ &= \mathbb{P}((X_1, Y_1) \in S)^k \\ &= \frac{\left(\int_a^b g(x) dx\right)^k}{c^k(b-a)^k}. \end{aligned}$$

Using the asymptotic frequency we are able to give a confidence interval for I^k .

- If one experiment consists in choosing a Poisson distributed number of points uniformly distributed on Ω we can compute the probability that no points are below the curve g .

$$\begin{aligned} p &= \sum_{k=0}^{\infty} \mathbb{P}(K = k, (X_1, Y_1) \notin S, \dots, (X_k, Y_k) \notin S) \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \left(1 - \frac{\int_a^b g(x) dx}{c(b-a)}\right)^k \\ &= \exp \left\{ \lambda - \frac{\lambda \int_a^b g(x) dx}{c(b-a)} - \lambda \right\} \\ &= \exp \left\{ -\frac{\lambda}{c(b-a)} \int_a^b g(x) dx \right\} \\ &= \exp \left\{ \int_a^b g(x) dx \right\} \quad \text{if } \lambda = c(b-a) \end{aligned}$$

so now I can use asymptotic frequency to estimate e^{-I} . This is cool since many processes are in the form of $e^{-\text{something}}$.

3.2.2 Sample mean method

Up to now, we have linked the integral to probability and just estimated the latter. But we may also consider the integral not as a probability but as an expectation. In this case we have

$$I \int_a^b g(x) dx \quad \infty \leq a < b \leq \infty$$

so there are no boundaries and it can be over ∞ ! The idea is to write I as an expectation:

$$I = \mathbb{E} \left[\frac{g(x)}{f_X(x)} \right]$$

where X is a random variable with probability distribution function f_X . There are conditions to write I as an expectation: we need that $f_X(x) = 0 \rightarrow g(x) = 0$ and that $\sup(g) \subset \sup(f_X) \in [a, b]$. Let us consider X_1, X_2, \dots, X_n with n random variables with probability density function f_X . Then

$$\vartheta_2 = \frac{1}{n} \sum_{k=1}^n \frac{g(x_k)}{f_X(x_k)}$$

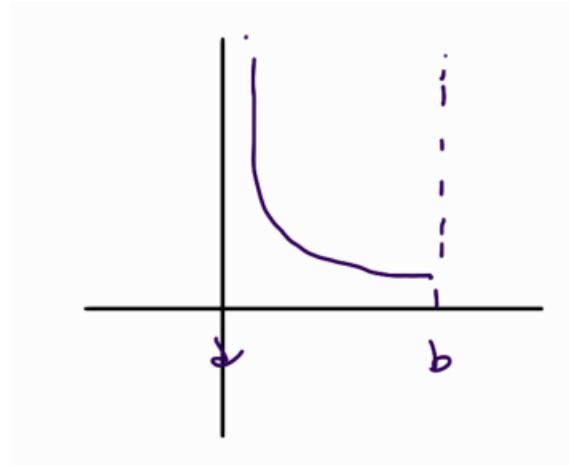


Figure 3.14: Here there is no c dominating g .

is an estimator of I ; more precisely, if $\mathbb{E} \left[\left| \frac{g(x)}{f_X(x)} \right| \right] < \infty$ then by LLN

$$\vartheta_2 \xrightarrow[n \rightarrow \infty]{\text{a.s.}} I.$$

We need to find the confidence interval. Using the CLT we have

$$\mathbb{E} \left[\left(\frac{g(x)}{f_X(x)} \right)^2 \right] < \infty \implies \sqrt{n} \frac{(\vartheta_2 - I)}{\sigma} \xrightarrow[n \rightarrow \infty]{\text{d}} N(0, 1)$$

where

$$\sigma^2 = \text{Var} \left(\frac{g(X)}{f_X(X)} \right).$$

In practice, we replace σ by the estimator S_n that comes from

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n \left(\frac{g(X_k)}{f_X(X_k)} - \vartheta_2 \right)^2.$$

By LLN

$$S_n^2 \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \sigma^2 \quad \text{and} \quad S_n \xrightarrow{\mathbb{P}} \sigma.$$

Corollary

If $\mathbb{E} \left[\frac{g(x)^2}{f_X(x)^2} \right] < \infty$ then

$$\sqrt{n} \frac{(\vartheta_2 - I)}{S_n} \xrightarrow{\text{d}} N(0, 1).$$

Proof

We use Slutsky's lemma:

$$\text{if } \begin{cases} A_n \xrightarrow{\text{d}} A \\ B_n \xrightarrow{\mathbb{P}} b \\ \text{deterministic} \end{cases} \text{ then } \frac{A_n}{B_n} \xrightarrow{\text{d}} \frac{A}{b}.$$

In our case $A = \sqrt{n} \frac{\vartheta_2 - I}{\sigma}$ and $B_n = \frac{S_n}{\sigma}$. □

About the confidence interval: with probability $1 - \alpha$, I belongs to

$$\left[\vartheta_2 - z_\alpha \frac{S_n}{\sqrt{n}}, \vartheta_2 + z_\alpha \frac{S_n}{\sqrt{n}} \right].$$

3.2.3 Variance reduction: Importance Sampling

We need to be careful: the CLT works only if the second moment is finite. For example: consider

$$\int_0^1 \frac{dx}{\sqrt{x+x^2}} = \mathbb{E} \left[\frac{1}{\sqrt{U+U^2}} \right] \quad U \sim \mathcal{U}([0,1]).$$

In this case

$$\vartheta_2 = \frac{1}{n} \sum_{k=1}^n \frac{1}{U_k + U_k^2} \quad \text{with } U_k \sim \mathcal{U}([0,1]).$$

We need to compute the second moment

$$\mathbb{E} \left[\left(\frac{1}{\sqrt{U_k + U_k^2}} \right)^2 \right] = \int_0^1 \frac{1}{x+x^2} dx = +\infty$$

and so we have infinite variance which means ∞ confidence interval which means kill yourself.
In this case we have

$$I = \int_0^1 \frac{2}{\sqrt{1+x}} \cdot \frac{1}{2\sqrt{x}} dx, \quad \text{take } f_X(x) = \frac{1}{2\sqrt{x}} \mathbf{1}_{[0,1]}(x).$$

Hence we have

$$I = \mathbb{E} \left[\frac{2}{\sqrt{1+x}} \right] \quad \text{with } X \sim f_X \sim U^2, \quad U \sim \mathcal{U}([0,1]).$$

This new expectation has finite variance and is thus safe for Monte-Carlo:

$$\mathbb{E} \left[\frac{2}{1+x} \right]^2 = 4 \int_0^1 \frac{1}{1+x} \cdot \frac{1}{2\sqrt{x}} dx < \infty$$

and we can apply the CLT for $\frac{2}{\sqrt{1+u^2}}$ with $U \sim \mathcal{U}([0,1])$. I is well approximated by

$$\vartheta_2 = \frac{1}{n} \sum_{k=1}^n \frac{1}{\sqrt{1+u_k^2}}$$

with $U_k \stackrel{i.i.d.}{\sim} \mathcal{U}([0,1])$.

To reduce the variance we can use **importance sampling**. The aim is to choose f_X such that $\mathbb{E} \left[\left(\frac{g(x)}{f_X(x)} \right)^2 \right] < \infty$ and $\text{Var}(\vartheta_2)$ is minimized.

