PROBABILITY THEORY - PART 4
BROWNIAN MOTION

MANJUNATH KRISHNAPUR

CONTENTS

1. Definition of Brownian motion and Wiener measure

Definition 1. *Brownian motion* is 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.

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 3. *(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.

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 $\mathbf{E}[W_tW_s] = t \wedge s$. 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 $F\mid_D = f$. 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 $w_f(\delta) = \sup\{|f(t) - f(s)| : |t-s| \le \delta\}$. 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 $w_f(\delta) \le C_f\delta$ and if f is Hölder(α) for some $0 < \alpha \le 1$, then $w_f(\delta) \le C_f\delta^\alpha$. 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 $D_n = \{k2^{-n} : 0 \le k \le 2^n\}$.

Lemma 4. *Let $f : [0,1] \to \mathbb{R}$. Let Define $\Delta_n(f) = \max\{|f(\frac{k+1}{2^n}) - f(\frac{k}{2^n})| : 0 \le k \le 2^n - 1\}$. 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 $w_f(\delta) \le 10\sum_{n \ge m_\delta} \Delta_n(f)$ 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, $W(\frac{k+1}{2^n}) - W(\frac{k}{2^n})$, $0 \le k \le 2^n - 1$, are i.i.d. $N(0, 2^{-n})$. Hence, by the tail estimate of the Gaussian distribution,

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

By the Borel Cantelli lemma, it follows that $\Delta_n \le 2\frac{\sqrt{n}}{\sqrt{2^n}}$ for all $n \ge N$ 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 $(W_{t_{1,n},\ldots,W_{t_{m,n}}}) \stackrel{a.s.}{\to} (W_{t_1}, \ldots, W_{t_m})$. 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

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

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 4. 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 5. *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. 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 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 4, 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 6. *Let $I_{n,k} = [\frac{k}{2^n}, \frac{k+1}{2^n}]$ for $0 \leq k \leq 2^n - 1$ and $n \geq 0$. Define $H_{n,k} : [0,1] \to \mathbb{R}$ by*

