1. GAUSSIAN PROCESSES: DEFINITIONS AND EXAMPLES

**Definition 1.1.** 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 \( F \subseteq T \) 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 \( \mathbb{R} \). A Gaussian process \( \{X_t\}_{t \in T} \) is said to be centered if \( EX_t = 0 \) for every \( t \in T \). The covariance function of a Gaussian process \( \{X_t\}_{t \in T} \) is the bivariate function
\[ R(s, t) = \text{cov}(X_s, X_t) = E(X_s - EX_s)(X_t - EX_t). \]

**Remark 1.** Any Gaussian distribution is completely determined by its mean vector and covariance matrix. In particular, if \( \{X_t\}_{t \in T} \) is a centered Gaussian process with covariance function \( R(s, t) \) then for any finite subset \( F = \{s_i\}_{i \leq d} \subseteq T \) the distribution of the random vector \( X_F \) has density
\[ p(x) = \frac{\exp\{-x^T \Sigma^{-1} x/2\}}{(2\pi)^{d/2} |\det \Sigma|} \]
where \( \Sigma \) is the \( d \times d \) matrix with entries \( R(s_i, s_j) \). Consequently, if two Gaussian processes have the same mean and covariance functions then they have the same finite-dimensional distributions.

We will be most interested in Gaussian processes indexed by an interval \( J \subseteq \mathbb{R} \), usually \( J = [0, \infty) \) or \( J = [0, T] \) for some \( 0 < T < \infty \). Under weak hypotheses there will always exist centered Gaussian processes with specified covariance functions, but it is not generally the case that such processes will have versions with continuous sample paths. Our primary objective in this chapter will be to develop a useful criterion for path-continuity, and to show that this criterion holds for a number of specific Gaussian processes of interest, notably the Wiener process, the Brownian bridge, and the Ornstein-Uhlenbeck process.

**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 \( \{Y_t\}_{t \in \mathbb{R}} \), and the Brownian
bridge \{ W_t \}_{t \in [0,1]}$. These are the centered Gaussian process processes with covariance functions
\begin{align}
EW_s W_t &= \min(s, t), \\
FY_s Y_t &= \exp\{-|t-s|\}, \\
EW_s^2 W_s^2 &= \min(s, t) - st.
\end{align}

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 \( t \geq 0 \). Also, in this chapter (only) we make no requirements of path-continuity on a Gaussian process.

**Exercise 1.1.** (A) Check that if \( W_t \) is a standard Wiener process, then the derived processes
\[ W_t^\circ := W_t - tW_0 \quad \text{and} \quad Y_t := e^{-t}W_{e^2t} \]
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. It follows that if there is a version of the Wiener process with continuous paths, then the same will be true for the Brownian bridge and the Ornstein-Uhlenbeck process. (B) Check that for any scalar \( \alpha > 0 \) the process
\[ \tilde{W}_t := \alpha^{-1}W_{\alpha^2t} \]
has the same covariance function, and therefore also the same finite-dimensional distributions, as \( W_t \). (This correspondence is called Brownian scaling.)

**Example 1.2.** Let \( \xi_1, \xi_2, \ldots \) be independent, identically distributed unit normals. Then for any finite set of frequencies \( \omega_i \geq 0 \), the process
\[ X_t := \sum_{i=1}^{m} \xi_i \cos(\omega_i t) \]
indexed by \( t \in \mathbb{R} \) 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 \( \{ W_s \}_{s \in \mathbb{R}^2} \) is the mean-zero Gaussian process indexed by ordered pairs \( s = (s_1, s_2) \) of nonnegative reals with covariance function
\[ EW_s W_t = \min(s_1, t_1) \min(s_2, t_2). \]
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 \( \{ X^H_t \}_{t \geq 0} \) with covariance function
\[ EX^H_t X^H_s = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \]
equivalently,
\[(1.8)\]
\[E|X^H_t - X^H_s|^2 = |t - s|^{2H}.\]

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
\[(1.9)\]
\[
\{a^{-H}X^H(at)\}_{t \geq 0} \overset{\mathcal{D}}{=} \{X^H(t)\}.
\]
(It is not immediately obvious that the covariance kernel (1.7) is nonnegative definite, but it is.)

