1

Gaussian Processes

Gaussian processes are a powerful and versatile class of stochastic processes fundamental to modern statistics, machine learning, and various scientific and engineering disciplines. Their inherent strength lies in their ability to model complex functions and time-dependent phenomena probabilistically, offering not only predictions but also a robust quantification of uncertainty.

Defined by the property that any finite collection of their random variables follows a multivariate Gaussian distribution, these processes are exceptionally well suited for capturing intricate dependencies over time or space, making them invaluable for tasks ranging from regression and classification to the analysis of complex dynamic systems.

To prepare for our study of advanced processes like the generalized fractional Brownian motion, we must first establish the foundational building blocks of this field. This chapter is dedicated to this essential groundwork.

We begin by exploring the fundamental properties of Gaussian random variables and vectors and their distributions. We then move to a comprehensive study of Gaussian processes, defining them and discussing the conditions for their existence. As a specific and important example, we will define the Wiener process (Brownian motion) as a particular case of a Gaussian process, establish its existence and recall its fundamental properties. The chapter will conclude with an introduction to Gaussian random measures and the core principles of Wiener integration, providing the necessary theoretical toolkit for the chapters to come.

1.1. Gaussian variables

To start, we will define what makes a random variable Gaussian.

DEFINITION 1.1.– *A random variable X is called a Gaussian variable (or normal variable) if its probability density function (PDF) with respect to the Lebesgue measure is given by:*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Here, $m \in \mathbb{R}$ represents the mean and $\sigma > 0$ represents the standard deviation. We denote this as $X \sim \mathcal{N}(m, \sigma^2)$. In the special case where $m = 0$ and $\sigma = 1$, X is called a standard Gaussian variable, denoted $X \sim \mathcal{N}(0, 1)$.

Among the many properties satisfied by such variables, we will cite only those which will be particularly useful in this book.

LEMMA 1.1.– *If $X \sim \mathcal{N}(m, \sigma^2)$, then:*

1) The expected value is $\mathbb{E}[X] = m$.

2) The variance is $\mathrm{Var}[X] = \sigma^2$.

3) The characteristic function is given by:

$$\varphi_X(t) = \mathbb{E}[\exp(itX)] = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right).$$

4) If $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ is independent of X, then their sum $X + Y$ is also a Gaussian random variable:

$$X + Y \sim \mathcal{N}(m + m_2, \sigma^2 + \sigma_2^2).$$

5) For any $(\alpha, \beta) \in (\mathbb{R} \setminus \{0\}) \times \mathbb{R}$, the linear transformation $\alpha X + \beta$ is also a Gaussian random variable:

$$\alpha X + \beta \sim \mathcal{N}(\alpha m + \beta, \alpha^2 \sigma^2).$$

Building on these fundamental characteristics, we proceed to a vital property concerning the *moments of Gaussian variables*.

LEMMA 1.2.– *If $X \sim \mathcal{N}(m, \sigma^2)$, then its central moments are:*

– $\mathbb{E}[(X-m)^n] = 0$ *if n is an odd positive integer.*

– $\mathbb{E}[(X-m)^n] = \sigma^n (n-1)!!$ *if n is an even positive integer, where the double factorial is given by:*

$$(n-1)!! = (n-1)(n-3)\cdots 1 = \frac{n!}{2^{n/2}(n/2)!}.$$

In particular, for a standard Gaussian variable $Z \sim \mathcal{N}(0, 1)$:

– $\mathbb{E}[Z^n] = 0$ *if n is an odd positive integer.*
– $\mathbb{E}[Z^n] = (n-1)!!$ *if n is an even positive integer.*

Proof. The proof of this lemma is standard; however, its derivation provides significant understanding and is therefore given as Exercise 1.2. □

1.1.1. *Chi-squared distribution*

Building upon the concept of Gaussian variables, we now introduce a distribution fundamental to statistical inference that arises directly from sums of squared Gaussian variables: the chi-squared distribution.

DEFINITION 1.2.– *Let $Z_1, Z_2, \ldots, Z_k$ be k independent and identically distributed (i.i.d.) standard Gaussian random variables, that is, $Z_i \sim \mathcal{N}(0,1)$ for $i = 1, \ldots, k$. A random variable Y defined as the sum of their squares:*

$$Y = \sum_{i=1}^{k} Z_i^2$$

is said to follow a chi-squared distribution with k degrees of freedom, denoted as $Y \sim \chi^2(k)$.

Its PDF for $y > 0$ is given by:

$$f_Y(y;k) = \frac{1}{2^{k/2}\Gamma(k/2)} y^{(k/2)-1} e^{-y/2},$$

where $\Gamma(\cdot)$ is the Gamma function, defined for real numbers $x > 0$ by the integral:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

The chi-squared distribution possesses several key characteristics important for its application.

LEMMA 1.3.– *If $Y \sim \chi^2(k)$ with k degrees of freedom, then:*

1) The expected value (mean) is $\mathbb{E}[Y] = k$.
2) The variance is $\mathrm{Var}[Y] = 2k$.
3) The characteristic function is given by: $\varphi_Y(t) = (1-2it)^{-k/2}$.

1.2. Gaussian vectors

Having defined Gaussian variables, we now move to multivariate settings by introducing the definition of a Gaussian vector.

DEFINITION 1.3.– *Let $\langle x, y\rangle = x_1y_1 + ... + x_ny_n$ denote the Euclidean inner product in $\mathbb{R}^n$ for $x, y \in \mathbb{R}^n$. A random vector $X = (X_1, X_2, ..., X_n)'$ is called a Gaussian vector if any linear combination of its components,*

$$\langle \lambda, X\rangle = \lambda_1 X_1 + \lambda_2 X_2 + ... + \lambda_n X_n,$$

is a Gaussian random variable for any vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)' \in \mathbb{R}^n$. Here, the prime notation $(.)'$ denotes the transpose of the vector.

If X is a Gaussian vector, its *expectation vector* is:

$$M_X = \mathbb{E}[X] = (\mathbb{E}[X_1], ..., \mathbb{E}[X_n])',$$

and its *covariance matrix* is:

$$K_X = \mathrm{Cov}(X) = (\mathrm{Cov}(X_i, X_j))_{1\le i,j\le n}.$$

We denote this as $X \sim \mathcal{N}_n(M_X, K_X)$. A special case is the *standard Gaussian vector*, where $M_X = \mathbf{0}$ (the zero vector) and $K_X = I_n$ (the identity matrix).

For Gaussian vectors, we will focus on their most relevant properties for this book.

LEMMA 1.4.– *If $X \sim \mathcal{N}_n(M_X, K_X)$, its characteristic function is given by:*

$$\varphi_X(\lambda) = \mathbb{E}\left[\exp\{i\lambda' X\}\right] = \exp\left(i\langle\lambda, M_X\rangle - \frac{1}{2}\langle\lambda, K_X\lambda\rangle\right).$$

This can also be written in matrix notation as:

$$\varphi_X(\lambda) = \exp\left(i\lambda' M_X - \frac{1}{2}\lambda' K_X\lambda\right)$$

for any $\lambda \in \mathbb{R}^n$.

A crucial property of Gaussian vectors is their behavior under linear transformations. Specifically, a linear transformation of a Gaussian vector remains a Gaussian vector, a fact we formally state and prove below.

LEMMA 1.5.– *If $\mathbf{Y}$ is a Gaussian vector, $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $S = \mathbf{b}'\mathbf{Y}$ where $\mathbf{b}$ is a constant vector, then S is a Gaussian random variable, specifically:*

$$S = \mathbf{b}'\mathbf{Y} \sim \mathcal{N}\left(\mathbf{b}'\boldsymbol{\mu}, \mathbf{b}'\boldsymbol{\Sigma}\mathbf{b}\right).$$

Proof. Since the scalar $S = \mathbf{b}'\mathbf{Y} = \sum_{j=1}^{n} b_j Y_j$ is a linear combination of the components of a Gaussian vector, S must be a Gaussian random variable by definition. To determine its specific distribution, we need to find its mean and variance.

For the mean of S, using the linearity of expectation:

$$\mathbb{E}[S] = \mathbb{E}[\mathbf{b}'\mathbf{Y}] = \mathbf{b}'\mathbb{E}[\mathbf{Y}] = \mathbf{b}'\boldsymbol{\mu}.$$

Now, the variance of a random variable S is $\text{Var}(S) = \mathbb{E}[(S-\mathbb{E}[S])^2]$. Substituting $S = \mathbf{b}'\mathbf{Y}$ and $\mathbb{E}[S] = \mathbf{b}'\boldsymbol{\mu}$:

$$\begin{aligned}\text{Var}(S) &= \mathbb{E}[(\mathbf{b}'\mathbf{Y} - \mathbf{b}'\boldsymbol{\mu})^2] \\ &= \mathbb{E}[(\mathbf{b}'(\mathbf{Y} - \boldsymbol{\mu}))^2] \\ &= \mathbb{E}[(\mathbf{b}'(\mathbf{Y} - \boldsymbol{\mu}))(\mathbf{Y} - \boldsymbol{\mu})'\mathbf{b}] \\ &= \mathbf{b}'\mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})']\mathbf{b}.\end{aligned}$$

By definition, $\mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})']$ is the covariance matrix of $\mathbf{Y}$, which is $\boldsymbol{\Sigma}$. Therefore,

$$\text{Var}(S) = \mathbf{b}'\boldsymbol{\Sigma}\mathbf{b}.$$

Thus, S follows a Gaussian distribution with mean $\mathbf{b}'\boldsymbol{\mu}$ and variance $\mathbf{b}'\boldsymbol{\Sigma}\mathbf{b}$. □

Extending this concept to transformations resulting in a vector, the following proposition demonstrates that any linear transformation of a Gaussian vector produces another Gaussian vector.

PROPOSITION 1.1.– *If $\mathbf{Z}$ is a Gaussian vector, $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \Gamma)$, and $\mathbf{V} = P'\mathbf{Z}$ where P is a constant $n \times n$ matrix and P' is its transpose, then $\mathbf{V}$ is also a Gaussian vector, specifically:*

$$\mathbf{V} = P'\mathbf{Z} \sim \mathcal{N}_n\left(\mathbf{0}, P'\Gamma P\right).$$

Proof. The proof proceeds by determining the mean vector and covariance matrix of $\mathbf{V}$ and then showing that its characteristic function has the Gaussian form.

The mean of $\mathbf{V}$ is given by the expectation of $\mathbf{V}$: $\mathbb{E}[\mathbf{V}] = \mathbb{E}[P'\mathbf{Z}]$.

Since P' is a constant matrix, we can take it out of the expectation:

$$\mathbb{E}[\mathbf{V}] = P'\mathbb{E}[\mathbf{Z}] = 0,$$

since $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \Gamma)$. Thus, the mean of $\mathbf{V}$ is also the zero vector.