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

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

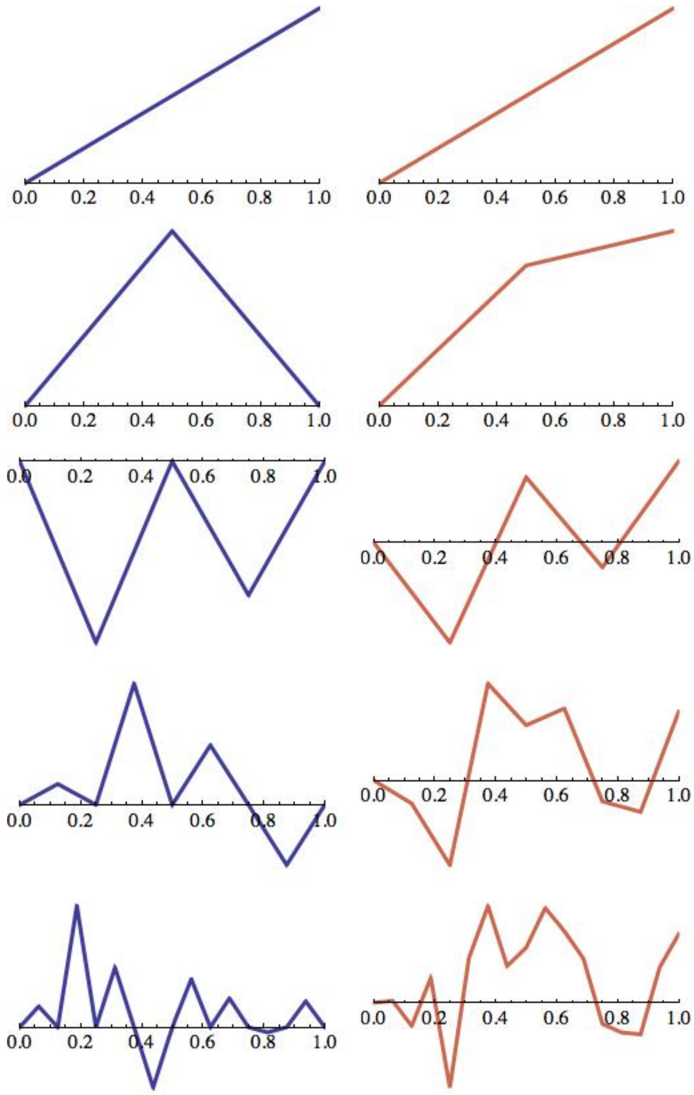

FIGURE 1. The first few steps in Lévy's construction. On the left are the functions F_n and on the right are the functions $F_0 + \ldots + F_n$, for $0 \leq n \leq 4$.

Then, together with the constant function $\mathbf{1}$*, the collection* $H_{n,k}$*,* $0 \leq k \leq 2^n - 1$*,* $0 \leq n$*, 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:

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

5. 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 $\mathbf{P}$-$a.e.\ \omega$,

(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 $\mathbf{E}[W_t W_s] = t \wedge s$.

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). 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

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

  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 $\lim_{t\o 0+} X(t) = 0$ 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

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

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.

6. 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: Let $W^{(k)}$, $k \geq 1$ be i.i.d. copies of standard Brownian motion. Then, for $d \geq 1$, define $W(t) = (W_1(t), \ldots, W_d(t))$. This is called multi-dimensional Brownian motion.

It may appear at first sight that this is somehow dependent on our choice of the standard orthonormal basis - what if we pick a different orthonormal basis? The answer is that the distribution of the process does not depend on the choice of basis. This can be seen using orthogonal invariance of standard Gaussian distribution. Alternately, do the following exercise (which clearly shows the basis-independence).

Exercise 7. *Let W be defined as above (so $W(t)$ is an $\mathbb{R}^d$-valued random vector). Show that (1) $W(0) = 0$ and $t \mapsto W(t)$ is continuous w.p.1., (2) for any $t_1 < \ldots < t_m$, the random vectors $W(t_1), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1})$ are independent, (3) For any $s < t$, $W(t) - W(s) \sim N_d(\mathbf{0}, (t-s)I_d)$ where I_d is the $d \times d$ identity matrix.*

