GAUSSIAN PROCESSES; KOLMOGOROV-CHENTSOV THEOREM

STEVEN P. LALLEY

1. Gaussian Processes: Definitions and Examples

Definition 1.1. A *standard (one-dimensional) Wiener process* (also called *Brownian motion*) is a stochastic process $\{W_t\}_{t \geq 0+}$ indexed by nonnegative real numbers t with the following properties:

(1) $W_0 = 0$.
(2) With probability 1, the function $t \to W_t$ is continuous in t.
(3) The process $\{W_t\}_{t \geq 0}$ has *stationary, independent increments*.
(4) The increment $W_{t+s} - W_s$ has the $\text{Normal}(0, t)$ distribution.

The term *independent increments* means that for every choice of nonnegative real numbers $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_n < t_n < \infty$, the *increment* random variables

$$W_{t_1} - W_{s_1}, W_{t_2} - W_{s_2}, \ldots, W_{t_n} - W_{s_n}$$

are jointly independent; the term *stationary increments* means that for any $0 < s, t < \infty$ the distribution of the increment $W_{t+s} - W_s$ has the same distribution as $W_t - W_0 = W_t$. In general, a stochastic process with stationary, independent increments is called a *Lévy process*; more on these later. The Wiener process is the intersection of the class of *Gaussian processes* with the *Lévy processes.* Using only elementary properties of the normal (Gaussian) distributions you should be able to verify the following in 30 seconds or less:

Proposition 1.1. *Let $\{W(t)\}_{t \geq 0}$ be a standard Brownian motion. Then each of the following processes is also a standard Brownian motion:*

$$\{-W(t)\}_{t \geq 0} \tag{1.1}$$

$$\{W(t+s) - W(s)\}_{t \geq 0} \tag{1.2}$$

$$\{aW(t/a^2)\}_{t \geq 0} \tag{1.3}$$

$$\{tW(1/t)\}_{t \geq 0}. \tag{1.4}$$

The first three of these, although elementary, are of crucial importance in stochastic calculus. In particular, property (1.3), the *Brownian scaling* law, is what accounts for so much of the strangeness in Brownian paths. More on this later. For now the issue is fundamental: does there exist a stochastic process that satisfies the conditions of Definition 1.1? That is, are the requirements on the distributions of the increments compatible with path-continuity? We will approach this by asking the more general question: which Gaussian processes have versions with continuous paths?

Definition 1.2. A *Gaussian process* $\{X_t\}_{t \in T}$ indexed by a set T is a family of (real-valued) random variables X_t, all defined on the same probability space, such that for any finite subset $F \subset T$ the random vector $X_F := \{X_t\}_{t \in F}$ has a (possibly degenerate) Gaussian distribution. Equivalently, $\{X_t\}_{t \in T}$ is Gaussian if every finite linear combination $\sum_{t \in F} a_t X_t$ is either identically zero or has a Gaussian distribution on $\mathbf{R}$. The *covariance function* of a Gaussian process $\{X_t\}_{t \in T}$ is the bivariate function

$$R(s,t) = \text{cov}(X_s, X_t) = E(X_s - EX_s)(X_t - EX_t). \tag{1.5}$$

Covariance Functions: You should recall that the covariance matrix of a multivariate Gaussian random vector is a symmetric, nonnegative definite matrix; the distribution is said to be *nondegenerate* if its covariance matrix is strictly positive definite. The mean vector and covariance matrix uniquely determine a Gaussian distribution; consequently, the mean function and covariance function of a Gaussian process completely determine all of the *finite-dimensional distributions* (that is, the joint distributions of finite subsets X_F of the random variables). Thus, if two Gaussian processes $\mathbf{X} = \{X_t\}_{t \in T}$ and $\mathbf{Y} = \{Y_t\}_{t \in T}$ have the same mean and covariance functions, then for any event B that depends on only finitely many coordinates,

$$P\{\mathbf{X} \in B\} = P\{\mathbf{Y} \in B\}. \tag{1.6}$$

Since any event can be arbitrarily well-approximated by events that depend on only finitely many coordinates, it follows that the equality (1.6) holds for *all* events B. Therefore, the processes $\mathbf{X}$ and $\mathbf{Y}$ are identical in law.

Gaussian processes: Examples

Example 1.1. The most important one-parameter Gaussian processes are the *Wiener process* $\{W_t\}_{t \geq 0}$ (Brownian motion), the *Ornstein-Uhlenbeck* process $\{Y_t\}_{t \in \mathbf{R}}$, and the *Brownian bridge* $\{W_t^\circ\}_{t \in [0,1]}$. These are the mean-zero processes with covariance functions

$$EW_sW_t = \min(s,t), \tag{1.7}$$

$$EY_sY_t = \exp\{-|t-s|\}, \tag{1.8}$$

$$EW_t^\circ W_s^\circ = \min(s,t) - st. \tag{1.9}$$

Note: In certain situations we truncate the parameter space T – in particular, sometimes we are interested in the Wiener process W_t only for $t \in [0,1]$, or in the Ornstein-Uhlenbeck process Y_t for $t \geq 0$.

