Chapter 1
Gaussian Variables and Gaussian Processes

Gaussian random processes play an important role both in theoretical probability and in various applied models. We start by recalling basic facts about Gaussian random variables and Gaussian vectors. We then discuss Gaussian spaces and Gaussian processes, and we establish the fundamental properties concerning independence and conditioning in the Gaussian setting. We finally introduce the notion of a Gaussian white noise, which will be used to give a simple construction of Brownian motion in the next chapter.

1.1 Gaussian Random Variables

Throughout this chapter, we deal with random variables defined on a probability space $(\Omega, \mathscr{F}, P)$. For some of the existence statements that follow, this probability space should be chosen in an appropriate way. For every real $p \geq 1$, $L^p(\Omega, \mathscr{F}, P)$, or simply L^p if there is no ambiguity, denotes the space of all real random variables X such that $|X|^p$ is integrable, with the usual convention that two random variables that are a.s. equal are identified. The space L^p is equipped with the usual norm.

A real random variable X is said to be a *standard Gaussian* (or *normal*) variable if its law has density

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

with respect to Lebesgue measure on $\mathbb{R}$. The complex Laplace transform of X is then given by

$$E[e^{zX}] = e^{z^2/2}, \qquad \forall z \in \mathbb{C}.$$

J.-F. Le Gall, *Brownian Motion, Martingales, and Stochastic Calculus*,
Graduate Texts in Mathematics 274, DOI 10.1007/978-3-319-31089-3_1

To get this formula (and also to verify that the complex Laplace transform is well defined), consider first the case when $z = \lambda \in \mathbb{R}$:

$$E[\mathrm{e}^{\lambda X}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\lambda x}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x = \mathrm{e}^{\lambda^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-(x-\lambda)^2/2}\, \mathrm{d}x = \mathrm{e}^{\lambda^2/2}.$$

This calculation ensures that $E[\mathrm{e}^{zX}]$ is well-defined for every $z \in \mathbb{C}$, and defines a holomorphic function on $\mathbb{C}$. By analytic continuation, the identity $E[\mathrm{e}^{zX}] = \mathrm{e}^{z^2/2}$, which is true for every $z \in \mathbb{R}$, must also be true for every $z \in \mathbb{C}$.

By taking $z = \mathrm{i}\xi$, $\xi \in \mathbb{R}$, we get the characteristic function of X:

$$E[\mathrm{e}^{\mathrm{i}\xi X}] = \mathrm{e}^{-\xi^2/2}.$$

From the expansion

$$E[\mathrm{e}^{\mathrm{i}\xi X}] = 1 + \mathrm{i}\xi E[X] + \cdots + \frac{(\mathrm{i}\xi)^n}{n!} E[X^n] + O(|\xi|^{n+1}),$$

as $\xi \to 0$ (this expansion holds for every $n \geq 1$ when X belongs to all spaces L^p, $1 \leq p < \infty$, which is the case here), we get

$$E[X] = 0, \quad E[X^2] = 1$$

and more generally, for every integer $n \geq 0$,

$$E[X^{2n}] = \frac{(2n)!}{2^n n!}. \quad E[X^{2n+1}] = 0.$$

If $\sigma > 0$ and $m \in \mathbb{R}$, we say that a real random variable Y is *Gaussian* with $\mathcal{N}(m, \sigma^2)$-distribution if Y satisfies any of the three equivalent properties:

(i) $Y = \sigma X + m$, where X is a standard Gaussian variable (i.e. X follows the $\mathcal{N}(0, 1)$-distribution);
(ii) the law of Y has density

$$p_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(y-m)^2}{2\sigma^2};$$

(iii) the characteristic function of Y is

$$E[\mathrm{e}^{\mathrm{i}\xi Y}] = \exp(\mathrm{i}m\xi - \frac{\sigma^2}{2}\xi^2).$$

We have then

$$E[Y] = m, \quad \mathrm{var}(Y) = \sigma^2.$$

By extension, we say that Y is Gaussian with $\mathcal{N}(m, 0)$-distribution if $Y = m$ a.s. (property (iii) still holds in that case).

Sums of independent Gaussian variables Suppose that Y follows the $\mathcal{N}(m, \sigma^2)$-distribution, Y' follows the $\mathcal{N}(m', \sigma'^2)$-distribution, and Y and Y' are independent. Then $Y + Y'$ follows the $\mathcal{N}(m + m', \sigma^2 + \sigma'^2)$-distribution. This is an immediate consequence of (iii).

Proposition 1.1 *Let $(X_n)_{n\geq 1}$ be a sequence of real random variables such that, for every $n \geq 1$, X_n follows the $\mathcal{N}(m_n, \sigma_n^2)$-distribution. Suppose that X_n converges in L^2 to X. Then:*

(i) *The random variable X follows the $\mathcal{N}(m, \sigma^2)$-distribution, where $m = \lim m_n$ and $\sigma = \lim \sigma_n$.*
(ii) *The convergence also holds in all L^p spaces, $1 \leq p < \infty$.*

Remark The assumption that X_n converges in L^2 to X can be weakened to convergence in probability (and in fact the convergence in distribution of the sequence $(X_n)_{n\geq 1}$ suffices to get part (i)). We leave this as an exercise for the reader.

Proof

(i) The convergence in L^2 implies that $m_n = E[X_n]$ converges to $E[X]$ and $\sigma_n^2 = \mathrm{var}(X_n)$ converges to $\mathrm{var}(X)$ as $n \to \infty$. Then, setting $m = E[X]$ and $\sigma^2 = \mathrm{var}(X)$, we have for every $\xi \in \mathbb{R}$,

$$E[e^{i\xi X}] = \lim_{n\to\infty} E[e^{i\xi X_n}] = \lim_{n\to\infty} \exp(im_n\xi - \frac{\sigma_n^2}{2}\xi^2) = \exp(im\xi - \frac{\sigma^2}{2}\xi^2),$$

showing that X follows the $\mathcal{N}(m, \sigma^2)$-distribution.