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

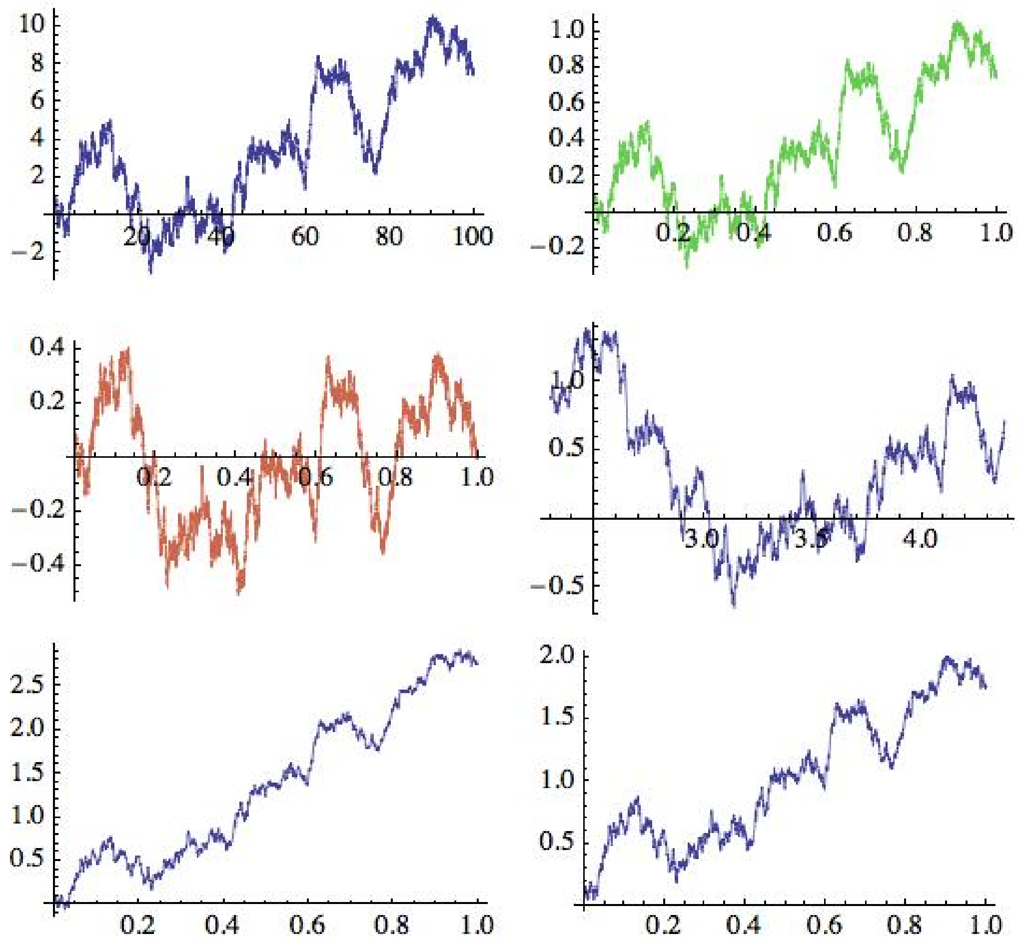

FIGURE 2. Top row left: Brownian motion run for time 10. Top row right: The same after a time-space scaling to time interval $[0, 1]$. Middle row left: A Brownian bridge. Middle row right: An Ornstein-Uhlenbeck sample path. Bottom row left: Brownian motion with linear drift $2t$. Bottom row right: $W_t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

$X(t) = W_1(t)$ 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: Define $X(t) = W(t) - tW(1)$ for $0 \le t \le 1$. Then X is continuous in t w.p.1., X_t are jointly Gaussian, and $\mathbf{E}[X_t X_s] = s(1-t)$ for $0 \le s < t \le 1$. Observe that $X(0) = X(1) = 0$ w.p.1. The process X is called *Brownian bridge*. It arises in many situations, but for now we simply motivate is 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.

7. 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, 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 . . .

8. Further continuity properties of Brownian motion

Let W denote standard Brownian motion in $[0,1]$. We have see that W is Hölder($\frac{1}{2} - \epsilon$) w.p1. for any $\epsilon > 0$. 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 $c' > c$).

Theorem 8 (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 $c < 0.6$, 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

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

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

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

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

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

We show for each fixed C that $\mathbf{P}\{W \in \mathcal{A}_{n,C}\} \to 0$ as $n \to \infty$. This implies[3] that $\mathbf{P}\{W \in \mathcal{A}\} = 0$. To show this,

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

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

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

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 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

[3] 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 $\mathbf{P}\{W \in \mathcal{A}_0\}$.

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 9 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 9. *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 10. *If $\xi \sim N(0,1)$, then $\sup\limits_{a\in\mathbb{R}} \mathbf{P}\{\xi \in [a-t, a+t]\} \le \frac{2t}{\sqrt{2\pi}}$.*

9. Summary of continuity properties

Let W be standard Brownian motion on $[0,1]$. First and foremost is the point that $\mathbf{E}[|W_t - W_s|^2] = |t-s|$ 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 $\limsup\limits_{h\downarrow 0} \max\limits_{t\in[0,1]} \frac{|W_t - W_s|}{\sqrt{h\log(1/h)}} < \infty$. We showed this (and if you follow our proof closely, you will see that the left hand side can be shown to be ≤ 10 w.p.1.).

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

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

(2) Almost surely W has no Hölder($\frac{1}{2}; c$) points for c sufficiently small. As a consequence, it is nowhere Hölder($\frac{1}{2}+\epsilon$) 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 $t \geq 0$,

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

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 $[-1, 1]$.