2. WIENER’S PERSPECTIVE: THE WIENER INTEGRAL

The difficult step in proving the existence of the Wiener process is establishing sample path continuity. This problem was settled by Wiener around 1920. His proof is far from the simplest of those now known, and so at a certain point we will part company with Wiener and follow a different route. But behind Wiener’s argument 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 on Brownian motion.

Suppose that Brownian motion exists, that is, suppose that on some probability space \((\Omega, \mathcal{F}, P)\) there is a centered Gaussian process \(\{W_t\}_{t \in [0,1]}\) with covariance \(EW_tW_s = \min(s,t)\). The random variables \(W_t\) are all elements of the space \(L^2(P)\) consisting of the real random variables defined on \((\Omega, \mathcal{F}, P)\) with finite second moments. This space is a Hilbert space with inner product \(\langle X, Y \rangle = \text{cov}(X,Y)\).

Now consider the Hilbert space \(L^2[0,1]\), consisting of all real-valued square-integrable functions on the unit interval, with inner product
\[
\langle f, g \rangle = \int_0^1 f(x)g(x) \, dx.
\]
Indicator functions \(1_{[0,t]}\) of intervals \([0, t]\) are elements of \(L^2[0,1]\), and obviously
\[
\langle 1_{[0,t]}, 1_{[0,s]} \rangle = \min(s,t).
\]
Thus, the indicators \(1_{[0,t]}\) have exactly the same inner products as do the random variables \(W_t\) in the Hilbert space \(L^2(P)\). Wiener’s key insight was that this identity between inner products implies that there is a linear isometry \(I_W\) from \(L^2[0,1]\) into \(L^2(P)\) mapping each indicator \(1_{[0,t]}\) to the corresponding random variable \(W_t\).

**Theorem 1.** (Wiener’s Isometry) Let \(\{w_t\}_{t \geq 0}\) 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 \(1_{(s,t)} \mapsto W_t - W_s\) 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 2.** The mapping \(I_W\) is called the Wiener isometry or, more commonly, the Wiener integral. The notation
\[(2.1)\]
\[I_W(\varphi) = \int \varphi \, dW\]
is often used. The integral is defined only for nonrandom integrands \( \varphi \), and in particular, only those functions \( \varphi \) that are square-integrable against Lebesgue measure. K. Itô later extended Wiener’s integral so as to allow random integrands – more on this later in the course.

**Proof.** Given the identity \( \langle 1_{[0,t]}, 1_{[0,s]} \rangle = \langle W_s, W_t \rangle \), the theorem 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 \( 1_A \). 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 2.1 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 \( \varphi_n \) 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. \(^1\) Since the second moment of \( I_W(\varphi) \) is \( \| \varphi \| _H^2 \), it is the zero random variable if and only if \( \varphi = 0 \) a.e. \( \square \)

**Proposition 2.1.** 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.** See any standard text (e.g., RUDIN, Real and Complex Analysis) that treats the basic theory of Hilbert spaces. \( \square \)

**Exercise 2.1.** Let \( \{ W(t) \}_{t \geq 0} \) be a standard Wiener process, and for each \( s \geq 0 \) define

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

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

The Hilbert space isometry \( I_W \) of Theorem 1 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 \( L^2(P) \). 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 \( 1_{[0,t]} \) in the basis \( \psi_n \) to the series expansion of \( W_t \) in the basis \( \xi_n \). Thus, with no further work we have the following.

**Theorem 2.** Assume that the probability space \( (\Omega, \mathcal{F}, P) \) supports an infinite sequence \( \xi_n \) 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

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

\(^1\)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).
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 (1.2). Here $\langle , \rangle$ denotes the $L^2$-inner product:

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

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

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

the series (2.2) 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:

(2.3) $$W_t = \xi_0 t + \sum_{k=1}^{2^n-1} \sum_{n=2^{k-1}}^{2^n-1} n^{-1} \xi_n \sqrt{2} \sin \pi nt.$$  

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 two different alternative constructions, the first due to Lévy, the second to Kolmogorov.

