Calculus on Gauss Space:
An Introduction to Gaussian Analysis

Tom Alberts and Davar Khoshnevisan

Version 5.0
September 25, 2018 (Davar)

Contents

Part I

The Finite-Dimensional Theory

Chapter 1

The Canonical Gaussian Measure on $\mathbb{R}^n$

?⟨ch:Canonical_Gaussian⟩?

1 Introduction

The main goal of this book is to study "Gaussian measures," the simplest example of which is the *canonical Gaussian measure* P_n on $\mathbb{R}^n$, where $n \geqslant 1$ is an arbitrary integer. The measure P_n is defined simply as

$$\mathrm{P}_n(A) := \int_A \gamma_n(x)\,\mathrm{d}x \qquad \text{for all Borel sets } A \subseteq \mathbb{R}^n,$$

where γ_n denotes the standard normal density function on $\mathbb{R}^n$, viz.,

$$\gamma_n(x) := \frac{\mathrm{e}^{-\|x\|^2/2}}{(2\pi)^{n/2}} \qquad [x \in \mathbb{R}^n]. \tag{1.1}$$

gamma_n

The function γ_1 describes the famous "bell curve," and γ_n looks like a suitable "rotation" of the curve of γ_1 when $n > 1$.

We frequently drop the subscript n from P_n when it is clear which dimension we are in.

Throughout, we consider the probability space $(\Omega\,, \mathcal{F}, \mathrm{P})$, where we have dropped the subscript n from P_n, and

$$\Omega := \mathbb{R}^n, \quad \text{and} \quad \mathcal{F} := \mathcal{B}(\mathbb{R}^n),$$

where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel subsets of $\mathbb{R}^n$.

Recall that measurable functions $f : \mathbb{R}^n \to \mathbb{R}$ are random variables, and measurable functions $F : \mathbb{R}^n \to \mathbb{R}^n$ can be regarded as random vectors. Throughout this book, we designate by $Z = (Z_1\,, \ldots, Z_n)$ the random vector

$$Z_j(x) := x_j \qquad \text{for all } x \in \mathbb{R}^n \text{ and } 1 \leqslant j \leqslant n. \tag{1.2}$$

Z

Thus, Z always denotes a random vector of n i.i.d. standard normal random variables on our probability space. In particular,

$$\mathrm{P}_n(A) = \mathrm{P}\{Z \in A\} \qquad \text{for all Borel sets } A \subseteq \mathbb{R}^n.$$

We also let $\mathrm{E} := \mathrm{E}_n$ denote the expectation operator for $\mathrm{P} = \mathrm{P}_n$, which allows us to write integrals, using shorthand, as

$$\mathrm{E}[f(Z)] = \int_{\mathbb{R}^n} f(x)\,\mathrm{P}(\mathrm{d}x) = \int_{\mathbb{R}^n} f(x)\gamma_n(x)\,\mathrm{d}x.$$

One of the elementary, though useful, properties of the measure P_n is that its "tails" are vanishingly small.

⟨lem:tails⟩ **Lemma 1.1.** *As $t \to \infty$,*

$$\mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| > t\right\} = \frac{2 + o(1)}{2^{n/2}\Gamma(n/2)}\, t^{n-2}\mathrm{e}^{-t^2/2},$$

where $\Gamma(\nu) := \int_0^\infty t^{\nu-1}\exp(-t)\,\mathrm{d}t$ is the gamma function evaluated at $\nu > 0$.

Proof. Define

$$S_n := \|Z\|^2 = \sum_{i=1}^n Z_i^2 \qquad \text{for all } n \geqslant 1. \tag{1.3}$$ S_n

Because S_n has a χ^2_n distribution,

$$\begin{aligned}\mathrm{P}\{x \in \mathbb{R}^n : \|x\| > t\} &= \mathrm{P}\{S_n > t^2\} \\ &= \frac{1}{2^{n/2}\Gamma(n/2)} \int_{t^2}^\infty x^{(n-2)/2}\mathrm{e}^{-x/2}\,\mathrm{d}x,\end{aligned} \tag{1.4}$$ eq:chi2:tail

for all $t \geqslant 0$. Now apply l'Hôpital's rule of calculus. ☐

The following large-deviations estimate is one of the ready consequences of Lemma 1.1: For every $n \geqslant 1$,

$$\lim_{t\to\infty} \frac{1}{t^2} \log \mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| > t\right\} = -\tfrac{1}{2}. \tag{1.5}$$ eq:LD

Of course, (1.5) is a weaker statement than Lemma 1.1. But it has the advantage of being "dimension independent." Dimension independence properties play a prominent role in the analysis of Gaussian measures. Here, for example, (1.5) teaches us that the tails of P_n behave roughly as do the tails of P_1 regardless of the value of $n \geqslant 1$.

Still, many of the more interesting properties of P_n are radically different from those of P_1 when n is large. In low dimensions—say $n = 1, 2, 3$—one can visualize the probability density function γ_n from (1.1). Based on that, or other methods, one knows that in low dimensions most of the mass of P_n concentrates near the origin. For example, an inspection of the standard normal table reveals that more than 68.26% of the total mass of P_1 is within one unit of the origin; see Figure 1.1.

In higher dimensions, however, the structure of P_n can be quite different. For example, let us first recall the random variable S_n from (1.3). Then, apply the weak law of large numbers XXX to deduce that S_n/n converges in probability to one, as $n \to \infty$.[1] Stated in other words,

$$\lim_{n\to\infty} \mathrm{P}\left\{x \in \mathbb{R}^n : (1-\varepsilon)n^{1/2} \leqslant \|x\| \leqslant (1+\varepsilon)n^{1/2}\right\} = 1, \tag{1.6}$$ pbm:CoM

for every $\varepsilon > 0$. The proof of (1.6) is short and can be reproduced right here: Recall that E denotes the expectation operator for $\mathrm{P} := \mathrm{P}_n$, and let Var be the corresponding variance operator. Since S_n has a χ^2 distribution with n degrees of freedom, simple computations show that $\mathrm{E}(S_n) = n$ and $\mathrm{Var}(S_n) = 2n$; see Problem 2 below. Therefore, Chebyshev's inequality yields $\mathrm{P}\{|S_n - \mathrm{E}(S_n)| > \varepsilon n\} \leqslant 2\varepsilon^{-2}n^{-1}$. Equivalently,

$$\mathrm{P}\left\{x \in \mathbb{R}^n : (1-\varepsilon)^{1/2}n^{1/2} \leqslant \|x\| \leqslant (1+\varepsilon)^{1/2}n^{1/2}\right\} \geqslant 1 - \frac{2}{n\varepsilon^2}. \tag{1.7}$$ WLLN

Thus we see that, when n is large, the measure P_n concentrates much of its total mass near the boundary of the centered ball of radius $n^{1/2}$, very far from the origin. A more careful examination shows that, in fact, very little of the total mass of P_n is elsewhere when n is large. The following theorem makes this statement much more precise. Theorem 1.2 is a simple consequence of a remarkable property of Gaussian measures that is known commonly as *concentration of measure* XXX. We will discuss this topic in more detail in due time.

⟨th:CoM:n⟩ **Theorem 1.2.** *For every $\varepsilon > 0$,*

$$\mathrm{P}\left\{x \in \mathbb{R}^n : (1-\varepsilon)n^{1/2} \leqslant \|x\| \leqslant (1+\varepsilon)n^{1/2}\right\} \geqslant 1 - 2\mathrm{e}^{-n\varepsilon^2}. \tag{1.8}$$ CoM:n

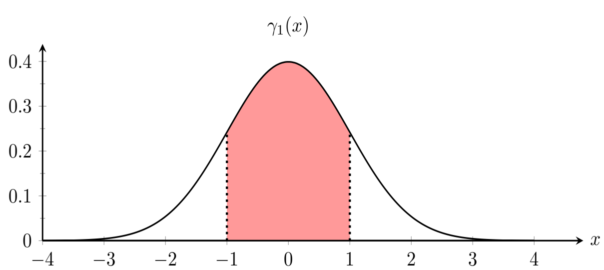

Figure 1.1. Area over $[-1\,,1]$ under the normal curve.
⟨fig:N(0,1)⟩

Theorem 1.2 does not merely improve the crude bound (1.7). Rather, it describes an entirely new phenomenon in high dimensions. To wit, let us consider the measure P_n when $n = 30,000$. When $\varepsilon = 0$, the left-hand side of (1.8) is equal to 0. But if ε is increased slightly, say to $\varepsilon = 0.01$, then the left-hand side of (1.8) increases to a probability > 0.9, whereas (1.7) reports a mere probability lower bound of $1/3$.

Proof. The result follows from a standard large-deviations argument that we reproduce next.

[1] In the present setting, it does not make sense to discuss almost-sure convergence since the underlying probability space is $(\mathbb{R}^n\,, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$.

Since $S_n := \sum_{i=1}^n Z_i^2$ has a χ_n^2 distribution, its moment generating function is

$$\mathrm{E}\,\mathrm{e}^{\lambda S_n} = (1-2\lambda)^{-n/2} \quad \text{for } -\infty < \lambda < 1/2, \tag{1.9}$$ mgf:chi2

and $\mathrm{E}\exp(\lambda S_n) = \infty$ when $\lambda \geqslant 1/2$. See Problem 2 below.

We use the preceding as follows: For all $t > 0$ and $\lambda \in (0\,,1/2)$,

$$\begin{aligned} \mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| > n^{1/2}t\right\} &= \mathrm{P}\left\{\mathrm{e}^{\lambda S_n} > \mathrm{e}^{\lambda n t^2}\right\} \\ &\leqslant (1-2\lambda)^{-n/2}\mathrm{e}^{-\lambda n t^2}, \end{aligned}$$

thanks to (1.9) and Chebyshev's inequality. The left-hand side is independent of $\lambda \in (0\,,1/2)$. Therefore, we may optimize the right-hand side over $\lambda \in (0\,,1/2)$ to find that

$$\begin{aligned} \mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| > n^{1/2}t\right\} &\leqslant \exp\left\{-n \sup_{0<\lambda<1/2}\left[\lambda t^2 + \frac{1}{2}\log(1-2\lambda)\right]\right\} \\ &= \exp\left\{-\frac{n}{2}\left[t^2 - 1 - 2\log t\right]\right\}. \end{aligned} \tag{1.10}$$ eq:tail:log

In particular, if $t > 1$, then it follows easily that the exponent of the right-most exponential in (1.10) is strictly positive, whence we have exponential decay of the probability as $n \to \infty$. This exponential decay is sharp; see Problem 3 below.

In any case, because $\log t < t - 1$ when $t > 1$, it follows from (1.10) that

$$\mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| > n^{1/2}t\right\} \leqslant \mathrm{e}^{-n(t-1)^2}. \tag{1.11}$$ BooBooBound

The special choice $t = 1 + \varepsilon$ yields (1.8) when $t > 1$.

When $t < 1$, we may argue similarly and write

$$\begin{aligned} \mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| < n^{1/2}t\right\} &= \mathrm{P}\left\{\mathrm{e}^{-\lambda S_n} > \mathrm{e}^{-\lambda n t^2}\right\} \qquad \text{for all } \lambda > 0 \\ &\leqslant \exp\left\{-n \sup_{\lambda>0}\left[-\lambda t^2 + \frac{1}{2}\log(1+2\lambda)\right]\right\} \\ &= \exp\left\{-\frac{n}{2}\left[t^2 - 1 - 2\log t\right]\right\}. \end{aligned}$$

Since $-2\log t > 2(1-t) + (1-t)^2$ when $t < 1$, it follows that

$$\begin{aligned} \mathrm{P}\left\{x \in \mathbb{R}^n : \|x\| < n^{1/2}t\right\} &\leqslant \exp\left\{-\frac{n}{2}\left[t^2 - 1 + 2(1-t) + (1-t)^2\right]\right\} \\ &\leqslant \mathrm{e}^{-n(1-t)^2}. \end{aligned}$$

Set $t = 1 - \varepsilon$ and combine with (1.11) to complete the proof. □

The preceding discussion shows that $\mathrm{P}\{\|Z\| \approx n^{1/2}\}$ is extremely close to one when n is large; that is, with very high probability, Z lies close to the boundary of the centered sphere of radius $n^{1/2}$. Of course, the latter is the sphere in the ℓ^2-norm, and it is worthwhile to consider how close Z lies to spheres in ℓ^p-norm instead, where $p \neq 2$. This problem too follows from the same analysis as above. In fact, another appeal to the weak law of large numbers shows that for all $\varepsilon > 0$,

$$\mathrm{P}\left\{(1-\varepsilon)\mu_p n^{1/p} \leqslant \|Z\|_p \leqslant (1+\varepsilon)\mu_p n^{1/p}\right\} \to 1 \qquad \text{as } n \to \infty, \tag{1.12}$$ 1_p

for all $p \in [1\,,\infty)$, where $\mu_p := \mathrm{E}(|Z_1|^p)$, and $\|\cdot\|_p$ denotes the ℓ^p-norm on $\mathbb{R}^n$; that is, $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for all $x \in \mathbb{R}^n$.

These results suggest that the n-dimensional Gauss space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$ has unexpected geometry when $n \gg 1$.

Interestingly enough, the analogue of (1.12) in the case that $p = \infty$ has a still different form. In this case, one is examining the maximum of n i.i.d. unbounded random variables. Naturally the maximum grows as $n \to \infty$. The following anticipates the rate of growth, which turns out to be only logarithmic. See also Problem 18.

⟨pr:max⟩ **Proposition 1.3.** *Let M_n denote either $\max_{1\leqslant i\leqslant n}|Z_i|$ or $\max_{1\leqslant i\leqslant n} Z_i$. Then,*

$$\mathrm{E}(M_n) = \sqrt{2\log n} + o(1) \qquad \textit{as } n \to \infty.$$

We will see later on (see XXX) that, in part because of Proposition 1.3, there exists a finite constant $c > 0$ such that

$$\mathrm{P}_n\left\{(1-\varepsilon)\sqrt{2\log n} \leqslant M_n \leqslant (1+\varepsilon)\sqrt{2\log n}\right\} \geqslant 1 - 2\mathrm{e}^{-c\varepsilon^2(\log n)^2}, \tag{1.13}$$?CoM:max?

simultaneously for all integers $n \geqslant 2$ and real numbers $\varepsilon \in (0\,,1)$. Thus, the measure P_n concentrates on ℓ^∞-balls of radius $\sqrt{2\log n}$ as $n \to \infty$.

Before we prove Proposition 1.3, let us mention only that it is possible to evaluate $\mathrm{E}(M_n)$ much more precisely than was done in Proposition 1.3; see Problem 19 below. However, Proposition 1.3 is strong enough for our present needs.

Proof of Proposition 1.3. Throughout the proof, define

$$\overline{M}_n := \max_{1\leqslant j\leqslant n}|Z_j| \quad \text{and} \quad \underline{M}_n := \max_{1\leqslant j\leqslant n} Z_j.$$

Since $\underline{M}_n \leqslant M_n \leqslant \overline{M}_n$, it suffices to study $\mathrm{E}(\overline{M}_n)$ for an upper bound and $\mathrm{E}(\underline{M}_n)$ for a lower bound. We begin with the former.

For all $t > 0$, the event $\{\overline{M}_n > t\}$ is equivalent to the event that some $|Z_i|$ exceeds t. Therefore, a simple union bound and Lemma 1.1 together yield a finite constant A such that

$$\mathrm{P}\left\{\overline{M}_n > t\right\} \leqslant \sum_{i=1}^n \mathrm{P}\{|Z_i| > t\} = n\mathrm{P}\{|Z_1| > t\} \leqslant Ant^{-1}\mathrm{e}^{-t^2/2},$$

valid uniformly for all $n \geqslant 1$, and for all sufficiently large $t > 1$. We will use this bound when $n \exp(-t^2/2) < 1$; that is, when $t > \sqrt{2 \log n}$. For smaller values of t, an upper bound of one is frequently a better choice. Thus, we write

$$\begin{aligned} \mathrm{E}\left(\overline{M}_n\right) &= \int_0^\infty \mathrm{P}\left\{\max_{1 \leqslant i \leqslant n} |Z_i| > t\right\} \mathrm{d}t \\ &\leqslant \sqrt{2\log n} + An \int_{\sqrt{2\log n}}^\infty t^{-1} \mathrm{e}^{-t^2/2}\, \mathrm{d}t \\ &= \sqrt{2 \log n} + o(1) \qquad \text{as } n \to \infty. \end{aligned}$$

This proves the upper bound.

The lower bound is also simple to establish. First, let us choose and fix an $\varepsilon > 0$ and then note that

$$\begin{aligned} \mathrm{E}\left(\underline{M}_n\right) &\geqslant \mathrm{E}\left[\underline{M}_n;\, \underline{M}_n > \sqrt{2\log n} - \varepsilon\right] \\ &\geqslant \left(\sqrt{2\log n} - \varepsilon\right) \mathrm{P}\left\{\underline{M}_n > \sqrt{2 \log n} - \varepsilon\right\}. \end{aligned} \tag{1.14}$$

eq:E(M_n):LB

We plan to prove that

$$\sqrt{\log n}\, \mathrm{P}\left\{\underline{M}_n \leqslant \sqrt{2 \log n} - \varepsilon\right\} \to 0 \quad \text{as } n \to \infty. \tag{1.15}$$

goal:LB:M_n

Indeed, (1.15) and (1.14) together imply that, as $n \to \infty$,

$$\begin{aligned} \mathrm{E}\left(\underline{M}_n\right) &\geqslant \left[\sqrt{2\log n} - \varepsilon\right] \left(1 - o\left(\frac{1}{\sqrt{\log n}}\right)\right) \\ &= \sqrt{2\log n} - \varepsilon + o(1). \end{aligned}$$

Since ε is arbitrary, it follows that $\mathrm{E}(\underline{M}_n) \geqslant \sqrt{2 \log n} + o(1)$, which completes the proof. It remains to verify (1.15).

Since $1 - a \leqslant \exp(-a)$ for all $a \in \mathbb{R}$, it follows from independence that

$$\begin{aligned} \mathrm{P}\left\{\underline{M}_n \leqslant \sqrt{2\log n} - \varepsilon\right\} &= \left(1 - \mathrm{P}\left\{Z_1 > \sqrt{2 \log n} - \varepsilon\right\}\right)^n \\ &\leqslant \exp\left(-n\mathrm{P}\left\{Z_1 > \sqrt{2\log n} - \varepsilon\right\}\right). \end{aligned} \tag{1.16}$$

eq:P(M_n):UB

According to Lemma 1.1, as $n \to \infty$,

$$\mathrm{P}\left\{Z_1 > \sqrt{2\log n} - \varepsilon\right\} = \frac{\mathrm{e}^{-\varepsilon^2/2} + o(1)}{2n\sqrt{\pi \log n}} \exp\left(\varepsilon \sqrt{2 \log n}\right).$$

Because $\exp(\varepsilon\sqrt{2\log n})$ grows faster than any given power of $\log n$, the preceding probability must exceed $n^{-1}(\log n)^2$ for all n sufficiently large.[2] In particular, (1.16) implies that, as $n \to \infty$,

$$\mathrm{P}\left\{\underline{M}_n \leqslant \sqrt{2\log n} - \varepsilon\right\} \leqslant \mathrm{e}^{-(\log n)^2} = o\left(\frac{1}{\sqrt{\log n}}\right). \tag{1.17}$$

eq:under:M

[2]Of course, the same sentence continues to hold if we replace $(\log n)^2$ by $(\log n)^p$ for an arbitrary $p > 0$. We need only to choose $p > 1$—here, $p = 2$—in order to ensure the final identity in (1.17).

This verifies (1.15), and completes the proof of the proposition. □

2 The Projective CLT

The following projective central limit theorem is a different way to say that a Gaussian vector in $\mathbb{R}^n$ lies very close to the ℓ^2-sphere of radius $n^{1/2}$.

⟨pr:ProjCLT⟩ **Proposition 2.1.** *Choose and fix an integer $k \geqslant 1$ and a bounded and continuous function $f : \mathbb{R}^k \to \mathbb{R}$. Let μ_n denote the uniform measure on $\sqrt{n}\,\mathbb{S}^{n-1}$. Then, as $n \to \infty$,*

$$\int_{\sqrt{n}\mathbb{S}^{n-1}} f(x_1, \ldots, x_k)\, \mu_n(\mathrm{d}x_1 \cdots \mathrm{d}x_n) \to \int_{\mathbb{R}^k} f(x_1, \ldots, x_k)\, \mathrm{P}_k(\mathrm{d}x_1 \cdots \mathrm{d}x_k).$$

Proposition 2.1 is a rigorous way to say that, when $n \gg 1$, the canonical Gaussian measure on $\mathbb{R}^n$ is very close to the uniform distribution on the ball $\sqrt{n}\,\mathbb{S}^{n-1}$ of radius $n^{1/2}$ in $\mathbb{R}^n$.

