9 Brownian Motion: Construction

9.1 Definition and Heuristics

The central limit theorem states that the standard Gaussian distribution arises as the weak limit of the rescaled partial sums $S_n/\sqrt{n}$ of independent, identically distributed random variables $X_i = S_i - S_{i-1}$ with mean 0 and variance 1. This has an important corollary: the family $\{\varphi_t\}_{t\geq 0}$ of normal densities is closed under convolution. To see this, observe that for any $0 < t < 1$ the sum $S_n = S_{[nt]} + (S_n - S_{[nt]})$ is obtained by adding two independent sums; the central limit theorem applies to each sum separately, and so by an elementary scaling we must have $\varphi_1 = \varphi_t * \varphi_{1-t}$. More generally, for any $s, t \geq 0$,

$$\varphi_s * \varphi_t = \varphi_{s+t}. \tag{9.1}$$

This law can, of course, be proved without reference to the central limit theorem, either by direct calculation ("completing the square") or by Fourier transform. However, our argument suggests a "dynamical" interpretation of the equation (9.1) that the more direct proofs obscure. For any finite set of times $0 = t_0 < t_1 < \cdots < t_m < \infty$ there exist (on some probability space) independent, mean-zero Gaussian random variables $W_{t_{i+1}} - W_{t_i}$ with variances $t_{i+1} - t_i$. The De Moivre-Laplace theorem implies that as $n \to \infty$,

$$\frac{1}{\sqrt{n}}(S_{[nt_1]}, S_{[nt_2]}, \ldots, S_{[nt_m]}) \xrightarrow{\mathscr{D}} (W_{t_1}, W_{t_2}, \ldots, W_{t_m}). \tag{9.2}$$

The convolution law (9.1) guarantees that the joint distributions of these limiting random vectors are mutually consistent, that is, if the set of times $\{t_i\}_{i\leq m}$ is enlarged by adding more time points, the joint distribution of $W_{t_1}, W_{t_2}, \ldots, W_{t_m}$ will not be changed. This suggests the possibility of defining a continuous-time stochastic process $\{W_t\}_{t\geq 0}$ in which all of the random vectors $(W_{t_1}, W_{t_2}, \ldots, W_{t_m})$ are embedded.

Definition 9.1. A *standard (one-dimensional) Wiener process* (also called *Brownian motion*) is a continuous-time stochastic process $\{W_t\}_{t\geq 0}$ (i.e., a family of real random variables indexed by the set of nonnegative real numbers t) with the following properties:

(A) $W_0 = 0$.
(B) With probability 1, the function $t \to W_t$ is continuous in t.
(C) The process $\{W_t\}_{t\geq 0}$ has *stationary, independent increments.*
(D) For each t the random variable W_t has the NORMAL$(0, t)$ distribution.

A continuous-time stochastic process $\{X_t\}_{t\geq 0}$ is said to have *independent increments* if for all $0 \leq t_0 < t_1 < \cdots < t_m$ the random variables $(X_{t_{i+1}} - X_{t_i})$ are mutually independent; it is said to have *stationary increments* if for any $s, t \geq 0$ the distribution of $X_{t+s} - X_s$ is the same as that of $X_t - X_0$. Processes with stationary, independent increments are known as *Lévy processes.*

Properties (C) and (D) are mutually consistent, by the convolution law (9.1), but it is by no means clear that there exists a stochastic process satisfying (C) and (D) that has continuous sample paths. That such a process *does* exist was first proved by N. WIENER in about 1920; we will give a different proof, due to P. LÉVY.

The convergence (9.2) shows that, in a certain sense, the Wiener process is an infinitesimal form of the simple random walk. This convergence also explains, at least in part, why the Wiener process is useful in the modeling of natural processes. Many stochastic processes behave, at least for long stretches of time, like random walks with small but frequent jumps. The argument above suggests that such processes will look, at least approximately, and on the appropriate time scale, like Brownian motion.

Notation and Terminology. A *Brownian motion with initial point x* is a stochastic process $\{W_t\}_{t\geq 0}$ such that $\{W_t - x\}_{t\geq 0}$ is a standard Brownian motion. Unless otherwise specified, *Brownian motion* means *standard Brownian motion.* To ease eyestrain, we will adopt the convention that whenever convenient the index t will be written as a functional argument instead of as a subscript, that is, $W(t) = W_t$.