Proposition 3.2.1

The minimal value equals

$$\left(\int_a^b |g(x)|^2 dx \right)^2 - I^2$$

and is obtained for

$$f_X(x) = \frac{|g(x)|}{\int_a^b |g(x)| dx}.$$

Proof

We have, by Cauchy-Schwartz,

$$\begin{aligned} \left(\int_a^b |g(x)|^2 dx \right)^2 &= \left(\int_a^b \frac{|g(x)|}{\sqrt{f_X(x)}} \sqrt{f_X(x)} dx \right)^2 \\ &\leq \int_a^b \frac{g(x)^2}{f_X(x)} \cdot \underbrace{\int_a^b f_X(x) dx}_{=1} \\ &= \int_a^b \left(\frac{g(x)}{f_X(x)} \right)^2 f_X(x) dx \\ &= \mathbb{E} \left[\left(\frac{g(x)}{f_X(x)} \right)^2 \right]. \end{aligned}$$

□

If $f_X(x) \approx |g(x)|$ then we are close to minimizing. “High values” based on g should be sampled more frequently.

Another way of controlling variance is using the **stratification method**.

3.2.4 Variance reduction: Control Variable method

Here we have $I = \mathbb{E}[Y]$ for some Y and $\vartheta_2 = \frac{1}{b} \sum_{k=1}^n Y_k$. Let C be a random variable such that we know the value of $\mu_C = \mathbb{E}[C]$. Thus

$$\mathbb{E}[Y] = \mathbb{E}[Y - \beta(C - \mu_C)] \quad \forall \beta \in \mathbb{R}.$$

We define

$$\vartheta_c = \frac{1}{n} \sum_{k=1}^n (Y_k - \beta(C_k - \mu_C)) \quad \text{with } (Y_k, C_k) \text{ i.i.d.}$$

We now compute the variance of ϑ_c :

$$\begin{aligned} \vartheta_c &\stackrel{\mathbb{E}}{=} \frac{1}{n^2} \sum_{k=1}^n \text{Var}(Y_k - \beta(C_k - \mu_C)) \\ &\stackrel{\text{i.d.}}{=} \frac{1}{n} \text{Var}(Y_1 - \beta(C - \mu_C)) \\ &= \frac{1}{n} (\text{Var}(Y) + \beta^2 \text{Var}(C) - 2\beta \text{Cov}(Y, C)). \end{aligned}$$

If $\beta^2 \text{Var}(C) - 2\beta \text{Cov}(Y, C) < 0$ then $\text{Var}(\vartheta_c) < \text{Var}(\vartheta_2)$ while $\mathbb{E}[\vartheta_c] = \mathbb{E}[\vartheta_2] = I$. The minimal value is obtained with

$$\beta^\star = \frac{\text{Cov}(Y, C)}{\text{Var}(C)} = \left(1 - \rho_{Y,C}^2\right) \text{Var}(Y) \quad \rho_{Y,C}^2 = \text{correlation coefficient.}$$

As an example, consider

$$I = \mathbb{E} [\mathbb{1}_{\{X>0\}} \cdot e^{5X}] \quad \text{with } X \sim \mathcal{N}(0, 1).$$

We have

$$\begin{aligned} I &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot e^{5x} dx \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{5x - \frac{x^2}{2}} dx \\ &= \frac{e^{\frac{25}{2}}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{(x-5)^2}{2}} dx \\ &= \frac{e^{\frac{25}{2}}}{\sqrt{2\pi}} \int_{-5}^\infty e^{-\frac{y^2}{2}} dy \\ &= \mathbb{E} \left[e^{\frac{25}{2}} \cdot \mathbb{1}_{\{X \geq -5\}} \right] \\ &= e^{\frac{25}{2}} \cdot \mathbb{P}(X \geq -5). \end{aligned}$$

We have used the fact that

$$\mathbb{E} [e^{5x} \cdot \mathbb{1}_{\{x>0\}}] = \mathbb{E} \left[e^{\frac{25}{2}} \cdot \mathbb{1}_{\{x \geq -5\}} \right].$$

What about the variance, which we want to reduce? The application showed that the margin of the confidence interval is high. Our candidate control variable is $C = e^{5x}$. Why is that? Because we know $\mathbb{E}[e^{5x}]$: it is a Laplace transform

$$\forall X \in L^1 \quad \mathcal{L}\{X\}(\lambda) = \mathbb{E} [e^{\lambda X}] \quad \forall \lambda \in \mathbb{C}.$$

For $X \sim \mathcal{N}(0, 1)$, $\mathcal{L}\{X\}$ is known:

$$\mathbb{E} [e^{\lambda X}] = e^{\frac{\lambda^2}{2}} \quad \forall \lambda \in \mathbb{C}.$$

3.3 Simulation of a Brownian motion

3.3.1 Definition and approach

Definition 3.3.1

Invariance property:

$$(B_{ct})_{t \in [0, \infty)} \sim (\sqrt{c} B_t)_{t \in [0, \infty]}.$$

The one-dimensional Brownian motion is the unique Gaussian process with continuous trajectories such that $\mathbb{E}[B_s B_t] = s \wedge t$.

Proof

This is the proof of both of these properties.

$$\begin{bmatrix} B_{t_1} \\ \vdots \\ B_{t_n} \end{bmatrix} \sim \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\prod_{j=1}^n (t_j - t_{j-1})}} e^{-\frac{1}{2} \sum_{j=1}^n \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}}} d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} e^{-\frac{1}{2} \langle \mathbf{x}, \Sigma^{-1} \mathbf{x} \rangle} d\mathbf{x}$$

with $\Sigma = [t_j \wedge t_k]_{j,k=1}^n$ (recall the Schilling book). \square

A d -dimensional Brownian motion is characterized by d independent one-dimensional Brownian motions.

The issue with simulating a Brownian path with a computer is that time is continuous. We will make use of the above properties to simulate a one-dimensional Brownian motion and a d -dimensional in general. We introduce a time grid on $[0, T]$

$$0 = t_0 < \frac{t_1}{\frac{T}{n}} < \dots < \frac{t_{n-1}}{\frac{n-1}{n}T} < \frac{t_n}{T}.$$

We generate B_{t_k} , $k = 1, \dots, n$ and then we construct a function $[0, T] \ni t \mapsto \hat{B}_t$. either by linear interpolation or piecewise filling.

We need to simulate a random vector $\begin{bmatrix} B_{t_1} \\ \vdots \\ B_{t_n} \end{bmatrix} \sim N\left(\mathbf{0}, \frac{T}{n} \begin{bmatrix} 1 & \cdots & \cdots & 1 \\ \vdots & 2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & \cdots & n \end{bmatrix}\right)$. How to do this

in practice?

Revise with Kotatsu!

Let Γ be symmetric positive semi-definite. Then $\exists A_{n \times n}$ such that $\Gamma = AA^\top$. So

$$X \sim N(\mathbf{0}, I) \implies Y := AX \sim N(\mathbf{0}, \Gamma).$$

A positive, semi definite matrix admits infinite square roots (i.e. such that $AA^\top = \Gamma$) but there exists a unique lower triangular matrix A such that $AA^\top = \Gamma$. This matrix is called the **Cholesky matrix**^a.