Proof. By the weak law of large numbers XXX,

$$\frac{\|Z\|}{\sqrt{n}} = \left[\frac{1}{n}\sum_{j=1}^n Z_j^2\right]^{1/2} \to 1 \qquad \text{in probability as } n \to \infty.$$

Therefore, for every integer $k \geqslant 1$,

$$\frac{\sqrt{n}(Z_1 \,\ldots, Z_k)}{\|Z\|} \Rightarrow (Z_1, \ldots, Z_k) \qquad \text{as } n \to \infty, \tag{1.18}$$ pCLT

where "⇒" denotes weak convergence in $\mathbb{R}^k$. Now the distribution of the random vector $\sqrt{n}\,Z/\|Z\|$ is rotationally invariant—see (1.1)—and supported on $\sqrt{n}\,\mathbb{S}^{n-1}$. Consequently, a classical fact about the uniqueness of Haar measures (see XXX) implies that the uniform measure μ_n coincides with the law of $\sqrt{n}\,Z/\|Z\|$. In other words, the proposition is just a paraphrase of the already-proved assertion (1.18). □

3 Anderson's Shifted-Ball Inequality

One of the defining features of P_n is that it is "unimodal." This property is sometimes called *Anderson's theorem*, which is in fact a theorem of convex analysis; see Anderson XXX. When $n = 1$, "unimodality" refers to the celebrated bell-shaped curve of γ_1, and can be seen for example in Figure 1.1 on page 5. There are similar, also visual, ways to think about "unimodality" in higher dimensions.

Anderson's theorem has many deep applications in probability theory, as well as multivariate statistics, which originally was one of the main motivations for Anderson's work. We will see some of these applications later on. For now we contend ourselves with a statement and proof.

The proof of Anderson's theorem requires some notions from convex analysis, which we develop first.

Recall that a set $E \subset \mathbb{R}^n$ is *convex* if for all $x, y \in E$ the line segment $\overline{xy}$ that joins x and y lies entirely in E. Equivalently put, E is convex if and only if $\lambda x + (1-\lambda)y \in E$ for all $x, y \in E$ and $\lambda \in [0\,,1]$. See Figure 1.2.

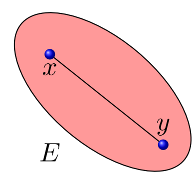

Figure 1.2. A bounded, convex set $E \subset \mathbb{R}^2$: $\overline{xy} \subseteq E \quad \forall x, y \in E$.

⟨fig:Cvx:Set⟩

One can check, using only first principles, that a set $E \subset \mathbb{R}^n$ is convex if and only if

$$E = \lambda E + (1-\lambda)E \qquad \text{for all } \lambda \in [0\,,1], \tag{1.19}$$

eq:EEE

where $\alpha A + \beta B$ denotes the *Minkowski sum* of αA and βB for all $\alpha, \beta \in \mathbb{R}$ and $A, B \subseteq \mathbb{R}^n$. That is,

$$\alpha A + \beta B := \{\alpha x + \beta y : \, x \in A\,, y \in B\}.$$

See Problem 12.

Convex sets are measurable sets as the following result shows.

⟨pr:convpbm:meas⟩ **Proposition 3.1.** *Every convex set $E \subset \mathbb{R}^n$ is Lebesgue measurable.*

Proposition 3.1 will be established en route the proof of Anderson's inequality. In order to state Anderson's inequality, we need to recall two standard definitions.

Definition 3.2. A set $E \in \mathbb{R}^n$ is *symmetric* if $E = -E$.

Definition 3.3. If $f : \mathbb{R}^n \to \mathbb{R}$ is a measurable function, then its *level set at level* $r \in \mathbb{R}$ is defined as $f^{-1}[r\,,\infty) := \{x \in \mathbb{R}^n : \, f(x) \geqslant r\} := \{f \geqslant r\}$. We say f is *symmetric* if $f(x) = f(-x)$ for all $x \in \mathbb{R}^n$, or equivalently if all of its level sets are symmetric.

We can finally state Anderson's theorem.

⟨th:Anderson⟩ **Theorem 3.4** (Anderson's inequality)**.** *Let $f \in L^1(\mathbb{R}^n)$ be a non-negative symmetric function that has convex level sets. Then,*

$$\int_E f(x - \lambda y)\,\mathrm{d}x \geqslant \int_E f(x-y)\,\mathrm{d}x,$$

for all symmetric convex sets $E \subset \mathbb{R}^n$, every $y \in \mathbb{R}^n$, and all $\lambda \in [0\,,1]$.

The proof will take up the remainder of this chapter. For now, let us remark briefly on how the Anderson inequality can be used to analyse the Gaussian measure P_n.

Recall γ_n from (1.1), and note that for every $r > 0$, the level set

$$\gamma_n^{-1}[r\,,\infty) = \left\{ x \in \mathbb{R}^n : \|x\| \leqslant \sqrt{2\log r + n\log(2\pi)} \right\}$$

is a closed, whence convex and symmetric, ball in $\mathbb{R}^n$. Therefore, we can apply Anderson's inequality with $\lambda = 0$ to immediately deduce the "unimodality" of P_n in the sense of the following result.

⟨co:Anderson⟩ **Corollary 3.5** (Unimodality of P_n)**.** *For all symmetric convex sets $E \subset \mathbb{R}^n$, $0 \leqslant \lambda \leqslant 1$, and $y \in \mathbb{R}^n$, $\mathrm{P}_n(E + \lambda y) \geqslant \mathrm{P}_n(E + y)$. In particular,*

$$\sup_{y \in \mathbb{R}^n} \mathrm{P}_n(E + y) = \mathrm{P}_n(E).$$

It is important to emphasize the remarkable fact that Corollary 3.5 is a "dimension-free theorem." Here is a typical consequence: $\mathrm{P}\{\|Z - a\| \leqslant r\}$ is maximized at $a = 0$ for all $r > 0$. For this reason, Corollary 3.5 is sometimes referred to as a "shifted-ball inequality."

One can easily generalize the preceding example with a little extra effort. Let us first note that if M is an $n \times n$ positive-semidefinite matrix, then $E := \{x \in \mathbb{R}^n : x'Mx \leqslant r\}$ is a symmetric convex set for every real number $r > 0$ (it is an ellipsoid).[3] Equivalently, E is the event—in our probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$—that $Z'MZ \leqslant r$. Therefore, Anderson's shifted-ball inequality implies that

$$\mathrm{P}\left\{(Z - \mu)'M(Z - \mu) \leqslant r\right\} \leqslant \mathrm{P}\left\{Z'MZ \leqslant r\right\} \quad \forall\, r > 0 \text{ and } \mu \in \mathbb{R}^n.$$

This inequality has applications in multivariate statistics; see, for example, the final section of Anderson XXX. We will see other interesting examples of Anderson's inequality later on.

The proof of Anderson's inequality takes up the rest of this section and is divided into four parts. The first three parts are self-contained and establish a series of ancillary results. Some readers may wish to accept the statements of the first three parts on faith, and then proceed directly to the proof of Anderson's inequality in the fourth part.

§3.1 Part 1. Measurability of Convex Sets

Here we prove Proposition 3.1. But first let us mention the following example.

Example 3.6. Suppose $n \geqslant 2$ and $E = B(0\,,1) \cup F$, where $B(0\,,1)$ is the usual notation for the Euclidean ball of radius one about $0 \in \mathbb{R}^n$, and $F \subset \partial B(0\,,1)$. The set E is convex, but it is not Borel measurable unless F is. Still, E is always Lebesgue measurable, in this case because F is Lebesgue null in $\mathbb{R}^n$.

[3] We frequently identify the elements of $\mathbb{R}^n$ with *column vectors*. In this way, we see that a quantity such as $x'Mx$ is a scalar for all $x \in \mathbb{R}^n$ and all $n \times n$ matrices M.

This example shows that, in general, one cannot hope to replace the Lebesgue measurability of convex sets by their Borel measurability.

Proof of Proposition 3.1. We will prove the following: *Every **bounded** convex set is measurable.*

This does the job since whenever E is convex and $n \geqslant 1$, $E \cap B(0\,,n)$ is a bounded convex set, which is measurable by the above. Therefore, $E = \cup_{n=1}^{\infty} E \cap B(0\,,n)$ is also measurable.

The closure $\overline{\partial E}$ of ∂E is manifestly closed; therefore, it is measurable. We will prove that $|\overline{\partial E}| = 0$. This shows that the difference between E and the open set E^0 is a subset of a null set, whence E is Lebesgue measurable. There are many proofs of this fact. Here is an elegant one, due to Lang XXX.

Define

$$\mathcal{M} := \left\{ B \in \mathcal{B}(\mathbb{R}^n) : \; \left| B \cap \overline{\partial E} \right| \leqslant \left(1 - 3^{-n}\right) |B| \right\}.$$

Then $\mathcal{M}$ is clearly a monotone class; that is, $\mathcal{M}$ is closed under countable, increasing unions and also closed under countable, decreasing intersections. We plan to prove that every upright *rectangle*, that is every nonempty set of of the form $\prod_{i=1}^{n}(a_i\,, b_i]$, is in $\mathcal{M}$. If this were so, then Sierpiński's monotone class theorem would imply that $\mathcal{M} = \mathcal{B}(\mathbb{R}^n)$. That would show, in turn, that $|\overline{\partial E}| = |\overline{\partial E} \cap \overline{\partial E}| \leqslant (1 - 3^{-n})|\overline{\partial E}|$, which proves the claim.

Choose and fix a *rectangle* $B := \prod_{i=1}^{n}(a_i\,, b_i]$, where $a_i < b_i$ for all $1 \leqslant i \leqslant n$. Subdivide each 1-dimensional interval $(a_i\,, b_i]$ into 3 equal-sized parts: $(a_i\,, a_i + r_i]$, $(a_i + r_i\,, a_i + 2r_i]$, and $(a_i + 2r_i\,, a_i + 3r_i]$ where $r_i := (b_i - a_i)/3$. We can write B as a disjoint union of 3^n equal-sized rectangles, each of which has the form $\prod_{i=1}^{n}(a_i + c_i r_i\,, a_i + (1 + c_i)r_i]$ where $c_i \in \{0\,, 1\,, 2\}$. Call these rectangles $B_1, \ldots, B_{3^n}$.

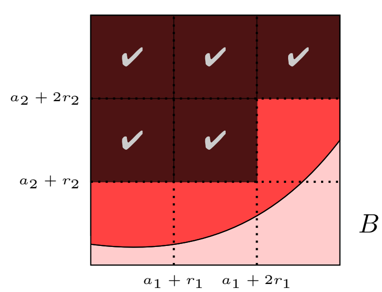

Figure 1.3. A subdivison of B. The region above the curved line belongs to the convex set E. The smaller, darker, checked boxes are those boxes in B that do not intersect ∂E.

⟨fig:Cvx:Set:1⟩

Direct inspection shows that, because E is assumed to be convex, there must exist an integer $1 \leqslant L \leqslant 3^n$ such that $\overline{\partial E} \cap B_L = \varnothing$. For otherwise the middle rectangle $\prod_{i=1}^{n}(a_i + r_i\,, a_i + 2r_i]$ would have to lie entirely in the interior E° and intersect $\overline{\partial E}$ at the same time; this would contradict the existence of a supporting hyperplane at every point of ∂E which is a defining feature of convexity (see Figure 1.3). Let us fix the integer L alluded to here. Since the B_j's are translates of one another they have the same measure. Therefore,

$$\left| B \cap \overline{\partial E} \right| \leqslant \sum_{\substack{1 \leqslant j \leqslant 3^n \\ j \neq L}} |B_j| = |B| - |B_L| = \left(1 - 3^{-n}\right) |B|.$$

This proves that every rectangle B is in $\mathcal{M}$, whence completes the proof. $\square$

§3.2 Part 2. The Brunn–Minkowski Inequality

In this subsection we state and prove the Brunn–Minkowski inequality. That inequality XXX is a classical result from convex analysis, and has profound connections to several other areas of research.

In order to partially motivate what is to come, define

$$B_r := \{x \in \mathbb{R}^n : \|x\| \leqslant r\} \tag{1.20}$$ def:B_r

to be the closed ball of radius $r > 0$ about the origin.

The *ε-enlargement* of a compact set $A \subset \mathbb{R}^n$ is defined as the set $A + B_\varepsilon = A + \varepsilon B_1$; see Figure 1.4. The Brunn–Minkowski inequality is one of the many ways in which we can describe how the volume A relates to the volume of the perturbed set $A + \varepsilon B_1$ when $\varepsilon > 0$ is small. See Problem 14 for a sampler.

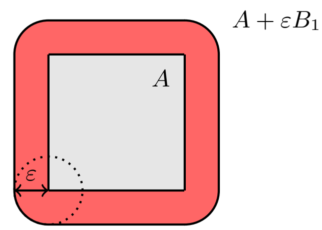

Figure 1.4. The ε-enlargement, $A + \varepsilon B_1$, of the inner square A.

More generally still, one can consider two compact sets $A, B \subset \mathbb{R}^n$ and ask about the relation between the volume of A and the volume of the perturbed set $A + B$. ⟨fig:Cvx:Set:2⟩

It is easy to see that if A and B are compact, then so is $A + B$, since the latter is clearly bounded and closed. In particular, $A + B$ is measurable. The Brunn–Minkowski inequality relates the Lebesgue measure of the Minkowski sum $A + B$ to those of A and B.

⟨th:BrunnMinkowski⟩ **Theorem 3.7** (The Brunn–Minkowski Inequality)**.** *For all compact sets $A, B \subset \mathbb{R}^n$,*

$$|A + B|^{1/n} \geqslant |A|^{1/n} + |B|^{1/n}.$$

We can replace A by αA and B by $(1 - \alpha)B$, where $0 \leqslant \alpha \leqslant 1$, and recast the Brunn–Minkowski inequality in the following equivalent form:

$$|\alpha A + (1 - \alpha)B|^{1/n} \geqslant \alpha|A|^{1/n} + (1 - \alpha)|B|^{1/n}, \tag{1.21}$$ eq:BrunnMink:bis

for all compact sets $A, B \subset \mathbb{R}^n$ and $\alpha \in [0\,, 1]$. Among other things, this formulation suggests the existence of deeper connections to convex analysis because if A and B are convex sets, then so is $\alpha A + (1 - \alpha)B$ for all $\alpha \in [0\,, 1]$. Problem 13 contains a small generalization of (1.21).

Proof. The proof is elementary but tricky. In order to clarify the underlying ideas, we will divide it up into 3 small steps.
Step 1. Say that $K \subset \mathbb{R}^n$ is a *rectangle* when K has the form,

$$K = [x_1\,, x_1 + k_1] \times \cdots \times [x_n\,, x_n + k_n],$$

for some $x := (x_1\,, \ldots, x_n) \in \mathbb{R}^n$ and $k_1, \ldots, k_n > 0$. We refer to the point x as the *lower corner* of K, and $k := (k_1\,, \ldots, k_n)$ as the *length* of K.

In this first step we verify the theorem in the case that A and B are rectangles with respective lengths a and b. In this case, we can see that $A+B$ is an rectangle of side length $a+b$. The Brunn–Minkowski inequality, in this case, follows from the following application of Jensen's inequality [the *arithmetic–geometric mean inequality*]:

$$\left(\prod_{i=1}^n \frac{a_i}{a_i+b_i}\right)^{1/n} + \left(\prod_{i=1}^n \frac{b_i}{a_i+b_i}\right)^{1/n} \leqslant \frac{1}{n}\sum_{i=1}^n \left(\frac{a_i}{a_i+b_i}\right) + \frac{1}{n}\sum_{i=1}^n \left(\frac{b_i}{a_i+b_i}\right)$$
$$= 1.$$

Step 2. Now we consider the case that A and B are interior-disjoint [or "ID"] finite unions of rectangles.

For every compact set K let us write $K^+ := \{x \in K : x_1 \geqslant 0\}$ and $K^- := \{x \in K : x_1 \leqslant 0\}$.

Now we apply a so-called "Hadwiger–Ohmann cut": Notice that if we translate A and/or B, then we do not alter $|A+B|$, $|A|$, or $|B|$. Therefore, after we translate the sets suitably, we can always ensure that: (a) A^+ and B^+ are rectangles; (b) A^- and B^- are ID unions of rectangles; and (c)

$$\frac{|A^+|}{|A|} = \frac{|B^+|}{|B|}.$$

With this choice in mind, we find that

$$|A+B| \geqslant |A^+ + B^+| + |A^- + B^-| \geqslant \left(|A^+|^{1/n} + |B^+|^{1/n}\right)^n + |A^- + B^-|,$$

thanks to Step 1 and the fact that $A^+ + B^+$ is disjoint from $A^- + B^-$. Now,

$$\left(|A^+|^{1/n} + |B^+|^{1/n}\right)^n = |A^+|\left(1 + \frac{|B^+|^{1/n}}{|A^+|^{1/n}}\right)^n = |A^+|\left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n,$$

whence

$$|A+B| \geqslant |A^+|\left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n + |A^- + B^-|.$$

Now split up, after possibly also translating, A^- into $A^{-,\pm}$ and B^- into $B^{-,\pm}$ such that:

1. $A^{-,\pm}$ are interior disjoint;
2. $B^{-,\pm}$ are interior disjoint; and
3. $|A^{-,+}|/|A^-| = |B^{-,+}|/|B^-|$.

Thus, we can apply the preceding to A^- and B^- in place of A and B in order to see that

$$|A+B| \geqslant |A^+|\left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n + |A^{-,+}|\left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n + |A^{-,-} + B^{-,-}|$$
$$= \left(|A^+| + |A^{-,-}|\right)\left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n + |A^{-,-} + B^{-,-}|.$$

And now continue to split and translate $A^{-,-}$ and $B^{-,-}$, etc. In this way we obtain a countable sequence $A_0 := A^+$, $A_1 := A^{-,+}$, $\ldots$, $B_0 := B^+$, $B_1 := B^{-,+}$, $\ldots$ of ID rectangles such that:

1. $\cup_{j=0}^{\infty} B_j = B$ [after translation];
2. $\cup_{j=0}^{\infty} A_j = A$ [after translation]; and
3. $|A+B|$ is bounded below by

$$\sum_{j=0}^{\infty} |A_j| \left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n = |A| \left(1 + \frac{|B|^{1/n}}{|A|^{1/n}}\right)^n = \left(|A|^{1/n} + |B|^{1/n}\right)^n.$$

This proves the result in the case that A and B are ID unions of rectangles.

Step 3. Every compact set can be written as an countable union of ID rectangles. In other words, we can find $A^1, A^2, \ldots$ and $B^1, B^2, \ldots$ such that:

1. Every A^j and B^k is a finite union of ID rectangles;
2. $A^j \subseteq A^{j+1}$ and $B^k \subseteq B^{k+1}$ for all $j, k \geqslant 1$; and
3. $A = \cup_{j=1}^{\infty} A^j$ and $B = \cup_{j=1}^{\infty} B^j$.

By the previous step,

$$|A+B|^{1/n} \geqslant |A^m + B^m|^{1/n} \geqslant |A^m|^{1/n} + |B^m|^{1/n} \qquad \text{for all } m \geqslant 1.$$

Let $m \uparrow \infty$ and appeal to the inner continuity of Lebesgue measure in order to deduce the theorem in its full generality. □

§3.3 Part 3. Change of Variables

In the second part of the proof we develop an elementary fact from integration theory.

Let $A \subseteq \mathbb{R}^n$ be a Borel set, and $g : A \to \mathbb{R}_+$ a Borel-measurable function.

Definition 3.8. The *distribution function* of g is the function $\bar{G} : [0\,,\infty) \to \mathbb{R}_+$, defined as

$$\bar{G}(r) := \left|g^{-1}[r\,,\infty)\right| := |\{x \in A : \, g(x) \geqslant r\}| := |\{g \geqslant r\}| \quad \text{for all } r \geqslant 0.$$

This is standard notation in classical analysis, and should not be mistaken with the closely-related definition of *cumulative distribution functions* in probability and statistics. In any case, the following ought to be familiar.

