PROBABILITY THEORY - PART 4
BROWNIAN MOTION

MANJUNATH KRISHNAPUR

CONTENTS

1. DEFINITION OF BROWNIAN MOTION AND WIENER MEASURE

Definition 1: Brownian motion

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

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

(2) For any $s < t$ the distribution of $W_t - W_s$ is $N(0, t-s)$. Also, $W_0 = 0$, $a.s.$

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

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.

Einstein and the physical Brownian motion: In 1820s, the botanist Brown observed under 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 (which at the time of Einstein, was not yet universally accepted). 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, 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 particle very slightly (since it is much larger), but the number of collisions in a second (say), is very high. 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 W_t denote the displacement of the x-coordinate of the particle, we have the following conclusions.

(1) The displacements in two disjoint intervals of time are independent. This is the first condition in the definition of Brownian motion.

(2) The displacement in a given interval (provided it is long enough that the number of collisions with water molecules is large) must have Normal distribution. This is a consequence of the central limit theorem.

(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 variance that depends only on the length of the time interval, say σ_t^2 for an interval of length t.

From the first and third conclusion, $\sigma_{t+s}^2 = \sigma_t^2 + \sigma_s^2$, which means that $\sigma_t^2 = D \cdot t$ 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 (or water) 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 $\mathbf{E}[W_1^2]$. Since $D = \mathbf{E}[W_1^2]$, this gives 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.

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 B. There are infinitely many pairwise disjoint balls of radius 1 inside $\sqrt{2}B$ (for example, take unit balls centered around each co-ordinate vector e_i, $i \geq 1$). Thus, if $\mu(B) > 0$, then by translation invariance, all these balls have the same measure and hence $\mu(\sqrt{2}B)$ 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 $\otimes_{t \in [0,1]}\mu$ on $\mathbb{R}^{[0,1]}$ (the space of all functions from $[0, 1]$ to $\mathbb{R}$) endowed with the product sigma-algebra. But this is very poor as a measure space as we have discussed in probability class. For example, the space $C[0, 1]$ is not a measurable subset of $\mathbb{R}^{[0,1]}$, since sets in the product sigma-algebra are determined by countably many co-ordinates.

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).

[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.

And then, he showed that it is possible to put a 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*.

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.

If X is a metric space, let $C_d(X)$ be the space of continuous functions from X to $\mathbb{R}^d$. If $d=1$, we just write $C(X)$. Of particular interest to us are $C[0,\infty)$, $C[0,1]$. When discussing d-dimensional Brownian motion, we shall need $C_d[0,\infty)$ and $C_d[0,1]$.

On $C[0,1]$, define the norm $\|f\|_{\sup} = \max\{|f(t)| : t \in [0,1]\}$ and the metric $d(f,g) = \|f-g\|_{\sup}$. It is a fact that $C[0,1]$ is complete under this metric and hence, it is a Banach space. 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

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

The metric is irrelevant, what matters is the topology and the fact that the topology is metrizable. In fact, many other metrics such as $\tilde{d}(f,g) = \sum_{n=1}^{\infty} \frac{1}{n^2} \min\{1, \|f-g\|_{\sup[0,n]}\}$ 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 $\mathbb{R}_+ = [0,\infty)$. For $t \in [0,\infty)$, define the projection map $\Pi_t : C[0,\infty) \to \mathbb{R}$ by $\Pi_t(f) = f(t)$. The topology on $C[0,\infty)$ can also be described as the smallest topology in which all the projections are continuous (exercise!).

Once the topology is defined, we have the Borel σ-algebra $\mathcal{B}(C[0,\infty))$ which is, by definition, the smallest sigma-algebra containing all open sets. Alternately, we may say that the Borel σ-algebra is generated by the collection of projection maps. Sets of the form $(\Pi_{t_1},\ldots,\Pi_{t_n})^{-1}(B)$ for $n \ge 1$ and $t_1 < \ldots < t_n$ and $B \in \mathcal{B}(\mathbb{R}^n)$, are called (finite dimensional) cylinder sets. Cylinder sets form a π-system that generate the Borel sigma-algebra. Thus, by the $\pi-\lambda$ theorem, any two Borel probability measures that agree on cylinder sets agree on the entire Borel σ-algebra $\mathcal{B}(C[0,\infty))$. All these considerations apply if we restrict our attention to $C[0,1]$.

Definition 2: Wiener measure

is the Borel probability measure μ on $C[0,\infty)$ such that for any $n \ge 1$ and any $t_1 < \ldots < t_n$, the measure $\mu \circ (\Pi_{t_1},\ldots,\Pi_{t_n})^{-1}$ (a Borel probability measure on $\mathbb{R}^n$) is the multivariate Gaussian distribution with zero means and covariance matrix equal to $(t_i \wedge t_j)_{1 \le i,j \le n}$.