^aIn Python, use `numpy.linalg.cholesky`

Proof

Γ is symmetric, so there exists an orthogonal U such that $\Gamma = UDU^\top$ with $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ eigenvalues of Γ . Recall that Γ is positive semi-definite and this is equivalent to saying that $\lambda_i \geq 0$ for $\forall i = 1, \dots, n$. If we define

$$D_1 = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{bmatrix}$$

and set $A := UD_1$ then

$$\begin{aligned} AA^\top &= UD_1(UD_1)^\top \\ &= U \underbrace{D_1 D_1^\top}_{=D} U^\top \\ &= UDU^\top = \Gamma. \end{aligned}$$

So the existence is proven. \square

Since $X \sim N(\mathbf{0}, I)$

$$\text{Var}(AX) = A(\text{Var } X)A^\top = AA^\top = \Gamma.$$

In the case of a Brownian motion the Cholesky of $[j \wedge k]_{j,k=1}^n$ is

$$\begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}.$$

Try with $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Hence

$$\Gamma = \sqrt{\frac{T}{n}} \cdot \underbrace{\begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}}_{n \times n} \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}}_{n \times n} \sqrt{\frac{T}{n}}.$$

The direct calculation is:

$$B_{t_1}, \dots, B_{t_n} = \left(B_{t_1}, B_{t_2} - B_{t_1} + B_{t_1}, \dots, B_{t_1} + \sum_{j=2}^n (B_{t_j} - B_{t_{j-1}}) \right)$$

so we only generate normal i.i.d random variables.

3.3.2 First approach

Theorem 3.3.1

If $(\hat{B}_t)_{t \in [0, T]}$ is the piecewise approximation of the Brownian motion there exists a constant c_t depending on time such that

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\hat{B}_t - B_t| \right] \leq \frac{c_t}{\sqrt{n}} \quad \forall n \geq 1.$$

Proof

① $\mathbb{P} \left(\sup_{t \in [0, T]} |B_t - \hat{B}_t| > \delta \right)$ for $\delta > 0$.

$$\begin{aligned} \mathbb{P} \left(\sup_{t \in [0, T]} |B_t - \hat{B}_t| > \delta \right) &= \mathbb{P} \left(\left\{ \sup_{t \in [0, T]} B_t - \hat{B}_t > \delta \right\} \cup \left\{ \sup_{t \in [0, T]} -B_t + \hat{B}_t > \delta \right\} \right) \\ &\leq 2 \mathbb{P} \left(\sup_{t \in [0, T]} B_t - \hat{B}_t > \delta \right) \quad \text{refl. principle} \\ &\leq 2 \mathbb{P} \left(\max_{1 \leq k \leq n} \sup_{t \in [t_k, t_{k-1}]} B_t - \hat{B}_t > \delta \right) \\ &\leq 2n \mathbb{P} \left(\sup_{t \in [0, t_1]} B_t - \hat{B}_{t_0} > \delta \right) \\ &\leq 2n \mathbb{P} \left(\sup_{t \in [0, \frac{T}{n}]} B_t > \delta \right) \\ &\leq 4n \mathbb{P} \left(B_{\frac{T}{n}} > \delta \right) \sim N \left(0, \frac{T}{n} \right). \end{aligned}$$

This is called Andre's reflection principle. From this it follows that

$$\mathbb{P} \left(\sup_{[0, T]} |B_t - \hat{B}_t| > \delta \right) \leq \frac{4n}{\sqrt{2\pi}} \int_{\frac{\delta\sqrt{n}}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} dy \xrightarrow[n \rightarrow \infty]{} 0$$

so

$$\sup_{[0, T]} |B_t - \hat{B}_t| \xrightarrow{\mathbb{P}} 0$$

as $n \rightarrow \infty$.

② By Holder inequality we have that

$$\mathbb{E} \left[\sup_{t \in [0, T]} |B_t - \hat{B}_t| \right] \leq \mathbb{E} \left[\sup_{t \in [0, T]} |B_t - \hat{B}_t|^p \right]^{\frac{1}{p}}.$$

Let A be a positive random variable with a density which decreases exponentially fast. Then

$$\begin{aligned} \mathbb{E} [A^p] &= \int_0^\infty y^p f_A(y) dy \\ &= p \int_0^\infty y^{p-1} \mathbb{P}(A \geq y) dy. \end{aligned}$$

We can use this fact to say that

$$\begin{aligned}
\mathbb{E} \left[\sup_{t \in [0, T]} |B_t - \hat{B}_t|^p \right]^{\frac{1}{p}} &\leq \left(p \int_0^\infty \delta^{p-1} \underbrace{\frac{4n}{\sqrt{2\pi}} \int_{\frac{\delta\sqrt{n}}{\sqrt{t}}}^\infty e^{-\frac{y^2}{2}} dy d\delta }_{\mathbb{P}(A \geq \delta)} \right)^{\frac{1}{p}} \\
\text{by Fubini} &\leq \left(p \int_0^\infty \int_0^\infty \delta^{p-1} \frac{4n}{\sqrt{2\pi}} \mathbf{1}_{\{\delta \leq \frac{y\sqrt{t}}{\sqrt{n}}\}} e^{-\frac{y^2}{2}} dy d\delta \right)^{\frac{1}{p}} \\
&\leq \left(\int_0^\infty \left(\frac{y\sqrt{t}}{\sqrt{n}} \right)^p \frac{4n}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right)^{\frac{1}{p}} \\
&\leq \frac{\sqrt{t}}{(2\pi)^{\frac{1}{2p}}} n^{\frac{1}{p}-\frac{1}{2}} \left(\int_0^\infty y^p e^{-\frac{y^2}{2}} dy \right)^{\frac{1}{p}} \\
&\leq c_t n^{\frac{1}{p}-\frac{1}{2}}.
\end{aligned}$$

So we have that

$$\sqrt{n} \mathbb{E} \left[\sup_{t \in [0, T]} |B_t - \hat{B}_t| \right] \leq c_t n^{\frac{1}{p}} \quad \forall p \geq 1.$$

For $p = n$

$$n^{\frac{1}{n}} = e^{\frac{\ln n}{n}} \rightarrow 1.$$

□

What the fuck?

Say we have to generate a n -dim random vector and use it to approximate a Brownian motion.
We have

$$(B_{\frac{T}{n}}, B_{\frac{2T}{n}}, \dots, B_T) \sim \left(\sqrt{\frac{T}{n}} X_1, \sqrt{\frac{T}{n}} X_1 + \sqrt{\frac{T}{n}} X_2, \dots, \sqrt{\frac{T}{n}} X_1 + \dots + \sqrt{\frac{T}{n}} X_n \right)$$

. What we need to do is:

1. extract a n -dimensional Gaussian vector;
2. use Cholesky algorithm to make the covariance matrix lower rectangular
3. calculate the cumulative sum (each sum will give a point of the Brownian motion).

3.3.3 Levy's original argument

Let us assume that we have simulated the Brownian motion on a time grid of size n . If we want to generate a Brownian motion on a thinner time grid we need to do everything again from scratch which is shit. To avoid this, we work with time grids of 2^n and use Brownian bridges.