⟨pr:ChangeofVar⟩ **Proposition 3.9** (Change of Variables Formula). *For every Borel measurable function $F : \mathbb{R}_+ \to \mathbb{R}_+$,*

$$\int_A F(g(x))\,\mathrm{d}x = -\int_0^{\infty} F(r)\,\mathrm{d}\bar{G}(r).$$

If, in addition, A is compact and F is absolutely continuous, then

$$\int_A F'(g(x))\,\mathrm{d}x = \int_0^{\infty} F'(r)\bar{G}(r)\,\mathrm{d}r.$$

Proof. First consider the case that $F = \mathbb{1}_{[a,b)}$ for some $b \geqslant a > 0$. In that case,

$$\begin{aligned}\int_0^\infty F(r)\,\mathrm{d}\bar{G}(r) &= \bar{G}(b-) - \bar{G}(a) \\ &= -\left|\{x \in A:\, a \leqslant g(x) < b\}\right| \\ &= -\int_A F(g(x))\,\mathrm{d}x.\end{aligned}$$

This proves our formula when F is a simple function. By linearity, it holds also when F is an elementary function. The general form of the first assertion of the proposition follows from this and Lebesgue's dominated convergence theorem. The second follows from the first and integration by parts for Stieldjes integrals. □

§3.4 Part 4. The Proof of Anderson's Inequality

Recall f, E, λ, and y from Theorem 3.4. Let us define a new number $\alpha \in [0\,,1]$ by $\alpha := (1+\lambda)/2$. The number α is chosen so that

$$\alpha y + (1-\alpha)(-y) = \lambda y.$$

Since E is convex, we have $E = \alpha E + (1-\alpha)E$. Therefore, the preceding display implies that

$$(E + \lambda y) \supseteq \alpha(E+y) + (1-\alpha)(E-y).$$

And because the intersection of two convex sets is a convex set, we may infer that

$$\begin{aligned}&(E+\lambda y) \cap f^{-1}[r\,,\infty) \\ &\qquad \supseteq \alpha\left[(E+y)\cap f^{-1}[r\,,\infty)\right] + (1-\alpha)\left[(E-y)\cap f^{-1}[r\,,\infty)\right].\end{aligned}$$

Now we apply the Brunn–Minkowski inequality (Theorem 3.7), in the form (1.21), in order to see that

$$\begin{aligned}&\left|(E+\lambda y)\cap f^{-1}[r\,,\infty)\right|^{1/n} \\ &\qquad \geqslant \alpha\left|(E+y)\cap f^{-1}[r\,,\infty)\right|^{1/n} + (1-\alpha)\left|(E-y)\cap f^{-1}[r\,,\infty)\right|^{1/n}.\end{aligned}$$

Since E is symmetric, $E - y = -(E+y)$. Because of this identity and the fact that f has symmetric level sets, it follows that

$$(E-y)\cap f^{-1}[r\,,\infty) = -\left[(E+y)\cap f^{-1}[r\,,\infty)\right].$$

Therefore,

$$\left|(E+y)\cap f^{-1}[r\,,\infty)\right|^{1/n} = \left|(E-y)\cap f^{-1}[r\,,\infty)\right|^{1/n},$$

whence

$$\bar{H}_\lambda(r) := \left|(E+\lambda y)\cap f^{-1}[r\,,\infty)\right| \geqslant \left|(E+y)\cap f^{-1}[r\,,\infty)\right| := \bar{H}_1(r).$$

Now two applications of the change of variables formula [Proposition 3.9] yield the following:

$$\begin{aligned}\int_E f(x+\lambda y)\,\mathrm{d}x - \int_E f(x+y)\,\mathrm{d}x &= \int_{E+\lambda y} f(x)\,\mathrm{d}x - \int_{E+y} f(x)\,\mathrm{d}x \\ &= -\int_0^\infty r\mathrm{d}\bar{H}_\lambda(r) + \int_0^\infty r\,\mathrm{d}\bar{H}_1(r) \\ &= \int_0^\infty \left[\bar{H}_\lambda(r) - \bar{H}_1(r)\right]\mathrm{d}r \geqslant 0.\end{aligned}$$

This completes the proof of Anderson's inequality for $-y$, which completes the proof overall since y was arbitrary. □

4 Gaussian Random Vectors

In the first three sections of this chapter we worked exclusively with the standard Gaussian distribution on $\mathbb{R}^n$, but as most readers are aware, there is an entire family of Gaussian distributions on $\mathbb{R}^n$, indexed by their mean vectors and covariance matrices. Rather than continue this discussion in this way, it turns out to be convenient to begin with a slightly different characterization: A random vector is Gaussian iff all linear combinations of its entries are real Gaussian random variables. For the formal definition let $(\Omega\,, \mathcal{F}, \mathrm{Q})$ be a general probability space, and recall the following.

Definition 4.1. A random n-vector $X = (X_1\,, \ldots, X_n)$ in $(\Omega\,, \mathcal{F}, \mathrm{Q})$ is *Gaussian* if $a'X$ has a normal distribution for every non-random n-vector a.

General theory ensures that we can always assume that $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, and $\mathrm{Q} = \mathrm{P}_n$, which we will do from now on without further mention in order to save on the typography.

If X is a Gaussian random vector in $\mathbb{R}^n$ and $a \in \mathbb{R}^n$ is fixed, then $a'X$ has a one-dimensional Gaussian distribution, and hence has finite moments of all orders. Let μ and Γ respectively denote the mean vector and the covariance matrix of X; that is,

$$\mu_i = \mathrm{E}(X_i), \quad \Gamma_{i,j} = \mathrm{Cov}(X_i\,, X_j),$$

where the expectation and covariance are computed with respect to the measure P. It is often convenient to write this in vector form as

$$\mu = \mathrm{E}(X), \quad \Gamma = \mathrm{E}\left[(X-\mu)(X-\mu)'\right],$$

where we regard X and μ as $n \times 1$ vectors.

It is easy to see that if $X = (X_1\,, \ldots, X_n)$ is Gaussian with mean μ and covariance Γ, then $a'X$ is necessarily distributed as $\mathrm{N}(a'\mu\,, a'\Gamma a)$. In particular, the characteristic function of X is described by

$$\mathrm{E}\left[\mathrm{e}^{ia'X}\right] = \exp\left(ia'\mu - \tfrac{1}{2}a'\Gamma a\right) \qquad \text{for all } a \in \mathbb{R}^n. \tag{1.22}$$ chf:Gauss

Definition 4.2. Let X be a Gaussian random vector in $\mathbb{R}^n$ with mean μ and covariance matrix Γ. The distribution of X is then called a *multivariate normal distribution on* $\mathbb{R}^n$ and is denoted by $\mathrm{N}_n(\mu\,,\Gamma)$.

When Γ is non singular we can invert the Fourier transform to find that the probability density function of X at any point $x \in \mathbb{R}^n$ is the following (see Problem 7):

$$p_X(x) = \frac{1}{(2\pi)^{n/2}\,|\det\Gamma|^{1/2}} \exp\left\{-\tfrac{1}{2}(x-\mu)'\Gamma^{-1}(x-\mu)\right\}. \tag{1.23}$$ pdf:Gauss

The identity (1.1) for γ_1 corresponds to the special case where μ is the zero vector and Γ the identity matrix.

When Γ is singular, the distribution of X is singular with respect to the Lebesgue measure on $\mathbb{R}^n$, and hence does not have a density.

Example 4.3. Suppose $n = 2$ and W has a $\mathrm{N}(0\,,1)$ distribution on the line [which you might recall is denoted by P_1]. Then, the distribution of $X = (W\,,W)$ is concentrated on the diagonal $\{(x\,,x) : x \in \mathbb{R}\}$ of $\mathbb{R}^2$. Since the diagonal has zero Lebesgue measure, it follows that the distribution of X is singular with respect to the Lebesgue measure on $\mathbb{R}^2$.

The following are a series of simple, though useful, facts from elementary probability theory.

⟨lem:G1⟩ **Lemma 4.4.** *If X has a $\mathrm{N}_n(\mu\,,\Gamma)$ distribution, then $AX + b$ is distributed as $\mathrm{N}_m(A\mu + b\,, A\Gamma A')$ for every $b \in \mathbb{R}^m$ and all $m \times n$ matrices A.*

⟨lem:G2⟩ **Lemma 4.5.** *Suppose X has a $\mathrm{N}_n(\mu\,,\Gamma)$ distribution. Choose and fix an integer $1 \leqslant K \leqslant n$, and suppose in addition that $I_1, \ldots, I_K$ are K disjoint subsets of $\{1\,,\ldots,n\}$ such that*

$$\mathrm{Cov}(X_i\,,X_j) = 0 \qquad \textit{whenever } i \textit{ and } j \textit{ lie in distinct } I_\ell\textit{'s.}$$

Then, $\{X_i\}_{i\in I_1}, \ldots, \{X_i\}_{i\in I_K}$ are independent, each having a multivariate normal distribution.

⟨lem:G3⟩ **Lemma 4.6.** *Suppose X has a $\mathrm{N}_n(0\,,\Gamma)$ distribution, where Γ is symmetric and non singular. Then $\Gamma^{-1/2}X$ has the same distribution $\mathrm{N}_n(0\,,I)$ as Z.*

We can frequently use one, or more, of these basic lemmas to study the general Gaussian distribution on $\mathbb{R}^n$ via the canonical Gaussian measure P_n. Here is a typical example.

⟨th:Anderson:Gauss⟩ **Theorem 4.7** (Anderson's Shifted-Ball Inequality)**.** *If X has a $\mathrm{N}_n(0\,,\Gamma)$ distribution and Γ is positive definite, then for all convex symmetric sets $F \subset \mathbb{R}^n$ and $a \in \mathbb{R}^n$,*

$$\mathrm{P}\{X \in a + F\} \leqslant \mathrm{P}\{X \in F\}.$$

Proof. Since $\Gamma^{-1/2}X$ has the same distribution as Z,

$$\mathrm{P}\{X \in a + F\} = \mathrm{P}\left\{Z \in \Gamma^{-1/2}a + \Gamma^{-1/2}F\right\}.$$

Now $\Gamma^{-1/2}F$ is symmetric and convex because F is. Apply Anderson's shifted-ball inequality for P_n [Corollary 3.5] to see that

$$\mathrm{P}\left\{Z \in \Gamma^{-1/2}a + \Gamma^{-1/2}F\right\} \leqslant \mathrm{P}\left\{Z \in \Gamma^{-1/2}F\right\}.$$

This proves the theorem. □

The following comparison theorem is one of the noteworthy corollaries of the preceding theorem.

⟨th:Anderson:Gauss:2⟩ **Corollary 4.8.** *Suppose X and Y are respectively distributed as $\mathrm{N}_n(0\,,\Gamma_X)$ and $\mathrm{N}_n(0\,,\Gamma_Y)$, where $\Gamma_X - \Gamma_Y$ is positive semidefinite. Then,*

$$\mathrm{P}\{X \in F\} \leqslant \mathrm{P}\{Y \in F\},$$

for all symmetric, closed convex sets $F \subset \mathbb{R}^n$.

Proof. First consider the case that Γ_X, Γ_Y, and $\Gamma_X - \Gamma_Y$ are positive definite. Let W be independent of Y and have a $\mathrm{N}_n(0\,,\Gamma_X - \Gamma_Y)$ distribution. The distribution of W has a probability density p_W, and $W + Y$ is distributed as X, whence

$$\mathrm{P}\{X \in F\} = \mathrm{P}\{W + Y \in F\} = \int_{\mathbb{R}^n} \mathrm{P}\{Y \in -a + F\} p_W(a)\,\mathrm{d}a \leqslant \mathrm{P}\{Y \in F\},$$

thanks to Theorem 4.7. This proves the theorem in the case that $\Gamma_X - \Gamma_Y$ is positive definite. If Γ_Y is positive definite but $\Gamma_X - \Gamma_Y$ is only positive semidefinite, then we define for all $0 < \delta < \varepsilon < 1$,

$$X(\varepsilon) := X + \varepsilon U, \quad Y(\delta) := Y + \delta U,$$

where U is independent of $(X\,,Y)$ and has the $\mathrm{N}_n(0\,,I)$ distribution. The respective distributions of $X(\varepsilon)$ and $Y(\delta)$ are $\mathrm{N}_n(0\,,\Gamma_{X(\varepsilon)})$ and $\mathrm{N}_n(0\,,\Gamma_{Y(\delta)})$, where $\Gamma_{X(\varepsilon)} := \Gamma_X + \varepsilon I$ and $\Gamma_{Y(\delta)} := \Gamma_Y + \delta I$. Since $\Gamma_{X(\varepsilon)}$, $\Gamma_{Y(\delta)}$, and $\Gamma_{X(\varepsilon)} - \Gamma_{Y(\delta)}$ are positive definite, the portion of the theorem that has been proved so far implies that $\mathrm{P}\{X(\varepsilon) \in F\} \leqslant \mathrm{P}\{Y(\delta) \in F\}$, for all symmetric convex sets $F \subset \mathbb{R}^n$. Let ε and δ tend down to zero, all the while ensuring that $\delta < \varepsilon$, to deduce the result from the fact that F is closed [$F = \bar{F}$]. □

Example 4.9 (Comparison of Moments). Recall that for $1 \leqslant p \leqslant \infty$, the ℓ^p-norm of $x \in \mathbb{R}^n$ is

$$\|x\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} & \text{if } p < \infty, \\ \max_{1 \leqslant i \leqslant n} |x_i| & \text{if } p = \infty. \end{cases}$$

It is easy to see that all centered ℓ^p-balls of the form $\{x \in \mathbb{R}^n : \|x\|_p \leqslant t\}$ are convex and symmetric. Therefore, it follows immediately from Corollary 4.8 that if $\Gamma_X - \Gamma_Y$ is positive semidefinite, then

$$\mathrm{P}\left\{\|X\|_p > t\right\} \geqslant \mathrm{P}\left\{\|Y\|_p > t\right\} \qquad \text{for all } t > 0 \text{ and } 1 \leqslant p \leqslant \infty.$$

Multiply both sides by rt^{r-1} and integrate both sides $[\mathrm{d}t]$ from $t = 0$ to $t = \infty$ in order to see that

$$\mathrm{E}\left[\|X\|_p^r\right] \geqslant \mathrm{E}\left[\|Y\|_p^r\right] \qquad \text{for } r > 0 \text{ and } 1 \leqslant p \leqslant \infty.$$

These are examples of moment comparison, and can sometimes be useful in estimating expectation functionals of X in terms of expectation functionals of a Gaussian random vector Y with a simpler covariance matrix than that of X. Similarly, $\mathrm{P}\{\|X+a\|_p > t\} \geqslant \mathrm{P}\{\|X\|_p > t\}$ for all $a \in \mathbb{R}^n$, $t > 0$, and $1 \leqslant p \leqslant \infty$ by Theorem 4.7. Therefore,

$$\mathrm{E}\left(\|X\|_p^r\right) = \inf_{a \in \mathbb{R}^n} \mathrm{E}\left(\|X + a\|_p^r\right) \qquad \text{for all } 1 \leqslant p \leqslant \infty \text{ and } r > 0.$$

This is a nontrivial generalization of the familiar fact that when $n = 1$, $\mathrm{Var}(X) = \inf_{a \in \mathbb{R}} \mathrm{E}(|X + a|^2)$.

5 The Isserlis Formula

The *Isserlis formula* XXX, later rediscovered by Wick XXX, is a computation of the joint product moment $m_n := \mathrm{E}(\prod_{i=1}^n X_i)$, where $X = (X_1, \ldots, X_n)$ is distributed as $\mathrm{N}_n(0, \Gamma)$ for an arbitrary $n \times n$ covariance matrix Γ. The answer is found immediately when n is odd: Because X and $-X$ have the same law, it follows from the parity of n that $m_n = -m_n$, whence $m_n = 0$ when n is odd. The Isserlis formula deals with the less elementary case that n is even.

⟨th:Isserlis⟩ **Theorem 5.1** (Isserlis, 1918). *Let X have a $\mathrm{N}_n(0, \Gamma)$ distribution, where $n \geqslant 2$ is an even integer and Γ is an $n \times n$ covariance matrix. Then,*

$$\mathrm{E}\left[\prod_{i=1}^n X_i\right] = \sum_{(i_1, \ldots, i_n)} \prod_{j=1}^{n/2} \Gamma_{i_{2j-1}, i_{2j}}$$

where $\sum_{(i_1, \ldots, i_n)}$ denotes the sum over all perfect matchings $((i_1, i_2), \ldots, (i_{n-1}, i_n))$ of pairs in $\{1, \ldots, n\}$.

Perfect matchings are matchings or pairings of every element in the set with exactly one other element. Clearly a perfect matching exists if and only if the number of elements in the set is even. Note that in using the symbols $((i_1, i_2), \ldots, (i_{n-1}, i_n))$ to denote a perfect matching there are many equivalent symbols that denote the same matching. For example, if $n = 4$ then $((1, 2), (3, 4))$ is equivalent to $((2, 1), (3, 4))$ which is equivalent to $((3, 4), (1, 2))$, etc. Whichever symbol is used the product $\prod \Gamma_{i_{2j-1},i_{2j}}$ is the same. Enumerating the number of non-equivalent symbols can be done by the following scheme. Regard any permutation σ of $\{1,\ldots,n\}$ as an ordered list of the elements, and group the first two elements of the list into a pair, the third and fourth into a pair, etc. Then equivalent symbols are produced by permuting the $n/2$ pairs arbitrarily, and the two elements within each pair can be listed in arbitrary order. Thus the total number of non-equivalent symbols is

$$\sum_{(i_1,\ldots,i_n)} 1 = \frac{n!}{2^{n/2}(n/2)!} \qquad \text{whenever } n \geqslant 2 \text{ is even.} \tag{1.24}$$

partition:count

Furthermore, this enumeration scheme allows us to rewrite the Isserlis formula as

$$\mathrm{E}\left[\prod_{i=1}^{n} X_i\right] = \frac{1}{2^{n/2}(n/2)!} \sum_{\sigma\in\Pi_n} \prod_{j=1}^{n/2} \Gamma_{\sigma(2j-1),\sigma(2j)}, \tag{1.25}$$

Isserlis:bis

where Π_n denotes the collection of all $n!$ permutations of $\{1,\ldots,n\}$.

Note that the Isserlis formula states that the product moment m_n can be fully expressed using only the pairwise covariances $\Gamma_{i,j}$ for $i,j=1,\ldots,n$, which is of course obvious since the covariance matrix uniquely determines the joint distribution of the X_i. The Isserlis formula basically says that the formula for m_n has the simplest possible form that can be described via the $\Gamma_{i,j}$'s. For example, it will follow immediately from the Theorem 5.1 that

$$\begin{aligned}
m_2 = \mathrm{E}(X_1X_2) &= \Gamma_{1,2},\\
m_4 = \mathrm{E}(X_1X_2X_3X_4) &= \Gamma_{1,2}\Gamma_{3,4} + \Gamma_{1,3}\Gamma_{2,4} + \Gamma_{1,4}\Gamma_{2,3},\\
m_6 = \mathrm{E}(X_1X_2X_3X_4X_5X_6) &= \Gamma_{1,2}\Gamma_{3,4}\Gamma_{5,6} + \Gamma_{1,2}\Gamma_{3,5}\Gamma_{2,6} + \Gamma_{1,2}\Gamma_{3,6}\Gamma_{2,4}\\
&\quad + \Gamma_{1,3}\Gamma_{2,4}\Gamma_{5,6} + \Gamma_{1,3}\Gamma_{2,5}\Gamma_{3,6} + \Gamma_{1,3}\Gamma_{2,6}\Gamma_{3,4}\\
&\quad + \Gamma_{1,4}\Gamma_{2,3}\Gamma_{5,6} + \Gamma_{1,4}\Gamma_{2,5}\Gamma_{3,6} + \Gamma_{1,4}\Gamma_{2,6}\Gamma_{3,5}\\
&\quad + \Gamma_{1,5}\Gamma_{2,3}\Gamma_{4,6} + \Gamma_{1,5}\Gamma_{2,3}\Gamma_{4,6} + \Gamma_{1,5}\Gamma_{2,4}\Gamma_{3,6}\\
&\quad + \Gamma_{1,6}\Gamma_{2,3}\Gamma_{4,5} + \Gamma_{1,6}\Gamma_{2,4}\Gamma_{3,5} + \Gamma_{1,6}\Gamma_{2,5}\Gamma_{4,6},
\end{aligned}$$

and so on. The number of individual summands in m_n is, thanks to (1.24), $c_n := n!/\{2^{n/2}(n/2)!\}$ for every even integer $n \geqslant 2$. In particular, we set $X_1 = X_2 = \cdots = X_n := X$ to deduce the well-known fact that if X has a standard normal distribution, then $\mathrm{E}[X^2] = c_2 = 1$, $\mathrm{E}[X^4] = c_4 = 3$, $\mathrm{E}[X^6] = c_6 = 15$, and in general

$$\mathrm{E}[X^n] = c_n = \frac{n!}{2^{n/2}(n/2)!}, \qquad \text{whenever } n \geqslant 2 \text{ is even.}$$

Theorem 5.1 will be proved using moment generating functions. The proof hinges on an elementary computation from multivariate calculus. For that let us introduce some notation.