10. Wiener's stochastic integral

Let W be standard Brownian motion on $(\Omega, \mathcal{F}, \mathbf{P})$. Let $f : [0,1] \to \mathbb{R}$. We want to make sense of $\int_0^1 f(t)dW(t)$ 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 $f \in C[0,1]$ we may define $\int_0^1 f(t)d\alpha(t)$ as $\int_0^1 f(t)\alpha'(t)dt$, the latter being the Riemann integral of a continuous function.
- More generally, if α is a function of bounded variation[4], then following ideas similar to that of Riemann intergral, Stieltjes showed that $\int_0^1 f(t)d\alpha(t)$ can be made sense of for any $f \in C[0,1]$.
- Suppose $\alpha \in C[0,1]$, not necessarily of bounded variation. Then it is no longer possible to define Stieltjes' integral. But for $f \in C[0,1]$, we can define

$$\int_0^1 f(t)d\alpha(t) := f(1)\alpha(1) - f(0)\alpha(0) - \int_0^1 \alpha(t)f'(t)dt.$$

The justification for this definition is that when α is of bounded variation, the expression on the right is equal to $\int_0^1 f d\alpha$, 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 $f \in C[0,1]$.

[4]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 $0 = t_0 < t_1 < \ldots < t_n = 1$. 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. A

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 $\int_0^1 f(t)dW(t)$ for every $f \in L^2[0,1]$! This is done as follows.

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

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

If $\mathcal{S}$ denotes the collection of all step functions on $[0,1]$, then $\mathcal{S}$ is a dense subspace of $L^2[0,1]$. What we have defined is a function $I : \mathcal{S} \to L^2(\Omega, \mathcal{F}, \mathbf{P})$.

Step 2: We claim that $I : \mathcal{S} \to L^2(\Omega, \mathcal{F}, \mathbf{P})$ 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

$$\|I(f)\|^2_{L^2(\mathbf{P})} = \mathbf{E}[|I(f)|^2] = \mathrm{Var}(\sum_{k=1}^n \lambda_k (W(b_k) - W(a_k))) = \sum_{k=1}^n \lambda_k^2 (b_k - a_k) = \|f\|^2_{L^2[0,1]}.$$

That $I(f)$ is Gaussian is clear. Therefore it has $N(0, \|f\|^2_{L^2[0,1]})$ distribution.

Step 3: I maps Cauchy sequences in $\mathcal{S}$ to Cauchy sequences in $L^2(\mathbf{P})$. Hence, if $f_n \in \mathcal{S}$ and $f \in L^2[0,1]$ and $f_n \to f$ in L^2, then $\{f_n\}$ is Cauchy in $L^2[0,1]$ and therefore $\{I(f_n)\}$ is Cauchy in $L^2(\mathbf{P})$. By completeness of $L^2(\mathbf{P})$, $I(f_n)$ 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 $L^2[0,1]$ into $L^2(\mathbf{P})$.

This defines the stochastic integral and we usualy write $\int_0^1 f(t)dW(t)$ for $I(f)$. 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, $\mathrm{Var}(I(f)) = \|f\|^2_{L^2[0,1]}$.

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 $f \in L^2[0,1]$ (even $f \in C[0,1]$) that cannot be integrated in any sense against the Brownian path.

(2) For a fixed $f \in L^2[0,1]$, 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.

We shall not really use the Wiener integral to deduce any properties of Brownian motion. We discussed it to mention an important topic that we shall not touch upon in this course. This is the subject of *Ito integral* which makes sense of integrals of random functions such as $\int_0^t W_s dW_s$ (it turns out that this integral is $\frac{1}{2}W_t^2 - \frac{1}{2}t$ in contrast to C^1 functions α for which we always have $\int_0^t \alpha(s)d\alpha(s) = \frac{1}{2}\alpha(t)^2$). Here is an exercise.

Exercise 11. *Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\int_0^t f^2 < \infty$ for all $t < \infty$. Then we may define $X_t = \int_0^t f(s)dW(s)$ exactly as above. Show that X is a martingale.*

Stochastic summation: an analogy or more: 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 $L_m(x) = \sum_{k\le m} a_k x_k$. Then L_m is a bounded linear functional with $\|L_m\|^2 = \sum_{k\le m} a_k^2$. 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 $\sum_{k\le m} a_k^2 \le C$ 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 $(\Omega, \mathcal{F}, \mathbf{P})$. 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 $x \in \ell^2$, using standard results on sums of independent random variables, it follows that $\sum_n \xi_n$ 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: $I : \mathcal{S} \mapsto L^2(\Omega, \mathcal{F}, \mathbf{P})$ 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 $\mathbf{E}[I(x)^2]$.

