Brownian motion

Manjunath Krishnapur

Contents

CHAPTER 1

Brownian motion

1. Definition of Brownian motion and Wiener measure

Definition 1: Brownian motion

A collection of random variables $W = (W_t)_{t \geqslant 0}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and satisfying the following properties.

(1) For any $n \geqslant 1$ and any $0 = t_0 < t_1 < \ldots < t_n$, the random variables $W_{t_k} - W_{t_{k-1}}$, $1 \leqslant k \leqslant n$, are independent.

(2) For any $s < t$ the distribution of $W_t - W_s$ is $N(0, t-s)$.

(3) For a.e. $\omega \in \Omega$, the function $t \mapsto W_t(\omega)$ is continuous and $W_0(\omega) = 0$.

That such a collection of random variables exists requires proof. But first, why such a definition? We give some semi-historical and semi-motivational explanation in this section.

1.1. Einstein and the physical Brownian motion. In 1820s, the botanist Robert Brown observed water under a microscope and noticed certain particles buzzing about in an erratic manner. There was no explanation of this phenomenon till about 1905 when Einstein and Smoluchowski (independently of each other) came up with an explanation using statistical mechanics. More precisely, in Einstein's paper, he predicted that a small particle suspended in a liquid undergoes a random motion of a specific kind, and tentatively remarked that this could be the same motion that Brown observed.

We give a very cut-and dried (and half-understood) summary of the idea. Imagine a spherical particle inside water. The particle is assumed to be small in size but observable under a microscope, and hence much larger than the size of water molecules (whose existence was not yet universally accepted at that time). According to the kinetic theory, at any temperature above absolute zero, molecules of water are in constant motion, colliding with each other, changing their direction, exchanging momenta, etc. (rather, it is this motion of molecules that defines the temperature). Now the suspended particle gets hit by agitating water molecules and hence gets pushed around. Each collision affects the larger particle very slightly, but the collisions are relentless and numerous. Hence, the total displacement of the particle in an interval of time is a sum of a large number of

random and mutually independent small displacements. Then, letting X_t denote the displacement of the x-coordinate of the particle, we have the following conclusions.

(1) Over disjoint intervals of time, the displacements or increments of X are independent.

(2) The displacement $X_t - X_s$ in the time interval $[s, t]$ must have Gaussian distribution, by the central limit theorem (assuming that the number of collisions in this interval is large).

(3) If the liquid is homogeneous and isotropic and kept at constant temperature, then the displacement in a given interval of time must have zero mean. And its variance is proportional to the number of collisions during the time interval, which in turn is proportional to the length of the time interval. Thus, $X_t - X_s \sim N(0, D(t-s))$ for $s < t$, for some constant D.

If we set $D = 1$, we get the first two defining properties of Brownian motion. In his paper, Einstein wrote a formula for D in terms of the size of the suspended particle, the ambient temperature, some properties of the liquid and the Avogadro number N. All of these can be measured except N. By measuring the displacement of a particle over a unit interval of time many times, we can estimate $\mathbb{E}[X_1^2]$. Since $D = \mathbb{E}[X_1^2]$, this gives an estimate for D and hence N. This was Einstein's proposal to calculate the Avogadro number by macroscopic observations and apparently this evidence convinced everyone of the reality of atoms.

1.2. Wiener and the mathematical Brownian motion. After the advent of measure theory in the first few years after 1900, mainly due to Borel and Lebesgue, mathematicians were aware of the Lebesgue measure and the Lebesgue integral on $\mathbb{R}^n$. The notion of abstract measure was also developed by Fréchet before 1915. Many analysts, particularly Gateaux, Lévy and Daniell and Wiener, pursued the question as to whether a theory of integration could be developed over infinite dimensional space[1]. One can always put an abstract measure on any space, but they were looking for something natural.

What is the difficulty? Consider an infinite dimensional Hilbert space such as ℓ^2, the space of square summable infinite sequences. Is there a translation invariant Borel measure on ℓ^2? Consider the unit ball $\mathbb{B}$. There are infinitely many pairwise disjoint balls of radius 1 inside $3\mathbb{B}$ (for example, take unit balls centered around each co-ordinate vector $\sqrt{2}e_i$, $i \geqslant 1$). Thus, if $\mu(\mathbb{B}) > 0$, then by

[1] In 1924 or so, Wiener himself realized that dimension is irrelevant in measure theory. Indeed, in probability theory class we have see that once Lebesgue measure on $[0, 1]$ is constructed, one can just push it forward by appropriate maps to get all measures of interest such as Lebesgue measure on $[0, 1]^n$ and even product uniform measure on $[0, 1]^{\mathbb{N}}$. All these spaces are the same in measure theory, in sharp contrast to their distinctness in topology. Therefore, today no one talks of integration in infinite dimension anymore (I think!). We just think that Wiener measure is interesting.

translation invariance, all these balls have the same measure and hence $\mu(3\mathbb{B})$ must be infinite! This precludes the existence of any natural measure such as Lebesgue measure.

What else can one do? One of the things that was tried essentially amounted to thinking of a function $f : [0,1] \to \mathbb{R}$ as an infinite vector $f = (f_t)_{t\in[0,1]}$. In analogy with $\mathbb{R}^n$, where we have product measures, we can consider a product measure $\otimes_t \mu$ on $\mathbb{R}^{[0,1]}$ (the space of all functions from $[0,1]$ to $\mathbb{R}$) endowed with the product sigma-algebra (and μ can be any probability measure on $\mathbb{R}$, e.g., $\mu = N(0,1)$). But this is very poor as a measure space. For example, the space $C[0,1]$ is not a measurable subset of $\mathbb{R}^{[0,1]}$. In fact, any set in the product sigma-algebra is determined by countably many co-ordinates which means that most questions of interest about the trajectory $t \mapsto f_t$ are forbidden.

Norbert Wiener took inspiration from Einstein's theory to ask for the independence of *increments* of f rather than of independence of the *values* of f (which is what product measure does). And then, he showed that it is possible to put a Borel probability measure on $C[0,\infty)$ such that the increments are independent across disjoint intervals. This is why, his 1923 paper that introduced Brownian motion is titled *Differential space*, emphasizing that independence is at the level of differences of the function values.

2. The space of continuous functions

It is most appropriate to think of Brownian motion as a $C[0,\infty)$-valued random variable. Hence we recall the topology and measure structure on this space. Often it is more convenient to work on the compact time interval $[0,1]$, so we also consider $C[0,1]$. We want to endow these spaces with appropriate sigma-algebras.

2.1. The metrics/topologies. On $C[0,1]$, define the norm $\|f\|_{\sup} = \max\{|f(t)| : t \in [0,1]\}$ and the metric $d(f,g) = \|f-g\|_{\sup}$. Convergence in this metric is just uniform convergence on $[0,1]$.

Obviously the sup-norm can be defined for $C[0,T]$ for any $T < \infty$, but not for $C[0,\infty)$, as the latter contains unbounded functions. The metric on $C[0,\infty)$ is defined by

$$d(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min\{1, \|f-g\|_{\sup[0,n]}\}.$$

The metric is irrelevant, what matters is the topology and the fact that the topology is metrizable. In fact, many other metrics such as

$$\tilde{d}(f,g) = \sum_{n=1}^{\infty} \frac{1}{n^2} \frac{\|f-g\|_{\sup[0,n]}}{1+\|f-g\|_{\sup[0,n]}}$$

induces the same topology on $C[0,\infty)$. In this topology, $f_n \to f$ if f_n converges to f uniformly on all compact sets of $\mathbb{R}_+ = [0,\infty)$. The following fact will be important later.

Exercise 1

Show that $C[0,1]$ and $C[0,\infty)$ are complete and separable metric spaces.

2.2. The sigma-algebras. For any metric space (X,d), the *Borel sigma-algebra* $\mathcal{B}_X$ is defined as the smallest sigma-algebra that contains all open subsets of X. Observe that if $F : X \to \mathbb{R}$ is continuous, then $\{F > t\}$ is open for all $t \in \mathbb{R}$. Conversely, any open set G can be written in this form, for example as $G = \{d(\bullet, G^c) > 0\}$. What these two statements show is that $\mathcal{B}_X$ can also be described as the smallest sigma-algebra for which all continuous functions $F : X \to \mathbb{R}$ are measurable.

These considerations apply to $C[0,1]$ and $C[0,\infty)$ and the corresponding Borel-sigma algebras are denoted $\mathcal{B}(C[0,1])$ and $\mathcal{B}(C[0,\infty))$. There is yet another description of these Borel sigma-algebras that will be useful. For $t \in \mathbb{R}_+$ define the projection map $\Pi_t : C[0,1] \to \mathbb{R}$ by $\Pi_t(f) = f(t)$. Let $\mathcal{C}$ denote the smallest sigma-algebra that makes all Π_t measurable. It is easy to see that

$$\mathcal{C} = \sigma\{\Pi_{t_1}^{-1}(A_1) \cap \ldots \cap \Pi_{t_n}^{-1}(A_n) : n \geqslant 1,\ 0 \leqslant t_1 < \ldots < t_n,\ A_1, \ldots, A_n \in \mathcal{B}_{\mathbb{R}}\}.$$

Sets of the form $\Pi_{t_1}^{-1}(A_1) \cap \ldots \cap \Pi_{t_n}^{-1}(A_n)$ are called (finite dimensional) *cylinder sets*[2]. Intersection of two cylinder sets is clearly a cylinder set. Thus, cylinder sets form a π-system that generates $\mathcal{C}$, which we may also call the cylinder sigma-algebra. For $C[0,1]$, we define $\mathcal{C}(C[0,1])$ in the same way, except that the projections Π_t are defined only for $t \in [0,1]$.

Any dilemma of whether to work with the Borel or cylinder sigma algebra is rendered moot by the following observation.

$$\mathcal{C}(C[0,\infty)) = \mathcal{B}(C[0,\infty)) \text{ and } \mathcal{C}(C[0,1]) = \mathcal{B}(C[0,1]). \tag{1}$$

We show the proof for the case of $C[0,1]$.

Proof. As $|\Pi_t(f) - \Pi_t(g)| = |f(t) - g(t)| \leqslant \|f - g\|$, we see that the projections are continuous on $C[0,1]$. Hence they are $\mathcal{B}(C[0,1])$-measurable. This shows that $\mathcal{C}(C[0,1]) \subseteq \mathcal{B}(C[0,1])$. For the converse, take a closed ball of the form $\overline{B(f_0,r)} = \{f \in C[0,1] : \|f - f_0\| \leqslant r\}$ for some $f_0 \in C[0,1]$ and $b > 0$. We can write it as a countable intersection of cylinder sets

$$\overline{B(f_0,r)} = \bigcap_{t \in \mathbb{Q} \cap [0,1]} \Pi_t^{-1}[f_0(t) - r, f_0(t) + r]$$

and hence $\overline{B(f_0,r)} \in \mathcal{C}(C[0,1])$. The open ball $B(f_0,r)$ is the countable union of closed balls $\overline{B(f_0, r - \frac{1}{n})}$ and hence in $\mathcal{C}(C[0,1])$. As $C[0,1]$ is separable, any open set is a countable union of open balls. Thus, $\mathcal{B}(C[0,1]) \subseteq \mathcal{C}(C[0,1])$. ■

[2]Some books use the same name for the more general sets of the form $(\Pi_{t_1}, \ldots, \Pi_{t_n})^{-1}(B)$ for some $B \in \mathcal{B}_{\mathbb{R}^n}$. For example, $\{f \in C[0,1] : f(0) + f(1/2) < 1\}$ is a cylinder set by this definition.

Exercise 2

Prove (1) for the case $C[0,\infty)$.

[*Hint:* Show that any open set in $C[0,\infty)$ can be written as a countable union of sets of the form $\{f \in C[0,\infty) : \|f - f_0\|_{\sup[0,n]} \leqslant r\}$, where $n \geqslant 1$ and $r > 0$. It can be very painful if you try to work with the ad-hoc metrics that were defined on $C[0,\infty)$!]

In summary, the Borel sigma-algebras on $C[0,1]$ and $C[0,\infty)$ are generated by any one of (a) open sets, (b) cylinder sets, (c) projection maps, (d) continuous functions into $\mathbb{R}$. As a cautionary remark, observe that the topology generated by the projection maps is *not* the topology on $C[0,1]$ or $C[0,\infty)$! Indeed, the smallest topology that makes the projections continuous is the topology of pointwise convergence.

2.3. Measures. Now that we have the sigma-algebras, we can ask for Borel probability measure on $C[0,1]$ or $C[0,\infty)$. If μ is a Borel probability measure, then the push-forward measure

$$\mu_{t_1,\ldots,t_n} := \mu \circ (\Pi_{t_1},\ldots,\Pi_{t_n})^{-1}$$

where $n \geqslant 1$ and $0 \leqslant t_1 < \ldots < t_n$, is a Borel probability measure on $\mathbb{R}^n$. These are called the finite dimensional marginals of μ. Knowing the finite dimensional marginals tells us the probabilities of all cylinder sets. Since cylinder sets form a π-system that generate the Borel sigma-algebra, it follows that the finite dimensional marginals determine the measure. In other words,

$$\mu_{t_1,\ldots,t_n} = \nu_{t_1,\ldots,t_n} \text{ for any } n \geqslant 1 \text{ and } t_1 < \ldots < t_n \implies \mu = \nu. \tag{2}$$

How to define interesting probability measures on $C[0,1]$ or $C[0,\infty)$? Recall that when we define probability measures on $\mathbb{R}$ (e.g., $N(0,1)$) or the countable product $\mathbb{R}^{\mathbb{N}}$ (e.g., product measures), we never define it on the whole sigma-algebra but specify the measures on a generating set (such as intervals) and use an abstract existence theorem to say that a measure exists on the sigma-algebra satisfying those specifications. For the same reasons, here we specify probability measures $\mu_{t_1,\ldots,t_n}$ on $\mathbb{R}^n$ for every $n \geqslant 1$ and $t_1 < \ldots < t_n$, and look for a measure having those as its finite dimensional marginals. What (2) shows is uniqueness: there is at most one such measure. The existence is a question whose difficulty is of a different order of magnitude and in fact there is no general necessary and sufficient condition! Here is one example where we can answer the existence question in the affirmative.

Definition 2: Wiener measure

Wiener measure is the Borel probability measure μ on $C[0,\infty)$ such that for any $n \geqslant 1$ and any $t_1 < \ldots < t_n$, the finite dimensional marginal $\mu_{t_1,\ldots,t_n}$ is the multivariate Gaussian distribution $N_n(0, (t_i \wedge t_j)_{i,j\leqslant n})$.

As already discussed, proof is needed that Wiener measure exists, but it is unique (if it exists). In fact, Wiener measure and Brownian motion are two sides of the same coin, related to each other in the same way as a Gaussian random variable and the Gaussian measure are. In other words, Wiener measure is the distribution of Brownian motion.

2.4. Random variables. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A measurable function X from Ω to $C[0,1]$ or $C[0,\infty)$ (these spaces are endowed with the Borel sigma algebras of course) is called a continuous random process, or simply a random variable taking values in these spaces. If we write $X_t = \Pi_t \circ X$, then X_t are real-valued random variables. The fact that the Borel sigma algebra is generated by the projection maps has the convenient consequence that measurability of X is no more than measurability of X_t for each t.

What this means is that a $C[0,\infty)$-valued random variable is the same as a collection $X = (X_t)_{t\geqslant 0}$, where X_t are real-valued random variables with the added condition[3] that the sample paths $t \mapsto X_t(\omega)$ are continuous for each $\omega \in \Omega$. We usually suppress ω, but t is indicated either as subsript or argument or superscript depending on our need and mood.

Example 1: Brownian motion

Let $W = (W_t)_{t\geqslant 0}$ be a Brownian motion. We can now think of it as a single $C[0,\infty)$-valued random variable by writing $(W(\omega))(t) = W_t(\omega)$. But a slight issue is that we allowed the sample paths to be not continuous for $\omega \in N$ for some $N \in \mathcal{F}$ with $\mathbb{P}(N) = 0$, and $W(\omega) \notin C[0,\infty)$ for $\omega \in N$. To rectify this, we redefine the random variables (if you wish you may say "change the random variable", but this is a harmless change) so that $W_t(\omega) = 0$ for all $t \geqslant 0$ and all $\omega \in N$, then the exceptional set goes away and W does become a $C[0,\infty)$-valued random variable.

All the probabilistic information about X is contained in its distribution, which is defined as the push-forward measure $\mu := \mathbb{P} \circ X^{-1}$ on $C[0,1]$ or $C[0,\infty)$. This just means that $\mu(A) = \mathbb{P}\{X^{-1}(A)\}$ for $A \in \mathcal{B}$. It is allso trivially true that every probability measure is the distribution of a random variable. To see this, given a probability measure ν on $C[0,\infty)$, consider the probability space $(C[0,\infty), \mathcal{B}(C[0,\infty)), \nu)$ and define the random variable $X : C[0,\infty) \to C[0,\infty)$ as the identity map $X(\omega) = \omega$. Then the distribution of X is ν. Thus, $C[0,1]$-valued random variables and probability measures on $C[0,1]$ are two sides of the same coin. Usually probabilists prefer to use the language of random variables (defined on an unspecified probability space), but ultimately the questions are all about the distributions of the random variables.

[3] If we drop the continuity condition, then X is called a stochastic process, which means no more than a collection of real valued random variables X_t, $t \geqslant 0$, or equivalently an $\mathbb{R}^{[0,\infty)}$-valued random variable X (here $\mathbb{R}^{[0,\infty)}$ is endowed with the cylinder sigma-algebra).

Example 2: Wiener measure is the distribution of Brownian motion

Let $W = (W_t)_{t\geqslant 0}$ be a Brownian motion. Fix $0 = t_0 < t_1 < t_2 < \ldots < t_n$. By definition, $X_k := W(t_k) - W(t_{k-1}) \sim N(0, t_k - t_{k-1})$, $1 \leqslant k \leqslant n$, are independent. Writing $W(t_k) = X_1 + \ldots X_k$, we see that $(W(t_1), \ldots, W(t_n))$ is a multivariate Gaussian with zero means and covariances

$$\begin{aligned}\mathbb{E}[W(t_j)W(t_k)] &= \mathbb{E}[(X_1 + \ldots + X_j)(X_1 + \ldots + X_k)] \\ &= \mathbb{E}[X_1^2] + \ldots + \mathbb{E}[X_j^2] \text{ if } j \leqslant k, \\ &= t_1 + (t_2 - t_1) + \ldots + (t_j - t_{j-1}) \\ &= t_j.\end{aligned}$$

Thus, if μ is the distribution of W, then its finite dimensional marginal $\mu_{t_1,\ldots,t_n}$ is $N_n(0, (t_j \wedge t_k)_{j,k\leqslant n})$. In other words, μ is the Wiener measure.

3. Other processes from Brownian motion

Brownian motion is not the only $C[0,\infty)$ or $C[0,1]$ valued random variable of interest and correspondingly Wiener measure is only one of the probability measures of interest on these spaces. But once we have Brownian motion, we can transform it to define various other $C[0,1]$ or $C[0,\infty)$ valued stochastic processes. It will be easy to see that these transformed random processes take values in appropriate spaces of continuous functions on $[0,1]$ or $[0,\infty)$. As such, their distributions are entirely determined by their finite dimensional distributions and their existence follows from the existence of Brownian motion.

3.1. Brownian motion started at any location. If W is standard Brownian motion and $x \in \mathbb{R}$, the process X defined by $X_t = x + W_t$ for $t \geqslant 0$, is called Brownian motion started at x.

3.2. Brownian motion with drift and scaling. Let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Then define $X_t = \mu t + \sigma W_t$. This process X is called Brownian motion with drift μ and scale σ. It is the unique process with continuous sample paths and independent increments satisfying $X_t - X_s \sim N(\mu(t-s), \sigma^2(t-s))$ for any $s < t$.

More generally, we can consider the process $t \mapsto f(t) + \sigma W_t$ for some fixed function f as a noisy version of f. Brownian motion moves very randomly, these processes have a deterministic motion on which a layer of randomness is added.

3.3. Multi-dimensional Brownian motion. Brownian motion in $\mathbb{R}^d$, started at $x \in \mathbb{R}^d$, is defined as the stochastic process $W = (W(t))_{t\geqslant 0}$ where $W(t)$ are $\mathbb{R}^d$-valued random variables and (a) $W(0) = x$ a.s., (b) for any $t_1 < \ldots < t_k$, the increments $W(t_1)$, $W(t_2) - W(t_1)$,$\ldots W(t_k) -$

$W(t_{k-1})$ are independent, (c) for any $s < t$ the distribution of $W(t) - W(s)$ is d-dimensional Gaussian with mean vector zero and covariance matrix $(t-s)I_d$, and (d) $t \mapsto W(t)$ is continuous w.p.1.

The existence of such a process need not be proved from scratch. Since we know that standard one-dimensional Brownian motion exists, we can find a probability space on which we have i.i.d. copies $W^{(k)}$, $k \geqslant 1$, of standard one-dimensional Brownian motion. Then define $W(t) = x + (W^{(1)}(t), \ldots, W^{(d)}(t))$. It is easy to check that this satisfies the properties stated above.

It is also worth noting that if we fix any orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_d$ of $\mathbb{R}^d$ and define $W(t) = x + W^{(1)}(t)\mathbf{v}_1 + \ldots + W^{(d)}(t)\mathbf{v}_d$, this also gives d-dimensional Brownian motion (check the properties!). Taking $x = 0$, this shows that standard Brownian motion W on $\mathbb{R}^d$ is invariant under orthogonal transformations, i.e., if $X(t) = PW(t)$ where P is a $d \times d$ orthogonal matrix, then $X \stackrel{d}{=} W$. In other words, it is directly associated to the space $\mathbb{R}^d$, not to the choice of co-ordinate system.

3.4. Ornstein-Uhlenbeck process. Let $X(t) = e^{-\frac{t}{2}} W(e^t)$ for $t \in \mathbb{R}$. Then X is called Ornstein-Uhlenbeck process and it takes values in $C(\mathbb{R})$ (whose topology and Borel signa-algebra are defined similarly to $C(\mathbb{R}_+)$). It is a continuous process and X_t, $t \in \mathbb{R}$ are jointly Gaussian with zero means and covariances $\mathbb{E}[X_t X_s] = e^{-\frac{1}{2}(s+t)}\mathbb{E}[W(e^s)W(e^t)] = e^{-\frac{1}{2}|s-t|}$. Note that X does not have independent increments property. However, it has the interesting property of *stationarity* or *shift-invariance*: Fix $t_0 \in \mathbb{R}$ and define $Y(t) = X(t_0 + t)$. Then, check that Y has the same distribution of X (you may use space-time scale invariance of W). In other words, for the process X the origin is not a special time-point, it is just like any other point.

3.5. Brownian bridge. Brownian bridge is the continuous Gaussian process $X = (X(t))_{t \in [0,1]}$ such that $\mathbb{E}[X_t X_s] = s(1-t)$ for $0 \leqslant s < t \leqslant 1$. Observe that $X(0) = X(1) = 0$ w.p.1. and $\mathbb{E}[X_0^2] = 0 = \mathbb{E}[X_1^2]$. It arises in many situations, but for now we simply motivate it as a possible model for a random surface in $1+1$ dimensions (the graph of X is to be thought of as a surface) that is pinned down at both endpoints.

How to prove the existence of Brownian bridge? Let W be a standard Brownian motion on $[0,1]$ and set $X(t) = W(t) - tW(1)$ for $0 \leqslant t \leqslant 1$. Check that X has the defining properties of Brownian bridge. This representation is also useful in working with Brownian bridge.

So far, we just applied transformations to Brownian motion. The next two are a little different, and will take more work to define.

3.6. Diffusions. Recall the physical motivation for Brownian motion as a particle in a fluid that is being bombarded on all sides by the molecules of the fluid. The mathematical definition that we have given assumes that the fluid is homogeneous in space and time (i.e., it is similar everywhere

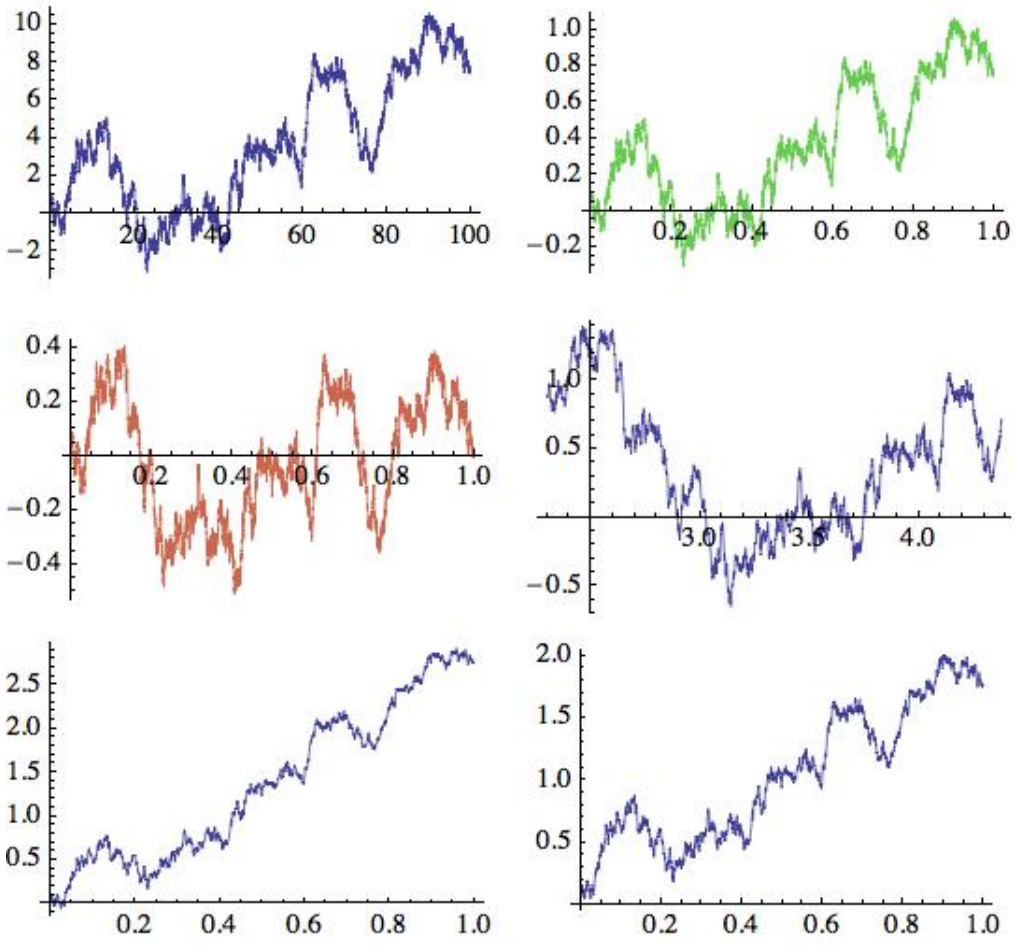

FIGURE 1. Top row left: Brownian motion run for time 100. Top row right: The same after a time-space scaling to time interval $[0, 1]$. Middle row left: A Brownian bridge. Middle row right: An Ornstein-Uhlenbeck sample path. Bottom row left: Brownian motion with linear drift $2t$. Bottom row right: $W_t + \sqrt{2t}$. Take note of the markings on both axes.

at all time) and the motion is isotropic (no preferred direction of motion, for example no gravity pulling the particle down). If one imagines motion in a non-homogeneous medium (assume that it remains homogenous in time), one arrives at the following kind of stochastic process.

For each $x \in \mathbb{R}^d$, let $m(x) \in \mathbb{R}^d$ and Σ_x be a positive definite $d \times d$ matrix. We want a $\mathbb{R}^d$-valued stochastic process $X = (X(t))_{t \geqslant 0}$ that has continuous sample paths, independent increments over disjoint intervals of time and such that conditional on $X(s)$, $s \leqslant t$, for small h, the distribution of $X(t+h) - X(t)$ is approximately Gaussian with mean vector $hm(X(t))$ and covariance matrix $h\Sigma_{X(t)}$. This last statement has to be interpreted in a suitable sense of $h \to 0$. Such a process is called a *diffusion*.

If $m(x) = 0$ and $\Sigma_x = I_d$, then we get back Brownian motion. If $m(x) = m$ (a constant vector) and $\Sigma_x = \Sigma$ (a constant matrix), then we can get such a process as $X(t) = tm + \Sigma^{\frac{1}{2}} W(t)$ where W is a standard d-dimensional Brownian motion. This is the higher dimensional version of Brownian motion with a drift and scale. A third and more interesting example is that of the Ornstein-Uhlenbeck process $X(t) = e^{-t/2} W(e^t)$, where W is a standard Brownian motion (let us stick to one-dimension for simplicity). Then

$$\begin{aligned} X(t+h) - X(t) &= (e^{-(t+h)/2} - e^{-t/2}) W(e^{t+h}) + e^{-t/2}(W(e^{t+h}) - W(e^t)) \\ &\approx -\frac{1}{2} h e^{-t/2} W(e^{t/2}) + e^{-t/2}(W(e^{t+h}) - W(e^t)) \end{aligned}$$

if h is small. Observe that if we have observed $X(s)$, $s \leqslant t$, the above says that in the next h (small) unit of time, the increment of X will be Gaussian with approximate mean $-\frac{1}{2} X(t) h$ and variance $e^{-t}(e^{t+h} - e^t) \approx h$. This fits the description of diffusion above with $m(x) = -\frac{1}{2} x$ and $\sigma^2(x) = 1$. The farther the process is from the origin, stronger is the drift and it is directed towards the origin (imagine a stretched spring).

But this description is still at a heuristic level. We shall only understand it if we can make sense of (i.e., construct) such processes for more general $m(\cdot)$ and $\sigma^2(\cdot)$. That is not easy, and forms the subject of Ito's theory of stochastic calculus.

3.7. New processes by conditioning. Let $W = (W_t)_{0 \leqslant t \leqslant 1}$ be a standard Brownian motion and let μ denote the Wiener measure on $C[0,1]$ (thus $W \sim \mu$). If $A \in \mathcal{B}(C[0,1])$ is any subset with $\mu(A) > 0$, we can define the conditional distribution of W given $W \in A$ as the probability measure on $C[0,1]$ given by

$$\mu_A(\bullet) = \frac{\mu(\bullet \cap A)}{\mu(A)}.$$

Of course, it lives on A in the sense that $\mu_A(A^c) = 0$. This for example can be done with $A = \{f : f(1) > 0\}$ (since $\mu(A) = \mathbb{P}\{W(1) > 0\} = \frac{1}{2}$) and $A = \{f \in C[0,1] : \|f\|_{\sup} \leqslant b\}$ for some $b > 0$ (it is easy to see that $\mu(A) > 0$ for large enough b, but not easy to see the same for all $b > 0$).

But the most interesting cases are when we condition on zero probability events such as $A = \{f : f(1) = 0\}$ (clear that $\mu(A) = \mathbb{P}\{W(1) = 0\} = 0$) or $A = \{f : f(t) \geqslant 0 \text{ for all } t \in [0,1]\}$ (it is not so clear that $\mu(A) = 0$ in this case!). In such cases, we are asking Brownian motion to do something it is unlikely to do on its own, and consequently the process we get will have rather different features.

The above definition of conditional probability does not work and we take recourse to regular conditional probabilities. For this it is not enough to have the event A, but a sigma algebra. For concreteness, assume that we can represent it as $A = M^{-1}\{x\}$ for some measurable function $M : C[0,1] \to \mathbb{R}$ and some $x \in \mathbb{R}$. Then (as $C[0,1]$ is a complete and separable metric space) there is a $\mathbb{Q} : \mathcal{B}(C[0,1]) \times \mathbb{R} \to [0,1]$ such that $\mathbb{Q}(\bullet, x)$ represents the conditional distribution $\mathbb{P}\{W \in$

$\bullet \,\big|\, M(W) = x\}$. Different choices of M and x can lead to different answers and care is needed to see that what we get is what we want. As $\mathbb{Q}(\bullet, x)$ is a probability measure on $C[0,1]$, it is determined completely by its finite dimensional distributions. In other words, it suffices to find the conditional distribution of $(W(t_1), \ldots, W(t_n))$ given $M(W) = x$, for any $n \geqslant 1$ and $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant 1$. But even that could be difficult.

In the very special case when M is a linear functional on $C[0,1]$, we see that $(W, M(W))$ is jointly Gaussian, and hence the conditional distributions are also Gaussian. This reduces our work to finding the mean $\mathbb{E}[W(t) \,\big|\, M(W) = x]$ and covariance $\mathbb{E}[W(t)W(s) \,\big|\, M(W) = x]$ for any $0 \leqslant s < t \leqslant 1$. This involves working with the 3-dimensional Gaussian vector $(W(s), W(t), M(W))$, which is often tractable.

