CHAPTER 1

CLASSICAL RESULTS ON THE REGULARITY OF PATHS

This initial chapter contains a number of elements that are used repeatedly in the book and constitute necessary background. We will need to study the paths of random processes and fields; the analytical properties of these functions play a relevant role. This raises a certain number of basic questions, such as whether the paths belong to a certain regularity class of functions, what one can say about their global or local extrema and about local inversion, and so on. A typical situation is that the available knowledge on the random function is given by its probability law, so one is willing to know what one can deduce from this probability law about these kinds of properties of paths. Generally speaking, the result one can expect is the existence of a *version* of the random function having good analytical properties. A version is a random function which, at each parameter value, coincides almost surely with the one given. These are the contents of Section 1.4, which includes the classical theorems due to Kolmogorov and the results of Bulinskaya and Ylvisaker about the existence of critical points or local extrema having given values. The essence of all this has been well known for a long time, and in some cases proofs are only sketched. In other cases we give full proofs and some refinements that will be necessary for further use.

As for the earlier sections, Section 1.1 contains starting notational conventions and a statement of the Kolmogorov extension theorem of measure theory, and Sections 1.2 and 1.3 provide a quick overview of the Gaussian distribution and some connected results. Even though this is completely elementary, we call the reader's attention to Proposition 1.2, the Gaussian regression formula, which

Level Sets and Extrema of Random Processes and Fields, By Jean-Marc Azaïs and Mario Wschebor

will appear now and again in the book and can be considered as the basis of calculations using the Gaussian distribution.

1.1. KOLMOGOROV'S EXTENSION THEOREM

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space and $(F, \mathcal{F})$ a measurable space. For any measurable function

$$Y : (\Omega, \mathcal{A}) \to (F, \mathcal{F}),$$

that is, a random variable with values in F, the image measure

$$Q(A) = \mathrm{P}(Y^{-1}(A)) \qquad A \in \mathcal{F}$$

is called the *distribution* of Y.

Except for explicit statements to the contrary, we assume that probability spaces are complete; that is, every subset of a set that has zero probability is measurable. Let us recall that if $(F, \mathcal{F}, \mu)$ is a measure space, one can always define its completion $(F, \mathcal{F}_1, \mu_1)$ by setting

$$\mathcal{F}_1 = \{A : \exists B, C, A = B \vartriangle C, \text{such that } B \in \mathcal{F}, C \subset D \in \mathcal{F}, \mu(D) = 0\}, \tag{1.1}$$

and for $A \in \mathcal{F}_1$, $\mu_1(A) = \mu(B)$, whenever A admits the representation in (1.1). One can check that $(F, \mathcal{F}_1, \mu_1)$ is a complete measure space and μ_1 an extension of μ.

A *real-valued stochastic process indexed by the set* T is a collection of random variables $\{X(t) : t \in T\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. In what follows we assume that the process is *bi-measurable*. This means that we have a σ-algebra $\mathcal{T}$ of subsets of T and a Borel-measurable function of the pair (t, ω) to the reals:

$$X : (T \times \Omega, \mathcal{T} \times \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

($\mathcal{B}_{\mathbb{R}}$ denotes the Borel σ-algebra in $\mathbb{R}$), so that

$$X(t)(\omega) = X(t, \omega).$$

Let T be a set and $\mathbb{R}^T = \{g : T \to \mathbb{R}\}$ the set of real-valued functions defined on T. (In what follows in this section, one may replace $\mathbb{R}$ by $\mathbb{R}^d$, $d > 1$.) For $n = 1, 2, \ldots,$ $t_1, t_2, \ldots, t_n$, n distinct elements of T, and $B_1, B_2, \ldots, B_n$ Borel sets in $\mathbb{R}$, we denote

$$C(t_1, t_2, \ldots, t_n; B_1, B_2, \ldots, B_n) = \{g \in \mathbb{R}^T : \ g(t_j) \in B_j, \ \ j = 1, 2, \ldots, n\}$$

and $\mathcal{C}$ the family of all sets of the form $C(t_1, t_2, \ldots, t_n; B_1, B_2, \ldots, B_n)$. These are usually called *cylinder sets depending on* $t_1, t_2, \ldots, t_n$. The smallest σ-algebra of parts of $\mathbb{R}^T$ containing $\mathcal{C}$ will be called the *Borel σ-algebra of* $\mathbb{R}^T$ and denoted by $\sigma(\mathcal{C})$.

Consider now a family of probability measures

$$\{P_{t_1,t_2,\ldots,t_n}\}_{t_1,t_2,\ldots,t_n \in T;\ n=1,2,\ldots} \tag{1.2}$$

as follows: For each $n = 1, 2, \ldots$ and each n-tuple $t_1, t_2, \ldots, t_n$ of distinct elements of T, $P_{t_1,t_2,\ldots,t_n}$ is a probability measure on the Borel sets of the product space $X_{t_1} \times X_{t_2} \times \cdots \times X_{t_n}$, where $X_t = \mathbb{R}$ for each $t \in T$ (so that this product space is canonically identified as $\mathbb{R}^n$).

We say that the probability measures (1.2) satisfy the *consistency condition* if for any choice of $n = 1, 2, \ldots$ and distinct $t_1, \ldots, t_n, t_{n+1} \in T$, we have

$$P_{t_1,\ldots,t_n,t_{n+1}}(B \times \mathbb{R}) = P_{t_1,\ldots,t_n}(B)$$

for any Borel set B in $X_{t_1} \times \cdots \times X_{t_n}$. The following is the basic Kolmogorov extension theorem, which we state but do not prove here.

Theorem 1.1 (Kolmogorov). $\{P_{t_1,t_2,\ldots,t_n}\}_{t_1,t_2,\ldots,t_n \in T;\ n=1,2,\ldots,}$ *satisfy the consistency condition if and only if there exists one and only one probability measure* P *on* $\sigma(\mathcal{C})$ *such that*

$$P(C(t_1, \ldots, t_n; B_1, \ldots, B_n)) = P_{t_1,\ldots,t_n}(B_1 \times \cdots \times B_n) \tag{1.3}$$

for any choice of $n = 1, 2, \ldots$*, distinct* $t_1, \ldots, t_n \in T$ *and* B_j *Borel sets in* X_{t_j}, $j = 1, \ldots, n$.

It is clear that if there exists a probability measure P on $\sigma(\mathcal{C})$ satisfying (1.3), the consistency conditions must hold since

$$C(t_1, \ldots, t_n, t_{n+1}; B_1, \ldots, B_n, X_{t_{n+1}}) = C(t_1, \ldots, t_n; B_1, \ldots, B_n).$$

So the problem is how to prove the converse. This can be done in two steps: (1) define P on the family of cylinders $\mathcal{C}$ using (1.3) and show that the definition is unambiguous (note that each cylinder has more than one representation); and (2) apply Caratheodory's theorem on an extension of measures to prove that this P can be extended in a unique form to $\sigma(\mathcal{C})$.

Remarks

1. Theorem 1.1 is interesting when T is an infinite set. The purpose is to be able to measure the probability of sets of functions from T to $\mathbb{R}$ (i.e., subsets of $\mathbb{R}^T$) which cannot be defined by means of a finite number of coordinates, which amounts to looking only at the values of the functions at a finite number of t-values.

Notice that in the case of cylinders, if one wants to know whether a given function $g : T \to \mathbb{R}$ belongs to $C(t_1, \ldots, t_n; B_1, \ldots, B_n)$, it suffices to look at the values of g at the finite set of points $t_1, \ldots, t_n$ and check if $g(t_j) \in B_j$ for $j = 1, \ldots, n$. However, if one takes, for example, $T = \mathbb{Z}$ (the integers) and considers the sets of functions

$$A = \{g : g : T \to \mathbb{R}, \lim_{t\to+\infty} g(t) \text{ exists and is finite}\}$$

or

$$B = \{g : g : T \to \mathbb{R}, \sup_{t\in T} |g(t)| \le 1\},$$

it is clear that these sets are in $\sigma(\mathcal{C})$ but are not cylinders (they "depend on an infinite number of coordinates").

2. In general, $\sigma(\mathcal{C})$ is strictly smaller than the family of all subsets of $\mathbb{R}^T$. To see this, one can check that

$$\sigma(\mathcal{C}) = \{A \subset \mathbb{R}^T : \exists T_A \subset T, T_A \text{ countable and } B_A \text{ a Borel set in } \mathbb{R}^{T_A}, \text{ such that } g \in A \text{ if and only if } g/T_A \in B_A\}. \quad (1.4)$$

The proof of (1.4) follows immediately from the fact that the right-hand side is a σ-algebra containing $\mathcal{C}$. Equation (1.4) says that a subset of $\mathbb{R}^T$ is a Borel set if and only if it "depends only on a countable set of parameter values." Hence, if T is uncountable, the set

$$\{g \in \mathbb{R}^T : g \text{ is a bounded function}\}$$

or

$$\{g \in \mathbb{R}^T : g \text{ is a bounded function}, |g(t)| \le 1 \text{ for all } t \in T\}$$

does not belong to $\sigma(\mathcal{C})$. Another simple example is the following: If $T = [0, 1]$, then

$$\{g \in \mathbb{R}^T : g \text{ is a continuous function}\}$$

is not a Borel set in $\mathbb{R}^T$, since it is obvious that there does not exist a countable subset of [0, 1] having the determining property in (1.4). These examples lead to the notion of *separable process* that we introduce later.

3. In the special case when $\Omega = \mathbb{R}^T$, $\mathcal{A} = \sigma(\mathcal{C})$, and $X(t)(\omega) = \omega(t)$, $\{X(t) : t \in T\}$ is called a *canonical process*.

4. We say that the stochastic process $\{Y(t) : t \in T\}$ is a version of the process $\{X(t) : t \in T\}$ if $P(X(t) = Y(t)) = 1$ for each $t \in T$.

1.2. REMINDER ON THE NORMAL DISTRIBUTION

Let μ be a probability measure on the Borel subsets of $\mathbb{R}^d$. Its Fourier transform $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is defined as

$$\widehat{\mu}(z) = \int_{\mathbb{R}^d} \exp(i\langle z, x\rangle)\mu(dx),$$

where $\langle\cdot,\cdot\rangle$ denotes the usual scalar product in $\mathbb{R}^d$.

We use *Bochner's theorem* (see, e.g., Feller, 1966): $\widehat{\mu}$ is the Fourier transform of a Borel probability measure on $\mathbb{R}^d$ if and only if the following three conditions hold true:

1. $\widehat{\mu}(0) = 1$.
2. $\widehat{\mu}$ is continuous.
3. $\widehat{\mu}$ is positive semidefinite; that is, for any $n = 1, 2, \dots$ and any choice of the complex numbers $c_1, \dots, c_n$ and of the points $z_1, \dots, z_n$, one has

$$\sum_{j,k=1}^{n} \widehat{\mu}(z_j - z_k)c_j\overline{c}_k \geq 0.$$

The random vector ξ with values in $\mathbb{R}^d$ is said to have the *normal distribution*, or the *Gaussian distribution*, with parameters (m, Σ)" [$m \in \mathbb{R}^d$ and Σ a $d \times d$ positive semidefinite matrix] if the Fourier transform of the probability distribution μ_ξ of ξ is equal to

$$\widehat{\mu_\xi}(z) = \exp\left[i\langle m, z\rangle - \tfrac{1}{2}\langle z, \Sigma z\rangle\right] \qquad z \in \mathbb{R}^d.$$

When $m = 0$ and $\Sigma = I_d$ = identity $d \times d$ matrix, the distribution of ξ is called *standard normal in* $\mathbb{R}^d$. For $d = 1$ we use the notation

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-(1/2)x^2} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^{x} \varphi(y)\,dy$$

for the density and the cumulative distribution function of a standard normal random variable, respectively.