3. Lévy’s Construction

Paul Lévy discovered that it is relatively easy to prove uniform convergence of the random series (2.2) if one uses the Haar wavelet basis instead of the Fourier basis. The Haar basis is defined as follows: first, let $\psi : \mathbb{R} \to \{-1, 1\}$ be the “mother wavelet” function

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

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

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

The function $\psi_{n,k}$ has support $[k/2^n, (k + 1)/2^n]$, and has absolute value equal to $2^{n/2}$ on this interval, so its $L^2$-norm is 1. Note that $\psi_{0,0} = 1$ on $[0,1]$. Moreover, the functions $\psi_{n,k}$ are mutually orthogonal:

(3.3) $$\langle \psi_{n,k}, \psi_{m,l} \rangle = \begin{cases} 1 & \text{if } n = m \text{ and } k = l; \\ 0 & \text{otherwise}. \end{cases}$$

**Exercise 3.1.** Prove that the Haar functions $\{\psi_{n,k}\}_{n\geq0, 0\leq k < 2^n}$ form a complete orthonormal basis of $L^2[0,1]$. HINT: Linear combinations of the indicator functions of dyadic intervals $[k/2^n, (k + 1)/2^n]$ are dense in $L^2[0,1]$.

In certain senses the Haar basis is better suited to Brownian motion than the Fourier basis, in part because the functions $\psi_{n,k}$ are “localized” (which fits with the independent increments property), and in part because the normalization of the functions forces the scale factor $2^{n/2}$ in (3.2) (which fits with the Brownian scaling law).
The series expansion (2.2) involves the inner products of the basis functions with the indicators $1_{[0,t]}$. For the Haar basis, these inner products define the Schauder functions $G_{n,k}$, which are defined as the indefinite integrals of the Haar functions:

$$G_{n,k}(t) = \int_0^t \psi_{n,k}(s) \, ds$$

The graphs of these functions are steeply peaked “hats” sitting on the dyadic intervals $[k/2^n, (k + 1)/2^n]$, with heights $2^{-n/2}$ and slopes $\pm 2^{n/2}$. Note that $G_{0,0}(t) = t$.

**Theorem 3.** (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

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

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

**Lemma 3.1.** If $Z$ is a standard normal random variable then for every $x > 0$,

$$P\{Z > x\} \leq \frac{2e^{-x^2/2}}{\sqrt{2\pi x}}. \tag{3.6}$$

**Proof.**

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

**Proof of Theorem 3.** By definition of the Schauder functions $G_{n,k}$, the series (3.5) is a particular case of (2.2), so the random variables $W(t)$ defined by (3.5) are centered Gaussian with covariances that agree with the covariances of a Wiener process. Hence, to prove that (3.5) 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 (3.5) 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 3.1 for the normal distribution to show that with probability one there is a (possibly random) $m_*$ such that

$$\max_k |\xi_{m,k}| \leq 2^{m_*/4} \quad \text{for all } m \geq m_* \tag{3.7}$$

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 (3.7) holds eventually, it suffices (by Borel-Cantelli) to show that the probabilities of the complementary events are summable. By Lemma 3.1,

$$P\{\xi_{m,k} \geq 2^{m/4}\} \leq \frac{4}{2^{m/4}\sqrt{2\pi}} e^{-2^{m/2}}.$$
Hence, by the Bonferroni inequality (i.e., the crude union bound),

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

Since this bound is summable in $m$, Borel-Cantelli implies that with probability 1, eventually (3.7) must hold. This proves that w.p.1 the series (3.5) converges uniformly, and therefore $W(t)$ is continuous.

\[ \square \]

4. KOLMOGOROV-CHENTSOV THEOREM

4.1. 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 $t \in [0, 1)$; it applies also to stochastic processes indexed by parameters that take values in a subset of a higher-dimensional Euclidean space. (Such processes are sometimes called random fields.)