Step 3: I extends as an isometry of ℓ^2 into $L^2(\Omega, \mathcal{F}, \mathbf{P})$. 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 "$\int x(n)dS(n)$".

11. 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 $(\Omega, \mathcal{F}, \mathbf{P})$. Let $\mathcal{F}_t := \sigma\{W_s : s \le t\}$ be the associated natural filtration. Define $\tau = \inf\{t : W(t) \ge 1\}$ 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 $\mathcal{F}_t^+ := \bigcap\limits_{s>t} \mathcal{F}_s$. The filtration $\mathcal{F}_\bullet^+$ is called the right-continuous version of $\mathcal{F}_\bullet$ because $\bigcap\limits_{s>t} \mathcal{F}_s^+ = \mathcal{F}_t^+$ for every $t \ge 0$. It is easy to see that τ' is indeed a stopping time with respect to $\mathcal{F}_\bullet^+$, since the event $\{\tau' \le t\} \in \mathcal{F}_s$ 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 $\mathbf{P}$-null sets. This is indeed true.

Lemma 12 (Blumenthal's zero-one law). *If $A \in \mathcal{F}_0^+$, then $\mathbf{P}(A)$ equals 0 or 1.*

In the class on 8th October 2015, we did the following:

- Defined the right continuous filtration $\mathcal{F}_t^+$ and explained why we need it.
- Proved Blumenthal's zero-one law and by time reversal that the tail sigma-algebra is also trivial.
- Derived some consequences: (1) Started from 0, Brownian motion crosses 0 infinitely many times in any positive length of time. (2) Local maxima of Brownian motion are all strict, dense in time, and the values at the local maxima are distinct.

We followed closely the presentation in the book of Mörters and Peres, hence there will be no notes on this. It is also suggested to read the proof of Blumenthal's zero-one law from Durrett's book.

Let W be a standard d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. 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 $t \geq 0$. 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 $t \geq 0$. Then B is a standard Brownian motion independent of $\mathcal{F}_\tau^+$. Recall that $\mathcal{F}_\tau^+ = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$.

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 $\mathbf{P}$-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 $n \geq 1$ and $t_1, \ldots, t_n \geq 0$, and any $u_1 \ldots, u_n \in \mathbb{R}$. Let E be the event that $B(t_j) \leq u_j$ for $1 \leq j \leq n$. Then,

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

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

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

Adding up and using $\mathbf{P}(A) = \sum_m \mathbf{P}\{A \cap \{\tau = s_m\}\}$ gives the identity $\mathbf{P}\{E \cap A\} = \mathbf{P}\{W(t_j) \leq u_j \text{ for } j \leq n\}\mathbf{P}\{A\}$. 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 $\ell \geq 1$ define $\tau_\ell = 2^{-\ell}\lceil 2^\ell \tau \rceil$. Then τ_ℓ is a stopping time, $\tau \leq \tau_\ell \leq \tau + 2^{-\ell}$. 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

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

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 $W(T+t)$, $t \geq 0$, 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 $\mathcal{F}_\tau^+$.

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 $\mathbf{P}_x$, $x \in \mathbb{R}$, on $C[0,\infty)$ such that $\mathbf{P}_x\{f : f(0)=x\} = 1$. This family is said to have (time-homogeneous) Markov property if:

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

Example 13. *Let $\mathbf{P}_x$ be the distribution of $(x+W_t+t)_t$ for $x \geq 0$ and the distribution of $(x+W_t-t)_{t\geq 0}$ for $x<0$. Then $\mathbf{P}_x$ does not have Markov property.*

Example 14. *Let $\mathbf{P}_x$ be the distribution of $x+W$ for $x \neq 0$ and let $\mathbf{P}_0 = \delta_0$ be the Dirac measure at the constant function zero. Then, $\mathbf{P}$ 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 $(W(T+t))_{t\geq 0}$ 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 $(W(\tau+t))_{t\geq 0}$ is the standard Brownian motion, which is not the same as $\mathbf{P}_{W(\tau)} = \mathbf{P}_0$.