9.2 Existence of the Wiener process

9.2.1 Preliminaries: Gaussian random variables

Lemma 9.2. *Almost sure limits of Gaussian random variables are Gaussian, that is, if $Y_1, Y_2, \ldots$ are Gaussian random variables, all defined on a common probability space, and if $Y = \lim_{n\to\infty} Y_n$ exists almost surely, then either Y is a constant random variable or Y is Gaussian.*

Proof. Characteristic functions. (Exercise.) □

Lemma 9.3. *Random variables $Y_1, Y_2, \ldots, Y_m$ defined on a common probability space are independent Gaussian random variables with means $EY_i = 0$ and variances $\sigma_i^2 = EY_i^2$ if and only if for any choice of real scalars $a_1, a_2, \ldots, a_m$ the random variable $S = \sum_{i=1}^m a_i Y_i$ is Gaussian with mean zero and variance $\sum_{i=1}^m a_i^2 \sigma_i^2$.*

Proof. The *only if* implication is easy. To prove the *if* implication, assume that the random vector $Y = (Y_1, Y_2, \ldots, Y_n)$ has the property that its one-dimensional projections $S = \sum_{i=1}^m a_i Y_i$ are mean-zero Gaussians with variances $\sum_{i=1}^m a_i^2 \sigma_i^2$. Let $\varphi(\theta) = Ee^{i\langle\theta, Y\rangle}$ be the characteristic function of the random vector Y; then the characteristic function of the random variable $S = \langle\theta, Y\rangle$ is

$$Ee^{i\beta S} = Ee^{i\beta\langle\theta, Y\rangle} = \varphi(\beta\theta).$$

Since S has a Gaussian distribution with mean zero and variance $\sum_{i=1}^m \theta_i^2\sigma_i^2$, its characteristic function must be

$$Ee^{i\beta S} = \exp\{-\beta^2 \sum_{k=1}^m \theta_k^2\sigma_k^2/2\}.$$

Hence, the characteristic function of Y splits as a product:

$$\varphi(\theta) = \prod_{k=1}^m e^{-\theta_k^2\sigma_k^2/2}.$$

It now follows that the component random variables $Y_1, Y_2, \dots, Y_m$ are independent, by the following lemma. $\square$

Lemma 9.4. *Random variables $Y_1, Y_2, \dots, Y_m$ defined on a common probability space are independent if and only if their joint characteristic function $\varphi(\theta) = Ee^{i\langle\theta,Y\rangle}$ splits as a product, that is,*

$$\varphi(\theta) = Ee^{i\langle\theta,Y\rangle} = \prod_{k=1}^m \psi_k(\theta_k). \tag{9.3}$$

If this equation holds then the factors $\psi_k(\theta_k)$ are the characteristic functions of the component random variables Y_k.

Proof. It is clear that if (9.3) holds then he factors $\psi_k(\theta_k)$ are the characteristic functions of the component random variables Y_k, because the characteristic function of any component Y_k can be recovered by setting $\theta_j = 0$ for all $j \neq k$. Let $u_1, u_2, \dots, u_m : \mathbb{R} \to \mathbb{R}$ be C^∞ functions all with compact support; then by the Fourier inversion formula,

$$\prod_{k=1}^m u_k(y_k) = (2\pi)^{-m} \iint \cdots \int e^{-i\langle\theta,y\rangle} \prod_{k=1}^m \hat{u}_k(\theta_k)\, d\theta_1 d\theta_2 \cdots d\theta_m.$$

Consequently, by Fubini, if the relation (9.3) holds then

$$\begin{aligned} E\prod_{k=1}^m u_k(Y_k) &= (2\pi)^{-m} \iint \cdots \int \prod_{k=1}^m \hat{u}_k(\theta_k)\psi_k(\theta_k)\, d\theta_1 d\theta_2 \cdots d\theta_m \\ &= \prod_{k=1}^m Eu_k(Y_k), \end{aligned}$$

the last by the Parseval relation. It now follows by a routine approximation argument that the identity

$$E\prod_{k=1}^m u_k(Y_k) = \prod_{k=1}^m Eu_k(Y_k)$$

holds for all indicator functions $u_k = \mathbf{1}_{B_k}$, where B_k is a one-dimensional Borel set, and so the independence of the random variables Y_k follows by definition. $\square$