It is not yet proved that Wiener measure exists. But if it exists, it must be unique, since any two such measures agree on all cylinder sets. In fact, Wiener measure and Brownian motion are two sides of the same coin, just as closely related as a Gaussian random variable and the Gaussian measure. In other words, Wiener measure is the distribution of Brownian motion, as the following exercise shows.

Exercise 1

(1) Suppose μ is the Wiener measure. Then, the collection of random variables $(\Pi_t)_{t\in\mathbb{R}_+}$ defined on the probability space $(C[0,\infty), \mathcal{B}(C[0,\infty)), \mu)$ is a Brownian motion.

(2) Suppose W is a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, then define the map $T : \Omega \to C[0,\infty)$ by

$$T(\omega) = \begin{cases} W_{\bullet}(\omega) & \text{if } t \mapsto W_t(\omega) \text{ is continuous,} \\ 0 & \text{otherwise.} \end{cases}$$

Then the push-forward measure $\mu := \mathbf{P} \circ T^{-1}$ is the Wiener measure.

Remark 1

At first one might think it more natural to consider the space of all functions, $\mathbb{R}^{[0,1]}$, endowed with the cylinder sigma-algebra (the one generated by the projections $\Pi_t(f) = f(t)$). But the only events that are measurable in this sigma-algebra are those that are functions of countably many co-ordinates. In particular, sets such as $C[0,1]$ are not measurable subsets. In all of probability, when we talk of stochastic processes, it is usually on a space of functions with some continuity properties. Although $C[0,\infty)$ is restrictive for some purposes (eg., point processes, or events that happen in a time instant), in this course this will suffice for us. More generally one works with the space of right continuous functions having left limits (RCLL).

However, some books start by considering a measure on this space with the finite dimensional distributions of Brownian motion (such a measure exists by Kolmogorov consistency) and then show that that the outer measure of $C[0,1]$ is 1. From there, it becomes possible to get the measure to sit on $C[0,1]$ to get Brownian motion. I feel that this involves unnecessary technical digressions than the proof we give in the next section.

3. 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 $\mathbf{E}[W_t] = 0$ and $\mathbf{E}[W_t W_s] = t \wedge s$. Here is the sketch of how it is done by the so called chaining method of Kolmogorov and Centsov.

(1) Let $D \subseteq [0,1]$ be a countable dense set. Because of countability, we know how to construct W_t, $t \in D$, on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, having a joint Gaussian distribution with zero means and covariance $t \wedge s$.

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

(3) By standard real analysis, this means that for each such ω, the function $t \mapsto W_t(\omega)$ extends to a continuous function on $[0,1]$.

(4) Since limits of Gaussians are Gaussian, the resulting W_t, $t \in [0,1]$, have joint Gaussian distribution with the prescribed covariances.

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 $F(t) = \lim\limits_{D \ni s \to t} f(s)$ 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 $w_f(\delta) = \sup\{|f(t) - f(s)| : |t-s| \leq \delta\}$. 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 $w_f(\delta) \leq C_f \delta$ and if f is Hölder(α) for some $0 < \alpha \leq 1$, then $w_f(\delta) \leq C_f \delta^\alpha$. 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 $D_n = \{k2^{-n} : 0 \leq k \leq 2^n\}$.

Lemma 1: Kolmogorov-Centsov

Let $f : [0,1] \to \mathbb{R}$. Let Define $\Delta_n(f) = \max\{|f(\frac{k+1}{2^n}) - f(\frac{k}{2^n})| : 0 \leq k \leq 2^n - 1\}$. 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 $w_f(\delta) \leq 10 \sum_{n \geq m_\delta} \Delta_n(f)$ 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, $W(\frac{k+1}{2^n}) - W(\frac{k}{2^n})$, $0 \leq k \leq 2^n - 1$, are i.i.d. $N(0, 2^{-n})$. Hence, by the tail estimate of the Gaussian distribution,

$$\mathbf{P}\left\{\Delta_n(f) \geq 2\frac{\sqrt{n}}{\sqrt{2^n}}\right\} \;\leq\; 2^n \mathbf{P}\left\{|\xi| \geq 2\sqrt{n}\right\} \;\leq\; 2^n \exp\left\{-\frac{1}{2}(4n)\right\} \;\leq\; 2^{-n}.$$

By the Borel-Cantelli lemma, it follows that $\Delta_n \leq 2\frac{\sqrt{n}}{\sqrt{2^n}}$ for all $n \geq N$ 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 $n \le N(\omega)$ and write

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

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

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

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 $0 \le t_1 < \ldots < t_m \le 1$, then find $t_{i,n} \in D$ that converge to t_i, for $1 \le i \le m$. Then $(W_{t_{1,n},\ldots,W_{t_{m,n}}}) \overset{a.s.}{\to} (W_{t_1},\ldots,W_{t_m})$. 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

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

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.