Define $f_1,\ldots,f_n : \mathbb{R}^n \to \mathbb{R}$ via

$$f_k(t) := (t'\Gamma)_k = \sum_{i=1}^{n} t_i\Gamma_{k,i} \qquad \text{for } k=1,\ldots,n \text{ and } t\in\mathbb{R}^n.$$

⟨lem:partial:exp⟩ **Lemma 5.2.** *Suppose $n \geqslant 2$. Then, for all $t \in \mathbb{R}^n$ and all $1 \leqslant \ell \leqslant n$*

$$\frac{\partial^\ell}{\partial t_1 \cdots \partial t_\ell} \mathrm{e}^{\frac{1}{2}t'\Gamma t} = \mathrm{e}^{\frac{1}{2}t'\Gamma t} \sum_{\substack{\text{partial matchings} \\ \text{of } \{1,\ldots,\ell\}}} \prod_{\substack{(i,j) \\ \text{matched}}} \Gamma_{i,j} \prod_{\substack{k \\ \text{unmatched}}} f_k(t),$$

where a partial matching of $\{1, \ldots, \ell\}$ matches together some of the elements of the set but can also leave some of the elements unmatched.

Proof. First observe that, for all integers $j, k = 1, \ldots, n$ and for every $t \in \mathbb{R}^n$,

$$\frac{\partial}{\partial t_j} f_k(t) = \Gamma_{j,k} = \Gamma_{k,j} \qquad \text{and} \qquad \frac{\partial}{\partial t_k} \mathrm{e}^{\frac{1}{2}t'\Gamma t} = f_k(t)\mathrm{e}^{\frac{1}{2}t'\Gamma t}.$$

With these two formulas in mind, we find that

$$\frac{\partial^2}{\partial t_1 \partial t_2} \mathrm{e}^{\frac{1}{2}t'\Gamma t} = \Gamma_{1,2}\mathrm{e}^{\frac{1}{2}t'\Gamma t} + f_1(t)f_2(t)\mathrm{e}^{\frac{1}{2}t'\Gamma t} = e^{\frac{1}{2}t'\Gamma t}[\Gamma_{1,2} + f_1(t)f_2(t)].$$

Thus the formula is true for the case $\ell = 2$, since the two elements in $\{1, 2\}$ are either matched or unmatched. Now proceed by induction on ℓ, i.e. assume the statement of the lemma is true for some $2 \leqslant \ell < n$. Then differentiate the right hand side with respect to $t_{\ell+1}$ using the product rule. The derivative of the exponential produces an extra factor of $f_{\ell+1}(t)$, which is equivalent to taking all the partial matchings of $\{1, \ldots, l\}$ in the summation and adding in $\ell + 1$ but not matching it to anything. Similarly, the derivative of the summation turns the $f_k(t)$ terms into $\Gamma_{k,\ell+1}$ terms, which is equivalent to taking a partial matching of $\{1, \ldots, \ell\}$, adding in $\ell + 1$, and matching it to the previously unmatched number k. Adding these two components of the integration-by-parts formula produces a summation over all partial matchings of $\{1, \ldots, \ell\}$ with $\ell + 1$ added onto it and then either matched to an unmatched element or left unmatched. This generates all partial matchings of $\{1, \ldots, \ell + 1\}$ in a unique way, which completes the inductive step and hence the proof. □

Once armed with the calculus lemma 5.2 we can easily dispense with the proof of Theorem 5.1.

Proof of Theorem 5.1. Let us write all n-vectors as column vectors. Then, by virtue of definition, $\mathrm{E}[\exp(t'X)] = \exp(\frac{1}{2}t'\Gamma t)$ for all $t \in \mathbb{R}^n$. Therefore, the dominated convergence theorem yields the following for all $t \in \mathbb{R}^n$:

$$\mathrm{E}\left[X_1 \cdots X_n \, \mathrm{e}^{t'X}\right] = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \mathrm{E}\left[\mathrm{e}^{t'X}\right] = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \mathrm{e}^{\frac{1}{2}t'\Gamma t}.$$

Set $t = 0$ in order to see that

$$\mathrm{E}[X_1 \cdots X_n] = \left. \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \mathrm{e}^{\frac{1}{2}t'\Gamma t} \right|_{t=0}.$$

We may now deduce the Isserlis theorem from Lemma 5.2 because, in that lemma $f_k(0) = 0$, so the only terms which contribute are those in which every number is matched. These are exactly the perfect matchings. □

Problems

1. Use Lemma 4.4 to show that if X has a $\mathrm{N}_n(\mu_X, \Gamma_X)$ distribution and Y has a $\mathrm{N}_n(\mu_Y, \Gamma_Y)$ distribution, with (X, Y) jointly Gaussian, then $X+Y$ has a $\mathrm{N}_n(\mu_X+\mu_Y, \Gamma_X+\Gamma_Y+2\Gamma_{X,Y})$ distribution, where $\Gamma_{X,Y}$ is the matrix
$$\Gamma_{X,Y} = \tfrac{1}{2}\,\mathrm{E}[(X-\mu_X)(Y-\mu_Y)' + (Y-\mu_Y)(X-\mu_X)'].$$
Is $\Gamma_{X,Y}$ symmetric and positive semi-definite? Would this result continue to hold if X and Y were not jointly Gaussian?

⟨pbm:chi:2⟩ 2. Prove that $\|Z\|^2 = Z_1^2 + \cdots + Z_n^2$ has a χ_n^2 distribution; that is, show that the probability density function of $\|Z\|^2$ is
$$p(x) = \frac{x^{(n-2)/2}\mathrm{e}^{-x/2}}{2^{n/2}\Gamma(n/2)} \qquad \text{for all } x \geqslant 0,$$
and $p(x)=0$ otherwise. Also verify that the mean and variance of $\|Z\|^2$ are n and $2n$, respectively, while the moment generating function of $\|Z\|$ is described by the formula (1.9).

⟨pbm:Laplace⟩ 3. Use (1.4) to prove that
$$\mathrm{P}\left\{x\in\mathbb{R}^n:\ \|x\|>t\right\} \geqslant \frac{(t^2/2)^{(n-2)/2}\mathrm{e}^{-t^2/2}}{\Gamma(n/2)} \qquad \text{for all } t\geqslant 1 \text{ and } n\geqslant 2.$$
Conclude from this fact that, for all $\varepsilon>0$ and $n\geqslant 2$,
$$\lambda(\varepsilon) := \liminf_{n\to\infty}\frac{1}{n}\log \mathrm{P}\left\{x\in\mathbb{R}^n:\ \|x\|>(1+\varepsilon)n^{1/2}\right\} > 0.$$
You may use, without proof, the following form of *Stirling's formula* for the gamma function: $\Gamma(\nu)\sim(2\pi/\nu)^{1/2}(\nu/\mathrm{e})^{\nu}$ as $\nu\to\infty$; see XXX. Compute $\lambda(\varepsilon)$, and show that "lim inf" is a bona fide limit.

⟨pbm:BM:QV⟩ 4. Recall that a collection of random variables $\{X_t\}_{t\geqslant 0}$ is a *Brownian motion* if for each collection of disjoint intervals $[s_i, t_i]\subset\mathbb{R}_+$, , $i=1,\ldots,n$, the random variables $X_{t_i}-X_{s_i}$ ($i=1,\ldots,n$) are independent, the ith one with a $\mathrm{N}(0, t_i-s_i)$ distribution. In addition, the sample functions $t\mapsto X_t$ are continuous with probability one, and for convenience we may assume that $X_0=0$.
   (a) Choose and fix an unbounded, non decreasing sequence $\{m_n\}_{n=1}^{\infty}$ of positive integers, and define $t_{i,n} := i/m_n$ for $i=0,1,\ldots,m_n$. Compute the mean and variance of the *quadratic variation* process,
   $$V_n := \sum_{i=1}^{m_n}(X_{t_{i,n}} - X_{t_{i-1,n}})^2 \qquad (n\geqslant 1).$$
   (b) Use your answer to preceding part in order to show that $V_n\to 1$ in probability as $n\to\infty$.
   (c) Apply the Borel–Cantelli lemma to prove that $\lim_{n\to\infty}V_n = 1$ almost surely if $\sum_{n=1}^{\infty}(1/m_n)<\infty$.
   (d) Improve the preceding, using Theorem 1.2 in place of Chebyshev's inequality, and deduce the following much stronger theorem, essentially observed first by Dudley XXX: If $\lim_{n\to\infty}(m_n/\log n)=\infty$, then $\lim_{n\to\infty}V_n=1$ almost surely.

5. Let $\{X_t\}_{t\geqslant 0}$ denote a Brownian motion, as in Problem 4. Prove that

$$\mathrm{P}\left\{\sup_{t\in[0,T]}|X_t|\leqslant r\right\}=\sup_{a\in\mathbb{R}}\mathrm{P}\left\{\sup_{t\in[0,T]}|X_t-a|\leqslant r\right\}\qquad\text{for all } r,T>0,$$

and, additionally for all real numbers $p\geqslant 1$,

$$\mathrm{P}\left\{\int_0^T|X_t|^p\,\mathrm{d}t<r\right\}=\sup_{a\in\mathbb{R}}\mathrm{P}\left\{\int_0^T|X_t-a|^p\,\mathrm{d}t<r\right\}.$$

6. Suppose that the random variables $X_0,X_1,\ldots,X_n$ are jointly Gaussian, $\mathrm{E}(X_i)=0$ and $\mathrm{Cov}(X_i\,,X_j)=\varrho(i-j)$ for all $i,j=0,\ldots,n$, where ϱ is a function from $\{-n\,,\ldots,n\}$ to $[-1\,,1]$ such that $\varrho(0)=1$.
   (a) Prove that $\mathrm{E}(X_i\mid X_0)=\kappa_iX_0$ for every $i=1,\ldots,n$. (Hint: Find κ_i such that $X_i-\kappa_iX_0$ and X_0 are independent.)
   (b) Compute $\sigma_i^2:=\mathrm{Var}(X_i\mid X_0)$ for $i=1,\ldots,n$. Is σ_i random?
   (c) Conclude from the previous part that $Y_1,\ldots,Y_n$ are jointly Gaussian, where $Y_i:=X_i-\kappa_iX_0$ for $i=1,\ldots,n$. Compute $\mathrm{E}(Y_i)$ and $\mathrm{Cov}(Y_i\,,Y_j)$ for all $i,j=1,\ldots,n$.
   (d) Prove that, for all $\lambda>0$,

$$\mathrm{P}\left\{\max_{1\leqslant i\leqslant n}|X_i|<\lambda\right\}\leqslant\mathrm{P}\left\{\max_{1\leqslant i\leqslant n}|Y_i|<\lambda\right\}.$$

⟨ex:pdf:chf:Gauss⟩ 7. Suppose $\mu\in\mathbb{R}^n$ and Γ is an $n\times n$, strictly positive definite matrix. Then, prove that the function p_X, defined in (1.23) on page 18, is a probability density function on $\mathbb{R}^n$, whose characteristic function is given in (1.22). In the particular case that $n=2$, $\mu=0$, and

$$\Gamma=\begin{bmatrix}1&\rho\\\rho&1\end{bmatrix}\qquad\text{for some }\rho\in(-1\,,1),$$

show that the expression for the probability density simplifies to the following:

$$p_X(x)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{x_1^2-2\rho x_1x_2+x_2^2}{2(1-\rho^2)}\right)\qquad\text{for }x:=(x_1\,,x_2)\in\mathbb{R}^2.$$

?⟨pbm:P(XY<0)⟩? 8. Suppose, as in Problem 7, that $(X\,,Y)$ has a $\mathrm{N}_2(0\,,\Gamma)$ distribution where $\Gamma_{1,1}=\Gamma_{2,2}=1$ and $\Gamma_{1,2}=\Gamma_{2,1}=\rho$ for a fixed number $\rho\in(-1\,,1)$. Show that

$$\mathrm{P}\{XY<0\}=\frac{1}{2}-\frac{1}{\pi}\arcsin\rho=\frac{1}{\pi}\arccos\rho.$$

9. Suppose that $X=(X_1\,,\ldots,X_n)$ is distributed as $\mathrm{N}_n(\mu\,,\Gamma)$ for some $\mu\in\mathbb{R}^n$ and a covariance matrix $\Gamma\in\mathbb{R}^{n\times n}$.
   (a) Prove that if Γ is non singular, then for every $k\in\{1\,,\ldots,n-1\}$ and $L\in\{k+1\,,\ldots,n\}$ there exist finite constants $c_1,\ldots,c_k$ such that

$$\mathrm{E}(X_L\mid X_1,\ldots,X_k)=\mu_L+c_1X_1+\cdots+c_kX_k\qquad\text{a.s.}$$

   (Hint: Use Lemma 4.6 to reduce the problem to one about i.i.d. Gaussian random variables.)
   (b) Use an approximation argument to remove the restriction on the non-singularity of Γ.

(c) Prove that $\mathrm{Var}(X_L \mid X_1, \ldots, X_k)$ is always non random.
(d) Prove that $\mathrm{Var}(X_L \mid X_1, \ldots, X_k) \leqslant \mathrm{Var}(X_L \mid X_1, \ldots, X_{k-1})$.

10. Verify Lemmas 4.4, 4.5, and 4.6 using Fourier analysis.
11. Improve Theorem 3.4 by demonstrating that $\lambda \mapsto \int_E f(x-\lambda y)\,\mathrm{d}x$ is non increasing on $[0\,,1]$.

⟨pbm:EEE⟩ 12. Verify that a set $E \subset \mathbb{R}^n$ is convex if and only if it satisfies (1.19).

⟨pbm:Brunn⟩ 13. Prove that $|\alpha A + \beta B|^{1/n} \geqslant \alpha|A|^{1/n} + \beta|B|^{1/n}$ for all $\alpha, \beta \in \mathbb{R}$ and all compact sets $A, B \subset \mathbb{R}^n$.

⟨pbm:Minkowski⟩ 14. Recall (1.20). Minkowski XXX has defined the *surface area* $|\partial A|$ of set $A \subset \mathbb{R}^n$ as

$$|\partial A| := \lim_{\varepsilon \downarrow 0} \frac{|A+\varepsilon B_1| - |A|}{\varepsilon} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}|A+\varepsilon B_1|\,\bigg|_{\varepsilon=0},$$

and proved that the limit exists whenever $A \subset \mathbb{R}^n$ is compact and convex. Moreover, the limit agrees with the usual notion of surface area. You may use these facts without proof in the sequel.

(a) Prove that, because $B_r = rB_1$, we have $|B_r| = r^n|B_1|$ and $|\partial B_r| = r^{n-1}|\partial B_1|$ for every $r > 0$. (Hint: Start by proving that $|rK| = r^n|K|$ for every closed and bounded set $K \subset \mathbb{R}^n$.)
(b) Integrate in spherical coordinates to justify the following:

$$1 = \int_{\mathbb{R}^n} \gamma_n(x)\,\mathrm{d}x = \int_0^\infty \mathrm{d}r \int_{\partial B_r} \mathrm{d}\sigma\; \frac{\mathrm{e}^{-r^2/2}}{(2\pi)^{n/2}};$$

whence deduce a formula for $|\partial B_1|$ in terms of the gamma function.
(c) Prove that $|\partial B_1| = n|B_1|$.
(d) Use the Brunn–Minkowski inequality and the previous parts of the problem in order to derive the *isoperimetric inequality for convex bodies*: If A is a compact, convex set and $|A| = |B_1|$, then $|\partial A| \geqslant |\partial B_1|$. In words, prove that balls have minimum surface area among all convex bodies of a given volume.

?⟨pbm:E(X_n):general⟩? 15. Suppose that $\{Y_j\}_{j=1}^n$ is an arbitrary sequence of standard normal random variables on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$ for every $n \geqslant 1$. Verify that, for every real number $p \geqslant 1$,

$$\mathrm{E}\left(\max_{1\leqslant j\leqslant n} |Y_j|^p\right) \leqslant (2\log n)^{p/2} + o(1) \qquad \text{as } n \to \infty.$$

16. Verify (1.24), and use it to prove that (1.25) is an equivalent formulation of Theorem 5.1.
17. Use Lemma 5.2 to find an alternative proof of the fact that if X has a $\mathrm{N}_n(0\,,\Gamma)$ distribution and n is odd, then $\mathrm{E}(X_1 \cdots X_n) = 0$.

The following problems are nontrivial variations of Proposition 1.3, and should not be missed. From here on, M_n denotes either $\max_{1\leqslant j\leqslant n} |Z_j|$ or $\max_{1\leqslant j\leqslant n} Z_j$ for every integer $n \geqslant 1$.

⟨pbm:1:M_n⟩ 18. Define

$$c = \begin{cases} 1 & \text{if } M_n = \max_{1\leqslant j\leqslant n} Z_j,\\ 2 & \text{if } M_n = \max_{1\leqslant j\leqslant n} |Z_j|.\end{cases} \tag{1.26}$$ $\boxed{\mathtt{c}}$

Then, use Lemma 1.1 to prove that, as $n \to \infty$,

$$M_n^2 - 2\log n + \log\log n + 2\log\left(\frac{c}{2\sqrt{\pi}}\right) \Rightarrow -2\log\mathcal{E},$$

where "$\Rightarrow$" denotes convergence in distribution, and $\mathcal{E}$ has a mean-one exponential distribution. Use this to prove that, as $n \to \infty$,

$$a_1\sqrt{\log n}\left(M_n - \sqrt{2\log n}\right) + a_2\log\log n + a_3 \Rightarrow -\log\mathcal{E}, \tag{1.27}$$?eq:2:M_n?

where a_1, a_2, and a_3 are numerical constants. Compute these constants.

⟨pbm:2:M_n⟩ 19. (Problem 18, continued) Check that $\log\mathcal{E}$ has finite moments of all orders. Then do the following:

(a) Prove that, as $n \to \infty$,

$$\mathrm{E}(M_n) = \sqrt{2\log n} + \frac{b_1\log\log n}{\sqrt{\log n}} + \frac{b_2}{\sqrt{\log n}} + o\left(\frac{1}{\sqrt{\log n}}\right),$$

for numerical constants b_1 and b_2, which you should also calculate in terms of the constant c – see (1.26) – and the moments of $\log\mathcal{E}$.

(b) Prove that

$$\mathrm{Var}(M_n) \sim \frac{\lambda}{\log n} \qquad \text{as } n \to \infty,$$

where λ is a numerical constant. Compute λ in terms of c and the moments of $\mathcal{E}$.

(c) Conclude that $\mathrm{E}(|M_n - \sqrt{2\log n}|^2)$ convergence to zero as $n \to \infty$, and estimate its exact rate of convergence.

Chapter 2

Calculus in Gauss Space

⟨ch:Calc_on_Gauss_Space⟩

In this section we develop the basics of calculus on the finite-dimensional Gauss space. The differences between this calculus and the "regular" calculus that we first learn (which we call calculus on Lebesgue space) are not that stark. At the end of the day we still compute integrals and derivatives in the same way, but there are some modifications that must be taken into account. The most important of these is the *integration-by-parts* formula, which must be modified to properly accomodate for the Gaussian background measure. On a computational level this modification is elementary. But we shall see that it has far-reaching consequences.

1 The Gradient Operator

The *n-dimensional Lebesgue space* is the measurable space $(\mathbb{E}^n, \mathcal{B}(\mathbb{E}^n))$—where $\mathbb{E} = [0\,,1)$ or $\mathbb{E} = \mathbb{R}$—endowed with the Lebesgue measure, and the "calculus of functions" on Lebesgue space is just "real and harmonic analysis."

The n-dimensional *Gauss space* is the same measure space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ as in the previous paragraph, but is now endowed with the Gauss measure P_n in place of the Lebesgue measure. Since the Gauss space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$ is a probability space, we can—and frequently will—think of a measurable function $f : \mathbb{R}^n \to \mathbb{R}$ as a random variable. Therefore,

$$\begin{aligned}
\mathrm{P}\{f \in A\} &= \mathrm{P}_n\{f \in A\} = \mathrm{P}_n\{x \in \mathbb{R}^n : f(x) \in A\},\\
\mathrm{E}(f) &= \mathrm{E}_n(f) = \int f \,\mathrm{d}\mathrm{P}_n = \int f \,\mathrm{d}\mathrm{P},\\
\mathrm{Cov}(f\,,g) &= \langle f\,,g\rangle_{L^2(\mathrm{P})} = \int fg \,\mathrm{d}\mathrm{P},
\end{aligned}$$

etc. Note, also, that $f = f(Z)$ for all random variables f, where Z is the standard normal random vector $Z(x) := x$ for all $x \in \mathbb{R}^n$, as before. In particular,

$$\mathrm{E}(f) = \mathrm{E}_n(f) = \mathrm{E}[f(Z)],$$

$$\mathrm{Var}(f) = \mathrm{Var}[f(Z)], \quad \mathrm{Cov}(f\,,g) = \mathrm{Cov}[f(Z)\,,g(Z)], \ldots$$

and so on, notation being typically obvious from context.

