Gaussian Process and Brownian Motion (가우시안 프로세스와 브라운 운동)

김지수 (Jisu KIM)

확률론 2 (Probability Theory 2), 2025 2nd semester (fall)

Stochastic Processes

A stochastic process is a random object taking values in a Euclidean space, and defined over a parameter space of dimensionality at least one. We shall usually denote parameter spaces by T, and elements of T will be denoted by t, in a notation harking back to the early days of stochastic processes when the parameter was always one-dimensional 'time'.

Definition ([1, Definition 1.1.1])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and T a topological space. Then a measurable mapping

$$X : \Omega \to \mathbb{R}^T$$

(the space of all real-valued functions on T) is called a real-valued stochastic process. Measurable mappings from Ω to $(\mathbb{R}^T)^d$, $d > 1$, are called vector-valued stochastic processes. If $T \subset \mathbb{R}^N$, we call X an (N, d) stochastic process, and if $d = 1$, simply an N-dimensional stochastic process.

Thus, $X(\omega)$ is a function, and $(X(\omega))(t)$ its value at time t. In general, however, we shall not distinguish among

$$X_t \equiv X(t) \equiv X(t, \omega) \equiv (X(\omega))(t),$$

etc., unless there is some special need to do so.

Before going further, we shall make one technical blanket assumption on all stochastic processes that will appear in this book, although we shall rarely mention it again. Throughout, we shall demand that all stochastic processes be separable.[1] If you are familiar with this assumption, you will know that it solves many measurability problems. For example, without separability, it is not necessarily the case that the supremum of a random field is a well-defined random variable. If you are not familiar with the concept, it probably would not have occurred to you that such problems might exist. Either way, you need not worry about it anymore.

Before leaving this section, there is one rather important, theoretical point that needs to be made. While [1, Definition 1.1.1] may effectively define stochastic processes/random fields, it does not really tell us how to differentiate between them. A celebrated theorem of Kolmogorov, known as his Consistency or Existence Theorem, says that the distributional properties of a (N, d) stochastic process over T are determined by its finite-dimensional distributions, under certain regularity conditions. These are the distributions

$$P\left(X(t_1) \in B_1, \cdots, X(t_n) \in B_n\right), \tag{1}$$

[1] This property, due originally to Doob, implies conditions on both T and X. In particular, an $\mathbb{R}^d$-valued stochastic process X on a topological space T is called separable if there exists a countable dense subset $D \subset T$ and a fixed event N with $\mathbb{P}(N) = 0$ such that, for any closed $B \subset \mathbb{R}^d$ and open $I \subset T$,

$$\{\omega : X(t, \omega) \in B \ \forall t \in I\} \triangle \{\omega : X(t, \omega) \in B \ \forall t \in I \cap D\} \subset N,$$

where $\triangle$ denotes the symmetric difference operator, so that $A \triangle B = (A \backslash B) \cup (B \backslash A)$.

for all $n \geq 1$ and all collections $\{t_j\}_{1\leq j\leq n}$ and Borel $\{B_j\}_{1\leq j\leq n}$ with $t_j \in T$ and $B_j \in \mathcal{R}^d$. If X is a process which possesses joint probability densities, then the probabilities in (1) can be expressed, in a self-explanatory notation, as

$$\int_{B_1} \cdots \int_{B_n} p_{t_1,\cdots,t_n}(x_1, \cdots, x_n) dx_1 \cdots dx_n,$$

and so Kolmogorov's theorem reduces to demanding that we know these densities.

We end this section by introducing the concept of stationary.

Definition. Suppose that X is an (N, d) random field defined over all of $\mathbb{R}^N$. Suppose furthermore that the mean function $m(t)$ is constant, and that the covariance function $\Sigma(s, t)$ is a function of the difference $t - s$ only. Then we say that X is homogeneous or stationary.

Two comments are called for following on from this definition. The first is that, in fact, we have only really defined what is usually called weak or second order stationarity , which depends only on first and second moments. A stronger form of stationarity is that of the finite dimensional distributions, that requires that joint distributions of the form

$$\{X(t_1 + \tau), \cdots, X(t_n + \tau)\}$$

be independent of τ, for all $n \geq 1$ and all $t_j \in \mathbb{R}^N$. It is obvious that the two definitions coincide when X is Gaussian.

The second is more an issue of notation than a comment. When X is stationary, we shall generally abuse notation and write

$$\Sigma(s, t) = \Sigma(s - t).$$

Gaussian Variables

A real-valued random variable X is said to be Gaussian (or normally distributed) if it has the density function

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R},$$

for some $m \in \mathbb{R}$ and $\sigma > 0$. It is elementary calculus that the mean of X is m and the variance σ^2, and that the characteristic function is given by

$$\phi(\theta) = \mathbb{E}[e^{i\theta X}] = e^{i\theta m - \frac{1}{2}\sigma^2\theta^2}.$$

We abbreviate this by writing $X \sim \mathcal{N}(m, \sigma^2)$. The case $m = 0$, $\sigma^2 = 1$ is rather special, and in this situation we say that X has a standard normal distribution. In general, if a random variable or process has zero mean, we call it centered.