Example 3: Brownian bridge

Let $A = \{f : f(1) = 0\} = \Pi_1^{-1}\{0\}$. In the notation above, this means that $M = \Pi_1$ and $x = 0$ and $M(W) = W(1)$. The regular conditional distribution $\mathbb{Q}(\bullet, x)$ of W given $W(1) = x$ is determined by its finite dimensional distributions. But for any $n \geqslant 1$ and $0 \leqslant t_1 < t_2 < \ldots < t_n < 1$, the random vector $(W(t_1), W(t_2), \ldots, W(t_n), W(1))$ has a multivariate Gaussian distribution. Hence the conditional distribution of $V = (W(t_1), W(t_2), \ldots, W(t_n))$ given $W(1) = x$ is also multivariate Gaussian, and we can use the formulas for the conditional means and conditional covariances of Gaussians. But the special form of the covariance allows us to shorten the process with a trick.

Let $Y = (W(t_1) - t_1 W(1), \ldots, W(t_n) - t_n W(1))$. Then Y is independent of $W(1)$ (as $(Y, W(1))$ are jointly Gaussian and $\mathbb{E}[Y_j W(1)] = 0$). But $V = Y + (t_1, \ldots, t_n)W(1)$, which show that the conditional distribution of V given $W(1) = x$ is the same as the unconditional distribution of $Y + (t_1, \ldots, t_n)x$. In particular, the conditional distribution of V given $W(1) = 0$ is the same as the unconditional distribution of Y. But Y is Gaussian with zero means and covariances $\mathbb{E}[Y_j Y_k] = t_j(1 - t_k)$ for $t_j < t_k$. Remembering that $t(1 - s)$ (for $t < s$) is the covariance of the Brownian bridge, we see that $\mathbb{Q}(\bullet, 0)$ is precisely the distribution of the Brownian bridge. Thus, the Brownian bridge may also be interpreted as W conditioned on $W(1) = 0$.

Exercise 3

Describe explicitly $\mathbb{Q}(\bullet, x)$, the conditional distribution of W given $W(1) = x$.

Exercise 4

Find the conditional distribution of the standard Brownian motion $W = (W_t)_{0 \leqslant t \leqslant 1}$ given that $\int_0^1 W(t)dt = 0$.

Exercise 5

Find the conditional distribution of the standard Brownian motion $W = (W_t)_{0 \leqslant t \leqslant 1}$ given that $W(0) = W(1)$.

4. Preview of some results to come

Accepting that it exists, what kind of a function is Brownian motion $W = (W_t)_{t \geqslant 0}$. Here is a non-exhaustive list of its almost sure properties, some of which are easy to prove and some difficult.

- W has no interval of increase or decrease: There does not exist $0 \leqslant a < b$ such that W is increasing (or decreasing) on (a, b).
- W has a dense set of local maxima (and minima): For any $0 \leqslant a < b$, there is point $t_\in(a, b)$ and a $\delta > 0$ such that $W(t_0) > W(t)$ if $t \neq t_0$ and $t \in (t_0 - \delta, t_0 + \delta)$.
- W is nowhere differentiable, and in fact nowhere Hölder$(\frac{1}{2} + \varepsilon)$ for any $\varepsilon > 0$: There does not exist $t_0 \geqslant 0$ for which $|W(t) - W(t_0)| \leqslant C|t - t_0|^{\frac{1}{2}+\varepsilon}$ for some $C < \infty$ and all $|t - t_0| < \delta$.
- W is Hölder$(\frac{1}{2} - \varepsilon)$ for all $\varepsilon > 0$, on any bounded interval.
- $\lim_{t\to\infty} \frac{W_t}{t} \overset{a.s.}{\to} 0$.
- $\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t \log\log t}} = 1$ and $\liminf_{t\to\infty} \frac{W_t}{\sqrt{2t \log\log t}} = -1$.
- If $0 = t_1^{(n)} < \ldots < t_{N_n}^{(n)} = t$ such that $\max_k |t_k^{(n)} - t_{k-1}^{(n)}| \to 0$ as $n \to \infty$, then
$$\sum_{k=1}^{n} (W(t_k^{(n)}) - W(t_{k-1}^{(n)}))^2 \overset{L^2, P}{\to} t.$$
- Range$(W) = \mathbb{R}$.
-

CHAPTER 2

Basic properties of Brownian motion

The natural thing to do next would be to prove the existence of Brownian motion, or equivalently of Wiener measure. However, all constructions of Brownian motion are a little involved. To make the point that in studying Brownian motion, we can directly appeal to its definition of Brownian rather than the details of its construction, we first prove several properties.

Throughout, let $W = (W_t)_{t\geqslant 0}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This means that

(a) $W_0 = 0$ a.s. (b) $W_t - W_s \sim N(0, t-s)$ for any $0 \leqslant s < t$. (c) Increments over pairwise disjoint intervals are independent. (d) W has continuous sample paths.

or equivalently,

(A) W_t, $t \geqslant 0$, are jointly Gaussian. (B) $\mathbb{E}[W_t] = 0$ for all t. (C) $\mathbb{E}[W_t W_s] = s$ for any $s < t$. (D) W has continuous sample paths.

1. Symmetries of Brownian motion

Let W be standard Brownian motion and let μ_W denote the Wiener measure. By a symmetry, we mean a transformation $T : C[0,\infty) \to C[0,\infty)$ such that $\mu_W \circ T^{-1} = \mu_W$ or in the language of random variables, $T(W) \stackrel{d}{=} W$. Brownian motion has many symmetries, some of which we mention now.

- (Reflection symmetry). $T(f) = -f$. That is, if $X_t = -W_t$, then X is standard Brownian motion. To see this, observe that X has continuous sample paths as W does, and that X_t are jointly Gaussian with zero means. It is also clear that $\mathbb{E}[X_t X_s] = \mathbb{E}[W_t W_s] = s$ if $s < t$.
- (Space-time scaling symmetry). Let $\alpha > 0$ and define $[T(f)](t) = \frac{1}{\sqrt{\alpha}} f(\alpha t)$. That is, if $X_t = \frac{1}{\sqrt{\alpha}} W_{\alpha t}$, then X is a standard Brownian motion. Again, it is obvious that X has continuous sample paths and that X_t are jointly Gaussian with zero means. Lastly, for $s < t$,
$$\mathbb{E}[X_t X_s] = \frac{1}{\alpha}\mathbb{E}[W_{\alpha t} W_{\alpha s}] = \frac{1}{\alpha}(\alpha s) = s.$$
- (Time-reversal symmetry) Let W be standard Brownian motion on $[0,1]$. Define $X(t) = W(1-t) - W(1)$ for $0 \leqslant t \leqslant 1$. Then X is standard Brownian motion on $[0,1]$. Continuity of sample paths, joint Gaussianity, and zero mean properties are all clear. The covariance

is

$$\begin{aligned}\mathbb{E}[X_t X_s] &= \mathbb{E}[W_{1-t}W_{1-s}] - \mathbb{E}[W_{1-t}W_1] - \mathbb{E}[W_1 W_{1-s}] + \mathbb{E}[W_1^2] \\ &= (1-t) - (1-t) - (1-s) + 1 \;=\; s.\end{aligned}$$

- (Time-inversion symmetry). Define $X_t = tW_{1/t}$ for $t \in (0,\infty)$. Then X_t are jointly Gaussian, continuous in t w.p.1., and for $s < t$ we have

$$\mathbb{E}[X_t X_s] = ts\mathbb{E}\left[W\left(\frac{1}{t}\right)W\left(\frac{1}{s}\right)\right] = ts\frac{1}{t} = s.$$

  Thus, $(X_s)_{s\in(0,\infty)}$ has the same distribution as $(W_s)_{s\in(0,\infty)}$. In particular, the random variables

$$M_k^X = \sup_{0<s\leqslant\frac{1}{k}} X_s \text{ and } M_k^W = \sup_{0<s\leqslant\frac{1}{k}} W_s$$

  have the same distribution. But $\lim_{k\to\infty} M_k^W = 0$ w.p.1., and hence $\lim_{k\to\infty} M_k^X = 0$ w.p.1. But that precisely means that $\lim_{t\to 0+} X(t) = 0$ w.p.1. The upshot is that if we set $X_0 = 0$, then X is standard Brownian motion.

- (Time-shift symmetry). Let $t_0 \geqslant 0$ and define $[Tf](t) = f(t+t_0) - f(t_0)$. That is, if $X_t = W_{t+t_0} - W_{t_0}$, then X is standard Brownian motion. Joint Gaussianity and continuity are clear. As for covariances, for $s < t$ we get

$$\begin{aligned}\mathbb{E}[X_t X_s] &= \mathbb{E}[W_{s+t_0}W_{t+t_0}] - \mathbb{E}[W_{t_0}W_{t+t_0}] - \mathbb{E}[W_{s+t_0}W_{t_0}] + \mathbb{E}[W_{t_0}W_{t_0}] \\ &= (s+t_0) - t_0 - t_0 + t_0 \\ &= s.\end{aligned}$$

  Thus X is a standard Brownian motion. Whether the time-shift invariance holds at random times t_0 is an important question that we shall ask later.

2. First remarks on the sample path properties of Brownian motion

By definition, the trajectories $t \mapsto W_t(\omega)$ are continuous for a.e. ω. Beyond that, Brownian motion is quite wild. For a first jolt, consider the following proposition. For reasons to be revealed later, we shall assume that the $(\Omega, \mathcal{F}, \mathbb{P})$ on which our Brownian motion $W = (W_t)_{t\geqslant 0}$ is defined is a complete probability space. That is, if $A \in \mathcal{F}$ and $\mathbb{P}(A) = 0$, then any subset $B \subseteq A$ is also in $\mathcal{F}$ (and then of course $\mathbb{P}(B) = 0$).

Proposition 1

Almost surely, there is no interval on which W is monotone.

Proof. Fix $a < b$ and consider an infinite sequence of points $a < t_1 < t_2 < \ldots < b$. Then $W(t_{k+1}) - W(t_k)$ are independent Gaussian random variables with zero means and strictly positive

variances. Clearly

$$\{W \text{ is increasing on } (a,b)\} \subseteq \bigcap_{k=1}^{\infty} \{W(t_{k+1}) - W(t_k) \geqslant 0\}.$$

The set on the right is a countable intersection of measurable sets and hence measurable. The events in the intersection are independent, and each has probability equal to $\frac{1}{2}$. Hence the intersection has probability 0.

By the assumption of completeness, it follows that the event on the left is also measurable and has probability 0. This was for fixed $a < b$. Take countable union over all rational $a < b$ to see that $\mathbb{P}\{W \text{ is increasing on some interval}\} = 0$ (because if it is increasing on any interval, it is increasing on an interval with rational endpoints). Similarly, $\mathbb{P}\{W \text{ is decreasing on some interval}\} = 0$. The theorem is proved. ■

Try to draw a function that is not monotone on any interval, you may have some difficulty! This brings out how wild the sample paths of Brownian motion are. Here is a corollary of the proposition. Recall that a point t_0 is a (weak) local maximum of f if $f(t_0) = \max\{f(t) : t \in [t_0 - \varepsilon, t_0 + \varepsilon]$ for some $\varepsilon > 0$.

Corollary 1

Almost surely, the set of local maxima of W is dense in $\mathbb{R}_+$. The same is true of local minima.

Proof. Let f be a continuous function that is not monotone on any interval. Fix $a < b$. As f is not increasing, there must be $a < s < t < b$ such that $f(t) < f(s)$. As f is not decreasing on (a,t), there must be $u \in (a,t)$ such that $f(u) < f(t)$. Then it is clear that the maximum of f on $[u,s]$ is attained inside (u,s). That point is a local maximum of f. Similarly (or applying this to $-f$) shows that f must also have a local minimum in (a,b).

Almost surely, W is not monotone on any interval, and therefore it has a local maximum and a local minimum in (a,b), a.s. Take intersection of these events over rational $a < b$ to see that almost surely, W has a local maximum and a local minimum in every interval. ■

Remark 1: About the completeness assumption

In the proof of Proposition 1 we used completeness to avoid the question of whether the set "there is an interval of time over which W is monotone" is measurable or not. In this case it was possible to show measurability by writing this set as $W^{-1}(A)$ where

$$A := \bigcap_{\substack{a<b \\ a,b\in\mathbb{Q}}} \bigcap_{\substack{s<t \\ s,t\in\mathbb{Q}\cap(a,b)}} \{f \in C[0,\infty) : f(t) - f(s) \geqslant 0\}$$

which is clearly a Borel set in $C[0,\infty)$. However, in many situations that we shall later encounter, it will be either difficult or not even true that a set is measurable, but we can still show that it is contained inside a measurable set of zero measure. Making the completeness assumption just makes it possible to avoid these unimportant hurdles.

3. Continuity properties of Brownian motion

Let W denote standard Brownian motion in $[0,1]$. The key point to remember is that for fixed t, h, we have $W(t+h) - W(t) \stackrel{d}{=} \sqrt{h}Z$, where $Z \sim N(0,1)$. Thus, Brownian motion travels a distance of order $\sqrt{h}$ in short times h. Contrast this with smooth functions for which $g(t+h) - g(t) = hg'(t_*)$ where $t_* \in [t, t+h]$. As $\sqrt{h}$ is way larger than h, Brownian motion oscillates much more than smooth functions. In fact, one is tempted to say that W is approximately Hölder$(\frac{1}{2})$. Many of the properties we shall prove are just more precise and stronger forms of this heuristic, and the proofs are often non-trivial. But this basic point should always be kept in mind to understand how Brownian motion behaves.

What are the difficulties in making this heuristic into a proof? Using the properties of $N(0,1)$ distribution, we see that

$$\mathbb{P}\{\varepsilon\sqrt{h} \leqslant W(t+h) - W(t) \leqslant \frac{1}{\varepsilon}\sqrt{h}\} \geqslant 1 - \varepsilon$$

for small enough ε. But this is for a fixed t, h. As there are many choices of t, h, it is very much probable that there are intervals with abnormally large or abnormally small increments. A slightly more refined argument is that if we fix h, then there are only about $1/h$ many independent increments of W in the time interval $[0,1]$, e.g., $\Delta W([kh,(k+1)h])$, $0 \leqslant k < \frac{1}{h}$. Any other interval of length h overlaps two of these intervals. Therefore we may expect that

$$\max\{\Delta W([t,t+h]) : 0 \leqslant t \leqslant 1-h\} \approx \max\{\Delta W([kh,(k+1)h] : 0 \leqslant k < \frac{1}{h}\}. \tag{3}$$

This is not a rigorous claim because even if we break $[t,t+h]$ into two intervals $[t,kh]$ and $[kh,t+h]$ for some k, the increments on these subintervals can be larger than the increments on the full intervals $[(k-1)h,kh]$ and $[kh,(k+1)h]$ respectively. However, as the variance of the increment

is proportional to the length, in a probabilistic sense smaller intervals should have smaller increments. Let us accept this heuristic for now.

The maximum of the right side of (3) has the same distribution as $\sqrt{h}\max\{Z_1, \ldots, Z_{1/h}\}$, where Z_k are i.i.d. $N(0,1)$. One of the fundamental facts about Gaussians is that the maximum of n i.i.d. standard Gaussians is of the order of $\sqrt{\log n}$ (more precisely, it is about $\sqrt{2\log n}$). Consequently, $\max\{\Delta W([kh,(k+1)h])\}$ is of the order of $\sqrt{h\log\frac{1}{h}}$ (or the more precise $\sqrt{2h\log\frac{1}{h}}$). Note that this is of a larger order than $\sqrt{h}$, but not by too much as $\log\frac{1}{h}$ grows slowly as $h \to 0$. For instance this is smaller than $h^{\frac{1}{2}-\varepsilon}$ for any $\varepsilon > 0$, provided h is small enough.

This reasoning makes the following results looks somewhat natural. First some notation: For $f \in C[0,1]$, define its *modulus of continuity* as the function $\omega_f(h) = \max\{|f(t) - f(s)| : |t-s| \leqslant h\}$. Uniform continuity of f is the statement that $\omega_f(h) \downarrow 0$ as $h \downarrow 0$. The rate at which it approaches zero says something about the smoothness of the function. For example, if $f \in C^1$, then $\omega_f(h) \leqslant Ch$ where $C = \|f'\|_{\sup}$. If $\omega_f(h) \leqslant Ch^\gamma$, we say that f is a Hölder(γ) function.

Theorem 1: Modulus of continuity of Brownian motion

Let W be standard Brownian motion on $[0,1]$. Almost surely

(1) (Paley-Wiener-Zygmund) W is a Hölder(γ) function for every $\gamma < \frac{1}{2}$, a.s.

(2) $\omega_W(h) \leqslant C\sqrt{h\log\frac{1}{h}}$ for a finite random variable C.

(3) (Lévy) $\lim_{h\downarrow 0} \frac{\omega_W(h)}{\sqrt{2h\log\frac{1}{h}}} = 1$ a.s.

The first statement is equivalent to saying that $\omega_W(h) \leqslant C_\varepsilon h^{\frac{1}{2}-\varepsilon}$ for a finite random variable C_ε, for any $\varepsilon > 0$. Then it is clear that it follows from the second statement, since $h^{-\varepsilon}$ is much larger than $\sqrt{\log\frac{1}{h}}$. The third statement implies the second, because the quantity

$$\sup_{h \geqslant h_0} \frac{\omega_W(h)}{\sqrt{2h\log\frac{1}{h}}}$$

is obviously finite (but random) for any $h_0 > 0$. The issue of finiteness of C comes from $h \downarrow 0$.

The natural question is whether the continuity of Brownian motion is better than what is shown by Theorem 1. For example, is it Hölder(γ) for $\gamma \geqslant \frac{1}{2}$? Since even for a single interval, $\Delta W([t, t+h])$ is of order $\sqrt{h}$, we do not expect this to be true and it is even easy to show it (the heuristic above is in fact rigorous on this side, the right side of (3) is of the order of $\sqrt{h\log\frac{1}{h}}$). But this does not prepare us for the stunning conclusion of the following result, which says that the trajectory is wild everywhere.

If $f : [0,1] \to \mathbb{R}$ and $0 < \alpha \leqslant 1$, we say that t is a Hölder(α) point for f if

$$\limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h^\alpha} < \infty.$$

If the lim sup on the left is less than or equal to c, then we say that t is a Hölder$(\alpha; c)$ point (then it is also a Hölder$(\alpha; c')$ point for any $c' > c$). Observe that if f is differentiable at t, then t is a Hölder(1) point.

Theorem 2: Paley, Wiener, Zygmund

With probability 1, the following statements hold.

(1) Brownian motion is nowhere differentiable.

(2) Brownian motion is nowhere Hölder(α) for any $\alpha > \frac{1}{2}$.

(3) If $c < 0.3$, then Brownian motion has no Hölder$(\frac{1}{2}; c)$ points.

In a first reading, it suffices to understand the proofs of the first two statements in both Theorem 1 and Theorem 2.

4. Proof of the theorem on Hölder continuity of Brownian motion

These statements are increasingly stronger, hence it suffices to prove the last one. The usual proof given in all books for the first two statements is a very elegant one due to Dvoretsky, Erdös and Kakutani. As far as I can see, that method cannot prove the third. I went back to the original proof of Paley, Wiener and Zygmund, and found that their proof, also very elegant, in fact gives the third statement! However, historically, it appears that such a statement only appeared much later in a paper of Dvoretsky, who proved the even stronger statement that Hölder$(\frac{1}{2}; c)$ points exist if and only if $c > 1$. I am a little confused by the history but anyway...

Proof of nowhere differentiability due to Dvoretksy, Erdös and Kakutani. If f is differentiable at t, then $|f(s) - f(t)| \leqslant C|s - t|$ for some $C < \infty$ and for all $s \in [0,1]$. Then, $|f(s) - f(u)| \leqslant C(|s-t| + |u-t|)$ for all $s, u \in [0,1]$. In particular, for any $n \geqslant 0$ and any $0 \leqslant k \leqslant 2^n - 1$, this holds when we take $s = k2^{-n}$ and $u = (k+1)2^{-n}$. In particular, if ℓ is such that $[\ell 2^{-n}, (\ell+1)2^{-n}] \ni t$, then this holds for $k = \ell + j$, $j = 1, 2, 3$, or for $k = \ell - j$, $j = 1, 2, 3$ (if t is too close to 1 then $\ell + 3$ may be greater than $2^n - 1$ and if t is too close to 0, $\ell - 3$ may be less than 0, hence we consider both possibilities). For such k, we get

$$\left| f\left(\frac{k+1}{2^n}\right) - f\left(\frac{k}{2^n}\right) \right| \leqslant C\frac{10}{2^n} \tag{4}$$

since $k2^{-n}$ and $(k+1)2^{-n}$ are all within distance 5.2^{-n} of t. Thus, if we define

$$\mathcal{A} = \{f : f \text{ is differentiable at some } t \in [0,1]\},$$

$$\mathcal{A}_{n,C} = \{f : (4) \text{ holds for at least three consecutive } k \text{ in } 0, 1, \ldots, 2^n - 1\},$$

then what we have shown is that $\mathcal{A} \subseteq \bigcup_{C=1}^{\infty} \bigcap_{n=1}^{\infty} \mathcal{A}_{n,C}$.

We show for each fixed C that $\mathbb{P}\{W \in \mathcal{A}_{n,C}\} \to 0$ as $n \to \infty$. This implies[1] that $\mathbb{P}\{W \in \mathcal{A}\} = 0$. To show this,

$$\begin{aligned}
\mathbb{P}\{W \in \mathcal{A}_n\} &= \sum_{\ell=0}^{2^n-3} \mathbb{P}\{(4) \text{ holds for } f = W \text{ for } k = \ell, \ell+1, \ell+2\} \\
&\leqslant (2^n - 2) \left(\mathbb{P}\left\{ |\xi| \leqslant \frac{10C}{\sqrt{2^n}} \right\} \right)^3 \\
&\leqslant (2^n - 2) \left(\frac{1}{\sqrt{2\pi}} \frac{10C}{\sqrt{2^n}} \right)^3 \\
&\leqslant 10^3 C^3 \frac{1}{\sqrt{2^n}}.
\end{aligned}$$

This proves the nowhere differentiability of Brownian motion. ■

By considering several increments in place of three, one can show that W has no Hölder$(\frac{1}{2}+\varepsilon)$ points. If $f : [0,1] \to \mathbb{R}$, we say that $t \in [0,1]$ is an α-Hölder point of f if there exists $C < \infty$ such that $|f(s) - f(t)| \leqslant C|t-s|^\alpha$ for all $s \in [0,1]$.

Exercise 6

Let $W = (W_t)_{0 \leqslant t \leqslant 1}$ be a standard Brownian motion. Show that for a.e.$[\omega]$, the function $t \mapsto W_t(\omega)$ has no Hölder(α) points for any $\alpha > \frac{1}{2}$.

Hölder$(\frac{1}{2}; c)$ points: Next we adapt the original proof of Paley, Wiener and Zygmund to show that there are no Hölder$(\frac{1}{2}; c)$ points if c is small. For convenience of notation, let $\Delta f(I) = f(b) - f(a)$ for $f : [0,1] \mapsto \mathbb{R}$ and $I = [a,b]$ a subinterval of $[0,1]$. Also, let $I_{n,k} = [k2^{-n}, (k+1)2^{-n}]$ for $n \geqslant 0$ and $0 \leqslant k \leqslant 2^n - 1$.

A branching process proof due to Paley, Wiener and Zygmund. Let t is a Hölder$(\frac{1}{2}; c)$ point, then there exists $M < \infty$ such that $|f(s) - f(t)| \leqslant c\sqrt{|s-t|}$ for all $s \in [t - 2^{-M}, t + 2^{-M}]$. In

[1] One issue: Is $\mathcal{A}$ a Borel subset of $C[0,1]$?! It is, but we don't bother to prove it. Instead, let us always work with the completion of Wiener measure. In other words, if $\mathcal{A}_1 \subseteq \mathcal{A}_0 \subseteq \mathcal{A}_2$ and $\mathcal{A}_1$ and $\mathcal{A}_2$ are Borel and $\mathbb{P}\{W \in \mathcal{A}_1\} = \mathbb{P}\{W \in \mathcal{A}_2\}$, then the same is deemed to be the value of $\mathbb{P}\{W \in \mathcal{A}_0\}$.

particular, if $n \geqslant M$ and $I_{n,k} = [k2^{-n}, (k+1)2^{-n}]$ is the dyadic interval that contains t, then

$$|\Delta f(I_{n,k})| \leqslant c\left\{\sqrt{(k+1)2^{-n} - t} + \sqrt{t - k2^{-n}}\right\} \leqslant \frac{\sqrt{2}c}{\sqrt{2^n}}. \tag{5}$$

In the last inequality we used the elementary fact that if $0 \leqslant x \leqslant a$, then $\sqrt{x} + \sqrt{a-x} \leqslant \sqrt{2a}$.

The collection of dyadic intervals carries a natural tree structure with $I_{0,0}$ being the root vertex and by declaring $I_{n+1,\ell}$ as a child of $I_{n,k}$ if $I_{n+1,\ell} \subseteq I_{n,k}$. This is a tree where each vertex has two children. Let us declare a dyadic interval $I_{n,k}$ to be alive if it satisfies $\Delta f(I_{n,k}) \leqslant c\sqrt{2}/\sqrt{2^n}$. Thus, if t is a Hölder$(\frac{1}{2}; c)$ point, then for some M, the tree beyond generation M has an infinite chain of descendents that are all alive (namely the dyadic intervals containing the point t).

The process of vertices alive is a Branching process that we shall prove will become extinct with probability 1. To do this, let $\mathcal{F}_n = \{\Delta W(I_{n,k}) : 0 \leqslant k \leqslant 2^n - 1\}$ so that these sigma-algebras are increasing. Whether an interval $I_{n,k}$ is alive or not is an event in $\mathcal{F}_n$. Condition on $\mathcal{F}_{n-1}$ and consider any live individual I in the $(n-1)$st generation. It has two children J, J' in the nth generation. Conditional on $\mathcal{F}_{n-1}$, we know the sum $\Delta W(J) + \Delta W(J') = \Delta W(I)$. From Exercise 4 we can write $\Delta W(J) = \frac{1}{2}\Delta W(I) + \frac{\xi}{\sqrt{2^{n+1}}}$ and $\Delta W(J') = \frac{1}{2}\Delta W(I) - \frac{\xi}{\sqrt{2^{n+1}}}$ where $\xi \sim N(0,1)$ is independent of $\mathcal{F}_{n-1}$. Now, J is alive if and only if $|\Delta W(J)| \leqslant \frac{c\sqrt{2}}{\sqrt{2^n}}$. This means that ξ must lie in an interval of length $4c$ centered at $\sqrt{2^{n-1}}\Delta W(I)$. By Exercise 4, irrespective of the value of $\Delta W(I)$, this probability is at most $4c/\sqrt{2\pi}$.

In summary, the expected number of offsprings of I is at most $\lambda = 8c/\sqrt{2\pi}$. If $c' < 1$, then the number of descendants of an interval $I_{M,k}$ in the generation $M + j$ is exactly λ^j. Thus the expected total number of live individuals live in the $M + j$ generation is $2^M\lambda^j$ which goes to zero as $j \to \infty$, provided $\lambda < 1$. Hence, for $c < \frac{\sqrt{2\pi}}{8} = 0.313\ldots$, the branching process goes extinct with probability 1.

Since this is true for every M, taking a countable union over positive integer M, it follows that for any $c < 0.31$, with probability 1, Brownian motion has no Hölder$(\frac{1}{2}; c)$ points. ■

We used two simple facts about Gaussian distribution in the proof. They are left as exercises.

Exercise 7

Let X, Y be i.i.d. $N(0,1)$. Then, the conditional distribution of (X, Y) given $X + Y = t$ is the same as the (unconditional) distribution of $(\frac{1}{2}t + \frac{1}{\sqrt{2}}\xi, \frac{1}{2}t - \frac{1}{\sqrt{2}}\xi)$ where $\xi \sim N(0,1)$.

Exercise 8

If $\xi \sim N(0,1)$, then $\sup_{a\in\mathbb{R}} \mathbb{P}\{\xi \in [a-t, a+t]\} \leqslant \frac{2t}{\sqrt{2\pi}}$.

5. Summary of continuity properties

Let W be standard Brownian motion on $[0,1]$. First and foremost is the point that $\mathbb{E}[|W_t - W_s|^2] = |t-s|$ from which we see that $W_{t+h} - W_t$ should behave like $\sqrt{h}$, typically. A summary of the basic continuity results is as follows.

(1) Almost surely $\limsup_{h\downarrow 0} \max_{t\in[0,1]} \frac{|W_t - W_s|}{\sqrt{h\log(1/h)}} < \infty$. We showed this (and if you follow our proof closely, you will see that the left hand side can be shown to be $\leqslant 10$ w.p.1.).

We did not show Paul Lévy's sharp result that in fact

$$\max_{t\in[0,1]} \limsup_{h\downarrow 0} \frac{|W_t - W_s|}{\sqrt{h\log(1/h)}} = \sqrt{2} \quad \text{a.s.}$$

(2) Almost surely W has no Hölder$(\frac{1}{2}; c)$ points for c sufficiently small. As a consequence, it is nowhere Hölder$(\frac{1}{2}+\varepsilon)$ and in particular, nowhere differentiable.

We showed this. We did not show the results of Dvoretzky (and Kahane?) that the sharp constant is 1. That is, for $c < 1$, there do not exist Hölder$(\frac{1}{2}; c)$ points while for $c > 1$, they do exist.

(3) We shall show later that at a fixed point, the continuity is faster than $\sqrt{h}$ and slower than $\sqrt{h\log(1/h)}$. This is the celebrated law of iterated logarithm which asserts that for any fixed $t \geqslant 0$,

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

In fact the set of limit points of $\frac{W(t+h)-W(t)}{\sqrt{2h\log\log(1/h)}}$ as $h \downarrow 0$ is almost surely equal to $[-1,1]$.

CHAPTER 3

Constructions of Brownian motion

1. Chaining method and the first construction of Brownian motion

We want to construct random variables W_t, indexed by $t \in \mathbb{R}_+$, that are jointly Gaussian and such that $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t W_s] = t \wedge s$. Here is the sketch of how it is done by the so called chaining method of Kolmogorov and Centsov.

(1) From the observation that $t \wedge s = \langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle_{L^2(\mathbb{R})}$, it follows that for any $n \geqslant 1$ and $t_1, \ldots, t_n$, the matrix $(t_i \wedge t_j)_{i,j \leqslant n}$ is a Gram matrix and hence positive semi-definite. Therefore, $\mu_{t_1,\ldots,t_n} := N(0, (t_i \wedge t_j)_{i,j \leqslant n})$ is a well-defined distribution.

(2) Let $D = \{t_1, t_2, \ldots\}$ be a countable dense subset of $[0,1]$. The family of probability distributions $\mu_{t_1,\ldots,t_n}$, $t_1 < \ldots < t_n$, $n \geqslant 1$, is consistent. Therefore, we can construct W_t, $t \in D$ (on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$) having a joint Gaussian distribution with zero means and covariance $t \wedge s$.

(3) We show that the function $t \mapsto W_t(\omega)$ is uniformly continuous for $\mathbb{P} - a.e.\ \omega$. This is the key step.

(4) By standard real analysis, this means that for each such ω, the function $t \mapsto W_t(\omega)$ extends to a continuous function on $[0,1]$. Since limits of Gaussians are Gaussian, the resulting W_t, $t \in [0,1]$, have joint Gaussian distribution with the prescribed covariances.

(5) We concatenate i.i.d. copies of Brownian motion on $[0,1]$ to get Brownian motion on $[0,\infty)$.

Actually our construction will give more information about the continuity properties of Brownian motion. We start with some basic real analysis issues.

Let $D \subseteq [0,1]$ be a countable dense set and let $f : [0,1] \mapsto \mathbb{R}$ be given. We say that f extends continuously to $[0,1]$ if there exists $F \in C[0,1]$ such that $F\big|_D = f$. Clearly, a necessary condition for this to be possible is that f be uniformly continuous on D to start with. It is also sufficient. Indeed, a uniformly continuous function maps Cauchy sequences to Cauchy sequences, and hence, if $t_n \in D$ and $t_n \to t \in [0,1]$, then $(t_n)_n$ is Cauchy and hence $(f(t_n))_n$ is Cauchy and hence $\lim f(t_n)$ exists. Clearly, the limit is independent of the sequence $(t_n)_n$. Hence, we may define $F(t) = \lim_{D \ni s \to t} f(s)$ and check that it is the required extension.

But we would like to prove a more quantitative version of this statement. Recall that the *modulus of continuity* of a function $f : [0,1] \to \mathbb{R}$ is defined as $w_f(\delta) = \sup\{|f(t) - f(s)| : |t - s| \leqslant \delta\}$. Clearly, f is continuous if and only if $w_f(\delta) \downarrow 0$ as $\delta \downarrow 0$. The rate at which $w_f(\delta)$ decays to 0 quantifies the level of continuity of f. For example, if f is Lipschitz, then $w_f(\delta) \leqslant C_f\delta$ and if f is Hölder(α) for some $0 < \alpha \leqslant 1$, then $w_f(\delta) \leqslant C_f\delta^\alpha$. For example, t^α is Hölder(α) (and not any better) on $[0,1]$.

Henceforth, we fix the countable dense set to be the set of dyadic rationals, i.e., $D = \bigcup_n D_n$ where $D_n = \{k2^{-n} : 0 \leqslant k \leqslant 2^n\}$.