Let $\partial_j := \partial/\partial x_j$ for all $1 \leqslant j \leqslant n$. From now on we will use the following.

Definition 1.1. Let $C_0^k(\mathrm{P}_n)$ denote the collection of all infinitely-differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f and all of its mixed derivatives of order $\leqslant k$ grow more slowly than $[\gamma_n(x)]^{-\varepsilon}$ for every $\varepsilon > 0$. We also define

$$C_0^\infty(\mathrm{P}_n) := \bigcap_{k=1}^{\infty} C_0^k(\mathrm{P}_n).$$

It is not hard to see that $f \in C_0^k(\mathrm{P}_n)$ if and only if for every $\epsilon > 0$

$$\lim_{\|x\|\to\infty} \mathrm{e}^{-\varepsilon\|x\|^2}|f(x)| = \lim_{\|x\|\to\infty} \mathrm{e}^{-\varepsilon\|x\|^2}\,|(\partial_{i_1}\cdots\partial_{i_m}f)(x)| = 0,$$

for all $1 \leqslant i_1, \ldots, i_m \leqslant n$ and $1 \leqslant m \leqslant k$ (see Problem 4).

We will frequently use the following result without explicit mention.

⟨lem:Ck_moments⟩ **Lemma 1.2.** *If $f \in C_0^k(\mathrm{P}_n)$, then*

$$\mathrm{E}\left(|f|^p\right) < \infty \quad \textit{and} \quad \mathrm{E}\left(|\partial_{i_1}\cdots\partial_{i_m}f|^p\right) < \infty,$$

for all $1 \leqslant p < \infty$, $1 \leqslant i_1, \ldots, i_m \leqslant n$, and $1 \leqslant m \leqslant k$.

The proof is relegated to Problem 1.

For every $f \in C_0^1(\mathrm{P}_n)$, define

$$\begin{aligned}\|f\|_{1,2}^2 :=& \int |f(x)|^2\,\mathrm{P}_n(\mathrm{d}x) + \int \|(\nabla f)(x)\|^2\,\mathrm{P}_n(\mathrm{d}x)\\ =& \ \mathrm{E}\left(|f|^2\right) + \mathrm{E}\left(\|\nabla f\|^2\right),\end{aligned}$$

where $\nabla := (\partial_1, \ldots, \partial_n)$ denotes the gradient operator. Notice that $\|\cdot\|_{1,2}$ is a *bona fide* Hilbertian norm on $C_0^1(\mathrm{P}_n)$ with Hilbertian inner product

$$\begin{aligned}\langle f\,, g\rangle_{1,2} :=& \int fg\,\mathrm{d}\mathrm{P}_n + \int (\nabla f)\cdot(\nabla g)\,\mathrm{d}\mathrm{P}_n\\ =& \ \mathrm{E}[fg] + \mathrm{E}[\nabla f\cdot\nabla g].\end{aligned}$$

We will soon see that $C_0^1(\mathrm{P}_n)$ is not a Hilbert space with the preceding norm and inner product because it is not complete; that is, there are Cauchy sequences in $C_0^1(\mathrm{P}_n)$ that fail to converge in $C_0^1(\mathrm{P}_n)$. Thus, we are led to the following.

Definition 1.3. The *Gaussian Sobolev space* $\mathbb{D}^{1,2}(\mathrm{P}_n)$ is the completion of $C_0^1(\mathrm{P}_n)$ in the norm $\|\cdot\|_{1,2}$.

In order to understand what the elements of $\mathbb{D}^{1,2}(\mathrm{P}_n)$ look like, let us consider a function $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$. By definition, we can find a sequence $f_1, f_2, \ldots \in C_0^1(\mathrm{P}_n)$ such that $\|f_\ell - f\|_{1,2} \to 0$ as $\ell \to \infty$. Since $L^2(\mathrm{P}_n)$ is complete, we can deduce also that

$$D_j f := \lim_{\ell\to\infty} \partial_j f_\ell \qquad \text{exists in } L^2(\mathrm{P}_n) \text{ for every } 1 \leqslant j \leqslant n.$$

It follows, by virtue of construction, that

$$Df = \nabla f \qquad \text{for all } f \in C_0^1(\mathrm{P}_n).$$

Therefore, D is an extension of the gradient operator from $C_0^1(\mathrm{P}_n)$ to $\mathbb{D}^{1,2}(\mathrm{P}_n)$. From now on, we will almost always write Df in favor of ∇f when $f \in C_0^1(\mathrm{P}_n)$. This is because Df can make sense even when f is not in $C_0^1(\mathrm{P}_n)$, as we will see in the next few examples.

In general, we can think of elements of $\mathbb{D}^{1,2}(\mathrm{P}_n)$ as functions in $L^2(\mathrm{P}_n)$ that have one weak derivative in $L^2(\mathrm{P}_n)$. We may refer to the linear operator D as the *Malliavin derivative*, and the random variable Df as the [generalized] *gradient of* f. We will formalize this notation further at the end of this section. For now, let us note instead that the standard Sobolev space $W^{1,2}(\mathbb{R}^n)$ is obtained in exactly the same way as $\mathbb{D}^{1,2}(\mathrm{P}_n)$ was, but the Lebesgue measure is used in place of P_n everywhere. Since $\gamma_n(x) = \mathrm{dP}_n(x)/\mathrm{d}x < 1$,[1] it follows that the Hilbert space $\mathbb{D}^{1,2}(\mathrm{P}_n)$ is richer than the Hilbert space $W^{1,2}(\mathbb{R}^n)$, whence the Malliavin derivative is an extension of Sobolev's [generalized] gradient. The extension is strict; see Problem 6.

It is a natural time to produce examples to show that the space $\mathbb{D}^{1,2}(\mathrm{P}_n)$ is strictly larger than the space $C_0^1(\mathrm{P}_n)$ endowed with the norm $\|\cdot\|_{1,2}$.

⟨ex:Smoothing:1⟩ *Example* 1.4 ($n = 1$). Consider the case $n = 1$ and let f denote the "tent function," $f(x) := (1 - |x|)_+$ on $\mathbb{R}$. We claim that $f \in \mathbb{D}^{1,2}(\mathrm{P}_1) \setminus C_0^1(\mathrm{P}_1)$. Moreover, we claim

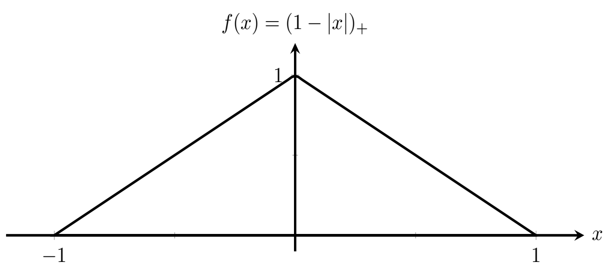

Figure 2.1. A tent function.

?⟨fig:tent⟩?

the P_1-a.s. identity,[2]

$$(Df)(x) = -\mathrm{sign}(x)\mathbb{1}_{[-1,1]}(x).$$

In a sense, this formula is obvious. We propose to derive it rigorously, thus emphasizing the fact that Df should be regarded as an element of $L^2(\mathrm{P}_n)$.

Let $\psi_1 \in C^\infty(\mathbb{R})$ be a symmetric probability density function on $\mathbb{R}$ such that $\psi_1 \equiv$ a positive constant on $[-1\,,1]$, and $\psi_1 \equiv 0$ off $[-2\,,2]$. For every real number $r > 0$, define $\psi_r(x) := r\psi_1(rx)$ and $f_r(x) := (f * \psi_r)(x)$. Then $\sup_x |f_N(x) - f(x)| \to 0$ as $N \to \infty$ because f is uniformly continuous. In particular, $\|f_N - f\|_{L^2(\mathrm{P}_n)} \to 0$ as $N \to \infty$. To complete the proof it remains to verify that

$$\lim_{N\to\infty} \int |f_N'(x) + \mathrm{sign}(x)\mathbb{1}_{[-1,1]}(x)|^2\, \mathrm{P}_n(\mathrm{d}x) = 0. \tag{2.1}$$ goal:n=1

[1]In other words, $\mathrm{E}(|f|^2) < \int_{\mathbb{R}^n} |f(x)|^2\,\mathrm{d}x$ for all $f \in L^2(\mathbb{R}^n)$ that are strictly positive on a set of positive Lebesgue measure.

[2]It might help to recall that Df is defined as an element of the Hilbert space $L^2(\mathrm{P}_1)$ in this case. Therefore, it does not make sense to try to compute $(Df)(x)$ for all $x \in \mathbb{R}$. This issue arises when one constructs any random variable on any probability space, of course. Also, note that P_1-a.s. equality is the same thing as Lebesgue-a.e. equality, since the two measures are mutually absolutely continuous.

By the dominated convergence theorem and integration by parts,

$$\begin{aligned} f_N'(x) &= \int_{-\infty}^{\infty} f(y)\psi_N'(x-y)\,\mathrm{d}y \\ &= -\int_0^1 \psi_N(x-y)\,\mathrm{d}y + \int_{-1}^0 \psi_N(x-y)\,\mathrm{d}y \\ &:= \; -A_N(x) + B_N(x). \end{aligned}$$

We now prove that $A_N \to \mathbb{1}_{[0,1]}$ as $N \to \infty$ in $L^2(\mathrm{P}_1)$; a small adaptation of this argument will also prove that $B_N \to \mathbb{1}_{[-1,0]}$ in $L^2(\mathrm{P}_1)$, from which (2.1) ensues.

By a change of variables, $A_N(x) = \int_{N(x-1)}^{Nx} \psi_1(y)\,\mathrm{d}y$. Because ψ_1 is a probability density function, it follows that $A_N(x) \to \mathbb{1}_{[0,1]}(x)$ as $N \to \infty$ for P_1-almost all x. Similarly $B_n(x) \to \mathbb{1}_{[-1,0]}(x)$ for P_1-almost all x, and therefore $f_N'(x) = -A_N(x) + B_N(x) \to -\mathrm{sign}(x)\mathbb{1}_{[-1,1]}(x)$ for P_1-almost all x. Since $f_N'(x) - \mathrm{sign}(x)\mathbb{1}_{[-1,1]}(x)$ is bounded uniformly by 2, the dominated convergence theorem implies that the convergence also takes place in $L^2(\mathrm{P}_1)$. This concludes our example.

⟨ex:Smoothing:2⟩ *Example* 1.5 ($n \geqslant 2$). Let us consider the case that $n \geqslant 2$. In order to produce a function $F \in \mathbb{D}^{1,2}(\mathrm{P}_n) \setminus C_0^1(\mathrm{P}_n)$ we use the construction of the previous example and set

$$F(x) := \prod_{j=1}^n f(x_j) \text{ and } \Psi_N(x) := \prod_{j=1}^n \psi_N(x_j) \quad \text{for all } x \in \mathbb{R}^n \text{ and } N \geqslant 1.$$

Then the calculations of Example 1.4 also imply that $F_N := F * \Psi_N \to F$ as $N \to \infty$ in the norm $\|\cdots\|_{1,2}$ of $\mathbb{D}^{1,2}(\mathrm{P}_n)$, $F_N \in C_0^1(\mathrm{P}_n)$, and $F \notin C_0^1(\mathrm{P}_n)$. Thus, it follows that $F \in \mathbb{D}^{1,2}(\mathrm{P}_n) \setminus C_0^1(\mathrm{P}_n)$. Furthermore,

$$(D_jF)(x) = -\mathrm{sign}(x_j)\mathbb{1}_{[-1,1]}(x_j) \times \prod_{\substack{1\leqslant \ell \leqslant n\\ \ell \neq j}} f(x_\ell),$$

for every $1 \leqslant j \leqslant n$ and P_n-almost every $x \in \mathbb{R}^n$.

⟨ex:Lipschitz:D12⟩ *Example* 1.6. The previous two examples are particular cases of a more general family of examples. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz continuous* if there exists a finite constant K such that

$$|f(x) - f(y)| \leqslant K\|x - y\| \qquad \text{for all } x, y \in \mathbb{R}^n.$$

The smallest such constant K is called the *Lipschitz constant of* f and is denoted by $\mathrm{Lip}(f)$; that is,

$$\mathrm{Lip}(f) := \sup_{x\neq y} \frac{|f(x) - f(y)|}{\|x - y\|}.$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz function. According to Rademacher's theorem XXX, f is almost everywhere [equivalently, P_n-a.s.] differentiable and $\|(\nabla f)(x)\| \leqslant \mathrm{Lip}(f)$ a.s. Also note that

$$|f(x)| \leqslant |f(0)| + \mathrm{Lip}(f)\|x\| \qquad \text{for all } x \in \mathbb{R}^n.$$

In particular, $\mathrm{E}(|f|^k) < \infty$ for all $k \geqslant 1$. A density argument, similar to the one that appeared in the preceding examples, shows that $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and

$$\|(Df)(x)\| \leqslant \mathrm{Lip}(f) \qquad \mathrm{P}_1\text{-almost all } x.$$

We will appeal to this fact several times in this book.

The generalized gradient D follows more or less the same general set of rules as does the more usual gradient operator ∇. And it frequently behaves as one expect it should even when it is understood as the Gaussian extension of ∇; see Examples 1.4 and 1.5, for instance. The following ought to reinforce this point of view.

⟨lem:ChainRule⟩ **Lemma 1.7** (Chain Rule). *For all $\psi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ and $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$,*

$$D(\psi \circ f) = [(D\psi) \circ f]\, D(f) \qquad a.s.$$

Proof. If f and ψ are smooth functions, then the chain rule of calculus ensures that $[\partial_j(\psi \circ f)](x) = \psi'(f(x))(\partial_j f)(x)$ for all $x \in \mathbb{R}^n$ and $1 \leqslant j \leqslant n$. That is,

$$D(g \circ f) = \nabla(\psi \circ f) = (\psi' \circ f)(\nabla f) = (D\psi)(f)D(f),$$

where $D\psi$ refers to the one-dimensional Malliavin derivative of ψ and $D(f) := Df$ refers to the n-dimensional Malliavin derivative of f. The general case follows from the smooth case and a density argument. □

Here is a final example that is worthy of mention.

⟨ex:DM⟩ *Example* 1.8. Let $M := \max_{1 \leqslant j \leqslant n} Z_j$ and note that

$$M(x) = \max_{1 \leqslant j \leqslant n} x_j = \sum_{j=1}^{n} x_j \mathbb{1}_{Q(j)}(x) \qquad \text{for } \mathrm{P}_n\text{-almost all } x \in \mathbb{R}^n,$$

where $Q(j)$ denotes the cone of all points $x \in \mathbb{R}^n$ such that $x_j \geqslant \max_{i \neq j} x_i$. We can approximate the indicator function of $Q(j)$ by a smooth function to see that $M \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $D_j M = \mathbb{1}_{Q(j)}$ a.s. for all $1 \leqslant j \leqslant n$. Let

$$J(x) := \arg\max(x).$$

Clearly, $J(x)$ is defined uniquely for P_n-almost every $x \in \mathbb{R}^n$. For all other values of x, redefine $J(x) := 0$ to be concrete. Our computation of $D_j M$ equivalently yields

$$(DM)(x) = \mathbf{e}_{J(x)} \qquad \text{for } \mathrm{P}_n\text{-almost all } x \in \mathbb{R}^n, \tag{2.2}$$ eq:DM

where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denote the standard basis of $\mathbb{R}^n$.

Let us end this section by introducing a little more notation.

The preceding discussion constructs, for every function $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$, the Malliavin derivative Df as an $\mathbb{R}^n$-valued function with coordinates in $L^2(\mathrm{P}_n)$. We will use the following natural notations exchangeably:

$$(Df)(x\,, j) := [(Df)(x)]_j = (D_j f)(x),$$

for every $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$, $x \in \mathbb{R}^n$, and $1 \leqslant j \leqslant n$. In this way we may also think of Df as a scalar-valued element of the real Hilbert space $L^2(\mathrm{P}_n \times \chi_n)$, where

Definition 1.9. χ_n always denotes the counting measure on $\{1\,, \ldots, n\}$.

We see also that the inner product on $\mathbb{D}^{1,2}(\mathrm{P}_n)$ is

$$\begin{aligned}\langle f\,, g\rangle_{1,2} &= \langle f\,, g\rangle_{L^2(\mathrm{P}_n)} + \langle Df\,, Dg\rangle_{L^2(\mathrm{P}_n \times \chi_n)} \\ &= \mathrm{E}(fg) + \mathrm{E}\,(Df \cdot Dg) \qquad \text{for all } f, g \in \mathbb{D}^{1,2}(\mathrm{P}_n).\end{aligned}$$

Definition 1.10. The random variable $Df \in L^2(\mathrm{P}_n \times \chi_n)$ is called the *Malliavin derivative* of the random variable $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$.

2 Higher-Order Derivatives

One can define higher-order weak derivatives just as easily as we obtained the directional weak derivatives.

Choose and fix $f \in C^2(\mathbb{R}^n)$ and two integers $1 \leqslant i, j \leqslant n$. The *mixed derivative* of f in direction $(i\,, j)$ is the function $x \mapsto (\partial^2_{i,j} f)(x)$, where

$$\partial^2_{i,j} f := \partial_i \partial_j f = \partial_j \partial_i f.$$

The *Hessian operator* ∇^2 is defined as

$$\nabla^2 := \begin{pmatrix} \partial^2_{1,1} & \cdots & \partial^2_{1,n} \\ \vdots & \ddots & \vdots \\ \partial^2_{n,1} & \cdots & \partial^2_{n,n} \end{pmatrix}.$$

With this in mind, we can define a Hilbertian inner product $\langle \cdot\,, \cdot \rangle_{2,2}$ via

$$\begin{aligned} \langle f\,, g\rangle_{2,2} &:= \int fg \, \mathrm{dP}_n + \int (\nabla f)\cdot(\nabla g)\, \mathrm{P}_n(\mathrm{d}x) + \int \mathrm{tr}\left[(\nabla^2 f)(\nabla^2 g)\right] \mathrm{dP}_n \\ &= \int f(x)g(x)\, \mathrm{P}_n(\mathrm{d}x) + \sum_{i=1}^n \int (\partial_i f)(x)(\partial_i g)(x)\, \mathrm{P}_n(\mathrm{d}x) \\ &\qquad\qquad + \sum_{i,j=1}^n \int (\partial^2_{i,j} f)(x)(\partial^2_{i,j} g)(x)\, \mathrm{P}_n(\mathrm{d}x) \\ &= \langle f\,, g\rangle_{1,2} + \int (\nabla^2 f)\cdot(\nabla^2 g)\, \mathrm{dP}_n \\ &= \mathrm{E}(fg) + \mathrm{E}\left[\nabla f \cdot \nabla g\right] + \mathrm{E}\left[\nabla^2 f \cdot \nabla^2 g\right] \qquad [f, g \in C^2_0(\mathrm{P}_n)], \end{aligned}$$

where $K \cdot M$ denotes the matrix—or Hilbert–Schmidt—inner product,

$$K \cdot M := \sum_{i,j=1}^n K_{i,j} M_{i,j} = \mathrm{tr}(K'M),$$

for all $n \times n$ matrices K and M.

We also obtain the corresponding Hilbertian norm $\|\cdot\|_{2,2}$ where:

$$\begin{aligned} \|f\|^2_{2,2} &= \|f\|^2_{L^2(\mathrm{P}_n)} + \sum_{i=1}^n \|\partial_i f\|^2_{L^2(\mathrm{P}_n)} + \sum_{i,j=1}^n \left\|\partial^2_{i,j} f\right\|^2_{L^2(\mathrm{P}_n)} \\ &= \|f\|^2_{1,2} + \left\|\nabla^2 f\right\|^2_{L^2(\mathrm{P}_n \times \chi^2_n)} \\ &= \mathrm{E}\left(f^2\right) + \mathrm{E}\left(\|\nabla f\|^2\right) + \mathrm{E}\left(\|\nabla^2 f\|^2\right) \qquad [f \in C^2_0(\mathrm{P}_n)]; \end{aligned}$$

$\chi^2_n := \chi_n \times \chi_n$ denotes the counting measure on $\{1\,, \cdots, n\}^2$; and

$$\|K\| := \sqrt{K \cdot K} = \sqrt{\sum_{i,j=1}^n K^2_{i,j}} = \sqrt{\mathrm{tr}(K'K)}$$

denotes the Hilbert–Schmidt norm of any $n \times n$ matrix K.