Proof of Lemma 1. A function on D and its extension to $[0,1]$ have the same modulus of continuity. Hence, it suffices to show that $|f(t)-f(s)| \le 10\sum_{n\ge m_\delta}\Delta_n(f)$ for $t,s \in D$, $|t-s| \le \delta$.

Let $0 < t-s \le \delta$, $s,t \in D$. We write $I = [s,t]$ as a union of dyadic intervals using the following greedy algorithm. First we pick the largest dyadic interval (by this we mean an interval of the form $[k2^{-n},(k+1)2^{-n}]$ for some n,k). contained in $[s,t]$. Call it, I_1 and observe that $|I_1| = 2^{-m}$ where $2^{-m} \le t-s \le 4.2^{-m}$. Then inside $I \setminus I_1$, pick the largest possible dyadic interval I_2. Then pick the largest possible dyadic interval in $I \setminus (I_1 \cup I_2)$ and so on. Since $t,s \in D_n$ for some n and hence, in a finite number of steps we end up with the empty set, i.e., we arrive at $I = I_1 \sqcup I_2 \sqcup \ldots \sqcup I_q$ for some positive integer q.

A little thought shows that for the lengths of I_j are non-increasing in j and that for any $n \ge m$, at most two of the intervals $I_1,\ldots,I_q$ can have length 2^{-n}. Write the intervals from left to right and express $f(t)-f(s)$ as a sum of the increments of f over these intervals to see that

$$|f(t)-f(s)| \le 2\sum_{n\ge m}\Delta_n(f).$$

Since $2^{-m} \le t-s$, we see that $m \ge \log_2\frac{1}{t-s} \ge m_\delta$ and hence the conclusion in the statement of the lemma follows. ∎

We put together the conclusions in the following theorem and extend the index set to $\mathbb{R}_+$.

Theorem 2

Standard Brownian motion $W = (W_t)_{t\in[0,\infty)}$ exists. Further, for any $\epsilon > 0$ and any $T < \infty$, w.p.1., the sample paths $t \mapsto W_t$ are uniformly Hölder$\left(\frac{1}{2} - \epsilon\right)$ on $[0, T]$.

Proof. We used countably many i.i.d. standard Gaussians to construct standard Brownian motion on $[0, 1]$. By using countably many such independent collections, we can construct (say on $([0, 1], \mathcal{B}, \lambda)$) a collection of independent Brownian motions $W^{(k)} = (W^k(t))_{t\in[0,1]}$. Then for $0 \le t < \infty$, define

$$W(t) = \sum_{k=1}^{m-1} W^{(k)}(1) + W^{(m)}(t - m)$$

if $m \le t < m + 1$ for $m \in \mathbb{N}$. In words, we just append the Brownian motions successively to the previous ones.

We leave it for you to check that W is indeed a standard Brownian motion. Each $W^{(k)}$ has modulus of continuity $O(\sqrt{\delta}\sqrt{\log \frac{1}{\delta}})$ which is of course $O(\delta^{\frac{1}{2}-\epsilon})$ for any $\epsilon > 0$. For finite T, only finitely many $w_{W[0,T]}(\delta) \le 2 \max\{w_{W^{(k)}[0,1]}(\delta) : k \le T + 1\}$. Hence, Hölder continuity holds on compact intervals. ∎

4. 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 $\mathbf{E}[X_t X_s] = K(t, s)$? 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 $h(\delta)^2 = \max\{K(t, t) + K(s, s) - 2K(t, s) : 0 \le t, s \le 1,\ |t - s| \le \delta\}$ (to understand what is happening, observe that if (X_t, X_s) has the prescribed bivariate Gaussian distribution, then $\mathbf{E}[(X_t - X_s)^2] = K(t, t) + K(s, s) - 2K(t, s)$). Then, each of $X(k + 12^n) - X(\frac{k}{2^n})$ 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 $\Delta_n \le \sqrt{10(1 + \delta)}\sqrt{n}h(2^{-n})$, 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 $h(\delta) \le C\delta^p$ for any positive p, then $\sum_n \sqrt{n}h(2^{-n}) < \infty$. In fact, it suffices if $h(\delta) \le (\log(1/\delta))^p$ for a sufficiently large p.

Beyond Gaussians: Now suppose for every $k \geq 1$ and every $0 \leq t_1 < t_2 < \ldots < t_k \leq 1$, 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 $\mathbf{P}\{Z > t\} \leq e^{-t^2/2}$. Now that we do not have that, assume that $\mathbf{E}[(X_t - X_s)^\alpha] \leq C|t-s|^{1+\beta}$ for some positive numbers C, α, β and for all $t, s \in [0,1]$. Observe that by $\mathbf{E}[|X_t - X_s|^\alpha]$ we mean the quantity $\int_{\mathbb{R}^2} |x-y|^\alpha d\mu_{t,s}(x,y)$. Then, it follows that

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

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

$$\mathbf{P}\{\Delta_n \geq u_n\} \leq C u_n^{-\alpha} 2^{-n\beta}.$$

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

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

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

Exercise 2

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

5. 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[2].