(ii) Since X_n has the same distribution as $\sigma_n N + m_n$, where N is a standard Gaussian variable, and since the sequences (m_n) and (σ_n) are bounded, we immediately see that

$$\sup_n E[|X_n|^q] < \infty, \qquad \forall q \geq 1.$$

It follows that

$$\sup_n E[|X_n - X|^q] < \infty, \qquad \forall q \geq 1.$$

Let $p \geq 1$. The sequence $Y_n = |X_n - X|^p$ converges in probability to 0 and is uniformly integrable because it is bounded in L^2 (by the preceding bound with $q = 2p$). It follows that this sequence converges to 0 in L^1, which was the desired result.

□

1.2 Gaussian Vectors

Let E be a d-dimensional Euclidean space (E is isomorphic to $\mathbb{R}^d$ and we may take $E = \mathbb{R}^d$, with the usual inner product, but it will be more convenient to work with an abstract space). We write $\langle u, v\rangle$ for the inner product in E. A random variable X with values in E is called a *Gaussian vector* if, for every $u \in E$, $\langle u, X\rangle$ is a (real) Gaussian variable. (For instance, if $E = \mathbb{R}^d$, and if $X_1, \ldots, X_d$ are are independent Gaussian variables, the property of sums of independent Gaussian variables shows that the random vector $X = (X_1, \ldots, X_d)$ is a Gaussian vector.)

Let X be a Gaussian vector with values in E. Then there exist $m_X \in E$ and a nonnegative quadratic form q_X on E such that, for every $u \in E$,

$$E[\langle u, X\rangle] = \langle u, m_X\rangle,$$
$$\mathrm{var}(\langle u, X\rangle) = q_X(u).$$

Indeed, let $(e_1, \ldots, e_d)$ be an orthonormal basis on E, and write $X = \sum_{i=1}^d X_j\, e_j$ in this basis. Notice that the random variables $X_j = \langle e_j, X\rangle$ are Gaussian. It is then immediate that the preceding formulas hold with $m_X = \sum_{j=1}^d E[X_j]\, e_j \overset{\text{(not.)}}{=} E[X]$, and, if $u = \sum_{j=1}^d u_j e_j$,

$$q_X(u) = \sum_{j,k=1}^{d} u_j u_k \,\mathrm{cov}(X_j, X_k).$$

Since $\langle u, X\rangle$ follows the $\mathcal{N}(\langle u, m_X\rangle, q_X(u))$-distribution, we get the characteristic function of the random vector X,

$$E[\exp(\mathrm{i}\langle u, X\rangle)] = \exp(\mathrm{i}\langle u, m_X\rangle - \frac{1}{2} q_X(u)). \tag{1.1}$$

Proposition 1.2 *Under the preceding assumptions, the random variables $X_1, \ldots, X_d$ are independent if and only if the covariance matrix $(\mathrm{cov}(X_j, X_k))_{1\le j,k\le d}$ is diagonal or equivalently if and only if q_X is of diagonal form in the basis $(e_1, \ldots, e_d)$.*

Proof If the random variables $X_1, \ldots, X_d$ are independent, the covariance matrix $(\mathrm{cov}(X_j, X_k))_{j,k=1,\ldots d}$ is diagonal. Conversely, if this matrix is diagonal, we have for every $u = \sum_{j=1}^d u_j e_j \in E$,

$$q_X(u) = \sum_{j=1}^{d} \lambda_j\, u_j^2\ ,$$

where $\lambda_j = \mathrm{var}(X_j)$. Consequently, using (1.1),

$$E\Big[\exp\Big(\mathrm{i}\sum_{j=1}^{d} u_j X_j\Big)\Big] = \prod_{j=1}^{d} \exp(\mathrm{i}u_j E[X_j] - \frac{1}{2}\lambda_j u_j^2) = \prod_{j=1}^{d} E[\exp(\mathrm{i}u_j X_j)],$$

which implies that $X_1, \ldots, X_d$ are independent. □

With the quadratic form q_X, we associate the unique symmetric endomorphism γ_X of E such that

$$q_X(u) = \langle u, \gamma_X(u) \rangle$$

(the matrix of γ_X in the basis $(e_1, \ldots, e_d)$ is $(\mathrm{cov}(X_j, X_k))_{1 \le j,k \le d}$ but of course the definition of γ_X does not depend on the choice of a basis). Note that γ_X is nonnegative in the sense that its eigenvalues are all nonnegative.

From now on, to simplify the statements, we restrict our attention to centered Gaussian vectors, i.e. such that $m_X = 0$, but the following results are easily adapted to the non-centered case.

Theorem 1.3

(i) *Let γ be a nonnegative symmetric endomorphism of E. Then there exists a Gaussian vector X such that $\gamma_X = \gamma$.*

(ii) *Let X be a centered Gaussian vector. Let $(\varepsilon_1, \ldots, \varepsilon_d)$ be a basis of E in which γ_X is diagonal, $\gamma_X \varepsilon_j = \lambda_j \varepsilon_j$ for every $1 \le j \le d$, where*

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_d$$

so that r is the rank of γ_X. Then,

$$X = \sum_{j=1}^{r} Y_j \varepsilon_j,$$

where Y_j, $1 \le j \le r$, are independent (centered) Gaussian variables and the variance of Y_j is λ_j. Consequently, if P_X denotes the distribution of X, the topological support of P_X is the vector space spanned by $\varepsilon_1, \ldots, \varepsilon_r$. Furthermore, P_X is absolutely continuous with respect to Lebesgue measure on E if and only if $r = d$, and in that case the density of X is

$$p_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \gamma_X}} \exp -\frac{1}{2}\langle x, \gamma_X^{-1}(x) \rangle.$$