Lemma 1: Kolmogorov-Centsov

Let $f : D \to \mathbb{R}$. Let Define $\Delta_n(f) = \max\{|f(\frac{k+1}{2^n}) - f(\frac{k}{2^n})| : 0 \leqslant k \leqslant 2^n - 1\}$. Assume that $\sum_n \Delta_n(f) < \infty$. Then, f extends to a continuous function on $[0,1]$ (we continue to denote it by f) and $w_f(\delta) \leqslant 10 \sum_{n \geqslant m_\delta} \Delta_n(f)$ where $m_\delta = \lfloor \log_2(1/\delta) \rfloor$.

Assuming the lemma, we return to the construction of Brownian motion.

Construction of Brownian motion. First construct W_t, $t \in D$, that are jointly Gaussian with zero means and covariance $t \wedge s$. Then, $W(\frac{k+1}{2^n}) - W(\frac{k}{2^n})$, $0 \leqslant k \leqslant 2^n - 1$, are i.i.d. $N(0, 2^{-n})$. Hence, by the tail estimate of the Gaussian distribution,

$$\mathbb{P}\left\{\Delta_n(f) \geqslant 2\frac{\sqrt{n}}{\sqrt{2^n}}\right\} \leqslant 2^n \mathbb{P}\left\{|\xi| \geqslant 2\sqrt{n}\right\} \leqslant 2^n \exp\left\{-\frac{1}{2}(4n)\right\} \leqslant 2^{-n}.$$

By the Borel-Cantelli lemma, it follows that $\Delta_n \leqslant 2\frac{\sqrt{n}}{\sqrt{2^n}}$ for all $n \geqslant N$ for some random variable N that is finite w.p.1. If $N(\omega) < \infty$, then we can se a large constant $C(\omega)$ to take care of $\Delta_n(W_\bullet(\omega))$ for $n \leqslant N(\omega)$ and write

$$\Delta_n(W_\bullet(\omega)) \leqslant C(\omega)\frac{\sqrt{n}}{\sqrt{2^n}} \text{ for all } n \geqslant 1$$

for a random variable C that is finite w.p.1.

Fix any ω such that $C(\omega) < \infty$. Then, by the lemma, we see that $(W_t(\omega))_{t \in D}$ extends continuously to a function $(W_t(\omega))_{t \in [0,1]}$ and that the extension has modulus of continuity

$$w(\delta) \leqslant \sum_{n \geqslant m_\delta} \frac{\sqrt{n}}{\sqrt{2^n}} \leqslant 10C(\omega)\frac{\sqrt{m_\delta}}{\sqrt{2^{m_\delta}}} \leqslant C'(\omega)\sqrt{\delta \log \frac{1}{\delta}}$$

using $m_\delta = \lfloor \log_2(1/\delta) \rfloor$. This shows that w.p.1., the extended function $t \mapsto W_t$ is not only uniformly continuous but has modulus of continuity $O(\sqrt{\delta}\sqrt{\log(1/\delta)})$.

It remains to check that the extended function has joint Gaussian distribution with the desired covariances. If $0 \leqslant t_1 < \ldots < t_m \leqslant 1$, then find $t_{i,n} \in D$ that converge to t_i, for $1 \leqslant i \leqslant m$. Then $(W_{t_{1,n}}, \ldots, W_{t_{m,n}}) \stackrel{a.s.}{\to} (W_{t_1}, \ldots, W_{t_m})$. But $(W_{t_{1,n}}, \ldots, W_{t_{m,n}})$ has joint Gaussian distribution. Hence, after taking limits, we see that $(W_{t_1}, \ldots, W_{t_m})$ has joint Gaussian distribution. In addition, the

covariances converge, hence

$$\mathbb{E}[W_{t_1}W_{t_2}] = \lim_{n\to\infty} \mathbb{E}[W_{t_{1,n}}W_{t_{2,n}}] = \lim_{n\to\infty} t_{1,n} \wedge t_{2,n} = t_1 \wedge t_2.$$

Thus, W_t, $t \in [0,1]$ is the standard Brownian motion (indexed by $[0,1]$, extension to $[0,\infty)$ is simple and will be shown later). ■

It only remains to prove the lemma. Before that it is good to see what does not work. The first thought would be to take dyadics $t = \frac{k}{2^N}$ and $s = \frac{\ell}{2^M}$ and if $M > N$, then write $[t,s]$ as a union of $|s-t|2^M$ many dyadic intervals of length 2^{-M} to get $|f(t)-f(s)| \leqslant |s-t|2^M\Delta_M(f)$. This is no good because M can be arbitrarily large (and if we expect Hölder(γ) continuity, then $\Delta_M(f)$ is about $2^{-M\gamma}$, which is overwhelmed by the 2^M factor as $\gamma < 1$).

This suggests that the right approach is to fill up $[s,t]$ with *large* dyadic intervals to the extent possible. One can write the proof that way, by choosing dyadic rationals $\ldots r_{-1} < r_0 < r_1 \ldots$ as follows: First pick r_0 as the dyadic in $[s,t]$ with the smallest denominator. Then proceed to find $r_0 < r_1 < r_2 < \ldots \leqslant t$ successively as the dyadics with the smallest denominators available. After a finite number of choices $r_k = t$. Similarly pick $r_0 > r_{-1} > r_{-2} > \ldots \geqslant s$. Now write $f(t) - f(s)$ as the sum over $f(r_j) - f(r_{j-1})$. This approach can be carried out, (Exercise!) but the presentation below looks a little different on the surface.

Proof. Fix dyadic rationals $t = \frac{k}{2^N}$ and $s = \frac{\ell}{2^M}$ and let p be such that $2^{-p-1} \leqslant |t-s| \leqslant 2^{-p}$ (without loss of generality we take $N, M \geqslant p$ since $D_m \subseteq D_{m+1}$). Fix $t_k \in D_k$ such that $|t - t_k| \leqslant 2^{-k}$ and observe that $t_N = t$ (we think of t_k as approximations to t in D_k). Define $s_k = t_k$ for $k \leqslant p$ and such that $|s_k - s| \leqslant 2^{-k}$ for $k > p$. We see that for $k \leqslant p$,

$$|s_k - s| \leqslant |t_k - t| + |t - s| \leqslant 2^{-k} + 2^{-p} \leqslant 2^{-k+1}.$$

Thus, $|s_k - s| \leqslant 2^{-k+1}$ for all k and $s_M = s$. Now write

$$f(t) = \sum_{j=1}^{N} f(t_j) - f(t_{j-1}) \qquad \text{and} \qquad f(s) = \sum_{j=1}^{M} f(s_j) - f(s_{j-1}).$$

When we take the difference, the first p terms cancel as $s_j = t_j$ for $j \leqslant p$ and we get

$$|f(t) - f(s)| \leqslant \sum_{j=p+1}^{N} |f(t_j) - f(t_{j-1})| + \sum_{j=p+1}^{M} |f(s_j) - f(s_{j-1})|. \tag{6}$$

Now observe that

$$|t_j - t_{j-1}| \leqslant |t_j - t| + |t_{j-1} - t| \leqslant 2^{-j+1},$$

$$|s_j - s_{j-1}| \leqslant |s_j - s| + |s_{j-1} - s| \leqslant 2^{-j+2}.$$

As t_j, t_{j-1} are both in D_j, we can write $f(t_j) - f(t_{j-1})$ as a sum of increments of at most two dyadic intervals of the form $[i/2^j, (i+1)/2^j]$ and hence $|f(t_j) - f(t_{j-1})| \leqslant 2\Delta_j(f)$. Similarly, $|f(s_j) - f(s_{j-1})| \leqslant$

$4\Delta_j(f)$. Plugging these into (6), we get (note that we want the bound to depend on $|s-t|$ but not on M, N, so there is no choice but to extend to the infinite sum)

$$|f(t) - f(s)| \leqslant 6 \sum_{j>p} \Delta_j(f).$$

As $p = \lfloor \log_2 \frac{1}{|s-t|} \rfloor$, we get that $w_f(\delta) \leqslant 6 \sum_{j \geqslant m_\delta} \Delta_j(f)$. As the tail of a convergent series, this goes to zero as $\delta \downarrow 0$, hence we get the uniform continuity of f. Therefore it extends continuously to $[0,1]$. It is obvious that the extended function has the same modulus of continuity as the original function on D. ■

2. Some insights from the proof

The proof of the construction can be used to extract valuable consequences.

Existence of continuous Gaussian processes with given covariance. Suppose $K : [0,1] \times [0,1] \mapsto \mathbb{R}$ is a postive semi-definite kernel. Do there exists random variables X_t, $t \in [0,1]$ having joint Gaussian distribution with zero means and covariance $\mathbb{E}[X_t X_s] = K(t,s)$? It is not difficult to see that continuity of K is a necessary condition (why?).

To get a sufficient condition, we may follow the same construction as before, and construct X_t, $t \in D$, having the prescribed joint distributions. How do we estimate Δ_n?

Set $h(\delta)^2 = \max\{K(t,t) + K(s,s) - 2K(t,s) : 0 \leqslant t, s \leqslant 1,\ |t-s| \leqslant \delta\}$ (to understand what is happening, observe that if (X_t, X_s) has the prescribed bivariate Gaussian distribution, then $\mathbb{E}[(X_t - X_s)^2] = K(t,t) + K(s,s) - 2K(t,s)$). Then, each of $X(\frac{k+1}{2^n}) - X(\frac{k}{2^n})$ is Gaussian with standard deviation less than or equal to $h(2^{-n})$. By a union bound and the standard estimate for the Gaussian tail, we see that $\Delta_n \leqslant \sqrt{10(1+\delta)}\sqrt{n}h(2^{-n})$, with probability $1 - 2^{-n}$ (observe that even though there is independence of increments in the Brownian case, we did not really use it in this step). Then the same steps as before show that X extends to a continuous function on $[0,1]$ provided $\sum_n \sqrt{n}h(2^{-n}) < \infty$.

In the case of Brownian motion, we had $h(\delta) = \sqrt{\delta}$. If $h(\delta) \leqslant C\delta^p$ for any positive p, then $\sum_n \sqrt{n}h(2^{-n}) < \infty$. In fact, it suffices if $h(\delta) \leqslant (\log(1/\delta))^p$ for a sufficiently large p.

Beyond Gaussians. Now suppose for every $k \geqslant 1$ and every $0 \leqslant t_1 < t_2 < \ldots < t_k \leqslant 1$, we are given a probability distribution $\mu_{t_1,\ldots,t_k}$ on $\mathbb{R}^n$ (in the Gaussian case it was enough to specify the means and covariances, but not in general). The question is whether there exist random variables X_t, $t \in [0,1]$, such that $(X(t_1), \ldots, X(t_k))$ has distribution $\mu_{t_1,\ldots,t_k}$ for every k and every $t_1 < \ldots < t_k$ and such that $t \mapsto X(t)$ is continuous a.s.? We shall of course need the consistency of the finite dimensional distributions, but that is not enough.

From the consistency, we can construct X_t, $t \in D$, as before. It remains to estimate Δ_n. The Gaussian distribution was used when we invoked the tail bound $\mathbb{P}\{Z > t\} \leqslant e^{-t^2/2}$. Now that we do not have that, assume that $\mathbb{E}[(X_t - X_s)^\alpha] \leqslant C|t-s|^{1+\beta}$ for some positive numbers C, α, β and for all $t, s \in [0,1]$. Observe that by $\mathbb{E}[|X_t - X_s|^\alpha]$ we mean the quantity $\int_{\mathbb{R}^2} |x-y|^\alpha d\mu_{t,s}(x,y)$. Then, it follows that

$$\mathbb{P}\left\{|X(\frac{k+1}{2^n}) - X(\frac{k}{2^n})| \geqslant u_n\right\} \leqslant u_n^{-\alpha}\mathbb{E}[|X(\frac{k+1}{2^n}) - X(\frac{k}{2^n})|^\alpha] \leqslant u_n^{-\alpha}2^{-n(1+\beta)}.$$

by the usual Chebyshev idea. Taking union over $0 \leqslant k \leqslant 2^n - 1$, we see that

$$\mathbb{P}\{\Delta_n \geqslant u_n\} \leqslant Cu_n^{-\alpha}2^{-n\beta}.$$

which is summable if $u_n = 2^{-\gamma n}$ for some $0 < \gamma < \frac{\beta}{\alpha}$. Therefore, we get a process with continuous sample paths having modulus of continuity given by the series

$$\sum_{n \geqslant \log_2(1/|t-s|)} u_n \asymp 2^{-\gamma \log_2(1/|t-s|)} = |t-s|^\gamma.$$

The paths are Hölder continuous for any exponent smaller than β/α. This is the original form of the Kolmogorov-Centsov theorem.

Exercise 9

Deduce that Brownian motion is Hölder continuous with any exponent less than $\frac{1}{2}$.

The method of proof clearly cannot give any Hölder exponent larger than $1/2$. In fact by a little analysis, it is easy to see that Brownian motion is *not* uniformly Hölder of any exponent larger than $1/2$. We outline this in the exercise below. Later we shall show a much stronger fact, that Brownian motion has no Hölder points of exponent greater than $1/2$.

Exercise 10

If $Z \sim N(0,1)$, check that $\mathbb{P}\{|Z| < \varepsilon\} \leqslant \varepsilon$ and hence deduce that $\mathbb{P}\{\Delta_n(W) \leqslant C2^{-n\alpha}\}$ is summable for $\alpha > \frac{1}{2}$ and $C < \infty$. Deduce that Brownian motion on $[0,1]$ is not uniformly Hölder(α) for any $\alpha > \frac{1}{2}$, almost surely.

3. Lévy's construction of Brownian motion

Our first construction involved first defining W_t, $t \in D$, having the specified covariances, and then proving uniform continuity of the resulting function. For constructing W_t, $t \in D$, we showed in general that a countable collection of Gaussians with specified covariances can be constructed by choosing appropriate linear combinations of i.i.d. standard Gaussians.

In the following construction, due to Lévy and Cisielski, the special form of the Brownian covariance is exploited to make this construction very explicitly[1].

Lévy's construction of Brownian motion: As before, we construct it on time interval $[0,1]$. Let $\xi_{n,k}$, $k, n \geqslant 0$ be i.i.d. standard Gaussians. Let $F_0(t) = \xi_0 t$. For $n \geqslant 1$, define the random functions F_n by

$$F_n(t) = \begin{cases} \xi_{n,k} 2^{-\frac{1}{2}(n+1)} & \text{if } 0 \leqslant k \leqslant 2^n - 1 \text{ is odd,} \\ 0 & \text{if } 0 \leqslant k \leqslant 2^n - 1 \text{ is even,} \end{cases}$$

and such that F_n is linear on each dyadic interval $[\frac{k}{2^n}, \frac{k+1}{2^n}]$. Then define

$$W_n = F_0 + F_1 + \ldots + F_n.$$

In Figure 3, you may see the first few steps of the construction.

We claim that $\|F_n\|_{\sup} \leqslant 10\frac{\sqrt{n}}{\sqrt{2^n}}$ with probability $\geqslant 1 - \frac{1}{2^n}$. This is because F_n attains its maximum at $k2^{-n}$ for some odd k, and by definition, these values are independent Gaussians with mean zero and variance $1/2^{n+1}$. The usual estimate for the maximum of Gaussians gives the claim.

From this, it follows that $\sum_n \|F_n\|_{\sup} < \infty$ a.s.Therefore, w.p.1., the series $\sum_{n=0}^{\infty} F_n$ converges uniformly on $[0,1]$ and defines a random continuous function W. Further, at any dyadic rational $t \in D_m$, since $F_n(t) = 0$ for $n > m$, the series defining $W(t)$ is a finite sum of independent Gaussians. From this, we see that $W(t)$, $t \in D$ are jointly Gaussian.

We leave it as an exercise to check that $\mathbb{E}[W(t)W(s)] = t \wedge s$ (for $t, s \in D$). Since W is already continuous, and limits of Gaussians are Gaussian, conclude that the Gaussianity and covariance formulas are valid for all $t, s \in [0,1]$. Thus, W is standard Brownian motion on $[0,1]$.

[1] If the following description appears too brief, consult the book of Mörter and Peres where it is explained beautifully.

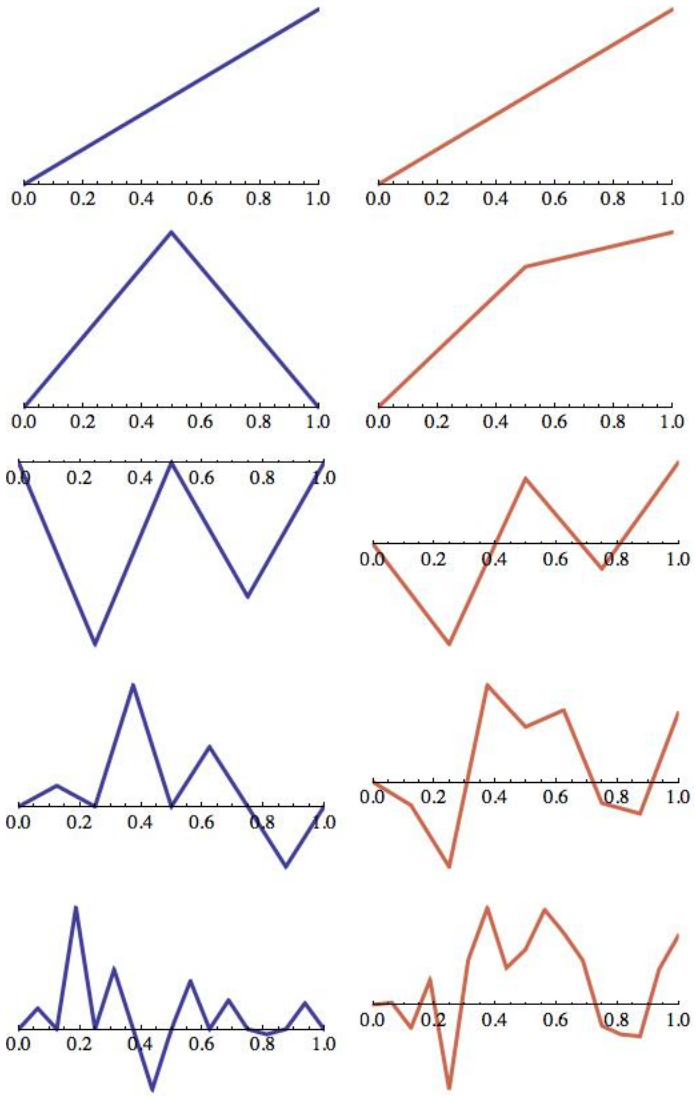

FIGURE 1. The first few steps in Lévy's construction. On the left are the functions F_n and on the right are the functions $F_0 + \ldots + F_n$, for $0 \leqslant n \leqslant 4$.

Remark 2

Let $I_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n}]$ for $0 \leqslant k \leqslant 2^n - 1$ and $n \geqslant 0$. Define $H_{n,k} : [0,1] \to \mathbb{R}$ by

$$H_{n,k}(x) = \begin{cases} +2^{-n/2} & \text{if } x \in [\frac{k}{2^n}, \frac{k+\frac{1}{2}}{2^n}), \\ -2^{-n/2} & \text{if } x \in [\frac{k+\frac{1}{2}}{2^n}, \frac{k+1}{2^n}], \\ 0 & \text{otherwise.} \end{cases}$$

Then, together with the constant function $\mathbf{1}$, the collection $H_{n,k}$, $0 \leqslant k \leqslant 2^n - 1$, $0 \leqslant n$, form an orthonormal basis for $L^2[0,1]$. It is easy to see that

$$F_{n+1}(t) = \sum_{k=0}^{2^n-1} \xi_{n+1,k} \int_0^t H_{n,k}(u)\,du.$$

Thus, the above construction gives the following "formula" for Brownian motion:

$$W(t) = \xi_0 \int_0^t \mathbf{1}(u)\,du + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \xi_{n+1,k} \int_0^t H_{n,k}(u)\,du.$$

4. Series constructions of Brownian motion

Let us do some formal (i.e., non-rigorous) manipulations that sheds a light on the construction of Brownian motion. We start with the idea of "differential space" as Wiener termed it: If W is Brownian motion, the differentials $dW(t)$, $0 \leqslant t \leqslant 1$, are i.i.d. Gaussians (we can't say with what variance, because this is a formal statement without meaning!). Now take any orthonormal basis $\{\phi_n\}$ for $L^2[0,1]$. We know that

$$\sum_n \langle f, \phi_n\rangle\langle g, \phi_n\rangle = \langle f, g\rangle \tag{7}$$

for any $f, g \in L^2[0,1]$. If we set $f = \delta_t$ and $g = \delta_s$, then formally we get $\sum_n \phi_n(t)\phi_n(s) = \langle \delta_t, \delta_s\rangle$, which is precisely the covariance structure we want for $dW(t)$. This suggests that we construct dW by setting $dW(t) = \sum_n X_n\phi_n(t)$, where X_n are i.i.d. $N(0,1)$ (because when we compute $\mathbb{E}[dW(t)dW(s)]$, all terms with $m \neq n$ vanish and we get $\sum_n \phi_n(t)\phi_n(s)$. If so, since we want $W(0) = 0$, we must have

$$W(t) = \sum_n X_n \int_0^t \phi_n(u)du \tag{8}$$

where X_n are i.i.d. standard Gaussians.

Now we can forget the means of derivation and consider the series on the right hand side of (8). If we can show that the series converges uniformly over $t \in [0,1]$ (with probability 1), then the resulting random function is continuous (since $t \mapsto \int_0^t \phi_n$ is), and $W(t)$s will be jointly Gaussian with zero means. To compute their covariances, write $\int_0^t \phi_n = \langle \phi_n, \mathbf{1}_{[0,t]}\rangle$ and hence by taking limits of covariances of partial sums, we see that

$$\begin{aligned}\mathbb{E}[W(t)W(s)] &= \sum_{m,n} \mathbb{E}[X_m X_n]\langle \phi_n, \mathbf{1}_{[0,t]}\rangle\langle \phi_m, \mathbf{1}_{[0,s]}\rangle \;=\; \sum_n \langle \phi_n, \mathbf{1}_{[0,t]}\rangle\langle \phi_n, \mathbf{1}_{[0,s]}\rangle \\ &= \langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]}\rangle \;=\; t \wedge s.\end{aligned}$$

In the first equality in the second line, we used (7).

This gives many new constructions (or new representations) of Brownian motion! The only remaining point is to show the uniform convergence. I do not know for what bases one gets uniform convergence, but here are a few important examples.

Haar basis: Consider the Haar basis, $\mathbf{1}, H_{0,0}, H_{1,0}, H_{1,1}, H_{2,0}, \ldots, H_{2,3}, \ldots$. In this case, it makes sense to index our i.i.d. Gaussian coefficients as $X, X_{0}, X_{1,0}, X_{1,1}, X_{2,0}, \ldots, X_{2,3}, \ldots$. The random function

$$\sum_{k=0}^{2^n-1} X_{n,k} \int_0^t H_{n,k}(u)du$$

is precisely what was called $F_{n+1}(t)$ in the previous section (see Remark 2). And it was shown that the series actually converges uniformly and has the correlations of the Brownian motion. What is special and helps here is that if t is a dyadic rational, then the series for $W(t)$ has only finitely many non-zero terms.

Trigonometric basis: $1, \sqrt{2}\cos(2\pi nt), \sqrt{2}\sin(2\pi nt), n \geqslant 1$, form an orthonormal basis[2] for $L^2[0,1]$. In this case, the series form (8) becomes

$$W(t) = X_0 t + \sqrt{2}\sum_{n=1}^{\infty} \frac{1}{2\pi n}[X_n \sin(2\pi nt) + Y_n(1-\cos(2\pi nt))]$$

where X_n, Y_n are i.i.d. standard Gaussian random variables. In this case it is possible (but not trivial at all) to show that the series converges uniformly with probability 1, and that the resulting random function is Brownian motion.

Another trigonometric basis: The functions $\sqrt{2}\cos[\pi(n+\frac{1}{2})t], n \geqslant 0$, form an orthonormal basis of $L^2[0,1]$. The series (8) then becomes

$$W(t) = \sqrt{2}\sum_{n\geqslant 0} X_n \frac{\sin[\pi(n+\frac{1}{2})t]}{\pi(n+\frac{1}{2})}. \tag{9}$$

Again, it can be shown that the series converges uniformly with probability 1, and gives back Brownian motion. This particular expansion is known as the Karhunen-Loeve expansion (it is an expansion first introduced by D. D. Kosambi. The orthonormal basis here are the eigenfunctions of the integral operator on $L^2[0,1]$ with kernel $K(t,s) = t \wedge s$).

Complex Brownian motion: By complex-valued Brownian motion we mean $W_{\mathbb{C}} = W(t) + iW'(t)$ where W, W' are i.i.d. Brownian motions on $[0,1]$. In the formal manipulation that we gave at the beginning of the section, if we allow complex valued functions and complex scalars, we end up with complex Brownian motion. In other words, the analogue of (8) is

$$W_{\mathbb{C}}(t) = \sum_n Z_n \int_0^t \phi_n(u)du$$

where $\{\phi_n\}$ is an orthonormal basis of $L^2[0,1]$ (now complex-valued functions) and Z_n are i.i.d. standard complex Gaussians (meaning that $\mathrm{Re}(Z_n)$ and $\mathrm{Im}(Z_n)$ are i.i.d. $N(0,1)$).

[2] You may have seen this in Fourier analysis class as an immediate consequence of Fejér's theorem. If not, consider the span of all these functions, and apply Stone-Weierstrass theorem to show that the span is dense in $C[0,1]$ with the sup-norm metric and hence in $L^2[0,1]$ with the L^2 metric.

Again, this may or may not be true for general orthonormal basis. We take the particular case of complex exponentials $\{e_n : n \in \mathbb{Z}\}$, where $e_n(t) = e^{2\pi i n t}$. Then the series becomes

$$W_{\mathbb{C}}(t) = Z_0 t + \sum_{n \neq 0} \frac{Z_n}{2\pi i n} e^{2\pi i n t}.$$

The series converges uniformly (the proof of this assertion is nontrivial) with probability 1 and gives complex Brownian motion.

4.1. Ideas of proofs. In the last three examples, we did not present proofs. There are two stages: First prove that the series converges uniformly on $[0,1]$ with probability 1. Then show that the resulting random function has the right correlations. The first step is similar in all three examples, so let us consider the last one.

Lemma 2

The series $\sum_n \frac{Z_n}{2\pi i n} e^{2\pi i n t}$ converges uniformly over $t \in [0,1]$, with probability 1.

If Z_n/n was absolutely summable with probability 1, then we would be done, but that is false! The main idea is to use cancellation between terms effectively by breaking the sum into appropriately large blocks. Another point worth noting is that for fixed t, the series converges almost surely, by Khinchine-Kolmogorov theorems on sums of independent random variables. One can adapt their proof to Hilbert-space valued random variables and show that the series converges in $L^2[0,1]$, with probability 1. The difficulty here is in getting uniform convergence.

Proof of Lemma 2. For $n \geqslant 1$ define

$$F_n(t) = \sum_{k=2^{n-1}+1}^{2^n} \frac{Z_k}{k} e^{2\pi i k t}.$$

We aim to show that $\sum_n \|F_n\|_{\sup} < \infty$ with probability 1, which of course implies that $\sum_n F_n$ converges uniformly. That implies that the sum over $n \geqslant 1$ of $\frac{Z_n}{n} e^{2\pi i n t}$ converges uniformly with probability 1.

To control $\|F_n\|_{\sup}$, write $M = 2^{n-1} + 1$ and $N = 2^n$ and observe that

$$\begin{aligned} |F_n(t)|^2 &= \sum_{r=M-N+1}^{N-M-1} e^{2\pi i r t} \sum_{k : M \leqslant k, k+r \leqslant N} \frac{\overline{Z}_k Z_{k+r}}{k(k+r)} \\ &\leqslant \frac{1}{M^2} \sum_{r=M-N+1}^{N-M-1} \Big| \sum_{M \leqslant k, k+r \leqslant N} \overline{Z}_k Z_{k+r} \Big| \end{aligned}$$

and hence writing $\|F_n\|$ for the sup-norm of F_n on $[0,1]$, we have

$$\mathbb{E}[\|F_n\|^2] \leqslant \frac{1}{M^2} \sum_{r=M-N+1}^{N-M-1} \mathbb{E}\left[\Big| \sum_{M \leqslant k, k+r \leqslant N} \overline{Z}_k Z_{k+r} \Big| \right].$$

Observe that $\mathbb{E}[\overline{Z}_k Z_\ell] = 2\delta_{k,\ell}$. Therefore, for $r = 0$, the summand is $\mathbb{E}[\sum_{k=M}^{N} |Z_k|^2] = 2(N - M + 1)$. For $r \neq 0$, we bound the summand by the square root of

$$\mathbb{E}\left[\Big|\sum_{M\leqslant k,k+r\leqslant N} \overline{Z}_k Z_{k+r}\Big|^2\right] = \mathbb{E}\left[\sum_{M\leqslant k,k+r\leqslant N}\ \sum_{M\leqslant \ell,\ell+r\leqslant N} \overline{Z}_k Z_{k+r} Z_\ell \overline{Z}_{\ell+r}\right] = 2(N-M+1)$$

because all terms with $k \neq \ell$ vanish. This shows that

$$\begin{aligned}\mathbb{E}[\|F_n\|^2] &\leqslant \frac{1}{M^2}\left\{2(N-M+1) + 2(N-M)\sqrt{2(N-M+1)}\right\}\\ &\leqslant 5\frac{N^{\frac{3}{2}}}{M^2} \ \leqslant\ \frac{20}{2^{\frac{n}{2}}}.\end{aligned}$$

Therefore $\mathbb{E}[\|F_n\|] \leqslant 5 \times 2^{-n/4}$ which is summable, showing that $\sum_n \|F_n\| < \infty$ w.p.1. Hence the series converges uniformly with probability 1. ■

The proofs of uniform convergence is similar in the other cases.

CHAPTER 4

Markov and Strong Markov properties of Brownian motion

1. Blumenthal's zero-one law

We move towards the Markov property of Brownian motion and its consequences. To give a quick preview, standard Brownian motion turns out to be a strong Markov process, and we shall find many martingales hidden in it. These, together with optional sampling theorems applied to certain stopping times will allow us to study very fine properties of Brownian motion in depth. But as may be expected, certain technical matters will crop up. We start with one such.

Let W be a standard Brownian motion in 1-dimension, defined on some $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_t := \sigma\{W_s : s \leqslant t\}$ be the associated natural filtration. Define $\tau = \inf\{t : W(t) \geqslant 1\}$ and let $\tau' = \inf\{t : W(t) > 1\}$. It is easy to see that τ is a stopping time for the natural filtration but τ' is not (just find two paths that agree up to τ but that have different values for τ').

We would like τ' to also be a stopping time. This can be done by enlarging the filtration to $\mathcal{F}_t^+ := \bigcap_{s>t} \mathcal{F}_s$. The filtration $\mathcal{F}_\bullet^+$ is called the right-continuous version of $\mathcal{F}_\bullet$ because $\bigcap_{s>t} \mathcal{F}_s^+ = \mathcal{F}_t^+$ for every $t \geqslant 0$ or in other words $(\mathcal{F}_\bullet^+)^+ = \mathcal{F}_\bullet^+$. It is easy to see that τ' is indeed a stopping time with respect to $\mathcal{F}_\bullet^+$, since the event $\{\tau' \leqslant t\} \in \mathcal{F}_s$ for each $s > t$.

Needless to say, τ remains a stopping time upon enlarging the filtration. What can go wrong with enlargement are Markov properties or martingale properties. For example, for any t we know that $W(\cdot + t) - W(t)$ is independent of $\mathcal{F}_t$. Does it remain true that $W(\cdot + t) - W(t)$ is independent of $\mathcal{F}_t^+$? If not, it is easy to imagine that the enlargement causes more difficulties than it solves.

The first and foremost task is to check that the enlargement is trivial - it adds only $\mathbb{P}$-null sets. This is indeed true.

Lemma 3: Blumenthal's zero-one law

If $A \in \mathcal{F}_0^+$, then $\mathbb{P}(A)$ equals 0 or 1.

Proof. We know that $W^T := W(T + \cdot) - W(T)$ is independent of $(W_t)_{0 \leqslant t \leqslant T}$, for any $T > 0$. As $T \downarrow 0$, the sigma-algebra generated by $(W_t)_{0 \leqslant t \leqslant T}$ decreases to $\mathcal{F}_0^+$. Further, $\sigma(\cup_{T>0} \sigma(W^T)) = \sigma(W)$, since $W_T \to 0$ (to be more precise, because $W(t) = \lim_{T \downarrow 0} W^T(t)$). Therefore, $\sigma(W)$ is independent of $\mathcal{F}_0^+$. But $\mathcal{F}_0^+ \subseteq \sigma(W)$, hence $\mathcal{F}_0^+$ is independent of itself. Hence any $A \in \mathcal{F}_0^+$ must have probability 0 or 1. ∎

In this proof, we used the following simple fact (observe that we could have worked with a sequence T_n decreasing to 0).

Exercise 11

Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ and $\mathcal{G}_1 \supseteq \mathcal{G}_2 \supseteq \ldots$ be sub-sigma algebras of $\mathcal{F}$ in $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{F}_n$ is independent of $\mathcal{G}_n$ for each n, then $\sigma\left(\bigcup_n \mathcal{F}_n\right)$ is independent of $\bigcap_n \mathcal{G}_n$.