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

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

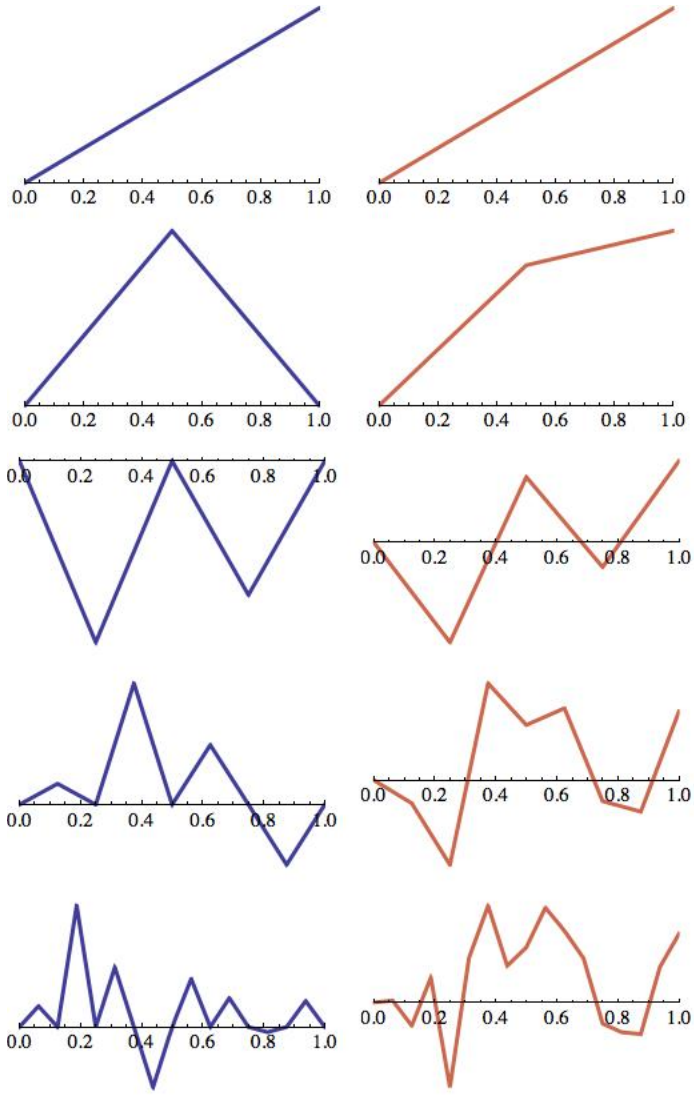

FIGURE 1. The first few steps in Lévy's construction. On the left are the functions F_n and on the right are the functions $F_0 + \ldots + F_n$, for $0 \leq n \leq 4$.

F_n by

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

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 5, you may see the first few steps of the construction.

We claim that $\|F_n\|_{\sup} \leq 10\frac{\sqrt{n}}{\sqrt{2^n}}$ with probability $\geq 1 - \frac{1}{2^n}$. 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 $\mathbf{E}[W(t)W(s)] = t \wedge s$ (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 $[0,1]$.

Remark 2

Let $I_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n}]$ for $0 \le k \le 2^n - 1$ and $n \ge 0$. 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 $H_{n,k}$, $0 \le k \le 2^n - 1$, $0 \le n$, 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:

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

6. Series constructions of Brownian motion

Let us do some formal (meaning, 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 $dW(t)$, $0 \le t \le 1$, 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 $\{\varphi_n\}$ for $L^2[0,1]$. We know that

$$\sum_n \langle f, \varphi_n \rangle \langle g, \varphi_n \rangle = \langle f, g \rangle$$

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

$W(0) = 0$, we must have

$$W(t) = \sum_n X_n \int_0^t \varphi_n(u)du \tag{1}$$

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 (1). 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 $t \mapsto \int_0^t \varphi_n$ is), and $W(t)$s will be jointly Gaussian and the means are zero. If we check that the covariances match those of Brownian motion, that gives a new construction (or a new representation) of Brownian motion! I do not know if this works for any orthonormal basis, but here we look at a few specific ones.

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: $1, \sqrt{2}\cos(2\pi nt), \sqrt{2}\sin(2\pi nt), n \geq 1$, form an orthonormal basis[3] for $L^2[0,1]$. In this case, the series form (1) 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 $\sqrt{2}\cos[\pi(n+\frac{1}{2})t]$, $n \geq 0$, form an orthonormal basis of $L^2[0,1]$. The series (1) then becomes

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

[3]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, 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 (1) is

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

where $\{\varphi_n\}$ 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 the real and imaginary parts are i.i.d. $N(0,1)$ random variables).