If Σ is nonsingular, μ_ξ is said to be nondegenerate and one can verify that it has a density with respect to Lebesgue measure given by

$$\mu_\xi(dx) = \frac{1}{(2\pi)^{d/2}(\det(\Sigma))^{1/2}} \exp\left[-\frac{1}{2}(x-m)^{\mathrm{T}}\Sigma^{-1}(x-m)\right] dx$$

x^{T} denotes the transpose of x. One can check that

$$m = \mathrm{E}(\xi), \qquad \Sigma = \mathrm{Var}(\xi) = \mathrm{E}((\xi - m)(\xi - m)^{\mathrm{T}}),$$

so m and Σ are, respectively, the mean and the variance of ξ.

From the definition above it follows that if the random vector ξ with values in $\mathbb{R}^d$ has a normal distribution with parameters m and Σ, A is a real matrix with n rows and d columns, and b is a nonrandom element of $\mathbb{R}^n$, then the random vector $A\xi + b$ with values in $\mathbb{R}^n$ has a normal distribution with parameters $(Am + b, A\Sigma A^{\mathrm{T}})$. In particular, if Σ is nonsingular, the coordinates of the random vector $\Sigma^{-1/2}(\xi - m)$ are independent random variables with standard normal distribution on the real line.

Assume now that we have a pair ξ and η of random vectors in $\mathbb{R}^d$ and $\mathbb{R}^{d'}$, respectively, having finite moments of order 2. We define the $d \times d'$ covariance matrix as

$$\mathrm{Cov}(\xi, \eta) = \mathrm{E}((\xi - \mathrm{E}(\xi)(\eta - \mathrm{E}(\eta)^{\mathrm{T}}).$$

It follows that if the distribution of the random vector (ξ, η) in $\mathbb{R}^{d+d'}$ is normal and $\mathrm{Cov}(\xi, \eta) = 0$, the random vectors ξ and η are independent. A consequence of this is the following useful formula, which is standard in statistics and gives a version of the conditional expectation of a function of ξ given the value of η.

Proposition 1.2. *Let ξ and η be two random vectors with values in $\mathbb{R}^d$ and $\mathbb{R}^{d'}$, respectively, and assume that the distribution of (ξ, η) in $\mathbb{R}^{d+d'}$ is normal and* $\mathrm{Var}(\eta)$ *is nonsingular. Then, for any bounded function $f : \mathbb{R}^d \to \mathbb{R}$, we have*

$$\mathrm{E}(f(\xi)|\eta = y) = \mathrm{E}(f(\zeta + Cy)) \tag{1.5}$$

for almost every y, where

$$C = \mathrm{Cov}(\xi, \eta)[\mathrm{Var}(\eta)]^{-1} \tag{1.6}$$

and ζ is a random vector with values in $\mathbb{R}^d$, having a normal distribution with parameters

$$\big(\mathrm{E}(\xi) - C\mathrm{E}(\eta), \mathrm{Var}(\xi) - \mathrm{Cov}(\xi, \eta)[\mathrm{Var}(\eta)]^{-1}[\mathrm{Cov}(\xi, \eta)]^{\mathrm{T}}\big). \tag{1.7}$$

Proof. The proof consists of choosing the matrix C so that the random vector

$$\zeta = \xi - C\eta$$

becomes independent of η. For this purpose, we need the fact that

$$\mathrm{Cov}(\xi - C\eta, \eta) = 0,$$

and this leads to the value of C given by (1.6). The parameters (1.7) follow immediately. □

In what follows, we call the version of the conditional expectation given by formula (1.5), *Gaussian regression*. To close this brief list of basic properties, we mention that a useful property of the Gaussian distribution is stability under passage to the limit (see Exercise 1.5).

Let $r : T \times T \to \mathbb{R}$ be a positive semidefinite function and $m : T \to \mathbb{R}$ a function. In this more general context, that r is a positive semidefinite function, means that for any $n = 1, 2, \dots$ and any choice of distinct $t_1, \dots, t_n \in T$, the matrix $((r(t_j, t_k)))_{j,k=1,\dots,n}$ is positive semidefinite. [This is consistent with the previous definition, which corresponds to saying that $r(s,t) = \widehat{\mu}(s-t)$, $s, t \in \mathbb{R}^d$ is positive semidefinite.]

Take now for $\mathrm{P}_{t_1,\dots,t_n}$ the Gaussian probability measure in $\mathbb{R}^n$ with mean

$$m_{t_1,\dots,t_n} := (m(t_1), \dots, m(t_n))^{\mathrm{T}}$$

and variance matrix

$$\Sigma_{t_1,\dots,t_n} := ((r(t_j, t_k)))_{j,k=1,\dots,n}.$$

It is easily verified that the set of probability measures $\{\mathrm{P}_{t_1,\dots,t_n}\}$ verifies the consistency condition, so that Kolmogorov's theorem applies and there exists a unique probability measure P on the measurable space $(\mathbb{R}^T, \sigma(\mathcal{C}))$, which restricted to the cylinder sets depending on $t_1, \dots, t_n$ is $\mathrm{P}_{t_1,\dots,t_n}$ for any choice of distinct parameter values $t_1, \dots, t_n$. P is called the *Gaussian measure generated by the pair* (m, r). If $\{X(t) : t \in T\}$ is a real-valued stochastic process with distribution P, one verifies that:

- For any choice of distinct parameter values $t_1, \dots, t_n$, the joint distribution of the random variables $X(t_1), \dots, X(t_n)$ is Gaussian with mean $m_{t_1,\dots,t_n}$ and variance $\Sigma_{t_1,\dots,t_n}$.
- $\mathrm{E}(X(t)) = m(t)$ for $t \in T$.
- $\mathrm{Cov}(X(s), X(t)) = \mathrm{E}((X(s) - m(s))(X(t) - m(t))) = r(s,t)$ for $s, t \in T$.

A class of examples that appears frequently in applications is the d-parameter real-valued Gaussian processes, which are centered and stationary, which means that

$$T = \mathbb{R}^d, \qquad m(t) = 0, \qquad r(s,t) = \Gamma(t-s).$$

A general definition of strictly stationary processes is given in Section 10.2.

If the function Γ is continuous, $\Gamma(0) \neq 0$, one can write

$$\Gamma(\tau) = \int_{\mathbb{R}^d} \exp(i\langle \tau, x\rangle)\mu(dx),$$

where μ is a Borel measure on $\mathbb{R}^d$ with total mass equal to $\Gamma(0)$. μ is called the *spectral measure* of the process. We usually assume that $\Gamma(0) = 1$: that is, that μ is a probability measure which is obtained simply by replacing the original process $\{X(t) : t \in \mathbb{R}^d\}$ by the process $\{X(t)/(\Gamma(0))^{1/2} : t \in \mathbb{R}^d\}$.

Example 1.1 (Trigonometric Polynomials). *An important example of stationary Gaussian processes is the following. Suppose that μ is a purely atomic probability symmetric measure on the real line; that is, there exists a sequence $\{x_n\}_{n=1,2,..}$ of positive real numbers such that*

$$\mu(\{x_n\}) = \mu(\{-x_n\}) = \tfrac{1}{2}c_n \text{ for } n = 1, 2, \dots ; \qquad \mu(\{0\}) = c_0; \qquad \sum_{n=0}^{\infty} c_n = 1.$$

Then a centered Gaussian process having μ as its spectral measure is

$$X(t) = c_0^{1/2}\xi_0 + \sum_{n=1}^{\infty} c_n^{1/2}(\xi_n \cos tx_n + \xi_{-n} \sin tx_n) \qquad t \in \mathbb{R}, \tag{1.8}$$

where the $\{\xi_n\}_{n\in Z}$ is a sequence of independent identically distributed random variables, each having a standard normal distribution. In fact, the series in (1.8) converges in $L^2(\Omega, \mathcal{F}, P)$ and

$$\mathrm{E}(X(t)) = 0 \quad \text{and} \quad \mathrm{E}(X(s)X(t)) = c_0 + \sum_{n=1}^{\infty} c_n \cos[(t-s)x_n] = \widehat{\mu}(t-s).$$

We use the notation

$$\lambda_k := \int_{\mathbb{R}} x^k \mu(dx) \qquad k = 0, 1, 2, \dots \tag{1.9}$$

whenever the integral exists. λ_k is the kth *spectral moment* of the process.

An extension of the preceding class of examples is the following. Let $(T, \mathcal{T}, \rho)$ be a measure space, $H = L^2_{\mathbb{R}}(T, \mathcal{T}, \rho)$ the Hilbert space of real-valued square-integrable functions on it, and $\{\varphi_n(t)\}_{n=1,2,\dots}$ an orthonormal sequence in H. We assume that each function $\varphi_n : T \to \mathbb{R}$ is bounded and denote $M_n = \sup_{t\in T} |\varphi_n(t)|$. In addition, let $\{c_n\}_{n=1,2,..}$ be a sequence of positive numbers such that

$$\sum_{n=1}^{\infty} c_n < \infty, \qquad \sum_{n=1}^{\infty} c_n M_n^2 < \infty$$

and $\{\xi_n\}_{n=1,2,\dots}$ a sequence of independent identically distributed (i.i.d.) random variables, each with standard normal distribution in $\mathbb{R}$.

Then the stochastic process

$$X(t) = \sum_{n=1}^{\infty} c_n^{1/2} \xi_n \varphi_n(t) \tag{1.10}$$

is Gaussian, centered with covariance

$$r(s,t) = \mathrm{E}(X(s)X(t)\} = \sum_{n=1}^{\infty} c_n \varphi_n(s)\varphi_n(t).$$

Formulas (1.8) and (1.10) are simple cases of spectral representations of Gaussian processes, which is an important subject for both theoretical purposes and for applications. A compact presentation of this subject, including the Karhunen–Loève representation and the connection with reproducing kernel Hilbert spaces, may be found in Fernique's lecture notes (1974).

1.3. 0–1 LAW FOR GAUSSIAN PROCESSES

We will prove a 0–1 law for Gaussian processes in this section without attempting full generality. This will be sufficient for our requirements in what follows. For a more general treatment, see Fernique (1974).

Definition 1.3. *Let* $\mathcal{X} = \{X(t) : t \in T\}$ *and* $\mathcal{Y} = \{Y(t) : t \in S\}$ *be real-valued stochastic processes defined on some probability space* $(\Omega, \mathcal{A}, P)$. $\mathcal{X}$ *and* $\mathcal{Y}$ *are said to be* independent *if for any choice of the parameter values* $t_1, \ldots, t_n \in T; s_1, \ldots, s_m \in S,\ n, m \geq 1$, *the random vectors*

$$(X(t_1), \ldots, X(t_n)),\ (Y(s_1), \ldots, Y(s_m))$$

are independent.