The covariance matrix of $\mathbf{V}$, denoted by $\Sigma_{\mathbf{V}}$, is defined as:

$$\Sigma_{\mathbf{V}} = \mathbb{E}\left[(\mathbf{V} - \mathbb{E}[\mathbf{V}])(\mathbf{V} - \mathbb{E}[\mathbf{V}])'\right].$$

Since $\mathbb{E}[\mathbf{V}] = \mathbf{0}$, this simplifies to:

$$\Sigma_{\mathbf{V}} = \mathbb{E}\left[\mathbf{V}\mathbf{V}'\right] = \mathbb{E}\left[(P'\mathbf{Z})(P'\mathbf{Z})'\right].$$

Using the properties $(AB)' = B'A'$ and $(A')' = A$, we have:

$$(P'\mathbf{Z})' = \mathbf{Z}'(P')' = \mathbf{Z}'P.$$

Thus,

$$\Sigma_{\mathbf{V}} = \mathbb{E}\left[P'\mathbf{Z}\mathbf{Z}'P\right].$$

Since P' and P are constant matrices, we can take them out of the expectation:

$$\Sigma_{\mathbf{V}} = P'\mathbb{E}\left[\mathbf{Z}\mathbf{Z}'\right]P.$$

The covariance matrix of $\mathbf{Z}$, denoted by $\Sigma_{\mathbf{Z}}$, is given by:

$$\Sigma_{\mathbf{Z}} = \mathbb{E}\left[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])'\right] = \mathbb{E}\left[\mathbf{Z}\mathbf{Z}'\right] = \Gamma.$$

Substituting this back into the expression for $\Sigma_{\mathbf{V}}$: $\Sigma_{\mathbf{V}} = P'\Gamma P$.

The characteristic function of $\mathbf{V}$ can be found by substituting its definition into the characteristic function of a random vector:

$$\phi_{\mathbf{V}}(\mathbf{t}) = \mathbb{E}\left[e^{i\mathbf{t}'\mathbf{V}}\right] = \mathbb{E}\left[e^{i\mathbf{t}'(P'\mathbf{Z})}\right].$$

Using the property $\mathbf{t}'(P'\mathbf{Z}) = (P\mathbf{t})'\mathbf{Z}$ and letting $\mathbf{s} = P\mathbf{t}$, we can express this in terms of the characteristic function of $\mathbf{Z}$:

$$\phi_{\mathbf{V}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{s}'\mathbf{Z}}) = \phi_{\mathbf{Z}}(\mathbf{s}).$$

Since $\mathbf{Z}$ is a Gaussian vector with mean $\mathbf{0}$ and covariance matrix Γ, its characteristic function is given by:

$$\phi_{\mathbf{Z}}(\mathbf{s}) = e^{-\frac{1}{2}\mathbf{s}'\Gamma\mathbf{s}}.$$

Substituting $\mathbf{s} = P\mathbf{t}$ into this expression gives us the characteristic function of $\mathbf{V}$:

$$\phi_{\mathbf{V}}(\mathbf{t}) = e^{-\frac{1}{2}(P\mathbf{t})'\Gamma(P\mathbf{t})} = e^{-\frac{1}{2}\mathbf{t}'P'\Gamma P\mathbf{t}}.$$

As the covariance matrix of $\mathbf{V}$ is $\Sigma_{\mathbf{V}} = P'\Gamma P$, the characteristic function simplifies to:

$$\phi_{\mathbf{V}}(\mathbf{t}) = e^{-\frac{1}{2}\mathbf{t}'\Sigma_{\mathbf{V}}\mathbf{t}}.$$

This is the characteristic function of a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathbf{V}}$.

This completes the proof. □

For a complete understanding of their distribution in Euclidean space, it is essential to define the PDF of a Gaussian vector. The following lemma provides this density for the nondegenerate case, followed by a crucial remark concerning singular covariance matrices.

LEMMA 1.6.– *If $X \sim \mathcal{N}_n(M_X, K_X)$ with a nondegenerate (invertible) covariance matrix K_X, then its PDF is:*

$$f_X(x) = (2\pi)^{-n/2}(\det K_X)^{-1/2} \exp\left(-\frac{1}{2}(x - M_X)' K_X^{-1}(x - M_X)\right)$$

where $\det K_X$ is the determinant of K_X, and K_X^{-1} is its inverse.

REMARK 1.1.– *It is important to note that the covariance matrix K_X of a Gaussian random vector is not always invertible. It can be singular (i.e. have a determinant of zero) under certain conditions. When K_X is singular, it signifies that the Gaussian random vector X is degenerate. This means there exist linear dependencies among its components. Specifically, one or more components X_i can be expressed as an affine (linear plus constant) transformation of the other components. For example, if $X_n = c_1X_1 + \cdots + c_{n-1}X_{n-1} + c_0$ for some constants c_i, then K_X would be singular. The direct consequence of this degeneracy is that the probability distribution of X is not spread out over the entire n-dimensional space $\mathbb{R}^n$. Instead, it is concentrated on a lower dimensional affine subspace of $\mathbb{R}^n$. As the distribution is confined to such a subspace (which has zero Lebesgue measure in $\mathbb{R}^n$), the Gaussian vector does not possess a PDF with respect to the Lebesgue measure on $\mathbb{R}^n$. In these cases, we might refer to its distribution as a degenerate Gaussian or note that its density exists only with respect to a lower dimensional Lebesgue measure on the affine subspace it occupies.*

1.2.1. *Isserlis' theorem for Gaussian moments*

When dealing with expectations of products of Gaussian random variables, especially for orders higher than two, a powerful tool known as *Isserlis' theorem* (also known as Wick's theorem or the Gaussian moment theorem) significantly simplifies calculations. It provides a general formula for the expected value of the product of any number of zero-mean jointly Gaussian random variables.

THEOREM 1.1.– *Let $X_1, X_2, \ldots, X_n$ be jointly Gaussian random variables with zero mean ($\mathbb{E}[X_i] = 0$ for all i). Let $K_{ij} = \mathbb{E}[X_iX_j]$ be their covariances.*

1) If n is an odd integer, then $\mathbb{E}[X_1X_2 \cdots X_n] = 0$.

2) If n is an even integer, then:

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \sum \prod_{pairs} \mathbb{E}[X_a X_b],$$

where the sum is over all distinct ways of partitioning the n variables into $n/2$ pairs. Each product consists of the expected values of these pairs.

Proof. The proof of Isserlis' theorem can be found in Vignat (2012) and the references therein. □

REMARK 1.2 (Isserlis' theorem for Gaussian variables).– *For a zero-mean multivariate Gaussian vector $X = (X_1, \ldots, X_n)$ with covariance matrix $K_X = \mathbb{E}[XX']$:*

1) Third-order moment: for any three components X_i, X_j, X_k, the third-order moment is zero.

$$\mathbb{E}[X_i X_j X_k] = 0.$$

2) Fourth-order moment: for any four components X_i, X_j, X_k, X_l, the fourth-order moment is the sum of all distinct pairings of the covariances.

$$\mathbb{E}[X_i X_j X_k X_l] = \mathbb{E}[X_i X_j]\mathbb{E}[X_k X_l] + \mathbb{E}[X_i X_k]\mathbb{E}[X_j X_l] + \mathbb{E}[X_i X_l]\mathbb{E}[X_j X_k]. \quad [1.1]$$

This corresponds to the three pairings: (X_i, X_j) and (X_k, X_l); (X_i, X_k) and (X_j, X_l); and (X_i, X_l) and (X_j, X_k).

As a direct consequence of Isserlis' theorem, we get the following corollary which provides a fundamental relationship between the covariance of two zero-mean Gaussian random variables of their squares and their original covariance.

COROLLARY 1.1.– *For any two jointly zero-mean Gaussian random variables X and Y, the covariance of their squares is given by:*

$$Cov(X^2, Y^2) = 2(Cov(X, Y))^2.$$

Proof. To prove this, we start with the definition of covariance:

$$\text{Cov}(X^2, Y^2) = \mathbb{E}[X^2 Y^2] - \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

Since X and Y are zero-mean Gaussian random variables, we know that:

$$\mathbb{E}[X^2] = \text{Var}(X) \text{ and } \mathbb{E}[Y^2] = \text{Var}(Y).$$

Also, by Theorem 1.1, for zero-mean Gaussian random variables, the fourth-order moment $\mathbb{E}[X_1 X_2 X_3 X_4]$ is the sum of products of pairwise covariances. For $\mathbb{E}(X^2 Y^2)$, we have $X_1 = X, X_2 = X, X_3 = Y, X_4 = Y$. The possible pairings are:

– (X_1, X_2) and $(X_3, X_4) \implies \mathbb{E}[X \cdot X]\mathbb{E}[Y \cdot Y] = \text{Var}(X)\text{Var}(Y)$;
– (X_1, X_3) and $(X_2, X_4) \implies \mathbb{E}[X \cdot Y]\mathbb{E}[X \cdot Y] = (\text{Cov}(X, Y))^2$;
– (X_1, X_4) and $(X_2, X_3) \implies \mathbb{E}[X \cdot Y]\mathbb{E}[X \cdot Y] = (\text{Cov}(X, Y))^2$.

Summing these pairings, we get:

$$\mathbb{E}[X^2Y^2] = \text{Var}(X)\text{Var}(Y) + 2(\text{Cov}(X, Y))^2$$

Now substitute this back into the covariance formula:

$$\begin{aligned}\text{Cov}(X^2, Y^2) &= \text{Var}(X)\text{Var}(Y) + 2(\text{Cov}(X, Y))^2 - \text{Var}(X)\text{Var}(Y) \\ &= 2(\text{Cov}(X, Y))^2,\end{aligned}$$

which completes the proof. □

Moving beyond linear transformations, understanding higher order moments of Gaussian vectors, especially those involving quadratic forms, is crucial. The following lemma provides a powerful and widely used result for the expectation of the product of two quadratic forms from a zero-mean Gaussian vector. Its technical proof relies on Isserlis' theorem, but given its direct utility in this book, we will state the result without derivation.

LEMMA 1.7.– *Let $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma})$ be an n-dimensional zero-mean Gaussian vector with covariance matrix $\mathbf{\Sigma}$. For any symmetric $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, the expectation of the product of two quadratic forms $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ and $\mathbf{Y}'\mathbf{B}\mathbf{Y}$ is given by:*

$$\mathbb{E}[(\mathbf{Y}'\mathbf{A}\mathbf{Y})(\mathbf{Y}'\mathbf{B}\mathbf{Y})] = \mathit{Tr}(\mathbf{A}\mathbf{\Sigma})\mathit{Tr}(\mathbf{B}\mathbf{\Sigma}) + 2\mathit{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}).$$

Here, $\mathit{Tr}(\mathbf{M})$ denotes the trace of a square matrix $\mathbf{M}$, which is defined as the sum of its diagonal elements: $\mathit{Tr}(\mathbf{M}) = \sum_{i=1}^{n} M_{ii}$.

As a direct application of Lemma 1.7, the following corollary derives the expected value of a quadratic form multiplied by a squared linear form.