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 $e_n(t) = e^{2\pi int}$. Then the series becomes

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

The series converges uniformly with probability 1 and gives complex Brownian motion.

6.1. **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 3

The series $\sum_n \frac{Z_n}{2\pi in} e^{2\pi int}$ 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 3. For $n \geq 1$ define

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

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 $n \geq 1$ 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

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

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

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

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

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

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

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

Therefore $\mathbf{E}[\|F_n\|] \leq 5 \times 2^{-n/4}$ 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. Then one must show that the resulting random continuous function has the covariance structure of Brownian motion. If the series is $\sum_n \xi_n \psi_n(t)$, then all we want is to show that

$$
\sum_n \psi_n(t)\psi_n(s) = t \wedge s.
$$

Let us carry this out in the series (2). We shall assume knowledge of spectral decomposition of compact operators (if you do not, omit the proof).

Lemma 4

The series $\sum_{n \geq 0} \frac{\sin[\pi(n+\frac{1}{2})s]}{\pi(n+\frac{1}{2})} \frac{\sin[\pi(n+\frac{1}{2})t]}{\pi(n+\frac{1}{2})}$ converges uniformly on $[0,1] \times [0,1]$ to $\frac{1}{2}(t \wedge s)$.

Proof. Let $K(t,s) = t \wedge s$ and define the operator $T : L^2[0,1] \mapsto L^2[0,1]$ by

$$Tf(t) = \int_0^1 K(t,s)f(s)ds.$$

It is well-known (and easy to check) that T is a compact operator (as it can be approximated by finite dimensional operators using Riemann sums) and is self-adjoint (as $K(t,s) = K(s,t)$). Hence, we know by the spectral theorem that $T\psi_n = \lambda_n \psi_n$ for an orthonormal set $\{\psi_n\}$ in $L^2[0,1]$ and a sequence λ_n (non-zero real numbers) that converges to zero and such that $\text{Ker}(T) = \text{span}\{\psi_1, \psi_2, \ldots\}^\perp$. We proceed to find these ψ_ns. As K is a positive definite kernel, it is also true that $\lambda_n > 0$ for all n.

Suppose $Tf = \lambda f$ for some $f \in L^2$ and $\lambda > 0$, then

$$\lambda f(t) = \int_0^t sf(s)ds + t\int_t^1 f(s)ds.$$

A priori, the equality is in L^2, but since $f \in L^2 \subseteq L^1$, the right hand side is continuous in t, and hence f is continuous. But then the right hand side becomes differentiable in t, hence f is differentiable. Inductively, we see that f is smooth and that the above identity holds pointwise for $t \in [0,1]$. Differentiate twice to get $\lambda f''(t) = -f(t)$ which implies that $f(t) = a\sin(t/\sqrt{\lambda}) + b\cos(t/\sqrt{\lambda})$ for some $a, b \in \mathbb{R}$. From the above identity, we also see that $f(0) = 0$ and $f'(1) = 1$. This forces $b = 0$ and $\cos(1/\sqrt{\lambda}) = 0$ or $\frac{1}{\sqrt{\lambda}} = (n + \frac{1}{2})\pi$ fr some $n \geq 0$. Thus we have $\psi_n(t) = \sqrt{2}\sin((n+\frac{1}{2})\pi t)$ (normalized so that $\int_0^1 \psi_n^2 = 1$) and $\lambda_n = \frac{1}{\pi^2(n+\frac{1}{2})^2}$, for $n = 0, 1, 2 \ldots$. Define,

$$L(t,s) = \sum_n \lambda_n \psi_n(t)\psi_n(s) = \sum_{n\geq 0} \frac{\sin[\pi(n+\frac{1}{2})s]}{\pi(n+\frac{1}{2})} \frac{\sin[\pi(n+\frac{1}{2})t]}{\pi(n+\frac{1}{2})}.$$

The last series clearly converges uniformly on $[0,1]^2$, since the nth term is uniformly bounded by $1/n^2$. But then $\int_0^1 L(t,s)f(s)ds = \sum_n \lambda_n \psi_n(t)\langle f, \psi_n\rangle \stackrel{L^2}{=} Tf$ (to see the last equality write $f = \sum_n \langle f, \psi_n\rangle \psi_n$ and apply T). Thus, K and L define the same integral operator and hence $K = L$ $a.e.$ Both are continuous on $[0,1]^2$, hence equal everywhere. That is $K(t,s)$ is given by the series in the statement of the lemma. ■

7. Basic properties of Brownian motion

We have given two constructions of Brownian motion (and outlined one more). However, in our further study of Brownian motion, we would not like to use the specifics of this construction, but only the defining properties of Brownian motion. To this end, let us recall that standard Brownian motion is a collection of random variables $W = (W_t)_{t\in[0,\infty)}$ on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that

(1) $t \mapsto W_t(\omega)$ is continuous for $\mathbf{P}$-$a.e.$ ω,

(2) Increments over disjoint intervals are independent,

(3) $W_t - W_s \sim N(0, t-s)$ for any $s < t$.