Exercise 1.1. Check that if W_t is a standard Wiener process, then the derived processes

$$W_t^\circ := W_t - tW_1 \quad \text{and} \quad Y_t := e^{-t} W_{e^{2t}}$$

have the same covariance functions as given above, and so these derived processes have the same "finite-dimensional distributions" as the Brownian bridge and Ornstein-Uhlenbeck process, respectively. Also, check that for any scalar $\alpha > 0$ the process

$$\tilde{W}_t := \alpha^{-1} W_{\alpha^2 t}$$

has the same covariance function, and therefore also the same finite-dimensional distributions, as W_t. (This correspondence is called *Brownian scaling*.)

Exercise 1.2. Let W_t be a standard Wiener process, and let $f(t)$ be any continuous (nonrandom) function. Define

$$Z_t = \int_0^t W_s f(s)\, ds$$

(The integral is well-defined because the Wiener process has continuous paths.) Show that Z_t is a Gaussian process, and calculate its covariance function. HINT: First show that if a sequence X_n of Gaussian random variables converges in distribution, then the limit distribution is Gaussian (but possibly degenerate).

Example 1.2. Let $\xi_1, \xi_2, \dots$ be independent, identically distributed unit normals. Then for any finite set of frequencies $\omega_i \geq 0$, the process

$$X_t := \sum_{i=1}^{m} \xi_i \cos(\omega_i t) \tag{1.10}$$

indexed by $t \in \mathbf{R}$ is a Gaussian process. This process has smooth sample paths (they are just random linear combinations of cosine waves). Note that for any finite set F of cardinality larger than m the random vector X_F has a degenerate Gaussian distribution (why?).

Example 1.3. The two-parameter *Brownian sheet* $\{W_s\}_{s \in \mathbf{R}_+^2}$ is the mean-zero Gaussian process indexed by ordered pairs $s = (s_1, s_2)$ of nonnegative reals with covariance function

$$EW_s W_t = \min(s_1, t_1) \min(s_2, t_2). \tag{1.11}$$

Observe that for each fixed $r > 0$, the *one*-parameter process $Z_s^r := W_{s,r}$ has the same covariance function as a standard Wiener process multiplied by $\sqrt{r}$. Thus, the Brownian sheet has slices in the two coordinate directions that look like scaled Wiener processes. For figures showing simulations of Brownian sheets, see Mandelbrot's book *Fractal Geometry of Nature*.

Example 1.4. The *fractional Brownian motion* with *Hurst parameter* $H \in (0, 1)$ is the mean-zero Gaussian process $\{X_t^H\}_{t \geq 0}$ with covariance function

$$EX_t^H X_s^H = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}), \tag{1.12}$$

equivalently,

$$E|X_t^H - X_s^H|^2 = |t - s|^{2H}. \tag{1.13}$$

The case $H = 1/2$ is just standard Brownian motion. When $H < 1/2$ the increments of X^H are *negatively* correlated; for $H > 1/2$ they are *positively* correlated. As for Brownian motion, increments are *stationary*; and as for Brownian motion, there is a scaling law

$$\{a^{-H} X^H(at)\}_{t \geq 0} \overset{\mathcal{D}}{=} \{X^H(t)\}. \tag{1.14}$$

It is not immediately obvious that the covariance kernel (1.12) is nonnegative definite. However, this will follow from an explicit construction of X_t^H from standard Brownian motion given later (see Exercise 3.4 in section 3 below).

Example 1.5. The *discrete Gaussian free field* is a mean-zero Gaussian process $\{X_v\}_{v \in V}$ indexed by the vertices v of a (let's say) finite, connected graph $\mathcal{G} = (V, \mathcal{E})$. The covariance function (which in this case can be viewed as a symmetric matrix $\Sigma = (r_{v,w})_{v,w \in V}$) is the inverse of the *Dirichlet form* associated with the graph; thus, the joint distribution of the random variables X_v is the multivariate normal distribution with density proportional to

$$\exp\{-H(x)\}$$

where

$$H(x) = \frac{1}{2} \sum_{(v,w):vw \in \mathcal{E}} (x_v - x_w)^2$$

with the sum being over all pairs of vertices v, w such that vw is an edge of the graph. The Gaussian free field and its continuum limit are basic objects of study in *quantum field theory*.

Distributions and Conditional Distributions of Random Processes

The Space $C(J)$. For any compact interval $J \subseteq \mathbb{R}$ (or more generally any compact metric space J), the space $C(J)$ of continuous, real-valued functions on J is a (complete, separable) metric space relative to the sup-norm distance:

$$d(f, g) := \|f - g\|_\infty := \max_{t \in J} |f(t) - g(t)|. \tag{1.15}$$

The *Borel $\sigma-$algebra* on $C(J)$ is the smallest $\sigma-$algebra containing all open sets. (More generally, the Borel $\sigma-$algebra on an arbitrary metric space $\mathcal{X}$ is the smallest $\sigma-$algebra containing all open sets.) The Borel $\sigma-$algebra on $C(J)$ is generated by the *cylinder sets*, that is, events of the form

$$\{x \in C(J) \,:\, x(t_i) \in A_i \; \forall \, 1 \leq i \leq k\} \tag{1.16}$$

where each A_i is an open interval of $\mathbb{R}$. (Exercise: check this.) This is quite useful to know, because it means that a Borel probability measure on $C(J)$ is uniquely determined by its *finite-dimensional distributions*, that is, its values on cylinder sets.