COROLLARY 1.2.– *Let $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma})$ be an n-dimensional zero-mean Gaussian vector with covariance matrix $\mathbf{\Sigma}$. For any symmetric $n \times n$ matrix $\mathbf{A}$ and any constant $n \times 1$ vector $\mathbf{b}$, the expected value of the product of the quadratic form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ and the squared linear form $(\mathbf{b}'\mathbf{Y})^2$ is given by:*

$$\mathbb{E}[(\mathbf{Y}'\mathbf{A}\mathbf{Y})(\mathbf{b}'\mathbf{Y})^2] = 2\mathit{Tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}) + \mathit{Tr}(\mathbf{A}\mathbf{\Sigma})\mathbf{b}'\mathbf{\Sigma}\mathbf{b}.$$

The full proof of Corollary 1.2 is provided as Exercise 1.6, where hints for its derivation can be found.

1.2.2. *Moments of quadratic and squared linear forms*

We now present a particularly useful result for calculating the expectation of the square of a linear form of a centered Gaussian random vector.

LEMMA 1.8.– *Let $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ be a centered Gaussian random vector (i.e. $\mathbb{E}[\mathbf{Z}] = \mathbf{0}$) with a non-singular covariance matrix $\mathbf{\Gamma}_Z$. Let $\mathbf{t} = (t_1, \ldots, t_n)'$ be a deterministic vector. Then:*

$$\mathbb{E}[(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})^2] = \mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{t}.$$

Proof. Let the linear form be $Y = \mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z}$. We want to compute $\mathbb{E}[Y^2]$.

Since Y is a scalar, $Y^2 = YY$. We can write this as:

$$Y^2 = (\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z}).$$

For any scalar c, $c = c'$. Applied to our scalar linear form, $(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})' = \mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z}$. By the property of transpose of a product $(AB)' = B'A'$, we also have:

$$(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})' = \mathbf{Z}'(\mathbf{\Gamma}_Z^{-1})'(\mathbf{t}')' = \mathbf{Z}'(\mathbf{\Gamma}_Z^{-1})'\mathbf{t}.$$

Given that $\mathbf{\Gamma}_Z$ is a covariance matrix, it is symmetric, that is, $\mathbf{\Gamma}_Z = \mathbf{\Gamma}_Z'$. Consequently, its inverse is also symmetric, $(\mathbf{\Gamma}_Z^{-1})' = \mathbf{\Gamma}_Z^{-1}$. Therefore, we can equate the scalar form with its transpose:

$$\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z} = \mathbf{Z}'\mathbf{\Gamma}_Z^{-1}\mathbf{t}.$$

Using this, we can rewrite the product within the expectation as:

$$\mathbb{E}[(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})^2] = \mathbb{E}[(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})(\mathbf{Z}'\mathbf{\Gamma}_Z^{-1}\mathbf{t})]. \quad [1.2]$$

The vectors $\mathbf{t}$ and the matrix $\mathbf{\Gamma}_Z^{-1}$ are deterministic constants. By the linearity of the expectation operator, these constant matrices and vectors can be moved outside the expectation:

$$\mathbb{E}[\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{\Gamma}_Z^{-1}\mathbf{t}] = \mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}']\mathbf{\Gamma}_Z^{-1}\mathbf{t}. \quad [1.3]$$

Since $\mathbf{Z}$ is a centered random vector $\mathbf{Z}$, its covariance matrix is defined as:

$$\mathrm{Cov}(\mathbf{Z}) = \mathbb{E}[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])'] = \mathbb{E}[\mathbf{Z}\mathbf{Z}'].$$

By definition, the covariance matrix of $\mathbf{Z}$ is $\mathbf{\Gamma}_Z$. Therefore, we have $\mathbb{E}[\mathbf{Z}\mathbf{Z}'] = \mathbf{\Gamma}_Z$.

Substitute this back into the expression [1.3] by [1.2], we get:

$$\mathbb{E}[(\mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{Z})^2] = \mathbf{t}^T\mathbf{\Gamma}_Z^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}']\mathbf{\Gamma}_Z^{-1}\mathbf{t} = \mathbf{t}'\mathbf{\Gamma}_Z^{-1}\mathbf{\Gamma}_Z\mathbf{\Gamma}_Z^{-1}\mathbf{t}.$$

Since $\mathbf{\Gamma}_Z^{-1}\mathbf{\Gamma}_Z = I_n$ (the identity matrix of dimension n), the expression simplifies to:

$$\mathbb{E}[(\mathbf{t}^T\mathbf{\Gamma}_Z^{-1}\mathbf{Z})^2] = \mathbf{t}^T\mathbf{\Gamma}_Z^{-1}\mathbf{t}. \qquad \square$$

To further characterize Gaussian vectors, we next derive the expected value of a quadratic form.

LEMMA 1.9.– *Let $X \sim \mathcal{N}_n(\mathbf{0}, K_X)$ be an n-dimensional zero-mean multivariate Gaussian vector with covariance matrix K_X. Let $\mathbf{A}$ be a constant $n \times n$ matrix. Then, the expected value of the quadratic form $X'\mathbf{A}X$ is given by:*

$$\mathbb{E}[X'\mathbf{A}X] = Tr(\mathbf{A}K_X).$$

Proof. Let $X = (X_1, \ldots, X_n)'$ be the Gaussian vector and $\mathbf{A}$ be an $n \times n$ matrix with elements A_{ij}. The quadratic form $X'\mathbf{A}X$ can be written in summation notation as:

$$X'\mathbf{A}X = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}X_iX_j.$$

Taking the expected value of both sides and using the linearity of expectation, we get:

$$\mathbb{E}[X'\mathbf{A}X] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}X_iX_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\mathbb{E}[X_iX_j].$$

Since X is a zero-mean vector, we have $\mathbb{E}[X_iX_j] = \mathrm{Cov}(X_i, X_j)$. By definition, $\mathrm{Cov}(X_i, X_j)$ is the (i,j)th element of the covariance matrix K_X, denoted as $(K_X)_{ij}$. Substituting this into the expectation equation, we have:

$$\mathbb{E}[X'\mathbf{A}X] = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(K_X)_{ij} \quad (*)$$

Now, let us consider the trace of the product $\mathbf{A}K_X$. Let $\mathbf{C} = \mathbf{A}K_X$. The elements of $\mathbf{C}$ are $C_{ik} = \sum_{p=1}^{n} A_{ip}(K_X)_{pk}$. The trace of $\mathbf{C}$ is the sum of its diagonal elements:

$$\mathrm{Tr}(\mathbf{A}K_X) = \sum_{i=1}^{n} C_{ii} = \sum_{i=1}^{n}\sum_{p=1}^{n} A_{ip}(K_X)_{pi}.$$

Since the covariance matrix K_X is symmetric (i.e. $(K_X)_{pi} = (K_X)_{ip}$), we can substitute this property into the trace sum:

$$\mathrm{Tr}(\mathbf{A}K_X) = \sum_{i=1}^{n}\sum_{p=1}^{n} A_{ip}(K_X)_{ip}.$$

By simply re-indexing the dummy summation variable p to j, we obtain:

$$\text{Tr}(\mathbf{A}K_X) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(K_X)_{ij} \quad (**)$$

Comparing equations $(*)$ and $(**)$, we see that:

$$\mathbb{E}[X'\mathbf{A}X] = \text{Tr}(\mathbf{A}K_X).$$

This completes the proof. □

A direct application of the previous lemma yields a particularly insightful result.

COROLLARY 1.3.– *If $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \Gamma)$ is an n-dimensional zero-mean Gaussian vector with an invertible covariance matrix Γ, then:*

$$\mathbb{E}\left[\mathbf{Z}'\Gamma^{-1}\mathbf{Z}\right] = n.$$

Proof. This result is a direct application of Lemma 1.9. We identify the terms from the corollary with those in the lemma: let $X = \mathbf{Z}$, the covariance matrix $K_X = \mathbf{\Gamma}$, and the constant matrix $\mathbf{A} = \mathbf{\Gamma}^{-1}$. Since $\mathbf{Z}$ is a zero-mean Gaussian vector, we can apply Lemma 1.9:

$$\mathbb{E}[\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z}] = \text{Tr}(\mathbf{\Gamma}^{-1}\mathbf{\Gamma}) = \text{Tr}(I_n),$$

where I_n is the $n \times n$ identity matrix. The trace of the identity matrix is the sum of its diagonal elements, which are all 1s, therefore $\text{Tr}(I_n) = n$. Thus, we conclude:

$$\mathbb{E}[\mathbf{Z}'\mathbf{\Gamma}^{-1}\mathbf{Z}] = n.$$ □

A particularly important result in the analysis of Gaussian random vectors, completing their characterization through moments, concerns the variance of their quadratic forms.

LEMMA 1.10.– *Let $X \sim \mathcal{N}_n(\mathbf{0}, K_X)$ be an n-dimensional zero-mean multivariate Gaussian vector with covariance matrix K_X. Let $\mathbf{A}$ be a constant $n \times n$ matrix. Then, the variance of the quadratic form $X'\mathbf{A}X$ is given by:*

$$\text{Var}(X'\mathbf{A}X) = 2\mathit{Tr}((\mathbf{A}K_X)^2).$$

Proof. The proof for this result is left as an exercise for the reader and can be found with guiding hints in Exercise 1.5. □

Building on the results for the variance and expectation of quadratic forms, we can derive the expected value of the square of a specific quadratic form.

COROLLARY 1.4.– *If* $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \Gamma)$ *is a zero-mean Gaussian vector with an invertible covariance matrix* Γ*, then:*

$$\mathbb{E}\left[(\mathbf{Z}'\Gamma^{-1}\mathbf{Z})^2\right] = n^2 + 2n.$$

Proof. Let $Q = \mathbf{Z}'\Gamma^{-1}\mathbf{Z}$. We have $\mathbb{E}[Q^2] = \mathrm{Var}(Q) + (\mathbb{E}[Q])^2$. From Corollary 1.3, we directly have:

$$\mathbb{E}[Q] = \mathbb{E}[\mathbf{Z}'\Gamma^{-1}\mathbf{Z}] = n.$$

From Lemma 1.10,

$$\mathrm{Var}(Q) = \mathrm{Var}(\mathbf{Z}'\Gamma^{-1}\mathbf{Z}) = 2\mathrm{Tr}((\Gamma^{-1}\Gamma)^2) = 2\,\mathrm{Tr}(I_n) = 2n.$$

Therefore,

$$\mathbb{E}[Q^2] = n^2 + 2n. \qquad \square$$

1.3. Gaussian processes

Having established the fundamental properties of Gaussian random variables and vectors, we now extend these concepts to the realm of *stochastic processes*. This extension leads us to Gaussian processes, a cornerstone of statistical modeling with wide-ranging applications.

1.3.1. *Generalities on stochastic processes*

Before diving into Gaussian processes specifically, let us set the stage with some essential definitions for stochastic processes. A stochastic process can be thought of as a collection of random variables indexed by some set, often representing time or space.

DEFINITION 1.4.– *Consider a non-empty set* $\mathbb{T}$*, a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and a function* $X : \mathbb{T} \times \Omega \to \mathbb{R}^d$*, where* $d \geq 1$*. If, for any fixed* $t \in \mathbb{T}$*,* $X(t, \cdot)$ *is an* $\mathbb{R}^d$*-valued random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$*, then* X *is called a stochastic process with values in* $\mathbb{R}^d$*. We commonly denote a stochastic process as* $X = (X_t)_{t\in\mathbb{T}}$*, where* $X_t(\omega) = X(t, \omega)$ *for* $\omega \in \Omega$ *and* $t \in \mathbb{T}$*.*

This fundamental definition comes with key associated concepts:

– For a fixed $\omega \in \Omega$, the function $X(\cdot, \omega)$ is termed a *trajectory* or a *sample path* of the process X. This represents a single realization of the process over its entire index set $\mathbb{T}$.