Proposition 1.4. *Let the processes* $\mathcal{X}$ *and* $\mathcal{Y}$ *be independent and* E *(respectively,* F*) belong to the* σ*-algebra generated by the cylinders in* $\mathbb{R}^T$ *(respectively,* $\mathbb{R}^S$*). Then*

$$\mathrm{P}(X(\cdot) \in E, Y(\cdot) \in F) = \mathrm{P}(X(\cdot) \in E)\mathrm{P}(Y(\cdot) \in F). \tag{1.11}$$

Proof. Equation (1.11) holds true for cylinders. Uniqueness in the extension theorem provides the result. □

Theorem 1.5 (0–1 Law for Gaussian Processes). *Let* $\mathcal{X} = \{X(t) : t \in T\}$ *be a real-valued centered Gaussian process defined on some probability space* $(\Omega, \mathcal{A}, \mathrm{P})$ *and* $(\mathrm{E}, \mathcal{E})$ *a measurable space, where* E *is a linear subspace of* $\mathbb{R}^T$ *and the* σ*-algebra* $\mathcal{E}$ *has the property that for any choice of the scalars* $a, b \in \mathbb{R}$, *the function* $(x, y) \rightsquigarrow ax + by$ *defined on* $\mathrm{E} \times \mathrm{E}$ *is measurable with respect to the product* σ*-algebra. We assume that the function* $X : \Omega \to \mathrm{E}$ *defined as* $X(\omega) = X(\cdot, \omega)$ *is measurable* $(\Omega, \mathcal{A}) \to (\mathrm{E}, \mathcal{E})$. *Then, if* L *is a measurable subspace of* E*, one has*

$$\mathrm{P}(X(\cdot) \in L) = 0 \quad or \quad 1.$$

Proof. Let $\{X^{(1)}(t) : t \in T\}$ and $\{X^{(2)}(t) : t \in T\}$ be two independent processes each having the same distribution as that of the given process $\{X(t) : t \in T\}$. For each λ, $0 < \lambda < \pi/2$, consider a new pair of stochastic processes, defined for $t \in T$ by

$$\begin{aligned} Z_\lambda^{(1)}(t) &= X^{(1)}(t)\cos\lambda + X^{(2)}(t)\sin\lambda \\ Z_\lambda^{(2)}(t) &= -X^{(1)}(t)\sin\lambda + X^{(2)}(t)\cos\lambda. \end{aligned} \tag{1.12}$$

Each of the processes $Z_\lambda^{(i)}(t)(i = 1, 2)$ has the same distribution as $\mathcal{X}$.

In fact, $\mathrm{E}(Z_\lambda^{(1)}(t)) = 0$ and since $\mathrm{E}(X^{(1)}(s)X^{(2)}(t)) = 0$, we have $\mathrm{E}(Z_\lambda^{(1)}(s)$ $Z_\lambda^{(1)}(t)) = \cos^2\lambda\ \mathrm{E}(X^{(1)}(s)X^{(1)}(t)) + \sin^2\lambda\ \mathrm{E}(X^{(2)}(s)X^{(2)}(t)) = \mathrm{E}(X(s)X(t))$.

A similar computation holds for $Z_\lambda^{(2)}$.

Also, the processes $Z_\lambda^{(1)}$ and $Z_\lambda^{(2)}$ are independent. To prove this, note that for any choice of $t_1, \dots, t_n; s_1, \dots, s_m, n, m \geq 1$, the random vectors

$$(Z_\lambda^{(1)}(t_1), \dots, Z_\lambda^{(1)}(t_n)), (Z_\lambda^{(2)}(s_1), \dots, Z_\lambda^{(2)}(s_m))$$

have a joint Gaussian distribution, so it suffices to show that

$$\mathrm{E}(Z_\lambda^{(1)}(t)Z_\lambda^{(2)}(s)) = 0$$

for any choice of $s, t \in T$ to conclude that they are independent. This is easily checked.

Now, if we put $q = \mathrm{P}(X(\cdot) \in L)$, independence implies that for any λ,

$$q(1-q) = \mathrm{P}(E_\lambda) \qquad \text{where} \quad E_\lambda = \{Z_\lambda^{(1)} \in L, Z_\lambda^{(2)} \notin L\}.$$

If $\lambda, \lambda' \in (0, \pi/2), \lambda \neq \lambda'$, the events E_λ and $E_{\lambda'}$ are disjoint. In fact, the matrix

$$\begin{pmatrix} \cos\lambda & \sin\lambda \\ \cos\lambda' & \sin\lambda' \end{pmatrix}$$

is nonsingular and (1.12) implies that if at the same time $Z_\lambda^{(1)} \in L, Z_{\lambda'}^{(1)} \in L$, then $X^{(1)}(\cdot), X^{(2)}(\cdot) \in L$ also, since $X^{(1)}(\cdot), X^{(2)}(\cdot)$ are linear combinations of $Z_\lambda^{(1)}$ and $Z_{\lambda'}^{(1)}$. Hence, $Z_\lambda^{(2)}, Z_{\lambda'}^{(2)} \in L$ and E_λ, $E_{\lambda'}$ cannot occur simultaneously. To finish, the only way in which we can have an infinite family $\{E_\lambda\}_{0<\lambda<\pi/2}$ of pairwise disjoint events with equal probability is for this probability to be zero. That is, $q(1-q) = 0$, so that $q = 0$ or 1. □

In case the parameter set T is countable, the above shows directly that any measurable linear subspace of $\mathbb{R}^T$ has probability 0 or 1 under a centered Gaussian law. If T is a σ-compact topological space, E the set of real-valued continuous functions defined on T, and $\mathcal{E}$ the σ-algebra generated by the topology of uniform convergence on compact sets, one can conclude, for example, that the subspace of E of bounded functions has probability 0 or 1 under a centered Gaussian measure. The theorem can be applied in a variety of situations similar to standard function spaces. For example, put a measure on the space (E, $\mathcal{E}$) and take for L an L^p of this measure space.

1.4. REGULARITY OF PATHS

1.4.1. Conditions for Continuity of Paths

Theorem 1.6 (Kolmogorov). *Let* $\mathcal{Y} = \{Y(t) : t \in [0, 1]\}$ *be a real-valued stochastic process that satisfies the condition*

(K) *For each pair* t *and* $t + h \in [0, 1]$,

$$\mathrm{P}\{|Y(t+h) - Y(t)| \geq \alpha(h)\} \leq \beta(h),$$

where α *and* β *are even real-valued functions defined on* $[-1, 1]$, *increasing on* $[0, 1]$, *that verify*

$$\sum_{n=1}^{\infty} \alpha(2^{-n}) < \infty, \qquad \sum_{n=1}^{\infty} 2^n \beta(2^{-n}) < \infty.$$

Then there exists a version $\mathcal{X} = \{X(t) : t \in T\}$ *of the process* $\mathcal{Y}$ *such that the paths* $t \rightsquigarrow X(t)$ *are continuous on* $[0, 1]$.

Proof. For $n = 1, 2, \ldots; k = 0, 1, \ldots, 2^n - 1$, let

$$E_{k,n} = \left\{ \left| Y\left(\frac{k+1}{2^n}\right) - Y\left(\frac{k}{2^n}\right) \right| \geq \alpha(2^{-n}) \right\}, \qquad E_n = \bigcup_{k=0}^{2^n-1} E_{k,n}.$$

From the hypothesis, $\mathrm{P}(E_n) \leq 2^n \beta(2^{-n})$, so that $\sum_{n=1}^{\infty} \mathrm{P}(E_n) < \infty$. The Borel–Cantelli lemma implies that $\mathrm{P}(\limsup_{n\to\infty} E_n) = 0$, where

$$\limsup_{n\to\infty} E_n = \{\omega : \omega \text{ belongs to infinitely many } E_n\text{'s}\}.$$

In other words, if $\omega \notin \limsup_{n\to\infty} E_n$, one can find $n_0(\omega)$ such that if $n \geq n_0(\omega)$, one has

$$\left| Y\left(\frac{k+1}{2^n}\right) - Y\left(\frac{k}{2^n}\right) \right| < \alpha(2^{-n}) \qquad \text{for all } k = 0, 1, \ldots, 2^n - 1.$$

Denote by $Y^{(n)}$ the function whose graph is the polygonal with vertices $(k/2^n, Y(k/2^n))$, $k = 0, 1, \ldots, 2^n$; that is, if $k/2^n \le t \le (k+1)/2^n$, one has

$$Y^{(n)}(t) = (k + 1 - 2^n t)Y\left(\frac{k}{2^n}\right) + (2^n t - k)Y\left(\frac{k+1}{2^n}\right).$$

The function $t \rightsquigarrow Y^{(n)}(t)$ is continuous. Now, if $\omega \notin \limsup_{n\to\infty} E_n$, one easily checks that there exists some integer $n_0(\omega)$ such that

$$\left\| Y^{(n+1)} - Y^{(n)} \right\|_\infty \le \alpha\left(2^{-(n+1)}\right) \qquad \text{for } n + 1 \ge n_0(\omega)$$

(here $\|\cdot\|_\infty$ denotes the sup norm on [0, 1]). Since $\sum_{n=1}^{\infty} \alpha(2^{-(n+1)}) < \infty$ by the hypothesis, the sequence of functions $\{Y^{(n)}\}$ converges uniformly on [0, 1] to a continuous limit function that we denote $X(t)$, $t \in [0, 1]$.

We set $X(t) \equiv 0$ when $\omega \in \limsup_{n\to\infty} E_n$. To finish the proof, it suffices to show that for each $t \in [0, 1]$, $\mathrm{P}(X(t) = Y(t)) = 1$.

- If t is a dyadic point, say $t = k/2^n$, then given the definition of the sequence of functions $Y^{(n)}$, it is clear that $Y^{(m)}(t) = Y(t)$ for $m \ge n$. Hence, for $\omega \notin \limsup_{n\to\infty} E_n$, one has $X(t) = \lim_{m\to\infty} Y^{(m)}(t) = Y(t)$. The result follows from $\mathrm{P}((\limsup_{n\to\infty} E_n)^C) = 1$ (A^C is the complement of the set A).
- If t is not a dyadic point, for each n, $n = 1, 2, \ldots$, let k_n be an integer such that $|t - k_n/2^n| \le 2^{-n}$, $k_n/2^n \in [0, 1]$. Set

  $$F_n = \left\{ \left| Y(t) - X\left(\frac{k_n}{2^n}\right) \right| \ge \alpha(2^{-n}) \right\}.$$

  We have the inequalities

  $$\mathrm{P}(F_n) \le \mathrm{P}\left(\left| Y(t) - X\left(\frac{k_n}{2^n}\right)\right| \ge \alpha\left(\left| t - \frac{k_n}{2^n}\right|\right)\right) \le \beta\left(\left| t - \frac{k_n}{2^n}\right|\right) \le \beta(2^{-n}),$$

  and a new application of the Borel–Cantelli lemma gives $\mathrm{P}(\limsup_{n\to\infty} F_n) = 0$. So if $\omega \notin [\limsup_{n\to\infty} E_n] \cup [\limsup_{n\to\infty} F_n]$, we have at the same time, $X(k_n/2^n)(\omega) \to X(t)(\omega)$ as $n \to \infty$ because $t \rightsquigarrow X(t)$ is continuous, and $X(k_n/2^n)(\omega) \to Y(t)(\omega)$ because $|Y(t) - X(k_n/2^n)| < \alpha(2^{-n})$ for $n \ge n_1(\omega)$ for some integer $n_1(\omega)$.