Distributions of Stochastic Processes with Continuous Paths. The *distribution* of a stochastic process $\{X_t\}_{t \in J}$ with continuous paths is the (Borel) probability measure μ on the space $C(J)$ of continuous functions indexed by $t \in J$ defined by

$$\mu(B) := P\{\{X_t\}_{t \in J} \in B\} \tag{1.17}$$

where B is a Borel subset of $C(J)$. To show that two stochastic processes $\{X(t)\}_{t \in J}$ and $\{Y(t)\}_{t \in J}$ with continuous paths have the same distribution, it suffices to check that they have the same *finite-dimensional distributions*, that is, that

$$P\{X_{t_i} \in A_i \; \forall \, 1 \leq i \leq k\} = P\{Y_{t_i} \in A_i \; \forall \, 1 \leq i \leq k\} \tag{1.18}$$

for all choices of t_i and A_i. Consequently, to check that two *Gaussian* processes with continuous paths have the distribution, it suffices to check that they have the same mean and covariance functions.

Regular Conditional Distributions. Let $\{X_t = X(t)\}_{t \in J}$ be a stochastic process with continuous paths. For any sub-$\sigma-$algebra $\mathcal{G}$ of the underlying probability space $(\Omega, \mathcal{F}, P)$, the conditional distribution of the stochastic process $\{X(t)\}_{t \in J}$ given $\mathcal{G}$ is defined in the obvious way:

$$P(\{X_t\}_{t \in J} \in B \,|\, \mathcal{G}). \tag{1.19}$$

It is a nontrivial *theorem* (see Wichura notes) that there is always a *regular* conditional distribution, that is, a version of the conditional distribution (1.19) such that for every $\omega \in \Omega$,

$$P(\{X_t\}_{t \in J} \in \cdot \,|\, \mathcal{G})(\omega)$$

is a Borel probability measure on $C(J)$. Henceforth, assume that conditional distributions of stochastic processes with continuous paths are *regular*. This makes it possible to talk about conditional distributions of stochastic processes given events of probability zero like $W(1) = x$.

Exercise 1.3. Let W_t be a standard Wiener process, and let $\mathcal{G}$ be the $\sigma-$algebra of events generated by the random variable W_1 (that is, the smallest $\sigma-$algebra containing all events $\{W_1 \in B\}$, where B is a Borel set). Show that the conditional distribution of $\{W_t\}_{0\leq t\leq 1}$ given $\mathcal{G}$ is the same as the distribution of the process

$$W_t^\circ + \xi t$$

where W° is a standard Brownian bridge and ξ is an independent standard normal random variable.

HINT: First prove the following general lemma: If $\{X_t\}_{t\in T}$ is a Gaussian process then for any finite set $\{t_j\}_{1\leq j\leq k} \subset T$ of indices and any matching set $\{y_j\}_{1\leq j\leq k}$ of real numbers, the *conditional* joint distribution of $\{X_t\}_{t\in T}$ *given* the values $X_{t_j} = y_j$ is itself the law of a Gaussian process.

Construction of Gaussian Processes. It is not at all obvious that the Gaussian processes in Examples 1.1 and 1.3 exist, nor what kind of sample paths/sheets they will have. The difficulty is that uncountably many random variables are involved. We will show that not only do all of the processes above exist, but that they have continuous sample functions. This will be done in two steps: First, we will show that Gaussian processes with *countable* index sets can always be constructed from i.i.d. unit normals. Then, in section 2, we will show that under certain restrictions on the covariance function a Gaussian process can be extended continuously from a countable dense index set to a continuum. The following example shows that some restriction on the covariance is necessary.

Exercise 1.4. Show that there is no Gaussian process $\{X_t\}_{t\in[0,1]}$ with continuous sample paths and covariance function

$$\begin{aligned} R(s,t) &= 0 \quad \text{for } s \neq t \quad \text{and} \\ R(s,s) &= 1. \end{aligned}$$

Processes with Countable Index Sets. For each $m = 1, 2, ..$, let F_m be a (Borel) probability distribution on $\mathbb{R}^m$. Assume that these are mutually *consistent* in the following sense: for each Borel subset B of $\mathbb{R}^m$,

$$F_m(B) = F_{m+1}(B \times \mathbb{R}); \tag{1.20}$$

that is, F_m is the marginal joint distribution of the first m coordinates of a random vector with distribution F_{m+1}. I will show that on some probability space are defined random variables X_i such that for each m, the random vector $(X_1, X_2, \ldots, X_m)$ has distribution F_{m+1}. In fact, any probability space that support an i.i.d. sequence U_i of uniform-[0,1] random variables will suffice.

Recall that any probability distribution F on $\mathbb{R}$ can be "simulated" using a single uniform-[0,1] random variable U, by the *quantile transform* method. Hence, there is a (Borel) function $\varphi_1(U_1) := X_1$ that has distribution F_1. Now suppose that X_i have been constructed for

$i \leq m$ using the values U_i, with $i \leq m$, in such a way that the joint distribution of the random vector $(X_i)_{i \leq m}$ is F_m. Let G_{m+1} be the conditional distribution on $\mathbb{R}^1$ of the $(m+1)$st coordinate of an $F_{m+1}-$ distributed random vector given that the first m coordinates have values $(X_i)_{i \leq m}$. Use U_{m+1} to produce a random variable X_{m+1} with conditional distribution G_{m+1} given the values $(X_i)_{i \leq m}$. Then the joint distribution of $(X_i)_{i \leq m+1}$ will be F_{m+1}, by the consistency hypothesis.