Definition 2.1. The *Gaussian Sobolev space* $\mathbb{D}^{2,2}(\mathrm{P}_n)$ is the completion of $C^2_0(\mathrm{P}_n)$ in the norm $\|\cdot\|_{2,2}$.

For every $f \in \mathbb{D}^{2,2}(\mathrm{P}_n)$ we can find functions $f_1, f_2, \ldots \in C_0^2(\mathrm{P}_n)$ such that $\|f_\ell - f\|_{2,2} \to 0$ as $\ell \to \infty$ Then $D_i f$ and $D_{i,j}^2 f := \lim_{\ell\to\infty} \partial_{i,j}^2 f$ exist in $L^2(\mathrm{P}_n)$ for every $1 \leqslant i, j \leqslant n$. Equivalently, $Df = \lim_{\ell\to\infty} \nabla f$ exists in $L^2(\mathrm{P}_n \times \chi_n)$ and $D^2 f = \lim_{\ell\to\infty} \nabla^2 f$ exists in $L^2(\mathrm{P}_n \times \chi_n^2)$.

Now we extend the definition to derivatives of order greater than two. Choose and fix an integer $k \geqslant 2$, and as a convenient shorthand introduce the notation $[n] = \{1, 2, \ldots, n\}$. As in the usual calculus the kth derivative is described by a k-tensor, which we recall is a function $K : [n]^k \to \mathbb{R}$, where $[n]^k$ is the set of vectors $(q_1, \ldots, q_k)$ of k integers in $\{1, 2, \ldots, n\}$. The 2-tensors are simply $n \times n$ matrices and the higher order ones are their natural generalizations. If $q = (q_1, \ldots, q_k) \in [n]^k$, then let

$$(\partial_q^k f)(x) := (\partial_{q_1} \cdots \partial_{q_k} f)(x) \qquad [f \in C^k(\mathbb{R}^n),\ x \in \mathbb{R}^n].$$

Thus ∂_q^k takes k successive derivatives of f in the order specified by the directions $(q_1, \ldots, q_k)$. For example, if $n = 3$ and $k = 5$ with $q = (3, 1, 2, 1, 2)$ then ∂_q^k computes the derivative of f in the direction of the second coordinate, then the first, then the second again, the first again, and finally the third. By equality of mixed partial derivatives all that matters is the number of times the derivative is taken in each direction, not the order in which they are taken.

Let ∇^k denote the formal k-tensor whose q-th coordinate is ∂_q^k. We define a Hilbertian inner product $\langle \cdot\,, \cdot \rangle_{k,2}$ inductively via

$$\langle f\,, g \rangle_{k,2} = \langle f\,, g \rangle_{k-1,2} + \int (\nabla^k f) \cdot (\nabla^k g)\, \mathrm{d}\mathrm{P}_n,$$

for all $f, g \in C_0^k(\mathrm{P}_n)$, where "$\cdot$" denotes the Hilbert–Schmidt inner product for k-tensors:

$$K \cdot M := \sum_{q \in [n]^k} K_q M_q,$$

for all k-tensors K and M. The corresponding norm is defined via $\|f\|_{k,2} := \langle f\,, f \rangle_{k,2}^{1/2}$.

Definition 2.2. The *Gaussian Sobolev space* $\mathbb{D}^{k,2}(\mathrm{P}_n)$ is the completion of $C_0^k(\mathrm{P}_n)$ in the norm $\|\cdot\|_{k,2}$. We also define $\mathbb{D}^{\infty,2}(\mathrm{P}_n) := \cap_{k\geqslant 1} \mathbb{D}^{k,2}(\mathrm{P}_n)$.

If $f \in \mathbb{D}^{k,2}(\mathrm{P}_n)$ then we can find a sequence of functions $f_1, f_2, \ldots \in C_0^k(\mathrm{P}_n)$ such that $\|f_\ell - f\|_{k,2} \to 0$ as $\ell \to \infty$. It then follows that

$$D^j f := \lim_{\ell\to\infty} \nabla^j f_\ell \quad \text{exists in } L^2(\mathrm{P}_n \times \chi_n^j),$$

for every $1 \leqslant j \leqslant k$, where $\chi_n^j := \chi_n \times \cdots \times \chi_n$ [$j - 1$ times] denotes the counting measure on $\{1\,, \ldots, n\}^j$. The operator D^k is called the kth *Malliavin derivative.*

It is easy to see that the Gaussian Sobolev spaces are nested; that is,

$$\mathbb{D}^{k,2}(\mathrm{P}_n) \subset \mathbb{D}^{k-1,2}(\mathrm{P}_n) \qquad \text{for all } 2 \leqslant k \leqslant \infty.$$

Also, whenever $f \in C_0^k(\mathrm{P}_n)$, the kth Malliavin derivative of f is just the classically-defined derivative $\nabla^k f$, which is a k-tensor. By equality of mixed partial derivatives $\nabla^k f$ is in fact a symmetric k-tensor, which for a general k-tensor K means that $K_q = K_{\sigma(q)}$ where σ is any permutation of $\{1, \ldots, k\}$ and $\sigma(q) = (q_{\sigma(1)}, \ldots, q_{\sigma(k)})$. That is, K_q depends only on the number of times each element in $[n]$ appears in the qth-coordinate, not the order of the elements in q. Because every polynomial in n

variables[3] is in $C_0^\infty(\mathrm{P}_n)$, it follows immediately that $\mathbb{D}^{\infty,2}(\mathbb{R}^n)$ contains all n-variable polynomials; and that all Malliavin derivatives acts as one might expect them to. This last fact will be important for the *Wiener chaos decomposition*, which is a way to write a fairly generic random variable as an infinite sum of polynomials, much like a Taylor series does. If the required sum converges properly then the last fact says that the Malliavin derivative acts on it as we expect it should.

More generally, we have the following.

?⟨def:D:k,p⟩? **Definition 2.3.** For every integer $k \geqslant 1$ and real $p \geqslant 1$, the Gaussian Sobolev spaces $\mathbb{D}^{k,p}(\mathrm{P}_n)$ is defined as the completion of the space $C_0^\infty(\mathrm{P}_n)$ in the norm

$$\|f\|_{\mathbb{D}^{k,p}(\mathrm{P}_n)} := \|f\|_{k,p} := \left[\|f\|^p_{L^p(\mathrm{P}_n)} + \sum_{j=1}^k \|D^j f\|^p_{L^p(\mathrm{P}_n\times\chi_n^j)}\right]^{1/p}.$$

Each $\mathbb{D}^{k,p}(\mathrm{P}_n)$ is a Banach space in the preceding norm. Note that, as usual, these norms are not induced by an inner product unless $p = 2$. Furthermore, for each fixed k the spaces $\mathbb{D}^{k,p}$ are non-increasing in p.

3 The Adjoint Operator

Recall the canonical Gaussian probability density function $\gamma_n := \mathrm{dP}_n/\mathrm{d}x$ from (1.1). Since $(D_j\gamma_n)(x) = -x_j\gamma_n(x)$, we can apply integration by parts and the product rule to see that for every $f, g \in C_0^1(\mathrm{P}_n)$,

$$\begin{aligned}\mathrm{E}\left[D_j(f)g\right] &= \int_{\mathbb{R}^n} (D_jf)(x)g(x)\gamma_n(x)\,dx\\ &= -\int_{\mathbb{R}^n} f(x)D_j\left[g(x)\gamma_n(x)\right]\mathrm{d}x\\ &= -\int_{\mathbb{R}^n} f(x)(D_jg)(x)\,\mathrm{P}_n(\mathrm{d}x) + \int_{\mathbb{R}^n} f(x)g(x)x_j\,\mathrm{P}_n(\mathrm{d}x),\end{aligned}$$

for $1 \leqslant j \leqslant n$. Using the $L^2(\mathrm{P}_n)$ inner product notation we can rewrite the latter identity as the "adjoint relation,"

$$\mathrm{E}\left[D_j(f)g\right] = \langle D_jf\,, g\rangle_{L^2(\mathrm{P}_n)} = \langle f\,, A_jg\rangle_{L^2(\mathrm{P}_n)} = \mathrm{E}\left[fA_j(g)\right], \tag{2.3}$$ IbP

where A is the formal adjoint of D; that is,

$$(Ag)(x) := -(Dg)(x) + xg(x). \tag{2.4}$$ A:g

Note that $g : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function, but $Ag : \mathbb{R}^n \to \mathbb{R}^n$, and

$$(A_jg)(x) = -(D_jg)(x) + x_jg(x).$$

Furthermore, (2.4) is defined pointwise whenever $g \in C_0^1(\mathrm{P}_n)$, but it also makes sense as an identity in $L^2(\mathrm{P}_n \times \chi_n)$ if, for example, $g \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $x \mapsto xg(x)$ is in $L^2(\mathrm{P}_n \times \chi_n)$.

[3] A function $f : \mathbb{R}^n \to \mathbb{R}$ is a polynomial in n variables if it can be written as a linear combination of monomials $x_1^{d_1}\cdots x_n^{d_n}$, where each d_j is a non-negative integer. The *degree* of each monomial is the sum of the exponents appearing in it, and the degree of the polynomial is the maximum degree of all monomials appearing in it. Thus, for example $g(x) = x_1x_2^3 - 2x_5$ is a polynomial of degree 4 in 5 variables.

Let us pause to emphasize that (2.3) can be stated equivalently as

$$\mathrm{E}[gD(f)] = \mathrm{E}[fA(g)], \tag{2.5}$$ D:delta

as n-vectors.[4]

If $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$, then we can always find functions $f_1, f_2, \ldots \in C_0^1(\mathrm{P}_n)$ such that $\|f_\ell - f\|_{1,2} \to 0$ as $\ell \to \infty$. Note that

$$\begin{aligned}\left\|\int gDf_\ell \,\mathrm{dP}_n - \int gDf \,\mathrm{dP}_n\right\| &\leqslant \|g\|_{L^2(\mathrm{P}_n)}\|Df_\ell - Df\|_{L^2(\mathrm{P}_n\times\chi_n)} \\ &\leqslant \|g\|_{L^2(\mathrm{P}_n)}\|f_\ell - f\|_{1,2} \to 0,\end{aligned} \tag{2.6}$$ DfDf

as $\ell \to \infty$. Also,

$$\begin{aligned}\left\|\int f_\ell Ag \,\mathrm{dP}_n - \int fAg \,\mathrm{dP}_n\right\| &\leqslant \|Ag\|_{L^2(\mathrm{P}_n\times\chi_n)}\|f_\ell - f\|_{L^2(\mathrm{P}_n)} \\ &\leqslant \|Ag\|_{L^2(\mathrm{P}_n\times\chi_n)}\|f_\ell - f\|_{1,2} \to 0,\end{aligned} \tag{2.7}$$ fDgfDg

whenever $g \in C_0^1(\mathrm{P}_n)$. We can therefore combine (2.5), (2.6), and (2.7) in order to see that (2.5) in fact holds for all $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $g \in C_0^1(\mathrm{P}_n)$.

Finally define

$$\mathrm{Dom}[A] := \left\{g \in \mathbb{D}^{1,2}(\mathrm{P}_n) : \; Ag \in L^2(\mathrm{P}_n \times \chi_n)\right\}. \tag{2.8}$$?Dom:A?

Since $C_0^1(\mathrm{P}_n)$ is dense in $L^2(\mathrm{P}_n)$, we may infer from (2.5) and another density argument the following.

?⟨pr:adjoint⟩? **Proposition 3.1.** *The adjoint relation* (2.5) *is valid for all* $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ *and* $g \in \mathrm{Dom}[A]$.

Definition 3.2. The linear operator A is the *adjoint operator*, and $\mathrm{Dom}[A]$ is called the *domain of the definition*—or just *domain*—of A.

The linear space $\mathrm{Dom}[A]$ has a number of nicely-behaved subspaces. The following records an example of such a subspace.

⟨pr:Subspace⟩ **Proposition 3.3.** *For every* $2 < p \leqslant \infty$,

$$\mathbb{D}^{1,2}(\mathrm{P}_n) \cap L^p(\mathrm{P}_n) \subset \mathrm{Dom}[A].$$

Proof. We apply Hölder's inequality to see that

$$\mathrm{E}\left(\|Z\|^2[g(Z)]^2\right) = \int \|x\|^2[g(x)]^2\,\mathrm{P}_n(\mathrm{d}x) \leqslant c_p\|g\|^2_{L^p(\mathrm{P}_n)},$$

where

$$c_p = \left[\mathrm{E}\left(\|Z\|^{2p/(p-2)}\right)\right]^{(p-2)/p} < \infty.$$

Therefore, $Zg(Z) \in L^2(\mathrm{P}_n \times \chi_n)$, and we may apply (2.4) to find that

$$\|Ag\|_{L^2(\mathrm{P}_n\times\chi_n)} \leqslant \|Dg\|_{L^2(\mathrm{P}_n\times\chi_n)} + c_p^{1/2}\|g\|_{L^p(\mathrm{P}_n)} \leqslant \|g\|_{1,2} + c_p^{1/2}\|g\|_{L^p(\mathrm{P}_n)} < \infty.$$

This proves that $g \in \mathrm{Dom}[A]$. □

[4] If $\zeta = (\zeta_1, \ldots, \zeta_m)$ is a random m-vector then $\mathrm{E}(\zeta)$ is the m-vector whose jth coordinate is $\mathrm{E}(\zeta_j)$.

Problems

1. ⟨pbm:Ck_moments⟩ Prove Lemma 1.2.
2. For which values of $s \in \mathbb{R}$ is $m(s) := \mathrm{E}(\|Z\|^s)$ finite? When it is finite, compute $\mathrm{E}(\|Z\|^s)$ in terms of the gamma function. These constants arose earlier during the course of the proof of Proposition 3.3.
3. Let $Z : \mathbb{R}^n \to \mathbb{R}^n$ denote the usual vector of independent, standard normal random variables, and define $X := \|MZ\|^s$, where M is a nonrandom $n \times n$ matrix and $s > 0$ is a non random real number. For which values of s and k is $X \in \mathbb{D}^{k,2}(\mathbb{R}^n)$?
4. ⟨pbm:C^k_0⟩ Prove that $f \in C_0^k(\mathrm{P}_n)$ if and only if f is infinitely differentiable in all of its variables, and
$$\lim_{\|x\|\to\infty} \mathrm{e}^{-\varepsilon\|x\|^2}|f(x)| = \lim_{\|x\|\to\infty} \mathrm{e}^{-\varepsilon\|x\|^2}\left|(\partial_{i_1}\cdots\partial_{i_m}f)(x)\right| = 0,$$
for all $1 \leqslant i_1, \ldots, i_m \leqslant n$ and $1 \leqslant m \leqslant k$.
5. Show directly from integration by parts that the standard Laplace operator
$$\Delta := D \cdot D = \sum_{i=1}^n \partial^2_{i,i}$$
is not self-adjoint on $L^2(\mathrm{P}_n)$, even though it is self-adjoint on the Lebesgue space $L^2(\mathbb{R}^n)$. What is the adjoint of Δ on $L^2(\mathrm{P}_n)$?
6. ⟨pbm:Malliavin:Sobolev⟩ Let $C_c^\infty(\mathbb{R}^n)$ denote the collection of all infinitely-differentiable functions of compact support from $\mathbb{R}^n$ to $\mathbb{R}$, and recall that the Sobolev space $W^{1,2}(\mathbb{R}^n)$ is the completion of $C_c^\infty(\mathbb{R}^n)$ in the norm $\|f\|_{L^2(\mathbb{R}^n)} + \|\nabla f\|_{L^2(\mathbb{R}^n)}$ for every $f \in C_c^\infty(\mathbb{R}^n)$. Construct an element of $\mathbb{D}^{1,2}(\mathrm{P}_n)$ that is not an element of $W^{1,2}(\mathbb{R}^n)$.
7. Fill in the details of the derivation of the identity (2.2).
8. ⟨pbm:Dmax||⟩ Define $N := \max_{1\leqslant j\leqslant n}|Z_j|$. Prove that $N \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and evaluate DN.
9. ⟨pbm:MG:transform⟩ Suppose $n \geqslant 2$ is an integer. A stochastic process $X_1, \ldots, X_n$ is said to be *adapted* if X_i is measurable with respect to the σ-algebra generated by $Z_1, \ldots, Z_i$ for every $i = 1, \ldots, n$. Given an adapted process X, define a new stochastic process M – a so-called *martingale transform* of Z – as follows:
$$M_0 := 0 \quad \text{and} \quad M_k := \sum_{i=2}^k X_{i-1}Z_i \quad \text{for } k = 2, \ldots, n.$$
Suppose $X_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ for every $i = 2, \ldots, n$.
   (a) Prove that M is a mean-zero martingale and $M_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ for every $i = 1, \ldots, n$.
   (b) Compute DM_i for all $i = 1, \ldots, n$.
10. (The "*Divergence Operator*" δ) If $G : \mathbb{R}^n \to \mathbb{R}^n$, then define $\delta G := A \cdot G = \sum_{i=1}^n A_i G_i$, when possible. Notice that $\delta D = -\mathcal{L}$.
   (a) Verify that if $G_1, \ldots, G_n \in C_0^1(\mathrm{P}_n)$, then
$$(\delta G)(Z) = -(\operatorname{div} G)(Z) + Z \cdot G(Z) \qquad \mathrm{P}_n\text{-a.s.}$$
   (b) Define $\mathrm{Dom}[\delta]$ to be the collection of all $G : \mathbb{R}^n \to \mathbb{R}^n$ such that $G_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ for all $i = 1, \ldots, n$ and $Z \cdot G(Z) \in L^2(\mathrm{P}_n)$. Prove the following *integration by parts formula*,
$$\mathrm{E}\left[G \cdot (Df)\right] = \mathrm{E}\left[\delta(G)f\right] \qquad \text{for all } f \in \mathbb{D}^{1,2}(\mathrm{P}_n) \text{ and } G \in \mathrm{Dom}[\delta].$$

(c) We say that $G : \mathbb{R}^n \to \mathbb{R}^n$ is *adapted* when G_1 is a constant and $G_i(x)$ depends only on $(x_1, \ldots, x_{i-1})$ for all $i = 2, \ldots, n$. [See also Problem 9.] Prove that, if in addition $G \in \mathrm{Dom}[\delta]$, then $\delta(G)$ is the "discrete Itô integral," $\delta(G) = G \cdot Z$. The random variable $\delta(F)$ is sometimes called the *Skorohod integral* of $F \in \mathrm{Dom}[\delta]$. And you have just shown that the Skorohod integral of an adapted process is the same as its Itô integral.

(d) Prove that

$$D_i(\delta G) = \delta(D_i G) + G_i \qquad \text{for all } i = 1, \ldots, n \text{ and } G \in \mathrm{Dom}[\delta],$$

where $D_i(F) = (D_i F_1, \ldots, D_i F_n)$ whenever $F : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $F_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ for all $i = 1, \ldots, n$.

(e) Use the preceding formula to prove that $\mathrm{E}[\delta(F)] = 0$ and

$$\mathrm{E}\left[\delta(F)\delta(G)\right] = \mathrm{E}[FG] + \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}\left[D_j(F_i) \cdot D_i(G_j)\right],$$

for all $F, G \in \mathrm{Dom}[\delta]$.

Chapter 3

Harmonic Analysis

⟨ch:Harmonic_Analysis⟩

Recall that if $f \in L^2([0\,,2\pi]^n)$, then we can write f as

$$f(x) = (2\pi)^{-n} \sum_{k\in\mathbb{Z}^n} \mathrm{e}^{-ik\cdot x} \hat{f}_k, \tag{3.1}$$ F:L

where $\hat{f}_k := \int_{[0,2\pi]^n} \mathrm{e}^{ik\cdot x} f(x)\,\mathrm{d}x$ denotes the "kth Fourier coefficient" of f, and convergence holds in $L^2([0\,,2\pi]^n)$; that is, $\int_{[0,2\pi]^n} |f(x) - f_N(x)|^2 \mathrm{d}x \to 0$ as $N \to \infty$, where $f_N(x) := \sum_{\|k\|\leqslant N} \exp\{ik\cdot x\}\hat{f}_k$.

Eq. (3.1) is one of the many possible starting points of the theory of harmonic analysis in the Lebesgue space $[0\,,2\pi]^n$. In this chapter we develop a parallel theory for the Gauss space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$. Problems 17 through 22 work out the analogous details for "Poisson" spaces. And other distributional spaces are also possible; see XXX for more discussion on this topic.