Since the indefinite integral of φ is not a simple function, we also need notation Φ for the distribution function and Ψ for the tail probability function of a standard normal variable:

$$\Psi(x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du.$$

While Φ and Ψ may not be explicit, there are simple, and rather important, bounds that hold for every $x > 0$ and become sharp very quickly as x grows. In particular, in terms of φ we have[2]

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) < \Psi(x) < \frac{1}{x}\varphi(x).$$

[2]The upper bound follows from

$$\int_x^\infty e^{-u^2/2} du \leq \int_x^\infty \frac{u}{x} e^{-u^2/2} du = \frac{1}{x} e^{-x^2/2},$$

A $\mathbb{R}^d$-valued random variable X is said to be multivariate Gaussian if, for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, the real-valued variable

$$\langle \alpha, X \rangle = \sum_{i=1}^{d} \alpha_i X_i$$

is Gaussian. In this case, there exists a mean vector $m \in \mathbb{R}^d$ with $m_j = \mathbb{E}[X_j]$ and a nonnegative definite $d \times d$ covariance matrix Σ with elements $\Sigma_{ij} = \mathbb{E}[(X_i - m_i)(X_j - m_j)]$, such that the probability density of X is given by

$$\varphi(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-m)^\top \Sigma^{-1}(x-m)\right). \tag{2}$$

where $|\Sigma| = \det \Sigma$ is the determinant of Σ. Consistently with the one dimensional case, we write this as $X \sim \mathcal{N}(m, \Sigma)$ or $X \sim \mathcal{N}_d(m, \Sigma)$ to emphasize the dimension.

In view of (2), Gaussian distributions are completely determined by their first- and second-order moments and that uncorrelated Gaussian variables are independent. Both of these facts will be of crucial importance later on.

While the definitions are fresh, note that it is relatively straightforward to check from (2) that the characteristic function of a multivariate Gaussian X is given by:

$$\phi(\theta) = \mathbb{E}[e^{i\langle \theta, X\rangle}] = \exp\left(i\langle \theta, m\rangle - \frac{1}{2}\theta^\top \Sigma \theta\right), \quad \theta \in \mathbb{R}^d. \tag{3}$$

One consequence of the simple structure of ϕ is the fact that if $\{X_n\}_{n\geq 1}$ is an L^2-convergent sequence of Gaussian vectors, then the limit X must also be Gaussian. Furthermore, if $X_n \sim \mathcal{N}(m_n, C_n)$, then

$$|m_n - m|^2 \to 0 \quad \text{and} \quad \|C_n - C\|_2 \to 0 \quad \text{as } n \to \infty, \tag{4}$$

where m and C are the mean and covariance matrix of the limiting Gaussian, and the norm $\|\cdot\|_2$ is any matrix norm (e.g., spectral norm, Frobenius norm).

One immediate consequence of either (2) or (3) is that if A is any $d \times d$ matrix and $X \sim \mathcal{N}_d(m, C)$, then:

$$AX \sim \mathcal{N}(Am, A\Sigma A^\top). \tag{5}$$

A judicious choice of A then allows us to compute conditional distributions as well. If $n < d$, make the partitions:

$$X = (X_1, X_2), \quad m = (m_1, m_2), \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $X_1 \in \mathbb{R}^n$, $X_2 \in \mathbb{R}^{d-n}$, and C_{11} is $n \times n$. Then the conditional distribution of X_1 given X_2 is also Gaussian, with mean vector:

$$m_{1|2} = m_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - m_2),$$

and covariance matrix:

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

---

and the lower bound follows by substituting $u = x + v/x$, leading to:

$$\int_x^\infty e^{-u^2/2}du = \frac{1}{x}e^{-x^2/2}\int_0^\infty e^{-v(1-v/(2x^2))}dv \geq \frac{1}{x}e^{-x^2/2}\int_0^\infty (1 - v/(2x^2))dv \geq \left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x).$$

Gaussian Process

We can now define a real-valued Gaussian (random) field or Gaussian (random) process to be a stochastic process X on a parameter set T for which the (finite-dimensional) distributions of $(X_{t_1}, \ldots, X_{t_n})$ are multivariate Gaussian for each $1 \leq n < \infty$ and each $(t_1, \ldots, t_n) \in T^n$.

Since multivariate Gaussian distributions are determined by means and covariances, it is immediate that Gaussian random fields are determined by their mean functions

$$m(t) = \mathbb{E}[X(t)],$$

and their covariance functions

$$\Sigma(s,t) = \mathbb{E}[(X_s - m_s)(X_t - m_t)]. \tag{6}$$

In fact, this is one of the main reasons, beyond ubiquitous but not always justified appeals to the central limit theorem, that Gaussian processes are such popular and useful choices for models for random processes on general spaces.

Multivariate Gaussian fields taking values in $\mathbb{R}^d$ are stochastic processes for which $\langle \alpha, X_t \rangle$ is a real-valued Gaussian process for every $\alpha \in \mathbb{R}^d$. In this case, $m(t)$ takes values in $\mathbb{R}^d$ and the covariance function of (6) is replaced by a function whosevalues are non-negative definite, $d \times d$, matrices. In particular,

$$\Sigma(s,t) = \mathbb{E}[(X_s - m_s)(X_t - m_t)^\top].$$

As in the real valued case, for Gaussian X the vector function m and the matrix function Σ determine all of its statistical properties.

In fact, one can also go in the other direction as well. Given any set T, a function $m : T \to \mathbb{R}$, and a nonnegative definite function $\Sigma : T \times T \to \mathbb{R}$, there exists[3] a Gaussian process on T with mean function m and covariance function Σ.

Putting all this together, we have the important principle that for a Gaussian process, *everything about it is determined by the mean and covariance functions*. The fact that no real structure is required of the parameter space T is what makes Gaussian fields such a useful model for random processes on general spaces. To build an appreciation for this, you may want to to look at some examples. However, you will get more out of that section if you first bear with us to answer one of the most fundamental questions in the theory of Gaussian processes: When are Gaussian processes (almost surely) bounded and/or continuous?