Now let $R(s,t)$ be a *positive definite function* indexed by a countable set T, that is, a symmetric function with the property that for every finite subset F of T, the matrix $\Sigma_F := (R(s,t))_{s,t \in F}$ is positive definite. Without loss of generality, assume that $T = \mathbb{N}$. Then the sequence of distributions

$$F_m = \text{Normal}(0, \Sigma_m)$$

is mutually consistent (why?). Therefore, by the preceding paragraph, there exists a sequence of random variables X_i such that, for each finite subset $F \subset T$ the joint distribution of X_F is Gaussian with mean zero and covariance Σ_F. $\square$

2. Kolmogorov-Chentsov Theorem

2.1. **Continuity of Sample Paths.** The Kolmogorov-Chentsov theorem provides a useful criterion for establishing the existence of versions of stochastic processes with continuous sample paths. It is not limited to Gaussian processes, nor is it limited to stochastic processes indexed by $t \in [0, \infty)$; it applies also to processes indexed by parameters that take values in a subset of a higher-dimensional Euclidean space. Such processes are called *random fields*. More generally, a *random field* is a family $\mathbf{X} = \{X_t\}_{t \in T}$ of real random variables indexed by the points $t \in T$ of a subset T of a topological space V.

Theorem 1. *(Kolmogorov-Chentsov) Let $\mathbf{X}$ be a random field whose index set T is a dense set of points t in an open domain $D \subseteq \mathbb{R}^d$. Suppose that there are positive constants α, β, C such that*

$$E|X_t - X_s|^\alpha \leq C|t-s|^{d+\beta} \quad \textit{for all } s,t \in T. \tag{2.1}$$

Then the random field can be extended to a random field $\{\tilde{X}_t\}_{t \in \bar{D}}$ indexed by the closure $\bar{D}$ of D in such a way that, with probability one, the mapping $t \mapsto \tilde{X}_t$ is continuous, and so that for every $t \in T$,

$$X_t = \tilde{X}_t \quad \textit{almost surely.} \tag{2.2}$$

Moreover, if $\gamma < \beta/\alpha$ then for every compact subset $K \subset \bar{D}$,

$$\max_{s \neq t \in K} \frac{|X_t - X_s|}{|t-s|^\gamma} < \infty \quad \textit{almost surely.} \tag{2.3}$$

Remark 1. The inequality (2.3) means that the sample functions X_t are *Hölder continuous* with Hölder exponent γ.

Remark 2. The hypothesis (2.1) means that if $t, s \in T$ are close, then the random variables X_t and X_s are close in the L^α norm. In particular, it follows that if $s_n \to t$ in T then

$$X_{s_n} \xrightarrow{P} X_t. \tag{2.4}$$

This by itself is not enough to imply that X_t can be extended continuously to $\bar{D}$, as the following example shows.

Example 2.1. Consider the standard Poisson counting process $\{N_t\}_{t\geq 0}$: this certainly does not have continuous paths, even though (2.4) holds. This shows that the moment hypothesis (2.1) requires an exponent $d+\beta$ strictly larger than d, at least when $d=1$, because for every $\alpha = 1, 2, \dots,$

$$E(N_{t+s} - N_t)^\alpha = C_\alpha s. \tag{2.5}$$

Exercise 2.1. Find similar examples to show that the moment hypothesis requires an exponent $d+\beta$ strictly larger than d in any dimension d.

Example 2.2. Next, consider the Wiener process W_t: the increment $W_t - W_s$ is normally distributed with mean 0 and variance $|t-s|$, equivalently, $W_t - W_s$ has the same distribution as $|t-s|^{1/2}Z$, where Z is standard normal. Hence,

$$E|W_t - W_s|^{2k} = C_k|t-s|^k \tag{2.6}$$

where C_k is the $2k$th moment of a standard normal. Thus, the Kolmogorov-Chentsov condition (2.1) holds with $\alpha = 2k$ and $\beta = k-1$ for all $k \geq 1$, and so Theorem 1 implies that there exists a version of the Wiener process with continuous sample paths. Moreover, since $(k-1)/2k \to 1/2$ as $k \to \infty$, it follows from (2.3) that for every $\gamma < 1/2$,

$$\max_{t\neq s\in[0,1]} \frac{|W_t - W_s|}{|t-s|^\gamma} < \infty \tag{2.7}$$

almost surely. (Note: Lévy proved a considerably sharper result, called *Lévy's modulus*. This states that the denominator $|t-s|^\gamma$ can be replaced by $\sqrt{|t-s|\log(1/|t-s|)}$.)

Exercise 2.2. Use the Kolmogorov-Chentsov criterion to prove that there is a continuous extension of the Brownian sheet to the parameter space $t \in [0,1]^2$.

Exercise 2.3. More generally, let $R(s,t)$ be any covariance function defined for s,t in an open subset of $\mathbb{R}^d$ that satisfies

$$R(s,s) + R(t,t) - 2R(s,t) \leq C|t-s|^\gamma \tag{2.8}$$

for some $\gamma > 0$. Show that there is a mean-zero Gaussian process X_t with covariance function $R(s,t)$ that has continuous sample functions almost surely.