This proves that $X(t)(\omega) = Y(t)(\omega)$ for almost every ω. □

Corollary 1.7 *Assume that the process $\mathcal{Y} = \{Y(t) : t \in [0, 1]\}$ satisfies one of the following conditions for $t, t + h \in [0, 1]$:*

(a)

$$\mathrm{E}(|Y(t+h)-Y(t)|^p) \leq \frac{K|h|}{|\log|h||^{1+r}}, \tag{1.13}$$

where p,r, and K are positive constants, $p < r$.

(b) $\mathcal{Y}$ *is Gaussian,* $m(t) := \mathrm{E}(Y(t))$ *is continuous, and*

$$\mathrm{Var}(Y(t+h)-Y(t)) \leq \frac{C}{|\log|h||^{a}} \tag{1.14}$$

for all t, sufficiently small h, C some positive constant, and $a > 3$.

Then the conclusion of Theorem 1.6 holds.

Proof

(a) Set

$$\alpha(h) = \frac{1}{|\log|h||^b} \qquad 1 < b < \frac{r}{p}$$

$$\beta(h) = \frac{|h|}{|\log|h||^{1+r-bp}}$$

and check condition (K) using a Markov inequality.

(b) Since the expectation is continuous, it can be subtracted from $Y(t)$, so that we may assume that $\mathcal{Y}$ is centered. To apply Theorem 1.6, take

$$\alpha(h) = \frac{1}{|\log|h||^b} \text{ with } 1 < b < (a-1)/2 \text{ and } \beta(h) = \exp\left[-\frac{1}{4C}|\log|h||^{a-2b}\right].$$

Then

$$\mathrm{P}(|Y(t+h)-Y(t)| \geq \alpha(h)) = \mathrm{P}\left(|\xi| \geq \frac{\alpha(h)}{\sqrt{\mathrm{Var}(Y(t+h)-Y(t))}}\right),$$

where ξ stands for standard normal variable. We use the following usual bound for Gaussian tails, valid for $u > 0$:

$$\mathrm{P}(|\xi| \geq u) = 2\mathrm{P}(\xi \geq u) = \sqrt{\frac{2}{\pi}}\int_u^{+\infty} e^{-(1/2)x^2}dx \leq \sqrt{\frac{2}{\pi}}\frac{1}{u}e^{-(1/2)u^2}.$$

With the foregoing choice of $\alpha(\cdot)$ and $\beta(\cdot)$, if $|h|$ is small enough, one has $\alpha(h)/\sqrt{\mathrm{Var}(Y(t+h)-Y(t))} > 1$ and

$$\mathrm{P}(|Y_{t+h} - Y(t)| \geq \alpha(h)) \leq (\mathrm{const})\,\beta(h).$$

where (const) denotes a generic constant that may vary from line to line. On the other hand, $\sum_1^\infty \alpha(2^{-n}) < \infty$ and $\sum_1^\infty 2^n\beta(2^{-n}) < \infty$ are easily verified. □

Some Examples

1. *Gaussian stationary processes.* Let $\{Y(t) : t \in \mathbb{R}\}$ be a real-valued Gaussian centered stationary process with covariance $\Gamma(\tau) = \mathrm{E}(Y(t)\,Y(t+\tau))$. Then condition (1.14) is equivalent to

$$\Gamma(0) - \Gamma(\tau) \leq \frac{C}{|\log|\tau||^a}$$

for sufficiently small $|\tau|$, with the same meaning for C and a.

2. *Wiener process.* Take $T = \mathbb{R}^+$. The function $r(s,t) = s \wedge t$ is positive semidefinite. In fact, if $0 \leq s_1 < \cdots < s_n$ and $x_1, \ldots, x_n \in \mathbb{R}$, one has

$$\sum_{j,k=1}^{n} (s_j \wedge s_k)\, x_j x_k = \sum_{k=1}^{n} (s_k - s_{k-1})(x_k + \cdots + x_n)^2 \geq 0, \qquad (1.15)$$

where we have set $s_0 = 0$.

Then, according to Kolmogorov's extension theorem, there exists a centered Gaussian process $\{Y(t) : t \in \mathbb{R}^+\}$ such that $\mathrm{E}(Y(s)Y(t)) = s \wedge t$ for $s, t \geq 0$. One easily checks that this process satisfies the hypothesis in Corollary 1.7(b), since the random variable $Y(t+h) - Y(t)$, $h \geq 0$ has the normal distribution $N(0,h)$ because of the simple computation

$$\mathrm{E}([Y(t+h) - Y(t)]^2) = t + h - 2t + t = h.$$

It follows from Corollary 1.7(b) that this process has a continuous version on every interval of the form $[n, n+1]$. The reader will verify that one can also find a version with continuous paths defined on all $\mathbb{R}^+$. This version, called the *Wiener process*, is denoted $\{W(t) : t \in \mathbb{R}^+\}$.

3. *Ito integrals.* Let $\{W(t) : t \geq 0\}$ be a Wiener process on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. We define the *filtration* $\{\mathcal{F}_t : t \geq 0\}$ as $\mathcal{F}_t = \tilde{\sigma}\{W(s) : s \leq t\}$, where the notation means the σ-algebra generated by the set of random variables $\{W(s) : s \leq t\}$ (i.e., the smallest σ-algebra with respect to which these random variables are all measurable) completed with respect to the probability measure P.

Let $\{a_t : t \geq 0\}$ be a stochastic process adapted to the filtration $\{\mathcal{F}_t : t \geq 0\}$. This means that a_t is $\mathcal{F}_t$-measurable for each $t \geq 0$. For simplicity we assume that $\{a_t : t \geq 0\}$ is uniformly locally bounded in the sense that for each $T > 0$ there exists a constant C_T such that $|a_t(\omega)| \leq C_T$ for every ω and all $t \in [0, T]$. For each $t > 0$, one can define the stochastic Ito integral

$$Y(t) = \int_0^t a_s \, dW(s)$$

as the limit in $L^2 = L^2(\Omega, \mathcal{A}, \mathrm{P})$ of the Riemann sums

$$S_Q = \sum_{j=0}^{m-1} \widetilde{a}_{t_j} (W(t_{j+1}) - W(t_j))$$

when $N_Q = \sup\{(t_{j+1} - t_j) : 0 \leq j \leq m-1\}$ tends to 0. Here Q denotes the partition $0 = t_0 < t_1 < \cdots < t_m = t$ of the interval $[0, t]$ and $\{\widetilde{a}_t : t \geq 0\}$ an adapted stochastic process, bounded by the same constant as $\{a_t : t \geq 0\}$ and such that

$$\sum_{j=0}^{m-1} \widetilde{a}_{t_j} \mathbb{1}_{\{t_j \leq s < t_{j+1}\}}$$

tends to $\{a_t : 0 \leq s \leq t\}$ in the space $L^2([0, t] \times \Omega, \lambda \times \mathrm{P})$ as $N_Q \to 0$. λ is a Lebesgue measure on the line.

Of course, the statements above should be proved to be able to define $Y(t)$ in this way (see, e.g., McKean, 1969). Our aim here is to prove that the process $\{Y(t) : t \geq 0\}$ thus defined has a version with continuous paths. With no loss of generality, we assume that t varies on the interval $[0, 1]$ and apply Corollary 1.7(a) with $p = 4$.

We will prove that

$$\mathrm{E}((Y(t+h) - Y(t))^4) \leq (\text{const})h^2.$$

For this, it is sufficient to see that if Q is a partition of the interval $[t, t+h]$, $h > 0$,

$$\mathrm{E}(S_Q^4) \leq (\text{const})h^2, \tag{1.16}$$

where (const) does not depend on t, h, and Q, and then apply Fatou's lemma when $N_Q \to 0$.

Let us compute the left-hand side of (1.16). Set $\Delta_j = W(t_{j+1}) - W(t_j)$. We have

$$\mathrm{E}(S_Q^4) = \sum_{j_1, j_2, j_3, j_4 = 0}^{m-1} \mathrm{E}(\widetilde{a}_{t_{j_1}} \widetilde{a}_{t_{j_2}} \widetilde{a}_{t_{j_3}} \widetilde{a}_{t_{j_4}} \Delta_{j_1} \Delta_{j_2} \Delta_{j_3} \Delta_{j_4}). \tag{1.17}$$

If one of the indices, say j_4, satisfies $j_4 > j_1, j_2, j_3$, the corresponding term becomes

$$\begin{aligned}\mathrm{E}\left(\prod_{h=1}^{4}(\widetilde{a}_{t_{j_h}} \Delta_{j_h})\right) &= \mathrm{E}\left(\mathrm{E}\left(\prod_{h=1}^{4}(\widetilde{a}_{t_{j_h}} \Delta_{j_h}) | \mathcal{F}_{t_{j_4}}\right)\right) \\ &= \mathrm{E}\left(\prod_{h=1}^{3}(\widetilde{a}_{t_{j_h}} \Delta_{j_h}) \widetilde{a}_{t_{j_4}} \mathrm{E}(\Delta_{j_4} | \mathcal{F}_{t_{j_4}})\right) = 0\end{aligned}$$

since

$$\mathrm{E}(\Delta_j | \mathcal{F}_{t_j}) = \mathrm{E}(\Delta_j) = 0 \quad \text{and} \quad \prod_{h=1}^{3}(\widetilde{a}_{t_{j_h}} \Delta_{j_h}) \widetilde{a}_{t_{j_4}} \ \text{is} \mathcal{F}_{t_{j_4}} - \text{measurable}.$$

In a similar way, if $j_4 < j_1 = j_2 = j_3$ (and similarly, if any one of the indices is strictly smaller than the others and these are all equal), the corresponding term vanishes since in this case

$$\begin{aligned} \mathrm{E}\left(\prod_{h=1}^{4}\left(\widetilde{a}_{t_{j_h}}\Delta_{j_h}\right)\right) &= \mathrm{E}\left(\mathrm{E}\left(\left(\widetilde{a}_{t_{j_1}}\Delta_{j_1}\right)^3\widetilde{a}_{t_{j_4}}\Delta_{j_4}|\mathcal{F}_{t_{j_1}}\right)\right) \\ &= \mathrm{E}\left(\widetilde{a}^3_{t_{j_1}}\widetilde{a}_{t_{j_4}}\Delta_{j_4}\mathrm{E}\left(\Delta^3_{j_1}|\mathcal{F}_{t_{j_1}}\right)\right) = 0 \end{aligned}$$

because

$$\mathrm{E}\left(\Delta_j^3|\mathcal{F}_{t_j}\right) = \mathrm{E}\left(\Delta_j^3\right) = 0.$$

The terms with $j_1 = j_2 = j_3 = j_4$ give the sum

$$\sum_{j=0}^{m-1}\mathrm{E}\left(\left(\widetilde{a}_{t_j}\Delta_j\right)^4\right) \le C_1^4\sum_{j=0}^{m-1}3\,(t_{j+1}-t_j)^2 \le 3\,C_1^4\,h^2.$$