Equivalently, we may define W as a $C[0,\infty)$-values random variable such that W_t, $t \geq 0$, are jointly Gaussian with mean zero and covariance $\mathbf{E}[W_tW_s] = t \wedge s$.

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 is continuous w.p.1., X_t are jointly Gaussian and $X_t - X_s = -(W_t - W_s)$ has $N(0, t-s)$ distribution by the symmetry of mean zero Gaussian distribution.
- (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.
- (Time-reversal symmetry) Let W be standard Brownian motion on $[0,1]$. Define $X(t) = W(1-t) - W(1)$ for $0 \leq t \leq 1$. Then X is standard Brownian motion on $[0,1]$.
- (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
$$\mathbf{E}[X_tX_s] = ts\mathbf{E}\left[W\left(\frac{1}{t}\right)W\left(\frac{1}{s}\right)\right] = ts\frac{1}{t} = s.$$
Thus, $(X_s)_{s\in(0,infty)}$ has the same distribution as $(W_s)_{s\in(0,\infty)}$. In particular, if $M_\delta^X = \sup_{0<s\leq\delta} X_s$ and $M_\delta^W = \sup_{0<s\leq\delta} W_s$, then $(M_{1/k}^X)_{k\geq 1}$ has the same distribution as $(M_{1/k}^W)_{k\geq 1}$. But $\lim_{k\to\infty} M_{1/k}^W = 0$ w.p.1., and hence $\lim_{k\to\infty} M_{1/k}^X = 0$ 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 $t_0 \geq 0$ 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
$$\begin{aligned}\mathbf{E}[X_tX_s] &= \mathbf{E}[W_{s+t_0}W_{t+t_0}] - \mathbf{E}[W_{t_0}W_{t+t_0}] - \mathbf{E}[W_{s+t_0}W_{t_0}] + \mathbf{E}[W_{t_0}W_{t_0}] \\ &= (s+t_0) - t_0 - t_0 + t_0 \\ &= s.\end{aligned}$$
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.

8. Other processes from Brownian motion

Having constructed Brownian motion, we can use it to define various other processes with behaviour modified in many ways.

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 $t \geq 0$, is called Brownian motion started at x.

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. 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 (especially if σ is small). Brownian motion moves very randomly, these processes have a deterministic motion on which a layer of randomness is added.

Multi-dimensional Brownian motion: Brownian motion in $\mathbb{R}^d$, started at $x \in \mathbb{R}^d$, is defined as the stochastic process $W = (W(t))_{t \geq 0}$ where $W(t)$ are $\mathbb{R}^d$-valued random variables,(a) $W(0) = x \ a.s.$, (b) for any $t_1 < \ldots < t_k$, the increments $W(t_1)$, $W(t_2) - W(t_1)$,... $W(t_k) - W(t_{k-1})$ are independent, (c) for any $s < t$ the distribution of $W(t) - W(s)$ is d-dimensional Gaussian with zero mean and covariance matrix $(t-s)I_d$, and (d) $t \mapsto W(t)$ is continuous with probability 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 $W^{(k)}$, $k \geq 1$, of standard 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 $v_1, \ldots, v_d$ of $\mathbb{R}^d$ and define $W(t) = x + W^{(1)}(t)v_1 + \ldots + W^{(d)}(t)v_d$, 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 $X \stackrel{d}{=} W$.

Ornstein-Uhlenbeck process: Is it possible to define Brownian motion indexed by $\mathbb{R}$ instead of $[0, \infty)$. An obvious thing is to take two independent standard Brownian motions and set $X(t) = W_1(t)$ for $t \geq 0$ and $X(t) = W_2(-t)$, then X may be called a two-sided Brownian motion. Somehow, it is not satisfactory, since the location 0 plays a special role (the variance of $X(t)$ increases on either side of it).

A better model is to set $X(t) = e^{-\frac{1}{2}t}W(e^t)$ for $t \in \mathbb{R}$. Then X is called Ornstein-Uhlenbeck process. It is a continuous process and X_t, $t \in \mathbb{R}$ are jointly Gaussian with zero means and covariances $\mathbf{E}[X_t X_s] = e^{-\frac{1}{2}(s+t)}\mathbf{E}[W(e^s)W(e^t)] = e^{-\frac{1}{2}|s-t|}$. 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.