2.2. **Differentiablity of Gaussian Random Fields.** Gaussian processes may or may not have differentiable sample paths. For instance, the Wiener process does not, but the integrated Wiener processes of Exercise 1.2 do, by the fundamental theorem of calculus. In general, the degree of smoothness of a Gaussian process is determined by the smoothness of its covariance function near the diagonal. I will not try to prove this in any generality, but will discuss the case of Gaussian processes with a one-dimensional parameter $t \in \mathbb{R}$.

Proposition 2.1. *Let $\{X(t)\}_{t\in\mathbb{R}}$ be a mean-zero Gaussian process with C^1 sample paths and covariance function $R(s,t)$. Then the derivative process $X'(t)$ is mean-zero Gaussian, with continuous covariance function $\tilde{R}(s,t) := \partial_s\partial_t R(s,t)$.*

Proof. First, if $X(t)$ is a continuous mean-zero Gaussian process, then its covariance function $EX(t)X(s)$ must be jointly continuous in s,t, because the collection $\{X(t)X(s)\}_{s,t}$ is uniformly integrable (why?). Now define

$$\begin{aligned} DX(t,\varepsilon) :&= \frac{X(t+\varepsilon)-X(t)}{\varepsilon} \quad \text{for } \varepsilon \neq 0, \\ &= X'(t) \quad \text{for } \varepsilon = 0. \end{aligned}$$

Since $X(t)$ is continuously differentiable, $DX(t,\varepsilon)$ is a continuous two-parameter process. I claim that this two-parameter process is Gaussian. To see this, first observe that any linear combination of random variables $DX(t,\varepsilon)$ with $\varepsilon \neq 0$ is Gaussian, because such a linear combination is nothing more than a linear combination of random variables $X(s)$, and the process $X(t)$ was assumed Gaussian. But $DX(t,0)$ is the limit of random variables $DX(t,\varepsilon)$, so a linear combination of $DX(t,\varepsilon)$ possibly including terms with $\varepsilon = 0$ is still Gaussian. (Exercise: Explain this.) Hence, the entire two-parameter process $DX(t,\varepsilon)$ is Gaussian, and so in particular the restriction $DX(t,0) = X'(t)$ is Gaussian. Finally, consider the covariance function. It is clear that for $\varepsilon \neq 0$,

$$\operatorname{cov}(DX(s,\varepsilon), DX(t,\varepsilon)) = \left(R(t+\varepsilon, s+\varepsilon) - R(t, s+\varepsilon) - R(t+\varepsilon, s) + R(t,s)\right)/\varepsilon^2.$$

Taking the limit as $\varepsilon \to 0$, you get $\partial_s \partial_t R(s,t)$ as the covariance function of the derivative process $X'(t)$. □

There is a partial converse, which you might try to prove as an *exercise*:

Proposition 2.2. *Let $R(s,t)$ be a positive definite function with continuous mixed partial $\tilde{R}(s,t) := \partial_s \partial_t R(s,t)$. Suppose also that $\tilde{R}(s,t)$ satisfies inequality* (2.8) *above. Then there is a* continuously differentiable, *mean-zero Gaussian process $X(t)$ with covariance $R(s,t)$.*

2.3. **Proof of the Kolmogorov-Chentsov Theorem.** For simplicity I will consider the special case where $D = (0,1)^d$ is the unit cube in d dimensions and T is the set of dyadic rationals in D. The general case can be proved by the same argument, with some obvious modifications. The basic tool is the following purely analytic fact:

Lemma 2.3. *Let $x(t)$ be a real-valued function defined on the set T of dyadic rationals in D. Write $T = \cup_{m=1}^{\infty} L_m$ where L_m is the set of points whose coordinates are of the form $k/2^m$ for $0 \leq k \leq 2^m$. Suppose that for some $\gamma > 0$ and $C < \infty$,*

$$|x(s) - x(t)| \leq C|t-s|^{\gamma} \tag{2.9}$$

for all neighboring[1] *pairs s,t in the same dyadic level L_m. Then $x(t)$ extends to a continuous function on the closed unit cube, and the extension is Hölder of exponent γ, that is, for some $C' < \infty$,*

$$|x(s) - x(t)| \leq C'|t-s|^{\gamma}. \tag{2.10}$$

Proof. This is based on a technique called *chaining*: the idea is that each point $t \in \bar{D}$ can be reached via a sequence ("chain") of points $s_m(t)$ of points in $\cup_m L_m$. For each point $t \in \bar{D}$, there is a point of L_m within distance $C_d 2^{-m}$ of t, with $C_d = 2^{d/2}$ (by Pythagoras' theorem). Choose such a point and label it $s_m(t)$. Clearly, the sequence $s_m(t)$ converges to t. In fact, the distance between successive points $s_m(t)$ and $s_{m+1}(t)$ is less than $2C_d 2^{-m}$, and so there

[1] Two points $s,t \in L_m$ are *neighbors* if $2^m s$ and $2^m t$ are neighbors in the $d-$dimensional integer lattice, that is, if $2^m s$ and $2^m t$ differ by a unit vector.