Lemma 9.5. *If Z is a standard normal random variable (i.e., Gaussian with mean* 0 *and variance* 1*) then for every $x > 0$,*

$$P\{Z > x\} \le \frac{2e^{-x^2/2}}{\sqrt{2\pi}x}. \tag{9.4}$$

Proof.

$$P\{Z > x\} = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}\,dy \le \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-xy/2}\,dy = \frac{2e^{-x^2/2}}{\sqrt{2\pi}x}.$$

□

9.2.2 The Wiener Isometry

Behind Wiener's construction lies a simple but cogent observation, that if a Wiener process could be built on a probability space $(\Omega, \mathscr{F}, P)$ then there would be a natural *linear isometry* of the Hilbert space $L^2([0,1],\lambda)$ into the Hilbert space $L^2(P)$. Recall that the inner product on a (real) L^2 space $L^2(\mu)$ is defined by

$$\langle f, g\rangle = \int fg\,d\mu.$$

The inner product of the indicator functions $\mathbf{1}_{[0,t]}$ and $\mathbf{1}_{[0,s]}$ is

$$\langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]}\rangle = \min(s,t);$$

more generally, the inner product of any two indicators $\mathbf{1}_A$ and $\mathbf{1}_B$ is $\lambda(A\cap B)$. If $\{W_t\}_{t\in[0,1]}$ is a Wiener process, then

$$EW_tW_s = \min(s,t); \tag{9.5}$$

therefore, Wiener reasoned, the assignment $\mathbf{1}_{[0,t]} \mapsto W_t$ should extend to a linear isometry I_W, now known as the *Wiener isometry* or *Wiener integral*, of $L^2([0,1],\lambda)$ into $L^2(P)$. The Wiener isometry suggests a natural approach to explicit representations of the Wiener process, via *orthonormal bases.*

Definition 9.6. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space[8] and let $L^2(\mu)$ be the Hilbert space of square-integrable real-valued random variables[9] with inner product

$$\langle f, g\rangle = E_\mu fg.$$

A set of random variables $B = \{u_\lambda\}_{\lambda\in\Lambda}$ is said to be *orthonormal* if for any two elements $u_1, u_2 \in B$,

$$\begin{aligned}\langle u_1, u_2\rangle &= 0 \quad \text{if } u_1 \ne u_2,\\ &= 1 \quad \text{if } u_1 = u_2.\end{aligned}$$

[8] or, more generally, any measure space.

[9] more precisely, the set of *equivalence classes* of square-integrable random variables, where two random variables are considered equivalent if they are equal almost surely.

An orthonormal set B is an *orthonormal basis* if the set of finite linear combinations of elements of B is dense in $L^2(\mu)$.

Proposition 9.7. *Let $B = \{u_i\}_{1\leq i\leq m}$ be a finite orthonormal set in $L^2(\mu)$. Then for any $f \in L^2(\mu)$ the unique element g in the linear span of B that minimizes the L^2-distance $\|f-g\|_2$ is*

$$g = \sum_{i=1}^{m} \langle f, u_i \rangle u_i. \tag{9.6}$$

Proof. Calculus. □

Corollary 9.8. *If $B = \{u_i\}_{i\mathbb{N}}$ is a countable orthonormal basis of $L^2(\mu)$ then for any $f \in L^2(\mu)$,*

$$f = L^2 - \lim_{m\to\infty} \sum_{i=1}^{m} \langle f, u_i \rangle u_i \quad \textit{and} \tag{9.7}$$

$$\|f\|_2^2 = \sum_{n=1}^{\infty} |\|f, u_n\||^2. \tag{9.8}$$

The idea behind Wiener's construction can now be explained. If $\{\psi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $L^2[0,1]$, then $\{I_W(\psi_n)\}_{n\in\mathbb{N}}$ must be an orthonormal set in $L^2(P)$. Since uncorrelated Gaussian random variables are necessarily independent, it follows that the random variables $\xi_n := I_W(\psi_n)$ must be i.i.d. standard normals. Finally, since I_W is a linear isometry, it must map the L^2-series expansion of $\mathbf{1}_{[0,t]}$ in the basis ψ_n to the series expansion of W_t in the basis ξ_n. Thus, with no further work we have the following.