Proof

(i) Let $(\varepsilon_1, \dots, \varepsilon_d)$ be an orthonormal basis of E in which γ is diagonal, $\gamma(\varepsilon_j) = \lambda_j \varepsilon_j$ for $1 \le j \le d$, and let $Y_1, \dots, Y_d$ be independent centered Gaussian variables with $\text{var}(Y_j) = \lambda_j$, $1 \le j \le d$. We set

$$X = \sum_{j=1}^{d} Y_j \varepsilon_j.$$

Then, if $u = \sum_{j=1}^{d} u_j \varepsilon_j$,

$$q_X(u) = E\Big[\Big(\sum_{j=1}^{d} u_j Y_j\Big)^2\Big] = \sum_{j=1}^{d} \lambda_j u_j^2 = \langle u, \gamma(u)\rangle.$$

(ii) Let $Y_1, \dots, Y_d$ be the coordinates of X in the basis $(\varepsilon_1, \dots, \varepsilon_d)$. Then the matrix of γ_X in this basis is the covariance matrix of $Y_1, \dots, Y_d$. The latter covariance matrix is diagonal and, by Proposition 1.2, the variables $Y_1, \dots, Y_d$ are independent. Furthermore, for $j \in \{r+1, \dots, d\}$, we have $E[Y_j^2] = 0$ hence $Y_j = 0$ a.s.

Then, since $X = \sum_{j=1}^{r} Y_j \varepsilon_j$ a.s., it is clear that $\text{supp}\, P_X$ is contained in the subspace spanned by $\varepsilon_1, \dots, \varepsilon_r$. Conversely, if O is a rectangle of the form

$$O = \{u = \sum_{j=1}^{r} \alpha_j \varepsilon_j : a_j < \alpha_j < b_j,\ \forall 1 \le j \le r\},$$

we have $P[X \in O] = \prod_{j=1}^{r} P[a_j < Y_j < b_j] > 0$. This is enough to get that $\text{supp}\, P_X$ is the subspace spanned by $\varepsilon_1, \dots, \varepsilon_r$.

If $r < d$, since the vector space spanned by $\varepsilon_1, \dots, \varepsilon_r$ has zero Lebesgue measure, the distribution of X is singular with respect to Lebesgue measure on E. Suppose that $r = d$, and write Y for the random vector in $\mathbb{R}^d$ defined by $Y = (Y_1, \dots, Y_d)$. Note that the bijection $\varphi(y_1, \dots, y_d) = \sum y_j \varepsilon_j$ maps Y to X. Then, writing $y = (y_1, \dots, y_d)$, we have

$$\begin{aligned}
E[g(X)] &= E[g(\varphi(Y))] \\
&= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} g(\varphi(y)) \exp\Big(-\frac{1}{2}\sum_{j=1}^{d} \frac{y_j^2}{\lambda_j}\Big) \frac{\mathrm{d}y_1 \dots \mathrm{d}y_d}{\sqrt{\lambda_1 \dots \lambda_d}} \\
&= \frac{1}{(2\pi)^{d/2}\sqrt{\det \gamma_X}} \int_{\mathbb{R}^d} g(\varphi(y)) \exp\Big(-\frac{1}{2}\langle \varphi(y), \gamma_X^{-1}(\varphi(y))\rangle\Big) \mathrm{d}y_1 \dots \mathrm{d}y_d \\
&= \frac{1}{(2\pi)^{d/2}\sqrt{\det \gamma_X}} \int_{E} g(x) \exp\Big(-\frac{1}{2}\langle x, \gamma_X^{-1}(x)\rangle\Big) \mathrm{d}x,
\end{aligned}$$

since Lebesgue measure on E is by definition the image of Lebesgue measure on $\mathbb{R}^d$ under φ (or under any other vector isometry from $\mathbb{R}^d$ onto E). In the second equality, we used the fact that $Y_1, \ldots, Y_d$ are independent Gaussian variables, and in the third equality we observed that

$$\langle \varphi(y), \gamma_X^{-1}(\varphi(y)) \rangle = \Big\langle \sum_{j=1}^{d} y_j \varepsilon_j, \sum_{j=1}^{d} \frac{y_j}{\lambda_j} \varepsilon_j \Big\rangle = \sum_{j=1}^{d} \frac{y_j^2}{\lambda_j}.$$

□

1.3 Gaussian Processes and Gaussian Spaces

From now on until the end of this chapter, we consider only centered Gaussian variables, and we frequently omit the word "centered".

Definition 1.4 A (centered) *Gaussian space* is a closed linear subspace of $L^2(\Omega, \mathscr{F}, P)$ which contains only centered Gaussian variables.

For instance, if $X = (X_1, \ldots, X_d)$ is a centered Gaussian vector in $\mathbb{R}^d$, the vector space spanned by $\{X_1, \ldots, X_d\}$ is a Gaussian space.

Definition 1.5 Let $(E, \mathscr{E})$ be a measurable space, and let T be an arbitrary index set. A *random process* (indexed by T) with values in E is a collection $(X_t)_{t \in T}$ of random variables with values in E. If the measurable space $(E, \mathscr{E})$ is not specified, we will implicitly assume that $E = \mathbb{R}$ and $\mathscr{E} = \mathscr{B}(\mathbb{R})$ is the Borel σ-field on $\mathbb{R}$.

Here and throughout this book, we use the notation $\mathscr{B}(F)$ for the Borel σ-field on a topological space F. Most of the time, the index set T will be $\mathbb{R}_+$ or another interval of the real line.

Definition 1.6 A (real-valued) random process $(X_t)_{t \in T}$ is called a (centered) *Gaussian process* if any finite linear combination of the variables X_t, $t \in T$ is centered Gaussian.