Finally, we have the sum of the terms corresponding to 4-tuples of indices j_1, j_2, j_3, and j_4 such that for some permutation (i_1, i_2, i_3, i_4) of $(1, 2, 3, 4)$, one has $j_{i_1}, j_{i_2} < j_{i_3} = j_{i_4}$. This is

$$6\sum_{j_3=1}^{m-1}\sum_{0\le j_1, j_2<j_3}\mathrm{E}\left(\widetilde{a}_{t_{j_1}}\widetilde{a}_{t_{j_2}}\widetilde{a}^2_{t_{j_3}}\Delta_{j_1}\Delta_{j_2}\Delta^2_{j_3}\right).$$

Conditioning on $\mathcal{F}_{t_{j_3}}$ in each term yields for this sum

$$\begin{aligned} &6\sum_{j_3=1}^{m-1}\sum_{0\le j_1, j_2<j_3}(t_{j_3+1}-t_{j_3})\mathrm{E}\left(\widetilde{a}_{t_{j_1}}\widetilde{a}_{t_{j_2}}\widetilde{a}^2_{t_{j_3}}\Delta_{j_1}\Delta_{j_2}\right) \\ &\qquad = 6\mathrm{E}\left(\sum_{j_3=1}^{m-1}(t_{j_3+1}-t_{j_3})\widetilde{a}^2_{t_{j_3}}\left(\sum_{j=0}^{j_3-1}\widetilde{a}_{t_j}\Delta_j\right)^2\right) \\ &\qquad \le 6\,C_1^2\sum_{j_3=1}^{m-1}(t_{j_3+1}-t_{j_3})\mathrm{E}\left(\left(\sum_{j=0}^{j_3-1}\widetilde{a}_{t_j}\Delta_j\right)^2\right) \\ &\qquad = 6\,C_1^2\sum_{j_3=1}^{m-1}(t_{j_3+1}-t_{j_3})\sum_{j=0}^{j_3-1}\mathrm{E}\left(\widetilde{a}^2_{t_j}\right)(t_{j+1}-t_j) \le 3C_1^4\,h^2. \end{aligned}$$

Using (1.17), one obtains (1.16), and hence the existence of a version of the Itô integral possessing continuous paths.

Separability. Next, we consider the separability of stochastic processes. The separability condition is shaped to avoid the measurability problems that we have already mentioned and to use, without further reference, versions of stochastic processes having good path properties. We begin with a definition.

Definition 1.8. *We say that a real-valued stochastic process* $\{X(t) : t \in T\}$, T *a topological space, is* separable *if there exists a fixed countable subset* D *of* T *such that with probability 1,*

$$\sup_{t\in V\cap D} X(t) = \sup_{t\in V} X(t) \quad \textit{and} \quad \inf_{t\in V\cap D} X(t) = \inf_{t\in V} X(t) \ \ \textit{for all open sets } V.$$

A consequence of Theorem 1.6 is the following:

Proposition 1.9. *Let* $\{Y(t) : t \in I\}$, I *an interval in the line, be a separable random process that satisfies the hypotheses of Theorem 1.6. Then, almost surely (a.s.), its paths are continuous.*

Proof. Denote by D the countable set in the definition of separability. With no loss of generality, we may assume that D is dense in I. The theorem states that there exists a version $\{X(t) : t \in I\}$ that has continuous paths, so that

$$\mathrm{P}(X(t) = Y(t) \text{ for all } t \in D) = 1.$$

Let

$$E = \{X(t) = Y(t) \text{ for all } t \in D\}$$

and

$$F = \bigcap_{J\subset I, J=(r_1,r_2), r_1,r_2\in Q} \left\{ \sup_{t\in J\cap D} Y(t) = \sup_{t\in J} Y(t) \quad \text{and} \quad \inf_{t\in J\cap D} Y(t) = \inf_{t\in J} Y(t) \right\}.$$

Since $\mathrm{P}(E \cap F) = 1$, it is sufficient to prove that if $\omega \in E \cap F$, then $X(s)(\omega) = Y(s)(\omega)$ for all $s \in I$.

So, let $\omega \in E \cap F$ and $s \in I$. For any $\varepsilon > 0$, choose $r_1, r_2 \in Q$ such that

$$s - \varepsilon < r_1 < s < r_2 < s + \varepsilon.$$

Then, setting $J = (r_1, r_2)$,

$$Y(s)(\omega) \le \sup_{t\in J} Y(t)(\omega) = \sup_{t\in J\cap D} Y(t)(\omega) = \sup_{t\in J\cap D} X(t)(\omega) \le \sup_{t\in J} X(t)(\omega).$$

Letting $\varepsilon \to 0$, it follows that

$$Y(s)(\omega) \leq \limsup_{t \to s} X(t)(\omega) = X(s)(\omega)$$

since $t \rightsquigarrow X(t)(\omega)$ is continuous.

In a similar way, one proves that $Y(s)(\omega) \geq X(s)(\omega)$. □

The separability condition is usually met when the paths have some minimal regularity (see Exercise 1.7). For example, if $\{X(t) : t \in \mathbb{R}\}$ is a real-valued process having a.s. *càd-làg paths* (i.e., paths that are right-continuous with left limits), it is separable. All processes considered in the sequel are separable.

Some Additional Remarks and References. A reference for Kolmogorov's extension theorem and the regularity of paths, at the level of generality we have considered here, is the book by Cramér and Leadbetter (1967), where the reader can find proofs that we have skipped as well as related results, examples, and details. For d-parameter Gaussian processes, a subject that we consider in more detail in Chapter 6, in the stationary case, necessary and sufficient conditions to have continuous paths are due to Fernique (see his St. Flour 1974 lecture notes) and to Talagrand (1987) in the general nonstationary case. In the Gaussian stationary case, Belayev (1961) has shown that either: with probability 1 the paths are continuous, or with probability 1 the supremum (respectively, the infimum) on every interval is $+\infty$ (respectively, $-\infty$). General references on Gaussian processes are the books by Adler (1990) and Lifshits (1995).

1.4.2. Sample Path Differentiability and Hölder Conditions

In this section we state some results, without detailed proofs. These follow the lines of the preceding section.

Theorem 1.10. *Let $\mathcal{Y} = \{Y(t) : t \in [0, 1]\}$ be a real-valued stochastic process that satisfies the hypotheses of Theorem 1.6 and additionally, for any triplet $t - h$, t, $t + h \in [0, 1]$, one has*

$$\mathrm{P}(|Y(t+h) + Y(t-h) - 2Y(t)| \geq \alpha_1(h)) \leq \beta_1(h),$$

where α_1 and β_1 are two even functions, increasing for $h > 0$ and such that

$$\sum_{n=1}^{\infty} 2^n \alpha_1(2^{-n}) < \infty, \qquad \sum_{n=1}^{\infty} 2^n \beta_1(2^{-n}) < \infty.$$

Then there exists a version $\mathcal{X} = \{X(t) : t \in T\}$ of the process $\mathcal{Y}$ such that almost surely the paths of $\mathcal{X}$ are of class C^1.

Sketch of the Proof. Consider the sequence $\{Y^{(n)}(t) : t \in [0, 1]\}_{n=1,2,\ldots}$ of polygonal processes introduced in the proof of Theorem 1.6. We know that a.s. this sequence converges uniformly to $\mathcal{X} = \{X(t) : t \in [0, 1]\}$, a continuous version of $\mathcal{Y}$. Define:

$$\widetilde{Y}^{(n)}(t) := Y^{(n)\prime}(t^-) \quad \text{for } 0 < t \leq 1 \text{ (left derivative)}$$
$$\widetilde{Y}^{(n)}(0) := Y^{(n)\prime}(0^+) \quad \text{(right derivative)}.$$

One can show that the hypotheses imply:

1. Almost surely, as $n \to \infty$, $\widetilde{Y}^{(n)}(\cdot)$ converges uniformly on [0, 1] to a function $\widetilde{X}(\cdot)$.
2. Almost surely, as $n \to \infty$, $\sup_{t\in[0,1]}|\widetilde{Y}^{(n)}(t^+) - \widetilde{Y}^{(n)}(t)| \to 0$.

To complete the proof, check that the function $t \rightsquigarrow \widetilde{X}(t)$ a.s. is continuous and coincides with the derivative of $X(t)$ at every $t \in [0, 1]$. □

Example 1.2 (Stationary Gaussian Processes). *Let $\mathcal{Y} = \{Y(t) : t \in \mathbb{R}\}$ be a centered stationary Gaussian process with covariance of the form*

$$\Gamma(\tau) = \mathrm{E}(Y(t)Y(t+\tau)) = \Gamma(0) - \frac{1}{2}\lambda_2\tau^2 + O\left(\frac{\tau^2}{|\log|\tau||^a}\right)$$

with $\lambda_2 > 0$, $a > 3$. Then there exists a version of $\mathcal{Y}$ with paths of class C^1. For the proof, apply Theorem 1.10.

A related result is the following. The proof is left to the reader.

Proposition 1.11 (Hölder Conditions). *Assume that*

$$\mathrm{E}(|Y(t+h) - Y(t)|^p) \leq K|h|^{1+r} \qquad \textit{for } t, t+h \in [0, 1], \tag{1.18}$$

where K, p, and r are positive constants, $r \leq p$. Then there exists a version of the process $\mathcal{Y} = \{Y(t) : t \in [0, 1]\}$ with paths that satisfy a Hölder condition with exponent α for any α such that $0 < \alpha < r/p$.

Note that, for example, this proposition can be applied to the Wiener process (Brownian motion) with $r = (p-2)/2$, showing that it satisfies a Hölder condition for every $\alpha < \frac{1}{2}$.

1.4.3. Higher Derivatives

Let $\mathcal{X} = \{X(t) : t \in \mathbb{R}\}$ be a stochastic process and assume that for each $t \in \mathbb{R}$, one has $X(t) \in L^2(\Omega, \mathcal{A}, \mathrm{P})$.

Definition 1.12. *$\mathcal{X}$ is* differentiable in quadratic mean *(q.m.) if for all $t \in \mathbb{R}$,*

$$\frac{X(t+h) - X(t)}{h}$$

converges in quadratic mean as $h \to 0$ to some limit that will be denoted $X'(t)$.

The stability of Gaussian random variables under passage to the limit implies that the derivative in q.m. of a Gaussian process remains Gaussian.

Proposition 1.13. *Let $\mathcal{X} = \{X(t) : t \in \mathbb{R}\}$ be a stochastic process with mean $m(t)$ and covariance $r(s,t)$ and suppose that m is $\mathcal{C}^1$ and that r is $\mathcal{C}^2$. Then $\mathcal{X}$ is differentiable in the quadratic mean.*

Proof. We use the following result, which is easy to prove: The sequence $Z_1, \dots, Z_n$ of real random variables converges in q.m. if and only if there exists a constant C such that $\mathrm{E}(Z_m Z_n) \to C$ as the pair (m, n) tends to infinity. Since $m(t)$ is differentiable, it can be substracted from $X(t)$ without changing its differentiability, so we can assume that the process is centered. Then for all real h and k,

$$\begin{aligned} &\mathrm{E}\left(\frac{X(t+h) - X(t)}{h} \frac{X(t+k) - X(t)}{k}\right) \\ &\qquad = \frac{1}{hk}\big[r(t+h, t+k) - r(t, t+k) - r(t, t+h) + r(t,t)\big] \\ &\qquad\qquad \to r_{11}(t,t) \quad \text{as } (k,h) \to (0,0), \end{aligned}$$

where $r_{11}(s,t) := \partial^2 r(s,t)/\partial s \partial t$. This shows differentiability in q.m. □

We assume, using the remark in the proof above, that $\mathcal{X}$ is centered and satisfies the conditions of the proposition. It is easy to prove that

$$\mathrm{E}(X(s)X'(t)) = r_{01}(s,t) := \frac{\partial r}{\partial t}(s,t),$$

and similarly, that the covariance of $\mathcal{X}' = \{X'(t) : t \in \mathbb{R}\}$ is $r_{11}(s,t)$. Now let $\mathcal{X}$ be a Gaussian process and $\mathcal{X}'$ its derivative in quadratic mean. If this satisfies, for example, the criterion in Corollary 1.7(b), it admits a continuous version $\mathcal{Y}' = \{Y'(t) : Y'(t); t \in \mathbb{R}\}$. Set

$$Y(t) := X(0) + \int_0^t Y'(s)\, ds.$$

Clearly, $\mathcal{Y}$ has $\mathcal{C}^1$-paths and $\mathrm{E}(X(s), Y(s)) = r(s, 0) + \int_0^s r_{01}(s, t)\, dt = r(s, s)$. In the same way, $\mathrm{E}(Y(s)^2) = r(s, s)$, so that $\mathrm{E}([X(s) - Y(s)]^2) = 0$. As a consequence, $\mathcal{X}$ admits a version with $\mathcal{C}^1$ paths.