Boundedness and Continuity

Now we turn to discussing basic sample path properties. For example, we would like to know when a stochastic process X is continuous, or continuously differentiable. These two issues are precisely what we shall look at in this section, mainly for Gaussian processes.

It is actually quite easy to describe what kinds of conditions are needed to ensure that these smoothness properties hold, at least in the Gaussian case. For simplicity for the introduction, suppose that X is Gaussian, centered, and stationary, with covariance function Σ. Since Σ determines the finite dimensional distributions of X, and these determine its properties, ultimately we are looking for conditions on Σ.

Now, if X is to be smooth, it must be true that for s and t close, the difference $X(t) - X(s)$ must be small.

[3]This is a consequence of the Kolmogorov existence theorem, which, at this level of generality, can be found in Dudley. Such a process is a random variable in $\mathbb{R}^T$ and may have terrible properties, including lack of measurability in t. However, it will always exist.

However, these differences have a known distribution,

$$\begin{aligned} X(t) - X(s) &\sim \mathcal{N}\left(0, \mathbb{E}\left[|X(t) - X(s)|^2\right]\right) \\ &= \mathcal{N}\left(0, 2\left|\Sigma(0) - \Sigma(t-s)\right|\right), \end{aligned}$$

the second line a consequence of stationarity. Thus, what we require is that Σ itself is smooth in the neighborhood of the origin. The only question is "exactly how smooth does it need to be?", and this is what we plan to answer.

Before going through, we first recall that there are several notions of continuity for stochastic processes over a parameter space T, among them

- Continuity in probability:

$$\lim_{s \to t} \mathbb{P}\left(|X(t) - X(s)| \geq \varepsilon\right) = 0, \qquad \text{for each } t \in T \text{ and for each } \varepsilon > 0.$$

- Continuity in mean square, or L^2 continuity:

$$\lim_{s \to t} \mathbb{E}\left[|X(t) - X(s)|^2\right] = 0, \qquad \text{for each } t \in T.$$

- Continuity with probability one, sample path, or almost sure (a.s.), continuity:

$$\mathbb{P}\left(\lim_{s \to t} |X(t) - X(s)| = 0, \qquad \text{for all } t \in T\right) = 1.$$

Now we develop a useful sufficient condition for a centered Gaussian process on a parameter space T to be almost surely bounded and/or continuous, i.e., to determine conditions for which

$$\mathbb{P}\left\{\sup_{t \in T} |X(t)| < \infty\right\} = 1 \quad \text{or} \quad \mathbb{P}\left\{\lim_{s \to t} |X(t) - X(s)| = 0, \ \forall t \in T\right\} = 1.$$

Of course, in order to talk about continuity—i.e., for the notation $s \to t$ above to have some meaning—it is necessary that T have some topology, so we assume that (T, τ) is a metric space, and that continuity is in terms of the τ-topology.

Our first step is to show that τ is irrelevant to the question of continuity. This is rather useful, since we shall also soon show that boundedness and continuity are essentially the same problem for Gaussian fields, and formulating the boundedness question requires no topological demands on T.

To start, define a new metric d on T by:

$$d(s,t) := \left\{\mathbb{E}\left[(X(s) - X(t))^2\right]\right\}^{1/2}. \tag{7}$$

Actually, d is only a pseudometric, since although it satisfies all the other demands of a metric, $d(s,t) = 0$ does not necessarily imply that $s = t$. Nevertheless, we shall abuse terminology by calling d the canonical metric for T and/or X.

We will see that, at least on compact T, boundedness and continuity are equivalent problems. Furthermore, both depend on how "large" the parameter set T is when size is measured in a metric that comes from the process itself.

We first need a notation for the ball, of radius ε, in the canonical metric d, centered at a point $t \in T$, which we denote by

$$B_d(t, \varepsilon) := \{s \in T : d(s,t) \leq \varepsilon\}.$$

To understand what this means about the random process, note from (7) that for, $s, t \in T$, the d-distance between any two points is the standard deviation of the difference $X(t) - X(s)$. Thus, in regions where this standard deviation is large, we expect the process to move around comparatively rapidly, and, as a consequence, the 'physical' balls of fixed d-radius will be smaller.

Hence roughly speaking, the number of balls needed to cover T must, in some way, measure the behavior of X over T. This brings us to the notion of metric entropy.

Definition ([1, Definition 1.3.2])**.** Let X be a centered Gaussian field on T, and d the canonical metric (7). Assume that T is d-compact, and write: for the d-ball centered at $t \in T$ and of radius ε. Let $N(T, d, \varepsilon) \equiv N(\varepsilon)$ denote the smallest number of such balls that cover T, and set:

$$H(T, d, \varepsilon) \equiv H(\varepsilon) = \ln N(\varepsilon).$$

Then N and H are called the metric entropy and log-entropy functions for T (or X). We shall refer to any condition or result based on N or H as an *entropy condition/result*.

Note that since we are assuming that T is d-compact, it follows that $H(\varepsilon) < \infty$ for all $\varepsilon > 0$. The same need not be (nor generally is) true for $\lim_{\varepsilon \to 0} H(\varepsilon)$, since as $\epsilon \to 0$ the d-balls get smaller and smaller, and we need more and more of them to cover T. It is precisely the growth rate of H at zero that captures the 'smoothness of the covariance at the origin'. Furthermore, note for later use that if we define:

$$\operatorname{diam}(T) := \sup_{s,t \in T} d(s, t), \tag{8}$$

then $N(\varepsilon) = 1$ and so $H(\varepsilon) = 0$ for all $\varepsilon \geq \operatorname{diam}(T)$.