is a short chain of neighboring pairs (no more than $2d$ steps) in level L_{m+1} connecting $s_m(t)$ to $s_{m+1}(t)$. Consequently, by the hypothesis (2.9),

$$|x(s_m(t)) - x(s_{m+1}(t))| \leq 2dC2^{-\gamma m}.$$

Since $\gamma > 0$ this is summable in m, and so for any t the sequence $x(s_n(t))$ is Cauchy. Hence, it converges, and so we can set

$$x(t) := \lim_{m\to\infty} x(s_m(t)).$$

Note that if $t \in T$ then the limit agrees with the original value $x(t)$ (why?). Note also that $|x(t) - x(s_m(t))| \leq C''2^{-m}$ for a suitable constant C'', because the "links" of the chain from $s_m(t)$ on are dominated by the terms of a geometric series with ratio $2^{-\gamma}$.

It remains to prove that the function $x(t)$ so defined is continuous, and that it satisfies the Hölder condition (2.10). Choose distinct points $s, t \in \bar{D}$; then for some m,

$$2^{-m-1} \leq |s - t| < 2^{-m}.$$

It follows that the distance between $s_m(s)$ and $s_m(t)$ is less than $C'''2^{-m}$, for a suitable constant C''' (I think $C''' = 6C_d$ will do). Thus, there is a short chain of neighboring pairs in level L_m connecting $s_m(s)$ and $s_m(t)$, and so for some $C'''' < \infty$,

$$|x(s_m(s)) - x(s_m(t))| \leq C''''2^{-m\gamma}.$$

Since $|x(u) - x(s_m(u))| \leq C''2^{-m}$ for all u, by construction, inequality (2.10) follows from the triangle inequality. □

Proof of Theorem 1. The crucial geometric fact is that the cardinality of L_m grows like

$$\#L_m \sim 2^{md}.$$

(It's not quite equal, because D is the *open* unit cube. What's important is the exponential growth rate, not the exact count.) Two points $s, t \in L_m$ are *neighbors* if $t - s$ is a unit vector times 2^{-m}. If s, t are neighbors in L_m, then by the hypothesis (2.1) and the Markov inequality,

$$\begin{aligned} P\{|X_t - X_s| \geq |t - s|^\gamma\} &\leq \frac{E|X_t - X_s|^\alpha}{|t - s|^{\gamma\alpha}} \\ &\leq C|t - s|^{d+\beta-\gamma\alpha} \\ &= C2^{-md-m\beta+m\gamma\alpha}. \end{aligned}$$

Since the number of neighboring pairs in L_m grows like a constant multiple of 2^{md}, it follows that

$$P(B_m) \leq C'2^{-m\beta+m\gamma\alpha} \quad \text{where} \tag{2.11}$$
$$B_m := \{|X_t - X_s| \geq |t - s|^\gamma \text{ for some neighboring } s, t \in L_m\}.$$

Since $\beta > 0$, it is possible to choose $\gamma > 0$ so small that $-\beta + \gamma\alpha < 0$. For such a choice, the probabilities (2.11) are summable in m, and so the Borel-Cantelli lemma guarantees that with probability one, the event B_m occurs for only finitely many m. Consequently, with probability one there exists a (random) $\xi < \infty$ such that

$$|X_t - X_s| \leq \xi|t - s|^\gamma$$

for all neighboring points t, s in the same dyadic level. The result now follows from Lemma 2.3. $\square$

3. Wiener's Perspective: The Wiener Integral

The existence of Brownian motion was proved by Wiener around 1920. His proof is far from the simplest of those now known – in fact, it is quite complicated. But behind it lies an extremely useful insight, that Hilbert spaces of Gaussian random variables are naturally isomorphic to Hilbert spaces of functions, and that the isomorphism gives a way of constructing Gaussian processes with specified covariance functions. Let's focus first on Brownian motion.

Theorem 2. *Let $W(t) = W_t$ be a standard Wiener process defined on a probability space $(\Omega, \mathcal{F}, P)$. Then for any nonempty interval $J \subseteq \mathbb{R}_+$ the mapping $\mathbf{1}_{(s,t]} \mapsto W_t - W_s$ extends to a linear isometry $I_W : L^2(J) \to L^2(\Omega, \mathcal{F}, P)$. For every function $\varphi \in L^2(J)$, the random variable $I_W(\varphi)$ is mean-zero Gaussian.*

Remark 3. The mapping I_W is called the *Wiener isometry* or, more commonly, the *Wiener integral*. The notation

$$I_W(\varphi) = \int \varphi \, dW \tag{3.1}$$

is often used. The integral is defined only for *nonrandom* integrands φ, and in particular, only those functions φ that are square-integrable against Lebesgue measure. K. Itô later extended Wiener's integral so as to allow *random* integrands – more on this later.

Proof. First, check, by direct calculation, that for any two intervals $A = (s, t]$ and $B = (u, v]$ the covariance of $I_W(\mathbf{1}_A)$ and $I_W(\mathbf{1}_B)$ equals the inner product $m(A \triangle B)$ of $\mathbf{1}_A$ and $\mathbf{1}_B$ in $L^2(J)$. It then follows, by linearity of the inner product and covariance operators, that

$$\operatorname{cov}(I_W(\varphi), I_W(\psi)) = \int_J \varphi\psi \, dm$$

for all finite linear combinations φ, ψ of interval indicators. Also, for every such finite linear combination φ, the random variable $I_W(\varphi)$ is mean-zero Gaussian, because it is a finite linear combination of random variables W_t.