Brownian bridge: Brownian bridge is the continuous Gaussian process $X = (X(t))_{t \in [0,1]}$ such that $\mathbf{E}[X_t X_s] = s(1-t)$ for $0 \leq s < t \leq 1$. Observe that $X(0) = X(1) = 0$ w.p.1. It arises in many

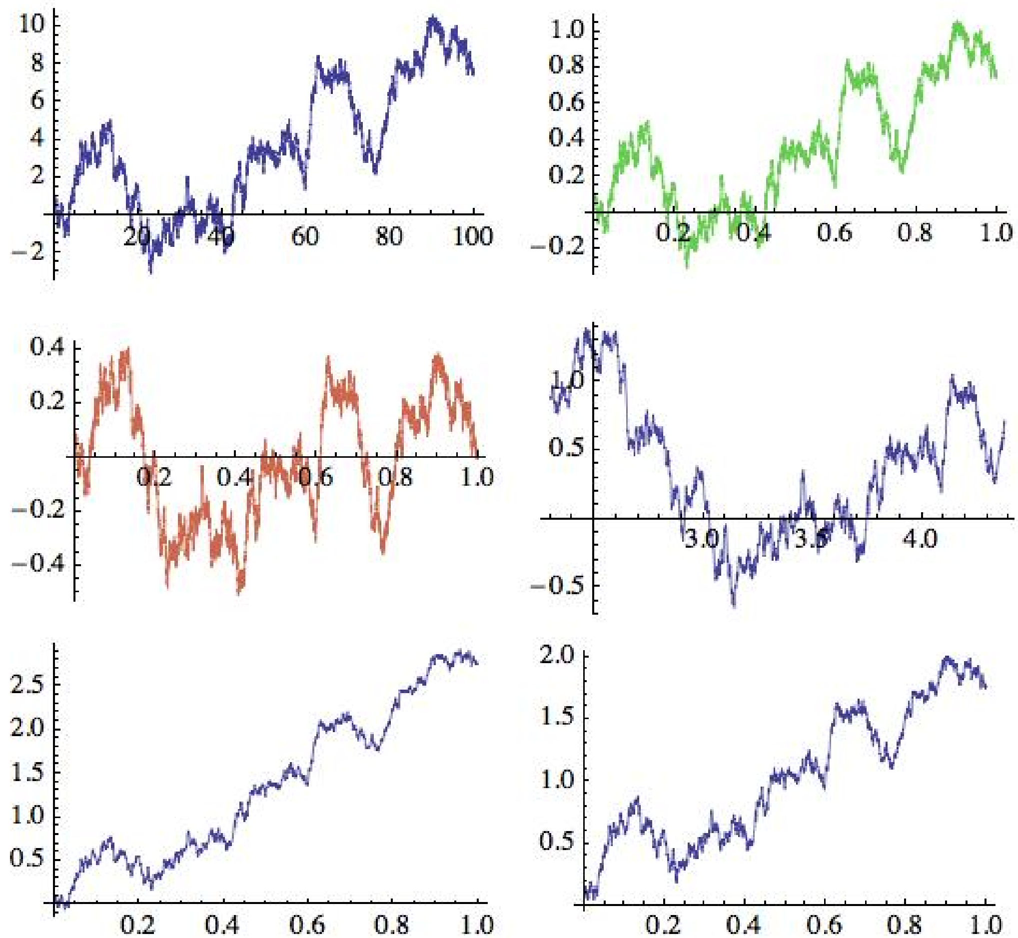

FIGURE 2. Top row left: Brownian motion run for time 10. Top row right: The same after a time-space scaling to time interval $[0,1]$. Middle row left: A Brownian bridge. Middle row right: An Ornstein-Uhlenbeck sample path. Bottom row left: Brownian motion with linear drift $2t$. Bottom row right: $W_t + \sqrt{2}t$. Take note of the markings on both axes.

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.

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

There is a third description of Brownian bridge. Consider standard Brownian motion $W = (W(t))_{t\in[0,1]}$ on some $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\mathcal{G} = \sigma\{W(1)\}$. Then, a regular conditional distribution of W given $\mathcal{G}$ exists. We may write it as $\mu(A, x)$, where $A \in \mathcal{B}(C[0,1])$ and $x \in \mathbb{R}$ (so $\mu(\cdot, x)$ is a probability measure that indicated the distribution of W given that $W(1) = x$). It can be checked that the conditional distributions are continuous in x. In fact, there is one measure μ_0 on $C[0,1]$

such that $\mu(A,x) = \mu_0\{g : t \mapsto g(t) + tx \text{ is in } A\}$. This is given in the homework and will be left as exercise.

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 (i.e., it is similar everywhere) and the motion is isotropic (there is no preferred direction of motion). If one imagines motion in a non-homogeneous medium, 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 $X = (X(t))_{t \geq 0}$ that has continuous sample paths, independent increments over disjoint intervals of time and such that conditional on $X(s)$, $s \leq t$, 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) 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. But more generally, it is not easy to show that such a process exists[4] and we shall not be able to touch upon this topic in this course.

9. Plan for the rest of the course

So far we have defined and constructed Brownian motion, and seen the most basic symmetries of it. We shall study the following aspects which cover only a small fraction (but reasonable enough for a first course) of things one could study about Brownian motion.