13. 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 15. *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 $[0,\epsilon)$. 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. ■

14. Reflection principle

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

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 $W(T_*) = \max_{s\le 1} W(s)$. 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 16 (Reflection principle)**.** *Let W be standard 1-dimensional Brownian motion. Fix $a > 0$ and define*

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

Then, B is a standard Brownian motion.

Proof. Let $X = (W_t)_{t\le\tau_a}$, $Y = (W_{t+\tau_a} - a)_{t\ge 0}$ 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 $B \stackrel{d}{=} W$. ∎

Distribution of the running maximum: Let $a > 0$ and $t > 0$. W is a standard Brownian motion and B is related to it as in the reflection principle. Then,

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

Therefore, $\mathbf{P}\{M_t > a\} = 2\mathbf{P}\{W_t > a\} = 2\bar{\Phi}(a/\sqrt{t})$ where $\bar{\Phi}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$ is the tail of the standard normal distribution. Differentiating, we get the density of M_t to be

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

Another way to say this is that for each fixed t we have $M_t \stackrel{d}{=} |W_t|$.

Distribution of the first passage times: Since $\tau_a \le t$ if and only if $M_t \ge a$, we see that $\mathbf{P}\{\tau_a \le t\} = 2\bar{\Phi}(a/\sqrt{t})$. The density of τ_a is

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

The density approaches zero at $t = 0$ and decayse like $t^{-3/2}$ as $t \to \infty$.

Exercise 17. *Deduce that $\mathbf{E}[\tau_a] = \infty$. In fact $\mathbf{E}[\tau_a^p] < \infty$ if and only if $p < \frac{1}{2}$.*

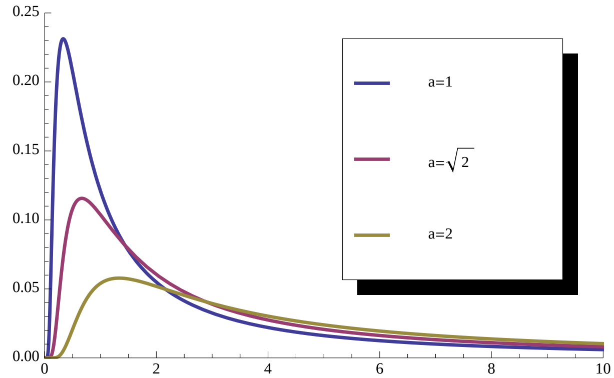

FIGURE 3. Densities of first passage time τ_a

Joint distribution of the Brownian motion and its running maximum: 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 $\mathbf{P}\{M_t > a \text{ and } W_t < b\} = \bar{\Phi}((2a-b)/\sqrt{t})$. Thus,

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

Some distributional identities: The process $(|W_t|)_{t\geq 0}$ is called reflected Brownian motion. We have the following distributional identities.

- $M_t \stackrel{d}{=} |W_t|$ for each t. We already saw this.
- $M_t - W_t \stackrel{d}{=} |W_t|$ for each t. This can be computed from the joint intensity, but here is a computation-free proof[5]. The process $X_s = W_{t-s} - W_t$ for $0 \leq s \leq t$ is a standard Brownian motion. Observe that $M_t^X = M_t^W - W_t$. Hence, the distributional identity follows!

Do these equalities in distribution extend to those of the processes? The first one does not, since M is an increasing process while $|W|$ is not. But it is a non-trivial theorem of Lévy that $M - W \stackrel{d}{=} |W|$. The key point is that $M - W$ is a Markov process. Once that is checked, the equality in distribution at fixed times easily extends to equality of finite dimensional distributions. Since both processes are continuous, this implies equality in distribution of the two processes.