Here then is the main result about Gaussian continuity and boundedness, due originally, more or less in the form given below, to Richard Dudley [30, 31]. However this result has a long and rich history, and is far from being the last word on the subject.

Theorem ([1, Theorem 1.3.3])**.** *Let X be a centered Gaussian field on a d-compact T, d the canonical metric, and H the corresponding entropy. Then there exists a universal constant K such that*

$$\mathbb{E}\left\{\sup_{t \in T} X_t\right\} \leq K \int_0^{\operatorname{diam}(T)/2} H^{1/2}(\varepsilon)\, d\varepsilon.$$

This result has immediate consequences for continuity. Define the modulus of continuity ω_F of a real-valued function F on a metric space (T, τ) as

$$\omega_F(\delta) \equiv \omega_{F,\tau}(\delta) := \sup_{\tau(s,t) \leq \delta} |F(t) - F(s)|, \qquad \delta > 0.$$

The modulus of continuity of X can be thought of as the supremum of the stochastic process

$$X_{s,t} = X_t - X_s$$

over a certain neighborhood of $T \times T$, in that

$$\omega_{X,\tau}(\delta) = \sup_{\substack{(s,t) \in T \times T \\ \tau(s,t) \leq \delta}} X(s, t).$$

(Note we can drop the absolute value sign since the supremum here is always nonnegative.)

More precisely, write $d^{(2)}$ for the canonical metric of $X_{s,t}$ on $T \times T$. Then:

$$d^{(2)}\big((s,t),(s',t')\big) = \Big[\mathbb{E}\Big\{\big((X_t - X_s) - (X_{t'} - X_{s'})\big)^2\Big\}\Big]^{1/2} \le 2\max(d(s,t), d(s',t')).$$

and so

$$N\Big(\{(s,t) : d(s,t) \le \delta\},\ d^{(2)},\ \delta\Big) \le N(T,d,\delta/2).$$

From these observations, [1, Theorem 1.3.3] immediately implies the following.

Corollary ([1, Corollary 1.3.4])**.** *Under the conditions of [1, Theorem 1.3.3], there exists a universal constant K such that*

$$\mathbb{E}\left\{\omega_{X,d}(\delta)\right\} \le K \int_0^\delta H^{1/2}(\varepsilon)\, d\varepsilon. \tag{9}$$

Note that this is not quite enough to establish the almost sure continuity of X. Continuity is, however, not far away, since the same construction used to prove [1, Theorem 1.3.3] will also give us the following, which, with the elementary tools we have at hand at the moment, neither follows from nor directly implies (9).

Theorem ([1, Theorem 1.3.5])**.** *Under the conditions of [1, Theorem 1.3.3] there exists a random $\eta \in (0,\infty)$ and a universal constant K such that*

$$\omega_{X,d}(\delta) \le K \int_0^\delta H^{1/2}(\varepsilon)\, d\varepsilon, \qquad \delta < \eta. \tag{10}$$

Note that (10) is expressed in terms of the d modulus of continuity. Translating this to a result for the τ modulus is trivial.

Reading above theorems carefully, you will note that there is no claim that the sufficient conditions given there are also necessary and, indeed, they generally are not. However, when X is a stationary, centered, Gaussian process, there is such a result, and then it can be shown that

$$\begin{aligned} X \text{ is a.s. continuous on } T &\iff X \text{ is a.s. bounded on } T \\ &\iff \int_0^\infty H^{1/2}(\varepsilon) d\varepsilon < \infty. \end{aligned} \tag{11}$$

We shall not go into proof, but we check two important observations.

Observation 1. If X is a separable process on T, then $\sup_{t\in T} X_t$ is a well-defined (i.e., measurable) random variable.

Measurability follows from the separability definition, which gave us a countable dense set $D \subset T$ such that

$$\sup_{t\in T} X_t = \sup_{t\in D} X_t.$$

The supremum of a countable set of measurable random variables is always measurable.

One can manage without separability for the proof by defining

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] := \sup\left\{\mathbb{E}[\sup_{t\in F} X_t] : F \subset T,\ F \text{ finite}\right\}.$$

Observation 2. If X is a separable process on T and ξ is a centered random variable (not necessarily independent of X), then

$$\mathbb{E}\left\{\sup_{t\in T}(X_t + \xi)\right\} = \mathbb{E}\left\{\sup_{t\in T} X_t\right\}.$$

As trivial as this is, it fails if we replace $\sup_t(X_t + \xi)$ with $\sup_t |X_t + \xi|$.

Examples

We shall start by looking at stochastic processes on $\mathbb{R}^N$, and then look at the Brownian family of processes. This family is also very instructive in understanding why the finiteness of entropy integrals is a natural condition for continuity.

Gaussian process on $\mathbb{R}^N$

Returning to Euclidean space after the abstraction of entropy on general metric spaces, it is natural to expect that conditions for continuity and boundedness will become so simple to both state and prove that there was really no need to introduce such abstruse general concepts.

This expectation is both true and false. It turns out that avoiding the notion of entropy does not make it any easier to establish continuity theorems, and indeed, reliance on the specific geometry of the parameter space often confounds the basic issues.