Remark 3

In the lectures, I indicated further enlargement by including all $\mathbb{P}$-null sets in each $\mathcal{F}_t$. More precisely, let $\mathcal{F}_\infty = \sigma\{W\} = \sigma\{\bigcup_{t\geqslant 0} \mathcal{F}_t\}$ and let

$$\mathcal{N} = \{A \subseteq \Omega : A \subseteq B \text{ for some } B \in \mathcal{F}_\infty \text{ with } \mathbb{P}(B) = 0\}.$$

Then define $\overline{\mathcal{F}}_t^+ = \sigma\{\mathcal{F}_t^+ \cup \mathcal{N}\}$. This is the completed, right-continuous filtration. All results stated below for the right-continuous filtration also hold for the completed right-continuous filtration.

2. Markov and strong Markov properties

Let W be a standard d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_\bullet$ be the natural filtration generated by W and let $\mathcal{F}_\bullet^+$ be the right-continuous filtration defined by $\mathcal{F}_t^+ = \bigcap_{s>t} \mathcal{F}_s$.

Here is a naive way to state the Markov and strong Markov properties.

- (Markov property). Fix T and define $B(t) = W(T+t) - W(T)$ for $t \geqslant 0$. Then, B is a standard Brownian motion that is independent of $\mathcal{F}_t^+$.
- (Strong Martov property). Fix an $\mathcal{F}_\bullet^+$-stopping time τ and define $B(t) = W(t+\tau) - W(\tau)$ for $t \geqslant 0$. Then B is a standard Brownian motion independent of $\mathcal{F}_\tau^+$. Recall that $\mathcal{F}_\tau^+ = \{A \in \mathcal{F} : A \cap \{\tau \leqslant t\} \in \mathcal{F}_t\}$.

We have already proved the Markov property when the filtration $\mathcal{F}_\bullet$ is used. By Blumenthal's zero-one law, $\mathcal{F}_t^+$ is got from $\mathcal{F}_t$ by augmenting some $\mathbb{P}$-null sets. Hence, independence of B from $\mathcal{F}_T$ is equivalent to independence of B from $\mathcal{F}_T^+$. Strong Markov property is slightly less obvious.

Proof of Strong Markov property. For simplicity we use the notation of 1-dimension. First assume that τ takes countably many values $s_0, s_1, s_2, \ldots$ for some $\delta > 0$. Fix any $A \in \mathcal{F}_\tau$, any $n \geqslant 1$ and $t_1, \ldots, t_n \geqslant 0$, and any $u_1 \ldots, u_n \in \mathbb{R}$. Let E be the event that $B(t_j) \leqslant u_j$ for $1 \leqslant j \leqslant n$. Then,

$$\begin{aligned}\mathbb{P}\{E \cap A\} &= \sum_{m=0}^{\infty} \mathbb{P}\{E \cap A \cap \{\tau = s_m\}\} \\ &= \mathbb{P}\left\{\{B(s_m + t_j) - B(s_m) \leqslant u_j \text{ for } j \leqslant n\} \cap A \cap \{\tau = s_m\}\right\}.\end{aligned}$$

For fixed m, by Markov property and the fact that $A \cap \{\tau = s_m\} \in \mathcal{F}^+_m$, the mth summand above is equal to

$$\mathbb{P}\{W(t_j) \leqslant u_j \text{ for } j \leqslant n\}\mathbb{P}\{A \cap \{\tau = s_m\}\}.$$

Adding up and using $\mathbb{P}(A) = \sum_m \mathbb{P}\{A \cap \{\tau = s_m\}\}$ gives the identity $\mathbb{P}\{E \cap A\} = \mathbb{P}\{W(t_j) \leqslant u_j \text{ for } j \leqslant n\}\mathbb{P}\{A\}$. This shows that B is independent of $\mathcal{F}^+_\tau$ and that B has the same distribution as W.

Now consider a general stopping time τ. For $\ell \geqslant 1$ define $\tau_\ell = 2^{-\ell}\lceil 2^\ell \tau \rceil$. Then τ_ℓ is a stopping time, $\tau \leqslant \tau_\ell \leqslant \tau + 2^{-\ell}$. Thus $\tau_\ell \downarrow \tau$. Let $V = (W(\tau + t_1), \ldots, W(\tau + t_n))$ and $V_\ell = (W(\tau_\ell + t_1), \ldots, W(\tau_\ell + t_n))$ so that by continuity of Brownian motion, we have $V_\ell \overset{a.s.}{\to} V$. Thus, for most choices of $u_1, \ldots, u_n$ (we need u_j to be a continuity point of $V(j)$) we get

$$\begin{aligned}\mathbb{P}\{\{V(j) \leqslant u_j \text{ for } j \leqslant n\} \cap A\} &= \lim_{\ell \to \infty} \mathbb{P}\{\{V_\ell(j) \leqslant u_j \text{ for } j \leqslant n\} \cap A\} \\ &= \lim_{\ell \to \infty} \mathbb{P}\{W(t_j) \leqslant u_j \text{ for } j \leqslant n\}\mathbb{P}\{A\}\end{aligned}$$

where the last line used the strong Markov property for stopping times τ_ℓ that takes countably many values. ■

For our purposes this is sufficient. Observe that Markov property can be stated as saying that the conditional distribution of $W(T + t)$, $t \geqslant 0$, given $\mathcal{F}^+_T$ is the same as that of Brownian motion started at the point $W(T)$. Similarly, strong Markov property says that the conditional distribution of $W(\tau + t)$ given $\mathcal{F}^+_\tau$ is Brownian motion started at $W(\tau)$.

This is a better way of stating these properties. In case of Brownian motion, because of symmetries ($W + x$ is the same as Brownian motion conditioned on starting at x). In general, we consider a family of probability measure $\mathbb{P}_x$, $x \in \mathbb{R}$, on $C[0, \infty)$ such that $\mathbb{P}_x\{f : f(0) = x\} = 1$. This family is said to have (time-homogeneous) Markov property if:

Fix any $x \in \mathbb{R}^d$ and let $X = (X_t)_{t \geqslant 0} \sim \mathbb{P}_x$. Then, conditional on $\mathcal{F}^+_T$, the process $(X(T + t))_{t \geqslant 0}$ has the same distribution as $\mathbb{P}_{X(T)}$. Strong Markov property is stated in a similar way.

Example 4

Let $\mathbb{P}_x$ be the distribution of $(x + W_t + t)_t$ for $x \geqslant 0$ and the distribution of $(x + W_t - t)_{t \geqslant 0}$ for $x < 0$. Then $\mathbb{P}_x$ does not have Markov property.

Example 5

Let $\mathbb{P}_x$ be the distribution of $x + W$ for $x \neq 0$ and let $\mathbb{P}_0 = \delta_0$ be the Dirac measure at the constant function zero. Then, $\mathbb{P}$ satisfies Markov property but not the strong Markov property.

Indeed, if $x = 0$, then conditional on $\mathcal{F}_T$, the distribution of the future path $(W(T+t))_{t\geqslant 0}$ is degenerate at zero. If $x \neq 0$, ignoring the zero probability event $W_T = 0$, we see that the future path $W(T+t)$ is that of Brownian motion stated at $W(T)$ (does not work if $W(T) = 0$ but zero probability events may be ignored).

But if $\tau = \min\{t : W_t = 0\}$, then the conditional distribution of $(W(\tau+t))_{t\geqslant 0}$ is the standard Brownian motion, which is not the same as $\mathbb{P}_{W(\tau)} = \mathbb{P}_0$.

3. Zero set of Brownian motion

Let W be standard 1-dimensional Brownian motion and let $Z = \{t : W_t = 0\}$. Clearly Z is a random closed set of $\mathbb{R}_+$.

Theorem 3

Z has no isolated points, w.p.1.

Proof. For $q \in \mathbb{Q}_+$, let $\tau_q = \min\{s > t : W(s) = 0\}$. By SMP, $W(\tau_q + t) - W(\tau_q) = W(\tau_q + t)$ is a standard Brownian motion. In particular, it has infinitely many zeros on any positive time interval $[0, \varepsilon)$. Hence, τ_q is an accumulation point (from the right) of Z, w.p.1. Take intersection over $q \in \mathbb{Q}_+$ to see that w.p.1., every τ_q, $q \in \mathbb{Q}$, is an accumulation point of Z.

Now, a zero $z \in Z$ is not of the form τ_q is and only if z is an accumulation point of Z from the left! Thus, all zeros of W are accumulation points. ■

4. Reflection principle

Let W be standard 1-dimensional Brownian motion. For $a > 0$ define the *running maximum* $M_t := \max_{0\leqslant s\leqslant t} W_s$ and the *first passage time* $\tau_a := \min\{t \geqslant 0 : W(t) \geqslant a\}$. These are closely interconnected, since $M_t \geqslant a$ if and only if $\tau_a \leqslant t$.

Many questions can be asked: What is the distribution of M, of τ_a? Let T_* be the (unique) time in $[0,1]$ such that $W(T_*) = \max_{s\leqslant 1} W(s)$. What is the distribution of T_*?

We shall answer all these questions. A basic tool is the reflection principle, a direct consequence of the strong Markov property.

Lemma 4: Reflection principle

Let W be standard 1-dimensional Brownian motion. Fix $a > 0$ and define

$$B(t) = \begin{cases} W(t) & \text{if } t \leqslant \tau_a, \\ 2W(\tau_a) - W(t) & \text{if } t > \tau_a. \end{cases}$$

Then, B is a standard Brownian motion.

PROOF. Let $X = (W_t)_{t \leqslant \tau_a}$, $Y = (W_{t+\tau_a} - a)_{t \geqslant 0}$ and $Z = -Y$. Then $Y \stackrel{d}{=} Z$, X is independent of Y (by strong Markov property) and hence X is independent of Z. Hence $(X, Y) \stackrel{d}{=} (X, Z)$.

Concatenating X with Y gives W while concatenating X with Z gives B. Thus $B \stackrel{d}{=} W$. ■

A far reaching generalization of the reflection principle is the Karlin-McGregor formula, discussed in a later section. For now, we focus on using the reflection principle (hence the strong Markov property ultimately) to find exact distributions of the following interesting functionals of (one-dimensional) Brownian motion.

(1) Running maximum: $M_t = \max\{W_s : 0 \leqslant s \leqslant t\}$.

(2) First passage times: $\tau_a = \inf\{t : W_t = a\}$.

(3) Exit time from an interval[1]: $\tau_{\{-a,b\}} = \tau_{-a} \wedge \tau_b$ for $a, b > 0$.

(4) Location of maximum: $T^* = \arg\max_{0 \leqslant t \leqslant 1} W(t)$.

(5) Last zero: $L = \max\{t \leqslant 1 : W(t) = 0\}$.

(6) Time spent on the positive axis: $R = \text{Leb.}\{t \leqslant 1 : W(t) \geqslant 0\}$.

When studying more general stochastic processes, these provide a model for what to ask for and limits for what one can hope to achieve.

5. Running maximum and the first passage time

The two are closely related, since $M_t \geqslant a$ if and only if $\tau_a \leqslant t$.

5.1. Running maximum. Let $a > 0$ and $t > 0$. Let W be a standard Brownian motion and let B be related to it as in the reflection principle. Then,

$$\begin{aligned} \{M_t > a\} &= \{M_t > a, W_t > a\} \sqcup \{M_t > a, W_t < a\} \\ &= \{W_t > a\} \sqcup \{B_t > a\}. \end{aligned}$$

[1] As a matter of fact, we shall do this with martingale methods later, not using the Strong Markov property.

Therefore, $\mathbb{P}\{M_t > a\} = 2\mathbb{P}\{W_t > a\} = \mathbb{P}\{|W_t| > a\}$. Thus, $M_t \stackrel{d}{=} |W_t| \stackrel{d}{=} \sqrt{t}|Z|$ where $Z \sim N(0,1)$. We may also write for $a \geqslant 0$,

$$\mathbb{P}\{M_t > a\} = 2\bar{\Phi}(a/\sqrt{t}) = \frac{2}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{1}{2}u^2} du.$$

Differentiating, we get the density of M_t to be

$$f_{M_t}(a) = -\frac{d}{dt}\mathbb{P}\{M_t > a\} = \frac{2}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{1}{2t}a^2}.$$

In fact, the proof gives us the joint distribution of the Brownian motion and its running maximum at any fixed time. Fix $a > 0$ and $-\infty < b < a$. Then, by the definition of B in terms of W,

$$\{M_t > a \text{ and } W_t < b\} = \{B_t > 2a - b\}.$$

Since B is standard Brownian motion, we get $\mathbb{P}\{M_t > a \text{ and } W_t < b\} = \bar{\Phi}((2a-b)/\sqrt{t})$. Thus,

$$f_{(M_t, W_t)}(a, b) = -\frac{d^2}{da\, db}\bar{\Phi}((2a-b)/\sqrt{t}) = \frac{2(2a-b)}{\sqrt{2\pi}\, t^{\frac{3}{2}}} e^{-\frac{1}{2t}(2a-b)^2}.$$

Two distributional identities:

(1) $M_t \stackrel{d}{=} |W_t|$ for each t. We already saw this.

(2) $M_t - W_t \stackrel{d}{=} |W_t|$ for each t. This can be computed from the joint intensity, but here is a computation-free proof[2]. The process $X_s = W_{t-s} - W_t$ for $0 \leqslant s \leqslant t$ is a standard Brownian motion and $M_t^X = M_t^W - W_t$.

Do these equalities in distribution extend to those of the processes? The first one does not, since M is an increasing process while $|W|$ (called *reflected Brownian motion*) is not. But it is a non-trivial theorem of Lévy that the second one does, i.e., $M - W \stackrel{d}{=} |W|$. The key point is that $M - W$ is a Markov process (we also need that $|W|$ is a Markov process, but that is easy to see). Once that is checked, the equality in distribution at fixed times easily extends to equality of finite dimensional distributions by successive conditioning. Since both $|W|$ and $M - W$ are continuous, this implies equality in distribution of the two processes.

5.2. First passage times. As $\tau_a \leqslant t$ if and only if $M_t \geqslant a$, we get $\mathbb{P}\{\tau_a \leqslant t\} = 2\bar{\Phi}(a/\sqrt{t})$. The density of τ_a is therefore,

$$f_{\tau_a}(t) = \frac{d}{dt}\mathbb{P}\{\tau_a \leqslant t\} = \frac{a}{\sqrt{2\pi}\, t^{\frac{3}{2}}} e^{-\frac{1}{2t}a^2}.$$

There are several observations worth making.

- The density may look unfamiliar. A small computation shows that it is the density of the reciprocal of a Gamma$(\frac{1}{2}, \frac{a^2}{2})$ random variable.

[2]Thanks to Arun Selvan for the nice proof!

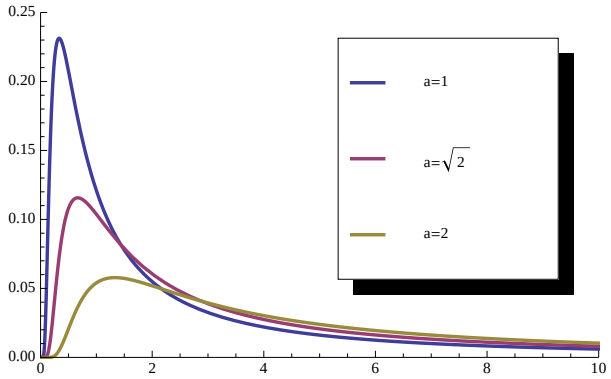

FIGURE 1. Densities of first passage time τ_a

- $\mathbb{P}\{\tau_a \leqslant t\} \to 2\bar{\Phi}(0)$ as $t \uparrow \infty$, showing that $\tau_a < \infty$ w.p.1. Taking intersection over all $a \in \mathbb{Q}$ and using continuity of Brownian paths, we see that almost surely, W hits every point in $\mathbb{R}$. We say that 1-dimensional Brownian motion is *point recurrent*. This fails in higher dimensions, as we shall see later.
- The density approaches decays like $t^{-3/2}$ and $\mathbb{P}\{\tau_a \geqslant t\}$ decays like $t^{-1/2}$. Thus the tail is quite heavy and $|\mathbb{E}[\tau_a^p] < \infty$ if and only if $p < \frac{1}{2}$. This is in contrast to the exit time of the two-sided interval $\tau_{\pm a}$ which has exponential tails because $\tau_{\pm a} \leqslant S := \min\{n : W(n) - W(n-1) > 2a\}$ (if W has not exited the interval $[-a, a]$ by time $N-1$, then it will exit during $[N, N+1]$) and $S \sim \text{Geo}(p)$ with $p = \bar{\Phi}(2a) < 1$.
- $\tau_a \stackrel{d}{=} a^2\tau_1$. One can see it from the density by a short calculation, but it is more illuminating to see it from the Brownian scaling. If $X(t) = aW(t/a^2)$, then $\tau_a^X = a^2\tau_1^W$.
- $\tau_{a+b} \stackrel{d}{=} \tau_a + \tilde{\tau}_b$, where $\tilde{\tau}_b = \inf\{t : W(\tau_a + t) - W(\tau_a) \geqslant b\}$ is an independent copy of τ_b. Together, with the scaling relation, this means that the sum of two i.i.d. copies of τ_a has the same distribution as $2^2\tau_a$. Thus, τ_a has a *Stable distribution*[3] with parameter $\frac{1}{2}$.
- One can also look at the process $(\tau_a)_{a \geqslant 0}$. What we see from above is that much like Brownian motion, it has independent, stationary increments. It also has RCLL paths, but not continuous. It is called a *Stable* $\frac{1}{2}$ *subordinator process*, a special class of *Lévy processes* (those with independent stationary increments).

[3] In general, we say that X has a Stable distribution with parameter α if the sum of two i.i.d. copies of X has the same distribution as $2^{1/\alpha}X$. Such distributions exist if and only if $0 < \alpha \leqslant 1$, and although not unique, form a small two-parameter family of distributions. Centered Gaussians are stable-2, while Cauchy is stable-1.

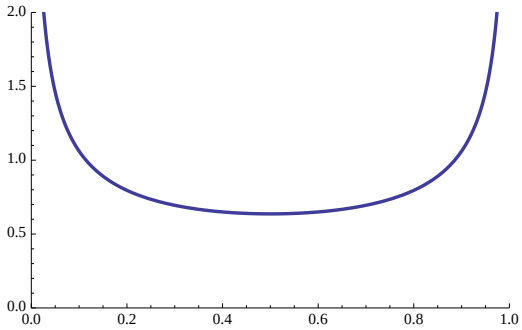

FIGURE 2. Arcsine density

5.3. Local times - a digression. [4] Recall Lévy's identity $M - W \stackrel{d}{=} |W|$ as processes. Imagine a probability space with two standard Brownian motions $W, \tilde{W}$ related such that $M^W - W = \tilde{W}$. Then, the process M^W is related to $\tilde{W}$ in a special way. Being an increasing function, M^W may be thought of as the distribution function of a random measure. Observe that M^W is constant on any interval (s,t) where $\tilde{W}$ has no zeros. This means that the random measure defined by M^W is supported on the zero set of $\tilde{W}$. It is called the *local time* of $\tilde{W}$, a clock that ticks only when the Brownian motion is at zero.

This is not entirely satisfactory. What we would like is to define a local time for the Brownian motion that we started with, in a canonical way. This is possible. Indeed, it can be shown that

$$L_t(0) := \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \text{Leb}\{s \leqslant t : |W_s| \leqslant \varepsilon\}$$

exists and defined the local time at 0. It is also possible to define $L_t(x)$ for $t > 0$ and $x \in \mathbb{R}$, simultaneously. But we shall not touch upon this matter in this course.

6. Lévy's arcsine laws

The arcsine law is another name for the Beta$(\frac{1}{2}, \frac{1}{2})$ distribution that has density $\frac{1}{\pi\sqrt{x(1-x)}}$ and the distribution function $\frac{2}{\pi}\arcsin(\sqrt{x})$ on $[0,1]$. In basic probability class, one may have seen that one way the Beta(p,q) distribution arises is as the distribution of $\frac{U}{U+V}$ where $U \sim \text{Gamma}(p,\lambda)$ and $V \sim \text{Gamma}(q,\lambda)$ are independent. Note that the scale parameter must be the same for both for it to cancels in the ratio. In particular, for arcsine we must take U, V i.i.d. Gamma$(\frac{1}{2}, \lambda)$.

6.1. Location of the maximum. As all the values of local maxima are distinct, T^* is well-defined (in fact the proof below shows this without having to invoke the result on distinctness of local maxima values). Fix $t \in (0,1)$. Then $T^* \leqslant t$ if and only if $\max_{0 \leqslant s \leqslant t} W_s \geqslant \max_{t \leqslant s \leqslant 1} W_s$ which is equivalent to $\max_{0 \leqslant s \leqslant t} W_s - W_t \geqslant \max_{t \leqslant s \leqslant 1} W_s - W_t$.

[4] Safe to omit.

If $\tilde{W}_s := W_{t+s} - W_t$ for $0 \leqslant s \leqslant 1-t$, then $\tilde{W}$ is a standard Brownian motion that is independent of $(W_s)_{s \leqslant t}$. Thus, putting everything together, we arrive at

$$\mathbb{P}\{T^* \leqslant t\} = \mathbb{P}\{M_t \geqslant \tilde{M}_{1-t}\}.$$

Because $M_t \stackrel{d}{=} |W_t|$, we may write $M_t = \sqrt{t}|X|$ and $\tilde{M}_{1-t} = \sqrt{1-t}|Y|$ where X, Y are i.i.d. standard Gaussians. Thus, $T^* \leqslant t$ if and only if $\sqrt{t}|X| \geqslant \sqrt{1-t}|Y|$ which upon squaring gives

$$\mathbb{P}\{T^* \leqslant t\} = \mathbb{P}\left\{\frac{Y^2}{X^2+Y^2} \leqslant t\right\}.$$

But X^2, Y^2 are i.i.d. Gamma$(\frac{1}{2}, \frac{1}{2})$ (also known as χ^2_1), hence the ratio here has Beta$(\frac{1}{2}, \frac{1}{2})$ distribution. Thus, T^* has arcsine distribution.

6.2. Last zero. $L \geqslant t$ if and only if W hits zero somewhere in $[t, 1]$. Let $\widetilde{W}_s = W_{t+s} - W_t$ for $0 \leqslant s \leqslant 1-t$ which is a Brownian motion independent of $\mathcal{F}_t$.

Now, W hits zero in $[t, 1]$ if and only if $\widetilde{M}_{1-t} \geqslant |W_t|$ (if $W_t < 0$) or $\min_{s \leqslant 1-t} \widetilde{W}_s \leqslant -|W_t|$ (if $W_t > 0$). Clearly either one has the same probability. Hence we arrive at

$$\mathbb{P}\{L \geqslant t\} = \mathbb{P}\{\widetilde{M}_{1-t} \geqslant |W_t|\}.$$

But we may write $\tilde{M}_{1-t} = \sqrt{1-t}|X|$ and $|W_t| = \sqrt{t}|Y|$ where X, Y are i.i.d. standard Gaussians. Hence we return to the same calculation as for T^*. Thus L must have arcsine distribution. ■

6.3. Time spent on the positive axis. TO BE WRITTEN. IGNORE. Consider $R = \{t \leqslant 1 : W_t > 0\}$, the proportion of time spent by the Brownian motion in the positive half-line.

Let $a > 0$ and let $0 = T_0 < T_1 < T_2 < \dots$ be the successive hitting times of $0, a, 0, a, 0, \dots$. Then $T_k - T_{k-1}$ are i.i.d. with the same distribution as τ_a. Let N_a be such that $T_{N_a} \leqslant 1 < T_{N_a+1}$ and let $M_a = \lfloor N_a/2 \rfloor$. Further, define $A = \{t \leqslant 1 : W(t) \geqslant 0\}$ and $B_a = \{t \leqslant 1 : W(t) \leqslant a\}$. Then

$$|A| \geqslant \sum_{j=1}^{M_a} T_{2j} - T_{2j-1} =: \xi_a, \qquad |B_a| \geqslant \sum_{j=1}^{M_a} T_{2j-1} - T_{2j-2} =: \eta_a$$

where ξ_a, η_a are i.i.d. with distribution same as the passage time τ_{aM_a}. Therefore, $\frac{1}{\xi_a}, \frac{1}{\eta_a}$ are i.i.d. Gamma$(\frac{1}{2}, \frac{a^2}{2})$

Further,

$$\mathbb{E}[|A \cap B_a|] = \int_0^1 \mathbb{P}\{0 \leqslant W_t \leqslant a\} \leqslant a$$

since the density of W_t is bounded by 1 (in fact by $1/\sqrt{2\pi}$). Therefore, $\mathbb{P}\{|A \cap B_a| \leqslant \sqrt{a}\} \leqslant \sqrt{a}$.

We need two facts (proved later):

(1) $N_a \uparrow \infty$ a.s. as $a \downarrow 0$.

(2) $\max_{k \leqslant N_a+1} \tau_k \xrightarrow{P} 0$ as $a \downarrow 0$.

Let and Then

Further,

$$\{t \leqslant 1 : W_t \geqslant a\} \subseteq \bigcup_{1 \leqslant j \leqslant \frac{1}{2}N_a} [T_{2j-1}, T_{2j}] \subseteq \{t \leqslant 1 : W_t \geqslant 0\}$$

and therefore,

$$|\{t \leqslant 1 : W_t \geqslant a\}| \leqslant \sum_{j \leqslant \frac{1}{2}N_a} \tau_{2j-1} \leqslant |\{t \leqslant 1 : W_t \geqslant 0\}|.$$

Similarly,

$$|\{t \leqslant 1 : W_t \leqslant 0\}| \leqslant \sum_{j \leqslant \frac{1}{2}(N_a+1)} \tau_{2j} \leqslant |\{t \leqslant 1 : W_t \leqslant 0\}|.$$

We shall prove this later.

Proof of Lévy's third arcsine law by chaos expansion:

Let γ be the standard Gaussian measure on $\mathbb{R}$. Applying Gram-Schmidt (without normalizing) procedure to $1, x, x^2, \ldots$ in $L^2(\gamma)$, we get a sequence of monic polynomials $H_0(x), H_1(x), \ldots$ that are orthogonal in $L^2(\gamma)$. Clearly H_n has degree n. These are known as *Hermite polynomials* and we can describe them explicitly in multiple ways:

(1) $H_n(x) = (-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2}$. This is clearly monic and has degree n. Hence it suffices to check that they are orthogonal in $L^2(\gamma)$, which follows by integrating by parts. If $m < n$,

$$\int H_n(x)H_m(x)d\gamma(x) = \int H_m(x)\frac{d^n}{dx^n}e^{-\frac{1}{2}x^2}\,dx = (-1)^n \int e^{-\frac{1}{2}x^2}\frac{d^n}{dx^n}H_m(x)dx = 0.$$

(2)

Proof of Lévy's third arcsine law by using passage time density:

7. Non-intersecting Brownian motions

In martingales class, we saw the Karlin-McGregor formula for the non-intersection probability of independent simple random walks on $\mathbb{Z}$. The proof we gave was using martingales. Here we give the analogous theorem for Brownian motion, but using the strong Markov property. In fact, the proof works for any Markov process with continuous sample paths but to not digress into definitions of these objects, we stick to Brownian motion.

Let $p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$. Then for a Brownian motion W, the conditional density of W_t given $W_s = y$ (known as the *transition density*) is given by $p_{t-s}(x-y)$.

Theorem 4: Karlin-McGregor formula

Let $W_1, \ldots, W_n$ be Brownian motions started at $x_1 < \ldots < x_n$. Fix $T > 0$. Let $\mathrm{NI}(T)$ denote the event that $W_i(s) \neq W_j(s)$ for all $i \neq j$ and all $s \leqslant T$. On the event $\mathrm{NI}(T)$, the density of $(W_1(T), \ldots, W_n(T))$ at $(y_1, \ldots, y_n)$ where $y_1 < \ldots < y_n$, is given by $\det(p_T(y_j - x_i))_{i,j \leqslant n}$.

The meaning of the statement is that for any $a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n$, we have

$$\mathbb{P}\{\mathrm{NI}(T) \text{ and } W_j(T) \in [a_j, b_j] \text{ for } 1 \leqslant j \leqslant n\} = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} \det(p_T(y_j - x_i))_{i,j \leqslant n} \, dy_n \ldots dy_1. \tag{10}$$

From this we can draw two conclusions:

(1) The probability of $\mathrm{NI}(T)$ is equal to

$$\mathcal{Z} := \int_{\mathbb{R}^n} \mathbf{1}_{y_1 < \ldots < y_n} \det(p_T(y_j - x_i))_{i,j \leqslant n} \, dy_1 \ldots dy_n. \tag{11}$$

(2) The conditional density of $(W_1(T), \ldots, W_n(T))$ given $\mathrm{NI}(T)$, at $y_1 < \ldots < y_n$, is equal to

$$\rho(y \,\big|\, \mathrm{NI}(T)) = \frac{1}{\mathcal{Z}} \det(p_T(y_j - x_i))_{i,j \leqslant n}. \tag{12}$$

Proof. Let $\mathcal{F}_\bullet$ denote the natural filtration of $W = (W_1, \ldots, W_n)$. Let $\tau = \min\{t \geqslant 0 : W_i(t) = W_j(t) \text{ for some } i \neq j\}$. If $\tau < \infty$, w.p.1., exactly two of the Brownian motions coincide at time τ (why?). Call their indices (I, J), with $I < J$. Now define

$$B_I(t) = \begin{cases} W_I(t) & t \leqslant \tau, \\ W_J(t) & t \geqslant \tau \end{cases}, \qquad B_J(t) = \begin{cases} W_J(t) & t \leqslant \tau, \\ W_I(t) & t \geqslant \tau \end{cases}, \qquad B_k = W_k \text{ for } k \neq I, J.$$

By the strong Markov property, $B = (B_1, \ldots, B_n)$ has the same distribution as W.

Let $C_k = [a_k, b_k]$, so the right side of (10) may be written as

$$\sum_{\pi \in \mathcal{S}_n} \mathrm{sgn}(\pi) \prod_{k=1}^n \int_{C_k} p_T(y_k - x_{\pi(k)}) dy_k = \mathbb{E}\left[\sum_{\pi \in \mathcal{S}_n} \mathrm{sgn}(\pi) \prod_{k=1}^n \mathbf{1}_{W_{\pi(k)}(T) \in C_k} \right].$$

Observe that in the sum inside the expectation, at most one term can be non-vanishing a.s. If we replace W by B, the expectation must remain the same, since $W \stackrel{d}{=} B$. Then both are equal to their average which is

$$\frac{1}{2}\mathbb{E}\left[\sum_{\pi \in \mathcal{S}_n} \mathrm{sgn}(\pi) \prod_{k=1}^n \mathbf{1}_{W_{\pi(k)}(T) \in C_k} + \sum_{\pi \in \mathcal{S}_n} \mathrm{sgn}(\pi) \prod_{k=1}^n \mathbf{1}_{B_{\pi(k)}(T) \in C_k} \right].$$

On the event $\tau \leqslant T$, we see that $B(T)$ is got by applying a transposition to $W(T)$, hence the integrand vanishes (if the first sum is ± 1, then the second is ∓ 1). On the event $\tau > T$, the two integrands are equal to 1, and it is the identity permutation that contributes. Thus, we get that the above quantity

is equal to

$$\mathbb{E}\left[\mathbf{1}_{\tau>T}\prod_{k=1}^{n}\mathbf{1}_{W_k(T)\in C_k}\right].$$

As $\{\tau > T\}$ is the same as $\mathrm{NI}(T)$, this expectation is precisely the left side of (10). ■

7.1. The case of equal starting points. If two of the starting locations are equal, then two rows of $(p_T(y_j - x_i))_{i,j\leqslant n}$ are equal and hence (11) shows that the probability of non-intersection is zero (this can also be seen directly: if $x_1 = x_2$, then infinitely often $W_1 - W_2$ oscillates across zero, so W_1 and W_2 do intersect). However, the conditional density can be still made sense of by approximation.