Proposition 1.7 *If $(X_t)_{t \in T}$ is a Gaussian process, the closed linear subspace of L^2 spanned by the variables X_t, $t \in T$, is a Gaussian space, which is called the Gaussian space generated by the process X.*

Proof It suffices to observe that an L^2-limit of centered Gaussian variables is still centered Gaussian, by Proposition 1.1. □

We now turn to independence properties in a Gaussian space. We need the following definition.

Definition 1.8 Let H be a collection of random variables defined on $(\Omega, \mathscr{F}, P)$. The σ-field *generated by H*, denoted by $\sigma(H)$, is the smallest σ-field on Ω such that all variables $\xi \in H$ are measurable for this σ-field. If $\mathscr{C}$ is a collection of subsets of Ω, we also write $\sigma(\mathscr{C})$ for the smallest σ-field on Ω that contains all elements of $\mathscr{C}$.

The next theorem shows that, in some sense, independence is equivalent to orthogonality in a Gaussian space. This is a very particular property of the Gaussian distribution.

Theorem 1.9 *Let H be a centered Gaussian space and let $(H_i)_{i\in I}$ be a collection of linear subspaces of H. Then the subspaces H_i, $i \in I$, are (pairwise) orthogonal in L^2 if and only the σ-fields $\sigma(H_i)$, $i \in I$, are independent.*

Remark It is crucial that the vector spaces H_i are subspaces of a common Gaussian space H. Consider for instance a random variable X distributed according to $\mathscr{N}(0,1)$ and another random variable ε independent of X and such that $P[\varepsilon = 1] = P[\varepsilon = -1] = 1/2$. Then $X_1 = X$ and $X_2 = \varepsilon X$ are both distributed according to $\mathscr{N}(0,1)$. Moreover, $E[X_1X_2] = E[\varepsilon]E[X^2] = 0$. Nonetheless X_1 and X_2 are obviously not independent (because $|X_1| = |X_2|$). In this example, (X_1, X_2) is not a Gaussian vector in $\mathbb{R}^2$ despite the fact that both coordinates are Gaussian variables.

Proof Suppose that the σ-fields $\sigma(H_i)$ are independent. Then, if $i \neq j$, if $X \in H_i$ and $Y \in H_j$,

$$E[XY] = E[X]E[Y] = 0,$$

so that the linear spaces H_i are pairwise orthogonal.

Conversely, suppose that the linear spaces H_i are pairwise orthogonal. From the definition of the independence of an infinite collection of σ-fields, it is enough to prove that, if $i_1, \ldots, i_p \in I$ are distinct, the σ-fields $\sigma(H_{i_1}), \ldots, \sigma(H_{i_p})$ are independent. To this end, it is enough to verify that, if $\xi_1^1, \ldots, \xi_{n_1}^1 \in H_{i_1}, \ \ldots \ , \xi_1^p, \ldots, \xi_{n_p}^p \in H_{i_p}$ are fixed, the vectors $(\xi_1^1, \ldots, \xi_{n_1}^1), \ldots, (\xi_1^p, \ldots, \xi_{n_p}^p)$ are independent (indeed, for every $j \in \{1, \ldots, p\}$, the events of the form $\{\xi_1^j \in A_1, \ldots, \xi_{n_j}^j \in A_{n_j}\}$ give a class stable under finite intersections that generates the σ-field $\sigma(H_{i_j})$, and the desired result follows by a standard monotone class argument, see Appendix A1). However, for every $j \in \{1, \ldots, p\}$ we can find an orthonormal basis $(\eta_1^j, \ldots, \eta_{m_j}^j)$ of the linear subspace of L^2 spanned by $\{\xi_1^j, \ldots, \xi_{n_j}^j\}$. The covariance matrix of the vector

$$(\eta_1^1, \ldots, \eta_{m_1}^1, \eta_1^2, \ldots, \eta_{m_2}^2, \ldots, \eta_1^p, \ldots, \eta_{m_p}^p)$$

is then the identity matrix (for $i \neq j$, $E[\eta_l^i \eta_r^j] = 0$ because H_i and H_j are orthogonal). Moreover, this vector is Gaussian because its components belong to H. By Proposition 1.2, the components of the latter vector are independent random variables. This implies in turn that the vectors $(\eta_1^1, \ldots, \eta_{m_1}^1), \ldots, (\eta_1^p, \ldots, \eta_{m_p}^p)$ are independent. Equivalently, the vectors $(\xi_1^1, \ldots, \xi_{n_1}^1), \ldots, (\xi_1^p, \ldots, \xi_{n_p}^p)$ are independent, which was the desired result. □

As an application of the previous theorem, we now discuss conditional expectations in the Gaussian framework. Again, the fact that these conditional expectations can be computed as orthogonal projections (as shown in the next corollary) is very particular to the Gaussian setting.

Corollary 1.10 *Let H be a (centered) Gaussian space and let K be a closed linear subspace of H. Let p_K denote the orthogonal projection onto K in the Hilbert space L^2, and let $X \in H$.*

(i) *We have*

$$E[X \mid \sigma(K)] = p_K(X).$$

(ii) *Let $\sigma^2 = E[(X - p_K(X))^2]$. Then, for every Borel subset A of $\mathbb{R}$, the random variable $P[X \in A \mid \sigma(K)]$ is given by*

$$P[X \in A \mid \sigma(K)](\omega) = Q(\omega, A),$$

where $Q(\omega, \cdot)$ denotes the $\mathcal{N}(p_K(X)(\omega), \sigma^2)$-distribution:

$$Q(\omega, A) = \frac{1}{\sigma\sqrt{2\pi}} \int_A \mathrm{d}y \, \exp\Big(- \frac{(y - p_K(X)(\omega))^2}{2\sigma^2} \Big)$$

(and by convention $Q(\omega, A) = \mathbf{1}_A(p_K(X))$ if $\sigma = 0$).