1 Hermite Polynomials in Dimension One

Before we discuss the general n-dimensional case, let us consider the special case that $n = 1$. We may observe the following elementary computations:

$$\gamma_1'(x) = -x\gamma_1(x), \quad \gamma_1''(x) = (x^2-1)\gamma_1(x), \quad \gamma_1'''(x) = -(x^3-3x)\gamma_1(x), \quad \text{etc.}$$

It follows from these computations, and from induction, that the kth derivative of γ_1 satisfies

$$\gamma_1^{(k)}(x) = (-1)^k H_k(x)\gamma_1(x) \qquad [k \geqslant 0,\ x \in \mathbb{R}], \tag{3.2}$$ def:Hermite

where H_k is a polynomial of degree at most k. Moreover,

$$\begin{gathered} H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \\ H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3, \quad \text{etc.} \end{gathered} \tag{3.3}$$ eq:H1-4

Definition 1.1. H_k is called the *Hermite polynomial* of degree $k \geqslant 0$.

Be warned that some authors normalize their Hermite polynomials differently than has been done here. Therefore, it might help to remember that our Hermite polynomials are monic; this will be explained shortly.

The following basic lemma records some of the salient features of Hermite polynomials.

⟨lem:Hermite⟩ **Lemma 1.2.** *For all $x \in \mathbb{R}$ and $k \in \mathbb{Z}_+$:*

⟨lem:Hermite:1⟩ *1.* $H_{k+1}(x) = xH_k(x) - H_k'(x)$*;*

⟨lem:Hermite:2⟩ *2.* $H_{k+1}'(x) = (k+1)H_k(x)$*; and*

⟨lem:Hermite:3⟩ *3.* $H_k(-x) = (-1)^k H_k(x)$*.*

This simple lemma teaches us a great deal about Hermite polynomials. For instance, we learn from part 1 and induction that

$$H_k \text{ is a polynomial of exact degree } k \text{ for every } k \geqslant 0,$$

and the following *Rodriguez formula* [or *reproduction formula*] holds:

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x) \qquad \text{for all } k \geqslant 0 \text{ and } x \in \mathbb{R}. \tag{3.4}$$ Rodriguez

Moreover, every H_k is *monic*; that is, the coefficient of x^k in $H_k(x)$ is one for all $k \geqslant 0$.

To observe another interesting property of Hermite polynomials, let us first recall the adjoint operator A from (2.4) on page 34. Presently, $n = 1$; therefore, in this case, A maps a scalar function to a scalar function. Then we can notice that, since polynomials are in the domain of the definition of A [Chapter 2, Proposition 3.3], parts 1 and 2 of Lemma 1.2 respectively say that:

$$H_{k+1} = AH_k \quad \text{and} \quad DH_{k+1} = (k+1)H_k \qquad \text{for all } k \geqslant 0. \tag{3.5}$$ A:D:H

In other words, we can remember the above as saying that H_k plays the same role in the Gauss space $(\mathbb{R}\,, \mathcal{B}(\mathbb{R})\,, \mathrm{P}_1)$ as does the monomial $x \mapsto x^k$ in the Lebesgue space: $DH_{k+1} = (k+1)H_k$ is the Gaussian analogue of the statement that $\mathrm{d}(x^{k+1})/\mathrm{d}x = (k+1)x^k$. As it turns out the adjoint operator behaves a little like an integral operator, and the identity $AH_k = H_{k+1}$ is the Gaussian analogue of the anti-derivative identity $\int x^k\,\mathrm{d}x \propto x^{k+1}$, valid in Lebesgue space.

Other properties of Hermite polynomials will unfold themselves in due time. For the time being, let us prove Lemma 1.2.

Proof. We prove part 1 of the lemma by direct computation:

$$\begin{aligned}
(-1)^{k+1} H_{k+1}(x)\gamma_1(x) &= \gamma_1^{(k+1)}(x) && \text{[by (3.2)]}\\
&= \frac{\mathrm{d}}{\mathrm{d}x}\gamma_1^{(k)}(x)\\
&= (-1)^k \frac{\mathrm{d}}{\mathrm{d}x}\left[H_k(x)\gamma_1(x)\right] && \text{[by (3.2)]}\\
&= (-1)^k \left[H_k'(x)\gamma_1(x) + H_k(x)\gamma_1'(x)\right]\\
&= (-1)^k \left[H_k'(x) - xH_k(x)\right]\gamma_1(x),
\end{aligned}$$

where the last line follows from a third appeal to (3.2), together with the fact that $H_1(x) = x$. Divide both sides by $(-1)^{k+1}\gamma_1(x)$ to complete the proof of part 1.

Part 2 is clearly correct when $k = 0$. We now apply induction: Suppose $H_{j+1}'(x) = (j+1)H_j(x)$ for all $0 \leqslant j \leqslant k$. We plan to prove this for $j = k+1$. By part 1 and the induction hypothesis, the Rodriguez formula (3.4) holds. Therefore, we can differentiate the latter formula in order to find that

$$\begin{aligned}
H_{k+1}'(x) &= H_k(x) + xH_k'(x) - kH_{k-1}'(x)\\
&= H_k(x) + kxH_{k-1}(x) - kH_{k-1}'(x),
\end{aligned}$$

thanks to a second appeal to the induction hypothesis. Because of Part 1, $xH_{k-1}(x)-H'_{k-1}(x)=H_k(x)$. This proves that $H'_{k+1}(x)=(k+1)H_k(x)$, and part 2 follows.

We apply parts 1 and 2 of the lemma, and induction, in order to see that H_k is odd [and H'_k is even] if and only if k is. This proves part 3. □

The following is the *raison d'être* for our study of Hermite polynomials. Specifically, it states that the sequence $\{H_k\}_{k=0}^{\infty}$ plays the same sort of harmonic-analyatic role in the 1-dimensional Gauss space $(\mathbb{R}\,,\mathcal{B}(\mathbb{R})\,,\mathrm{P}_1)$ as do the complex exponentials in Lebesgue spaces.

⟨th:Hermite:1⟩ **Theorem 1.3.** *The normalized Hermite polynomials $\{H_k/\sqrt{k!}\}_{k=0}^{\infty}$ form a complete, orthonormal basis for $L^2(\mathrm{P}_1)$.*

Before we prove Theorem 1.3 let us mention the following corollary.

⟨co:Hermite:1⟩ **Corollary 1.4.** *For every $f\in L^2(\mathrm{P}_1)$,*

$$f=f(Z)=\sum_{k=0}^{\infty}\frac{1}{k!}\,\langle f\,,H_k\rangle_{L^2(\mathrm{P}_1)}\,H_k(Z)=\sum_{k=0}^{\infty}\frac{1}{k!}\,\mathrm{E}\,[fH_k]\,H_k \qquad a.s.$$

To prove this we merely apply Theorem 1.3 and the Riesz–Fischer theorem. Next is another corollary which also has a probabilistic flavor.

⟨co:Hermite:Wiener:1⟩ **Corollary 1.5** (Wiener XXX)**.** *For all $f,g\in L^2(\mathrm{P}_1)$,*

$$\mathrm{E}[fg]=\sum_{k=0}^{\infty}\frac{1}{k!}\,\mathrm{E}[fH_k]\,\mathrm{E}[gH_k] \quad \textit{and} \quad \mathrm{Cov}(f\,,g)=\sum_{k=1}^{\infty}\frac{1}{k!}\,\mathrm{E}[fH_k]\,\mathrm{E}[gH_k].$$

Proof. Multiply both sides of the first identity of Corollary 1.4 by $g(x)$ and integrate $[\mathrm{dP}_1]$ in order to obtain the identity,

$$\langle g\,,f\rangle_{L^2(\mathrm{P}_1)}=\sum_{k=0}^{\infty}\frac{1}{k!}\,\langle f\,,H_k\rangle_{L^2(\mathrm{P}_1)}\,\langle g\,,H_k\rangle_{L^2(\mathrm{P}_1)}.$$

The exchange of sums and integrals is justified by Fubini's theorem. The preceding is another way to say the first result. The second follows from the first and the fact that $H_0\equiv 1$. □

We now prove Theorem 1.3.

Proof of Theorem 1.3. Thanks to (3.5) and the fact that A is the adjoint to D,

$$\mathrm{E}(H_k^2)=\mathrm{E}\,[H_k\cdot A(H_{k-1})]=\mathrm{E}\,[D(H_k)\cdot H_{k-1}]=k\,\mathrm{E}\left[H_{k-1}^2\right].$$

Since $\mathrm{E}(H_0^2)=1$, induction shows that $\mathrm{E}(H_k^2)=k!$ for all integers $k\geqslant 0$.

Next we prove that

$$\mathrm{E}(H_kH_{k+\ell})=0 \qquad \text{for integers } \ell>0, k\geqslant 0.$$

By (3.5),

$$\mathrm{E}(H_kH_{k+\ell})=\mathrm{E}\,[H_k\cdot A(H_{k+\ell-1})]=\mathrm{E}\,[D(H_k)H_{k+\ell-1}]=k\,\mathrm{E}\,[H_{k-1}H_{k+\ell-1}]\,.$$

Now iterate this identity to find that

$$\mathrm{E}\left(H_k H_{k+\ell}\right) = k!\,\mathrm{E}\left[H_0 H_\ell\right] = k! \int_{-\infty}^{\infty} H_\ell(x)\gamma_1(x)\,\mathrm{d}x = 0,$$

since $H_\ell\,\gamma_1 = (-1)^\ell \gamma_1^{(\ell)}$, thanks to (3.2). It follows that $\{H_k/\sqrt{k!}\}_{k=0}^\infty$ is an orthonormal sequence of elements of $L^2(\mathrm{P}_1)$.

In order to complete the proof, we need to show the orthonormal basis is complete. We do this in a standard way. Namely, we suppose that $f \in L^2(\mathrm{P}_1)$ is orthogonal in $L^2(\mathrm{P}_1)$ to H_k for all $k \geqslant 0$, and then proceed to prove that, as a consequence, $f = 0$ almost surely $[\mathrm{P}_1]$.

Part 1 of Lemma 1.2 shows that $H_k(x) = x^k - p(x)$ where p is a polynomial of degree $k-1$ for every $k \geqslant 1$. Consequently, the span of $H_0, \ldots, H_k$ is the same as the span of the monomials $1, x, \cdots, x^k$ for all $k \geqslant 0$. In particular, $\int_{-\infty}^\infty f(x)x^k\gamma_1(x)\,\mathrm{d}x = 0$ for all $k \geqslant 0$. Multiply both sides by $(-it)^k/k!$ and add over all $k \geqslant 0$ in order to see that

$$\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-itx}\gamma_1(x)\,\mathrm{d}x = 0 \qquad \text{for all } t \in \mathbb{R}. \tag{3.6}$$ pre:Hermite

If the Fourier transform $\hat{g}$ of a function $g \in C_c(\mathbb{R})$ is absolutely integrable, then by the inversion theorem of Fourier transforms,

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\hat{g}(t)\,\mathrm{d}t \qquad \text{for all } x \in \mathbb{R}.$$

Multiply both sides of (3.6) by $\hat{g}(t)$ and integrate $[\mathrm{d}t]$ in order to see from Fubini's theorem that $\int fg\,\mathrm{dP}_1 = 0$ for all $g \in C_c(\mathbb{R})$ such that $\hat{g} \in L^1(\mathbb{R})$. Since the class of such functions g is dense in $L^2(\mathrm{P}_1)$, it follows that $\int fg\,\mathrm{dP}_1 = 0$ for every $g \in L^2(\mathrm{P}_1)$. Set $g \equiv f$ to see that $f = 0$ a.s. □

Finally, let us mention one more important corollary.

⟨co:Nash⟩ **Corollary 1.6** (A Poincaré Inequality)**.** $\mathrm{Var}(f) \leqslant \mathrm{E}(|Df|^2)$ *for all* $f \in \mathbb{D}^{1,2}(\mathrm{P}_1)$.

Thus, we see from the Poincaré inequality that, if the [Malliavin] derivative of f is "small," then f is close to the constant $\mathrm{E}(f)$ with high probability.

Proof. By Corollary 1.5 and (3.5),

$$\begin{aligned}\mathrm{Var}(f) &= \sum_{k=0}^\infty \frac{1}{(k+1)!}\,\left|\mathrm{E}[fH_{k+1}]\right|^2 = \sum_{k=0}^\infty \frac{1}{(k+1)!}\,\left|\mathrm{E}[fA(H_k)]\right|^2\\ &= \sum_{k=0}^\infty \frac{1}{(k+1)!}\,\left|\mathrm{E}[D(f)H_k]\right|^2 \leqslant \sum_{k=0}^\infty \frac{1}{k!}\,\left|\mathrm{E}[D(f)H_k]\right|^2.\end{aligned}$$

The right-most quantity is equal to $\mathrm{E}(|Df|^2)$, thanks to Corollary 1.5. □

2 Hermite Polynomials in General Dimensions

One can easily extend the domain of definition of Hermite polynomials from $\mathbb{R}$ to $\mathbb{R}^n$ by tensorization: For every $k \in \mathbb{Z}_+^n$ let $\mathcal{H}_k := H_{k_1} \otimes \cdots \otimes H_{k_n}$. Or written out the long way,

$$\mathcal{H}_k(x) := \prod_{j=1}^n H_{k_j}(x_j) \qquad [k \in \mathbb{Z}_+^n,\ x \in \mathbb{R}^n].$$

As it turns out, these are the natural n-variable extensions of Hermite polynomials on $\mathbb{R}$. Though, when $n = 1$, we will continue to write $H_k(x)$ in place of $\mathcal{H}_k(x)$ in order to distinguish the multi-dimensional case from the case $n = 1$.

Clearly, $x \mapsto \mathcal{H}_k(x)$ is a polynomial, in n variables, of degree k_j in the variable x_j. For instance, when $n = 2$,

$$\mathcal{H}_{(0,0)}(x) = 1, \qquad \mathcal{H}_{(1,0)}(x) = x_1, \qquad \mathcal{H}_{(0,1)}(x) = x_2, \tag{3.7 \{?\}}$$

$$\mathcal{H}_{(1,1)}(x) = x_1 x_2, \qquad \mathcal{H}_{(1,2)}(x) = x_1(x_2^2 - 1), \ldots . \tag{3.8 \{?\}}$$

Because each measure P_n has the product form $\mathrm{P}_n = \mathrm{P}_1 \times \cdots \times \mathrm{P}_1$, Theorem 1.3 immediately extends to the following.

⟨th:Hermite⟩ **Theorem 2.1.** *Define $k! := \prod_{\nu=1}^n k_\nu!$ for all $k \in \mathbb{Z}_+^n$. Then, for every integer $n \geqslant 1$, the collection $\{\mathcal{H}_k/\sqrt{k!}\}_{k\in\mathbb{Z}_+^n}$ is a complete, orthonormal basis in $L^2(\mathrm{P}_n)$.*

Corollary 1.4 has the following immediate extension.

⟨co:Hermite⟩ **Corollary 2.2.** *For every $n \geqslant 1$ and $f \in L^2(\mathrm{P}_n)$,*

$$f = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{E}(f\mathcal{H}_k)}{k!}\, \mathcal{H}_k \quad \textit{almost surely,}$$

where the infinite sum converges in $L^2(\mathrm{P}_n)$.

Similarly, the following immediate extension of Corollary 1.5 computes the covariance between two arbitrary square-integrable random variables in the Gauss space.

⟨co:Hermite:Wiener⟩ **Corollary 2.3** (Wiener XXX)**.** *For all $n \geqslant 1$ and $f, g \in L^2(\mathrm{P}_n)$,*

$$\mathrm{E}[fg] = \sum_{k\in\mathbb{Z}_+^n} \frac{1}{k!}\, \mathrm{E}[f\mathcal{H}_k]\,\mathrm{E}[g\mathcal{H}_k] \quad \textit{and} \quad \mathrm{Cov}(f\,, g) = \sum_{\substack{k\in\mathbb{Z}_+^n\\ k\neq 0}} \frac{1}{k!}\, \mathrm{E}(f\mathcal{H}_k)\,\mathrm{E}(g\mathcal{H}_k).$$

And the following generalizes Corollary 1.6 to several dimensions.

⟨pr:Nash⟩ **Proposition 2.4** (The Poincaré Inequality)**.** $\mathrm{Var}(f) \leqslant \mathrm{E}(\|Df\|^2)$ *for all* $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$.

Proof. By Corollary 2.2, the following holds a.s. for all $1 \leqslant q \leqslant n$:

$$D_q f = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{E}[D_q(f)\mathcal{H}_k]}{k!}\, \mathcal{H}_k = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{E}[f A_q(\mathcal{H}_k)]}{k!}\, \mathcal{H}_k,$$

where we recall A_q denotes the qth coordinate of the vector-valued adjoint operator. By orthogonality and (3.5),

$$\begin{aligned}
\mathrm{E}\left(\|Df\|^2\right) &= \sum_{q=1}^n \sum_{k\in\mathbb{Z}_+^n} \frac{1}{k!}\, \left|\mathrm{E}[f A_q(\mathcal{H}_k)]\right|^2 \\
&= \sum_{q=1}^n \sum_{k\in\mathbb{Z}_+^n} \frac{1}{k!} \left|\mathrm{E}\left[f(Z) H_{k_q+1}(Z_q) \prod_{\substack{1\leqslant \ell\leqslant n\\ \ell\neq q}} H_{k_\ell}(Z_\ell)\right]\right|^2 .
\end{aligned}$$

Fix an integer $1 \leqslant q \leqslant n$ and relabel the inside sum as $j_\ell := k_\ell$ if $\ell \neq q$ and $j_q := k_q + 1$. In this way we find that

$$\begin{aligned}
\mathrm{E}\left(\|Df\|^2\right) &\geqslant \sum_{q=1}^{n} \sum_{\substack{j \in \mathbb{Z}_+^n \\ j_q \geqslant 1}} \frac{1}{j_1! \cdots j_n!} \left| \mathrm{E}\left[f(Z) H_{j_q}(Z_q) \prod_{\substack{1 \leqslant \ell \leqslant n \\ \ell \neq q}} H_{j_\ell}(Z_\ell) \right] \right|^2 \\
&= \sum_{q=1}^{n} \sum_{\substack{j \in \mathbb{Z}_+^n \\ j_q \geqslant 1}} \frac{1}{j_1! \cdots j_n!} \left| \mathrm{E}\left[f \mathcal{H}_j \right] \right|^2 .
\end{aligned}$$

using only the fact that $1/(j_q - 1)! > 1/j_q!$. This completes the proof since the right-hand side is simply

$$\sum_{j \in \mathbb{Z}_+^n} \frac{1}{j_1! \cdots j_n!} \left| \mathrm{E}\left[f \mathcal{H}_j \right] \right|^2 - \left| \mathrm{E}[f \mathcal{H}_0] \right|^2 ,$$

which is equal to the variance of $f(Z)$ [Corollary 2.3]. □

Consider a Lipschitz-continuous function $f : \mathbb{R}^n \to \mathbb{R}$. Recall [Example 1.6, page 30] that this means that $\mathrm{Lip}(f) < \infty$, where

$$\mathrm{Lip}(f) := \sup_{\substack{x, y \in \mathbb{R}^n \\ x \neq y}} \frac{|f(x) - f(y)|}{\|x - y\|}.$$

Since $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $\|Df\| \leqslant \mathrm{Lip}(f)$ a.s., the Poincaré inequality has the following ready consequence.

⟨co:Nash:Lip⟩ **Corollary 2.5.** *For every Lipschitz-continuous function* $f : \mathbb{R}^n \to \mathbb{R}$*,*

$$\mathrm{Var}(f) \leqslant |\mathrm{Lip}(f)|^2.$$

Thus, Corollary 2.5 implies the intuitively-pleasant fact that if $\mathrm{Lip}(f)$ is small, then $f \approx \mathrm{E}(f)$ with high probability.

Let us now mention two concrete examples.

Example 2.6. The function $f(x) := n^{-1} \sum_{i=1}^n x_i$ is Lipschitz continuous and $\mathrm{Lip}(f) = 1/\sqrt{n}$. In this case, Corollary 2.5 implies that

$$\mathrm{Var}\left(n^{-1} \textstyle\sum_{i=1}^n Z_i\right) \leqslant n^{-1},$$

which is in fact an identity. This example shows that the bound in the Poincaré inequality can be saturated.