On the other hand, the basic results for Gaussian processes on $\mathbb{R}^N$ are easy to state without specifically referring to any abstract notions. We shall do this for continuity. We assume that X is a centered Gaussian process with continuous covariance function Σ defined on a compact $T \subset \mathbb{R}^N$.

We start by defining a function which captures the size of increments, by setting

$$\begin{aligned} p^2(u) &= \sup_{|s-t|\le u} \mathbb{E}\left[|X_s - X_t|^2\right], \\ &= \sup_{|s-t|\le u} \left[\Sigma(t,t) + \Sigma(s,s) - 2\Sigma(s,t)\right], \end{aligned}$$

where $|\cdot|$ is the usual Euclidean norm. If X is stationary, then

$$p^2(u) = 2 \sup_{|t|\le u} \left[\Sigma(0) - \Sigma(t)\right].$$

And so one sees that the rate of convergence on p to zero as $u \to 0$ is closely related to the smoothness of Σ at the origin discussed above.

Here is the main result on continuity of Gaussian random fields on $\mathbb{R}^N$.

Theorem ([1, Theorem 1.4.1]). *If, for some $\delta > 0$, either*

$$\int_0^\delta (-\ln u)^{1/2}\, dp(u) < \infty \qquad \text{or} \qquad \int_\delta^\infty p\left(e^{-u^2}\right) du < \infty, \tag{12}$$

then X is continuous and bounded on T with probability one.

A sufficient condition for either integral in (12) *to be finite is that for some $0 < C < \infty$ and $\alpha, \eta > 0$,*

$$\begin{aligned} \mathbb{E}\left[|X_s - X_t|^2\right] &= \Sigma(t,t) + \Sigma(s,t) - 2\Sigma(s,t) \\ &\le \frac{C}{|\log|s-t||^{1+\alpha}}, \end{aligned} \tag{13}$$

for all s, t with $|s-t| < \eta$. *Furthermore, there exists a constant K, dependent only on the dimension N, and a random $\delta_0 > 0$ such that for all $\delta < \delta_0$,*

$$\omega_X(\delta) \le K \int_0^{p(\delta)} (-\ln u)^{1/2}\, dp(u), \tag{14}$$

where the modulus of continuity ω_X is taken with respect to the Euclidean metric. A similar bound, in the spirit of (9)*, holds for* $\mathbb{E}[\omega_X(\delta)]$.

The various sufficient conditions for continuity of [1, Theorem 1.4.1] are quite sharp, but not necessary. Still, the results of [1, Theorem 1.4.1] are, from a practical point of view, reasonably definitive. For example, if X is stationary, then, following on from (11), it is possible to check that if

$$\frac{C_1}{(-\log|t|)^{1+\alpha_1}} \leq \Sigma(0) - \Sigma(t) \leq \frac{C_2}{(-\log|t|)^{1+\alpha_2}}, \tag{15}$$

for $|t|$ small enough, then X will be continuous if $\alpha_2 > 0$ and discontinuous if $\alpha_1 < 0$.

In practical situations, it is rare indeed that one even gets close to the logarithmic behavior of (13) or (15). The more common situation in the applications is that the covariance function has a power series representation of the form

$$\Sigma(s,t) = \Sigma(t,t) - (t-s)^\top \Lambda_t (t-s) + o(|t-s|^{2+\delta}),$$

for $|t-s|$ small and some $\delta > 0$, or, in the stationary case

$$\Sigma(t) = \Sigma(0) - t^\top \Lambda t + o(|t|^{2+\delta}),$$

for t in the neighborhood of the origin. The matrices Λ_t and Λ are $N \times N$ and positive definite.

The Brownian Family of Processes

Perhaps the most basic of all random fields is a collection of independent Gaussian random variables. While it is simple to construct such random fields for finite and even countable parameter sets, deep technical difficulties obstruct the construction for uncountable parameter sets. The path that we shall take around these difficulties involves the introduction of random measures, which, at least in the Gaussian case, are straightforward to formulate.

Let $(T, \mathcal{T}, \nu)$ be a σ-finite measure space and denote by T_ν the collection of sets of T of finite ν-measure. A Gaussian noise based on ν, or "Gaussian ν-noise,'' is a random field $W : T_\nu \to \mathbb{R}$ such that for all $A, B \in T_\nu$,

$$W(A) \sim \mathcal{N}(0, \nu(A)), \tag{16}$$

$$A \cap B = \emptyset \Rightarrow W(A \cup B) = W(A) + W(B) \quad \text{a.s.}, \tag{17}$$

$$A \cap B = \emptyset \Rightarrow W(A) \text{ and } W(B) \text{ are independent.} \tag{18}$$

Property (17) encourages one to think of W as a random (signed) measure, although it is not generally σ-finite. We describe (18) by saying that W has independent increments.

Theorem ([1, Theorem 1.4.3])**.** *If $(T, \mathcal{T}, \nu)$ is a measure space, then there exists a real-valued Gaussian noise, defined for all $A \in T_\nu$, satisfying* (16)*,* (17)*,* (18)*.*

A particularly simple example of Gaussian noise arises when $T = \mathbb{Z}$. Take ν a discrete measure of the form $\nu(A) = \sum_k a_k \delta_k(A)$, where the a_k are nonnegative constants and δ_k is the Dirac measure on $\{k\}$. For T take all subsets of $\mathbb{Z}$. In this case, the Gaussian noise can actually be defined on points $t \in T$ and extended to a signed measure on sets in T_ν by additivity. What we get is a collection $\{W_k\}_{k\in\mathbb{Z}}$ of independent, centered, Gaussian variables, with $\mathbb{E}\{W_k^2\} = a_k$. If the a_k are all equal, this is classical Gaussian "white'' noise on the integers.