Remarks

(a) Part (ii) of the statement can be interpreted by saying that the conditional distribution of X knowing $\sigma(K)$ is $\mathcal{N}(p_K(X), \sigma^2)$.

(b) For a general random variable X in L^2, one has

$$E[X \mid \sigma(K)] = p_{L^2(\Omega, \sigma(K), P)}(X).$$

Assertion (i) shows that, in our Gaussian framework, this orthogonal projection coincides with the orthogonal projection onto the space K, which is "much smaller" than $L^2(\Omega, \sigma(K), P)$.

(c) Assertion (i) also gives the principle of linear regression. For instance, if (X_1, X_2, X_3) is a (centered) Gaussian vector in $\mathbb{R}^3$, the best approximation in L^2 of X_3 as a (not necessarily linear) function of X_1 and X_2 can be written $\lambda_1 X_1 + \lambda_2 X_2$ where λ_1 and λ_2 are computed by saying that $X_3 - (\lambda_1 X_1 + \lambda_2 X_2)$ is orthogonal to the vector space spanned by X_1 and X_2.

Proof

(i) Let $Y = X - p_K(X)$. Then Y is orthogonal to K and, by Theorem 1.9, Y is independent of $\sigma(K)$. Then,

$$E[X \mid \sigma(K)] = E[p_K(X) \mid \sigma(K)] + E[Y \mid \sigma(K)] = p_K(X) + E[Y] = p_K(X).$$

(ii) For every nonnegative measurable function f on $\mathbb{R}_+$,

$$E[f(X) \mid \sigma(K)] = E[f(p_K(X) + Y) \mid \sigma(K)] = \int P_Y(\mathrm{d}y) f(p_K(X) + y),$$

where P_Y is the law of Y, which is $\mathcal{N}(0, \sigma^2)$ since Y is centered Gaussian with variance σ^2. In the second equality, we also use the following general fact: if Z is a $\mathscr{G}$-measurable random variable and if Y is independent of $\mathscr{G}$ then, for every nonnegative measurable function g, $E[g(Y, Z) \mid \mathscr{G}] = \int g(y, Z)\, P_Y(\mathrm{d}y)$. Property (ii) immediately follows.

□

Let $(X_t)_{t \in T}$ be a (centered) Gaussian process. The covariance function of X is the function $\Gamma : T \times T \longrightarrow \mathbb{R}$ defined by $\Gamma(s,t) = \mathrm{cov}(X_s, X_t) = E[X_s X_t]$. This function characterizes the collection of *finite-dimensional marginal distributions* of the process X, that is, the collection consisting for every choice of the distinct indices $t_1, \ldots, t_p$ in T of the law of the vector $(X_{t_1}, \ldots, X_{t_p})$. Indeed this vector is a centered Gaussian vector in $\mathbb{R}^p$ with covariance matrix $(\Gamma(t_i, t_j))_{1 \le i,j \le p}$.

Remark One can define in an obvious way the notion of a non-centered Gaussian process. The collection of finite-dimensional marginal distributions is then characterized by the covariance function and the mean function $t \mapsto m(t) = E[X_t]$.

Given a function Γ on $T \times T$, one may ask whether there exists a Gaussian process X whose Γ is the covariance function. The function Γ must be symmetric ($\Gamma(s,t) = \Gamma(t,s)$) and of positive type in the following sense: if c is a real function on T with finite support, then

$$\sum_{T \times T} c(s)c(t)\, \Gamma(s,t) \ge 0.$$

Indeed, if Γ is the covariance function of the process X, we have immediately

$$\sum_{T \times T} c(s)c(t)\, \Gamma(s,t) = \mathrm{var}\Big(\sum_T c(s) X_s\Big) \ge 0.$$

Note that when T is finite, the problem of the existence of X is solved under the preceding assumptions on Γ by Theorem 1.3.

The next theorem solves the existence problem in the general case. This theorem is a direct consequence of the Kolmogorov extension theorem, which in the particular case $T = \mathbb{R}_+$ is stated as Theorem 6.3 in Chap. 6 below (see e.g. Neveu [64, Chapter III], or Kallenberg [47, Chapter VI] for the general case). We omit the proof as this result will not be used in the sequel.

Theorem 1.11 *Let Γ be a symmetric function of positive type on $T \times T$. There exists, on an appropriate probability space $(\Omega, \mathscr{F}, P)$, a centered Gaussian process whose covariance function is Γ.*

Example Consider the case $T = \mathbb{R}$ and let μ be a finite measure on $\mathbb{R}$, which is also symmetric (i.e. $\mu(-A) = \mu(A)$). Then set, for every $s, t \in \mathbb{R}$,

$$\Gamma(s,t) = \int e^{i\xi(t-s)}\,\mu(d\xi).$$

It is easy to verify that Γ has the required properties. In particular, if c is a real function on $\mathbb{R}$ with finite support,

$$\sum_{\mathbb{R}\times\mathbb{R}} c(s)c(t)\,\Gamma(s,t) = \int |\sum_{\mathbb{R}} c(s)e^{i\xi s}|^2\,\mu(d\xi) \geq 0.$$

The process Γ enjoys the additional property that $\Gamma(s,t)$ only depends on $t - s$. It immediately follows that any (centered) Gaussian process $(X_t)_{t\in\mathbb{R}}$ with covariance function Γ is stationary (in a strong sense), meaning that

$$(X_{t_1+t}, X_{t_2+t}, \ldots, X_{t_n+t}) \overset{(d)}{=} (X_{t_1}, X_{t_2}, \ldots, X_{t_n})$$

for any choice of $t_1, \ldots, t_n, t \in \mathbb{R}$. Conversely, any stationary Gaussian process X indexed by $\mathbb{R}$ has a covariance of the preceding type (this is Bochner's theorem, which we will not use in this book), and the measure μ is called the spectral measure of X.