Using this construction inductively, one can prove the following:

- Let $\mathcal{X}$ be a Gaussian process with mean $\mathcal{C}^k$ and covariance $\mathcal{C}^{2k}$ and such that its kth derivative in quadratic mean satisfies the weak condition of Corollary 1.7(b). Then $\mathcal{X}$ admits a version with paths of class $\mathcal{C}^k$.
- If $\mathcal{X}$ is a Gaussian process with mean of class $\mathcal{C}^\infty$ and covariance of class $\mathcal{C}^\infty$, $\mathcal{X}$ admits a version with paths of class $\mathcal{C}^\infty$.

In the converse direction, regularity of the paths implies regularity of the expectation and of the covariance function. For example, if $\mathcal{X}$ has continuous sample paths, the mean and the variance are continuous. In fact, if $t_n, n = 1, 2, \ldots$ converges to t, then $X(t_n)$ converges a.s. to $X(t)$, hence also in distribution. Using the form of the Fourier transform of the Gaussian distribution, one easily proves that this implies convergence of the mean and the variance. Since for Gaussian variables, all the moments are polynomial functions of the mean and the variance, they are also continuous. If the process has differentiable sample paths, in a similar way one shows the convergence

$$\frac{m(t+h) - m(t)}{h} \to \mathrm{E}(X'(t))$$

as $h \to 0$, showing that the mean is differentiable.

For the covariance, restricting ourselves to stationary Gaussian processes defined on the real line, without loss of generality we may assume that the process is centered. Put $\Gamma(t) = r(s, s+t)$. The convergence in distribution of $(X(h) - X(0))/h$ to $X'(0)$ plus the Gaussianity imply that $\mathrm{Var}\big((X(h) - X(0))/h\big)$ has a finite limit as $h \to 0$. On the other hand,

$$\mathrm{Var}\left(\frac{X(h) - X(0)}{h}\right) = 2 \int_{-\infty}^{+\infty} \frac{1 - \cos hx}{h^2} \mu(dx),$$

where μ is the spectral measure.

Letting $h \to 0$ and applying Fatou's lemma, it follows that

$$\lambda_2 = \int_{-\infty}^{+\infty} x^2 \mu(dx) \le \liminf_{h \to 0} \mathrm{Var}\left(\frac{X(h) - X(0)}{h}\right) < \infty.$$

Using the result in Exercise 1.4, Γ is of class $\mathcal{C}^2$.

This argument can be used in a similar form to show that if the process has paths of class $\mathcal{C}^k$, the covariance is of class $\mathcal{C}^{2k}$. As a conclusion, roughly speaking, for Gaussian stationary processes, the order of differentiability of the sample paths is half of the order of differentiability of the covariance.

1.4.4. More General Tools

In this section we consider the case when the parameter of the process lies in $\mathbb{R}^d$ or, more generally, in some general metric space. We begin with an extension of Theorem 1.6.

Theorem 1.14. *Let $\mathcal{Y} = \{Y(t) : t \in [0, 1]^d\}$ be a real-valued random field that satisfies the condition*

(K_d) *For each pair $t, t + h \in [0, 1]^d$,*

$$P\{|Y(t + h) - Y(t)| \geq \alpha(\bar{h})\} \leq \beta(\bar{h}),$$

where $h = (h_1, \ldots, h_d)$, $\bar{h} = \sup_{1 \leq i \leq d} |h_i|$, and α, β are even real-valued functions defined on $[-1, 1]$, increasing on $[0, 1]$, which verify

$$\sum_{n=1}^{\infty} \alpha(2^{-n}) < \infty, \qquad \sum_{n=1}^{\infty} 2^{dn} \beta(2^{-n}) < \infty.$$

Then there exists a version $\mathcal{X} = \{X(t) : t \in [0, 1]^d\}$ of the process $\mathcal{Y}$ such that the paths $t \rightsquigarrow X(t)$ are continuous on $[0, 1]^d$.

Proof. The main change with respect to the proof of Theorem 1.6 is that we replace the polygonal approximation, adapted to one-variable functions by another interpolating procedure. Denote by $\mathcal{D}_n$ the set of dyadic points of order n in $[0, 1]^d$; that is,

$$\mathcal{D}_n = \left\{ t = (t_1, \ldots, t_d) : t_i = \frac{k_i}{2^n}, k_i \text{ integers }, 0 \leq k_i \leq 2^n, i = 1, \ldots, d \right\}.$$

Let $f : [0, 1]^d \to \mathbb{R}$ be a function. For each $n = 1, 2, \ldots$, one can construct a function $f^{(n)} : [0, 1]^d \to \mathbb{R}$ with the following properties:

- $f^{(n)}$ is continuous.
- $f^{(n)}(t) = f(t)$ for all $t \in \mathcal{D}_n$.
- $\|f^{(n+1)} - f^{(n)}\|_\infty = \max_{t \in \mathcal{D}_{n+1} \setminus \mathcal{D}_n} |f(t) - f^{(n)}(t)|$, where $\| \cdot \|_\infty$ denotes sup-norm on $[0, 1]^d$.

A way to define $f^{(n)}$ is the following: Let us consider a cube $\mathcal{C}_{t,n}$ of the nth-order partition of $[0, 1]^d$; that is,

$$\mathcal{C}_{t,n} = t + \left[0, \frac{1}{2^n}\right]^d,$$

where $t \in \mathcal{D}_n$ with the obvious notation for the sum. For each vertex τ, set

$$f^{(n)}(\tau) = f(\tau).$$

Now, for each permutation π of $\{1, 2, \ldots, d\}$, let $\mathcal{S}_\pi$ be the simplex

$$\mathcal{S}_\pi = \left\{ t + s : s = (s_{\pi(1)}, \ldots, s_{\pi(d)}), 0 \le s_{\pi(1)} \le \cdots \le s_{\pi(d)} \le \frac{1}{2^n} \right\}.$$

It is clear that $\mathcal{C}_{t,n}$ is the union of the $\mathcal{S}_\pi$'s over all permutations. In a unique way, extend $f^{(n)}$ to $\mathcal{S}_\pi$ as an affine function. It is then easy to verify the afore mentioned properties and that

$$\| f^{(n+1)} - f^{(n)} \|_\infty \le d \sup_{s,t \in \mathcal{D}_{n+1}, |t-s| = 2^{-(n+1)}} | f(s) - f(t)|.$$

The remainder of the proof is essentially similar to that of Theorem 1.6. □

From this we deduce easily

Corollary 1.15. *Assume that the process $\mathcal{Y} = \{Y(t) : t \in [0, 1]^d\}$ verifies one of two conditions:*

(a)
$$\mathrm{E}(|Y(t+h) - Y(t)|^p) \le \frac{K_d |h|^d}{|\log |h||^{1+r}}, \tag{1.19}$$

where p, r, and K are positive constants, $p < r$.

(b) *If $\mathcal{Y}$ is Gaussian, $m(t) = \mathrm{E}(Y(t))$ is continuous and*

$$\mathrm{Var}(Y(t+h) - Y(t)) \le \frac{C}{|\log |h||^a} \tag{1.20}$$

for all t and sufficiently small h and $a > 3$.

Then the process has a version with continuous paths.

Note that the case of processes with values in $\mathbb{R}^{d'}$ need not to be considered separately, since continuity can be addressed coordinate by coordinate. For Hölder regularity we have

Proposition 1.16. *Let $\mathcal{Y} = \{Y(t) : t \in [0, 1]^d\}$ be a real-valued stochastic process with continuous paths such that for some $q > 1$, $\alpha > 0$,*

$$\mathrm{E}\big(|Y(s) - Y(t)|^q\big) \le (\mathrm{const}) \|s - t\|^{d+\alpha}.$$

Then almost surely, $\mathcal{Y}$ has Hölder paths with exponent $\alpha/2q$.

Until now, we have deliberately chosen elementary methods that apply to general random processes, not necessarily Gaussian. In the Gaussian case, even when the parameter varies in a set that does not have a restricted geometric structure, the question of continuity can be addressed using specific methods. As we have remarked several times already, we only need to consider centered processes.

Let $\{X(t) : t \in T\}$ be a centered Gaussian process taking values in $\mathbb{R}$. We assume that T is some metric space with distance denoted by τ. On T we define the canonical distance d,

$$d(s,t) := \sqrt{\mathrm{E}(X(t) - X(s))^2}.$$

In fact, d is a pseudodistance because two distinct points can be at d distance zero. A first point is that when the covariance $r(s,t)$ function is τ-continuous, which is the only relevant case (otherwise there is no hope of having continuous paths), d-continuity and τ-continuity are equivalent. The reader is referred to Adler (1990) for complements and proofs.

Definition 1.17. *Let (T, d) be a metric space. For $\varepsilon > 0$ denote by $N(\varepsilon) = N(T, d, \varepsilon)$ the minimum number of closed balls of radius ε with which we can cover T (the value of N_ε can be $+\infty$).*

We have the following theorem:

Theorem 1.18 (Dudley, 1973). *A sufficient condition for $\{X(t) : t \in T\}$ to have continuous sample paths is*

$$\int_0^{+\infty} \left(\log(N(\varepsilon))\right)^{1/2} d\varepsilon < \infty.$$

$\log(N(\varepsilon))$ *is called the* entropy *of the set T.*

A very important fact is that this condition is necessary in some relevant cases:

Theorem 1.19 (Fernique, 1974). *Let $\{X(t) : t \in T\}$, T compact, a subset of $\mathbb{R}^d$, be a stationary Gaussian process. Then the following three statements are equivalent:*

- *Almost surely, $X(\cdot)$ is bounded.*
- *Almost surely, $X(\cdot)$ is continuous.*
- $\displaystyle\int_0^{+\infty} \left(\log(N(\varepsilon))\right)^{1/2} d\varepsilon < \infty.$

This condition can be compared with Kolmogorov's theorem. The reader can check that Theorem 1.19 permits us to weaken the condition of Corollary 1.7(b) to $a > 1$. On the other hand, one can construct counterexamples (i.e., processes not having continuous paths) such that (1.14) holds true with $a = 1$. This shows that the condition of Corollary 1.7(b) is nearly optimal and sufficient for most applications. When the Gaussian process is no longer stationary, M. Talagrand has given necessary and sufficient conditions for sample path continuity in terms of the existence of majorizing measures (see Talagrand, 1987).