Theorem 9.9. *Assume that the probability space $(\Omega, \mathscr{F}, P)$ supports an infinite sequence ξ_n of independent, identically distributed $N(0,1)$ random variables, and let $\{\psi_n\}_{n\in\mathbb{N}}$ be any orthonormal basis of $L^2[0,1]$. Then for every $t \in [0,1]$ the infinite series*

$$W_t := \sum_{n=1}^{\infty} \xi_n \langle \mathbf{1}_{[0,t]}, \psi_n \rangle \tag{9.9}$$

converges in the L^2-metric, and the resulting stochastic process $\{W_t\}_{t\in[0,1]}$ is a mean-zero Gaussian process whose covariance function satisfies (9.5)*. Consequently, the process $\{W_t\}_{t\in[0,1]}$ satisfies properties (A), (B), and (C) of Definition 9.1.*

Proof. (A) Convergence of the series. For any t the indicator function $\mathbf{1}_{[0,t]}$ is square-integrable, with L^2-norm $t^{1/2}$, so by equation (9.8),

$$t = \sum_{n=1}^{\infty} |\langle \mathbf{1}_{[0,t]}, \psi_n \rangle|^2.$$

Thus, the sequence $\langle \mathbf{1}_{[0,t]}, \psi_n \rangle$ is square-summable. It now follows by Theorem 4.25 that the sequence of random variables

$$\sum_{n=1}^{m} \langle \mathbf{1}_{[0,t]}, \psi_n \rangle \xi_n$$

converges in L^2 and almost surely to a limit random variable W_t, and that the limit random variable has mean $EW_t = 0$ and variance $EW_t^2 = t$. By Lemma 9.2, the distribution of W_t is the Gaussian law with mean 0 and variance t.

(B) Independence of the increments. Any finite linear combination $\sum a_i W_{t_i}$ is the L^2- and a.s. limit of the series

$$\sum_{n=1}^{\infty} \xi_n \langle \sum_i \mathbf{1}_{[0,t]}, \psi_n \rangle.$$

Since each finite partial sum is a finite linear combination of the Gaussian random variables ξ_n, it is itself Gaussian, and so Lemma 9.2 implies that the linear combination $\sum a_i W_{t_i}$ is Gaussian.

Suppose now that $0 = t_0 < t_1 < t_2 < \cdots < t_m = 1$, and consider the joint distribution of the increments $W_{t_i - W_{t_{i-1}}}$. By the previous paragraph, any linear combination of these increments is Gaussian; hence, by Lemma 9.3, it will follow that the increments $W_{t_i - W_{t_{i-1}}}$ are independent Gaussians with mean 0 and variances $t_i - t_{i-1}$ if we show that for any scalars $a_1, a_2, \ldots, a_m$,

$$E\left(\sum_{i=1}^{m} a_i (W_{t_i - W_{t_{i-1}}})\right)^2 = \sum_{i=1}^{m} a_i^2 (t_i - t_{i-1}).$$

But this follows directly from the fact that the mapping

$$\mathbf{1}_{0,t} \mapsto W_t$$

is an L^2-isometry.

□

Because the convergence is in the L^2-metric, rather than the sup-norm, there is no way to conclude directly that the process so constructed has a version with continuous paths. Wiener was able to show by brute force that for the particular basis

$$\psi_n(x) = \sqrt{2} \cos \pi n x$$

the series (9.9) converges (along an appropriate subsequence) not only in L^2 but also uniformly in t, and therefore gives a version of the Wiener process with continuous paths:

$$W_t = \xi_0 t + \sum_{k=1}^{\infty} \sum_{n=2^{k-1}}^{2^k - 1} n^{-1} \xi_n \sqrt{2} \sin \pi n t. \tag{9.10}$$

The argument for the uniform convergence of the series is somewhat technical, though, and moreover, it is in many ways unnatural. Thus, rather than following Wiener's construction, we will describe an alternative construction, the due to Lévy.