Local times - a digression: Further, consider 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

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

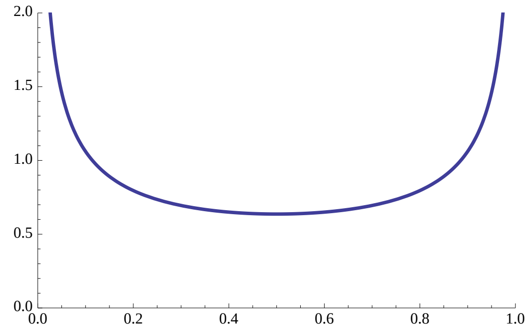

FIGURE 4. Arcsine density

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

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

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.

15. ARCSINE LAWS

Let $T^* = \arg \max_{0 \le t \le 1} W_t$ be the location of the global maximum of Brownian motion in unit time. As all the values of local maxima are distinct, T^* is well-defined. Also, define $L := \max\{t \le 1 : W_t = 0\}$ be the time of last return to the origin.

Let us also recall the *arc-sine* distribution that has CDF $\frac{2}{\pi} \arcsin(\sqrt{t})$ and density $\frac{1}{\pi\sqrt{t(1-t)}}$, for $0 < t < 1$.

Theorem 18 (Lévy). *T^* and L have the arcsine distribution.*

Proof. (1) Fix $t \in (0, 1)$. Then $T^* \le t$ if and only if $\max_{0 \le s \le t} W_s \ge \max_{t \le s \le 1} W_s$ which is equivalent to $\max_{0 \le s \le t} W_s - W_t \ge \max_{t \le s \le 1} W_s - W_t$.

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

$$\mathbf{P}\{T^* \ge t\} = \mathbf{P}\{M_t \ge \tilde{M}_{1-t}\}.$$

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,

$$\mathbf{P}\{T^* \geq t\} = \mathbf{P}\{\sqrt{t}|X| \geq \sqrt{1-t}|Y|\} = \mathbf{P}\left\{\left|\frac{Y}{X}\right| \leq \frac{\sqrt{t}}{\sqrt{1-t}}\right\}.$$

It is an easy exercise that Y/X has standard Cauchy distribution and hence the last probability is equal to $\frac{2}{\pi}\arctan(\sqrt{t}/\sqrt{1-t})$ which is equal to $\frac{2}{\pi}\arcsin(\sqrt{t})$. This shows that T^* has arcsine distribution.

(2) $L \geq t$ if and only if W hits zero somewhere in $[t, 1]$. Let $\tilde{W}_s = W_{t+s} - W_t$ for $0 \leq s \leq 1-t$ which is a Brownian motion independent of $\mathcal{F}_t$.

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

$$\mathbf{P}\{L \geq t\} = \mathbf{P}\{\tilde{M}_{1-t} \geq |W_t|\}.$$

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^* and deduce that L must have arcsine distribution. ■

Lévy prove a third arcsine law. This is for $\{t \leq 1 : W_t > 0\}$, the proportion of time spent by the Brownian motion in the positive half-line. We shall prove this later.

16. 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 $\tau_{-a,a} \leq t$ if and only if $\max_{s \leq t} |W_s| \geq a$, 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. We know that W_t itself is a martingale. But W_t^2 is not. Indeed,

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

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

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

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. What is the general pattern?

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

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

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

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. The martingales that we got earlier are precisely $t^{n/2}H_n(W_t/\sqrt{t})$.

Exercise 19. *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 $n \geq 0$.*

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.

Standard normal: A standard normal or Gaussian random variable is one with density $\varphi(x) := \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$. Its distribution function is $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$ 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 $\prod_{i=1}^{n} \varphi(x_i) = (2\pi)^{-n/2} \exp\{-|\mathbf{x}|^2/2\}$ and the distribution is denoted by γ_n, so that for every Borel set A in $\mathbb{R}^n$ we have $\gamma_n(A) = (2\pi)^{-n/2} \int\limits_A \exp\{-|\mathbf{x}|^2/2\}d\mathbf{x}$.

Exercise 20. *[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 20). It is a simple exercise that $\mu_i = \mathbf{E}[X_i]$ and $\sigma_{i,j} = \text{Cov}(X_i, X_j)$. 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 $\mathbf{v}^t\Sigma\mathbf{v} \geq 0$ 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 21. *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 multi-variate Gaussian. You may prove this using characteristic functions, for example. The characteristic function of Gaussian distribution is given in the exercise below.