The problem of differentiability can be addressed in the same manner as for $d = 1$. A sufficient condition for a Gaussian process to have a version with $\mathcal{C}^k$ sample paths is for its mean to be $\mathcal{C}^k$, its covariance $\mathcal{C}^{2k}$, and its kth derivative in quadratic mean to satisfy some of the criteria of continuity above.

1.4.5. Tangencies and Local Extrema

In this section we give two classical results that are used several times in the book. The first gives a simple sufficient condition for a one-parameter random process not to have a.s. critical points at a certain specified level. The second result states that under mild conditions, a Gaussian process defined on a quite general parameter set with probability 1 does not have local extrema at a given level. We will use systematically the following notation: If ξ is a random variable with values in $\mathbb{R}^d$ and its distribution has a density with respect to Lebesgue measure, this density is denoted as

$$p_\xi(x) \qquad x \in \mathbb{R}^d.$$

Proposition 1.20 (Bulinskaya, 1961). *Let $\{X(t) : t \in I\}$ be a stochastic process with paths of class C^1 defined on the interval I of the real line. Assume that for each $t \in I$, the random variable $X(t)$ has a density $p_{X(t)}(x)$ which is bounded as t varies in a compact subset of I and x in a neighborhood v of $u \in \mathbb{R}$. Then*

$$\mathrm{P}(T_u^X \neq \emptyset) = 0,$$

where $T_u^X = \{t : t \in I, X(t) = u, X'(t) = 0\}$ is the set of critical points with value u of the random path $X(\cdot)$.

Proof. It suffices to prove that $\mathrm{P}(T_u^X \cap J \neq \emptyset) = 0$ for any compact subinterval J of I. Let ℓ be the length of J and $t_0 < t_1 < \cdots < t_m$ be a uniform partition of J (i.e., $t_{j+1} - t_j = \ell/m$ for $j = 0, 1, \cdots, m-1$). Denote by $\omega_{X'}(\delta, J)$ the modulus of continuity X' on the interval J and $E_{\delta,\varepsilon}$ the event

$$E_{\delta,\varepsilon} = \{\omega_{X'}(\delta, J) \geq \varepsilon\}.$$

Let $\varepsilon > 0$ be given; choose $\delta > 0$ so that $\mathrm{P}(E_{\delta,\varepsilon}) < \varepsilon$ and m so that $\ell/m < \delta$, and $[u - l/m, u + l/m] \subset v$. We have

$$\mathrm{P}(T_u^X \cap J \neq \emptyset) \leq \mathrm{P}(E_{\delta,\varepsilon}) + \sum_{j=0}^{m-1} \mathrm{P}(\{T_u^X \cap [t_j, t_{j+1}] \neq \emptyset\} \cap E_{\delta,\varepsilon}^C)$$
$$< \varepsilon + \sum_{j=0}^{m-1} \mathrm{P}\left(|X(t_j) - u| \leq \varepsilon \frac{\ell}{m}\right) = \varepsilon + \sum_{j=0}^{m-1} \int_{|x-u| \leq \varepsilon(\ell/m)} p_{X(t_j)}(x)\, dx.$$

If C is an upper bound for $p_{X(t)}(x)$, $t \in J$, $|x - u| \leq \varepsilon\, l/m$, we obtain

$$\mathrm{P}(T_u^X \cap J \neq \emptyset) \leq \varepsilon + C\varepsilon\ell.$$

Since $\varepsilon > 0$ is arbitrary, the result follows. □

The second result is an extension of Ylvisaker's theorem, which has the following statement:

Theorem 1.21 (Ylvisaker, 1968). *Let $\{Z(t) : t \in T\}$ be a real-valued Gaussian process indexed on a compact separable topological space T having continuous paths and* $\mathrm{Var}(Z(t)) > 0$ *for all $t \in T$. Then, for fixed $u \in \mathbb{R}$, one has* $\mathrm{P}(E_u^Z \neq \emptyset) = 0$, *where E_u^Z is the set of local extrema of $Z(\cdot)$ having value equal to u.*

The extension is the following:

Theorem 1.22 *Let $\{Z(t) : t \in T\}$ be a real-valued Gaussian process on some parameter set T and denote by $M^Z = \sup_{t \in T} Z(t)$ its supremum (which takes values in $\mathbb{R} \cup \{+\infty\}$). We assume that there exists a nonrandom countable set $\mathcal{D}$, $\mathcal{D} \subset T$, such that a.s. $M^Z = \sup_{t \in \mathcal{D}} Z(t)$. Assume further that there exist $\sigma_0^2 > 0, m_- > -\infty$ such that*

$$m(t) = \mathrm{E}(Z(t)) \geq m_-$$
$$\sigma^2(t) = \mathrm{Var}(Z(t)) \geq \sigma_0^2 \qquad \textit{for every } t \in T.$$

Then the distribution of the random variable M^Z is the sum of an atom at $+\infty$ and a (possibly defective) probability measure on $\mathbb{R}$ which has a locally bounded density.

Proof. Step 1. Suppose first that $\{X(t) : t \in T\}$ satisfies the hypotheses of the theorem, and, moreover,

$$\mathrm{Var}(X(t)) = 1, \qquad \mathrm{E}(X(t)) \geq 0$$

for every $t \in T$. We prove that the supremum M^X has a density p_{M^X}, which satisfies the inequality

$$p_{M^X}(u) \le \psi(u) := \frac{\exp(-u^2/2)}{\int_u^{\infty} \exp(-v^2/2)\, dv} \qquad \text{for every u} \in \mathbb{R}. \tag{1.21}$$

Let $\mathcal{D} = \{t_k\}_{k=1,2,\ldots}$. Almost surely, $M^X = \sup\{X(t_1) \ldots\ X(t_n) \ldots\}$. We set

$$M_n := \sup_{1 \le k \le n} X(t_k).$$

Since the joint distribution of $X(t_k), k = 1, \ldots, n$, is Gaussian, for any choice of $k, \ell = 1, \ldots, n; k \ne \ell$, the probability $\mathrm{P}\{X(t_k) = X(t_\ell)\}$ is equal to 0 or 1. Hence, possibly excluding some of these random variables, we may assume that these probabilities are all equal to 0 without changing the value of M_n on a set of probability 1. Then the distribution of the random variable M_n has a density $g_n(\cdot)$ that can be written as

$$g_n(x) = \sum_{k=1}^{n} \mathrm{P}(X(t_j) < x, j = 1, \ldots, n;\ j \ne k | X(t_k) = x) \times \frac{e^{-(1/2)(x - m(t_k))^2}}{\sqrt{2\pi}} = \varphi(x) G_n(x),$$

where φ denotes the standard normal density and

$$G_n(x) = \sum_{k=1}^{n} \mathrm{P}(Y_j < x - m(t_j), j = 1, \ldots, n;\ j \ne k | Y_k = x - m(t_k)) e^{x m(t_k) - (1/2) m^2(t_k)} \tag{1.22}$$

with

$$Y_j = X(t_j) - m(t_j) \qquad j = 1, \ldots, n.$$

Let us prove that $x \rightsquigarrow G_n(x)$ is an increasing function.

Since $m(t) \ge 0$, it is sufficient that the conditional probability in each term of (1.22) be increasing as a function of x. Write the Gaussian regression

$$Y_j = Y_j - c_{jk} Y_k + c_{jk} Y_k \qquad \text{with} \quad c_{jk} = \mathrm{E}(Y_j Y_k),$$

where the random variables $Y_j - c_{jk} Y_k$ and Y_k are independent. Then the conditional probability becomes

$$\mathrm{P}(Y_j - c_{jk} Y_k < x - m(t_j) - c_{jk}(x - m(t_k)), j = 1, \ldots, n;\ j \ne k).$$

This probability increases with x because $1 - c_{jk} \geq 0$, due to the Cauchy–Schwarz inequality. Now, if $a, b \in \mathbb{R}$, $a < b$, since $M_n \uparrow M^X$,

$$P\{a < M^X \leq b\} = \lim_{n\to\infty} P(a < M_n \leq b).$$

Using the monotonicity of G_n, we obtain

$$G_n(b) \int_b^{+\infty} \varphi(x)\ dx \leq \int_b^{+\infty} G_n(x)\varphi(x)\ dx = \int_b^{+\infty} g_n(x)\ dx \leq 1,$$

so that

$$\begin{aligned} P\{a < M_n \leq b\} &= \int_a^b g_n(x)\ dx \leq G_n(b) \int_a^b \varphi(x)\, dx \\ &\leq \int_a^b \varphi(x)\, dx \left(\int_b^{+\infty} \varphi(x)\, dx \right)^{-1}. \end{aligned}$$

This proves (1.21).

STEP 2. Now let Z satisfy the hypotheses of the theorem without assuming the added ones in step 1. For given $a, b \in \mathbb{R}$, $a < b$, choose $A \in \mathbb{R}^+$ so that $|a| < A$ and consider the process

$$X(t) = \frac{Z(t) - a}{\sigma(t)} + \frac{|m_-| + A}{\sigma_0}.$$

Clearly, for every $t \in T$,

$$E\big(X(t)\big) = \frac{m(t) - a}{\sigma(t)} + \frac{|m_-| + A}{\sigma_0} \geq -\frac{|m_-| + |a|}{\sigma_0} + \frac{|m_-| + A}{\sigma_0} \geq 0$$

and

$$\mathrm{Var}\big(X(t)\big) = 1,$$

so that (1.21) holds for the process X.

On the other hand,

$$\{a < M^Z \leq b\} \subset \{\mu_1 < M^X \leq \mu_2\},$$

where

$$\mu_1 = \frac{|m_-| + A}{\sigma_0}, \qquad \mu_2 = \frac{|m_-| + A}{\sigma_0} + \frac{b - a}{\sigma_0}.$$

It follows that

$$P\big\{a < M^Z \leq b\big\} \leq \int_{\mu_1}^{\mu_2} \psi(u)\ du = \int_a^b \frac{1}{\sigma_0} \psi\Big(\frac{v - a + |m_-| + A}{\sigma_0}\Big)\ dv,$$

which proves the statement. □

Theorem 1.21 follows directly from Theorem 1.22, since under the hypotheses of Theorem 1.21, we can write

$$\{E_u^X \neq \emptyset\} \subset \bigcup_{U \in \mathcal{F}} (\{M_U = u\} \cup \{m_U = u\}),$$

where M_U (respectively, m_u) is the maximum (respectively, the minimum) of the process on the set U and $\mathcal{F}$ denotes a countable family of open sets being a basis for the topology of T.

Remark. We come back in later chapters to the subject of the regularity properties of the probability distribution of the supremum of a Gaussian process.

EXERCISES

1.1. Let $T = \mathcal{N}$ be the set of natural numbers. Prove that the following sets belong to $\sigma(\mathcal{C})$.

(**a**) c_0 (the set of real-valued sequences $\{a_n\}$ such that $a_n \to 0$). *Suggestion*: Note that $c_0 = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} \{|a_n| < 1/k\}$.

(**b**) ℓ^2 (the set of real-valued sequences $\{a_n\}$ such that $\sum_n |a_n|^2 < \infty$).