A more interesting case is $T = \mathbb{R}^N$, $\mathcal{T} = \mathcal{B}^N$ (the Borel σ-algebra on $\mathbb{R}^N$), and $\nu(\cdot) = |\cdot|$, Lebesgue measure. This gives us a Gaussian white noise defined on the Borel subsets of $\mathbb{R}^N$ of finite Lebesgue measure, which is also a field with orthogonal increments, in the sense of (18). It is generally called the set-indexed Brownian sheet. It is not possible, in this case, to assign nontrivial values to given points $t \in \mathbb{R}^N$, as was the case in the previous example.

It also turns out that working with the Brownian sheet on all of $\mathcal{B}^N$ is not really the right thing to do, since, as will follow from [1, Theorem 1.4.4], this process is rather badly behaved. Restricting the parameter space to various classes of subsets of $\mathcal{B}^N$ is the right approach. Doing so gives us a number of interesting examples, with which the remainder of this section is concerned.

As a first step, restrict W to rectangles of the form $[0,t] \subset \mathbb{R}^N_+$, where $t \in \mathbb{R}^N_+ = \{(t_1,\dots,t_N) : t_i \geq 0\}$. It then makes sense to define a random field on $\mathbb{R}^N_+$ itself via the equivalence

$$W(t) = W([0,t]). \tag{19}$$

W_t is called the Brownian sheet on $\mathbb{R}^N_+$, or multiparameter Brownian motion. It is easy to check that this field is the centered Gaussian process with covariance

$$\mathbb{E}\{W_s W_t\} = (s_1 \wedge t_1)\cdots(s_N \wedge t_N), \tag{20}$$

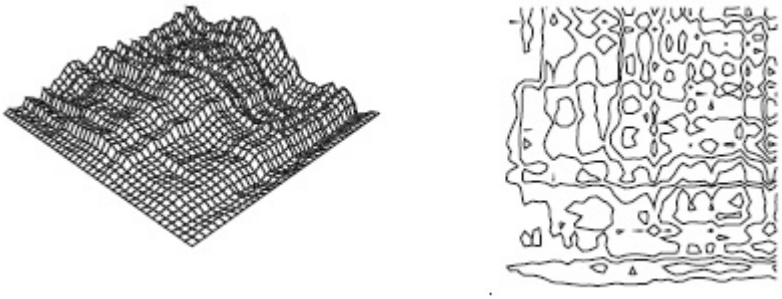
where $s \wedge t = \min(s,t)$.

When $N = 1$, W is the standard Brownian motion on $[0,\infty)$. For this case, (16), (17), (18), and [1, Theorem 1.4.4] can be rewritten as follows: $\{W_t : t \geq 0\}$ is a stochastic process with

- $W_0 = 0$ almost surely
- For $0 \leq s < t$, the increment $W_t - W_s \sim \mathcal{N}(0, t-s)$.
- The increment $W_{t_i} - W_{t_{i-1}}$ are independent for $t_0 < t_1 < \cdots < t_n$.
- The sample paths $t \mapsto W_t$ are continuous almost surely.

The above four conditions are regarded as a definition of one-dimensional standard Brownian motion.

When $N > 1$, if we fix $N-k$ of the indices, it is a scaled k-parameter Brownian sheet in the remaining variables. (This is easily checked via the covariance function.)



Above figure ([1, Figure 1.4.1]) is a simulated Brownian sheet on $[0,1]^2$, along with its contour lines at the zero level. One of the rather interesting aspects of the contour lines of the above figure is that they are predominantly parallel to the axes. There is a rather deep reason for this, and it has generated a rather massive literature.

The most basic of the sample path properties of the Brownian sheets are the continuity results of the following theorem. Introduce a partial order on $\mathbb{R}^k$ by writing $s \prec (\precsim) t$ if $s_i < (\leq) t_i$ for each $i = 1,\dots,k$, and for $s \prec t$, let $\Delta(s,t) = \prod_1^N [s_i, t_i]$. Although $W(\Delta(s,t))$ is already well defined via the original set-indexed process, it is also helpful to think of it as the "increment" of the point-indexed W_t over (s,t), , that is,

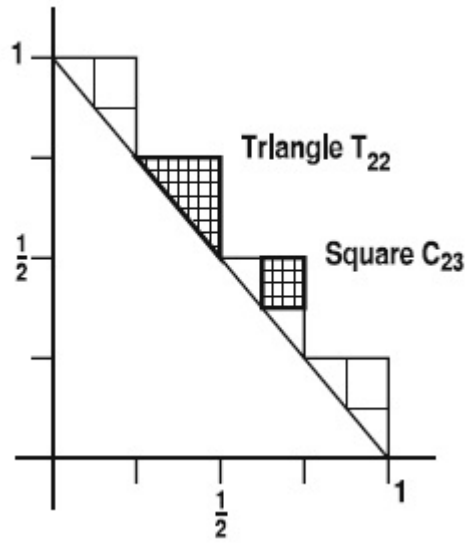
$$W(\Delta(s,t)) = \sum_{\alpha \in \{0,1\}^N} (-1)^{N - \sum_{i=1}^N \alpha_i} W\left(s + \sum_{i=1}^N \alpha_i (t_i - s_i)\right). \tag{21}$$

We call $W(\Delta(s,t))$ the rectangle-indexed version of W.