1.4 Gaussian White Noise

Definition 1.12 Let $(E, \mathscr{E})$ be a measurable space, and let μ be a σ-finite measure on $(E, \mathscr{E})$. A *Gaussian white noise* with intensity μ is an isometry G from $L^2(E, \mathscr{E}, \mu)$ into a (centered) Gaussian space.

Hence, if $f \in L^2(E, \mathscr{E}, \mu)$, $G(f)$ is centered Gaussian with variance

$$E[G(f)^2] = \|G(f)\|^2_{L^2(\Omega,\mathscr{F},P)} = \|f\|^2_{L^2(E,\mathscr{E},\mu)} = \int f^2\,d\mu.$$

If $f, g \in L^2(E, \mathscr{E}, \mu)$, the covariance of $G(f)$ and $G(g)$ is

$$E[G(f)G(g)] = \langle f, g\rangle_{L^2(E,\mathscr{E},\mu)} = \int fg\,\mathrm{d}\mu.$$

In particular, if $f = \mathbf{1}_A$ with $\mu(A) < \infty$, $G(\mathbf{1}_A)$ is $\mathscr{N}(0, \mu(A))$-distributed. To simplify notation, we will write $G(A) = G(\mathbf{1}_A)$.

Let $A_1, \ldots, A_n \in \mathscr{E}$ be disjoint and such that $\mu(A_j) < \infty$ for every j. Then the vector

$$(G(A_1), \ldots, G(A_n))$$

is a Gaussian vector in $\mathbb{R}^n$ and its covariance matrix is diagonal since, if $i \neq j$,

$$E[G(A_i)G(A_j)] = \left\langle \mathbf{1}_{A_i}, \mathbf{1}_{A_j}\right\rangle_{L^2(E,\mathscr{E},\mu)} = 0.$$

From Proposition 1.2, we get that the variables $G(A_1), \ldots, G(A_n)$ are independent.

Suppose that $A \in \mathscr{E}$ is such that $\mu(A) < \infty$ and that A is the disjoint union of a countable collection $A_1, A_2, \ldots$ of measurable subsets of E. Then, $\mathbf{1}_A = \sum_{j=1}^{\infty} \mathbf{1}_{A_j}$ where the series converges in $L^2(E, \mathscr{E}, \mu)$, and by the isometry property this implies that

$$G(A) = \sum_{j=1}^{\infty} G(A_j)$$

where the series converges in $L^2(\Omega, \mathscr{F}, P)$ (since the random variables $G(A_j)$ are independent, an easy application of the convergence theorem for discrete martingales also shows that the series converges a.s.).

Properties of the mapping $A \mapsto G(A)$ are therefore very similar to those of a measure depending on the parameter $\omega \in \Omega$. However, one can show that, if ω is fixed, the mapping $A \mapsto G(A)(\omega)$ does not (in general) define a measure. We will come back to this point later.

Proposition 1.13 *Let $(E, \mathscr{E})$ be a measurable space, and let μ be a σ-finite measure on $(E, \mathscr{E})$. There exists, on an appropriate probability space $(\Omega, \mathscr{F}, P)$, a Gaussian white noise with intensity μ.*

Proof We rely on elementary Hilbert space theory. Let $(f_i, i \in I)$ be a total orthonormal system in the Hilbert space $L^2(E, \mathscr{E}, \mu)$. For every $f \in L^2(E, \mathscr{E}, \mu)$,

$$f = \sum_{i\in I} \alpha_i f_i$$

where the coefficients $\alpha_i = \langle f, f_i \rangle$ are such that

$$\sum_{i \in I} \alpha_i^2 = \|f\|^2 < \infty.$$

On an appropriate probability space $(\Omega, \mathscr{F}, P)$ we can construct a collection $(X_i)_{i \in I}$, indexed by the same index set I, of independent $\mathscr{N}(0, 1)$ random variables (see [64, Chapter III] for the existence of such a collection – in the sequel we will only need the case when I is countable, and then an elementary construction using only the existence of Lebesgue measure on $[0, 1]$ is possible), and we set

$$G(f) = \sum_{i \in I} \alpha_i X_i.$$

The series converges in L^2 since the X_i, $i \in I$, form an orthonormal system in L^2. Then clearly G takes values in the Gaussian space generated by X_i, $i \in I$. Furthermore, G is an isometry since it maps the orthonormal basis $(f_i, i \in I)$ to an orthonormal system. □

We could also have deduced the previous result from Theorem 1.11 applied with $T = L^2(E, \mathscr{E}, \mu)$ and $\Gamma(f, g) = \langle f, g \rangle_{L^2(E,\mathscr{E},\mu)}$. In this way we get a Gaussian process $(X_f, f \in L^2(E, \mathscr{E}, \mu))$ and we just have to take $G(f) = X_f$.

Remark In what follows, we will only consider the case when $L^2(E, \mathscr{E}, \mu)$ is separable. For instance, if $(E, \mathscr{E}) = (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ and μ is Lebesgue measure, the construction of G only requires a sequence $(\xi_n)_{n \geq 0}$ of independent $\mathscr{N}(0, 1)$ random variables, and the choice of an orthonormal basis $(\varphi_n)_{n \geq 0}$ of $L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), \mathrm{d}t)$: We get G by setting

$$G(f) = \sum_{n \geq 0} \langle f, \varphi_n \rangle \, \xi_n.$$

See Exercise 1.18 for an explicit choice of $(\varphi_n)_{n \geq 0}$ when $E = [0, 1]$.

Our last proposition gives a way of recovering the intensity $\mu(A)$ of a measurable set A from the values of G on the atoms of finer and finer partitions of A.