9.2.3 Lévy's argument

Paul Lévy realized that Wiener's construction could be greatly simplified by using instead the simplest "wavelet" basis, consisting of the *Haar functions.*

Definition 9.10. The Haar functions $\psi_{n,k}$ are defined as follows. First, let $\psi : \mathbb{R} \to \{-1,1\}$ be the function

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \le t \le \frac{1}{2}; \\ -1 & \text{if } \frac{1}{2} < t \le 1; \\ 0 & \text{otherwise.} \end{cases} \tag{9.11}$$

Then for any integers $n \ge 0$ and $0 \le k < 2^n$ define the (n,k)th Haar function by

$$\psi_{n,k}(t) = 2^{n/2}\psi(2^n t - k). \tag{9.12}$$

Proposition 9.11. *The Haar functions $\psi_{n,k}$ form an orthonormal basis of $L^2(\lambda)$, where λ is Lebesgue measure on $[0,1]$.*

Proof. Exercise. □

Theorem 9.12. *(Lévy) If the random variables $\xi_{m,k}$ are independent, identically distributed with common distribution $N(0,1)$, then with probability one, the infinite series*

$$W(t) := \xi_{0,1}\, t + \sum_{m=1}^{\infty} \sum_{k=0}^{2^m-1} \xi_{m,k} G_{m,k}(t) \tag{9.13}$$

converges uniformly for $0 \le t \le 1$ and the limit function $W(t)$ is a standard Wiener process.

For any $t \in [0,1]$ the indicator function $\mathbf{1}_{[0,t]}$ is an element of $L^2(\lambda)$, and so by Corollary 9.8 has the representation

$$\mathbf{1}_{[0,t]} = \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \langle \mathbf{1}_{[0,t]}, \psi_{n,k} \rangle \psi_{n,k}. \tag{9.14}$$

The functions

$$G_{n,k}(t) = \langle \mathbf{1}_{[0,t]}, \psi_{n,k} \rangle = \int_0^t \psi_{n,k}(s)\, ds \tag{9.15}$$

aer called the *Schauder functions.* For any $n \ge 1$, the graph of $G_{n,k}$ is a "hat" (isosceles triangle) of height $2^{-n/2}$ and base $[k/2^n, (k+1)/2^n]$. The function $G_{0,0}$ is just $G_{0,0}(t) = t$.

Proof of Theorem 9.12. By definition of the Schauder functions $G_{n,k}$, the series (9.13) is a particular case of (9.9), so the random variables $W(t)$ defined by (9.13) are centered Gaussian with covariances that agree with the covariances of a Wiener process. Hence, to prove that (9.13) defines a Brownian motion, it suffices to prove that with probability one the series converges *uniformly* for $t \in [0,1]$.

The Schauder function $G_{m,k}$ has maximum value $2^{-m/2}$, so to prove that the series (9.13) converges uniformly it is enough to show that

$$\sum_{m=1}^{\infty} \sum_{k=1}^{2^m} |\xi_{m,k}|/2^{m/2} < \infty$$

with probability 1. To do this we will use the Borel-Cantelli Lemma and the tail estimate of Lemma 9.5 for the normal distribution to show that with probability one there is a (possibly random) m_* such that

$$\max_k |\xi_{m,k}| \leq 2^{m/4} \quad \text{for all } m \geq m_*. \tag{9.16}$$

This will imply that almost surely the series is eventually dominated by a multiple of the geometric series $\sum 2^{-(m+2)/4}$, and consequently converges uniformly in t.

To prove that (9.16) holds eventually, it suffices (by Borel-Cantelli) to show that the probabilities of the complementary events are summable. By Lemma 9.5,

$$P\{|\xi_{m,k}| \geq 2^{m/4}\} \leq \frac{4}{2^{m/4}\sqrt{2\pi}} e^{-2^{m/2}}.$$

Hence, by the Bonferroni inequality (i.e., the crude union bound),

$$P\{\max_{1\leq k\leq 2^m} |\xi_{m,k}| \geq 2^{m/4}\} \leq 2^m 2^{-m/4}\sqrt{2/\pi} e^{-2^{m-1}}.$$

Since this bound is summable in m, Borel-Cantelli implies that with probability 1, eventually (9.16) must hold. This proves that w.p.1 the series (9.13) converges uniformly, and therefore $W(t)$ is continuous.

□