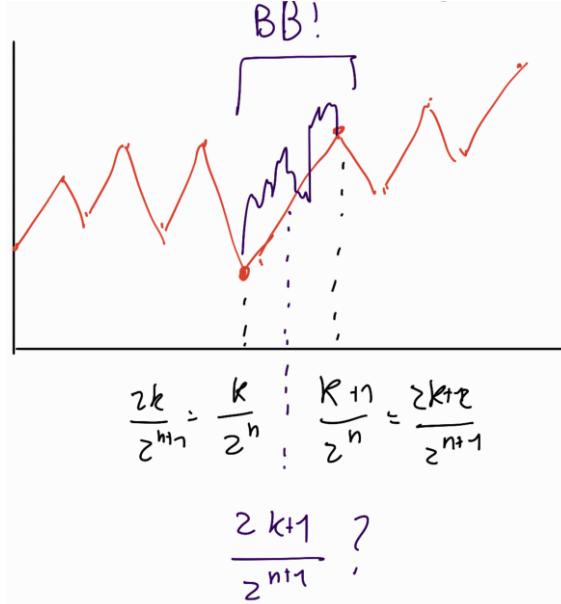


Figure 3.15: To “add” new information between two points we use a Brownian bridge.

We want to describe the value of $B_{\frac{k+1}{2^{n+1}}}$ when $B_{\frac{k}{2^n}}$ and $B_{\frac{k+1}{2^n}}$ are known.

Proposition 3.3.1

If $X \sim N(0, \sigma_1^2)$ and $Y \sim N(0, \sigma_2^2)$ then

$$X|X+Y \sim N\left(\frac{(X+Y)\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right).$$

Proof

We have $X+Y \sim N(0, \sigma_1^2 + \sigma_2^2)$ so

$$\mathbb{E}[f(X)g(X+Y)] = \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \int \int f(x)g(x+y)e^{-\frac{x^2}{2\sigma_1^2}-\frac{(y-x)^2}{2\sigma_2^2}} dx dy.$$

Perform a change of variable

$$\begin{cases} u = x \\ v = x + y \end{cases}$$

and get

$$\mathbb{E}[f(X)g(X+Y)] = \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \int \int f(y)g(v)e^{-\frac{u^2}{2\sigma_1^2}-\frac{(v-y)^2}{2\sigma_2^2}} dv du$$

so the conditional distribution is

$$\begin{aligned} & \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2} \exp\left\{-\frac{u^2}{2\sigma_1^2} + \frac{v^2}{2(\sigma_1^2 + \sigma_2^2)} - \frac{v^2}{2\sigma_2^2} - \frac{u^2}{2\sigma_1^2} + \frac{2uv}{2\sigma_2}\right\} \\ & \Downarrow \\ & \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left(u - \frac{v\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)^2 \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\sigma_2^2}\right\}. \end{aligned}$$

□

In the Brownian case we have that the sum $B_{t_3} - B_{t_1}$ is known and

$$\frac{B_{t_3} - B_{t_2}}{\mathcal{N}(0, t_3 - t_2)} \perp\!\!\!\perp \frac{B_{t_2} - B_{t_1}}{\mathcal{N}(0, t_2 - t_1)}.$$

Proposition 3.3.2

We have that

$$B_{t_2} - B_{t_1} \mid B_{t_3} - B_{t_1} \sim \mathcal{N}\left(\frac{(B_{t_3} - B_{t_1})(t_2 - t_1)}{t_3 - t_1}, \frac{(t_2 - t_1)(t_3 - t_1)}{t_3 - t_1}\right).$$

So we have

$$B_{t_2} \sim B_{t_1} + \frac{t_2 - t_1}{t_3 - t_1} (B_{t_3} - B_{t_1}) + \sqrt{\frac{(t_2 - t_1)(t_3 - t_2)}{t_3 - t_1}} G \sim \mathcal{N}(0, 1).$$

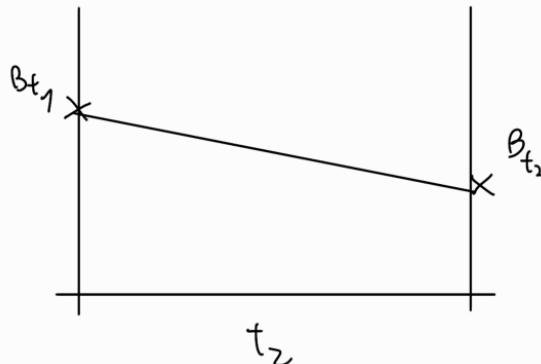


Figure 3.16: Another very necessary figure

3.3.4 Karhunen-Loeve theorem

We know that if $(X_t)_{t \geq 0}$ is a Brownian motion $\implies (X_t)$ is a Gaussian continuous process with covariance function $K(s, t) = s \wedge t$. Let us construct the linear operator $T_k : L^2([a, b]) \rightarrow L^2([a, b])$:

$$T_k(f)(t) = \int_a^b K(s, t)f(s) ds.$$

We can see that T_k is a probability kernel. Since it is a linear operator we can use the spectral decomposition and compute eigenfunctions and eigenvalues (e_k, λ_k) .

Theorem 3.3.2

(Mercer). Since K is symmetric positive semidefinite then $(e_k)_k$ is an orthogonal basis and K has the representation

$$K(s, t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t).$$

We construct

$$X_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k} G_k e_k(t)$$

where G_k is a sequence of i.i.d. Gaussian random variable $N(0, 1)$. X_t converges in L^2 . Moreover

$$\begin{aligned} \mathbb{E}[X_t X_s] &= \sum_{k,j=1}^{\infty} \sqrt{\lambda_k \lambda_j} \underbrace{\mathbb{E}[G_k G_j]}_{0 \text{ if } k \neq j} e_k(t) e_j(s) \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s) \\ &= K(s, t). \end{aligned}$$

Appendix A

Code

A.1 Simulation of random variables

Simulation of random variables using a linear congruence generator.