Proposition 1.14 *Let G be a Gaussian white noise on $(E, \mathscr{E})$ with intensity μ. Let $A \in \mathscr{E}$ be such that $\mu(A) < \infty$. Assume that there exists a sequence of partitions of A,*

$$A = A_1^n \cup \ldots \cup A_{k_n}^n$$

whose "mesh" tends to 0*, in the sense that*

$$\lim_{n\to\infty}\left(\sup_{1\le j\le k_n}\mu(A_j^n)\right)=0.$$

Then,

$$\lim_{n\to\infty}\sum_{j=1}^{k_n}G(A_j^n)^2=\mu(A)$$

in L^2.

Proof For every fixed n, the variables $G(A_1^n),\ldots,G(A_{k_n}^n)$ are independent. Furthermore, $E[G(A_j^n)^2]=\mu(A_j^n)$. We then compute

$$E\left[\left(\sum_{j=1}^{k_n}G(A_j^n)^2-\mu(A)\right)^2\right]=\sum_{j=1}^{k_n}\mathrm{var}(G(A_j^n)^2)=2\sum_{j=1}^{k_n}\mu(A_j^n)^2,$$

because, if X is $\mathcal{N}(0,\sigma^2)$, $\mathrm{var}(X^2)=E(X^4)-\sigma^4=3\sigma^4-\sigma^4=2\sigma^4$. Then,

$$\sum_{j=1}^{k_n}\mu(A_j^n)^2\le\left(\sup_{1\le j\le k_n}\mu(A_j^n)\right)\mu(A)$$

tends to 0 as $n\to\infty$ by assumption. □

Exercises

Exercise 1.15 Let $(X_t)_{t\in[0,1]}$ be a centered Gaussian process. We assume that the mapping $(t,\omega)\mapsto X_t(\omega)$ from $[0,1]\times\Omega$ into $\mathbb{R}$ is measurable. We denote the covariance function of X by K.

1. Show that the mapping $t\mapsto X_t$ from $[0,1]$ into $L^2(\Omega)$ is continuous if and only if K is continuous on $[0,1]^2$. In what follows, we assume that this condition holds.
2. Let $h:[0,1]\to\mathbb{R}$ be a measurable function such that $\int_0^1|h(t)|\sqrt{K(t,t)}\,\mathrm{d}t<\infty$. Show that, for a.e. ω, the integral $\int_0^1 h(t)X_t(\omega)\mathrm{d}t$ is absolutely convergent. We set $Z=\int_0^1 h(t)X_t\mathrm{d}t$.
3. We now make the stronger assumption $\int_0^1|h(t)|\mathrm{d}t<\infty$. Show that Z is the L^2-limit of the variables $Z_n=\sum_{i=1}^n X_{\frac{i}{n}}\int_{\frac{i-1}{n}}^{\frac{i}{n}}h(t)\mathrm{d}t$ when $n\to\infty$ and infer that Z is a Gaussian random variable.

4. We assume that K is twice continuously differentiable. Show that, for every $t \in [0,1]$, the limit

$$\dot{X}_t := \lim_{s \to t} \frac{X_s - X_t}{s - t}$$

exists in $L^2(\Omega)$. Verify that $(\dot{X}_t)_{t \in [0,1]}$ is a centered Gaussian process and compute its covariance function.

Exercise 1.16 (*Kalman filtering*) Let $(\epsilon_n)_{n \geq 0}$ and $(\eta_n)_{n \geq 0}$ be two independent sequences of independent Gaussian random variables such that, for every n, ϵ_n is distributed according to $\mathcal{N}(0, \sigma^2)$ and η_n is distributed according to $\mathcal{N}(0, \delta^2)$, where $\sigma > 0$ and $\delta > 0$. We consider two other sequences $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ defined by the properties $X_0 = 0$, and, for every $n \geq 0$, $X_{n+1} = a_n X_n + \epsilon_{n+1}$ and $Y_n = cX_n + \eta_n$, where c and a_n are positive constants. We set

$$\hat{X}_{n/n} = E[X_n \mid Y_0, Y_1, \ldots, Y_n],$$
$$\hat{X}_{n+1/n} = E[X_{n+1} \mid Y_0, Y_1, \ldots, Y_n].$$

The goal of the exercise is to find a recursive formula allowing one to compute these conditional expectations.

1. Verify that $\hat{X}_{n+1/n} = a_n \hat{X}_{n/n}$, for every $n \geq 0$.
2. Show that, for every $n \geq 1$,

$$\hat{X}_{n/n} = \hat{X}_{n/n-1} + \frac{E[X_n Z_n]}{E[Z_n^2]} Z_n,$$

where $Z_n := Y_n - c\hat{X}_{n/n-1}$.
3. Evaluate $E[X_n Z_n]$ and $E[Z_n^2]$ in terms of $P_n := E[(X_n - \hat{X}_{n/n-1})^2]$ and infer that, for every $n \geq 1$,

$$\hat{X}_{n+1/n} = a_n \left(\hat{X}_{n/n-1} + \frac{cP_n}{c^2 P_n + \delta^2} Z_n \right).$$

4. Verify that $P_1 = \sigma^2$ and that, for every $n \geq 1$, the following induction formula holds:

$$P_{n+1} = \sigma^2 + a_n^2 \frac{\delta^2 P_n}{c^2 P_n + \delta^2}.$$

Exercise 1.17 Let H be a (centered) Gaussian space and let H_1 and H_2 be linear subspaces of H. Let K be a closed linear subspace of H. We write p_K for the orthogonal projection onto K. Show that the condition

$$\forall X_1 \in H_1, \forall X_2 \in H_2, \quad E[X_1X_2] = E[p_K(X_1)p_K(X_2)]$$

implies that the σ-fields $\sigma(H_1)$ and $\sigma(H_2)$ are conditionally independent given $\sigma(K)$. (This means that, for every nonnegative $\sigma(H_1)$-measurable random variable X_1, and for every nonnegative $\sigma(H_2)$-measurable random variable X_2, one has $E[X_1X_2|\sigma(K)] = E[X_1|\sigma(K)]\,E[X_2|\sigma(K)]$.) *Hint*: Via monotone class arguments explained in Appendix A1, it is enough to consider the case where X_1, resp. X_2, is the indicator function of an event depending only on finitely many variables in H_1, resp. in H_2.