The rest is a straightforward use of standard results in Hilbert space theory. Let H_0 be the set of all finite linear combinations of interval indicator functions $\mathbf{1}_A$. Then H_0 is a dense, linear subspace of $L^2(J)$, that is, every function $f \in L^2(J)$ can be approximated arbitrarily closely in the L^2-metric by elements of H_0. Since I_W is a linear isometry of H_0, it extends uniquely to a linear isometry of $L^2(J)$, by Proposition 3.1 below. Furthermore, for any $\varphi \in L^2(J)$, the random variable $I_W(\varphi)$ must be (mean-zero) Gaussian (or identically zero, in case $\varphi = 0$ a.e.). This can be seen as follows: Since H_0 is dense in L^2, there exists a sequence φ_n in H_0 such that $\varphi_n \to \varphi$ in L^2. Since I_W is an isometry, $I_W(\varphi_n) \to I_W(\varphi)$ in L^2, and therefore also in distribution. Since each $I_W(\varphi_n)$ is Gaussian, the limit $I_W(\varphi)$ must be Gaussian or zero. [2] Since the second moment of $I_W(\varphi)$ is $\|\varphi\|_2^2$, it is the zero random variable if and only if $\varphi = 0$ a.e. $\square$

[2] If Y_n is a sequence of Gaussian random variables and if $Y_n \to Y$ in distribution, then Y must be either constant or Gaussian. If you don't already know this you should prove it as an exercise. Hint: Use characteristic functions (Fourier transforms).

Proposition 3.1. *Let H_0 be a dense, linear subspace of a Hilbert space H, and let $J : H_0 \to H'$ be a linear isometry mapping H_0 into another Hilbert space H'. Then J extends uniquely to a linear isometry $J : H \to H'$.*

Proof. Exercise. If you don't know what a Hilbert space is, just assume that H and H' are closed linear subspaces of L^2 spaces. □

Exercise 3.1. Let $f \in L^2[0,1]$ be a strictly positive function with L^2-norm 1. For $t \in [0,1]$ define

$$\begin{aligned} F(t) &= \int_0^t f(s)^2 \, ds, \\ \tau(t) &= F^{-1}(t) = \min\{s \,:\, F(s) = t\}, \quad \text{and} \\ Y(t) &= I_W(f\mathbf{1}_{[0,t]}). \end{aligned}$$

Show that $Y(\tau(t))$ is a Wiener process, that is, a mean-zero Gaussian process with covariance function (1.7). Interpretation: $f(s)dW_s$ is a Brownian increment whose "volatility" is multiplied by $|f(s)|$. Thus, $\tau(t)$ runs the integral until the total accumulated squared volatility (variance) reaches t.

Exercise 3.2. Let $g \in L^2[0,1]$ and define $G(t) = \int_0^t g(s)\, ds$. Note that $G(t)$ is continuous (why?). Also, if $W(t)$ is a standard Wiener process with continuous paths, then the integral $\int_0^1 g(s)W(s)\, ds$ is well-defined as a Riemann integral. Show that

$$G(1)W(1) - I_W(G) = \int_0^1 g(s)W(s)\, ds.$$

Exercise 3.3. Let $\{W(t)\}_{t\geq 0}$ be a standard Wiener process, and for each $s \geq 0$ define

$$Y_s := \sqrt{2}e^s \int e^{-t}\mathbf{1}_{[s,\infty]}(t)\, dW(t).$$

Show that $\{Y_s\}_{s\geq 0}$ is a standard Ornstein-Uhlenbeck process.

Exercise 3.4. Let $\{W_t\}_{t\geq 0}$ and $\{W_t^*\}_{t\geq 0}$ be independent standard Wiener processes. Fix $H \in (0,1)$, and define X_t^H by

$$\Gamma(H + \frac{1}{2})X_t^H = \int \left((t+s)^{H-1/2} - s^{H-1/2}\right) dW_s^* + \int (t-s)_+^{H-1/2}\, dW_s. \tag{3.2}$$

(Here $\Gamma(z)$ is Euler's Gamma function, and y_+ denotes the positive part of y, i.e., y if $y > 0$ and 0 otherwise.) Check that X_t^H is a fractional Brownian motion with Hurst parameter H.