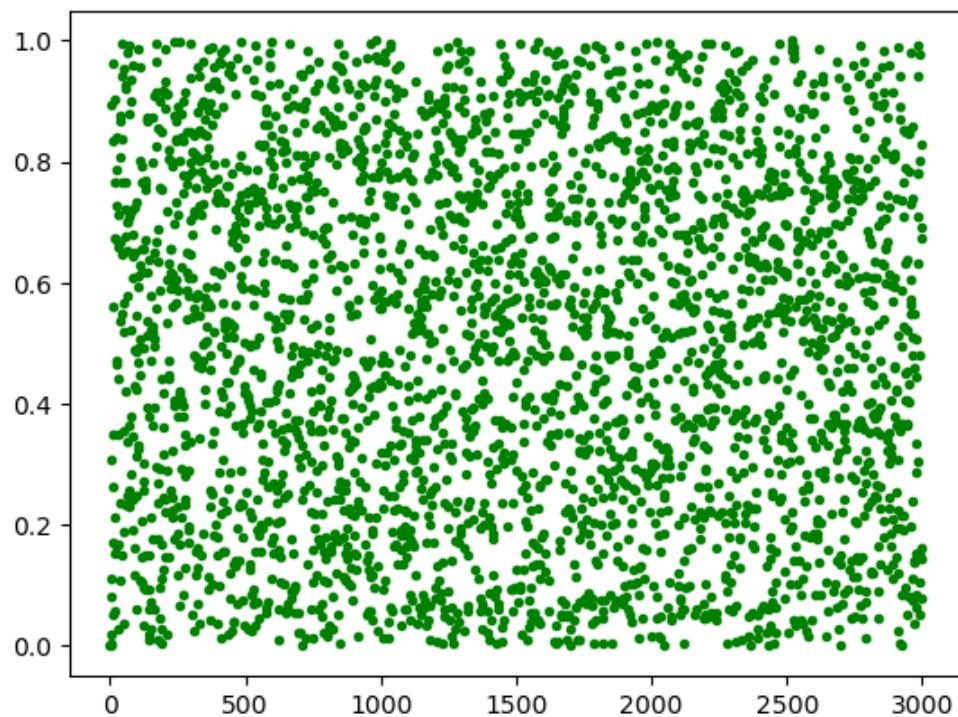
Using Fortran's congruence generator (IBM)

```
In [3]: alea=[];
alea.append(50);
for i in range(0,50):
    alea.append((2**16+3)*alea[i]%(2**32));
alea_norm=[alea[i]/(2**32) for i in range(0,50)];
print(alea_norm)
#print(alea)
#20 then 3000

[1.1641532182693481e-08, 0.0007629743777215481, 0.004577741492539644, 0.020599679555743933, 0.0823984039016068, 0.3089933074079454, 0.11237420933321118, 0.8933054893277586, 0.34846505196765065, 0.051040907856076956, 0.17005997942760587, 0.5609917058609426, 0.8354104203172028, 0.9635371691547334, 0.26252923207357526, 0.9033408700488508, 0.057282131630927324, 0.21362495934590697, 0.7662105713970959, 0.6746387942694128, 0.15193762304261327, 0.8398765898309648, 0.6718209316022694, 0.47203628113493323, 0.7858293023891747, 0.4666492841206491, 0.7274319832213223, 0.1647483422420919, 0.4416022044606507, 0.16687814658507705, 0.026849039364606142, 0.6591909169219434, 0.7135041472502053, 0.34830663120374084, 0.6683024619705975, 0.8750550909899175, 0.23560838820412755, 0.5381545103155077, 0.10845156805589795, 0.8073188154958189, 0.8678487804718316, 0.9412233433686197, 0.836701035965234, 0.5491961254738271, 0.7648674291558564, 0.6464394456706941, 0.9948298116214573, 0.15102385869249701, 0.952674847561866, 0.35683435713872313]
```

In [4]:

```
import matplotlib.pyplot as plt
plt.plot(range(0,3000),alea_norm,'.',color='g')
plt.show()
```



In [6]:

```
%matplotlib
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
alea1=[alea_norm[3*i] for i in range(0,1000)];
alea2=[alea_norm[3*i+1] for i in range(0,1000)];
alea3=[alea_norm[3*i+2] for i in range(0,1000)];
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(alea1, alea2, alea3, c='r', marker='o')
ax.set_xlabel('alea1')
ax.set_ylabel('alea2')
ax.set_zlabel('alea3')

plt.show()
```

Using matplotlib backend: MacOSX

In [7]:

```
import numpy.random as npr
a=npr.rand()
print('random value:',a)
print('random value: {:.7f}'.format(a))
npr.seed(2025)
a=npr.rand()
b=npr.rand()
c=npr.rand()
print(a,b,c)
npr.seed(2025)
a=npr.rand()
b=npr.rand()
c=npr.rand()
print(a,b,c)
```

```

npr.seed() # it uses the current time as the seed value
a=npr.rand()
b=npr.rand()
c=npr.rand()
print(a,b,c)

```

```

random value: 0.32762316486049825
random value: 0.3276232
0.1354881636779618 0.887851702730378 0.9326056398865025
0.1354881636779618 0.887851702730378 0.9326056398865025
0.3097299374127067 0.4158611883324189 0.968085811840334

```

```

In [8]: import numpy.random as npr
import numpy as np
m=200
outcome=[]
proba=[0, 0.1, 0.6, 0.3]
values=[1,7,8]
cumprob=np.cumsum(proba)
coin=npr.rand(m)
for i in range(0,m):
    k=sum((coin[i]>cumprob)*1)-1
    outcome.append(values[k])
print(outcome)
import matplotlib.pyplot as plt
plt.hist(outcome,density=True)
plt.show()

```

```

[7, 8, 8, 7, 8, 8, 8, 7, 8, 8, 7, 7, 7, 7, 7, 7, 7, 7, 7, 1, 8, 7, 7,
7, 8, 1, 1, 8, 7, 8, 7, 8, 7, 8, 8, 7, 7, 7, 7, 7, 8, 8, 8, 1, 8, 7, 1, 7,
1, 7, 7, 1, 8, 1, 7, 7, 7, 7, 7, 8, 7, 8, 7, 7, 7, 7, 8, 7, 7, 7, 7, 1, 8, 8,
7, 1, 7, 8, 7, 8, 7, 7, 8, 7, 8, 8, 7, 7, 7, 7, 8, 7, 7, 7, 8, 8, 7, 8, 8, 8,
8, 1, 7, 7, 8, 1, 7, 7, 7, 7, 8, 7, 1, 7, 8, 7, 7, 7, 7, 1, 8, 7, 7, 7, 8, 7,
7, 8, 7, 8, 8, 7, 8, 7, 1, 8, 8, 7, 8, 7, 7, 7, 8, 7, 7, 7, 8, 7, 7, 7, 8, 1,
8, 7, 8, 7, 7, 1, 8, 7, 1, 7, 7, 8, 7, 7, 7, 8, 7, 7, 7, 8, 1, 7, 7, 1, 7, 7,
7, 7, 7, 8, 8, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7]

```

```

In [9]: p=0.2;
print((npr.rand()<=p))
print((npr.rand()<=p)*1)

```

```

True
0

```

```

In [10]: p=0.2;
n=10;
a=(npr.rand(n)<p)*1
binom=sum(a)
print(a)
print(binom)

```

```

[1 0 0 0 0 0 1 1 0 1]
4

```

```

In [11]: n=10;
lambda0=1;
var=-np.log(npr.rand(n))/lambda0;
print(var)

```

```

[1.66633425 1.52111289 1.28888837 1.37214109 0.54077533 0.3734536
1.18128997 1.42418584 0.85980459 0.21336129]

```

```
In [13]: n=100;
p=0.1;
var=np.floor(np.log(np.random(n))/np.log(1-p))+1;
print(var)

[13. 19. 10. 22. 21. 20. 4. 4. 25. 6. 3. 12. 3. 13. 4. 16. 33. 4.
 4. 2. 7. 12. 3. 3. 29. 20. 6. 42. 18. 5. 7. 11. 5. 3. 5. 9.
 27. 7. 13. 2. 6. 3. 9. 12. 2. 3. 31. 3. 5. 17. 7. 8. 11. 1.
 12. 6. 18. 6. 11. 7. 3. 8. 4. 7. 17. 9. 3. 7. 10. 24. 7. 27.
 3. 7. 8. 3. 5. 9. 4. 13. 14. 3. 10. 8. 6. 21. 16. 1. 3. 1.
 6. 15. 30. 15. 4. 3. 21. 12. 22. 3.]
```



```
In [16]: n=10;
mu=0;
sigma=1;
print(np.random.normal(mu,sigma,n))

[-0.93397138  1.58991113 -1.07445361 -0.4209596   2.06633613 -1.91859686
 0.10345337 -0.12858846  0.67393238 -0.76958424]
```