Exercise 1.18 (*Lévy's construction of Brownian motion*) For every $t \in [0,1]$, we set $h_0(t) = 1$, and then, for every integer $n \geq 0$ and every $k \in \{0, 1, \ldots, 2^n - 1\}$,

$$h_k^n(t) = 2^{n/2}\,\mathbf{1}_{[(2k)2^{-n-1},(2k+1)2^{-n-1})}(t) - 2^{n/2}\,\mathbf{1}_{[(2k+1)2^{-n-1},(2k+2)2^{-n-1})}(t).$$

1. Verify that the functions h_0, $(h_k^n)_{n\geq 1, 0\leq k\leq 2^n-1}$ form an orthonormal basis of $L^2([0,1], \mathscr{B}([0,1]), \mathrm{d}t)$. (*Hint*: Observe that, for every fixed $n \geq 0$, any function $f : [0,1) \to \mathbb{R}$ that is constant on every interval of the form $[(j-1)2^{-n}, j2^{-n})$, for $1 \leq j \leq 2^n$, is a linear combination of the functions h_0, $(h_k^m)_{0\leq m<n, 0\leq k\leq 2^m-1}$.)
2. Suppose that N_0, $(N_k^n)_{n\geq 1, 0\leq k\leq 2^n-1}$ are independent $\mathscr{N}(0,1)$ random variables. Justify the existence of the (unique) Gaussian white noise G on $[0,1]$, with intensity $\mathrm{d}t$, such that $G(h_0) = N_0$ and $G(h_k^n) = N_k^n$ for every $n \geq 0$ and $0 \leq k \leq 2^n - 1$.
3. For every $t \in [0,1]$, set $B_t = G([0,t])$. Verify that

$$B_t = t\,N_0 + \sum_{n=0}^{\infty}\Big(\sum_{k=0}^{2^n-1} g_k^n(t)\,N_k^n\Big),$$

   where the series converges in L^2, and the functions $g_k^n : [0,1] \to [0,\infty)$ are given by

$$g_k^n(t) = \int_0^t h_k^n(s)\,\mathrm{d}s.$$

   Note that the functions g_k^n are continuous and satisfy the following property: For every fixed $n \geq 0$, the functions g_k^n, $0 \leq k \leq 2^n - 1$, have disjoint supports and are bounded above by $2^{-n/2}$.
4. For every integer $m \geq 0$ and every $t \in [0,1]$ set

$$B_t^{(m)} = t\,N_0 + \sum_{n=0}^{m-1}\Big(\sum_{k=0}^{2^n-1} g_k^n(t)\,N_k^n\Big).$$

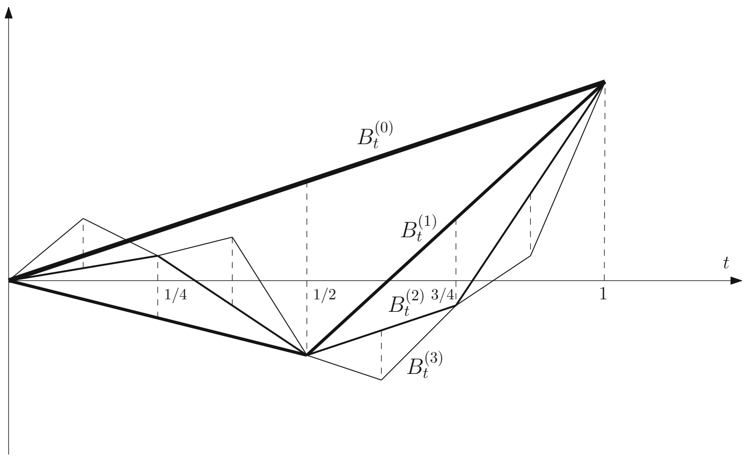

Fig. 1.1 Illustration of the construction of $B_t^{(m)}$ in Exercise 1.18, for $m = 0, 1, 2, 3$. For the clarity of the figure, lines become thinner when m increases. The lengths of the dashed segments are determined by the values of N_0 and N_k^m for $m = 0, 1, 2$

See Fig. 1.1 for an illustration. Verify that the continuous functions $t \mapsto B_t^{(m)}(\omega)$ converge uniformly on $[0, 1]$ as $m \to \infty$, for a.a. ω. (*Hint*: If N is $\mathcal{N}(0, 1)$-distributed, prove the bound $P(|N| \geq a) \leq e^{-a^2/2}$ for $a \geq 1$, and use this estimate to bound the probability of the event $\{\sup\{|N_k^n| : 0 \leq k \leq 2^n - 1\} > 2^{n/4}\}$, for every fixed $n \geq 0$.)

5. Conclude that we can, for every $t \geq 0$, select a random variable B'_t which is a.s. equal to B_t, in such a way that the mapping $t \mapsto B'_t(\omega)$ is continuous for every $\omega \in \Omega$.

Notes and Comments

The material in this chapter is standard. We refer to Adler [1] and Lifshits [55] for more information about Gaussian processes. The more recent book [56] by Marcus and Rosen develops striking applications of the known results about Gaussian processes to Markov processes and their local times. Exercise 1.16 involves a simple particular case of the famous Kalman filter, which has numerous applications in technology. See [49] or [62] for the details of the construction in Exercise 1.18.