Taking $T = 1$ for simplicity, we write

$$\begin{aligned}\det(p_1(y_j - x_i))_{i,j\leqslant n} &= e^{-\frac{1}{2}\sum_{j=1}^n y_j^2 - \frac{1}{2}\sum_{i=1}^n x_i^2}\det\left(e^{y_j x_i}\right)_{i,j\leqslant n}\\ &= e^{-\frac{1}{2}\sum_{j=1}^n y_j^2 - \frac{1}{2}\sum_{i=1}^n x_i^2}\det\left(\sum_{p=0}^{\infty}\frac{y_j^p x_i^p}{p!}\right)_{i,j\leqslant n}\\ &= e^{-\frac{1}{2}\sum_{j=1}^n y_j^2 - \frac{1}{2}\sum_{i=1}^n x_i^2}\det(AB)\end{aligned}$$

where A is an $n\times\infty$ matrix with entries $A_{i,p} = \frac{x_i^p}{\sqrt{p!}}$ and B is an $\infty\times n$ matrix with entries $B_{p,j} = \frac{y_j^p}{\sqrt{p!}}$. Applying the Cauchy-Binet formula, we see that

$$\det(AB) = \sum_{0\leqslant p_1<p_2<\ldots<p_n}\det(A_{i,p_j})_{i,j\leqslant n}\times\det(B_{p_i,j})_{i,j\leqslant n}.$$

Now take $x_i = i\varepsilon$, where $\varepsilon > 0$ will be driven to zero. Then $\det(A_{i,p_j})_{i,j\leqslant n}$ has a factor of $\varepsilon^{p_1+\ldots+p_n}$. Consequently, the dominant contribution comes by taking $p_i = i-1$, $1\leqslant i\leqslant n$, in which case

$$\det(B_{i-1,j})_{i,j\leqslant n} = \det(y_j^{i-1})_{i,j\leqslant n} = \Delta(y_1,\ldots,y_n) := \prod_{1\leqslant i<j\leqslant n}(y_j - y_i),$$

$$\det(A_{i,j-1})_{i,j\leqslant n} = \varepsilon^{0+1+\ldots+(n-1)}\det(i^{j-1})_{i,j\leqslant n} = \varepsilon^{(n-1)n/2}\Delta(1,2,\ldots,n).$$

Thus, noting that the $e^{-x_i^2/2}$ factors converge to 1, we get

$$\mathbb{P}\{\mathrm{NI}(1)\} = (1+O(\varepsilon))\varepsilon^{(n-1)n/2}\Delta(1,2,\ldots,n)\int_{\mathbb{R}^n}\mathbf{1}_{y_1<\ldots<y_n}\times e^{-\frac{1}{2}\sum_{j=1}^n y_j^2}\Delta(y_1,\ldots,y_n)\,dy_1\ldots dy_n$$

which vanishes even faster as $\varepsilon\downarrow 0$ than one might have thought. Further, in the conditional density, the main contribution is the first summand (divided by the non-intersection probability)

$$\rho_\varepsilon(y_1,\ldots,y_n\,\big|\,\mathrm{NI}(1)) = (1+O(\varepsilon))\frac{1}{Z}e^{-\frac{1}{2}\sum_{j=1}^n y_j^2}\Delta(y_1,\ldots,y_n)$$

where $Z = \Delta(1,2,\ldots,n)\int_{\mathbb{R}^n}\mathbf{1}_{y_1<\ldots<y_n}\times e^{-\frac{1}{2}\sum_{j=1}^n y_j^2}\Delta(y_1,\ldots,y_n)\,dy_1\ldots dy_n$ is a normalization constant (it can be found explicitly, but that is not important here). When $\varepsilon\downarrow 0$, we get the conditional density of $(W_1(1),\ldots,W_n(1))$ for n standard Brownian motions conditioned to not intersect

up to time 1:

$$\rho(y_1, \ldots, y_n \mid \mathrm{NI}(1)) = \frac{1}{Z} e^{-\frac{1}{2}\sum_{j=1}^{n} y_j^2} \Delta(y_1, \ldots, y_n).$$

This remarkable density arises in entirely unrelated looking context of random matrix theory! Look up GOE (Gaussian orthogonal ensemble), if you want to read more about it.

CHAPTER 5

Martingales in Brownian motion

Using the strong Markov property, we found the distribution of the first passage times τ_a. It can be thought of as the exit time of a half-infinite interval. A natural question is to find the distribution of the exit time $\tau_{b,a}$ of a finite interval $[b, a]$ for $b < 0 < a$. In particular, since $\tau_{-a,a} \leqslant t$ if and only if $\max_{s \leqslant t} |W_s| \geqslant a$, this will also tell us the distribution of the running maximum of a reflected Brownian motion.

The tools we use are martingales inside Brownian motion. In this section we see several such martingales.

1. Martingales in Brownian motion

In this section we shall find a few martingales constructed from Brownian motion. A whole class of martingales arise from Wiener integral and the Ito integral which we shall see later.

1.1. Polynomial martingales. We know that W_t itself is a martingale. But W_t^2 is not. Indeed,

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

From this, we can deduce that $W_t^2 - t$ is a martingale. Similarly,

$$\begin{aligned}\mathbb{E}[W_t^3 \,\big|\, \mathcal{F}_s] &= \mathbb{E}[(W_s + (W_t - W_s))^3 \,\big|\, \mathcal{F}_s] \\ &= \mathbb{E}[W_s^3 + 3W_s^2(W_t - W_s) + 3W_s(W_t - W_s)^2 + (W_t - W_s)^3 \,\big|\, \mathcal{F}_s] \\ &= W_s^3 + 3W_s(t - s).\end{aligned}$$

From this, we deduce that $W_t^3 - 3tW_t$ is a martingale. Continuing, we find that $W_t^4 - 6tW_t^2 + 3t^2$ is a martingale. The four martingales we have found can be written as $t^{d/2}P_d(W_t/\sqrt{t})$ where (1) $P_1(x) = x$, (2) $P_2(x) = x^2 - 1$, (3) $P_3(x) = x^3 - 3x$, (4) $P_4(x) = x^4 - 6x^2 + 3$. This suggests that for each $d \geqslant 1$, there must be polynomials P_d of degree d so that $t^{d/2}P_d(W_t/\sqrt{t})$ is a martingale. See if you can spot a pattern.

To not lose sight of the wood for the trees, note that the most important of all these martingales are W_t and $W_t^2 - t$. The exponential martingale that comes next is also very important.

1.2. Exponential martingales. Let $\lambda \in \mathbb{R}$ and define $M_\lambda(t) := e^{\lambda W_t - \frac{1}{2}\lambda^2 t}$. Then,

$$\mathbb{E}[M_\lambda(t) \,\big|\, \mathcal{F}_s^+] = e^{\lambda W_s - \frac{1}{2}\lambda^2 t}\mathbb{E}[e^{\lambda(W_t - W_s)}] = e^{\lambda W_s - \frac{1}{2}\lambda^2 t} e^{\frac{1}{2}\lambda^2(t-s)} = M_\lambda(s).$$

Thus, for each $\lambda \in \mathbb{R}$ we have a martingale $M_\lambda(t)$, $t \geqslant 0$.

Consider the power series expansion of function $e^{\lambda x - \frac{1}{2}\lambda^2} = \sum_{n=0}^{\infty} \frac{1}{n!} H_n(x)\lambda^n$ where

$$H_n(x) = \frac{d^n}{d\lambda^n} e^{\lambda x - \frac{1}{2}\lambda^2}\Big|_{\lambda=0}.$$

It is easy to see that $H_n(x)$ is a polynomial of degree n in x. These are called *Hermite polynomials*. By explicit computation one can see that $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, etc. These are precisely the polynomials we saw in the previous section!

Exercise 12

Use differentiation under the integral sign and the fact that M_λ is a martingale, to show that $t^{n/2} H_n(W_t/\sqrt{t})$ is a martingale for every $n \geqslant 0$.

The following exercise shows the exponential martingale for higher dimensional Brownian motion.

Exercise 13

Let W be d-dimensional Brownian motion. For any $v \in \mathbb{R}^d$, show that $M(t) = e^{\langle W(t), v\rangle - \frac{1}{2}\|v\|^2 t}$ is a martingale.

1.3. Functions of a Brownian motion. The previous martingales we found are of the form $F(W_t)$ or $F(t, W_t)$. Can we find general conditions for a function to yield a martingale? We proceed to do so.

Let W be d-dimensional Brownian motion (started anywhere). Let $F : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ be a function which we assume is sufficiently nice to allow various manipulations below (we can write the precise conditions later). A clearly necessary condition for $F(t, W_t)$ to be a martingale is that $\mathbb{E}[F(t, W_t)]$ is constant in time. To see what that condition implies, let us first recall the *heat kernel* $p_t(x) = \frac{1}{(2\pi t)^{d/2}} e^{-\|x\|^2/2t}$, which is the density of W_t. The most important fact about $p_t(x)$ is that it satisfies the *heat equation*

$$\frac{\partial}{\partial t} p_t(x) = \frac{1}{2}\Delta p_t(x), \qquad \text{where } \Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} \text{ is the } \textit{Laplacian} \text{ on } \mathbb{R}^d.$$

Therefore, freely differentiating under the integral sign (justifications later), the constancy of $\mathbb{E}_y[F(t, W_t)] = \int_{\mathbb{R}^d} F(t, x-y)p_t(x)dx$ can be written as

$$\begin{aligned}
0 &= \frac{d}{dt}\int_{\mathbb{R}^d} F(t,x-y)p_t(x)dx \\
&= \int_{\mathbb{R}^d} (\partial_t F(t,x-y)p_t(x) + F(t,x-y)\partial_t p_t(x))dx \\
&= \int_{\mathbb{R}^d} \partial_t F(t,x-y)p_t(x)dx + \frac{1}{2}\int_{\mathbb{R}^d} F(t,x-y)\Delta p_t(x)dx
\end{aligned}$$

using the heat equation. Now integrate by parts twice to get

$$\int_{\mathbb{R}^d} F(t,x-y)\Delta p_t(x)dx = \int_{\mathbb{R}^d} p_t(x)\Delta F(t,x-y)dx$$

where the boundary terms vanish provided F and its first partial derivatives are $e^{o(\|x\|^2)}$ as $\|x\| \to \infty$. Hence we get the condition

$$\begin{aligned}
0 &= \int_{\mathbb{R}^d} \left(\partial_t F(t,x-y) + \frac{1}{2}\Delta F(t,x-y)\right) p_t(x)dx \\
&= \int_{\mathbb{R}^d} \left(\partial_t F(t,x) + \frac{1}{2}\Delta F(t,x)\right) p_t(x-y)dx.
\end{aligned}$$

Observe that $N(0, tI_d)$ converges weakly to δ_0. Using this, we see that for all x we must have $\partial_t F(t,x) + \frac{1}{2}\Delta F(t,x) = 0$.

Now that we have found the condition, we can go back and redo the calculations (which we skip now) to arrive at the following result.

Theorem 5

Let $F \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ be such that $|F(t,x)| + \sum_{k=1}^d |\partial_{x_k} F(t,x)| \leqslant C_t e^{\varepsilon\|x\|^2}$ for any $\varepsilon > 0$ and $C_t < \infty$. If $\partial_t F(t,x) + \frac{1}{2}\Delta F(t,x) = 0$ for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, then $F(t, W_t)$ is a martingale.

An important class of examples is when $F(t,x) = u(x)$, where u is a harmonic function on $\mathbb{R}^d$ (recall that a harmonic function is one whose Laplacian vanishes identically). Then $u(W_t)$ is a martingale[1]. The examples we saw in the previous sections may also be seen as special cases of the above theorem.

Example 6

Let $F(t,x) = e^{\langle \nu, x\rangle - \frac{1}{2}\|\nu\|^2 t}$ where $\nu \in \mathbb{R}^d$ is fixed. Then $\partial_t F(t,x) = -\frac{1}{2}\|\nu\|^2 F(t,x)$ and $\partial^2_{x_k} F(t,x) = F(t,x)\nu_k^2$, from which we see that $\partial_t F(t,x) + \frac{1}{2}\Delta F(t,x) = 0$.

[1] And $u(X_t)$ is a submartingale/supermartingale if u is subharmonic/superharmonic. This is in fact the origin of the terminology of sub and super martingales.

The usefulness of martingales is via the optional sampling theorem. We showed in class how to analyse the exit time of an interval by one-dimensional Brownian motion. And also how to find martingales for multi-dimensional Brownian motion. For instance, any $u : \mathbb{R}_+ \times \mathbb{R}^d$ that satisfies $\partial_t u(t,x) + \frac{1}{2}\Delta u(t,x) = 0$ and some growth conditions gives a martingale $u(t, W_t)$. In particular, harmonic functions v satisfying some growth conditions give the martingales $v(W_t)$.

We used these to prove recurrence and transience properties of Brownian motion. We may touch upon the Dirichlet problem in the last lecture (if we have time).

Read up on these in the books we have been referring to.

Like with discrete time martingales, optional stopping theorem is a great tool. We state a basic version.

Theorem 6: Optional stopping theorem

Let $X = (X_t)_{t\geqslant 0}$ be a martingale w.r.t. a filtration $\mathcal{F}_\bullet = (\mathcal{F}_t)_{t\geqslant 0}$. Assume that the sample paths of X are continuous. Let τ be a stopping time for $\mathcal{F}_\bullet$ and define $X^\tau(t) := X(\tau \wedge t)$. If X^τ is uniformly integrable, then $\mathbb{E}[X(\tau)] = \mathbb{E}[X(0)]$.

1.4. Recurrence and transience of Brownian motion.

2. Wiener integral

Let W be standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $f : \mathbb{R}_+ \to \mathbb{R}$ where $\mathbb{R}_+ = [0,\infty)$. We want to make sense of $\int_0^t f(s)dW(s)$ with extra conditions on f if necessary.

Let us first review what can be done in the non-random situation, where the integrating function is fixed.

- Let $\alpha \in C^1(\mathbb{R}_+)$. Then for any $f \in C(\mathbb{R}_+)$ we may define $\int_0^t f(s)d\alpha(s)$ as $\int_0^t f(s)\alpha'(s)ds$, the latter being the Riemann integral of a continuous function.
- More generally, if α is a function of bounded variation[2] on $[0,T]$ for every T, then following ideas similar to that of Riemann intergral, Stieltjes showed that $\int_0^T f(t)d\alpha(t)$ can be made sense of for any $f \in C[0,1]$.
- Suppose $\alpha \in C(\mathbb{R}_+)$, not necessarily of bounded variation. Then it is no longer possible to define Stieltjes' integral. But for $f \in C^1(\mathbb{R}_+)$, we can define

$$\int_0^t f(s)d\alpha(s) := f(t)\alpha(t) - f(0)\alpha(0) - \int_0^t \alpha(s)f'(s)ds.$$

[2]By definition, α is said to have bounded variation if $\sup \sum_{k=1}^n |\alpha(t_k) - \alpha(t_{k-1})|$ is finite, where the supremum is over all $0 = t_0 < t_1 < \ldots < t_n = T$. It is a fact that a function is of bounded variation if and only if it can be written as a difference of two increasing functions.

The justification for this definition is that when α is of bounded variation, the expression on the right is equal to $\int_0^t f d\alpha$, known as the integration by parts formula.

This simple observation has considerable reach, and lies at the base of the theory of distributions in functional analysis. Any continuous function acts on smooth enough functions as above.

Now fix a sample path of Brownian motion. It is not of bounded variation, hence the first two approaches do not work. That is, we cannot make sense of $\int_0^1 f(t)dW(t)$ for all $f \in C[0,1]$. However, the sample path is indeed continuous, hence we can use the third approach and define $\int_0^1 f(t)dW(t)$ for $f \in C^1$ by the integration by parts formula.

But we can do more - we shall in fact define $\int_0^t f(s)dW(s)$ for every $f \in L^2[0,t]$! This is done as follows.

Step 1: Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a step function, $f(t) = \sum_{k=1}^n \lambda_k \mathbf{1}_{[a_k,b_k]}(t)$ for some $0 \leqslant a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n$ for some $n \geqslant 1$. Then we define

$$I(f) = \sum_{k=1}^n \lambda_k (W(b_k) - W(a_k)).$$

Let $\mathcal{S}$ denotes the collection of all step functions on $\mathbb{R}_+$, a dense subspace of $L^2(\mathbb{R}_+)$. What we have defined is a function $I : \mathcal{S} \to L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Step 2: We claim that $I : \mathcal{S} \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a linear isometry. Further, $I(f)$ is a Gaussian random variable for each $f \in \mathcal{S}$.

Linearity is clear. To check isometry, by the independent increments property of W, we get

$$\|I(f)\|^2_{L^2(\mathbb{P})} = \mathbb{E}[|I(f)|^2] = \mathrm{Var}\left(\sum_{k=1}^n \lambda_k (W(b_k) - W(a_k))\right) = \sum_{k=1}^n \lambda_k^2 (b_k - a_k) = \|f\|^2_{L^2(\mathbb{R}_+)}.$$

That $I(f)$ is a mean zero Gaussian is clear. Therefore it has $N(0, \|f\|^2_{L^2(\mathbb{R}_+)})$ distribution.

Step 3: Thus I maps Cauchy sequences in $\mathcal{S}$ to Cauchy sequences in $L^2(\mathbb{P})$. Hence, if $f_n \in \mathcal{S}$ and $f \in L^2(\mathbb{R}_+)$ and $f_n \to f$ in $L^2(\mathbb{R}_+)$, then $\{f_n\}$ is Cauchy in $L^2(\mathbb{R}_+)$ and therefore $\{I(f_n)\}$ is Cauchy in $L^2(\mathbb{P})$. By completeness of $L^2(\mathbb{P})$, $I(f_n)$ has a limit. Clearly this limit depends only on f and not on the sequence $\{f_n\}$. Therefore, we can unambiguously extend I to a linear isometry of $L^2(\mathbb{R}_+)$ into $L^2(\mathbb{P})$.

Since L^2-limits of Gaussians are Gaussians, it follows that for any $f, g in L^2[0,1]$, the distribution of $I(f)$ and $I(g)$ is bivariate Gaussian with zero means and $\mathrm{Cov}(I(f), I(g)) = \int_0^1 fg$. In particular, $\mathrm{Var}(I(f)) = \|f\|^2_{L^2[0,1]}$.

Step 4: This defines the stochastic integral and we usually write $\int_0^\infty f(t)dW(t)$ for $I(f)$. We can extend its domain to $L^2_{loc}(\mathbb{R}_+)$ (i.e., $f : \mathbb{R}_+ \to \mathbb{R}$ such that $\int_0^T f^2(s)ds < \infty$ for all $T < \infty$) by defining

$$\int_0^t f(s)dW(s) = I(f\mathbf{1}_{[0,t]}) \text{ for } t \in \mathbb{R}_+.$$

This completes the construction of the Wiener integral.

Wiener integral gives rise to many martingales as the following exercise shows.

Exercise 14

Let $f \in L^2_{loc}(\mathbb{R}_+)$ and define $X_t = \int_0^t f(s)dW(s)$ and $A_t = \int_0^t f^2(s)ds$. Show that X is a martingale and that $X^2 - A$ is a martingale.

But even more than that, it is a precursor to the more powerful *Ito integral* which makes sense of integrals of random functions such as $\int_0^t W_s\,dW_s$ and gives rise to many more martingales. For example, it turns out that this integral is $\frac{1}{2}W_t^2 - \frac{1}{2}t$ in contrast to C^1 functions α with $\alpha(0) = 0$ for which we always have $\int_0^t \alpha(s)d\alpha(s) = \frac{1}{2}\alpha(t)^2$).

2.1. Some remarks. How was it possible to integrate every L^2 function? The point to remember is that when talking about Brownian motion, we are not talking of one function, but an entire ensemble of them. Therefore,

(1) For any given Brownian path, there is a function $f \in L^2(\mathbb{R}_+)$ (even $f \in C[0,1]$) that cannot be integrated in any sense against the Brownian path.

(2) For a fixed $f \in L^2(\mathbb{R}_+)$, this problem does not arise for almost every Brownian path and we can integrate f with respect to W.

(3) For almost every Brownian path, the integrals of all C^1 functions can be simultaneously defined (using the integration by parts formula).

In this sense, Brownian motion is better than a distribution, it can integrate functions with hardly any smoothness. A good analogy is to consider sums instead of integrals.

Stochastic summation: Consider $\ell^2 = \{x = (x_n)_{n\in\mathbb{N}} : \sum_n x_n^2 < \infty\}$. Let $a = (a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$. Two observations.

- If $a \in \ell^2$, then $\sum_n a_n x_n$ converges for every $x \in \ell^2$. This is the inner product in ℓ^2, well-defined because of the Cauchy-Schwarz inequality.
- Suppose $\sum_n a_n x_n$ converges for each $x \in \ell^2$. Then $a \in \ell^2$. To see this, define $L_m : \ell^2 \to \mathbb{R}$ by $L_m(x) = \sum_{k\leqslant m} a_k x_k$. Then L_m is a bounded linear functional with $\|L_m\|^2 = \sum_{k\leqslant m} a_k^2$. By the hypothesis, for ach $x \in \ell^2$, the sequence $\{L_m(x)\}_m$ is convergent in $\mathbb{R}$,

and hence bounded in $\mathbb{R}$. By the uniform boundedness principle, $\{\|L_m\|\}$ is bounded. Thus $\sum_{k\leqslant m} a_k^2 \leqslant C$ for some C and for all m which implies that $a \in \ell^2$.

Now consider $\xi = (\xi_n)_{n\in\mathbb{N}}$, where ξ_n are i.i.d. $N(0,1)$ random variables on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, $\xi_n > 1$ infinitely often, w.p.1. and hence $\xi \notin \ell^2$, w.p.1. Thus, for almost every ω, there is an $x \in \ell^2$ such that $\sum_n \xi_n(\omega)x_n$ does not converge. However, for each $x \in \ell^2$, by Khinchine's theorem on sums of independent random variables, it follows that $\sum_n \xi_n x_n$ converges w.p.1. But let us do it in a more roundabout way to bring out the analogy with the Wiener integral.

Step 1: Let $\mathcal{S} = \{x \in \ell^2 : x_n = 0 \text{ for all large } n\}$, a dense subspace of ℓ^2. For $x \in \mathcal{S}$, the sum $I(x) := \sum_n \xi_n x_n$ is a finite sum and therefore well-defined.

Step 2: $I : \mathcal{S} \mapsto L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a linear isometry. In fact, for each $x \in \ell^2$, $I(x) \sim N(0, \|x\|^2)$. This is easy to see by computing $\mathbb{E}[I(x)^2]$.

Step 3: I extends as an isometry of ℓ^2 into $L^2(\Omega, \mathcal{F}, \mathbb{P})$. This step is carried out exactly the same way.

Interpret $I(x)$ as $\sum_n \xi_n x_n$ (in fact, the latter series converges almost surely, using standard theorems on sums of independent random variables). Thus, we have a very close analogy with the previous situation. To make the analogy even closer, you may want to define $S_n = \xi_1 + \ldots + \xi_n$ so that $\sum_n \xi_n x_n = \sum_n (S_n - S_{n-1})x_n$ looks like "$\int x(n)dS(n)$".

2.2. Wiener integral as a process. Let W be a $\mathcal{F}_\bullet$-Brownian motion. Fix $f \in L^2_{\text{loc}}(\mathbb{R}_+)$ and let $X_t = \int_0^t f(s)dW(s) = I(f\mathbf{1}_{[0,t]})$. We want to view $X = (X_t)_{t\geqslant 0}$ as a process. But the way we constructed Wiener integral, we fixed the integrand, and then defined its integral as an L^2-limit. Here we have uncountably many functions, $f\mathbf{1}_{[0,t]}$, one for each t. In view of the problems in defining $I(f)$ for all $f \in L^2(\mathbb{R}_+)$ discussed above, can we define $X_t(\omega)$ for all t, for a.e. ω? If so, what kind of regularity of sample paths may we expect?

Step L^2 functions: For $f \in \mathcal{S}$, say $f(t) = \sum_{k=1}^n \lambda_k \mathbf{1}_{[a_k,b_k]}(t)$, we take

$$X_t = \sum_{k=1}^n \lambda_k (W(b_k \wedge t) - W(a_k \wedge t)).$$

From this expression, the following properties easily follow (exercise!).

(1) Defined as above, X has continuous sample paths and is adapted to $\mathcal{F}_\bullet$.

(2) X has independent increments, i.e., $X_t - X_s$ is independent of $\mathcal{F}_s$ for any $s < t$.

(3) $X_t - X_s \sim N(0, A_t - A_s)$ where $A_t = \int_0^t f^2(u)du$.

(4) X_t and $X_t^2 - A_t$ are martingales.

General $L^2_{\text{loc}}(\mathbb{R}_+)$ functions:. Let $f \in L^2_{\text{loc}}(\mathbb{R}_+)$. Find $f_n \in \mathcal{S}$ such that $\|f_n - f\|_{L^2[0,T]} \to 0$ for all $T < \infty$. Let $X_n = \int_0^t f_n(s)dW(s)$ be defined as above for step L^2 functions. Then $X_n - X_m$ is a square integrable martingale, and by Doob's maximal inequality we see that

$$\mathbb{E}\left[\|X_n(t) - X_m(t)\|^2_{\sup[0,T]}\right] \leqslant C\mathbb{E}[|X_n(T) - X_m(T)|^2] \;=\; C\|f_n - f_m\|^2_{L^2[0,T]}$$

which goes to zero as $n, m \to \infty$. Pick a sequence $n_1 < n_2 < \dots$ so that $\|f_{n_k} - f_{n_{k+1}}\|^2_{L^2[0,T]} \leqslant \frac{1}{2^k}$. Then summing the above inequality we see that

$$\sum_{k \geqslant 1} \|X_n - X_m\|_{\sup[0,T]} < \infty \quad \text{a.s.}$$

as the left side quantity has finite expectation. This shows that X_{n_k} is uniformly Cauchy on $[0, T]$. Applying this to $T \in \mathbb{N}$ and using a diagonal argument, we can extract a subsequence which is uniformly Cauchy on every bounded interval. The limiting process, denoted X, is then a continuous adapted process. For each t, we have the almost sure and hence L^2 convergence of $X_n(t)$ to $X(t)$, which shows that $X(t)$ is a version of $\int_0^t f(s)dW(s)$. Thus, we have defined $\int_0^t f(s)dW(s)$ as a continuous adapted process.

The rest of the properties easily follow by taking limits: (1) X has independent increments, i.e., $X_t - X_s$ is independent of $\mathcal{F}_s$ for any $s < t$. (2) $X_t - X_s \sim N(0, A_t - A_s)$ where $A_t = \int_0^t f^2(u)du$. (3) X_t and $X_t^2 - A_t$ are martingales.

2.3. Time change. Let $f \in L^2_{\text{loc}}(\mathbb{R}_+)$ and let $X(t) = \int_0^t f(s)dW(s)$, defined as a continuous adapted process. Observe that $t \mapsto A_t = \int_0^t f^2(s)ds$ is an increasing, continuous function. Let $\theta(u) = \inf\{t : A_t > u\}$, which is a sort of inverse of A. Define $Y_u = X_{\theta(u)}$ for $u < A(\infty) = \|f\|^2_{L^2(\mathbb{R}_+)}$ (which could be ∞), a time-change of X. Even though θ need not be continuous (because A need not be strictly increasing), it is right continuous and the process Y has continuous sample paths. As a time reparameterization of X, it inherits the independent increment property. Lastly, $Y_v - Y_u = X_{\theta(v)} - X_{\theta(u)}$ is Gaussian with zero mean and variance equal to $A(\theta(v)) - A(\theta(u)) = v - u$. In other words, Y is just a Brownian motion!

All the martingales we can get as Wiener integrals are in fact Brownian motions in disguise! But in reality, this fact is even more general. We shall see that any continuous martingale is a time change of some Brownian motion.

3. A peek into Ito's integration theory

Let W be standard Brownian motion in one dimension, defined on some $(\Omega, \mathcal{F}, \mathbb{P})$. The Wiener integral made sense of $\int_0^1 f(t)dW(t)$ for $f \in L^2[0, 1]$ as a random variable on the same probability

space. Can we also integrate random integrands. In other words, suppose $X = (X(s))_{s \geqslant 0}$ is a stochastic process on the same probability space. Can we make sense of $\int_0^1 X(s)dW(s)$? If X is independent of W, then by conditioning on X, we reduce this to the situation of deterministic integrands. Hence we are more interested in the situation where X depends on W.

In the basic theory of Stieltjes' integral of f with respect to α, one starts out by considering partitions $0 = t_0 < t_1 < \ldots < t_N = 1$ and forming the Riemann sum $\sum_{k=0}^{N-1} f(t_k^*)(\alpha(t_{k+1}) - \alpha(t_k))$ where $t_k^* \in [t_k, t_{k+1}]$ are arbitrary points. If the limit of these sums exists, as the partitions get finer and over arbitrary choices of t_k^*s, then the limit value is defined to be $\int_0^1 f(t)d\alpha(t)$.

Taking inspiration from this (all integration must be limit of summation, after all), we can try to make sense of $\int_0^1 X(s)dW(s)$ by considering

$$\sum_{k=0}^{N-1} X(t_k^*)(W(t_{k+1}) - W(t_k)).$$

It turns out that the choice of t_k^* matters a lot! Henceforth, we consider the prototypical case of $X = W$. By taking t_k^* to be the left end-point or the right end-point or the mid-point of $[t_{k-1}, t_k]$, we get the following sums.

$$I_N := \sum_{k=0}^{N-1} W(\frac{k}{2^n})(W(\frac{k+1}{2^n}) - W(\frac{k}{2^n})),$$

$$J_N := \sum_{k=0}^{N-1} W(\frac{k+1}{2^n})(W(\frac{k+1}{2^n}) - W(\frac{k}{2^n})),$$

$$K_N := \sum_{k=0}^{N-1} W(\frac{k+\frac{1}{2}}{2^n})(W(\frac{k+1}{2^n}) - W(\frac{k}{2^n})).$$

Here is a simple calculation that shows that even if they have limits, the limits must be different.

$$J_N - I_N = \sum_{k=0}^{N-1} (W(\frac{k+1}{2^n}) - W(\frac{k}{2^n}))^2 \overset{a.s.}{\to} 1$$

where the almost sure convergence was a homework exercise (if we consider a general sequence of partitions, we only get convergence in probability, but that is not relevant for the point we are about to make). Thus, the limits of J_N and of I_N, if they exist, must differ by 1. In fact the limits exist because

$$J_N + I_N = \sum_{k=0}^{N-1} (W(\frac{k+1}{2^n})^2 - W(\frac{k}{2^n})^2) = W(1)^2.$$

Therefore, we deduce that

$$J_N \overset{a.s.}{\to} \frac{1}{2}W(1)^2 + \frac{1}{2}, \quad I_N \overset{a.s.}{\to} \frac{1}{2}W(1)^2 - \frac{1}{2}.$$

We leave it as an exercise to check that $K_N \overset{a.s.}{\to} \frac{1}{2}W(1)^2$. There is no difficulty in extending this for any fixed t and get three possible candidates $\frac{1}{2}W(t)^2 + \frac{1}{2}t$, $\frac{1}{2}W(t)^2 - \frac{1}{2}t$ and $\frac{1}{2}W(t)^2$ as possible definitions of $\int_0^t W(s)dW(s)$. Of course, choosing other t_k^*, one can get other candidates. Which is the right one?

It is tempting to choose the limit of K_N, since choosing the mid-point makes one seem less prejudiced to the temptations of left and right, and also because it agrees with the result for Stieltjes' integrals, $\int_0^t \alpha(t)d\alpha(t) = \frac{1}{2}\alpha(t)^2 - \frac{1}{2}\alpha(0)^2$. But when we consider the integral as a process in t (caution: We only showed that for fixed t, the limit exists, but let us suppose that the integral makes sense as a process in t), the leftist choice $\frac{1}{2}W(t)^2 - \frac{1}{2}t$ is more inviting, since it alone is a martingale! Why did it become a martingale? A hint is there already in the definition of I_N. If you consider the discrete-time martingale $W(k/2^n)$, $k = 0, 1, 2, \ldots$, then I_N is like enhancing the kth game by betting a *predictable* amount $W(k/2^n)$. The choices in J_N and K_N need knowledge of the future to make the bet, hence they fail to be martingales.

Ito integral: Let W be standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ and let X be a continuous stochastic process that is adapted to $\overline{\mathcal{F}_\bullet}^+$ (right-continuos, completed filtration). Let us assume that X is uniformly bounded (this can be relaxed if X is unlikely to be very large on bounded intervals). Then the limits

$$I_X(t) := \lim_{n\to\infty} \sum_{k=0}^{\lfloor 2^n t\rfloor} X(k/2^n)(W(\frac{k+1}{2^n}) - W(\frac{k}{2^n}))$$

exists as a process in t, and I_X is a continuous martingale. It is called the *Ito integral* of X with respect to W and denoted $I_X(t) = \int_0^t X(s)dW(s)$.

We have just stated this fact without proof. In specific cases, one can do this by hand. If nothing else, this may be thought of as a systematic way to obtain many martingales from Brownian motion!

Exercise 15

At least for a few small values of $p = 1, 2, \ldots$, show that

$$\int_0^t W(s)^p dW(s) = \frac{1}{p+1}W(t)^{p+1} - \frac{p}{2}\int_0^t W(s)^{p-1}ds.$$

3.1. Doob-Meyer decomposition.

4. Cameron-Martin formula

You are shown a random trajectory, which is either the path of a Brownian motion $(W_t)_{0\leqslant t\leqslant 1}$ or that of $(W_t + t)_{0\leqslant t\leqslant 1}$. Can you tell for sure which it is? This is a question of whether the two

distributions on $C[0,1]$ are singular or not. Generally speaking, on large spaces measures tend to sit on different regions of the space. On infinite dimensional spaces, being singular is the default behaviour. For example, any two i.i.d. product measures are singular, unless they are the same. On $C[0,1]$, the distributions of $(2W_t)_{0\leqslant t\leqslant 1}$ or $(W_{t/2})_{0\leqslant t\leqslant 1}$ are singular to Wiener measure. However, in very special situations, two measures can be mutually absolutely continuous. We discuss one such situation here.