Wiener's discovery of Theorem 2 was probably an afterthought to his original construction of the Wiener measure in the early 1920s. In hindsight, we can see that Theorem 2 suggests a natural approach to explicit representations of the Wiener process, via *orthonormal bases*. The idea is this: 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 $L^2(\Omega, \mathcal{F}, P)$. Since each $I_W(\varphi)$ is a mean-zero Gaussian random variable, it follows that the random variables $\xi_n := I_W(\psi_n)$ must be i.i.d. standard normals. (Exercise: why?) 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 3. *Let ξ_n be 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*

$$W_t := \sum_{n=1}^{\infty} \xi_n \langle \mathbf{1}_{[0,t]}, \psi_n\rangle \tag{3.3}$$

converges in the L^2-metric, and the resulting stochastic process $\{W_t\}_{t\in[0,1]}$ is a Wiener process, that is, a mean-zero Gaussian process with covariance (1.7). *Here $\langle,\rangle$ denotes the L^2-inner product:*

$$\langle f,g\rangle = \int_{[0,1]} fg\,dm$$

Remark 4. 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 (3.3) 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:

$$W_t = \xi_0 t + \sum_{k=1}^{\infty} \sum_{n=2^{k-1}}^{2^k-1} n^{-1}\xi_n \sqrt{2}\sin\pi n t. \tag{3.4}$$

4. Lévy's Construction

Lévy subsequently discovered that a much simpler, and in probabilistic terms more natural, construction could be given using the *Haar basis* (the simplest instance of a *wavelet* basis). This basis is defined as follows. First, set

$$\begin{aligned} \psi(x) &= 1 \quad \text{if } \ 0\le x<1/2; \\ &= -1 \quad \text{if } \ 1/2\le x\le 1; \\ &= 0 \quad \text{if } \ x\in\mathbb{R}\setminus[0,1]; \end{aligned} \tag{4.1}$$

this is the *mother wavelet*. The Haar functions $\psi_{m,k}$ are all scaled versions of ψ: for each dyadic interval $J = J_{m,k} = [k2^{-m},(k+1)2^{-m})$ set

$$\begin{aligned} \psi_{m,k}(x) &= \psi(2^m x - k) \quad \text{and} \\ \varphi_{m,k}(x) &= 2^{m/2}\psi_{m,k}(x). \end{aligned} \tag{4.2}$$

Go to Mathworld for a picture. The functions $\psi_{m,k}$ are mutually orthogonal, and together with the constant function 1 span $L^2[0,1]$ (with $m\ge 0$ amd $0\le k<2^m$). As defined here they haven't been normalized; the functions $\varphi_{m,k}$ are, though. Hence, $\{1\}\cup\{\varphi_{m,k}\}_{m\ge 0, 0\le k<2^m}$ is an orthonormal basis for $L^2[0.1]$, and so Theorem 3 implies that if $\xi_{-1},\xi_{0,0},\xi_{1,0},\xi_{1,1},\dots$ are i.i.d. Normal-$(0,1)$ then the series

$$W_t := \xi_{-1}t + \sum_{m=0}^{\infty}\sum_{k=0}^{2^m-1} \xi_{m,k}\langle \mathbf{1}_{[0,t]},\varphi_{m,k}\rangle \tag{4.3}$$

converges in L^2 and defines a Gaussian process with the mean and covariance functions of Brownian motion. Observe that even though the Haar functions $\psi_{m,k}$ are discontinuous, the *integrated* Haar functions $\langle \mathbf{1}_{[0,t]}, \varphi_{m,k}\rangle$ *are* continuous: they are the "hat" functions

$$\begin{aligned} H_{m,k}(t) = \langle \mathbf{1}_{[0,t]}, \varphi_{m,k}\rangle &= 2^{-m/2}(t - k/2^m) \quad \text{for } \frac{k}{2^m} < t \leq \frac{k}{2^m} + \frac{1}{2^{m+1}}; \\ &= 2^{-m/2}((k+1)/2^m - t) \quad \text{for } \frac{k}{2^m} + \frac{1}{2^{m+1}} 1/2^{m+1} < t \leq \frac{k+1}{2^m}; \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Consequently, the series (4.3) can be rewritten as

$$W_t := \xi_{-1} t + \sum_{m=0}^{\infty} \sum_{k=0}^{2^m - 1} \xi_{m,k} H_{m,k}(t). \tag{4.4}$$

Theorem 4. *With probability* 1*, the series* (4.3) *converges uniformly for* $t \in [0,1]$*. Therefore, the process* W_t *defined by* (4.3) *has continuous paths.*

Proof. Fix $m \geq 0$ and $0 \leq k < 2^m$, and consider the event

$$F_{m,k} := \{\max_{0 \leq t \leq 1} |\xi_{m,k} H_{m,k}(t)| > 2^{-m/4}\} = \{|\xi_{m,k}| > 2^{m/2}\}.$$

This event involves only the magnitude of the standard normal variable $\xi_{m,k}$, so its probability can be obtained by integrating the standard normal probability density over the tail region $x > 2^{m/2}$. Using a crude upper bound for the density in the region gives the inequality

$$P(F_{m,k}) \leq \frac{2}{\sqrt{2\pi}} \int_{2^{m/2}}^{\infty} e^{-2^{m/2} x/2}\, dx \leq \sqrt{2/\pi} e^{-2^{m-1}} / 2^{m/2}$$

From this we conclude that

$$\sum_{m=0}^{\infty} \sum_{k=0}^{2^m} P(F_{m,k}) \leq \sum_{m=0}^{\infty} 2^{m/2} e^{-2^{m-1}} < \infty,$$

and so Borel-Cantelli implies that with probability 1 only finitely many of the events $F_{m,k}$ occur. Consequently, all but finitely many terms of the series (4.4) are dominated in absolute value *uniformly in* t by the terms of

$$\sum_{m=0}^{\infty} 2^{-m/4} < \infty.$$

$\square$

UNIVERSITY OF CHICAGO, DEPARTMENT OF STATISTICS, USA