A.2 Monte Carlo methods

Monte Carlo methods

```
In [2]: import numpy.random as npr
import numpy as np
a=1
b=1.5
c=(np.log(3)-np.log(2))**2
N=1000000
print('number of variates:', N)
X=npr.rand(2,N)
Z=[(np.log(a+(b-a)*X[0,i])**2>c*X[1,i])*1 for i in range(0,N)]
theta=np.mean(Z)*c*(b-a)
print("estimated value:", theta)
print("estimated value {:.7f}".format(theta))
theta_b=theta-(1/np.sqrt(N))*1.96*np.std(Z)*c*(b-a);
theta_h=theta+(1/np.sqrt(N))*1.96*np.std(Z)*c*(b-a);
print("confidence interval: lower {:.7f}, upper {:.7f}".format(theta_b,theta_h))

number of variates: 1000000
estimated value: 0.030210503047408098
estimated value 0.0302105
confidence interval: lower 0.0301328, upper 0.0302882
```

```
In [3]: N=100000;
X=npr.randn(N)
Z=[np.exp(5*X[i])*(X[i]>0) for i in range(N)]
theta=np.mean(Z)
sigma=np.std(Z)
marge=1.96*sigma/np.sqrt(N)
print("estimated value: {:.4f}".format(theta))
print("margin: {:.4f}".format(marge))
print("confidence interval: lower {:.4f}, upper {:.4f}".format(theta-marge,theta+mage))

estimated value: 142079.3472
margin: 162065.4521
confidence interval: lower -19986.1050, upper 304144.7993
```

```
In [4]: N=100000;
X=npr.randn(N)
W=[np.exp(25/2)*(X[i]+5>0) for i in range(N)]
A=[(X[i]+5>0) for i in range(N)]
theta1=np.mean(W)
sigma1=np.std(W)
marge1=1.96*sigma1/np.sqrt(N)
print("estimated value: {:.15f}".format(theta1))
print("margin:", marge1)
print("confidence interval: lower", theta1-marge1, ", upper", theta1+marge1)

estimated value: 268337.286520874535199
margin: 3.6077482008433567e-13
confidence interval: lower 268337.28652087454 , upper 268337.28652087454
```

```
In [6]: N=100000;
X=npr.randn(N)
R=[np.exp(25/2)-np.exp(5*X[i])*(X[i]<0) for i in range(N)]
#print(R)
theta2=np.mean(R)
sigma2=np.std(R)
```

```
marge2=1.96*sigma2/np.sqrt(N)
print("confidence interval: lower", theta2-marge2, ", upper",theta2+marge2)
print("comparison of the margins :", 1.96*sigma/np.sqrt(N), 1.96*sigma1/np.sqrt(
confidence interval: lower 268337.2088204195 , upper 268337.2110815811
comparison of the margins : 518524.4469478968 3.6077482008433567e-13 0.0011305807
631627979
```

In []:

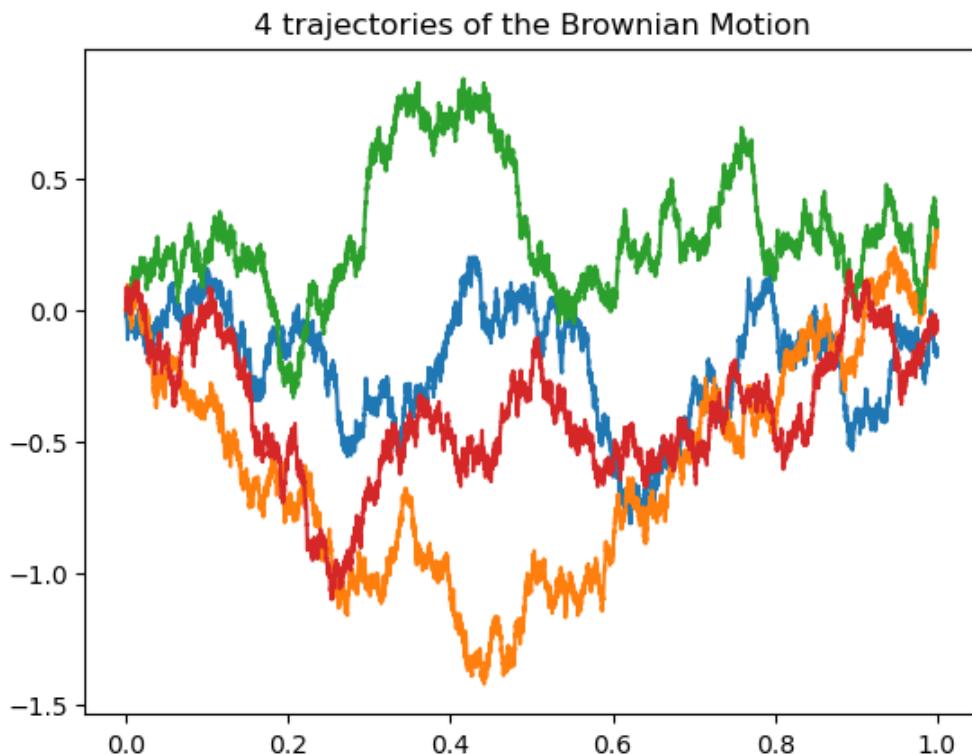
A.3 Simulation of Brownian motion

1. Simulation of the Brownian motion as a Gaussian process

Brownian motion (B_t , $t \in [0, 1]$)

```
In [12]: from math import *
import numpy.random as npr
import numpy as np
import matplotlib.pyplot as plt
import numpy.linalg as alg
```

```
In [15]: # number of steps of the time grid
N=10000
delta=1/N
t=np.linspace(0,1,N+1)
for i in range(4):
    X=npr.normal(0,1,size=N)
    Brownian=np.array([0])
    Brownian.append(Brownian,np.cumsum(np.sqrt(delta)*X))
    plt.plot(t,Brownian)
plt.title('4 trajectories of the Brownian Motion')
plt.show()
```



```
In [16]: # construction of the covariance matrix
def CovMatrix(n):
    K=np.zeros((n,n))
    for i in range(n):
        for j in range(n):
```

```
if i < j:  
    K[i,j]=(1+i)/n  
else:  
    K[i,j]=(1+j)/n  
  
return K  
print(CovMatrix(10))
```