– The *mean function* of X is defined by $m : t \in \mathbb{T} \longmapsto \mathbb{E}[X_t]$. This function exists if, for all $t \in \mathbb{T}$, every component of X_t is integrable (i.e. $\mathbb{E}[|X_{t,i}|] < \infty$ for all $i = 1, \ldots, d$). This function describes the average behavior of the process at each point $t \in \mathbb{T}$.

– The *covariance function* of X is defined by:

$$C : (s,t) \in \mathbb{T}^2 \longmapsto C(s,t) = \mathrm{Cov}(X_s, X_t).$$

This function exists if, for all $t \in \mathbb{T}$, every component of X_t is square-integrable (i.e. $\mathbb{E}[|X_{t,i}|^2] < \infty$ for all $i = 1, \ldots, d$). $C(s,t)$ is a $d \times d$ matrix that quantifies the linear dependence between the process values at any two points s and t in the index set.

We now introduce the formal concept of the finite-dimensional distributions.

DEFINITION 1.5.– *The finite-dimensional distributions (f.d.d.'s) of a stochastic process $X = (X_t)_{t\in\mathbb{T}}$ are the joint probability distributions of the random vectors $(X_{t_1}, \ldots, X_{t_n})$, for every choice of $n \geq 1$ and any set of indices $(t_1, \ldots, t_n) \in \mathbb{T}^n$. These distributions are crucial as they capture the probabilistic relationships among any finite number of points in the process*[1].

When comparing different stochastic processes, we use specific terms to describe their relationships, focusing on how their sample paths or distributions align.

DEFINITION 1.6 (relationships between stochastic processes).– *Consider two stochastic processes $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$.*

– The processes X and Y are said to be indistinguishable if:

$$\mathbb{P}(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega), \forall t \in \mathbb{T}\}) = 1.$$

In simpler terms, X and Y are indistinguishable if they share the exact same sample paths almost surely. This is the strongest form of equivalence.

– The processes X and Y are said to be modifications if for every $t \in \mathbb{T}$:

$$\mathbb{P}(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}) = 1.$$

This means that for any specific time t, the random variables X_t and Y_t are almost surely equal. While indistinguishable processes are always modifications, the reverse is not necessarily true (e.g. if paths can differ on a set of measure zero that varies with t).

1 While the finite-dimensional distributions are essential for defining the process's existence, they are generally insufficient to uniquely characterize properties of the sample paths, such as continuity or differentiability.

– The processes X and Y are said to be versions if they have the same finite-dimensional distributions. This implies that for any $n \geq 1$ and any $(t_1, \ldots, t_n) \in \mathbb{T}^n$, the random vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(Y_{t_1}, \ldots, Y_{t_n})$ have the same joint distribution.

– The processes X and Y are said to be independent if for any positive integers $m, n \geq 1$, any finite set of time points $s_1, \ldots, s_m \in \mathbb{T}$ for X, and any finite set of time points $t_1, \ldots, t_n \in \mathbb{T}$ for Y, the random vector $(X_{s_1}, \ldots, X_{s_m})$ is stochastically independent of the random vector $(Y_{t_1}, \ldots, Y_{t_n})$.

– The process $Y = (Y_t)_{t\in\mathbb{T}}$ is an independent copy of X if it possesses the same finite-dimensional distributions as X and is stochastically independent of X. This means that for any finite collection of random variables from X (e.g. $X_{s_1}, \ldots, X_{s_m}$) and any finite collection of random variables from Y (e.g. $Y_{t_1}, \ldots, Y_{t_n}$), these two collections are jointly independent.

To better understand the nuances of these definitions, let us establish the precise relationships between the three types of process equivalences.

PROPOSITION 1.2.– *Let X and Y be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with index set $\mathbb{T}$. The following implications hold:*

1) If X and Y are indistinguishable, then they are modifications.

2) If X and Y are modifications, then they are versions.

In summary, we have the hierarchy:

$$\textit{Indistinguishable} \implies \textit{Modifications} \implies \textit{Versions}$$

The reverse implications do not hold in general.

Proof. The proof of this proposition directly follows from the definitions and can be found in standard stochastic process literature (see, e.g. Karatzas and Shreve (1998)). The fact that the reverse implications do not hold in general is demonstrated by the counterexamples provided in Exercises 1.8 and 1.9. □

However, a significant special case arises when processes possess continuous sample paths, allowing us to establish a stronger equivalence.

PROPOSITION 1.3.– *Let $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$ be two stochastic processes, where $\mathbb{T}$ is an interval of $\mathbb{R}$ (e.g. $[0, \infty)$ or $\mathbb{R}$). If X and Y are modifications of each other and both processes have almost surely continuous sample paths, then they are indistinguishable.*

Proof. Let $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Since X and Y are modifications of each other, by definition, for every $t \in \mathbb{T}$, we have $\mathbb{P}(X_t = Y_t) = 1$.

Let $Q = \mathbb{Q} \cap \mathbb{T}$ be the countable set of rational numbers within the index set $\mathbb{T}$. For each $q \in Q$, we know $\mathbb{P}(X_q = Y_q) = 1$. Consider the event:

$$A_q = \{\omega \in \Omega : X_q(\omega) \neq Y_q(\omega)\}.$$

We have $\mathbb{P}(A_q) = 0$ for all $q \in Q$. Now consider the event $A = \bigcup_{q \in Q} A_q$, which represents the event that $X_q \neq Y_q$ for at least one rational q. By the subadditivity of probability for a countable union, we have:

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{q \in Q} A_q\right) \leq \sum_{q \in Q} \mathbb{P}(A_q) = \sum_{q \in Q} 0 = 0.$$

Therefore, $\mathbb{P}(A) = 0$. This implies that:

$$\mathbb{P}(A^c) = \mathbb{P}(\{\omega \in \Omega : X_q(\omega) = Y_q(\omega), \forall q \in Q\}) = 1.$$

Let $\Omega_0 = A^c$.

Next, both processes X and Y have almost surely continuous sample paths. This means there exists a set $\Omega_1 \subseteq \Omega$ with $\mathbb{P}(\Omega_1) = 1$ such that for every $\omega \in \Omega_1$, both $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are continuous functions on $\mathbb{T}$.

Let $\Omega^* = \Omega_0 \cap \Omega_1$. Since $\mathbb{P}(\Omega_0) = 1$ and $\mathbb{P}(\Omega_1) = 1$, it follows that $\mathbb{P}(\Omega^*) = 1$. Now, consider any $\omega \in \Omega^*$. For this ω: on the one hand, the functions $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are continuous on $\mathbb{T}$ (because $\omega \in \Omega_1$). On the other hand, $X_q(\omega) = Y_q(\omega)$ for all $q \in Q = \mathbb{Q} \cap \mathbb{T}$ (because $\omega \in \Omega_0$).

Since the set of rational numbers Q is dense in $\mathbb{T}$, and continuous functions that agree on a dense set must agree everywhere, we can conclude that for every $\omega \in \Omega^*$, $X_t(\omega) = Y_t(\omega)$ for all $t \in \mathbb{T}$.

Since $\mathbb{P}(\Omega^*) = 1$, this means that the event $\{\omega \in \Omega : X_t(\omega) = Y_t(\omega), \forall t \in \mathbb{T}\}$ has probability 1. By definition, this means X and Y are indistinguishable. □

Beyond their basic definitions, some stochastic processes exhibit particular temporal dependencies, which are vital for modeling dynamic systems.

DEFINITION 1.7.– *Consider $X = (X_t)_{t \in \mathbb{T}}$ a stochastic process, where $\mathbb{T}$ denotes a non-empty sub-interval of $\mathbb{R}$ (often representing time).*

– The process X has stationary increments if for any $n \geq 1$, any sequence of time points $t_1, \ldots, t_n \in \mathbb{T}$, and any time shift $h \in \mathbb{R}$ such that $t_i + h \in \mathbb{T}$ for all $i = 1, \ldots, n$, the joint distribution of the increments:

$$(X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$$

is the same as the joint distribution of the shifted increments:

$$(X_{t_2+h} - X_{t_1+h}, \ldots, X_{t_n+h} - X_{t_{n-1}+h}).$$

This means that the joint distribution of the increments is invariant under time shifts.

– The process X has independent increments if for any $n \geq 1$ and any sequence of ordered time points $t_1 < t_2 < \cdots < t_n$ in $\mathbb{T}$, the random variables $X_{t_2} - X_{t_1}$, $X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent. This implies that changes over non-overlapping time intervals are independent.

– The process X is said to be H-self-similar (or to possess H-self-similarity) for some constant $H > 0$, if for any $a > 0$, the process $(X_{at})_{t\in\mathbb{T}}$ has the same finite-dimensional distributions as $(a^H X_t)_{t\in\mathbb{T}}$. Formally, this means:

$$(X_{at})_{t\in\mathbb{T}} \overset{d}{=} (a^H X_t)_{t\in\mathbb{T}}.$$

This property implies that if we rescale time by a factor a, the process itself scales by a factor of a^H, maintaining its statistical properties.

1.3.2. *Gaussian processes: definition and main properties*

With the general concepts of stochastic processes in place, we can now precisely define Gaussian processes, which are characterized by the Gaussian nature of all their finite-dimensional distributions.

Consider a non-empty set $\mathbb{T}$, which serves as the index set for our process (e.g. time and spatial coordinates).

DEFINITION 1.8.– *A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called a Gaussian process if for every $n \in \mathbb{N}$ and any finite collection of indices $(t_1, t_2, \ldots, t_n) \in \mathbb{T}^n$, the corresponding random vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is a Gaussian vector.*

Equivalently, X is a Gaussian process if, for every $n \in \mathbb{N}$, any $(t_1, t_2, \ldots, t_n) \in \mathbb{T}^n$, and any coefficients $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, the linear combination $a_1 X_{t_1} + a_2 X_{t_2} + \cdots + a_n X_{t_n}$ is a Gaussian random variable. This definition directly leverages our prior understanding of Gaussian vectors and their linear transformations.

REMARK 1.3.– *A standout feature of Gaussian processes, stemming directly from the characteristics of Gaussian vectors (see, e.g., Lemma 1.4), is that their entire probability law is uniquely and completely determined by just two functions:*

– *its mean function* $m : t \in \mathbb{T} \longmapsto \mathbb{E}[X_t]$*;*
– *its covariance function* $C : (s,t) \in \mathbb{T}^2 \longmapsto \mathrm{Cov}(X_s, X_t)$.

This unique characterization of Gaussian processes by their mean and covariance functions, as highlighted in the preceding remark, profoundly simplifies their analysis and specification. This fundamental property not only facilitates their application in modeling complex phenomena but also implies that crucial stochastic attributes, such as self-similarity and independence, are fully determined by these two functions. The precise conditions governing these properties are formally presented in the following proposition.