Theorem ([1, Theorem 1.4.4]). *The point and rectangle-indexed Brownian sheets are continuous over compact* $T \subset \mathbb{R}^N$.

In the framework of the general set-indexed sheet, [1, Theorem 1.4.4] states that the Brownian sheet is continuous over $A = \{\text{all rectangles in } T\}$ for compact T, and so bounded. This is far from a trivial result, for enlarging the parameter set, for the same process, can lead to unboundedness. The easiest way to see this is with an example.

An interesting, but quite simple example is given by the class of *lower layers* in $[0,1]^2$. A set A in $\mathbb{R}^N$ is called a lower layer if $s \prec t$ and $t \in A$ implies $s \in A$. In essence, restricted to $[0,1]^2$ these are sets bounded by the two axes and a nonincreasing line. See below figure ([1, Figure 1.4.2]) for examples for lower layers.



Theorem ([1, Theorem 1.4.4]). *The Brownian sheet on lower layers in* $[0,1]^2$ *is discontinuous and unbounded with probability one.*

From now on, we consider $N = 1$ and W is the standard Brownian motion on $[0, \infty)$.

Filtrations and Martingales

Throughout this section, we consider a probability space $(\Omega, \mathcal{F}, P)$.In this section, we introduce some general notions that will be of constant use later.

Definition ([2, Definition 3.1]). A filtration on $(\Omega, \mathcal{F}, P)$ is a collection $(\mathcal{F}_t)_{0 \le t \le \infty}$ indexed by $[0, \infty]$ of sub-σ-fields of $\mathcal{F}$, such that $\mathcal{F}_s \subset \mathcal{F}_t$ for every $s \le t \le \infty$.

We have thus, for every $0 \le s < t$,

$$\mathcal{F}_0 \subset \mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}_\infty \subset \mathcal{F}.$$

We also say that $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ is a filtered probability space.

If B is a Brownian motion, we consider the filtration

$$\mathcal{F}_t = \sigma(B_s, 0 \le s \le t), \qquad \mathcal{F}_\infty = \sigma(B_s, s \ge 0).$$

Let $(\mathcal{F}_t)_{0 \le t \le \infty}$ be a filtration on $(\Omega, \mathcal{F}, P)$. We set, for every $t \ge 0$,

$$\mathcal{F}_{t+} = \bigcap_{s > t} \mathcal{F}_s,$$

and $\mathcal{F}_{\infty+} = \mathcal{F}_\infty$. Note that $\mathcal{F}_t \subset \mathcal{F}_{t+}$ for every $t \in [0,\infty]$. The collection $(\mathcal{F}_{t+})_{0\le t\le\infty}$ is also a filtration.We say that the filtration $\mathcal{F}_t$ is right-continuous if

$$\mathcal{F}_{t+} = \mathcal{F}_t, \qquad \forall t \ge 0.$$

By construction, the filtration $(\mathcal{F}_{t+})$ is right-continuous.

Let $(\mathcal{F}_t)$ be a filtration and let $\mathcal{N}$ be the class of all $(\mathcal{F}_\infty, P)$-negligible sets (i.e. $A \in \mathcal{N}$ if there exists an $A' \in \mathcal{F}_\infty$ such that $A \subset A'$ and $P(A') = 0$). The filtration is said to be complete if $\mathcal{N} \subset \mathcal{F}_0$ (and thus $\mathcal{N} \subset \mathcal{F}_t$ for every t).

Definition ([2, Definition 3.3])**.** A random process $(X_t)_{t\ge0}$ is called adapted if, for every $t \ge 0$, X_t is $\mathcal{F}_t$-measurable.

Definition ([2, Definition 3.10])**.** An adapted real-valued process $(X_t)_{t\ge0}$ such that $X_t \in L^1$ for every $t \ge 0$ is called

- a martingale if, for every $0 \le s < t$, $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$;
- a supermartingale if, for every $0 \le s < t$, $\mathbb{E}[X_t|\mathcal{F}_s] \le X_s$;
- a submartingale if, for every $0 \le s < t$, $\mathbb{E}[X_t|\mathcal{F}_s] \ge X_s$.

Definition ([2, Definition 3.11])**.** A real-valued process $B = (B_t)_{t\ge0}$ is an $(\mathcal{F}_t)$-Brownian motion if B is a Brownian motion and if B is adapted and has independent increments with respect to $(\mathcal{F}_t)$. Similarly, a process $B = (B_t)_{t\ge0}$ with values in $\mathbb{R}^d$ is a d-dimensional $(\mathcal{F}_t)$-Brownian motion if B is a d-dimensional Brownian motion and is adapted with independent increments relative to $(\mathcal{F}_t)$.

Note that if B is a (possibly completed) Brownian motion and $(\mathcal{F}^B_t)$ is its canonical filtration, then B is a $(\mathcal{F}^B_t)$-Brownian motion.

Let B be an $(\mathcal{F}_t)$-Brownian motion started from 0 (or from any $a \in \mathbb{R}$). Then it follows from the above observations that the processes

$$B_t, \quad B_t^2 - t, \quad e^{\lambda B_t - \frac{\lambda^2}{2}t}$$

are martingales with continuous sample paths. The processes $e^{\lambda B_t - \frac{\lambda^2}{2}t}$ are called exponential martingales of Brownian motion.

We have the analogous of maximal inequality and Doob's inequality for continuous martingales as well.