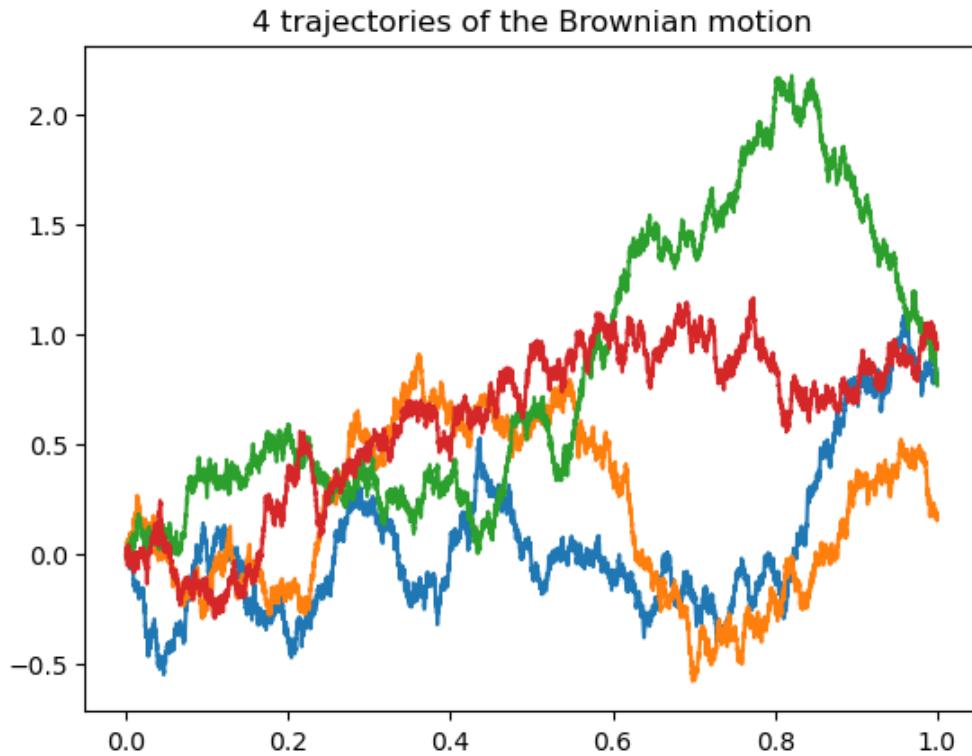

[[0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1]
[0.1 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2]
[0.1 0.2 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3]
[0.1 0.2 0.3 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4]
[0.1 0.2 0.3 0.4 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5]
[0.1 0.2 0.3 0.4 0.5 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6]
[0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.7 0.7 0.7 0.7 0.7 0.7]
[0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.8 0.8 0.8 0.8 0.8]
[0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.9 0.9 0.9 0.9]
[0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.]]

```
In [18]: # construction of N Gaussian vectors centered with a given covariance matrix
def SqrCovMatrix(n):
    return alg.cholesky(CovMatrix(n))
param_n=10
print(SqrCovMatrix(param_n))
def VGauss(n,Nbr):
    X=SqrCovMatrix(n)@npr.randn(n,Nbr)
    return X

[[0.31622777 0.          0.          0.          0.
  0.          0.          0.          0.          ]
 [0.31622777 0.31622777 0.          0.          0.
  0.          0.          0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.          0.
  0.          0.          0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.
  0.          0.          0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.          0.          0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.          0.          0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.31622777 0.          0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.31622777 0.31622777 0.          0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.31622777 0.31622777 0.31622777 0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.31622777 0.31622777 0.31622777 0.31622777 0.          ]
 [0.31622777 0.31622777 0.31622777 0.31622777 0.31622777
  0.31622777 0.31622777 0.31622777 0.31622777 0.31622777]]
```

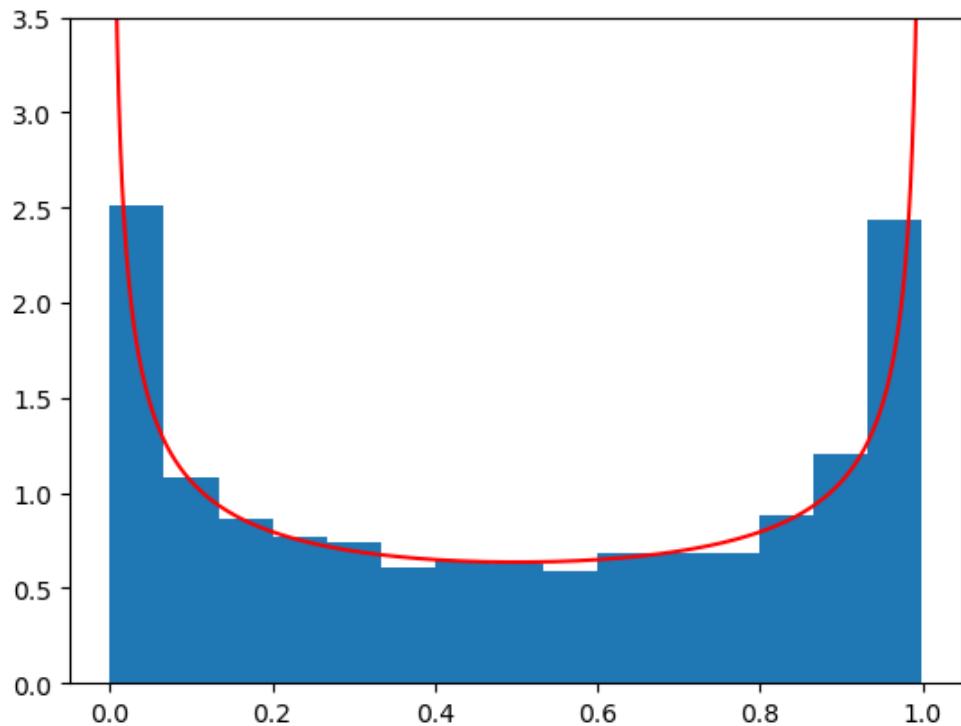
```
In [19]: # représentation graphique
n=10000
N_traj=4
X=VGauss(n,N_traj)
B0=np.zeros([1,N_traj])
X=np.concatenate((B0,X),axis=0)
t=np.linspace(0,1,n+1)
plt.plot(t,X[:,0])
plt.plot(t,X[:,1])
plt.plot(t,X[:,2])
plt.plot(t,X[:,3])
plt.title('4 trajectories of the Brownian motion')
plt.show
```

Out[19]: <function matplotlib.pyplot.show(close=None, block=None)>

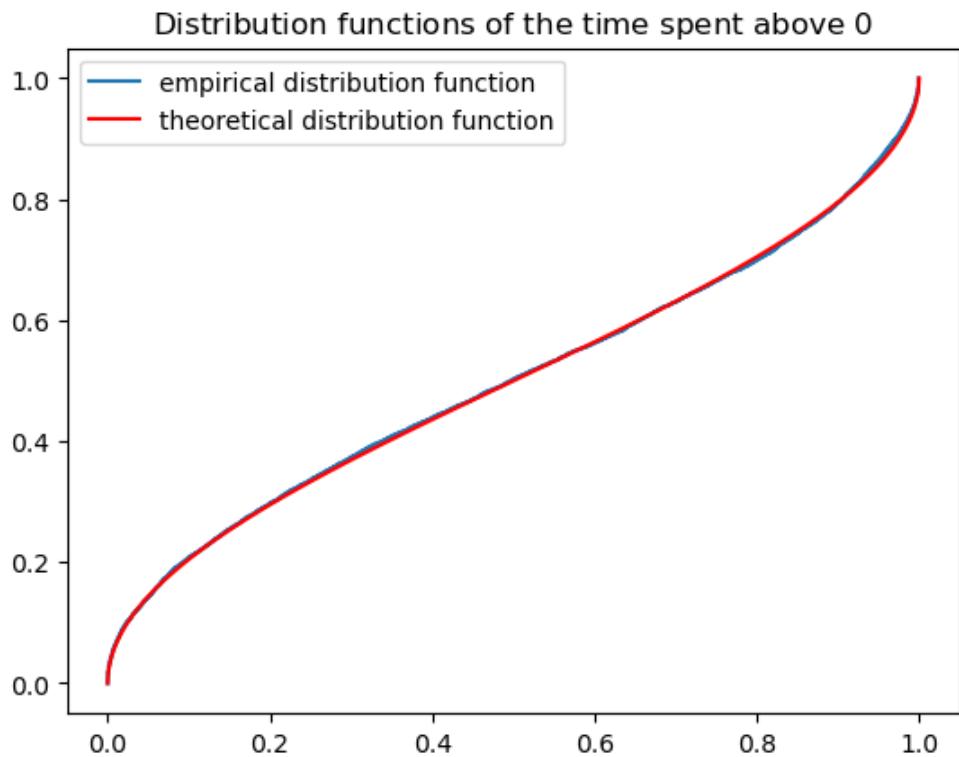


Distribution of the time spent by the Brownian motion in \mathbb{R}_+

```
In [23]: n=1000
N_traj=10000
X=VGauss(n,N_traj)
B0=np.zeros([1,N_traj])
X=np.concatenate((B0,X),axis=0)
positive=np.mean((X>0)*1,0)
plt.hist(positive,density=True,bins=15)
t=np.linspace(0.001,0.999,999)
th_density=1/(np.pi*np.sqrt(t*(1-t)))
plt.plot(t,th_density,'r')
plt.ylim(0,3.5)
plt.show()
```