PROPOSITION 1.4.– *Let* $X = (X_t)_{t\in\mathbb{T}}$ *and* $Y = (Y_t)_{t\in\mathbb{T}}$ *be two Gaussian processes, with respective mean functions* $\mu_X(t) = \mathbb{E}[X_t]$ *and* $\mu_Y(t) = \mathbb{E}[Y_t]$*, and covariance functions* $C_X(s,t) = Cov(X_s, X_t)$ *and* $C_Y(s,t) = Cov(Y_s, Y_t)$.

1) A Gaussian process X *is H-self-similar (as defined in Definition 1.7) if and only if its mean function satisfies* $\mu_X(at) = a^H \mu_X(t)$ *and its covariance function satisfies:*

$$C_X(as, at) = a^{2H} C_X(s,t)$$

for all $s,t \in \mathbb{T}$ *and* $a > 0$*. In particular, for a centered Gaussian process (i.e.* $\mu_X(t) = 0$ *for all* t*), it is H-self-similar if and only if its covariance function* $C_X(s,t) = \mathbb{E}[X_s X_t]$ *satisfies* $C_X(as, at) = a^{2H} C_X(s,t)$.

2) The Gaussian processes X *and* Y *are independent if and only if their cross-covariance function is identically zero. That is, for all* $s,t \in \mathbb{T}$*:*

$$Cov(X_s, Y_t) = \mathbb{E}[(X_s - \mu_X(s))(Y_t - \mu_Y(t))] = 0.$$

If X *and* Y *are centered Gaussian processes (i.e.* $\mu_X(t) = 0$ *and* $\mu_Y(t) = 0$ *for all* t*), this condition simplifies to* $\mathbb{E}[X_s Y_t] = 0$ *for all* $s,t \in \mathbb{T}$.

A particularly useful property for constructing and manipulating Gaussian processes is their closure under linear combinations.

PROPOSITION 1.5.– *If* $X = (X_t)_{t\in\mathbb{T}}$ *and* $Y = (Y_t)_{t\in\mathbb{T}}$ *are two independent Gaussian processes, then any linear combination* $Z = aX + bY = (aX_t + bY_t)_{t\in\mathbb{T}}$ *(for* $a, b \in \mathbb{R}$*) is also a Gaussian process.*

Proof. To show that Z is a Gaussian process, we need to demonstrate that any finite linear combination of its components is a Gaussian random variable. Consider an arbitrary linear combination $\sum_{i=1}^{n} c_i Z_{t_i}$ for $c_i \in \mathbb{R}$ and $t_i \in \mathbb{T}$. Substituting the definition of Z_{t_i}:

$$\sum_{i=1}^{n} c_i Z_{t_i} = \sum_{i=1}^{n} c_i (aX_{t_i} + bY_{t_i}) = a \sum_{i=1}^{n} c_i X_{t_i} + b \sum_{i=1}^{n} c_i Y_{t_i}.$$

Let $U = a\sum_{i=1}^{n} c_i X_{t_i}$ and $V = b\sum_{i=1}^{n} c_i Y_{t_i}$. Since X is a Gaussian process, the linear combination $\sum_{i=1}^{n} c_i X_{t_i}$ is a Gaussian random variable. Consequently, U, which is a scalar multiple of a Gaussian random variable, is also Gaussian. Similarly, since Y is a Gaussian process, the linear combination $\sum_{i=1}^{n} c_i Y_{t_i}$ is a Gaussian random variable. Therefore, V is also Gaussian. Furthermore, because X and Y are independent processes, the random variables U and V are independent Gaussian random variables. A fundamental property of Gaussian distributions is that the sum of two independent Gaussian random variables is itself a Gaussian random variable (see Lemma 1.1). Thus, $U + V = \sum_{i=1}^{n} c_i Z_{t_i}$ is a Gaussian random variable, which completes the proof. □

1.3.3. *Existence of gaussian processes with prescribed covariance function*

Having defined Gaussian processes as collections of random variables whose finite-dimensional distributions are Gaussian, a natural and critical question arises: *Under what conditions can such a process actually exist* for a given mean and covariance function? This question is answered by a fundamental result rooted in the broader theory of stochastic processes.

First, we need a specific property for functions that can serve as covariance functions.

DEFINITION 1.9.– *A real-valued function $f : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ is called positive semi-definite if for any $n \in \mathbb{N} \setminus \{0\}$, and for any choice of $t_1, t_2, \ldots, t_n \in \mathbb{T}$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$, the following inequality holds:*

$$\sum_{j,k=1}^{n} a_j a_k f(t_j, t_k) \geq 0.$$

The existence of a Gaussian process with a specified mean and covariance function is guaranteed by a specialized version of the celebrated *Kolmogorov existence theorem*. While the general Kolmogorov theorem ensures the existence of a stochastic process, given its finite-dimensional distributions, for Gaussian processes, this simplifies considerably. As we have established, the entire probabilistic structure of a Gaussian process is uniquely determined by its mean and covariance functions. This leads to the following powerful result:

THEOREM 1.2 (Kolmogorov existence theorem for Gaussian processes).– *Let $\mathbb{T}$ be any non-empty set. Let $m : \mathbb{T} \to \mathbb{R}$ be any real-valued function (serving as the mean function). Let $C : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ be a real-valued function (serving as the covariance function) such that: C is symmetric: $C(s,t) = C(t,s)$ for all $s,t \in \mathbb{T}$, and C is positive semi-definite. Then, there exists a Gaussian process $X = (X_t)_{t\in\mathbb{T}}$ with mean function $m(t)$ and covariance function $C(s,t)$.*

Proof. The proof of this fundamental theorem, which relies on constructing the appropriate probability space and random variables, can be found in standard texts on stochastic processes, such as Dudley (2002) or Mishura and Shevchenko (2017). □

To illustrate the concept of a positive semi-definite function and its direct application in defining a Gaussian process, let us consider a classic example:

EXAMPLE 1.1.– *Consider the case where the index set is $\mathbb{T} = \mathbb{R}^+ = [0,\infty)$, and let the candidate covariance function be $f(s,t) = s \wedge t := \min(s,t)$; $s,t \in \mathbb{R}^+$. We will demonstrate that this function f is indeed positive semi-definite.*

The key insight is to express $s \wedge t$ as an inner product of functions in an L_2 space. For any $s,t \in \mathbb{R}^+$ and a sufficiently large $T \geq s \vee t$ (where $s \vee t := \max(s,t)$), we can write:

$$s \wedge t = \int_0^T \mathbf{1}_{[0,s]}(u)\mathbf{1}_{[0,t]}(u)\,du = \left\langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \right\rangle_{L_2([0,T],\lambda_1)},$$

where λ_1 is the Lebesgue measure on $\mathbb{R}$, and $\langle\cdot,\cdot\rangle_{L_2([0,T],\lambda_1)}$ denotes the standard scalar product in the Hilbert space $L_2([0,T],\lambda_1)$ (the space of square-integrable functions on $[0,T]$). Here, $\mathbf{1}_{[0,x]}$ is the indicator function, which is 1 on $[0,x]$ and 0 otherwise.

Now, for any $n \in \mathbb{N}$, any real coefficients $a_1,\ldots,a_n \in \mathbb{R}$, and any ordered time points $0 \leq t_1 \leq \cdots \leq t_n$, choosing $T \geq t_n$, we can examine the sum:

$$\begin{aligned}
\sum_{j,k=1}^n a_j a_k f(t_j,t_k) &= \sum_{j,k=1}^n a_j a_k \left\langle \mathbf{1}_{[0,t_j]}, \mathbf{1}_{[0,t_k]} \right\rangle_{L_2([0,T],\lambda_1)} \\
&= \left\langle \sum_{j=1}^n a_j \mathbf{1}_{[0,t_j]}, \sum_{k=1}^n a_k \mathbf{1}_{[0,t_k]} \right\rangle_{L_2([0,T],\lambda_1)} \\
&= \left\| \sum_{j=1}^n a_j \mathbf{1}_{[0,t_j]} \right\|^2_{L_2([0,T],\lambda_1)} \\
&\geq 0.
\end{aligned}$$

The second equality follows from the bilinearity of the inner product, and the third from the definition of the squared norm in an L_2 space. Since the squared norm of

any function is always non-negative, the condition for positive semi-definiteness is satisfied.

This example provides the foundation for defining one of the most important and widely studied stochastic processes: the Wiener process.

DEFINITION 1.10.– *A stochastic process $(W_t)_{t\geq 0}$ is called a Wiener process (or often, Brownian motion) if it is a Gaussian process with:*

– a zero mean function: $m(t) := \mathbb{E}[W_t] = 0$ for all $t \geq 0$;

– a covariance function given by: $C(s,t) := s \wedge t = \min(s,t)$ for all $s, t \geq 0$.

REMARK 1.4.– *It follows directly from Example 1.1, where we proved the positive semi-definiteness of $f(s,t) = s \wedge t$, and Theorem 1.2, that Definition 1.10 correctly establishes the existence of such a Gaussian process, ensuring that the Wiener process is a well-defined mathematical object.*

The Wiener process has several crucial properties that make it invaluable in modeling random phenomena.

PROPOSITION 1.6 (properties of the Wiener process).– *A Wiener process $(W_t)_{t\geq 0}$ possesses the following fundamental properties:*

1) $W_0 = 0$ almost surely.

2) For any $0 \leq t_1 < t_2 < \cdots < t_n$, the increments $W_{t_2} - W_{t_1}$, $W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$ are mutually independent random variables.

3) For any $0 \leq s < t$, the increment $W_t - W_s$ is a Gaussian random variable with mean 0 and variance $t - s$, that is, $W_t - W_s \sim \mathcal{N}(0, t-s)$.

While the definitions above guarantee the existence of Gaussian processes based on their mean and covariance functions, they do not inherently tell us about the smoothness or continuity of their sample paths. For many applications, particularly in stochastic calculus, the regularity of these paths (trajectories) is paramount. This is where the powerful Kolmogorov continuity theorem comes into play.

THEOREM 1.3 (Kolmogorov continuity theorem).– *Let $X = \{X_t, t \in [0,T]\}$ be a stochastic process defined on a compact interval $[0,T]$.*

1) If there exist strictly positive constants K, α and β such that for all $0 \leq s < t \leq T$:

$$\mathbb{E}|X_t - X_s|^\alpha \leq K|t-s|^{1+\beta}, \qquad [1.4]$$

then X possesses a continuous modification. That is, there exists a process $Y = (Y_t)_{t\in[0,T]}$ such that Y is a modification of X (i.e. $P(X_t = Y_t) = 1$ for all t) and almost all of its sample paths $t \mapsto Y_t(\omega)$ are continuous functions.

2) Furthermore, if stochastic process $X = \{X_t, t \in [0,T]\}$ *satisfies condition* [1.4]*, then it possesses a modification whose trajectories are Hölder continuous of any order* $\gamma < \beta/\alpha$*. This implies that for almost all sample paths, there exists a constant* $C(\omega)$ *such that* $|Y_t(\omega) - Y_s(\omega)| \leq C(\omega)|t-s|^\gamma$ *for all* $s, t \in [0,T]$.