Proposition ([2, Proposition 3.15])**.** *(i) (Maximal inequality) Let $(X_t)_{t\ge0}$ be a supermartingale with right-continuous sample paths. Then, for every $t > 0$ and every $\lambda > 0$,*

$$\lambda\,\mathbb{P}\left(\sup_{0\le s\le t}|X_s| > \lambda\right) \le \mathbb{E}[|X_0|] + 2\mathbb{E}[|X_t|].$$

(ii) (Doob's inequality in L^p) Let $(X_t)_{t\ge0}$ be a martingale with right-continuous sample paths. Then, for every $t > 0$ and every $p > 1$,

$$\mathbb{E}\left[\sup_{0\le s\le t}|X_s|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_t|^p].$$

Stochastic Calculus

While Brownian motion serves as the fundamental model for stochastic dynamics, many natural and applied systems exhibit structured randomness beyond pure Brownian fluctuations. In particular, real-world stochastic processes often incorporate deterministic drift, governing systematic trends, and state-dependent volatility, modulating stochastic fluctuations.

A natural extension is given by diffusion processes, which generalize Brownian motion by satisfying the stochastic differential equation (SDE)

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t,$$

where $b(X_t)$ represents the deterministic drift, and $\sigma(X_t)$ is the diffusion coefficient controlling local variability.

These processes arise in diverse applications, from physics (particle transport, thermodynamics) to finance (stochastic modeling of asset prices) and biology (population dynamics, neural activity).

The rigorous mathematical formulation of diffusion processes requires a well-defined framework for stochastic integration and differentiation, which extends classical calculus to non-differentiable paths. This leads to the development of *Itô calculus*, a fundamental tool for analyzing stochastic dynamics.

Now we develop the Itô integral and the Itô formula. Itô calculus provides the necessary framework for rigorously defining stochastic integrals and differential equations that describe the evolution of processes with both deterministic drift and stochastic components.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions (increasing, right-continuous, and complete). Let $(W_t)_{t\geq 0}$ be a standard Brownian motion adapted to $(\mathcal{F}_t)$.

Definition ([3, Definition 2.1])**.** Stochastic Integral. For a process $f : [0,T]\times\Omega \to \mathbb{R}$ that is $\mathcal{F}_t$-adapted and belongs to the space $L^2([0,T]\times\Omega)$, i.e.,

$$\mathbb{E}\left[\int_0^T |f(t,\omega)|^2 dt\right] < \infty,$$

the Itô integral of f with respect to W_t over $[0,T]$ is defined as the limit of stochastic sums:

$$I(f) = \int_0^T f(t)\,dW_t := \lim_{|P|\to 0} \sum_{i=0}^{n-1} f(t_i)(W_{t_{i+1}} - W_{t_i}),$$

in probability, where $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a partition of $[0,T]$, $|P|$ is the maximum subinterval length, and $t_i \in [t_i, t_{i+1}]$.

Theorem ([3, Theorem 2.1])**.** *Fundamental Properties of the Itô Integral. Let $f, g \in L^2([0,T]\times\Omega)$ be $\mathcal{F}_t$-adapted processes and let $a, b \in \mathbb{R}$. Then:*

1. Linearity:
$$\int_0^T (af(t) + bg(t))\,dW_t = a\int_0^T f(t)\,dW_t + b\int_0^T g(t)\,dW_t.$$

2. Isometry:
$$\mathbb{E}\left[\left(\int_0^T f(t)\,dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T f(t)^2\,dt\right].$$

3. Martingale Property: The process
$$M_t = \int_0^t f(s)\,dW_s, \quad t \in [0,T],$$
is a martingale with respect to $(\mathcal{F}_t)$.

4. Quadratic Variation:
$$\left[\int_0^t f(s)\,dW_s\right]_t = \int_0^t f(s)^2\,ds.$$

5. Zero-Mean Property:
$$\mathbb{E}\left[\int_0^T f(t)\,dW_t\right] = 0.$$

Brownian motion serves as the fundamental model for continuous-time stochastic processes, but many real-world phenomena exhibit structured randomness beyond its purely diffusive nature. Iro processes generalize Brownian motion by incorporating a deterministic drift term and a statedependent diffusion coefficient. The rigorous definition of Ito processes relies on Ito integration, which provides a well-defined framework for handling stochastic integrals. In this section, we present the formal definition of an Ito process and establish its properties within the broader framework of stochastic calculus.

Definition ([3, Definition2 .3])**.** Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space and let W_t be a standard m-dimensional Brownian motion adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. A vector-valued stochastic process $\{X_t\}_{t\geq 0} \subset \mathbb{R}^d$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, is called a d-dimensional **Itô process** if it can be expressed as:

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s, \quad t \geq 0,$$

for some $\mathcal{F}_0$-measurable random vector $X_0 \in \mathbb{R}^d$, and for functions:

$$\mu : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \quad \text{(drift vector)}, \quad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times m} \quad \text{(diffusion matrix)},$$

such that all expressions are well-defined and

$$\int_0^T \|\mu(s, X_s)\|\, ds < \infty, \quad \int_0^T \|\sigma(s, X_s)\|_F^2\, ds < \infty, \quad \mathbb{P}\text{-a.s., for all } T > 0.$$

If $\{X_t\}_{t\geq 0}$ is an Itô process of the form above, it is sometimes written in the shorthand differential notation as the **stochastic differential equation (SDE)**:

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t.$$

References

[1] Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York, 2007.

[2] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer, [Cham], french edition, 2016.

[3] Helder Rojas. Stochastic processes and diffusion equations, 2025.