**Theorem 4.** (Kolmogorov-Chentsov) Let $\{X_t\}_{t \in T}$ be a stochastic process whose index set $T$ is a countable, dense set of points $t$ in a compact set $K \subseteq \mathbb{R}^d$. Suppose that there are positive constants $\alpha, \beta, C$ such that

$$E|X_t - X_s|^\alpha \leq C||t - s|^{d+\beta} \text{ for all } s, t \in T. \quad (4.1)$$

Then $\{X_t\}_{t \in T}$ can be extended to a stochastic process $\{\tilde{X}_t\}_{t \in K}$ with index set $K$, and in such a way that with probability one, the mapping $t \mapsto \tilde{X}_t$ is continuous. Moreover, the sample paths of the extension are H"{o}lder continuous, that is,

$$\max_{s \neq t \in K} \frac{|X_t - X_s|}{|t - s|^\gamma} < \infty \text{ almost surely.} \quad (4.2)$$

**Remark 3.** This is often used to prove that a stochastic process $\{X_t\}_{t \in K}$ with compact index set $K \subseteq \mathbb{R}$ has a version with continuous sample paths. Here is how it works: if $\{X_t\}_{t \in K}$ satisfies the hypothesis (4.1) then its restriction to any countable, dense subset $T$ must also satisfy (4.1). The theorem then implies that the restriction $\{X_t\}_{t \in T}$ has an extension to a stochastic process $\{\tilde{X}_t\}_{t \in T}$ with continuous sample paths. But the hypothesis (4.1) then guarantees that for every $t \in K$,

$$X_t = \tilde{X}_t \text{ a.s.}$$

(Exercise: Prove this.) Thus, $\{\tilde{X}_t\}_{t \in T}$ is a version of $\{X_t\}_{t \in K}$.

**Definition 4.1.** Let $\{X_t\}_{t \in T}$ and $\{Y_t\}_{t \in T}$ be stochastic processes defined on the same probability space, with the same index set $T$. Then $\{Y_t\}_{t \in T}$ is said to be a version of $\{X_t\}_{t \in T}$ if for every $t \in T$,

$$Y_t = X_t \text{ almost surely.}$$

Note: the null sets on which $X_t \neq Y_t$ may be different for different values of $t \in T$. Thus, if $T$ is uncountable, the union of these null sets might be a set of positive probability.

**Remark 4.** The crucial hypothesis of the Kolmogorov-Chentsov theorem is that the exponent $d + \beta$ in (4.1) is strictly greater than the ambient dimension $d$. The following example shows that the conclusion of the theorem need not hold if (4.1) holds with $\beta = 0$.  

Example 4.1. Consider the standard Poisson counting process \( \{ N_t \}_{t \geq 0} \): this certainly does not have continuous paths. This shows that the moment hypothesis (4.1) requires an exponent \( d + \beta \) strictly larger than \( d \), at least when \( d = 1 \), because for every \( \alpha = 1, 2, \ldots \),
\[
E (N_{t+s} - N_t)^\alpha = C_\alpha s.
\]

Example 4.2. Next, consider the Wiener process. We have shown in section 2 that on any probability space that supports an infinite sequence \( \{ \xi_n \}_{n \geq 1} \) of independent, identically distributed unit Gaussian random variables there is a centered Gaussian process \( \{ W_t \}_{t \in [0,1]} \) with covariance function (1.2). Consider the restriction of this Gaussian process to the set of dyadic rationals \( t \in [0,1] \). For any \( s, t \), the increment \( W_t - W_s \) is normally distributed with mean 0 and variance \( |t - s|^2 \), equivalently, \( W_t - W_s \) has the same distribution as \( |t - s|^{1/2} Z \), where \( Z \) is standard normal. Hence,
\[
E |W_t - W_s|^{2k} = C_k |t - s|^k
\]
where \( C_k \) is the \( 2k \)th moment of a standard normal. Thus, the Kolmogorov-Chentsov condition (4.1) holds with \( \alpha = 2k \) and \( \beta = k - 1 \) for all \( k \geq 1 \). Therefore, Theorem 4 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 (4.2) that for every \( \gamma < 1/2 \),
\[
\max_{t \neq s \in [0,1]} \frac{|W_t - W_s|}{|t - s|^{\gamma}} < \infty
\]
almost surely. (Note: Lévy proved a sharper result, called Lévy’s modulus, which states that the denominator \( |t - s|^\gamma \) can be replaced by \( (|t - s| \log(1/|t - s|))^{1/2} \).