Building upon the general Kolmogorov continuity theorem, a more specialized and often more direct result exists for Gaussian processes, which are central to many stochastic models. This specialized theorem, often referred to as the Kolmogorov–Chentsov theorem for Gaussian processes, leverages their unique properties to establish sample path regularity based solely on the second moments of their increments.

THEOREM 1.4 (Kolmogorov–Chentsov theorem for Gaussian processes).– *Let* $X = \{X_t\}_{t\in[0,T]}$ *be a centered Gaussian process defined on a compact interval* $[0,T]$*. If there exist a strictly positive constant* C *and an exponent* $\nu > 0$ *such that for all* $0 \leq s < t \leq T$*:*

$$\mathbb{E}[(X_t - X_s)^2] \leq C|t-s|^\nu. \qquad [1.5]$$

Then, for any $\gamma \in (0, \nu/2)$*, the process* X *possesses a modification whose sample paths are almost surely* γ*-Hölder continuous. That is, for almost all sample paths, there exists a random constant* $C(\omega)$ *such that* $|Y_t(\omega) - Y_s(\omega)| \leq C(\omega)|t-s|^\gamma$ *for all* $s, t \in [0,T]$.

Proof. Let $X = \{X_t\}_{t\in[0,T]}$ be a centered Gaussian process satisfying the condition [1.5]. Since X is a Gaussian process, its increments $Z_{s,t} = X_t - X_s$ are centered Gaussian random variables. Therefore, by Lemma 1.2, for any $p \geq 1$, there exists a positive constant C_p such that:

$$\mathbb{E}[|Z|^p] = C_p(\mathbb{E}[Z^2])^{p/2}.$$

Applying this to the increment $Z_{s,t} = X_t - X_s$, we have:

$$\mathbb{E}[|X_t - X_s|^p] = C_p(\mathbb{E}[(X_t - X_s)^2])^{p/2}.$$

Now, substituting the given variance bound $\mathbb{E}[(X_t - X_s)^2] \leq C|t-s|^\nu$:

$$\mathbb{E}[|X_t - X_s|^p] \leq C_p(C|t-s|^\nu)^{p/2} = (C_p C^{p/2})|t-s|^{p\nu/2}.$$

This inequality holds for any $p \geq 1$.

Now, we compare this with the condition of Kolmogorov continuity Theorem 1.3:

$$\mathbb{E}|X_t - X_s|^\alpha \leq K|t-s|^{1+\beta}.$$

We choose $\alpha = p$. Then, we can identify:

– $K = C_p C^{p/2}$. Since $C > 0$ and $C_p > 0$ for any $p \geq 1$, we have $K > 0$.
– $1 + \beta = p\nu/2$. This implies $\beta = p\nu/2 - 1$.

For Kolmogorov's Theorem 1.3 to guarantee Hölder continuity (Part 2), it requires the constant β to be strictly positive. The condition $\beta > 0$ translates to $p\nu/2 - 1 > 0$, or $p\nu/2 > 1$, which means $p > 2/\nu$. Since $\nu > 0$, we can always choose an integer p sufficiently large to satisfy this condition (e.g. take $p = \lfloor 2/\nu \rfloor + 1$, or simply choose a large integer).

With $\alpha = p$ and $\beta = p\nu/2 - 1$, Kolmogorov's Theorem 1.3 states that X possesses a modification whose trajectories are Hölder continuous of any order $\gamma < \beta/\alpha$. Substituting our values for α and β:

$$\gamma < \frac{p\nu/2 - 1}{p} = \frac{\nu}{2} - \frac{1}{p}.$$

This holds for any p satisfying $p > 2/\nu$. To obtain the strongest possible Hölder exponent, we can let p be arbitrarily large. As $p \to \infty$, the term $1/p$ approaches 0. Therefore, for any $\gamma \in (0, \nu/2)$, we can always choose a sufficiently large integer p such that $\gamma < \frac{\nu}{2} - \frac{1}{p}$, thus satisfying the conditions of the theorem.

Hence, the Gaussian process X possesses a modification whose sample paths are almost surely γ-Hölder continuous for any $\gamma \in (0, \nu/2)$. □

REMARK 1.5.– *The Kolmogorov continuity theorem, as stated above, is commonly presented for processes indexed by a one-dimensional time interval. However, for applications involving stochastic processes on higher dimensional spatial domains, a generalized version is often required. The following theorem extends this fundamental result to processes defined on compact subsets of* $\mathbb{R}^d$.

THEOREM 1.5 (Kolmogorov continuity theorem for multi-dimensional index sets).– *Let* $X = \{X_x, x \in S\}$ *be a stochastic process defined on a compact subset* $S \subset \mathbb{R}^d$.

1) If there exist strictly positive constants K, $p \geq 1$, *and* $\epsilon > 0$ *such that for all* $x, y \in S$*:*

$$\mathbb{E}[|X_x - X_y|^p] \leq K|x - y|^{d+\epsilon}, \qquad [1.6]$$

then X *possesses a continuous modification. That is, there exists a process* $Y = \{Y_x, x \in S\}$ *such that* Y *is a modification of* X *(i.e.* $P(X_x = Y_x) = 1$ *for all* $x \in S$*) and almost all of its sample paths* $x \mapsto Y_x(\omega)$ *are continuous functions on* S.

2) Furthermore, if the stochastic process $X = \{X_x, x \in S\}$ *satisfies the condition* [1.6]*, then it possesses a modification whose trajectories are Hölder continuous of any order* γ *such that* $0 < \gamma < \epsilon/p$. *This implies that for almost all sample paths, there exists a constant* $C(\omega)$ *such that* $|Y_x(\omega) - Y_y(\omega)| \leq C(\omega)|x - y|^\gamma$ *for all* $x, y \in S$.

1.3.4. *Nondegenerate Gaussian processes*

Among the diverse family of Gaussian processes, nondegenerate Gaussian processes form a crucial subclass. These processes are not only fully characterized by their mean and covariance functions but also possess an additional, desirable property: their finite-dimensional distributions are well-behaved, preventing perfect linear dependencies between random variables at distinct time points.

This "nicely behaved" property is formally captured by the invertibility of their covariance matrices.

DEFINITION 1.11.– *A Gaussian process $\{X_t\}_{t\in\mathbb{T}}$ is defined as nondegenerate if, for every finite set of distinct time points $t_1, t_2, \ldots, t_n \in \mathbb{T}$ (for any $n \geq 1$), the corresponding covariance matrix $\Sigma = (Cov(X_{t_i}, X_{t_j}))_{1\leq i,j\leq n}$ is invertible.*

This definition implies that for nondegenerate Gaussian processes, all of their finite-dimensional distributions are *non-singular*, meaning their PDFs exist and are well defined. In practical terms, this means there are no perfect linear dependencies among the random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ for any distinct set of time points.

Such dependencies would indicate that one variable could be perfectly predicted from a linear combination of others, which is often an undesirable characteristic in modeling.

The following lemma establishes an important equivalence, showing that the concept of a nondegenerate Gaussian process is intrinsically linked to the positive definiteness of its covariance matrices.

LEMMA 1.11.– *A Gaussian process $\{X_t\}_{t\in\mathbb{T}}$ is nondegenerate if, and only if, for every finite set of distinct time points $t_1, t_2, \ldots, t_n \in \mathbb{T}$ (for any $n \geq 1$), the corresponding covariance matrix $\Sigma = (Cov(X_{t_i}, X_{t_j}))_{1\leq i,j\leq n}$ is strictly positive definite. That is, for any non-zero vector $\mathbf{a} \in \mathbb{R}^n$, the quadratic form $\mathbf{a}'\Sigma\mathbf{a} > 0$.*

Proof. Let $\Sigma = (\mathrm{Cov}(X_{t_i}, X_{t_j}))_{1\leq i,j\leq n}$ be the $n \times n$ covariance matrix.

Assume that the process X is non-degenrate; that is the covariance matrix Σ is invertible. By definition of matrix invertibility, this means that Σ has a trivial null space; in other words, if $\Sigma\mathbf{a} = \mathbf{0}$ for some vector $\mathbf{a} \in \mathbb{R}^n$, then it must necessarily be that $\mathbf{a}$ is the zero vector ($\mathbf{a} = \mathbf{0}$).

We also know that by its very definition, Σ is symmetric, that is $\Sigma_{ij} = \Sigma_{ji}$ for all i, j, and positive semi-definite, that is for any vector $\mathbf{a} \in \mathbb{R}^n$, the quadratic form $\mathbf{a}'\Sigma\mathbf{a} \geq 0$. This holds because

$$\mathbf{a}'\Sigma\mathbf{a} = \text{Var}\left(\sum_{i=1}^{n} a_i X_{t_i}\right),$$

and variance is inherently nonnegative.

Now, let us proceed by contradiction. Suppose that Σ is *not* strictly positive definite. This would mean that there exists at least one non-zero vector $\mathbf{a}_0 \in \mathbb{R}^n$ (i.e. $\mathbf{a}_0 \neq \mathbf{0}$) such that the quadratic form is zero:

$$\mathbf{a}_0'\Sigma\mathbf{a}_0 = 0.$$

Since Σ is a positive semi-definite matrix, a fundamental property states that if $\mathbf{x}'M\mathbf{x} = 0$ for a positive semi-definite matrix M, then it must follow that $M\mathbf{x} = \mathbf{0}$ (see Lemma A.1). Applying this property to our covariance matrix Σ (which is positive semi-definite) and the vector $\mathbf{a}_0$, we get that $\Sigma\mathbf{a}_0 = \mathbf{0}$. However, this result, $\Sigma\mathbf{a}_0 = \mathbf{0}$ with $\mathbf{a}_0 \neq \mathbf{0}$, means that $\mathbf{a}_0$ is a non-zero vector in the null space of Σ. This is precisely the definition of a singular (non-invertible) matrix. This directly contradicts our initial assumption that Σ is invertible. Therefore, our supposition that Σ is not strictly positive definite must be false. Consequently, if the Gaussian process is nondegenerate, its covariance matrix must be strictly positive definite.