Exercise 22. *Irrespective of whether Σ is non-singular or not, the characteristic function of Y is given by*

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

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

Exercise 23. *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).

Appendix B. More about the univariate normal distribution

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

Exercise 24. *For all $x > 0$, we have*

$$\frac{1}{\sqrt{2\pi}}\frac{x}{1+x^2}e^{-\frac{1}{2}x^2} \le \bar{\Phi}(x) \le \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{1}{2}x^2}. \tag{3}$$

In particular[6]*, $\bar{\Phi}(x) \sim x^{-1}\varphi(x)$ as $x \to \infty$. Most often the following simpler bound, valid for $x \ge 1$, suffices.*

$$\frac{1}{10x}e^{-\frac{1}{2}x^2} \le \bar{\Phi}(x) \le e^{-\frac{1}{2}x^2}. \tag{4}$$

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 $M_n \ge t$ if and only if $X_i \ge t$ for some $i \le n$. Therefore,

$$\mathbf{P}\{M_n \ge t\} \le \sum_{k=1}^n \mathbf{P}\{X_k \ge t\} = n\bar{\Phi}(t).$$

Using the upper bound in (4), and setting $t = \sqrt{2A\log n}$ with $A > 1$, we get (since $t \ge 1$ for $n \ge 2$),

$$\mathbf{P}\{M_n \ge \sqrt{2A\log n}\} \le ne^{-A\log n} = \frac{1}{n^{A-1}}. \tag{5}$$

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

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

Exercise 25. *Use the lower bound for the tail of the Normal distribution from (4), show that* $\mathbf{P}\left\{\frac{1}{\sqrt{2\log n}}M_n \leq 1-\delta\right\} \to 0$ *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 $\sigma^2 \geq 0$. If this is not clear, take it as an exercise!

Gaussian density and heat equation: For $t > 0$, let $p_t(x) := \frac{1}{\sqrt{t}}\varphi(x/\sqrt{t})$ 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 26. *Show that* $p_t \star p_s = p_{t+s}$*, i.e.,* $\int\limits_{\mathbb{R}} p_t(x-y)p_s(y)dy = p_{t+s}(x)$.

Exercise 27. *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 28. *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* $f \in L^1(\mathbb{R})$*).*

APPENDIX C. EXISTENCE OF COUNTABLY MANY GAUSSIANS WITH GIVEN COVARIANCES

Let $\Sigma = (\sigma_{i,j})_{i,j\geq 1}$ 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 $\mathbf{E}[X_i] = 0$ and $\mathbf{E}[X_iX_j] = \sigma_{i,j}$ for all $i,j \geq 1$?

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 $\sum_{i,j=1}^n u_iu_j\sigma_{i,j} \geq 0$ for all $n \geq 1$ and all $\mathbf{u} \in \mathbb{R}^n$. Symmetry is clearly necessary. As for the second condition, observe that

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

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

Proof. Let ξ_n, $n \geq 1$, 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 $a_{n,j}$, $1 \leq j \leq n$, will be chosen so as to satisfy the covariance conditions. That X_n, $n \geq 1$, 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 $\sigma_{1,1} \geq 0$.

Next, from $\mathbf{E}[X_1X_2] = \sigma_{1,2}$ 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 $(\sigma_{i,j})_{i,j\leq 2}$ is p.s.d., we certainly have $\sigma_{1,1} \geq 0$ and $\sigma_{2,2}\sigma_{1,1} - \sigma_{1,2}^2 \geq 0$. 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 $a_{k,j}$, $1 \leq j \leq k \leq n-1$. We want to solve for $a_{n,j}$, $j \leq n$. Let us use matrix notation and write $B = (a_{k,j})_{j,k\leq n-1}$ (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 $k \leq n-1$, 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 $X = (\sigma_{i,j})_{i,j\leq n-1}$, $\mathbf{v}$ 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 $\det(\sigma_{i,j})_{i,j\leq n} > 0$ 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...] ■