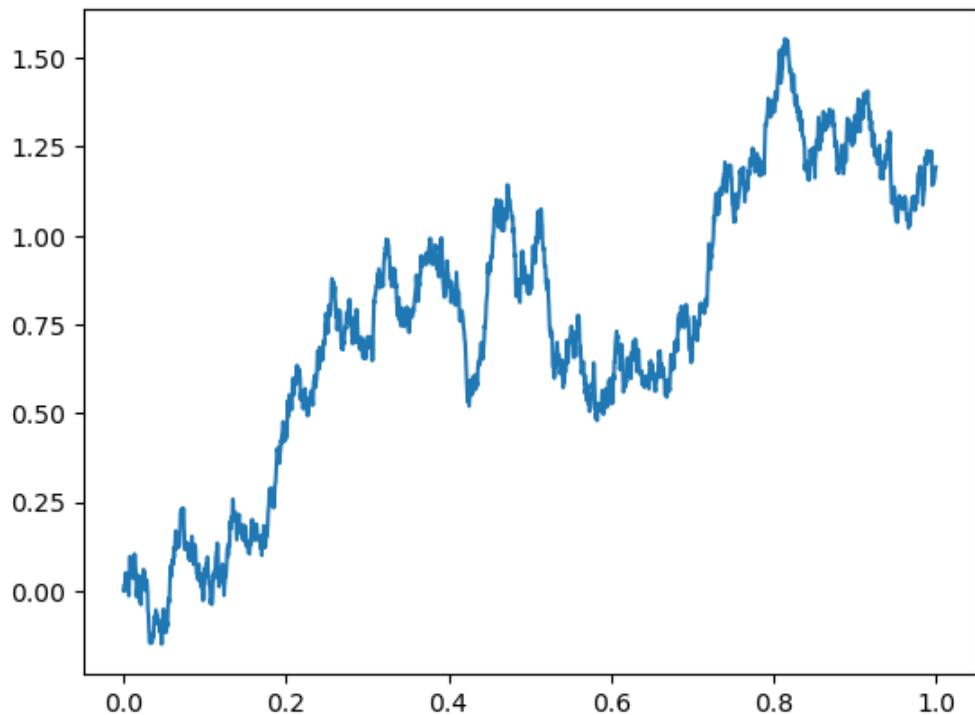
```
In [24]: X = np.sort(positive)
F = np.array(range(N_traj))/N_traj
plt.plot(X, F, label='empirical distribution function')
plt.title('Distribution functions of the time spent above $0$')
t=np.linspace(0,1,1001)
Fth=(2/np.pi)*np.arcsin(np.sqrt(t))
plt.plot(t,Fth,'r',label='theoretical distribution function')
plt.legend()
plt.show()
```



2. Brownian motion - Lévy's argument

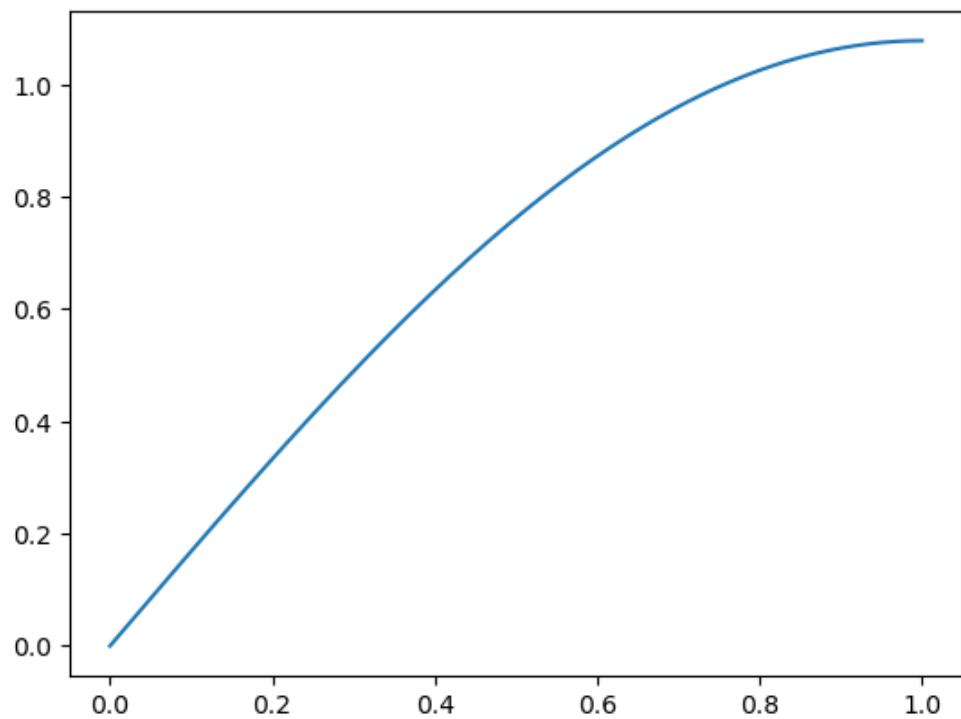
```
In [29]: # size of the step: 2 to the power - n
n=11
B=np.zeros((2**n+1))
B[1]=np.random(1)
B_interm=np.zeros((2**n+1))
#print(B)
for i in range(n):
    for j in range(2**i+1):
        B_interm[2*j]=B[j]
    for j in range(2**i):
        B_interm[2*j+1]=(B[j]+B[j+1])/2+np.random(1)/(2**((i/2)+1))
    B=B_interm
    #print(B_interm)
    B_interm=np.zeros((2**n+1))
t=np.linspace(0,1,2**n+1)
plt.plot(t,B)
```

```
Out[29]: [<matplotlib.lines.Line2D at 0x112491450>]
```



3. Karhunen-Loeve construction

```
In [10]: def funct_e(k,t):
    return np.sqrt(2)*np.sin((k-1/2)*np.pi*t)
def lamb(k):
    return 1/((k-1/2)*np.pi)**2
# number of terms in the truncated sum
def traj(N,t,G):
    x=0
    for i in range(N):
        x=x+G[i]*funct_e(i+1,t)*np.sqrt(lamb(i+1))
    return x
# number of terms in the truncated sum
N=1
nb_step=1000
G=np.random.randn(N)
t=np.linspace(0,1,nb_step)
plt.plot(t,traj(N,t,G))
plt.show()
```



In []:

Appendix B

Homeworks