Now let us assume that Σ is strictly positive definite. This means that for any non-zero vector $\mathbf{a} \neq \mathbf{0}$, $\mathbf{a}'\Sigma\mathbf{a} > 0$. Suppose, for contradiction, that Σ is *not* invertible. This implies that Σ is singular, meaning its null space is non-trivial. Thus, there exists a non-zero vector $\mathbf{a}_0 \neq \mathbf{0}$ such that $\Sigma\mathbf{a}_0 = \mathbf{0}$. Multiplying by $\mathbf{a}_0'$ from the left yields:

$$\mathbf{a}_0'\Sigma\mathbf{a}_0 = \mathbf{a}_0'\mathbf{0} = 0.$$

However, this directly contradicts our assumption that Σ is strictly positive definite (which requires $\mathbf{a}_0^T\Sigma\mathbf{a}_0 > 0$ for any non-zero $\mathbf{a}_0$). Therefore, if Σ is strictly positive definite, it must be invertible. This implies the Gaussian process is non-degenerate. □

EXAMPLE 1.2.– *The Brownian motion on* $(0, +\infty)$ *is a nondegenerate Gaussian process, because for any finite set of distinct time points* $t_1, t_2, \ldots, t_n \in (0, \infty)$*, the covariance matrix* $\Sigma = (Cov(B_{t_i}, B_{t_j}))_{1\leq i,j\leq n} = (\min(t_i, t_j))_{1\leq i,j\leq n}$ *is invertible.*

Indeed, let us assume, without loss of generality, that $0 < t_1 < t_2 < \cdots < t_n$. This gives the matrix:

$$\Sigma = \begin{pmatrix} t_1 & t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & t_2 & \ldots & t_2 \\ t_1 & t_2 & t_3 & \ldots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & \ldots & t_n \end{pmatrix}$$

To show that Σ is invertible, we will demonstrate that its determinant is non-zero. Upon applying the sequence of elementary row operations (specifically, $R_i \leftarrow R_i - R_{i-1}$ for $i = n, n-1, \ldots, 2$) to the covariance matrix Σ, we come back to the following upper triangular matrix, Σ_m:

$$\Sigma_m = \begin{pmatrix} t_1 & t_1 & t_1 & t_1 & \ldots & t_1 \\ 0 & t_2 - t_1 & t_2 - t_1 & t_2 - t_1 & \ldots & t_2 - t_1 \\ 0 & 0 & t_3 - t_2 & t_3 - t_2 & \ldots & t_3 - t_2 \\ 0 & 0 & 0 & t_4 - t_3 & \ldots & t_4 - t_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & t_n - t_{n-1} \end{pmatrix}.$$

Since these row operations do not change the determinant of the matrix:

$$\det(\Sigma) = \det(\Sigma_m) = t_1 \cdot (t_2 - t_1) \cdot (t_3 - t_2) \cdot \cdots \cdot (t_n - t_{n-1}) = t_1 \prod_{k=2}^{n} (t_k - t_{k-1}).$$

Given that the time points are distinct and ordered ($0 < t_1 < t_2 < \cdots < t_n$), it implies that $t_1 > 0$ and all differences $(t_k - t_{k-1})$ are strictly positive, ensuring that $\det(\Sigma) > 0$. This confirms the invertibility of the covariance matrix and, consequently, that the Brownian motion on $(0, +\infty)$ is a nondegenerate Gaussian process.

1.4. Gaussian random measures and Wiener integration

1.4.1. *Gaussian random measures*

In this section, we lay the foundational concepts of Gaussian random measures, which are essential for defining various types of noise in stochastic analysis. We then introduce specific instances, including the crucial Brownian measure and its higher dimensional generalization, space–time white noise.

DEFINITION 1.12.– *A Gaussian random measure $\mathcal{M}$ on a measurable space $(S, \mathcal{S})$ is a mapping that assigns a real-valued random variable $\mathcal{M}(A)$ to every set $A \in \mathcal{S}$ such that the following two conditions hold:*

1) For any finite collection of pairwise disjoint sets $A_1, A_2, \ldots, A_n \in \mathcal{S}$, the measure is additive:

$$\mathcal{M}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathcal{M}(A_i) \quad \textit{almost surely (a.s.).}$$

2) For any choice of $n \geq 1$ and any collection of sets $A_1, A_2, \ldots, A_n \in \mathcal{S}$, the random vector

$$(\mathcal{M}(A_1), \mathcal{M}(A_2), \ldots, \mathcal{M}(A_n))$$

is a jointly Gaussian vector.

REMARK 1.6.– *The entire distribution of a Gaussian random measure is completely determined by its mean and covariance structure.*

– We will primarily focus on centered Gaussian random measures, for which the mean is zero: $\mathbb{E}[\mathcal{M}(A)] = 0$ *for all* $A \in \mathcal{S}$.

– For any two sets $A, B \in \mathcal{S}$*, the covariance function* $\Gamma(A, B)$ *completely specifies the joint distribution:*

$$\mathbb{E}[\mathcal{M}(A)\mathcal{M}(B)] = \Gamma(A, B).$$

The function $\Gamma(A, B)$ *must satisfy properties of a covariance function (symmetry and positive semi-definiteness).*

1.4.2. *Brownian measure on* $\mathbb{R}$

The concept of a Gaussian random measure provides a rigorous foundation for understanding various types of "white noise". A primary example is the Brownian measure:

DEFINITION 1.13.– *A Brownian measure* W *on* $\mathbb{R}$ *is a specific type of Gaussian random measure defined on* $(\mathbb{R}, \mathcal{B}_b(\mathbb{R}))$ *(where* $\mathcal{B}_b(\mathbb{R})$ *is the set of bounded Borel sets on* $\mathbb{R}$*) whose covariance function* $\Gamma(A, B)$ *takes the form:*

$$\mathbb{E}[W(A)W(B)] = \lambda(A \cap B),$$

where λ *denotes the Lebesgue measure on* $\mathbb{R}$.

This particular covariance structure implies two fundamental properties for a Brownian measure:

1) The variance of $W(A)$ is simply the Lebesgue measure of the set A:

$$\text{Var}(W(A)) = \mathbb{E}[W(A)^2] = \lambda(A).$$

2) If two sets A and B are disjoint (i.e. $A \cap B = \emptyset$), then $\lambda(A \cap B) = 0$, implying $\mathbb{E}[W(A)W(B)] = 0$. Since $W(A)$ and $W(B)$ are Gaussian, zero covariance implies independence. Thus, the increments of a Brownian measure over disjoint sets are independent.

In integral notation, the expression "$dW(s)$" is a shorthand, referring to integration with respect to this Brownian measure. When we write an integral like $\int f(s)dW(s)$,

it represents a stochastic integral of the function f with respect to the Brownian measure W.

1.4.3. *Space–time white noise*

Crucial for stochastic partial differential equations, including the one rigorously studied in this book, is the concept of *space–time white noise*, which generalizes the 1D Brownian measure to a product space.

DEFINITION 1.14.– *A space–time white noise, often denoted by $W(ds, dy)$, is a specific type of Gaussian random measure defined on the product space $(\mathbb{R} \times \mathbb{R}^d, \mathcal{B}_b(\mathbb{R} \times \mathbb{R}^d))$. Its covariance function for any two bounded Borel sets $A, B \in \mathcal{B}_b(\mathbb{R} \times \mathbb{R}^d)$ is given by the Lebesgue measure of their intersection:*

$$\mathbb{E}(W(A)W(B)) = \lambda_{1+d}(A \cap B),$$

where λ_{1+d} is the $(1+d)$-dimensional Lebesgue measure on $\mathbb{R} \times \mathbb{R}^d$.

For product sets $A = I_A \times B_A$ and $B = I_B \times B_B$, where $I_A, I_B \in \mathcal{B}_b(\mathbb{R})$ (often representing time intervals) and $B_A, B_B \in \mathcal{B}_b(\mathbb{R}^d)$ (representing spatial regions), the covariance can equivalently be written as:

$$\mathbb{E}(W(I_A \times B_A)W(I_B \times B_B)) = \lambda_1(I_A \cap I_B)\lambda_d(B_A \cap B_B),$$

where λ_1 and λ_d are the Lebesgue measures on $\mathbb{R}$ and $\mathbb{R}^d$, respectively. This detailed form highlights its "whiteness" in both temporal and spatial dimensions.

Similar to the one-dimensional Brownian measure, this implies that the increments of a space–time white noise over disjoint space–time regions are independent. This characteristic makes space–time white noise a standard model for driving noise in stochastic partial differential equations.

1.4.4. *Gaussian random fields*

Building upon the concept of Gaussian processes, where the random variables are indexed by a one-dimensional parameter (typically, time $t \in \mathbb{R}$ or $t \in [0, \infty)$), we now extend this notion to higher dimensions. A Gaussian random field is a collection of random variables indexed by a multi-dimensional parameter space.

DEFINITION 1.15.– *A Gaussian random field $X = \{X_{\mathbf{x}}, \mathbf{x} \in \mathcal{D}\}$ is a collection of random variables indexed by a set $\mathcal{D} \subset \mathbb{R}^d$ for some integer $d \geq 1$ (or a more general topological space), such that every finite collection of these random variables, $\{X_{\mathbf{x}_1}, X_{\mathbf{x}_2}, \ldots, X_{\mathbf{x}_n}\}$, has a multivariate normal (Gaussian) distribution.*

Similar to Gaussian processes, a Gaussian random field is completely characterized by its *mean function* $m(\mathbf{x}) = \mathbb{E}[X_{\mathbf{x}}]$ and its *covariance function*

$$C(\mathbf{x}, \mathbf{y}) = \mathbb{E}[(X_{\mathbf{x}} - m(\mathbf{x}))(X_{\mathbf{y}} - m(\mathbf{y}))]$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{D}$. If $m(\mathbf{x}) = 0$ for all $\mathbf{x}$, the field is said to be *centered*.

Examples:

– Consider the temperature at various points on a geographical map. Here, $\mathbf{x}$ could represent the (latitude, longitude) coordinates of a location, and $X_{\mathbf{x}}$ would be the temperature at that specific location. The entire collection of temperatures across the map forms a Gaussian random field if its statistical properties are Gaussian.

– Another example could be the intensity of pixels in a grayscale image, where $\mathbf{x}$ represents the (x, y) coordinates of a pixel, and $X_{\mathbf{x}}$ is the pixel's intensity value.

1.4.5. *Wiener integration: definitions and isometries*

This section provides the formal definitions of the Wiener integral with respect to both 1D Wiener processes and space–time white noise, along with their crucial Itô isometry properties. These concepts are fundamental for building stochastic models and performing calculations in infinite-dimensional stochastic calculus.

1.4.5.1. *Wiener integral with respect to a 1D Wiener process* $W(y)$

Here, we recall the definition of the Itô (or Wiener) integral with respect to a standard one-dimensional Wiener process, also called Brownian motion (see Definition 1.10).

Let $W = \{W(y); y \in \mathbb{R}\}$ be a standard 1D Wiener process. The Wiener integral of a deterministic function $g : \mathbb{R} \to \mathbb{R}$ with respect to $W(y)$ is defined in stages:

1) For a simple function $g(y) = \sum_{j=1}^{n} c_j \mathbf{1}_{[a_j, b_j)}(y)$, where $c_j \in \mathbb{R}$ and $[a_j, b_j)$ are disjoint intervals, the integral is defined as:

$$\int_{\mathbb{R}} g(y) dW(y) := \sum_{j=1}^{n} c_j (W(b_j) - W(a_j)).$$

This definition ensures linearity and consistency with the properties of Brownian motion increments.

2) For any function $g \in L^2(\mathbb{R})$ (i.e. $\int_{\mathbb{R}} |g(y)|^2 dy < \infty$), there exists a sequence of simple functions g_n such that $g_n \to g$ in $L^2(\mathbb{R})$. The Wiener integral of g is then defined as the $L^2(\Omega)$ limit of the integrals of the simple functions:

$$\int_{\mathbb{R}} g(y) dW(y) := \lim_{n \to \infty} \int_{\mathbb{R}} g_n(y) dW(y).$$

This limit exists and is independent of the choice of approximating sequence g_n. The resulting integral is a centered Gaussian random variable.