Cameron Martin space: Let H denote the space of $f : [0,1] \to \mathbb{R}$ that have an L^2 derivative and satisfy $f(0) = 0$. What this means is that there is some $g \in L^2[0,1]$ such that $f(t) = \int_0^t g(s)ds$. For a given $f \in H$, clearly there is a unique $g \in L^2[0,1]$ (i.e., any two candidates are equal a.e.) which we call the derivative and denote as f'. It is clear that $C^1[0,1] \subseteq H \subseteq C[0,1]$, but functions in H need not have pointwise derivatives. For example, $t \mapsto |t - 0.2|$ is in H, the derivative being $\mathbf{1}_{[0.2,1]} - \mathbf{1}_{[0,0.2]}$. The function $t \mapsto t^p$ is in H if and only if $p > \frac{1}{2}$.

Theorem 7

Let μ denote the Wiener measure on $C[0,1]$. Let $f \in C[0,1]$ and let $\nu(\cdot) = \mu(\cdot - f)$ denote the distributions of $W + f$, where W is SBM on $[0,1]$.

(1) If $f \notin H$, then $\mu \perp \nu$.

(2) If $f \in H$, then μ and ν are mutually absolutely continuous and

$$\frac{d\nu}{d\mu}(\omega) = e^{-\int_0^1 f'(s)d\omega(s) - \frac{1}{2}\int_0^1 f^2(s)ds}.$$

To clarify the statement here, the term $\int_0^1 f'(s)d\omega(s)$ is the Wiener integral on f', which not defined pointwise for $\omega \in C[0,1]$, but well-defined a.s.$[\mu]$. And Radon-Nikodym derivative is also defined only a.s.$[\mu]$, so there is no problem of interpretation. Another way of stating this is that if W is SBM defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and we define a new probability measure $\mathbb{Q}$ on $\mathcal{F}$ by

$$d\mathbb{Q} = e^{\int_0^1 f'(s)dW(s) - \frac{1}{2}\int_0^1 |f'(s)|^2 ds} d\mathbb{P},$$

then $W + f$ is a SBM under $\mathbb{Q}$ (if $f \in H$, of course).

Proof. Let $D_n = \{k2^{-n} : 0 \leqslant k \leqslant 2^n\}$ and $D = \bigcup_{n\geqslant 0} D_n$. On $(C[0,1], \mathcal{B}_{\mathbb{C}[0,1]})$ with projection maps Π_t for $0 \leqslant t \leqslant 1$, define the filtration $\mathcal{F}_n = \sigma\{\Pi_t : t \in D_n\}$. Then $\sigma(\cup_n \mathcal{F}_n) = \mathcal{B}_{\mathbb{C}[0,1]}$. Let $\mu_n = \mu|_{\mathcal{F}_n}$ and $\nu_n = \nu|_{\mathcal{F}_n}$.

For $g \in C[0,1]$, let $\Delta_k g = g((k+1)/2^n) - g(k/2^n)$. If W is SBM, then $\Delta_k W$ are i.i.d. $N(0, 2^{-n})$, while $\Delta_k(W+f)$ are independent $N(\Delta_k f, 2^{-n})$. Therefore,

$$Y_n := \frac{d\nu_n}{d\mu_n}(\omega) = \exp\left\{-\frac{1}{2}\left[2^n \sum_{k=0}^{2^n-1} (\Delta_k \omega - \Delta_k f)^2 + 2^n \sum_{k=0}^{2^n-1} (\Delta_k \omega)^2\right]\right\}$$

$$= \exp\left\{2^n \sum_{k=0}^{2^n-1} (\Delta_k \omega)(\Delta_k f) - \frac{1}{2} 2^n \sum_{k=0}^{2^n-1} (\Delta_k f)^2\right\}. \tag{13}$$

As Y_n is a μ-martingale, we know that $Y_n \to Y$ a.s.$[\mu]$ for some Y and we know from martingales class that (to be pedantic, $\{Y = \infty\}$ below is a short form for $\{\lim Y_n = \infty\}$)

$$\nu(A) = \int_A Y d\mu + \nu(A \cap \{Y = \infty\}) \text{ for all } A. \tag{14}$$

The goal is to show that only one of the two components in (14) survive, depending on whether $f \in H$ or not. To this end, define $G_n = 2^n \sum_{k=0}^{2^n-1} (\Delta_k f) \mathbf{1}_{J_{n,k}}$ where $J_{n,k} = [k2^{-n}, (k+1)2^{-n}]$. Then the exponent in Y_n is equal to $\int_0^1 G_n d\omega - \frac{1}{2}\int_0^1 G_n^2$.

Case $f \in H$: We claim that $G_n \to f'$ in $L^2[0,1]$ which is an exercise in real analysis[3]. Accepting this, the isometry property of Wiener integral implies that the exponent in Y_n converges to $\int_0^1 f' d\omega - \frac{1}{2}\int_0^1 |f'|^2$ in $L^2(\mu)$, and hence in probability. As almost sure and in probability limits must be same, we get $Y(\omega) = \exp\{\int_0^1 f' d\omega - \frac{1}{2}\int_0^1 |f'|^2\}$. But under μ, we know that $\int_0^1 f' d\omega$ has $N(0, \int_0^1 |f'|^2)$ distribution, hence $\mathbb{E}_\mu[Y] = 1$ (recall the moment generating function of Gaussian). Therefore, $Y d\mu$ is already a probability measure, and the singular part in (14) must vanish. In other words, $d\nu = Y d\mu$, as claimed.

Case $f \notin H$: We use again the fact that under μ, the variables $\Delta_k \omega$ are i.i.d. $N(0, 2^{-n})$. Hence from (13), we get (use the moment generating function of Gaussian again)

$$\mathbb{E}_\mu[Y_n^p] = e^{2^{n-1} p(p-1) \sum_k (\Delta_k f)^2}.$$

If $f \notin H$, then $2^n \sum_k (\Delta_k f)^2 \to \infty$ a.s.$[\mu]$, which is another exercise in real analysis[4]. Take $p = \frac{1}{2}$ to see that $\mathbb{E}_\mu[\sqrt{Y_n}] \to 0$, and hence (by Markov's inequality) $Y_n \xrightarrow{P} 0$ w.r.t. μ. As $Y_n \xrightarrow{a.s.} Y$ a.s.$[\mu]$, it

[3] First assume that $g \in C[0,1]$. Then $G_n \to g$ pointwise. This is because if $x \in J_{n,k_n}$, then $G_n(x)$ is the average of g on J_{n,k_n}, and this interval shrinks to $\{x\}$. This also shows that $|G_n(x)| \leqslant \|g\|_{sup}$ pointwise. By DCT, $G_n \to g$ in $L^2[0,1]$. Now consider a general $g \in L^2$. Find $h \in C[0,1]$ such that $\|g-h\|_2 < \varepsilon$. On $J_{n,k}$, we have $|G_n - H_n|^2 \leqslant \frac{1}{|J_{n,k}|}\int_{J_{n,k}} (g-h)^2$, by Cauchy-Schwarz. Integrate over $J_{n,k}$ and sum over k to get $\|G_n - H_n\|_2 \leqslant \|g-h\|_2$. Now write

$$\|G_n - g\|_2 \leqslant \|G_n - H_n\|_2 + \|H_n - h\|_2 + \|h - g\|_2.$$

The first and third terms are bounded by ε. The middle term goes to 0 as $n \to \infty$, because $h \in C[0,1]$ (the first case). Thus, $\limsup \|G_n - g\|_2 \leqslant 2\varepsilon$. As ε is arbitrary, we see that $G_n \to g$ in $L^2[0,1]$.

[4]

follows that $Y = 0$ a.s.$[\mu]$. Thus the absolutely continuous part in (14) vanishes and we get $\nu \perp \mu$. ■

4.1. Brownian motion as a Gibbs measure. Given a set S with a reference measure ν and an "energy function" $H : S \to \mathbb{R}$, a Gibbs measure on S is a probability measure with density (w.r.t. ν) that is proportional to $e^{-H(x)}$. Many naturally occurring probability measures are profitably thought of this way. For example the standard Gaussian measure corresponds to $S = \mathbb{R}$, $\nu =$ Lebesgue measure and $H(x) = \frac{1}{2}x^2$. We now wish to interpret Wiener measure this way.

Let $W = (W_t)_{0 \leqslant t \leqslant 1}$ be a SBM. The density of $(W(k/2^n))_{1 \leqslant k \leqslant 2^n}$ w.r.t. Lebesgue measure on $\mathbb{R}^{2^n}$ at $u \in \mathbb{R}^{2^n}$ is

$$\frac{1}{(2\pi)^{2^{n-1}}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{2^n} 2^n (u_k - u_{k-1})^2\right\}.$$

If $f \in H$ with $f(k/2^n) = u_k$, then the exponent is approximately $\frac{1}{2}\int_0^1 |f'(s)|^2 ds$, as we saw in the proof of the Cameron-Martin theorem. This makes it tempting to say that Brownian motion is the Gibbs measure on $C[0,1]$ with the *Dirichlet energy* $H(f) = \frac{1}{2}\int_0^1 |f'(s)|^2 ds$. But this is meaningless, as there is no reference measure w.r.t. which we can express the density (there is no Lebesgure measure on $C[0,1]$ or any other infinite dimensional vector space)! All we have is the Wiener measure. Nevertheless, it is a useful and not entirely incorrect way to think about Brownian motion. Cameron-Martin formula gives us one way to make it precise.

Let μ be the Wiener measure. If $f \in H$, by the Cameron-Martin formula,

$$\mathbb{P}\{W \in B(f,\varepsilon)\} = \mathbb{P}\{W - f \in B(0,\varepsilon)\} = \int_{B(0,\varepsilon)} \rho_f(\omega) d\mu(\omega)$$

where $\rho_f(\omega) = \exp\{-\int_0^1 f'(s)d\omega(s) - \frac{1}{2}\int_0^1 |f'(s)|^2 ds\}$ is the Radon-Nikodym derivative of the distribution of $W - f$ w.r.t. Wiener measure. We claim that $|\int_0^1 f'(s)d\omega(s)| \leqslant 2\varepsilon\|f'\|_2$ for $\omega \in B(0,\varepsilon)$ a.s.$[\mu]$. If $f \in C^1[0,1]$,

$$\begin{aligned}\Big|\int_0^1 f'(s)d\omega(s)\Big| = \Big|f(1)\omega(1) - \int_0^1 \omega(s)f'(s)ds\Big| &= \Big|\int_0^1 (\omega(1) - \omega(s))f'(s)ds\Big| \\ &\leqslant \int_0^1 |\omega(1) - \omega(s)||f'(s)|ds.\end{aligned}$$

If $\omega \in B(0,\varepsilon)$, then $|\omega(1) - \omega(s)| \leqslant 2\varepsilon$ and using Cauchy-Schwarz we see that the above quantity is bounded by $2\varepsilon\|f\|_2$. We leave it as an exercise to prove the same assuming only that $f \in H$.

Then it follows that

$$e^{-\frac{1}{2}\int_0^1 |f'(s)|^2 ds - 2\varepsilon\|f\|_2} \leqslant \rho_f(\omega) \leqslant e^{-\frac{1}{2}\int_0^1 |f'(s)|^2 ds + 2\varepsilon\|f\|_2}.$$

Consequently,

$$e^{-\frac{1}{2}\int_0^1 |f'(s)|^2 ds - 2\varepsilon\|f\|_2} \leqslant \frac{\mathbb{P}\{W \in B(f,\varepsilon)\}}{\mathbb{P}\{W \in B(0,\varepsilon)\}} \leqslant e^{-\frac{1}{2}\int_0^1 |f'(s)|^2 ds + 2\varepsilon\|f\|_2}$$

which implies that

$$\lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}\{W \in B(f,\varepsilon)\}}{\mathbb{P}\{W \in B(0,\varepsilon)\}} = e^{-\frac{1}{2}\int_0^1 |f'(s)|^2 ds}.$$

This conclusion remains true for any $f \in C[0,1]$, provided we interpret $\int_0^1 |f'(s)|^2 ds$ as $+\infty$ for $f \notin H$ (so the right side is zero then). Thus the relative probabilities of Brownian motion being near different functions is proportional to $e^{-\frac{1}{2}\int_0^1 |f'(s)|^2 ds}$, as required of a Gibbs measure with Dirichlet energy.

CHAPTER 6

Random walks and Brownian motion

Part of the original motivation for Brownian motion was that it was a kind of random walk in continuous time and continuous space. We shall now make this precise and show that random walks converge to Brownian motion in the sense of distribution. For this we need to see them both as objects in the same space.

Random walk as a continuous stochastic process: Let $x_1, x_2, \ldots, x_n$ be real numbers. Define the continuous W_n by

$$W_n(t) = \begin{cases} x_1 + \ldots + x_k & \text{if } t = \frac{k}{n} \text{ for some } 0 \leqslant k \leqslant n, \\ \text{linear in each interval } [\frac{k}{n}, \frac{k+1}{n}]. \end{cases}$$

For later purposes, let us introduce the notation $W_n = \mathcal{T}(x_1, \ldots, x_n)$, so that $\mathcal{T} : \mathbb{R}^n \mapsto C[0,1]$. If x_i are random variables, then W_n is a continuous stochastic process whose distribution is the push-forward of the distribution of T under $\mathcal{T}$. When x_i are i.i.d. random variables, W_n is essentially the random walk with these steps, except that we interpolate continuously to make it a continuous process of continuous time.

Weak convergence in $C[0,1]$: Suppose μ_n, μ are Borel probability measures on a complete, separable metric space (X, d). We say that $\mu_n \xrightarrow{d} \mu$ if $\int f d\mu_n \to \int f d\mu$ for all $f \in C_b(X)$ (the space of bounded continuous functions on X). If $X_n \sim \mu_n$ and $X \sim \mu$ are random variables (not necessarily on the same probability space), we abuse notation and write $X_n \xrightarrow{d} X$ to mean $\mu_n \xrightarrow{d} \mu$. This is the notion of convergence in distribution or weak convergence (that we have studied extensively when the metric space is Euclidean space). In particular, this applies to probability measures on $C[0,1]$.

For now, let us only make the observation that if $F : X \mapsto \mathbb{R}$ is a continuous function, then $\mu_n \xrightarrow{d} \mu$ implies that $\mu_n \circ F^{-1} \xrightarrow{d} \mu \circ F^{-1}$ (these are probability measures on $\mathbb{R}$). Thus, convergence in distribution of one sequence of measures on $C[0,1]$ encodes innumerable convergence in distribution statements on the real line (just by varying F). A convergence in distribution statement on $C[0,1]$ (or other such "large spaces") are often called a *functional limit theorem*.

With these definitions, we are ready to state one of the foundational theorems of probability theory. For one, it is a far-reaching generalization of the central limit theorem.

Theorem 8: Donsker's invariance principle

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}[X_k] = 0$ and $\mathbb{E}[X_k^2] = 1$. Let $W_n = \mathcal{T}(\frac{X_1}{\sqrt{n}}, \ldots, \frac{X_n}{\sqrt{n}})$. Then $W_n \xrightarrow{d} W$, a standard Brownian motion on $[0,1]$.

How is this a generalization of the central limit theorem? Just consider the continuous function $F : C[0,1] \mapsto \mathbb{R}$ defined by $F(\phi) = \phi(1)$. As $F(W_n) = (X_1 + \ldots + X_n)/\sqrt{n}$ and $F(W) = W(1) \sim N(0,1)$, the standard central limit theorem follows. There are innumerable other functions one can use, and that gives us an amazing machinery to transfer results from random walks to Brownian motion or vice versa. Often from one particular random walk (e.g., simple symmetric random walk) to Brownian motion and hence to all other random walks with steps of zero mean and unit variance.

Why Donsker's theorem makes one break into a song: For example, let $F(\phi) = \max_{0 \leqslant t \leqslant 1} \phi(t)$. Then F is a continuous function from $C[0,1]$ to $\mathbb{R}$. From Donsker's theorem, it follows that

$$\max_{0 \leqslant t \leqslant 1} W_n(t) \xrightarrow{d} \max_{0 \leqslant t \leqslant 1} W(t).$$

The left hand side is just the maximum of $\{\frac{1}{\sqrt{n}} S_0, \ldots, \frac{1}{\sqrt{n}} S_n\}$ and the right hand side is what we have been calling M_1. Since we worked out that $M_1 \stackrel{d}{=} |Z|$ where $Z \sim N(0,1)$, we now have limiting distribution of $\max\{S_0, \ldots, S_n\}/\sqrt{n}$. Observe that the special tricks that we used to compute the distribution of M_1, namely the reflection principle, is not available for general random walks, hence a direct proof of this statement about random walks may not be so easy. This shows how Brownian motion can be useful even for proving things about random walks!

To see the usefulness in the opposite direction, let us observe that the reflection principle is in fact available for the simple symmetric random walk (steps ± 1 with equal probability), hence by some combinatorics (Feller's vol. 1, chapter 3 remains the best resource for this topic) one can actually show that $\frac{1}{\sqrt{n}} \max\{S_0, \ldots, S_n\} \xrightarrow{d} |Z|$. Now use Donsker's theorem to conclude that $M_1 \stackrel{d}{=} |Z|$. Furthermore, once you have it for Brownian motion, you have the result for random walk with any step distribution (with zero mean, unit variance). Thus, by proving it for one particular random walk, we can conclude it for all random walks, by passing through Brownian motion!

We shall use these ideas and revisit the three arcsine laws (we only proved two of them), and prove them for simple symmetric random walk (by combinatorics), for Brownian motion (by Donsker's theorem) and hence for general random walks. We shall also prove the Khinchine-Hartman-Wintner law of iterated logarithm and a functional form of it, by first doing it for Brownian motion and then deducing it for random walks.

1. Proof of Donsker's invariance principle

The method of proof is even simpler than that of CLT, since we shall essentially get convergence in distribution by coupling random variables so that there is convergence in probability! Here is the precise statement.

Lemma 5

Let μ_n, μ be probability measures on a metric space (X, d). Assume that μ is *tight*, in the sense that given $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq X$ such that $\mu(K_\varepsilon) > 1 - \varepsilon$.
Suppose we can construct X-valued random variables $Y_n \sim \mu_n$ and $Z_n \sim \mu$ on some probability space so that $d(Y_n, Z_n) \xrightarrow{P} 0$. Then, $\mu_n \xrightarrow{d} \mu$.

Proof. Let $f \in C_b(X)$. Then for any $\varepsilon > 0$, $\delta > 0$,

$$\begin{aligned}
&\Big|\int_X f d\mu_n - \int_X f d\mu\Big| \leqslant \mathbb{E}[|f(Y_n) - f(Z_n)|] \\
&\leqslant \mathbb{E}[|f(Y_n) - f(Z_n)|\mathbf{1}_{Z_n \notin K_\varepsilon}] + \mathbb{E}[|f(Y_n) - f(Z_n)|\mathbf{1}_{d(Y_n,Z_n) \geqslant \delta}] + \mathbb{E}[|f(Y_n) - f(Z_n)|\mathbf{1}_{Z_n \in K_\varepsilon}\mathbf{1}_{d(Y_n,Z_n)<\delta}] \\
&\leqslant 2\|f\|_{\sup}\varepsilon + 2\|f\|_{\sup}\mathbb{P}\{d(Y_n, Z_n) \geqslant \delta\} + \mathbb{E}[|f(Y_n) - f(Z_n)|\mathbf{1}_{Z_n \in K_\varepsilon}\mathbf{1}_{d(Y_n,Z_n)<\delta}].
\end{aligned}$$

We observe that if δ is small enough, then $|f(y) - f(z)| < \varepsilon$ whenever $z \in K_\varepsilon$ and $d(y, z) < \delta$. If not, there would be $y_n \in X$, $z_n \in K_\varepsilon$ such that $d(y_n, z_n) \to 0$ and $|f(y_n) - f(z_n)| \geqslant \varepsilon$. By compactness we may assume $z_n \to z \in K_\varepsilon$, then $y_n \to z$ too, and by continuity $f(y_n) - f(z_n) \to 0$. Thus, the third term above may be bounded by ε.

Now let $n \to \infty$ and then $\varepsilon \to 0$ to see that $\int f d\mu_n \to \int f d\mu$. As this holds for any $f \in C_b(X)$, $\mu_n \xrightarrow{d} \mu$. ■

The argument above may be slightly extended to extend its applicability considerably.

Corollary 2

In the setting of the above lemma, the the conclusion $\int f d\mu_n \to \int f d\mu$ holds for any f that is continuous on a σ-compact set S_f such that $\mu(S_f) = 1$.

Proof. Write $S = \cup_m L_m$ where $L_n$are compact and repeat the proof with K_ε replaced by $L_m \cap K_\varepsilon$. Then let $m \to \infty$ along with $n \to \infty$ and $\varepsilon \to 0$. ■

When μ is the Wiener measure, we know that it is tight (for example, take K_ε to be the set of all Hölder$(1/4)$ functions with sufficiently large Hölder constant). Further, for any set $A \in \mathcal{B}_{C[0,1]}$, there are compact sets $L_m \subseteq A$ such that $\mu(L_m) \uparrow \mu(A)$.

Proof of Donsker's theorem. Let W be standard Brownian motion on $[0,\infty)$. By Skorokhod embedding theorem, there are stopping times $0 = \tau_0 \leqslant \tau_1 \leqslant \tau_2 \leqslant \dots$ such that $(\tau_{i+1}-\tau_i, W(\tau_{i+1})-W(\tau_i))$, $i \geqslant 0$, are i.i.d., $W(\tau_{i+1})-W(\tau_i)$ are i.i.d. with the same distribution as X_1 and $\mathbb{E}[\tau_{i+1}-\tau_i] = 1$. In particular, $(W(\tau_0), W(\tau_1), \dots) \stackrel{d}{=} (S_0, S_1, \dots)$. Now define two $C[0,1]$-valued random variables (so $0 \leqslant t \leqslant 1$).

$$W_n(t) = \frac{1}{\sqrt{n}} W(nt), \qquad Y_n(t) = \begin{cases} \frac{W(\tau_k)}{\sqrt{n}} & \text{if } t = \frac{k}{n},\ 0 \leqslant k \leqslant n, \\ \text{linear in between.} \end{cases}$$

Then Y_n has the same distribution as the rescaled random walk and W_n is a standard Brownian motion. The key claim is that

$$\|W_n - Y_n\|_{\sup[0,1]} \xrightarrow{P} 0. \tag{15}$$

To show this, we observe that as Y_n is piecewise linear on each $J_{n,k} := [k/n, (k+1)/n]$,

$$\|W_n - Y_n\| = \left(\max_{0\leqslant k\leqslant n-1} \sup_{t\in J_{n,k}} |W_n(t) - Y_n(k/n)| \right) \vee \left(\max_{0\leqslant k\leqslant n-1} \sup_{t\in J_{n,k}} |W_n(t) - Y_n((k+1)/n)| \right).$$

For $\frac{k}{n} \leqslant t \leqslant \frac{k+1}{n}$, we have

$$W_n(t) - Y_n(k/n) = W_n(t) - W_n(\tau_k/n), \qquad W_n(t) - Y_n((k+1)/n) = W_n(t) - W_n(\tau_{k+1}/n)$$

hence if $\delta > \frac{1}{n}$, we have

$$\mathbb{P}\{\|W_n - Y_n\| > \varepsilon\} \leqslant \mathbb{P}\{\omega_{W_n}(2\delta) > \varepsilon\} + \mathbb{P}\left\{ \max_{0\leqslant k\leqslant n} |\frac{\tau_k}{n} - \frac{k}{n}| \geqslant \delta \right\}.$$

For fixed ε we can find δ so that the first probability is smaller than ε. As for the second, the probability goes to zero as $n \to \infty$. To see this, observe that if $(x_k)_k$ is a real sequence such that $\frac{x_n}{n} \to 1$, then $\max_{k\leqslant n} |x_k - k|/n \to 0$. Apply this to the sequence (τ_k) which satisfies $\frac{\tau_n}{n} \stackrel{a.s.}{\to} 1$ by SLLN. This completes the proof of (15). By Lemma 5, this proves Donsker's theorem. ∎

Implications of Donsker's theorem. Let $F : C[0,1] \to \mathbb{R}$ be a function that is continuous on a set of probability 1 under Wiener measure. Then $F(W_n) \xrightarrow{d} F(W)$. To see this, let $g \in C_b(\mathbb{R} \to \mathbb{R})$, and observe that $f = g \circ F$ satisfies the conditions of Corollary 2. Therefore, $\mathbb{E}[g(F(W_n))] \to \mathbb{E}[g(F(W))]$. As this is true for all $g \in C_b(\mathbb{R} \to \mathbb{R})$, we conclude that $F(W_n) \xrightarrow{d} F(W)$. In short, Donsker's theorem encodes within many univariate convergence in distribution statements.

2. Lévy's arcsine laws

For $f \in C[0,1]$, we define

(1) $T(f) = \arg\max f$, the smallest t such that $f(t) = \max_{0\leqslant s\leqslant 1} f(s)$. We say smallest, to remove ambiguity in the definition.

(2) $L(f) = \max\{t \in [0,1] : f(t) = 0\}$ where the maximum is 1 if the set is empty.

(3) $A(f) = \text{Leb.}\{t \in [0,1] : f(t) \geqslant 0\}$.

Theorem 9: Lévy's arcsine laws

If W is standard Browninan motion on $\mathbb{R}$ run for unit time, $T(W)$, $L(W)$ and $A(W)$ have arcsine distribution having density $\frac{1}{\pi\sqrt{x(1-x)}}$.

The analogous asymptotic statements for random walks are as follows.

Theorem 10

Let $X_1, X_2, \ldots$ be i.i.d. random variables with zero mean and unit variance. Let $S_0 = 0$ and $S_n = X_1 + \ldots + X_n$. Define

(1) $T_n = \min\{0 \leqslant k \leqslant n : S_k = \max_{0\leqslant j\leqslant n} S_j\}$,

(2) $L_n = \max\{0 \leqslant k \leqslant n : S_k S_{k+1} \leqslant 0\}$,

(3) $A_n = \max\{0 \leqslant k \leqslant n : S_k \geqslant 0\}$.

Then, $\frac{T_n}{n}$, $\frac{L_n}{n}$ and $\frac{A_n}{n}$ converge in distribution to the arcsine law.

One can make slight variations in the definitions of these random variables without changing the validity of the statement. For example, in the definition of L_n we can ask for strict inequality $S_k S_{k+1} < 0$ and similarly in A_n one can count strictly positive ones among $S_0, \ldots, S_n$. And in T_n one may take the last time S_k equals the global maximum.

First we deduce Theorem 10 from Theorem 9. This is a little less straightforward than what we discussed before (e.g., the maximum value), because none of T, W, A is a continuous function on $C_0[0,1]$. However, they are continuous a.e. on $C[0,1]$, with respect to Wiener measure.

Lemma 6

T, W, A are continuous a.e. with respect to Wiener measure.

Proof. To show this, define subsets of $C[0,1]$ that will be shown to have full Wiener measure and on which the corresponding functional will be shown to be continuous.

(1) $\mathcal{A}_1$: All f such that $L(f) < 1$ and for every $\varepsilon > 0$, there exist $s, t \in [L(f) - \varepsilon, L(f) + \varepsilon]$ such that $f(s) < 0 < f(t)$.

If f belongs to $\mathcal{A}_1$, fix $\varepsilon > 0$ and find $\delta > 0$ such that there exist $s, t \in [L(f) - \varepsilon, L(f) + \varepsilon]$ such that $f(s) < -\delta$ and $f(t) > \delta$ and such that $|f(u)| > \delta$ for all $u \geqslant L(f) + \varepsilon$. Then if $g \in C[0,1]$ and $\|g - f\| < \delta$, it is clear that g has no zeros in $[L(f) + \varepsilon, 1]$, while it does have a zero in $[L(f) - \varepsilon, L(f) + \varepsilon]$ (because $g(s) < 0 < g(t)$), hence $|L(g) - L(f)| \leqslant \varepsilon$.

Since $W(1) \neq 0$ and $W(0) = 0$ with probability 1, it follows that $L(W) < 1$ with probability 1. Since $L(f)$ is an accumulation point of the zero set of W (with probability 1), it follows that W has strict sign changes in $[L(W) - \varepsilon, L(W)]$ for any $\varepsilon > 0$, while it has no sign changes in $[L(W), 1]$. This implies that $W \in \mathcal{A}_1$ with probability 1.

(2) $\mathcal{A}_2$: All f for which $\{t : f(t) = \max_{0 \leqslant s \leqslant 1} f(s)\}$ is a singleton.

If $f \in \mathcal{A}_2$, then given $\varepsilon > 0$, there is a $\delta > 0$ such that the maximum of f outside $[T(f) - \delta, T(f) + \delta]$ is less than the global maximum by at least δ. Hence if $g \in C[0,1]$ and $\|g - f\| < \delta$, then $T(g) \in [T(f) - \delta, T(f) + \delta]$.

Fix any $T \in [0,1]$ and observe that $\max_{[0,T]} W - W(T)$ and $\max_{[T,1]} W - W(T)$ are independent random variables having the same distributions as $\sqrt{T}|Z_1|$ and $\sqrt{1-T}|Z_2|$, where Z_i are standard Gaussians. From this, it is clear that there is no chance that these two random variables are equal. But that is the same as saying that $\mathbb{P}\{\max_{[0,T]} W = \max_{[T,1]} W\} = 0$. As this is true for each fixed T, it is true for the union over all $T \in \mathbb{Q} \cap [0,1]$. But if W has two distinct global maxima, then $\max_{[0,T]} W = \max_{[T,1]} W$ for any rational T for some rational T (any T between the two global maxima). Hence, $\mathbb{P}\{W \in \mathcal{A}_2\} = 1$.

(3) $\mathcal{A}_3$: All f such that $\{t : f(t) = 0\}$ has zero Lebesgue measure.

If $f \in \mathcal{A}_3$, then given $\varepsilon > 0$, there exists $\delta > 0$ such that the Lebesgue measure of $\{f \geqslant \delta\}$ and $\{f \leqslant -\delta\}$ are within ε of the Lebesgue measures of $\{f \geqslant 0\}$ and $\{f \leqslant 0\}$ respectively. If $g \in C[0,1]$ and $\|f - g\| < \delta$, then $g > 0$ on $\{f \geqslant \delta\}$ and $g < 0$ on $\{f \leqslant -\delta\}$, from which it easily follows that $|\mathcal{A}(g) - \mathcal{A}(f)| \leqslant \varepsilon$.

We have already shown that the zero set of W has zero Lebesgue measure, hence $\mathbb{P}\{W \in \mathcal{A}_3\} = 1$.

All claims in the theorem are proved. ■

CHAPTER 7

Miscellaneous topics in Brownian motion

1. Laws of iterated logarithm

Let W be 1-dimensional SBM. There are several laws of iterated logarithm that are of interest.

▶ Law of iterated logarithm: $\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = 1$ a.s. Equivalently, the set of limit points of $\frac{W_t}{\sqrt{2t\log\log t}}$ as $t \to \infty$ is equal to $[-1, 1]$ a.s.

▶ Sharp form of the LIL: Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be increasing. Let E_ϕ be the event that the set $\{t \geqslant : W(t) \geqslant \sqrt{t}\phi(t)\}$ is unbounded (so there are arbitrarily large times when $W(t)$ is above $\sqrt{t}\phi(t)$). Let $g(x) = \int_0^x e^{u^2/2}du$ and say that ϕ belongs to the upper class $\mathcal{U}$ (respectively to the lower class $\mathcal{L}$) if $\int_*^\infty \frac{1}{g(\phi(t))}dt$ is finite (respectively infinite). Then,

$$\mathbb{P}\{E_\phi\} = 1 \iff \phi \in \mathcal{U}, \qquad \mathbb{P}\{E_\phi\} = 0 \iff \phi \in \mathcal{L}.$$

▶ Functional law of iterated logarithm: Let $W_T(t) = \frac{W(tT)}{\sqrt{2T\log\log T}}$ for $0 \leqslant t \leqslant 1$. Then the set of limit points of $W_T(\cdot)$ as $T \to \infty$ is equal to the unit ball in the Cameron Martin space, namely

$$\mathbb{B}_{CM} := \{f \in C[0,1] : f(t) = \int_0^1 g(s)ds \text{ for some } g \in L^2[0,1] \text{ with } \|g\|_{L^2} \leqslant 1\}.$$

▶ Sharp form of the functional LIL??

▶ LIL for random walks: Let $(S_k)_{k\geqslant 0}$ be a random walk with i.i.d. steps having zero mean and unit variance. Then $\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1$ a.s.

▶ Functional LIL for random walks: Let $(S_k)_{k\geqslant 0}$ be as above and let $\mathcal{S}_n(\cdot)$ be the random continuous function that is linear on $[\frac{k-1}{n}, \frac{k}{n}]$ for each $1 \leqslant k \leqslant n$ and with $\mathcal{S}_n(k/n) = \frac{S_k}{\sqrt{2n\log\log n}}$. Then the set of limit points of $\mathcal{S}_n$ is equal to $\mathbb{B}_{CM}$, the unit ball of the Cameron-Martin space, a.s.