(**c**) The set of real-valued sequences $\{a_n\}$ such that $\overline{\lim}_{n\to\infty} a_n \leq 1$.

1.2. Take $T = \mathbb{R}$, $\mathcal{T} = \mathcal{B}_{\mathbb{R}}$. Then if for each $\omega \in \Omega$ the function

$$t \rightsquigarrow X(t, \omega), \qquad (1.23)$$

the *path corresponding to* ω, is a continuous function, the process is bi-measurable. In fact, check that

$$X(t, \omega) = \lim_{n \to +\infty} X^{(n)}(t, \omega),$$

where for $n = 1, 2, \ldots,$ $X^{(n)}$is defined by

$$X^{(n)}(t, \omega) = \sum_{k=-\infty}^{k=+\infty} X_{k/2^n}(\omega) \mathbb{1}_{\{k/2^n \leq t < (k+1)/2^n\}},$$

which is obviously measurable as a function of the pair (t, ω). So the limit function X has the required property. If one replaces the continuity of the path (1.23) by some other regularity properties such as right continuity, bi-measurability follows in a similar way.

1.3. Let U be a random variable defined on some probability space $(\Omega, \mathcal{A}, \mathrm{P})$, having uniform distribution on the interval [0, 1]. Consider the two stochastic processes

$$Y(t) = \mathbb{1}_{t=U}$$
$$X(t) \equiv 0.$$

The process $Y(t)$ is sometimes called the *random parasite*.

(**a**) Prove that for all $t \in [0, 1]$, a.s. $X(t) = Y(t)$.

(**b**) Deduce that the processes $X(t)$ and $Y(t)$ have the same probability distribution P on $\mathbb{R}^{[0,1]}$ equipped with its Borel σ-algebra.

(**c**) Notice that for each ω in the probability space, $\sup_{t\in[0,1]} Y(t) = 1$ and $\sup_{t\in[0,1]} X(t) = 0$, so that the suprema of both processes are completely different. Is there a contradiction with the previous point?

1.4. Let μ be a Borel probability measure on the real line and Γ its Fourier transform; that is,

$$\Gamma(\tau) = \int_{\mathbb{R}} \exp(i\tau x)\mu(dx).$$

(**a**) Prove that if

$$\lambda_k = \int_{\mathbb{R}} |x|^k \mu(dx) < \infty$$

for some positive integer k, the covariance $\Gamma(\cdot)$ is of class C^k and

$$\Gamma^{(k)}(\tau) = \int_{\mathbb{R}} (ix)^k \exp(i\tau x)\mu(dx).$$

(**b**) Prove that if k is even, $k = 2p$, the reciprocal is true: If Γ is of class C^{2p}, then λ_{2p} is finite and

$$\Gamma(t) = 1 - \lambda_2 \frac{t^2}{2!} + \lambda_4 \frac{t^4}{4!} + \cdots + (-1)^{2p}\lambda_{2p}\frac{t^{2p}}{(2p)!} + o(t^{2p}).$$

Hint: Using induction on p and supposing that λ_k is infinite, then for every $A > 0$, one can find some $M > 0$ such that

$$\int_M^M x^k \mu(dx) \geq A.$$

Show that it implies that

$$(-1)^k \frac{k!}{t^k}\left[\Gamma(t) - \left(1 - \lambda_2 \frac{t^2}{2!} + \cdots + (-1)^{k-2}\lambda_{k-2}\frac{t^{k-2}}{(k-2)!}\right)\right]$$

has a limit, when t tends to zero, greater than A, which contradicts differentiability.

(**c**) When k is odd, the result is false [see Feller, 1966, Chap. XVII, example (c)].

1.5. Let $\{\xi_n\}_{n=1,2,\ldots}$ be a sequence of random vectors defined on some probability space taking values in $\mathbb{R}^d$, and assume that $\xi_n \to \xi$ in probability for some random vector ξ. Prove that if each ξ_n is Gaussian, ξ is also Gaussian.

1.6. Prove the following statements on the process defined by (1.10).

(**a**) For each $t \in T$ the series (1.10) converges a.s.

(**b**) Almost surely, the function $t \rightsquigarrow X(t)$ is in H and $\|X_{(\cdot)}\|_H^2 = \sum_{n=1}^{\infty} c_n \xi_n^2$.

(**c**) $\{\varphi_n\}_{n=1,2,\ldots}$ are eigenfunctions—with eigenvalues $\{c_n\}_{n=1,2,\ldots}$, respectively—of the linear operator $A : H \to H$ defined by

$$(Af)(s) = \int_T r(s,t) f(t) \rho(dt).$$

1.7. Let $\{X(t) : t \in T\}$ be a stochastic process defined on some separable topological space T.

(**a**) Prove that if $X(t)$ has continuous paths, it is separable.

(**b**) Let $T = \mathbb{R}$. Prove that if the paths of $X(t)$ are càd-làg, $X(t)$ is separable.

1.8. Let $\{X(t) : t \in \mathbb{R}^d\}$ be a separable stochastic process defined on some (complete) probability space $(\Omega, \mathcal{A}, \mathrm{P})$.

(**a**) Prove that the subset of Ω $\{X(\cdot)$ is continuous$\}$ is in $\mathcal{A}$.

(**b**) Prove that the conclusion in part (a) remains valid if one replaces "continuous" by "upper continuous", "lower continuous," or "continuous on the right" [a real-valued function f defined on $\mathbb{R}^d$ is said to be *continuous on the right* if for each t, $f(t)$ is equal to the limit of $f(s)$ when each coordinate of s tends to the corresponding coordinate of t on its right].

1.9. Show that in the case of the Wiener process, condition (1.18) holds for every $p \geq 2$, with $r = p/2 - 1$. Hence, the proposition implies that a.s., the paths of the Wiener process satisfy a Hölder condition with exponent α, for every $\alpha < \frac{1}{2}$.

1.10. (*Wiener integral*) Let $\{W_1(t) : t \geq 0\}$ and $\{W_2(t) : t \geq 0\}$ be two independent Wiener processes defined on some probability space $(\Omega, \mathcal{A}, \mathrm{P})$, and denote by $\{W(t) : t \in \mathbb{R}\}$ the process defined as

$$W(t) = W_1(t) \text{ if } t \geq 0 \quad \text{and} \quad W(t) = W_2(-t) \text{ if } t \leq 0.$$

$L^2(\mathbb{R}, \lambda)$ denotes the standard L^2-space of real-valued measurable functions on the real line with respect to Lebesgue measure and $L^2(\Omega, \mathcal{A}, \mathrm{P})$ the L^2 of the probability space. $\mathcal{C}^1_K(\mathbb{R})$ denotes the subspace of $L^2(\mathbb{R}, \lambda)$ of $\mathcal{C}^1$-functions with compact support. Define the function $I : \mathcal{C}^1_K(\mathbb{R}) \to L^2(\Omega, \mathcal{A}, \mathrm{P})$ as

$$I(f) = -\int_{\mathbb{R}} f'(t) W(t)\, dt \tag{1.24}$$

for each nonrandom $f \in \mathcal{C}^1_K(\mathbb{R})$. Equation (1.24) is well defined for each $\omega \in \Omega$ since the integrand is a continuous function with compact support.

(**a**) Prove that I is an isometry, in the sense that $\int_{\mathbb{R}} f^2(t)\, dt = \mathrm{E}\big(I^2(f)\big)$.

(**b**) Show that for each f, $I(f)$ is a centered Gaussian random variable. Moreover, for any choice of $f_1, \ldots, f_p \in \mathcal{C}^1_K(\mathbb{R})$, the joint distribution of $(I(f_1), \ldots., I(f_p))$ is centered Gaussian. Compute its covariance matrix.

(**c**) Prove that I admits a unique isometric extension $\tilde{I}$ to $L^2(\mathbb{R}, \lambda)$ such that:

(1) $\tilde{I}(f)$ is a centered Gaussian random variable with variance equal to $\int_{\mathbb{R}} f^2(t)\, dt$; similarly for joint distributions.

(2) $\displaystyle\int_{\mathbb{R}} f(t) g(t)\, dt = \mathrm{E}\big(\tilde{I}(f)\tilde{I}(g)\big).$

Comment: $\tilde{I}(f)$ is called the *Wiener integral of* f.

1.11. (*Fractional Brownian motion*) Let H be a real number, $0 < H < 1$. We use the notation and definitions of Exercise 1.10.

(**a**) For $t \geq 0$, define the function $K_t : \mathbb{R} \to \mathbb{R}$:

$$K_t(u) = \big[(t-u)^{H-1/2} - (-u)^{H-1/2}\big]\mathbb{I}_{u<0} + (t-u)^{H-1/2}\mathbb{I}_{0<u<t}.$$

Prove that $K_t \in L^2(\mathbb{R}, \lambda)$.

(**b**) For $t \geq 0$, define the Wiener integral $\tilde{I}(K_t)$, and for $s, t \geq 0$, prove the formula

$$\mathrm{E}\big(\tilde{I}(K_s)\tilde{I}(K_t)\big) = \frac{C_H}{2}\big[s^{2H} + t^{2H} - |t-s|^{2H}\big],$$

where C_H is a positive constant depending only on H. Compute C_H.

(**c**) Prove that the stochastic process $\{C_H^{-1/2}\tilde{I}(K_t) : t \geq 0\}$ has a version with continuous paths. This normalized version with continuous paths is usually called the *fractional Brownian motion with Hurst exponent* H and is denoted $\{W_H(t) : t \geq 0\}$.

(**d**) Show that if $H = \frac{1}{2}$, then $\{W_H(t) : t \geq 0\}$ is the standard Wiener process.

(**e**) Prove that for any $\delta > 0$, almost surely the paths of the fractional Brownian motion with Hurst exponent H satisfy a Hölder condition with exponent $H - \delta$.

1.12. (*Local time*) Let $\{W(t) : t \geq 0\}$ be a Wiener process defined in a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. For $u \in \mathbb{R}$, I an interval $I \subset [0, +\infty]$ and $\delta > 0$, define

$$\mu_\delta(u, I) = \frac{1}{2\delta} \int_I \mathbb{I}_{|W(t)-u|<\delta} dt = \frac{1}{2\delta} \lambda(\{t \in I : |W(t) - u| < \delta\}).$$

(**a**) Prove that for fixed u and I, $\mu_\delta(u, I)$ converges in $L^2(\Omega, \mathcal{A}, \mathrm{P})$ as $\delta \to 0$. Denote the limit by $\mu_0(u, I)$. *Hint:* Use Cauchy's criterion.

(**b**) Denote $Z(t) = \mu_0(u, [0, t])$. Prove that the random process $Z(t) : t \geq 0$ has a version with continuous paths. We call this version the *local time of the Wiener process at the level* u, and denote it by $L^W(u, t)$.

(**c**) For fixed u, $L^W(u, t)$ is a continuous increasing function of $t \geq 0$. Prove that a.s. it induces a measure on $\mathbb{R}^+$ that is singular with respect to Lebesgue measure; that is, its support is contained in a set of Lebesgue measure zero.

(**d**) Study the Hölder continuity properties of $L^W(u, t)$. For future reference, with a slight abuse of notation, we will write, for any interval $I = [t_1, t_2], 0 \leq t_1 \leq t_2$:

$$L^W(u, I) = L^W(u, t_2) - L^W(u, t_1).$$