1.4.5.2. *Itô isometry for 1D Wiener process*

THEOREM 1.6.– *The Wiener integral is an L^2-isometry. For any two deterministic functions $g_1, g_2 \in L^2(\mathbb{R})$, the following holds:*

$$\mathbb{E}\left[\left(\int_{\mathbb{R}} g_1(y)dW(y)\right)\left(\int_{\mathbb{R}} g_2(y)dW(y)\right)\right] = \int_{\mathbb{R}} g_1(y)g_2(y)dy.$$

In particular, for $g_1 = g_2$, this yields the L^2-norm isometry:

$$\mathbb{E}\left[\left(\int_{\mathbb{R}} g(y)dW(y)\right)^2\right] = \int_{\mathbb{R}} g(y)^2 dy.$$

This isometry is a cornerstone of stochastic calculus, allowing the computation of variances and covariances of Wiener integrals by simple integration in the L^2 space of the integrands.

1.4.5.3. *Wiener integral with respect to space–time white noise*

Consider a standard space–time white noise $W = \{W(I \times A); I \subset \mathbb{R}, A \subset \mathbb{R}^d,$ measurable$\}$.

When considering integrals of the form $\int f(s,y)W(ds,dy)$, or more simply $\int f(s,y)dW(s,y)$, the white noise $W(s,y)$ is understood as the underlying random measure for the measure space $(\mathbb{R} \times \mathbb{R}^d, \mathcal{B}(\mathbb{R} \times \mathbb{R}^d), dsdy)$.

Let $S = \mathbb{R} \times \mathbb{R}^d$ and μ be the Lebesgue measure $dsdy$ on S. The Wiener integral of a deterministic function $F : S \to \mathbb{R}$ with respect to the standard space–time white noise W is constructed in stages:

1) For a simple function $F(s,y) = \sum_{j=1}^{n} c_j \mathbf{1}_{I_j \times A_j}(s,y)$, where $I_j \times A_j$ are disjoint measurable rectangles, the integral is defined as:

$$\int_S F(s,y)dW(s,y) := \sum_{j=1}^{n} c_j W(I_j \times A_j).$$

2) For any function $F \in L^2(S,\mu)$ (i.e. $\int_S |F(s,y)|^2 dsdy < \infty$), there exists a sequence of simple functions F_n such that $F_n \to F$ in $L^2(S,\mu)$. The Wiener integral of F is then defined as the $L^2(\Omega)$ limit of the integrals of the simple functions:

$$\int_S F(s,y)dW(s,y) := \lim_{n\to\infty} \int_S F_n(s,y)dW(s,y).$$

This limit exists and is independent of the choice of approximating sequence F_n. The resulting integral is a centered Gaussian random variable.

1.4.5.4. *Itô isometry for space–time white noise*

THEOREM 1.7.– *The Wiener integral is an L^2-isometry. For any two deterministic functions $F, G \in L^2(\mathbb{R} \times \mathbb{R}^d, dsdy)$, the following holds:*

$$\mathbb{E}\left[\left(\int_{\mathbb{R}}\int_{\mathbb{R}^d} F(s,y)dW(s,y)\right)\left(\int_{\mathbb{R}}\int_{\mathbb{R}^d} G(s,y)dW(s,y)\right)\right]$$
$$= \int_{\mathbb{R}}\int_{\mathbb{R}^d} F(s,y)G(s,y)dsdy.$$

In particular, for $F = G$, this yields the L^2-norm isometry:

$$\mathbb{E}\left[\left(\int_{\mathbb{R}}\int_{\mathbb{R}^d} F(s,y)dW(s,y)\right)^2\right] = \int_{\mathbb{R}}\int_{\mathbb{R}^d} F(s,y)^2 dsdy.$$

This isometry is fundamental for computing covariances and variances of Wiener integrals.

1.5. Exercises

EXERCISE 1.1.– *Show that if X is a Gaussian vector, then $aX + b$ is also a Gaussian vector for any scalar $a \in \mathbb{R}$ and vector $b \in \mathbb{R}^n$.*

EXERCISE 1.2.– *Provide a rigorous proof for Lemma 1.2, which details the central moments of a Gaussian variable $X \sim \mathcal{N}(m, \sigma^2)$. Ensure your proof covers both the odd and even cases for n, and clearly demonstrates how the double factorial arises.*

Hint: consider first the case of a standard Gaussian variable $Z \sim \mathcal{N}(0,1)$. The characteristic function or integration by parts can be helpful.

EXERCISE 1.3.– *For a Gaussian random variable $X \sim \mathcal{N}(m, \sigma^2)$, derive its first few raw moments:*

- $\mathbb{E}[X] = m$
- $\mathbb{E}[X^2] = m^2 + \sigma^2$
- $\mathbb{E}[X^3] = m^3 + 3m\sigma^2$
- $\mathbb{E}[X^4] = m^4 + 6m^2\sigma^2 + 3\sigma^4$

Hint: these can be derived using the characteristic function or by expanding $(m + \sigma Z)^n$ where $Z \sim \mathcal{N}(0,1)$ and utilizing the known moments of a standard Gaussian variable.

EXERCISE 1.4.– *Consider a two-dimensional random vector $X = (X_1, X_2)'$. Let $X_1 \sim \mathcal{N}(0,1)$ be a standard Gaussian variable, and let $X_2 = X_1$.*

1) Show that X is indeed a Gaussian vector.

2) Determine the expectation vector $M_X = \mathbb{E}[X]$.

3) Compute the covariance matrix $K_X = \mathrm{Cov}(X)$.

4) Show that K_X is singular by calculating its determinant.

EXERCISE 1.5.– *Let $X \sim \mathcal{N}_n(\mathbf{0}, K_X)$ be an n-dimensional zero-mean multivariate Gaussian vector with covariance matrix K_X, and let $\mathbf{A}$ be a constant $n \times n$ matrix.*

1) Show that $\mathbb{E}[(X'\mathbf{A}X)^2] = \sum_{i,j,k,l=1}^{n} A_{ij}A_{kl}\mathbb{E}[X_iX_jX_kX_l]$.

2) Applying Isserlis' theorem and manipulating the summations, show that $\mathbb{E}[(X'\mathbf{A}X)^2]$ can be expressed as:

$$\mathbb{E}[(X'\mathbf{A}X)^2] = (Tr(\mathbf{A}K_X))^2 + 2Tr((\mathbf{A}K_X)^2).$$

3) Using the expression obtained in the previous part and Lemma 1.9, prove that the variance of the quadratic form $X'\mathbf{A}X$ is given by:

$$\mathrm{Var}(X'\mathbf{A}X) = 2\,\mathrm{Tr}((\mathbf{A}K_X)^2).$$

Hint: remember that K_X is a symmetric matrix ($K_{uv} = K_{vu}$) and that trace properties include $Tr(\mathbf{MN}) = Tr(\mathbf{NM})$.

EXERCISE 1.6.– *This exercise guides you through the proof of Corollary 1.2, building on fundamental matrix identities and properties of Gaussian vectors.*

1) Prove the following matrix trace identity: for any $n \times 1$ column vector $\mathbf{b}$ and any $n \times n$ matrix $\mathbf{\Sigma}$, show that:

$$Tr((\mathbf{b}\mathbf{b}')\mathbf{\Sigma}) = \mathbf{b}'\mathbf{\Sigma}\mathbf{b}.$$

Hint: you may find the cyclic property of the trace, $Tr(\mathbf{CD}) = Tr(\mathbf{DC})$, particularly useful. Also, remember that the trace of a 1×1 matrix is simply its single element.

2) Let $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma})$ be an n-dimensional zero-mean Gaussian vector with covariance matrix $\mathbf{\Sigma}$. For any symmetric $n \times n$ matrix $\mathbf{A}$ and any constant $n \times 1$ vector $\mathbf{b}$, show that the expected value of the product of the quadratic form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ and the squared linear form $(\mathbf{b}'\mathbf{Y})^2$ is given by:

$$\mathbb{E}[(\mathbf{Y}'\mathbf{A}\mathbf{Y})(\mathbf{b}'\mathbf{Y})^2] = 2Tr(\mathbf{A}\mathbf{\Sigma}\mathbf{b}\mathbf{b}'\mathbf{\Sigma}) + Tr(\mathbf{A}\mathbf{\Sigma})\mathbf{b}'\mathbf{\Sigma}\mathbf{b}.$$

Hint: begin by rewriting the squared linear form $(\mathbf{b}'\mathbf{Y})^2$ as a quadratic form involving a symmetric matrix $\mathbf{B}$. Then, apply Lemma 1.7 with appropriate matrices and utilize the trace identity proven in part (a) to simplify the resulting expression.

EXERCISE 1.7.– *Consider the function $C : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ defined by $C(s,t) = s \cdot t$ for $s, t \in \mathbb{R}^+ = [0, \infty)$. Can a zero-mean Gaussian process exist with $C(s,t) = s \cdot t$ as its covariance function? Justify your answer.*

EXERCISE 1.8.– *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let U be a random variable uniformly distributed on the interval $[0,1]$. Let $\mathbb{T} = [0,1]$ be our index set. Define two stochastic processes, $X = (X_t)_{t\in\mathbb{T}}$ and $Y = (Y_t)_{t\in\mathbb{T}}$, as follows:*

– *$X_t(\omega) = 0$ for all $t \in \mathbb{T}$ and all $\omega \in \Omega$. (This is the zero process)*

– *$Y_t(\omega) = \begin{cases} 1 & \text{if } t = U(\omega) \\ 0 & \text{if } t \neq U(\omega) \end{cases}$ for all $t \in \mathbb{T}$ and all $\omega \in \Omega$.*

Show that X and Y are modifications of each other, but they are not indistinguishable.

EXERCISE 1.9.– *Consider two independent standard Brownian motions, $X = (B_t^{(1)})_{t\geq 0}$ and $Y = (B_t^{(2)})_{t\geq 0}$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Show that X and Y are versions of each other, but they are not modifications.*

EXERCISE 1.10.– *Two-sided Brownian motion.*

Let $(W_t)_{t\geq 0}$ be a standard Brownian motion on $[0, \infty)$ with $W_0 = 0$. We define a process $B = (B_t)_{t\in\mathbb{R}}$ on the entire real line as follows:

$$B_t = \begin{cases} W_t & \text{for } t \geq 0 \\ \tilde{W}_{-t} & \text{for } t < 0 \end{cases}$$

where $\tilde{W} = (\tilde{W}_t)_{t\geq 0}$ is an independent copy of the standard Brownian motion W.

1) Demonstrate that the process $B = (B_t)_{t\in\mathbb{R}}$ is a centered Gaussian process.

2) Show that the covariance function of B is given by:

$$\mathbb{E}[B_s B_t] = \frac{1}{2}\left(|t| + |s| - |t - s|\right)$$

for all $s, t \in \mathbb{R}$. Explicitly verify this formula for all necessary cases: ($s, t \geq 0$; $s, t < 0$; and $s \geq 0, t < 0$).

3) Verify that the increments of B are Gaussian: for any $s < t$, $B_t - B_s$ is a Gaussian random variable with mean 0 and variance $|t - s|$.

4) Discuss the properties of stationary and independent increments for this two-sided Brownian motion, explaining why these hold based on its construction and the properties of standard Brownian motion.