Exercise 4.1. Use the Kolmogorov-Chentsov criterion to prove that there is a continuous version of the Brownian sheet with parameter space \( t \in [0,1]^2 \).

Exercise 4.2. More generally, let \( R(s, t) \) be any covariance function defined for \( s, t \) in an open subset of \( \mathbb{R}^d \) that satisfies
\[
R(s, s) + R(t, t) - 2R(s, t) \leq C |t - s|^\gamma
\]
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.

4.2. Proof of the Kolmogorov-Chentsov Theorem. For simplicity I will consider the special case where \( K = [0,1] \) is the unit interval and \( T \) is the set of points with dyadic rational coordinates in \( K \). The general case can be proved by the same argument, with some routine modifications.

We begin by developing some tools from real analysis.

Definition 4.2. Let \( (K, d) \) be an arbitrary metric space, and let \( f : K \to \mathbb{R} \) be a real-valued function on \( K \). A modulus of continuity of \( f \) is any function \( \omega : [0,1] \to \mathbb{R}_+ \) such that for all \( \delta \in (0,1) \),
\[
\sup_{d(x,y) \leq \delta} \{ |f(x) - f(y)| \} \leq \omega(\delta).
\]
The function \( f \) is Hölder continuous with Hölder exponent \( \gamma > 0 \) if for some \( C < \infty \) the function \( \omega(\delta) = C\delta^\gamma \) is a modulus of continuity of \( f \).
**Lemma 4.1.** Let $\omega : [0, 1] \to \mathbb{R}_+$ be any non-increasing, continuous function such that $\lim_{\delta \to 0} \omega(\delta) = 0$. If $T$ is a dense subset of a metric space $(K, d)$ and $f : T \to \mathbb{R}$ has modulus of continuity $\omega$ (with respect to the metric $d$ on $T$) then $f$ has a unique extension to $K$ with the same modulus of continuity.

**Proof.** Exercise. Hint: For every point $x \in K$ there is a sequence $x_n$ in $T$ with limit $x$. Show that for any such sequence, $\lim f(x_n)$ exists, and show that the limit depends only on $x$, and not on the particular sequence $x_n$ chosen. This specifies the extension. Now show that the extension has modulus of continuity $\omega$. 

Now let $K = [0, 1]$, and for each integer $m \geq 1$ let $L_m$ be the set of rational numbers in $[0, 1]$ of the form $n/2^m$, where $n$ is an integer between 1 and $2^m$. The union $T = \bigcup_{m \geq 1} L_m$ is the set of dyadic rationals in $[0, 1]$. If $s, t \in L_m$ for some $m \geq 1$, say that $s$ and $t$ are nearest neighbors in $L_m$ if $|s - t| = 2^m$. (Thus, for instance, the points $3/4$ and $7/8$ are nearest neighbors in $L_3$, but $5/8$ and $7/8$ are not.)

**Lemma 4.2.** Let $x : T \to \mathbb{R}$ be a real-valued function such that for some $\gamma > 0$ and $C < \infty$,

$$
|x(s) - x(t)| \leq C|t - s|^\gamma
$$

for all nearest neighbors $s, t \in L_m$ and all levels $m \geq 1$. Then $x(t)$ extends to a continuous function on the closed unit cube whose modulus of continuity satisfies

$$
\omega(\delta) \leq C' \delta^{\gamma}
$$

for a constant $C' < \infty$ depending only on $C$ and $\gamma$.

**Proof.** **Claim 1:** Suppose that $s, t \in L_m$ are at distance $< 2^{-k}$, for some $k \leq m$. Then $s, t$ are connected by a path $s = s_0, s_1, \ldots, s_n = t$ in $\bigcup_{j \leq m} L_m$ such that (a) any two adjacent points $s_i, s_{i+1}$ in the path are nearest neighbors in $L_j$ for some $k \leq j \leq m$; and (b) for each level $k \leq j \leq m$ the number of nearest neighbor pairs in $L_j$ that occur along the path is bounded by 2.