2. Sets hit by Brownian motion

Let W be d-dimensional standard Brownian motion. For $A \subseteq \mathbb{R}^d$, let $\tau_A = \inf\{t : W(t) \in A\}$. For closed and open sets A, we know that τ_A is a stopping time.

▶ If $d = 1$, then the range of W is $\mathbb{R}$, a.s. In other words, $\tau_x < \infty$ for all $x \in \mathbb{R}$, w.p.1. We say that Brownian motion is *recurrent*.

▶ If $d = 2$, then $\tau_x = \infty$ a.s. for any fixed $x \neq 0$. Further, $\tau_A < \infty$ a.s. for any open set A. We say that planar Brownian motion is *neighbourhood recurrent*, but *not point recurrent*.

► If $d \geqslant 3$, then for any bounded open set $A \not\ni 0$, we have $0 < \mathbb{P}(\tau_A < \infty) < 1$. An equivalent statement is that $W[t, \infty) \cap A = \emptyset$ for large enough t, a.s. (i.e., eventually exits every bounded set). Brownian motion is *transient*.

► In $d \geqslant 2$, what are the compact sets A for which $\mathbb{P}_x(\tau_A < \infty) > 0$ for all $x \in \mathbb{R}^d$? This way of phrasing it allows the set to be anywhere (for e.g., if it is true for A, it is true for all translates of A). The answer is that it is true for A if and only if it has *positive capacity*. By definition, A has positive capacity if there is a probability measure μ on A such that

$$\mathcal{E}_{G_d}(\mu) \int_A \int_A G_d(x, y) d\mu(x) d\mu(y) < \infty$$

where $G_d(x, y)$ is the Green's function on $\mathbb{R}^d$, given by

$$G_d(x, y) = \begin{cases} c_2 \log \frac{1}{\|x-y\|} & \text{if } d = 2, \\ c_d \frac{1}{\|x-y\|^{d-2}} & \text{if } d \geqslant 3. \end{cases}$$

The infimum of $\mathcal{E}_{G_d}(\mu)$ over all probability measures μ on A is called the *Greens capacity* of A.

► One can go further and define the Martin kernel $K_d(x, y) = \frac{G_d(x,y)}{G_d(0,y)}$ and also define $\mathrm{Cap}_{K_d}(A)$ as the minimum energy of all probability measures on A, except that energy $\mathcal{E}_K(\mu)$ is now defined using K_d instead of G_d. Then, there are constants $c_d > 0$ and $C_d < \infty$ such that

$$c_d \mathrm{Cap}_{K_d}(A) \leqslant \mathbb{P}_x\{\tau_A\} < \infty \leqslant C_d \mathrm{Cap}_{K_d}(A)$$

3. Large deviations

► Let $W = (W_t)_{0 \leqslant t \leqslant 1}$. We know that $\varepsilon W \to 0$ a.s. as $\varepsilon \downarrow 0$. What is the chance that εW is close to a different function f? The probability is small, but how small? The answer is given by *Schilder's theorem*: The rough statement is that the probability is about $\exp\{-\frac{1}{\varepsilon^2}\|f\|^2_{CM}\}$. The precise formulation is that

$$-\frac{1}{\varepsilon^2} \log \mathbb{P}\{\varepsilon W \in B\} \to \inf_{f \in B} \|f\|^2_{CM}$$

for any Borel set $B \subseteq C[0, 1]$ such that ∂B has zero Wiener measure. Like in weak convergence, we can allow B to be an open set (respectively closed set), but the conclusion is that the lim sup of the left side (respectively lim inf) is at most (respectively at least) the right side. In all this, $\|f\|^2_{CM} = \infty$ if f is not in the Cameron-Martin space.

CHAPTER 8

Appendix: Miscellaneous background material

1. Gaussian random variables

Standard normal: A standard normal or Gaussian random variable is one with density $\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Its distribution function is $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ and its tail distribution function is denoted $\bar{\Phi}(x) := 1 - \Phi(x)$. If X_i are i.i.d. standard normals, then $X = (X_1, \ldots, X_n)$ is called a standard normal vector in $\mathbb{R}^n$. It has density $\prod_{i=1}^{n} \phi(x_i) = (2\pi)^{-n/2} \exp\{-|\mathbf{x}|^2/2\}$ and the distribution is denoted by γ_n, so that for every Borel set A in $\mathbb{R}^n$ we have $\gamma_n(A) = (2\pi)^{-n/2} \int_A \exp\{-|\mathbf{x}|^2/2\} d\mathbf{x}$.

Exercise 16

[Rotation invariance] If $P_{n\times n}$ is an orthogonal matrix, then $\gamma_n P^{-1} = \gamma_n$ or equivalently, $PX \stackrel{d}{=} X$. Conversely, if a random vector with independent co-ordinates has a distribution invariant under orthogonal transformations, then it has the same distribution as cX for some (non-random) scalar c.

Multivariate normal: If $Y_{m\times 1} = \mu_{m\times 1} + B_{m\times n} X_{n\times 1}$ where $X_1, \ldots, X_n$ are i.i.d. standard normal, then we say that $Y \sim N_m(\mu, \Sigma)$ with $\Sigma = BB^t$. Implicit in this notation is the fact that the distribution of Y depends only on Σ and not on the way in which Y is expressed as a linear combination of standard normals (this follows from Exercise 24). It is a simple exercise that $\mu_i = \mathbb{E}[X_i]$ and $\sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$. Henceforth, for simplicity, we take the mean to be zero everywhere.

Since matrices of the form BB^t are precisely positive semi-definite matrices (defined as those $\Sigma_{m\times m}$ for which $\mathbf{v}^t \Sigma \mathbf{v} \geqslant 0$ for all $\mathbf{v} \in \mathbb{R}^m$), it is clear that covariance matrices of normal random vectors are precisely p.s.d. matrices. Clearly, if $Y \sim N_m(\mu, \Sigma)$ and $Z_{p\times 1} = C_{p\times m} Y + \theta_{p\times 1}$, then $Z \sim N_p(\theta + C\mu, C\Sigma C^t)$. Thus, affine linear transformations of normal random vectors are again normal.

Exercise 17

The random vector Y has density if and only if Σ is non-singular, and in that case the density is

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}\mathbf{y}^t\Sigma^{-1}\mathbf{y}\right\}.$$

If Σ is singular, then X takes values in a lower dimensional subspace in $\mathbb{R}^n$ and hence does not have density.

In particular, if $\mathbf{v} \in \mathbb{R}^m$, then $\mathbf{v}^tY$ is univariate normal with mean $\mathbf{v}^t\mu$ and variance $\mathbf{v}^t\Sigma\mathbf{v}$. The covariance of two different linear combinations $\mathbf{v}^tY$ and $\mathbf{u}^tY$ is $\mathbf{v}^t\Sigma\mathbf{u}$. The converse is also true. If $\mathbf{v}^tY$ is univariate Gaussian for every $\mathbf{v} \in \mathbb{R}^m$, then it is necessarily the case that Y is multivariate Gaussian. You may prove this using characteristic functions, for example. The characteristic function of Gaussian distribution is given in the exercise below.

Exercise 18

Irrespective of whether Σ is non-singular or not, the characteristic function of Y is given by

$$\mathbb{E}\left[e^{i\langle\lambda,Y\rangle}\right] = e^{-\frac{1}{2}\lambda^t\Sigma\lambda}, \text{ for } \lambda \in \mathbb{R}^m.$$

In particular, if $X \sim N(0,\sigma^2)$, then its characteristic function is $\mathbb{E}[e^{i\lambda X}] = e^{-\frac{1}{2}\sigma^2\lambda^2}$ for $\lambda \in \mathbb{R}$.

Exercise 19

If $U_{k\times 1}$ and $V_{(m-k)\times 1}$ are such that $Y^t = (U^t, V^t)$, and we write $\mu = (\mu_1, \mu_2)$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ are partitioned accordingly, then

(1) $U \sim N_k(\mu_1, \Sigma_{11})$.

(2) $U\Big|_V \sim N_k(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1/2}V,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ (assume that Σ_{22} is invertible).

2. More about the univariate normal distribution

Tail of the standard Gaussian distribution: Recall the standard Gaussian density $\phi(x)$. The corresponding cumulative distribution function is denoted by Φ and the tail is denoted by $\bar{\Phi}(x) := \int_x^\infty \phi(t)dt$. The following estimates will be used very often.

Exercise 20

For all $x > 0$, we have

$$\frac{1}{\sqrt{2\pi}}\frac{x}{1+x^2}e^{-\frac{1}{2}x^2} \leqslant \bar{\Phi}(x) \leqslant \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{1}{2}x^2}. \tag{16}$$

In particular[a], $\bar{\Phi}(x) \sim x^{-1}\phi(x)$ as $x \to \infty$. Most often the following simpler bound, valid for $x \geqslant 1$, suffices.

$$\frac{1}{10x}e^{-\frac{1}{2}x^2} \leqslant \bar{\Phi}(x) \leqslant e^{-\frac{1}{2}x^2}. \tag{17}$$

[a]The notation $f(x) \sim g(x)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Maximum of independent standard Gaussians: Let $X_1, \ldots, X_n$ be (not necessarily independent) random variables with each having $N(0,1)$ distribution. Let $M_n = \max\{X_1, \ldots, X_n\}$. How big is M_n? In general, the maximum of correlated Gaussians is a very important question of great current interest. The i.i.d. case is a very special and easy case where we can extract the right answer easily.

Observe that $M_n \geqslant t$ if and only if $X_i \geqslant t$ for some $i \leqslant n$. Therefore,

$$\mathbb{P}\{M_n \geqslant t\} \leqslant \sum_{k=1}^{n} \mathbb{P}\{X_k \geqslant t\} = n\bar{\Phi}(t).$$

Using the upper bound in (17), and setting $t = \sqrt{2A\log n}$ with $A > 1$, we get (since $t \geqslant 1$ for $n \geqslant 2$),

$$\mathbb{P}\{M_n \geqslant \sqrt{2A\log n}\} \leqslant ne^{-A\log n} = \frac{1}{n^{A-1}}. \tag{18}$$

We shall use this quantitative bound many times in the lectures. In particular, for every $\delta > 0$, the above inequality implies that $\mathbb{P}\left\{\frac{1}{\sqrt{2\log n}}M_n \geqslant 1+\delta\right\} \to 0$ as $n \to \infty$. This bound is actually tight if the random variables are independent.

Exercise 21: U

e the lower bound for the tail of the Normal distribution from (17), show that $\mathbb{P}\left\{\frac{1}{\sqrt{2\log n}}M_n \leqslant 1-\delta\right\} \to 0$ for any $\delta > 0$. Conclude that in this case $\frac{1}{\sqrt{2\log n}}M_n \xrightarrow{P} 1$.

Convergence and Gaussians: Distributional limits of Gaussians are Gaussians. In other words, if $\mu_n \to \mu$ and $\sigma_n^2 \to \sigma^2$, then $N(\mu_n, \sigma_n^2) \xrightarrow{d} N(\mu, \sigma^2)$. Conversely, if $N(\mu_n, \sigma_n^2) \xrightarrow{d} \nu$ for some probability measure ν, then $\nu = N(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma^2 \geqslant 0$. If this is not clear, take it as an exercise!

Gaussian density and heat equation: For $t > 0$, let $p_t(x) := \frac{1}{\sqrt{t}}\phi(x/\sqrt{t})$ be the $N(0,t)$ density. We interpret $p_0(x)dx$ as the degenerate measure at 0. These densities have the following interesting properties.

Exercise 22: S

ow that $p_t \star p_s = p_{t+s}$, i.e., $\int_{\mathbb{R}} p_t(x-y)p_s(y)dy = p_{t+s}(x)$.

Exercise 23: S

ow that $p_t(x)$ satisfies the heat equation: $\frac{\partial}{\partial t}p_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p_t(x)$ for all $t > 0$ and $x \in \mathbb{R}$.

Remark 4: P

t together, these facts say that $p_t(x)$ is the *fundamental solution* to the heat equation. This just means that the heat equation $\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(t,x)$ with the initial condition $u(0,x) = f(x)$ can be solved simply as $u(t,x) = (f\star p_t)(x) := \int_{\mathbb{R}} f(y)p_t(x-y)dy$. This works for reasonable f (say $f \in L^1(\mathbb{R})$).

3. Existence of countably many Gaussians with given covariances

Let $\Sigma = (\sigma_{i,j})_{i,j\geqslant 1}$ be a semi-infinite matrix. Do there exist random variables $X_1, X_2, \ldots$ that are jointly Gaussian (by which we mean that any finite sub-collection of them has joint Gaussian distribution) and such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_iX_j] = \sigma_{i,j}$ for all $i, j \geqslant 1$?

A necessary condition is that Σ is (symmetric and) positive semi-definite. This means that $\sigma_{i,j} = \sigma_{j,i}$ for all i, j and $\sum_{i,j=1}^n u_iu_j\sigma_{i,j} \geqslant 0$ for all $n \geqslant 1$ and all $\mathbf{u} \in \mathbb{R}^n$. Symmetry is clearly necessary. As for the second condition, observe that

$$\sum_{i,j=1}^n u_iu_j\sigma_{i,j} = \mathbb{E}\left[\left(\sum_{i=1}^n u_iX_i\right)^2\right]$$

by expanding the square and interchanging expectation with the sum. From this, the p.s.d property is clear. Note that we did not require Gaussian property here - covariance matrix of any collection of random variables is p.s.d.

Claim 1

Let $\Sigma = (\sigma_{i,j})_{i,j\geqslant 1}$ be a symmetric p.s.d. matrix. Then, there exist random variables (on some probability space) X_i, $i \geqslant 1$, that are jointly Gaussian, have zero means and covariance matrix Σ.

PROOF. Let ξ_n, $n \geqslant 1$, be i.i.d. $N(0,1)$ random variables (on your favourite probability space, for example, $([0,1], \mathcal{B}, \lambda)$). We shall define $X_n = a_{n,1}\xi_1 + \ldots + a_{n,n}\xi_n$, where the coefficients $a_{n,j}$, $1 \leqslant j \leqslant n$, will be chosen so as to satisfy the covariance conditions. That X_n, $n \geqslant 1$, have a joint Gaussian distribution is clear.

First, we define $a_{1,1} = \sqrt{\sigma_{1,1}}$ so that $X_1 \sim N(0,1)$. This definition is valid since p.s.d. property implies that $\sigma_{1,1} \geqslant 0$.

Next, from $\mathbb{E}[X_1X_2] = \sigma_{1,2}$ we get the equation $a_{1,2}\sqrt{\sigma_{1,1}} = \sigma_{1,2}$ and $a_{2,2}^2 + a_{2,1}^2 = \sigma_{2,2}$. As the 2×2 matrix $(\sigma_{i,j})_{i,j\leqslant 2}$ is p.s.d., we certainly have $\sigma_{1,1} \geqslant 0$ and $\sigma_{2,2}\sigma_{1,1} - \sigma_{1,2}^2 \geqslant 0$. If $\sigma_{1,1} > 0$, then the unique solutions are

$$a_{2,1} = \frac{\sigma_{1,2}}{\sqrt{\sigma_{1,1}}}, \quad a_{2,2} = \sqrt{\sigma_{2,2} - \frac{\sigma_{1,2}^2}{\sigma_{1,1}}}.$$

What if $\sigma_{1,1} = 0$. Then use p.s.d property to show that $\sigma_{1,i} = 0$ for all i (in general, if a diagonal entry vanishes, the entire row and column containing it must also vanish). But then the first equation is vacuous and we may set $a_{1,2} = 0$ (or anything else, it does not matter since X_1 is the zero random variable!) and $a_{2,2} = \sqrt{\sigma_{2,2}}$.

Now suppose we have solved for $a_{k,j}$, $1 \leqslant j \leqslant k \leqslant n-1$. We want to solve for $a_{n,j}$, $j \leqslant n$. Let us use matrix notation and write $B = (a_{k,j})_{j,k\leqslant n-1}$ (with $a_{k,j} = 0$ if $j > k$). Let $\mathbf{u}^t = (a_{n,1}, \ldots, a_{n,n-1})$ and let $\mathbf{v}^t = (\sigma_{n,1}, \ldots, \sigma_{n,n-1})$. Then, the equations that we must solve are $B\mathbf{u} = \mathbf{v}$ and $a_{n,n}^2 + \|u\|^2 = \sigma_{n,n}$. If $a_{k,k} > 0$ for $k \leqslant n-1$, then B is non-singular and we get the unique solutions $\mathbf{u} = B^{-1}\mathbf{v}$ and $a_{n,n} = \sqrt{\sigma_{n,n} - \|\mathbf{u}\|^2}$. The last square root makes sense because of the matrix theory fact that

$$\det \begin{bmatrix} X & \mathbf{v} \\ \mathbf{v}^t & c \end{bmatrix} = \det(X).(c - \mathbf{v}^t X^{-1}\mathbf{v})$$

whenever X is a non-singular matrix. Here we apply it with $X = (\sigma_{i,j})_{i,j\leqslant n-1}$, $\mathbf{v}$ as before and $c = \sigma_{n,n}$. Positive definiteness implies that both determinants are positive. Hence $c - \mathbf{v}^t X^{-1}\mathbf{v} > 0$ (in our case this is precisely $\sigma_{n,n} - \|u\|^2$.

All this is fine if Σ is strictly positive definite, for then $\det(\sigma_{i,j})_{i,j\leqslant n} > 0$ for every n. Hence, inductively, we see that $a_{n,n} > 0$ for all n and the above procedure continues without any difficulty. If $a_{n,n} = 0$ for some n, then we need to modify the procedure.

[Will write this, too tired now...] ■

4. Gaussian random variables

Standard normal: A standard normal or Gaussian random variable is one with density $\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Its distribution function is $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ and its tail distribution function is denoted $\bar{\Phi}(x) := 1 - \Phi(x)$. If X_i are i.i.d. standard normals, then $X = (X_1, \ldots, X_n)$ is called a standard normal vector in $\mathbb{R}^n$. It has density $\prod_{i=1}^{n} \phi(x_i) = (2\pi)^{-n/2} \exp\{-|\mathbf{x}|^2/2\}$ and the distribution is denoted by γ_n, so that for every Borel set A in $\mathbb{R}^n$ we have $\gamma_n(A) = (2\pi)^{-n/2} \int_A \exp\{-|\mathbf{x}|^2/2\} d\mathbf{x}$.

Exercise 24

[Rotation invariance] If $P_{n\times n}$ is an orthogonal matrix, then $\gamma_n P^{-1} = \gamma_n$ or equivalently, $PX \stackrel{d}{=} X$. Conversely, if a random vector with independent co-ordinates has a distribution invariant under orthogonal transformations, then it has the same distribution as cX for some (non-random) scalar c.

Multivariate normal: If $Y_{m\times 1} = \mu_{m\times 1} + B_{m\times n} X_{n\times 1}$ where $X_1, \ldots, X_n$ are i.i.d. standard normal, then we say that $Y \sim N_m(\mu, \Sigma)$ with $\Sigma = BB^t$. Implicit in this notation is the fact that the distribution of Y depends only on Σ and not on the way in which Y is expressed as a linear combination of standard normals (this follows from Exercise 24). It is a simple exercise that $\mu_i = \mathbb{E}[X_i]$ and $\sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$. Since matrices of the form BB^t are precisely positive semi-definite matrices (defined as those $\Sigma_{m\times m}$ for which $\mathbf{v}^t \Sigma \mathbf{v} \geqslant 0$ for all $\mathbf{v} \in \mathbb{R}^m$), it is clear that covariance matrices of normal random vectors are precisely p.s.d. matrices. Clearly, if $Y \sim N_m(\mu, \Sigma)$ and $Z_{p\times 1} = C_{p\times m} Y + \theta_{p\times 1}$, then $Z \sim N_p(\theta + C\mu, C\Sigma C^t)$. Thus, affine linear transformations of normal random vectors are again normal.

Exercise 25

The random vector Y has density if and only if Σ is non-singular, and in that case the density is

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}\mathbf{y}^t \Sigma^{-1} \mathbf{y}\right\}.$$

If Σ is singular, then X takes values in a lower dimensional subspace in $\mathbb{R}^n$ and hence does not have density.

Exercise 26

Irrespective of whether Σ is non-singular or not, the characteristic function of Y is given by

$$\mathbb{E}\left[e^{i\langle\lambda,Y\rangle}\right] = e^{-\frac{1}{2}\lambda^t\Sigma\lambda}, \text{ for } \lambda \in \mathbb{R}^m.$$

In particular, if $X \sim N(0,\sigma^2)$, then its characteristic function is $\mathbb{E}[e^{i\lambda X}] = e^{-\frac{1}{2}\sigma^2\lambda^2}$ for $\lambda \in \mathbb{R}$.

Exercise 27: I

$U_{k\times 1}$ and $V_{(m-k)\times 1}$ are such that $Y^t = (U^t, V^t)$, and we write $\mu = (\mu_1, \mu_2)$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ are partitioned accordingly, then

(1) $U \sim N_k(\mu_1, \Sigma_{11})$.

(2) $U\Big|_V \sim N_k(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1/2}V,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ (assume that Σ_{22} is invertible).

Moments: All questions about a centered Gaussian random vector must be answerable in terms of the covariance matrix. In some cases, there are explicit answers.

Exercise 28

Prove the *Wick formula* (also called *Feynman diagram formula*) for moments of centered Gaussians.

(1) Let $X \sim N_n(0,\Sigma)$. Then, $\mathbb{E}[X_1 \dots X_n] = \sum_{M\in\mathcal{M}_n} \prod_{\{i,j\}\in M} \sigma_{i,j}$, where $\mathcal{M}_n$ is the collection of all matchings of the set $[n]$ (thus $\mathcal{M}_n$ is empty if n is odd) and the product is over all matched pairs. For example, $\mathbb{E}[X_1X_2X_3X_4] = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23}$.

(2) If $\xi \sim N(0,1)$, then $\mathbb{E}[\xi^{2n}] = (2n-1)(2n-3)\dots(3)(1)$.

Cumulants: Let X be a real-valued random variable with $\mathbb{E}[e^{tX}] < \infty$ for t in a neighbourhood of 0. Then, we can write the power series expansions

$$\mathbb{E}[e^{i\lambda X}] = \sum_{k=0}^{\infty} m_n(X)\frac{\lambda^n}{n!}, \qquad \log \mathbb{E}[e^{i\lambda X}] = \sum_{k=1}^{\infty} \kappa_n[X]\frac{\lambda^n}{n!}.$$

Here $m_n[X] = \mathbb{E}[X^n]$ are the moments while $\kappa_n[X]$ is a linear combination of the first n moments ($\kappa_1 = m_1$, $\kappa_2 = m_2 - m_1^2$, etc). Then κ_n is called the nth cumulant of X. If X and Y are independent, then it is clear that $\kappa_n[X+Y] = \kappa_n[X] + \kappa_n[Y]$.

Exercise 29: (optional)

Prove the following relationship between moments and cumulants. The sums below are over partitions Π of the set $[n]$ and $\Pi_1, \ldots, \Pi_{\ell_\Pi}$ denote the blocks of Π.

$$m_n[X] = \sum_{\Pi} \prod_i \kappa_{|\Pi_i|}[X], \qquad \kappa_n[X] = \sum_{\Pi} (-1)^{\ell_\Pi - 1} \prod_i m_{|\Pi_i|}[X].$$

Thus $\kappa_1 = m_1$, $\kappa_2 = m_2 - m_1^2$,

Exercise 30

If $\xi \sim N(0,1)$, then $\kappa_1 = 0$, $\kappa_2 = 1$ and $\kappa_n = 0$ for all $n \geqslant 3$.

The converse of this result is also true and often useful in proving that a random variable is normal. For instance, the theorem below implies that to show that a sequence of random variables converges to normal, it suffices to show that cumulants $\kappa_m[X_n] \to 0$ for all $m \geqslant m_0$ for some m_0.

Result 1: Marcinkiewicz

If X is a random variable with finite moments of all orders and $\kappa_n[X] = 0$ for all $n \geqslant n_0$ for some n_0, then X is Gaussian.

Convergence and Gaussians:

Exercise 31

The family of distributions $N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $0 \leqslant \sigma^2 < \infty$, is closed under convergence in distribution (for this statement to be valid we include $N(\mu, 0)$ which means δ_μ). Indeed, $N(\mu_n, \sigma_n^2) \xrightarrow{d} N(\mu, \sigma^2)$ if and only if $\mu_n \to \mu$ and $\sigma_n^2 \to \sigma^2$.

A vector space of Gaussian random variables: Let $Y \sim N_m(0, \Sigma)$ be a random vector in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for every vector $\mathbf{v} \in \mathbb{R}^m$, define the random variable $Y_\mathbf{v} := \mathbf{v}^t Y$. Then, for any $\mathbf{v}_1, \ldots, \mathbf{v}_j$, the random variables $Y_{\mathbf{v}_1}, \ldots, Y_{\mathbf{v}_j}$ are jointly normal. The joint distribution of $\{Y_\mathbf{v}\}$ is fully specified by noting that $Y_\mathbf{v}$ have zero mean and $\mathbb{E}[Y_\mathbf{v} Y_\mathbf{u}] = \mathbf{v}^t \Sigma \mathbf{u}$.

We may interpret this as follows. If Σ is p.d. (p.s.d. and non-singular), then $(\mathbf{v}, \mathbf{u})_\Sigma := \mathbf{v}^t \Sigma \mathbf{u}$ defines an inner product on $\mathbb{R}^m$. On the other hand, the set $L_0^2(\Omega, \mathcal{F}, \mathbb{P})$ of real-valued random variables on Ω with zero mean and finite variance, is also an inner product space under the inner product $\langle U, V \rangle := \mathbb{E}[UV]$. The observation in the previous paragraph is that $\mathbf{v} \to Y_\mathbf{v}$ is an isomorphism of $(\mathbb{R}^m, (\cdot,\cdot)_\Sigma)$ into $L_0^2(\Omega, \mathcal{F}, \mathbb{P})$.

In other words, given any finite dimensional inner-product space $(V, \langle \cdot, \cdot \rangle)$, we can find a collection of Gaussian random variables on some probability space, such that this collection is isomorphic to the given inner-product space. Later we shall see the same for Hilbert spaces[1].

5. The Gaussian density

Recall the standard Gaussian density $\phi(x)$. The corresponding cumulative distribution function is denoted by Φ and the tail is denoted by $\bar{\Phi}(x) := \int_x^\infty \phi(t)dt$. The following estimate will be used very often.

Exercise 32

For all $x > 0$, we have $\frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} e^{-\frac{1}{2}x^2} \leqslant \bar{\Phi}(x) \leqslant \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{1}{2}x^2}$ In particular[a], $\bar{\Phi}(x) \sim x^{-1}\phi(x)$ as $x \to \infty$. Most often the following simpler bound, valid for $x \geqslant 1$, suffices.

$$\frac{1}{10x} e^{-\frac{1}{2}x^2} \leqslant \bar{\Phi}(x) \leqslant e^{-\frac{1}{2}x^2}.$$

[a]The notation $f(x) \sim g(x)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

For $t > 0$, let $p_t(x) := \frac{1}{\sqrt{t}}\phi(x/\sqrt{t})$ be the $N(0,t)$ density. We interpret $p_0(x)dx$ as the degenerate measure at 0. These densities have the following interesting properties.

Exercise 33

Show that $p_t \star p_s = p_{t+s}$, i.e., $\int_{\mathbb{R}} p_t(x-y)p_s(y)dy = p_{t+s}(x)$.

Exercise 34

Show that $p_t(x)$ satisfies the heat equation: $\frac{\partial}{\partial t}p_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p_t(x)$ for all $t > 0$ and $x \in \mathbb{R}$.

Remark 5

Put together, these facts say that $p_t(x)$ is the *fundamental solution* to the heat equation. This just means that the heat equation $\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(t,x)$ with the initial condition $u(0,x) = f(x)$ can be solved simply as $u(t,x) = (f \star p_t)(x) := \int_{\mathbb{R}} f(y)p_t(x-y)dy$. This works for reasonable f (say $f \in L^1(\mathbb{R})$).

We shall have many occasions to use the following "integration by parts" formula.

[1]This may seem fairly pointless, but here is one thought-provoking question. Given a vector space of Gaussian random variables, we can multiply any two of them and thus get a larger vector space spanned by the given normal random variables and all pair-wise products of them. What does this new vector space correspond to in terms of the original $(V, \langle \cdot, \cdot \rangle)$?

Exercise 35

Let $X \sim N_n(0, \Sigma)$ and let $F : \mathbb{R}^n \to \mathbb{R}$. Under suitable conditions on F (state sufficient conditions), show that $\mathbb{E}[X_i F(X)] = \sum_{j=1}^n \sigma_{ij} \mathbb{E}[\partial_j F(X)]$. As a corollary, deduce the Wick formula of Exercise 28.

Stein's equation: Here we may revert to $t = 1$, thus $p_1 = \phi$. Then, $\phi'(x) = -x\phi(x)$. Hence, for any $f \in C_b^1(\mathbb{R})$, we integrate by parts to get $\int f'(x)\phi(x)dx = -\int f(x)\phi'(x)dx = \int f(x)x\phi(x)dx$. If $X \sim N(0,1)$, then we may write this as

$$\mathbb{E}[(Tf)(X)] = 0 \quad \text{for all } f \in C_b^1(\mathbb{R}), \text{ where } (Tf)(x) = f'(x) - xf(x). \tag{19}$$

The converse is also true. Suppose (19) holds for all $f \in C_b^1(\mathbb{R})$. Apply it to $f(x) = e^{i\lambda x}$ for any fixed $\lambda \in \mathbb{R}$ to get $\mathbb{E}[Xe^{i\lambda X}] = i\lambda \mathbb{E}[e^{i\lambda X}]$. Thus, if $\psi(\lambda) := \mathbb{E}[e^{i\lambda X}]$ is the characteristic function of X, then $\psi'(\lambda) = -\lambda\psi(\lambda)$ which has only one solution, $e^{-\lambda^2/2}$. Hence X must have standard normal distribution.

Digression - central limit theorem: One reason for the importance of normal distribution is of course the central limit theorem. The basic central limit theorem is for $W_n := (X_1 + \ldots + X_n)/\sqrt{n}$ where X_i are i.i.d. with zero mean and unit variance. Here is a sketch of how central limit theorem can be proved using Stein's method. Let $f \in C_b^1(\mathbb{R})$ and observe that $\mathbb{E}[W_n f(W_n)] = \sqrt{n}\mathbb{E}[X_1 f(W_n)]$. Next, write

$$f\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right) \approx f\left(\frac{X_2 + \ldots + X_n}{\sqrt{n}}\right) + \frac{X_1}{\sqrt{n}} f'\left(\frac{X_2 + \ldots + X_n}{\sqrt{n}}\right)$$

where we do not make precise the meaning of the approximation. Let $\hat{W}_n = \frac{X_2 + \ldots + X_n}{\sqrt{n}}$. Then,

$$\mathbb{E}[W_n f(W_n)] \approx \sqrt{n}\mathbb{E}[X_1]\mathbb{E}[f(\hat{W}_n)] + \mathbb{E}[X_1^2]\mathbb{E}[f'(\hat{W}_n)] = \mathbb{E}[f'(\hat{W}_n)].$$

Since $\hat{W}_n \approx W_n$, this shows that $\mathbb{E}[Tf(W_n)] \approx 0$. We conclude that $W_n \approx N(0,1)$.

There are missing pieces here, most important being the last statement - that if a random variable satisfies Stein's equation approximately, then it must be approximately normal. When included, one does get a proof of the standard CLT.

CHAPTER 9

Stochastic integral and Ito's formula

1. Quadratic variation

If $0 = t_0 < t_1 < \ldots < t_n = t$, we say that $\Pi = (t_0, \ldots, t_n)$ is a partition of $[0, t]$ with size $\|Pi\| := \max(t_i - t_{i-1})$. For a function $f : [a, b] \to \mathbb{R}$, the variation is defined as

$$V_f([a, b]) := \sup_{\Pi=(a=t_0,\ldots,t_n=b)} \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|.$$

If $V_f([a, b]) < \infty$, we say that f is of *bounded variation*. A Lipschitz function is easily seen to be of bounded variation. On the other hand, if f is more rough, say a Hölder$(1/2)$ function, then in general we do not expect it to have bounded variation. What we can say is that $\sup_\Pi \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^2$ is finite.

Brownian motion is almost Hölder$(1/2)$, but not quite. But since it is a random function (i.e., we are dealing with an ensemble of functions, not one), it turns out that quadratic variation is the appropriate one. For $\Pi = (0 = t_0, \ldots, t_n = t)$ and a SBM W, let $Q(\Pi) = \sum_{k=1}^{n} |W(t_k) - W(t_{k-1})|^2$.

Theorem 11

Let W be standard Brownian motion. Fix $t > 0$. Let Π_n be a sequence of partitions of $[0, t]$.

(1) If $\|\Pi_n\| \to 0$, then $Q(\Pi_n) \xrightarrow{L^2,\, P} t$.

(2) If $\|\Pi_n\| \to 0$ and Π_{n+1} is a refinement of Π_n for each n, then $Q(\Pi_n) \xrightarrow{a.s.} t$.