► Continuity properties of Brownian motion. The modulus of continuity is $O(\sqrt{\delta \log(1/\delta)})$ and hence it is Hölder($\frac{1}{2} - \epsilon$) for any $\epsilon > 0$. We shall see that W is nowhere Hölder($\frac{1}{2} + \epsilon$) for any $\epsilon > 0$.

► Markov property and martingales in Brownian motion. Brownian motion will be shown to have Markov and strong Markov property. We shall extract many martingales out of it. All this will be used to get substantial information about the maximum of a Brownian motion, the zero set, the time to exit a given set, recurrence and transience, etc. If time permits, we shall see the relationship between multi-dimensional Brownian motion and harmonic functions and the Dirichlet problem.

► Brownian motion as a limiting object. We shall see that random walks converge to Brownian motion (Donsker's theorem). We shall use the connection between random walks and Brownian motion to deduce results about each from results about the other (eg., law of iterated logarithm,

[4] One will have to either develop stochastic calculus first or a theory of general Markov processes and some existence theorems for Elliptic partial differential equations.

some arc-sine laws). If time permits we relate the difference between empirical distribution of an i.i.d. sample and the true distribution to a Brownian bridge.

► There are many other aspects we may not have time for. Some of them are the ideas of Wiener integral with respect to Brownian motion, Cameron-Martin formula, Hausdorff dimensions of random fractal sets coming from Brownian motion, stochastic Calculus ...

10. Further continuity properties of Brownian motion

Let W denote standard Brownian motion in $[0,1]$. We have see that W is Hölder($\frac{1}{2}-\epsilon$) for any $\epsilon > 0$ with probability 1. We shall show in this section that it is nowhere Hölder($\frac{1}{2}+\epsilon$) for any $\epsilon > 0$, in particular, the paths are nowhere differentiable.

If $f : [0,1] \to \mathbb{R}$ and $0 < \alpha \le 1$, 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 $\limsup$ 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 5: Paley, Wiener, Zygmund

With probability 1, the following statements hold.

(1) Standard Brownian motion is nowhere differentiable.

(2) Standard 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.

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 but anyway...

Proof of nowhere differentiability due to Dvoretksy, Erdös and Kakutani. If f is differentiable at t, then $|f(s)-f(t)| \le C|s-t|$ for some $C < \infty$ and for all $s \in [0,1]$. Then, $|f(s)-f(u)| \le C(|s-t|+|u-t|)$ for all $s, u \in [0,1]$. In particular, for any $n \ge 0$ and any $0 \le k \le 2^n - 1$, 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, $\ell + 3$ 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

$$\left| f\left(\frac{k+1}{2^n}\right) - f\left(\frac{k}{2^n}\right) \right| \le C\frac{10}{2^n} \tag{3}$$

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

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

$$\mathcal{A}_{n,C} = \{f : (3) \text{ holds for at least three consecutive } k \text{ in } 0, 1, \dots, 2^n - 1\},$$

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

We show for each fixed C that $\mathbf{P}\{W \in \mathcal{A}_{n,C}\} \to 0$ as $n \to \infty$. This implies[5] that $\mathbf{P}\{W \in \mathcal{A}\} = 0$. To show this,

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

This proves the nowhere differentiability of Brownian motion. ■

By considering several increments in place of three, one can show that W has no Hölder($\frac{1}{2} + \epsilon$) points.

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 $n \ge 0$ and $0 \le k \le 2^n - 1$.

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 $|f(s) - f(t)| \le c\sqrt{|s-t|}$ for all $s \in [t - 2^{-M}, t + 2^{-M}]$. In particular, if $n \ge M$ and $I_{n,k}$ is the dyadic interval containing t, then

$$|\Delta f(I)| \ \le \ c\left\{\sqrt{(k+1)2^{-n} - t} + \sqrt{t - k2^{-n}}\right\} \ \le \ \frac{\sqrt{2}c}{\sqrt{2^n}}. \tag{4}$$

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

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

[5]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 $\mathbf{P}\{W \in \mathcal{A}_1\} = \mathbf{P}\{W \in \mathcal{A}_2\}$, then the same is deemed to be the value of $\mathbf{P}\{W \in \mathcal{A}_0\}$.

children. Let us declare a dyadic interval $I_{n,k}$ to be alive if it satisfies $\Delta f(I_{n,k}) \le c\sqrt{2}/\sqrt{2^n}$. 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 $\mathcal{F}_n = \{\Delta W(I_{n,k}) : 0 \le k \le 2^n - 1\}$ 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 10 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 $|\Delta W(J)| \le \frac{c\sqrt{2}}{\sqrt{2^n}}$. This means that ξ must lie in an interval of length $4c$ centered at $\sqrt{2^{n-1}}\Delta W(I)$. By Exercise 10, 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 3

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 4

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