**Proof of Claim 1 for $d = 1$.** Let $s = r_1/2^m$ and $t = r_2/2^m$ and assume that $r_1 < r_2$ and $r_2 - r_1 \leq 2^{m-k}$. Construct the path by first moving up one dyadic level at a time, until it is no longer possible to go up without exceeding $s$, and then moving down one dyadic level at a time until reaching $t$. For example, if $s = 1/16$ and $t = 5/16$ then the connecting path is

$$
\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{5}{16}.
$$

By construction, each adjacent pair along the path will be nearest neighbors in their common dyadic level, and furthermore, no dyadic level will be touched more than twice (once going up, once on the way down). Finally, if $r_2 - r_1 < 2^{m-k}$ then no level $L_j$ with $j \leq k$ can be visited along the path, because this would force the distance between the two successive points on the path at this level to be $2^{-k}$, which is larger than the distance from $s$ to $t$.

**Claim 2:** There is a constant $C' < \infty$ such that for any two points $s, t \in \bigcup_{m=1}^\infty L_m$,

$$
|x(s) - x(t)| \leq C'|t - s|^\gamma.
$$
Proof of Claim 2. If \( s, t \in \bigcup_{m=1}^{\infty} L_m \) then there is some \( m \geq 1 \) such that \( s, t \in L_m \). Assume that \( s, t \in L_m \) are at distance \( < d 2^{-k} \), for some \( k \leq m \). Then they are connected by a path \( s = s_0, s_1, \ldots, s_n = t \) with the properties spelled out in Claim 1. Consequently, by the triangle inequality,

\[
|x(s) - x(t)| \leq \sum_{i=1}^{n} |x(s_i) - x(s_{i+1})|
\]

\[
\leq \sum_{i=1}^{n} C|s_i - s_{i+1}|^\gamma
\]

\[
\leq 2Cd \sum_{l=k}^{m} 2^{-\gamma l}
\]

\[
\leq 2Cd 2^{-k\gamma}/(1 - 2^{-\gamma})
\]

\[
\leq C'|s - t|^\gamma
\]

for a suitable constant \( C' \). The lemma now follows directly from Lemma 1. \( \square \)

Remark 5. The same result holds for functions on the unit cube \([0, 1]^d\), provided the definition of the levels \( L_m \) and the notion of a nearest neighbor pair are extended in the obvious way. The connecting paths (Claim 1) are constructed by moving one coordinate at a time. Other than this, the proof in the higher-dimensional case is identical to that for \( d = 1 \).

Proof of Theorem 4. The crucial geometric fact is that the cardinality of \( L_m \) grows like

\[
\#L_m = 2^m.
\]

(What is important is the exponential growth rate, not the exact count.) If \( s, t \) are neighbors in \( L_m \), then by the hypothesis (4.1) and the Markov inequality,

\[
P\{|X_t - X_s| \geq |t - s|^\gamma \} \leq \frac{E|X_t - X_s|^\alpha}{|t - s|^\gamma \alpha}
\]

\[
\leq C|t - s|^{1+\beta - \gamma \alpha}
\]

\[
= C 2^{-m \beta + m \gamma \alpha}.
\]

Since the number of neighboring pairs in \( L_m \) grows like a constant multiple of \( 2^m \), it follows that

\[
P(B_m) \leq C' 2^{-m \beta + m \gamma \alpha} \quad \text{where}
\]

\[
B_m := \{|X_t - X_s| \geq |t - s|^\gamma \text{ for some neighboring } s, t \in L_m \}.
\]

Since \( \beta > 0 \), it is possible to choose \( \gamma > 0 \) so small that \( -\beta + \gamma \alpha < 0 \). For such a choice, the probabilities (4.10) 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 4.2. \( \square \)
4.3. Differentiability 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 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 4.3.** 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

$$DX(t, \varepsilon) := \frac{X(t + \varepsilon) - X(t)}{\varepsilon} \quad \text{for} \quad \varepsilon \neq 0,$$

$$= X'(t) \quad \text{for} \quad \varepsilon = 0.$$

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 $\varepsilon \neq 0$,

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

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

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

**Proposition 4.4.** 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 (4.6) above. Then there is a continuously differentiable, mean-zero Gaussian process $X(t)$ with covariance $R(s, t)$. 

*University of Chicago, Department of Statistics, USA*