Proof. If $\Pi = (0 = t_0, t_1, \ldots, t_n = t)$, then by the independence of increments,

$$\begin{aligned}
\mathbb{E}[Q(\Pi)] &= \sum_{k=1}^{n-1} t_k - t_{k-1} \;=\; t, \\
\mathrm{Var}(Q(\Pi)) &= \sum_{k=1}^{n-1} (t_k - t_{k-1})^2 \mathrm{Var}(Z^2) \quad \text{where } Z \sim N(0, 1), \\
&\leqslant \mathrm{Var}(Z^2) \|\Pi\| \sum_{k=1}^{n} (t_k - t_{k-1}) \\
&= 2\|\Pi\| t.
\end{aligned}$$

Hence if $\|Pi_n\| \to 0$, then $\mathbb{E}[(Q(\Pi_n) - t)^2] \to 0$ and hence the first part is proved.

For the second part, the key observation is that if Π_n are got by refining the previous one, then $Q(\Pi_n)$ is a reverse martingale. Accepting that, the reverse martingale theorem tells us that $Q(\Pi_n)$ converges a.s. But the almost sure limit must be the same as the limit in probability, which is t. To check the claim about $Q(\Pi_n)$ being a reverse martingale, it suffices to observe that for $s < t < u$,

$$\mathbb{E}[(W(s) - W(u))^2 \,\big|\, (W(s) - W(t))^2 + (W(t) - W(u))^2] = (W(s) - W(t))^2 + (W(t) - W(u))^2.$$

We leave this as an exercise. ■

Here is a more general result (the case $X = 1$ recovers the first part of previous theorem) that will be of use later.

Lemma 7

Let W be a $\mathcal{F}_\bullet$-Brownian motion. Let X be a continuous stochastic process adapted to $\mathcal{F}_\bullet$ such that $\sup_{t \leqslant T} \mathbb{E}[X_t^2] < \infty$. If $\Pi_n = (t_{n,0}, \ldots, t_{n,n})$ is a sequence of partitions of $[0, T]$ such that $\|\Pi_n\| \to 0$, then

$$\sum_{k=0}^{n-1} X(t_k)(W(t_{k+1}) - W(t_k))^2 \xrightarrow{L^2,\, P} \int_0^T X(t)\,dt.$$

Proof. Write $\Delta_{n,j} = (W(t_{n,j+1}) - W(t_{n,j}))^2$ and $\delta_{n,j} = t_{n,j+1} - t_{n,j}$ and let $Q_n := \sum_{k=0}^{n-1} X(t_{n,k})(\Delta_{n,k} - \delta_{n,k})$. Observe that

(1) $\mathbb{E}[\Delta_{n,k} - \delta_{n,k} \,\big|\, \mathcal{F}_{t_{n,k}}] = 0$,

(2) $\mathbb{E}[(\Delta_{n,k} - \delta_{n,k})^2 \,\big|\, \mathcal{F}_{t_{n,k}}] = 2(t_{n,k+1} - t_{n,k})^2$,

(3) If $j < k$, then $\mathbb{E}[(\Delta_{n,k} - \delta_{n,k})(\Delta_{n,j} - \delta_{n,j}) \,\big|\, \mathcal{F}_{t_{n,k}}] = (\Delta_{n,j} - \delta_{n,j})\mathbb{E}[\Delta_{n,k} - \delta_{n,k} \,\big|\, \mathcal{F}_{t_{n,k}}] = 0$.

Using these and the adaptedness of X, we see that $\mathbb{E}[Q_n] = 0$ and

$$\begin{aligned}\mathbb{E}[Q_n^2] &= 2\sum_{k=0}^{n-1} \mathbb{E}[X(t_{n,k})^2](t_{n,k+1} - t_{n,k})^2 \leqslant 2\|\Pi_n\| \sum_{k=0}^{n-1} \mathbb{E}[X(t_{n,k})^2](t_{n,k+1} - t_{n,k}) \\ &\leqslant 2\|\Pi_n\| \sup_{t \leqslant T} \mathbb{E}[X_t^2]T\end{aligned}$$

which goes to 0 as $n \to \infty$. Hence $Q_n \to 0$ in $L^2(\mathbb{P})$ and in probability. But observe that $Q_n = Q_n' - Q_n''$ where Q_n' is the expression in the statement of the lemma and $Q_n'' = \sum_{k=0}^{n-1} X(t_{n,k})\delta_{n,k}$. By the usual convergence of Riemann sums to integral, Q_n'' converges almost surely to $\int_0^T X(t)\,dt$. Hence $Q_n' \to \int_0^T X(t)\,dt$ in $L^2(\mathbb{P})$ and in probability. ■

2. Ito's stochastic integral

Let $(\Omega, \mathcal{F}, \mathcal{F}_\bullet, \mathcal{P})$ be a filtered probability space (assume that $\mathcal{F}_\bullet$ satisfies the usual conditions) and let W be a SBM in one dimension. Let $\mathcal{L}^2_{\text{loc}}$ denote the set of all progressively measurable stochastic processes $X = (X_t)_{t \geqslant 0}$ such that $\int_0^t \mathbb{E}[X_s^2] ds < \infty$ for all $t < \infty$.

The main result is that for every $X \in \mathcal{L}^2_{\text{loc}}$ is (uniquely)associated a process $I(X) = (I_t(X))_{t \geqslant 0}$ (usually denotes as $\int_0^t X(s) ds$ or as $(X.W)_t$), caled the Ito integral of X w.r.t. W, having the following properties:

(1) $I_t(X)$ is a continuous, adapted process.

(2) $I_t(X)$ and $I_t(X)^2 - \langle I(X) \rangle_t$ are martingales, where $\langle I(X) \rangle_t := \int_0^t X_s^2 ds$.

(3) $\mathbb{E}[I_t(X)] = 0$ and $\mathbb{E}[I_t(X)^2] = \int_0^t \mathbb{E}[X_s^2] ds$.

(4) $I_t(aX + bY) = aI_t(X) + bI_t(Y)$ for $X, Y \in \mathcal{L}^2_{\text{loc}}$ and $a, b \in \mathbb{R}$.

(5) If $X = \xi \mathbf{1}_{(a,b]}$ where $\xi \in \mathcal{F}_a$ and $\mathbb{E}[\xi^2] < \infty$, we have $I_t(X) = \xi(W(t \wedge b) - W(t \wedge a))$.

Construction of the Ito integral and showing its uniqueness is carried out in most books on the subject. Let us just give a brief sketch to make it look plausible. More importantly, the point is that the above properties suffice to work with the Ito integral, we do not need the details of its construction.

Sketch of the construction: From the last two requirements, it is clear that for a process of the form $X = \xi_0 \mathbf{1}_{(t_0,t_1]} + \ldots + \xi_n \mathbf{1}_{(t_{n-1},t_n]}$ with $0 = t_0 < \ldots < t_n < \infty$ and $\xi_k \in \mathcal{F}_{t_k}$ with $\mathbb{E}\xi_k^2 < \infty$ (such processes are called *elementary*), we see that there is no choice but to define its Ito integral as

$$I_t(X) = \sum_{k=0}^{n-1} \xi_k (W(t \wedge t_{k+1}) - -W(t \wedge t_k)).$$

It is also important to note that this definition is consistent in that even though X can be represented in the above form with a different set of t_ks and ξ_ks, that does not change $I_t(X)$. A key point is to check that for elementary processes, all five properties hold.

Once we have the isometry property (third in the list) for elementary processes, it is easy to see that if X is any process that can be approximated by elementary processes $X^{(n)}$ in the sense that

$$\int_0^t \mathbb{E}[|X^{(n)}(t) - X(t)|^2] dt \to 0 \text{ for all } t, \tag{20}$$

then $(I_t(X^{(n)}))_n$ is Cauchy in $L^2(\mathbb{P})$, and hence has a limit that can be defined as $I_t(X)$. For such processes X, the last four properties carry over easily. But the first one, continuity of sample paths of $I_t(X)$, needs justification. In fact, if $I_t(X)$ is defined separately for each t, we cannot expect to have continuity. To achieve that, we just need to take care and use not only the isometry property but

the martingale property of the Ito integral for elementary processes. Indeed, as $I_t(X^{(n)}) - I_t(X^{(m)})$ is a continuous martingale, by Doob's maximal inequality we see that

$$\mathbb{E}\left[\sup_{s\leqslant t}(I_s(X^{(n)}) - I_s(X^{(m)}))^2\right] \leqslant C\mathbb{E}[(I_t(X^{(n)}) - I_t(X^{(m)}))^2].$$

Now choose $n_1 < n_2 < \dots$ so that with $n = n_k$ and $m = n_{k+1}$, the quantity on the right is at most $1/4^k$, which implies that the expectation without the square is at most $1/2^k$ (Cauchy-Schwarz). Thus,

$$\sum_k \sup_{s\leqslant t} |I_s(X^{(n)}) - I_s(X^{(m)})| < \infty \text{a.s.}$$

as the expectation of the quantity on the left is finite. By the Weierstrass-M test, this shows that $I_s(X^{(n_k)}) = \sum_{j=0}^{k-1} I_s(X^{(n_{j+1})}) - I_s(X^{(n_j)})$ converges uniformly over $s \in [0,t]$, a.s. Hence $(I_s(X))_{s\leqslant t}$ is defined as continuous process. Take intersection over $t \in \mathbb{N}$ to see that $I_t(X)$ may be defined as a process with continuous sample paths.

The last remaining point is to identify the processes that can be approximated by simple processes in the sense of (20) with the collection $\mathcal{L}^2_{\text{loc}}$.

Lemma 8

Let X be a progressively measurable process with $\int_0^t \mathbb{E}[X_s^2]ds < \infty$ for all t. Then there exist elementary processes $X^{(n)}$ such that $\int_0^t \mathbb{E}[|X_s - X_s^{(n)}|^2]ds \to 0$ for all t.

Again we sketch it here, referring the reader to the books for details

Proof. It suffices to show the conclusion for fixed t, because then for each n we can find $X^{(n)}$ such that $\int_0^t \mathbb{E}[|X_s - X_s^{(n)}|^2]ds \leqslant \frac{1}{2^n}$. The sequence $X^{(n)}$ approximates X on all bounded intervals. Hence we fix t henceforth.

(1) Suppose X has continuous and uniformly bounded sample paths. Any continuous function $f \in C[0,t]$ can be uniformly approximated by step functions $f^{(n)}$ that also satisfy $|f^{(n)}| \leqslant |f|$. Applying this pointwise to $X[0,t]$, and using DCT, we get elementary processes $X^{(n)}$ such that $\int_0^t \mathbb{E}\left[|X_s - X_s^{(n)}|^2\right] ds \to 0$.

(2) Suppose X is progressively measurable and has uniformly bounded sample paths. Define $X^{(n)}(t) = \int_{t-n\frac{1}{n}}^t X(s)ds$. Then $X^{(n)}$ has continuous and uniformly bounded sample paths. By the Lebesgue differentiation theorem (applied to $s \mapsto X_s(\omega)$ for each ω), we conclude that $X_s^{(n)}(\omega) - X_s(\omega) \to 0$ for a.e. $(s,\omega) \in [0,t] \times \Omega$ w.r.t. Leb $\otimes \mathbb{P}$. By DCT, we see that $\int_0^t \mathbb{E}[|X_s - X_s^{(n)}|^2]ds \to 0$. Since $X^{(n)}$ are approximable by elementary processes (first step), it follows that the X is also approximable by elementary processes.

(3) If X is only progressively measurable, we define $X_s^{(n)} = X_s \mathbf{1}_{|X_s| \leqslant n}$. Clearly $X_s^{(n)}(\omega) - X_s(\omega) \to X_s(\omega)$ for for a.e. $(s, \omega) \in [0, t] \times \Omega$ w.r.t. Leb $\otimes \mathbb{P}$. Further, $|X_s^{(n)}(\omega) - X_s(\omega)|^2$ is dominated by $4|X_s(\omega)|^2$ which is integrable w.r.t. Leb $\otimes \mathbb{P}$. By DCT, it follows that $\int_0^t \mathbb{E}[|X_s - X_s^{(n)}|^2] ds \to 0$. As $X^{(n)}$ is progressively measurable and has uniformly bounded paths, by the previous step we may approximate is using elementary processes. The conclusion carries over to X. ■

2.1. An important extension of the Ito integral. Let X be a progressively measurable process such that $\int_0^t X_s^2 ds < \infty$ for all t, a.s. We show that it is still possible to define $I_t(X)$ as a continuous process, except that the martingale properties are weakened (there is no guarantee that expectations even exist).

Let $X_s^{(n)} = X_s \mathbf{1}_{|X_s| \leqslant n}$. Then $X^{(n)}$ is a progressively measurable process with uniformly bounded sample paths. Hence $I_t(X^{(n)})$ is well-defined and has all the properties discussed above. We claim that $I_t(X^{(n)})$ converges in probability. Then the limit random variable will be defined as $I_t(X)$.

To prove the claim, we show that the sequence $I_t(X^{(n)}) - I_t(X^{(m)}) \xrightarrow{P} 0$ as $m, n \to \infty$. The difference from the above is that we do not have L^2 convergence, and hence we need an "in-probability-estimate"[1].

Lemma 9

Let Y be a progressively measurable process with $\int_0^t \mathbb{E}[Y_s^2] ds < \infty$ for all t. Then, for any positive ε, δ,

$$\mathbb{P}\{I_t(Y) \geqslant \varepsilon\} \leqslant \frac{\delta}{\varepsilon^2} + \mathbb{P}\left\{\int_0^t Y_s^2 ds \geqslant \delta\right\}.$$

Proof. ■

3. Ito's formula

We rarely compute integrals from first principles as limits of Riemann sums. Most explicit computations are done using the fundamental theorem of Calculus. A similar role is played by Ito's formula for stochastic integrals. Of course, it is important in other ways (as is the fundamental theorem of Calculus), but this suffices as motivation.

Let $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$, usually written as $f(t, x)$ where $t \in \mathbb{R}_+$ is interpreted as time and $x \in \mathbb{R}^d$ is interpreted as space. We say that $f \in C^{1,2}$ if the first partial derivative ∂_0 w.r.t. t and the second partial derivative ∂_k^2 w.r.t. x_k for $1 \leqslant k \leqslant d$, all exist and are continuous.

[1] We have taken this from the book of Kallianpur and Sundar, which gives a really down-to-earth presentation of the topic.

Theorem 12: Ito's formula

Let $f : C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$. Let W be d-dimensional Brownian motion. Then

$$f(W_t) - f(W_0) = \int_0^t \nabla f(W_s).dW_s + \int_0^t \left[\partial_0 f(W_s) + \frac{1}{2}\Delta f(W_s)\right] ds.$$

Here $\nabla f = (\partial_1 f, \ldots, \partial_d f)$ and $\Delta f = \sum_{k=1}^d \partial_k^2 f$ denote the gradient and Laplacian of f w.r.t. the space variables. The first term on the right is just a short form for

$$\sum_{k=1}^d \int_0^t \partial_k f(W_s) dW_k(s).$$

Note that these are all Ito integrals w.r.t. one-dimensional Brownian motions (no higher dimensional generalization is needed). Often Ito's formula is written succinctly as

$$df(W_t) = \nabla f(W_t).dW_t + \left[\partial_0 f(W_t) + \frac{1}{2}\Delta f(W_t)\right] dt,$$

but this is just notation and has no independent meaning other than the integral form of the equation.

PROOF. Let $\Pi_n = (t_0, \ldots, t_n)$ (we should write $t_{n,k}$ strictly speaking) be a sequence of partitions of $[0, t]$ such that $\|\Pi_n\| \to 0$. Then we write the difference between the left and right sides of the Ito's formula as a sum of $J([t_0, t_1]) + \ldots + J([t_{n-1}, t_n])$ where

$$J([a, b]) = f(W_b) - f(W_a) - \int_a^b [\partial_0 f(W_s) + \frac{1}{2}\Delta f(W_s)] ds - \int_a^b \nabla f(W_s).dW_s.$$

By Taylor's expansion,

$$f(W(b)) - f(W(a)) = \nabla f(W(a)).(b - a) + \sum_{j,k=1}^d \partial_{j,k} f(W(a))(W_k(b) - W_k(a)) + \ldots$$

■

4. List of topics to cover

(1) Weak convergence theory on complete, separable metric spaces. Particular case of C and D spaces. (4 hours).

(2) Recap of discrete time martingales with some applications. (2+2 hours).

(3) Brownian motion: Definition and first construction (3 hours). Continuity properties (2 hours). Symmetries (1 hour). Markov and Strong Markov property (3 hours). Running maximum and First passage times (2 hours). Weiner integral (1 hour). Recurrence and Transience (1 hour). Arcsine laws (2 hours). Donsker's theorem (2 hours). Associated process (Brownian bridge etc.) (1 hour).

(4) Martingales in continuous time: Definition, examples, filtrations (2 hours). Regularization (2 hours). Doob-Meyer decomposition for continuous square integrable martingales (skip proof). Stopping times and Local martingales (2 hours).

CHAPTER 10

Probability measures on metric spaces

In basic probability we largely study real-valued random variables or at most $\mathbb{R}^d$-valued random variables. From the point of view of applications of probability, it is clear that there are more complex random objects. For example, consider the graph of the daily value of the rupee versus the dollar over a calendar year. For each year we get a different graph, and in some ways, the ups and downs appear to be random. While one can consider it as a vector of length 365, it may be more meaningful to think of it as defined at each time point. Hence we need the notion of a random function. There are situations where one may also want the notion of a discontinuous random function or random functions on the plane (eg., random surfaces), or random measures (eg., the length measure of the zero set of a random function from $\mathbb{R}^2$ to $\mathbb{R}$) or the set of locations of an epidemic, etc.

Probabilists have found that all applications of interest so far can be captured by allowing random variables to take values in a general complete and separable metric space. The distribution of such a random variables is a probability measure on the metric space. A key part of the theory is the notion of weak convergence of measures on such spaces. In this section, we summarize (mostly without proofs), the basic facts[1].

Let (X, d) be a complete and separable metric space. Let $\mathcal{B}_X$ denote the Borel sigma-algebra of X and let $\mathcal{P}(X)$ denote the set of all probability measures on $(X, \mathcal{B}_X)$. For $\mu, \nu \in \mathcal{P}(X)$, define

$$d(\mu, \nu) = \inf\{r > 0 : \mu(A_r) + r \geqslant \nu(A) \text{ and } \nu(A_r) + r \geqslant \mu(A) \text{ for all } A \in \mathcal{B}_X\}$$

where $A_r = \bigcup_{x \in A} B(x, r)$ is the r-neighbourhood of A (it is an open set, hence measurable).

Lemma 10: Prohorov metric

d defines a metric on $\mathcal{P}(X)$.

Observe that $x \mapsto \delta_x$ is an isometry from X to $\mathcal{P}(X)$, hence using the same letter d for the metric can be excused. If $d(\mu_n, \mu) \to 0$ for $\mu_n, \mu \in \mathcal{P}(X)$, we say that μ_n converges in distribution to μ and write $\mu_n \xrightarrow{d} \mu$.

[1] Billingsley's book

em Convergence of probability measures or K. R. Parthasarathy's *Probability measures on metric spaces* are excellent sources to know more. Of course, Kallenberg's book has everything succinctly.

Lemma 11: Portmanteau theorem

For $\mu_n, \mu \in \mathcal{P}(X)$, the following are equivalent.

(1) $\mu_n \xrightarrow{d} \mu$.

(2) $\int f d\mu_n \to \int f d\mu$ for all $f \in C_b(X)$.

(3) $\liminf_{n\to\infty} \mu_n(G) \geqslant \mu(G)$ for all open $G \subseteq X$.

(4) $\limsup_{n\to\infty} \mu_n(F) \leqslant \mu(F)$ for all closed $F \subseteq X$.

(5) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all $A \in \mathcal{B}_X$ satisfying $\mu(\partial A) = 0$.

Except for the use of distribution functions (which is not available on general metric spaces), the similarity to the situation in $\mathbb{R}$ is readily seen. The Prohorov metric also agrees with the Lévy-Prohorov distance that we had defined, except that the class of sets over which the infimum is taken was only right-closed intervals (in general metric spaces, many books take infimum only over closed sets).

Following the usual definition in metric spaces, a subset $\mathcal{A} \subseteq \mathcal{P}(X)$ is said to be relatively compact (or precompact) if every subsequence has a convergent subsequence. This is the same as saying that $\bar{\mathcal{A}}$ is compact in $(\mathcal{P}(X), d)$. The fundamental theorem is a characterization of relatively compact sets (analogous to Helly's theorem for probability measures on $\mathbb{R}$).

Definition 3: Tightness

We say that $\mathcal{A} \subseteq \mathcal{P}(X)$ is *tight* if, for any $\varepsilon > 0$, there is a compact $K_\varepsilon \subseteq X$ such that $\mu(K_\varepsilon) \geqslant 1 - \varepsilon$ for all $\mu \in \mathcal{A}$.

Theorem 13: Prokhorov's theorem

A subset $\mathcal{A} \subseteq \mathcal{P}(X)$ is relatively compact if and only if it is tight.

Corollary 3

If (X, d) is compact, then $(\mathcal{P}(X), d)$ is also compact. In general for any complete, separable (X, d), the metric space $(\mathcal{P}(X), d)$ is also complete and separable.

That completes all we want to know in general. When it comes to a specific metric space, a key thing is to be able to check tightness of a subset of measures, which involves understanding compact subsets on the metric space itself. We work out a couple of examples below and write out the conditions for checking tightness. But before that let us indicate another exceedingly useful approach to showing convergence in distribution that avoids having to know all this machinery.

Lemma 12

Let μ_n, μ belong to $\mathcal{P}(X)$. Suppose X_n, X are X-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ such that $\mathbb{P} \circ X_n^{-1} = \mu_n$, $\mathbb{P} \circ X^{-1} = \mu$ and $X_n \to X$ a.s.$[\mathbb{P}]$. Then, $\mu_n \xrightarrow{d} \mu$

Skorokhod showed the converse, that whenever $\mu_n \xrightarrow{d} \mu$, there is a probability space and random variables X_n, X having these distributions such that $X_n \xrightarrow{d} X$. However, the useful part is the above direction, although the proof is trivial!

Proof. Let $f \in C_b(X)$. Then $f(X_n) \xrightarrow{a.s.} f(X)$ and these are bounded real-valued random variables. Hence by the dominated convergence theorem $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ as $n \to \infty$. But $\mathbb{E}[f(X_n)] = \int f d\mu_n$ and $\mathbb{E}[f(X)] = \int f d\mu$, hence $\mu_n \xrightarrow{d} \mu$. ■

Observe that almost sure convergence also makes it trivial to say that for any continuous function $\phi : X \mapsto \mathbb{R}$, we have $\phi(X_n) \to \phi(X)$ almost surely and hence also in distribution. Thus, various "features" of μ_n also converge in distribution to the corresponding feature of μ (i.e., $\mu_n \circ \phi^{-1} \xrightarrow{d} \mu \circ \phi^{-1}$, as probability measures on $\mathbb{R}$).

Example 7

Let $X = \mathbb{R}^{\mathbb{N}}$. This is a complete and separable metric space with the metric $d(x, y) = \sum_n 2^{-n}(1 \wedge |x_n - y_n|)$ for $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$.

Example 8

Let $X = C[0, 1]$ with the sup-norm metric. Arzela-Ascoli theorem tell us that $K \subseteq C[0, 1]$ is compact if and only if it is closed and there is an $M < \infty$ such that $|f(0)| \leqslant M$ for all $f \in K$ and for each $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(y)| \leqslant \varepsilon$ for any $x, y \in [0, 1]$ with $|x - y| \leqslant \delta$ and for any $f \in K$. The last condition of *equicontinuity* is the crucial one.

1. Notes to be expanded

▶ Assume that (X, d) is a complete and separable metric space. Let $\mathcal{G}_X, \mathcal{F}_X, \mathcal{K}_X$ denote the collections of open, closed and compact subsets of X, respectively.

▶ The most important examples for this course are $C[0,1]$, $C[0,\infty)$ and $C([0,\infty) \to \mathbb{R}^d)$. What are the metrics?

▶ The Borel sigma-algebra $\mathcal{B}_X := \sigma(\mathcal{G}_X)$ is also equal $\sigma(C(X))$.

Proof: If $f \in C(X)$, then $\{f < t\}$ is open for any $t \in \mathbb{R}$, hence f is $\mathcal{B}_X$-measurable. Therefore $\sigma(C(X)) \subseteq \mathcal{B}_X$. For the converse, let $A \in \mathcal{F}_X$. Then $g := d(\cdot, A)$ is a continuous function on X and $\{g = 0\} = A$. Therefore, if g is measurable, so is A. Thus $\sigma(C(X)) \supseteq \mathcal{B}_X$.

▶ Borel isomorphism theorem: If X is complete and separable, then it is isomorphic to a Borel subset of $[0,1]$. This is a fundamental fact that we shall not use explicitly in what follows.

▶ Let $\mathcal{P}(X)$ (more usual notation is $\mathcal{M}(X)$) denote the collection of all Borel probability measures on X.

▶ If $\mu \in \mathcal{P}(X)$, then μ is regular, i.e., given $A \in \mathcal{B}_X$ and $\varepsilon > 0$, there exist $C \in \mathcal{F}$ and $U \in \mathcal{G}$ such that $C \subseteq A \subseteq U$ and $\mu(U \setminus C) < \varepsilon$.

Proof: Consider the collection $\mathcal{H}$ be the collection of all $A \in \mathcal{B}_X$ for which, given any $\varepsilon > 0$, there exist C and U as in the definition of regularity. The proof will be executed by showing that $\mathcal{H}$ is a sigma-algebra that contains $\mathcal{F}$.

$\mathcal{H}$ is a sigma-algebra: If $C \subseteq A \subseteq U$ and $C \in \mathcal{F}$ and $U \in \mathcal{G}$, then $U^c \subseteq A^c \subseteq C^c$ and $U^c \in \mathcal{F}$ and $C^c \in \mathcal{G}$. Further, $\mu(C^c \setminus U^c) = \mu(U \setminus C)$. Hence $A \in \mathcal{H}$ implies $A^c \in \mathcal{H}$. Next suppose that $A_n \in \mathcal{H}$, $n \geqslant 1$, and let $A = \cup_n A_n$. Find $C_n \in \mathcal{F}$ and $U_n \in \mathcal{G}$ such that $C_n \subseteq A_n \subseteq U_n$ and $\mu(U_n \setminus C_n) \leqslant \frac{1}{2^n}$. Let $C = \cup_n C_n$ and $U = \cup_n U_n$. Then $C \subseteq A \subseteq U$ and $\mu(U \setminus C) \leqslant \sum_n \mu(U_n \setminus C_n) \leqslant \varepsilon$. Clearly U is open, but C need not be closed. However, $C_N = C_1 \cup \ldots \cup C_N$ is closed for every N, then $\mu(C_N) \uparrow \mu(C)$, so for a sufficiently large N we get $C_N \subseteq A \subseteq U$ and $\mu(U \setminus C_N) \leqslant 2\varepsilon$.

$\mathcal{H}$ contains closed sets: Let $A \in \mathcal{F}$. Then we can take $C = A$. For $n \geqslant 1$, define $U_n = \{d(\bullet, A) < \frac{1}{n}\}$. Then U_n are open sets that decrease to A (this uses closedness of A) and hence $\mu(U_n) \downarrow A$. Thus, for large enough N, we have $C \subseteq A \subseteq U_N$ and $\mu(U_N \setminus A) < \varepsilon$. Thus $A \in \mathcal{H}$.

▶ If and $\mu \in \mathcal{P}(X)$, then μ is tight. That is, given $\varepsilon > 0$, there is a $K \in \mathcal{K}$ such that $\mu(K) \geqslant 1 - \varepsilon$. This requires X to be complete and separable.

Proof: Let $\mu \in \mathcal{P}(X)$ and $\varepsilon > 0$. For any $n \geqslant 1$, there exists a countable collection $x_{n,1}, x_{n,2}, \ldots$ such that $\cup_j B(x_{n,j}, \frac{1}{n}) = X$. Therefore, there exist $J_n < \infty$ such that if $C_n = \cup_{j \leqslant J_n} B(x_{n,j}, \frac{1}{n})$, then $\mu(C_n^c) \leqslant \varepsilon/2^{-n}$. Let $K = \cap_n C_n$ so that $\mu(K^c) \leqslant \sum_n \mu(C_n^c) \leqslant \varepsilon$. Each C_n is closed, hence so is K. To show that it is compact, take a sequence $z_m \in K$. By a diagonal argument, we can get $j_k \leqslant J_k$

for each k and a subsequence z_{m_k} so that $z_{m_i} \in B(x_{k,j_k}, \frac{1}{k})$ for all $i \geqslant k$. This shows that $(z_{m_k})_k$ is Cauchy, and hence converges. Thus K is compact.

- For $\mu, \nu \in \mathcal{P}(X)$ define the Lévy-Prohorov metric

$$D(\mu,\nu) = \inf\{r > 0 : \mu(A_r) + r \geqslant \nu(A),\ \nu(A^r) + r \geqslant \mu(A) \text{ for all } A \in \mathcal{F}\}$$

The symmetry and triangle inequality are straightforward (note that $A_{r+s} \supseteq (\overline{A_r})_s$). Suppose $D(\mu,\nu) = 0$. Then fixing A and letting $r \downarrow 0$, we see that $\mu(A) = \nu(A)$ for closed A, and hence also for open A (taking complements). By regularity, it follows that $\mu = \nu$. Hence D is a metric.

- For $\mu, \nu \in \mathcal{P}(X)$ define

$$D'(\mu,\nu) = \sup\{\Big|\int f d\mu - \int f d\nu\Big| : f \in \mathrm{Lip}(1),\ \|f\|_{\sup} \leqslant 1\}.$$

It is easy to see that D' is symmetric and satisfies triangle inequality. We show that $D'(\mu,\nu) > 0$ if $\mu \neq \nu$, by proving that $2D'(\mu,\nu) \geqslant D(\mu,\nu)$. To this end, for any closed set A and $0 < r < 1$, let $\phi_{A,r} = d(\bullet, A) \wedge r$. This is is a $\mathrm{Lip}(1)$ function and $\|\phi_{A,r}\|_{\sup} \leqslant r$. Hence $|\int \phi d\mu - \int \phi d\nu| \leqslant D'(\mu,\nu)$. As $\mathbf{1}_A \leqslant \phi_{A,r} \leqslant \mathbf{1}_{A_r}$, it follows that $\mu(A) \leqslant \nu(A_r) + D'(\mu,\nu)$ and $\nu(A) \leqslant \mu(A_r) + D'(\mu,\nu)$. Taking $r = D'(\mu,\nu)$, we see that $D(\mu,\nu) \leqslant 2D'(\mu,\nu)$.

In fact D' is equivalent to D. To see this, suppose $f \in \mathrm{Lip}(1)$ and $\|f\|_{\sup} \leqslant 1$. Assume that $f \geqslant 0$. Then $\int f d\mu = \int_0^1 \mu\{f \geqslant t\} dt$. For any $t > 0$, note that $A = \{f \geqslant t\}$ is closed and $\{f \geqslant t + r\} \subseteq A_r \subseteq \{f \geqslant t - r\}$ because $f \in \mathrm{Lip}(1)$. Consequently, if we set $r = D(\mu,\nu)$, then $\mu\{f \geqslant t - r\} + r \geqslant \nu\{f \geqslant t\}$. Integrating over t, we see that

$$\int f d\mu \geqslant \int_r^1 \mu\{f \geqslant t - r\} dt \ \geqslant\ \int_r^1 (\nu\{f \geqslant t\} - r) dt \geqslant \int f d\nu - 2r.$$

The same holds with μ and ν reversed. For a general f, we split into positive and negative parts and apply the above. We therefore get $D'(\mu,\nu) \leqslant 4D(\mu,\nu)$.

- We say that $\mu_n \xrightarrow{d} \mu$ if $D(\mu_n, \mu) \to 0$ which is the same as $D'(\mu_n, \mu) \to 0$.
- Portmanteau theorem: The following are equivalent. (A) $\mu_n \xrightarrow{d} \mu$. (B) $\int f d\mu_n \to \int f d\mu$ for all $f \in C_b(X)$. (C) $\limsup \mu_n(G) \geqslant \mu(G)$ for all open G. (D) $\liminf \mu_n(F) \leqslant \mu(F)$ for all closed F. (E) $\lim \mu_n(A) = \mu(A)$ for all $A \in \mathcal{B}_X$ for which $\mu(\partial A) = 0$.

Proof: (A) $\implies$ (B): Let $f \in C_b(X)$. Find K compact such that $\mu(K^c) \leqslant \varepsilon$. Then $\int_{K^c} f d\mu \leqslant \varepsilon \|f\|_{\sup}$.

- Prohorov's theorem: Let $\mathcal{A} \subseteq \mathcal{P}(X)$. Then $\mathcal{A}$ is precompact if and only if $\mathcal{A}$ is tight. Recall that we say that $\mathcal{A}$ is tight if for any $\varepsilon > 0$, there is a compact set K_ε such that $\mu(K_\varepsilon) \geqslant 1 - \varepsilon$ for all $\mu \in \mathcal{A}$.
- If X is C+S then $\mathcal{P}(X)$ is also C+S. If X is compact, so is $\mathcal{P}(X)$.