Example 2.7. For a more interesting example consider either the function $f(x) := \max_{1 \leqslant i \leqslant n} |x_i|$ or the function $g(x) := \max_{1 \leqslant i \leqslant n} x_i$. Both f and g are Lipschitz-continuous functions with Lipschitz constant 1. The Poincaré inequality implies that $\mathrm{Var}(M_n) \leqslant 1$,[1] where M_n denotes either $\max_{1 \leqslant i \leqslant n} Z_i$ or $\max_{1 \leqslant i \leqslant n} |Z_i|$. This is a non-trivial result about, for example, the absolute size of the *centered* random variable $M_n - \mathrm{E}\, M_n$. The situation changes completely once we remove the centering. Indeed by Proposition 1.3 (p. 7) and Jensen's inequality,

$$\mathrm{E}(M_n^2) \geqslant |\mathrm{E}(M_n)|^2 \sim 2 \log n \qquad \text{as } n \to \infty.$$

[1] This bound is sub optimal. The optimal bound is $\mathrm{Var}(M_n) = O(1/\log n)$. For more information on this see part (b) of Problem 19 on page 26.

Similar examples can be constructed for more general Gaussian random vectors than Z, thanks to the following.

⟨pr:Poincare:X⟩ **Proposition 2.8.** *Let Q be a positive semidefinite matrix, and define λ_* to be its largest eigenvalue. If X is distributed as $\mathrm{N}_n(0\,,Q)$, then*

$$\mathrm{Var}[f(X)] \leqslant \lambda_* \,\mathrm{E}\left(\|(Df)(X)\|^2\right) \qquad \textit{for every } f \in \mathbb{D}^{1,2}(\mathrm{P}_n).$$

Proof. We can write $Q = S^2$ where S is a symmetric $n \times n$ matrix; that is, S is a square root of Q. Define $g(x) := f(Sx)$ for every $x \in \mathbb{R}^n$, and observe that: (i) X has the same distribution as SZ; and therefore (ii) $\mathrm{Var}(f(X)) = \mathrm{Var}(g(Z)) \leqslant \mathrm{E}(\|(Dg)(Z)\|^2)$ thanks to Proposition 2.4. By the chain rule, $(Dg)(Z) = (Df)(SZ)S$, whence

$$\|(Dg)(Z)\|^2 = \langle (Df)(SZ)S\,,(Df)(SZ)S\rangle_{\mathbb{R}^n} = \langle (Df)(SZ)\,,(Df)(SZ)Q\rangle_{\mathbb{R}^n},$$

thanks to the facts that $Q = S^2$ and S is symmetric.[2] Since Q is symmetric, Rayleigh's principle yields $\langle x\,,xQ\rangle_{\mathbb{R}^n} \leqslant \lambda_*\|x\|^2$ for all $x \in \mathbb{R}^n$. Set $x := (Df)(SZ)$ to see that

$$\mathrm{E}(\|(Dg)(Z)\|^2) \leqslant \lambda_*\,\mathrm{E}(\|(Df)(SZ)\|^2),$$

which is equal to $\lambda_*\,\mathrm{E}(\|(Df)(X)\|^2)$. □

The above proposition is in general unimproveable: For example, the case that Df is P_n-a.s. a constant multiple of the eigenvector that corresponds to the largest eigenvalue of Q. Still, the proposition can be sharpened for certain specific choices of f. The next proposition highlights this assertion in a particular case.

⟨pr:Var:max⟩ **Proposition 2.9.** *Suppose X has a $\mathrm{N}_n(0\,,Q)$ distribution, and let M_n denote either $\max_{1\leqslant i\leqslant n} X_i$ or $\max_{1\leqslant i\leqslant n}|X_i|$. Then,*

$$\mathrm{Var}(M_n) \leqslant \max_{1\leqslant i\leqslant n} \mathrm{Var}(X_i).$$

Proof. Let S denote a symmetric square root of Q, and define $f(x) = \max_{1\leqslant i\leqslant n} x_i$. Because $f(X)$ has the same distribution as $f(SZ)$, the Poincaré inequality (Proposition 2.4) implies that

$$\mathrm{Var}\left(\max_{1\leqslant i\leqslant n} X_i\right) \leqslant \mathrm{E}\left[\langle (Df)(SZ)\,,(Df)(SZ)Q\rangle_{\mathbb{R}^n}\right].$$

See also the proof of Proposition 2.8. Because Df is P_n-a.s. a standard basis vector of $\mathbb{R}^n$ (see Example 1.8, p. 31), it follows that $\langle (Df)(SZ)\,,(Df)(SZ)Q\rangle_{\mathbb{R}^n}$ is P_n-almost surely a diagonal entry of Q and hence is bounded above by $\max_{1\leqslant i\leqslant n} Q_{i,i}$ a.s. This verifies the proposition in the case that $M_n = \max_{1\leqslant i\leqslant n} X_i$. The case that $f(x) = \max_{1\leqslant i\leqslant n}|x_i|$ is handled the same way, except that now Df is ± 1 times some standard basis element of $\mathbb{R}^n$; see Problem 8 on page 36. □

[2] Needless to say, $\langle a\,,b\rangle_{\mathbb{R}^n} := a \cdot b = \sum_{i=1}^n a_i b_i$ for all $a, b \in \mathbb{R}^n$.

3 Wick's Formula

Theorem 1.3 asserts that, after they are suitably normalized, the Hermite polynomials form a complete, orthonormal basis for $L^2(\mathrm{P}_n)$. Using this fact, a change of variables allows us to find a basis for $L^2(\mathrm{Q}_n)$, where Q_n denotes the law of an $\mathrm{N_n}(0\,,\Gamma)$ random variable provided that Γ has full rank (equivalently, is positive definite). Here is a precise statement of this fact.

⟨co:Hermite:correlated⟩ **Corollary 3.1.** *Let Q_n denote the law of a $\mathrm{N}_n(0\,,\Gamma)$ random variable, where Γ is a positive-definite $n\times n$ matrix. Define $\mathcal{C}_k(x) := \mathcal{H}_k(\Gamma^{-1/2}x)$ for every $k \in Z^n_+$ and $x \in \mathbb{R}^n$. Then $\{\mathcal{C}_k/\sqrt{k!}\}_{k\in\mathbb{Z}^n_+}$ is a complete, orthonormal basis for $L^2(\mathrm{Q}_n)$. Hence,*

$$f = \sum_{k\in\mathbb{Z}^n_+} \frac{\mathrm{E}[f\mathcal{C}_k]}{k!}\mathcal{C}_k \qquad \textit{for every } f \in L^2(\mathrm{Q}_n).$$

Corollary 3.1 follows readily from Theorem 2.1 and Corollary 2.2. In practice, this particular representation of a general function $f \in L^2(\mathrm{Q}_n)$ is not always useful. This is because it is in general hard to calculate $\mathrm{E}[f\mathcal{C}_k]$ as its computation involves working with unwieldly multi-dimensional integrals. Nevertheless, when f is a polynomial, additional combinatorial ideas can be used to help organize the above representation of f in a useful way. The resulting representation is then called *Wick's formula*. Before we state and prove Wick's formula, let us introduce some notation.

For every integer $j \geqslant 1$, let $\mathcal{P}_j$ denote the following linear subspace of the Gaussian Hilbert space $L^2(\mathrm{Q}_n)$:

$$\mathcal{P}_j := \mathrm{Span}\{\mathcal{C}_k : k \in \mathbb{Z}^n_+,\ |k| = j\};$$

where we recall $|k| = \sum_{m=1}^n k_m$ for every $k \in \mathbb{Z}^n_+$. We also define $\mathcal{P}_0$ to be the space of all real-valued, constant functions on $\mathbb{R}^n$. In this way, Corollary 3.1 can be rephrased succintly as an orthogonal decomposition of $L^2(\mathrm{Q}_n)$ in terms of the subspaces $\{\mathcal{P}_j\}_{j=0}^\infty$. Or, equivalently,

$$L^2(\mathrm{Q}_n) = \bigoplus_{j=0}^\infty \mathcal{P}_j. \tag{3.9}$$ eq:Wiener_chaos_defn

In other words, suppose X is distributed as Q_n and explicitly construct the Gaussian Hilbert space $L^2(\mathrm{Q}_n)$ as $L^2(\Omega\,,\sigma(X),\mathrm{Q}_n)$. Then $\mathcal{P}_j$ is the linear subspace of all degree-j polynomials in the variables $X_1,\ldots,X_n$ that are uncorrelated with all degree $< j$ polynomials in $X_1,\ldots,X_n$. This interpretation of (3.9) is particularly nice because it does not suppose knowledge of the explicit polynomial basis $\{\mathcal{C}_k : k \in \mathbb{Z}^n_+,\ |k| = j\}$ in order to define $\{\mathcal{P}_j\}_{j=0}^\infty$. Another nice feature of this "basis-free definition" is that it does not require the covariance matrix $\Gamma = (\mathrm{Cov}(X_i\,,X_j))_{1\leqslant i,j\leqslant n}$ of X to be positive definite (Γ is of course always positive semi-definite). In particular, this basis-free definition allows for some of the X_i's to be equal to one another. Nonetheless the decomposition (3.9) continues to remain valid, and simply states that *every Borel function of X with finite variance can be written uniquely as a (possibly) infinite sum of polynomials, one from every $\mathcal{P}_j$.*

Definition 3.2. The orthogonal decomposition (3.9) is referred to as the *Wiener chaos decomposition* of the Gaussian Hilbert space $L^2(\mathrm{Q}_n)$. Moreover, for every integer $j \geqslant 0$, $\mathcal{P}_j$ is called the jth *Wiener chaos* of $L^2(\mathrm{Q}_n)$. We define $\pi_j : L^2(Q_n) \to \mathcal{P}_j$ to be the orthogonal projection operator onto $\mathcal{P}_j$ in the $L^2(Q_n)$ inner product. Thus, every random variable $Y \in L^2(Q_n)$ can be decomposed uniquely as $Y = \pi_j Y + Y - \pi_j Y$, with $\mathrm{E}[\pi_j Y(Y - \pi_j Y)] = 0$.

Recall that that the *degree* of a monomial is the sum of its exponents, and the degree of a polynomial is the maximum degree of all monomials that appear in that polynomial. Then we learn immediately from (3.9) that a polynomial $f(X)$ of degree $\ell \in \mathbb{Z}_+$ in $X_1, \ldots, X_n$ can be decomposed uniquely into a sum of elements from $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_\ell$. As was noted above, the part in $\mathcal{P}_j$ is simply the orthogonal projection $\pi_j(f(X))$ of $f(X)$ in $L^2(\mathrm{Q}_n)$ onto $\mathcal{P}_j$.[3]

Wick's theorem, or Wick's formula, is a combinatorial expression for $\prod_{i=1}^n X_i$ and essentially contains the Isserlis theorem (p. 20) as a special case; see Remark 3.6 below for details. As such, Wick's formula inevitably involving the notion of matchings. With this in mind, let A denote a finite set. We write a partial matching of A as $\mathfrak{m} \oplus \mathfrak{u}$, where $\mathfrak{m}$ is a perfect matching of a subset of A and $\mathfrak{u}$ denotes the remaining unmatched elements. In the statement of Wick's theorem we also adopt the conventions that $\sum_\varnothing := 0$ and $\prod_\varnothing = 1$.

⟨th:Wick⟩ **Theorem 3.3** (Wick, YYYY)**.** *Suppose X has a $\mathrm{N}_n(0\,,\Gamma)$ distribution, where Γ is an arbitrary $n \times n$ covariance matrix. Then,*

$$\pi_n\left(\prod_{j=1}^n X_j\right) = \sum_{\mathfrak{m}\oplus\mathfrak{u}} (-1)^{|\mathfrak{m}|} \prod_{(i,j)\in\mathfrak{m}} \Gamma_{i,j} \prod_{k\in\mathfrak{u}} X_k, \tag{3.10}$$ eqn:Wick_projection

where the sum ranges over all partial matchings $\mathfrak{m} \oplus \mathfrak{u}$ of pairs in $\{1\,, \ldots, n\}$ and $|\mathfrak{m}|$ denotes the number of matched pairs in the partial matching $\mathfrak{m} \oplus \mathfrak{u}$. Furthermore, the Wiener chaos decomposition of $\prod_{j=1}^n X_j$ into orthogonal parts can be written as

$$\prod_{j=1}^n X_j = \sum_{\mathfrak{m}\oplus\mathfrak{u}} \prod_{(i,j)\in\mathfrak{m}} \Gamma_{i,j}\, \pi_{|\mathfrak{u}|}\left(\prod_{k\in\mathfrak{u}} X_k\right), \tag{3.11}$$ eqn:Wick_decomposition

where $\pi_{|\mathfrak{u}|}$ denote the projection operator onto the Wiener chaos $\mathcal{P}_{|\mathfrak{u}|}$ corresponding to the number of unmatched terms $|\mathfrak{u}|$.

Wick's theorem makes two distinct assertions: The first is that every projection is itself a polynomial in the variables $X_1\,, \ldots, X_n$. Whereas the second assertion of the theorem is that every polynomial in Gaussian variables $X_1, \ldots, X_n$ can be written uniquely as a sum of polynomials in the same variables with the property that any two terms in the sum are uncorrelated.

Let us consider two examples. The first is a "trivial" one.

Example 3.4. A single Gaussian variable X_1 has to be left unmatched, and hence $X_1 = \pi_1(X_1)$. Consequently, $\pi_1(aX_1 + b) = a\pi_1(X_1)$ for all real numbers a and b.

For a more interesting example consider the following.

Example 3.5. If $(X_1\,, X_2)$ has a Gaussian distribution, then X_1 and X_2 can either be matched or unmatched. This observation readily leads us to the formula

$$\pi_2(X_1X_2) = X_1X_2 - \mathrm{E}[X_1X_2].$$

[3] Wick's theorem was introduced originally in the context of quantum field theory in order to reduce a complicated product of creation and annihilation operators to sums of products of pairs of such operators. Physicists prefer to write $:Y:$ for the projection $\pi_n(Y)$ of a random variable $Y \in L^2(\mathrm{Q}_n)$.

Equivalently, $X_1X_2 = \pi_2(X_1X_2) + \mathrm{E}[X_1X_2]$, which is the orthogonal decomposition of X_1X_2. In particular, we may specialize to the case that $X_1 = X_2 = X$ in order to see that

$$\pi_2(X^2) = X^2 - \mathrm{E}[X^2] = X^2 - \mathrm{Var}(X). \tag{3.12}$$ pi2:Var

One can deduce from the above, especially from (3.12), that if a real-valued random variable X has a standard normal distribution, then $\pi_1(X) = X$ and $\pi_2(X^2) = X^2 - 1$. Since $\pi_0(X^0) = 1$, one might hope to have stumbled upon the general pattern,

$$\pi_n(X^n) = H_n(X) \qquad \text{for every } n \geqslant 0. \tag{3.13}$$ eq:Hermite_projection_for

It turns out that (3.13) is true. Let us verify it for $n = 3$ and $n = 4$, and leave the general case to the interested reader (Problem 7).

In order to verify (3.13) for the case $n = 3$, we first appeal to the Wick theorem (Theorem 3.3) to see that for every Gaussian random variable $(X_1\,, X_2\,, X_3)$,

$$\pi_3(X_1X_2X_3) = X_1X_2X_3 - \mathrm{E}[X_1X_2]X_3 - \mathrm{E}[X_1X_3]X_2 - \mathrm{E}[X_2X_3]X_1.$$

Then, we specialize this formula to the degenerate case where $X_1 = X_2 = X_3 = X$ in order to see that $\pi_3(X^3) = X^3 - 3X$, which agrees with (3.13) for $n = 3$. See (3.3) on page 39.

Similarly, for four variables, we have

$$\begin{aligned}\pi_4(X_1X_2X_3X_4) = {} & X_1X_2X_3X_3 - \mathrm{E}[X_1X_2]X_3X_4 - \mathrm{E}[X_1X_3]X_2X_4 - \mathrm{E}[X_1X_4]X_2X_3\\ & - \mathrm{E}[X_2X_3]X_1X_4 - \mathrm{E}[X_2X_4]X_1X_3 - \mathrm{E}[X_3X_4]X_1X_2\\ & + \mathrm{E}[X_1X_2]\,\mathrm{E}[X_3X_4] + \mathrm{E}[X_1X_3]\,\mathrm{E}[X_2X_4] + \mathrm{E}[X_1X_4]\,\mathrm{E}[X_2X_3],\end{aligned}$$

which specializes to $\pi_4(X) = X^4 - 6X^2 + 3 = H_4(X)$. See (3.3) on page 39.

⟨rem:Wick->Isserlis⟩ *Remark* 3.6. As was alluded to earlier, equation (3.11) contains the Isserlis formula (Theorem 5.1, p. 20) as a special case. In order to see why let $\mathfrak{m}\oplus\mathfrak{u}$ denote a [non void] partial matching of pairs in $\{1\,,\ldots,n\}$ and notice that $\sum_{(i,j)\in\mathfrak{m}\oplus\mathfrak{u}}\Gamma_{i,j}\prod_{k\in\mathfrak{u}}X_k$ lies in $\cup_{i=1}^n\mathcal{P}_j$ and is therefore orthogonal to $\mathcal{P}_0$. Equivalently, it has mean zero. Thus, we see that if we take expectations of both sides of (3.11), then all terms vanish except the ones that come from a perfect matching $\mathfrak{m}$ [i.e., $\mathfrak{u} = \varnothing$]. The Isserlis formula then follows from the identity $\pi_0(\prod_{k\in\varnothing}X_k) = 1$, which is a tautology.

Although the statement of Wick's theorem contains the statement of the Isserlis theorem, the proof of Wick's theorem actually relies on the Isserlis theorem. It also hinges on the following elementary combinatorial fact which we leave as exercise (see Problem 5).

⟨lem:comb⟩ **Lemma 3.7.** *For every non-empty finite set A and for all $x, y \in \mathbb{R}$,*

$$(x+y)^{|A|} = \sum_{B\subseteq A} x^{|A|}y^{|A\setminus B|},$$

where $|\cdots|$ denotes cardinality.

Proof of Wick's theorem (Theorem 3.3). We prove Wick's formula via a series of claims. First, let Y denote the quantity on the right-hand side of (3.10). One of the contributing terms to the expression for Y is $\prod_{i=1}^n X_i$ [corresponding to $\mathfrak{m} = \varnothing$], and all the other terms are polynomials in $X_1,\ldots,X_n$ of degree $< n$. Therefore, Y is a polynomial in $X_1, \ldots, X_n$ of degree n.

Claim 1. $\mathrm{E}[Y] = 0$*; that is,* Y *is orthogonal to* $\mathcal{P}_0$.

In order to see why, let us note that $\mathrm{E}[Y] = \sum_{\mathfrak{m}\oplus\mathfrak{u}}(-1)^{|\mathfrak{m}|}\prod_{(i,j)\in\mathfrak{m}}\Gamma_{i,j}\,\mathrm{E}[\prod_{k\in\mathfrak{u}}X_k]$, and $\mathrm{E}[\prod_{k\in\mathfrak{u}}X_k]=0$ unless $|\mathfrak{u}|$ is even. Remember also that a sum over the empty set is zero. We can combine these remarks with the Isserlis formula (Theorem 5.1, p. 20) in order to see that

$$\begin{aligned}\mathrm{E}[Y] &= \sum_{\mathfrak{m}\oplus\mathfrak{u}}(-1)^{|\mathfrak{m}|}\prod_{(i,j)\in\mathfrak{m}}\Gamma_{i,j}\sum_{\substack{\text{perfect matchings}\\ \mathfrak{m}' \text{ of } \mathfrak{u}}}\prod_{(i',j')\in\mathfrak{m}'}\Gamma_{i',j'}\\ &= \sum_{\mathfrak{m}\oplus\mathfrak{u}}\sum_{\substack{\text{perfect matchings}\\ \mathfrak{m}' \text{ of } \mathfrak{u}}}(-1)^{|\mathfrak{m}|}\prod_{(i,j)\in\mathfrak{m}}\Gamma_{i,j}\prod_{(i',j')\in\mathfrak{m}'}\Gamma_{i',j'}.\end{aligned}$$

The first two sums are over all partial matchings of $[n]$ and perfect matchings of the remaining unmatched terms within the partial matching, where we are using the standard combinatorial notation,

$$[L] := \{1\,,\ldots,L\}\qquad\text{for all } L\in\mathbb{N}.$$

This is equivalent to summing over all perfect matchings of the original set $[n]$, so we may rearrange the sum in order to see that

$$\mathrm{E}[Y] = \sum_{\substack{\text{perfect matchings}\\ \mathfrak{m} \text{ of } [n]}}\prod_{(i,j)\in\mathfrak{m}}\Gamma_{i,j}\sum_{\mathfrak{m}'\subseteq\mathfrak{m}}(-1)^{|\mathfrak{m}'|}.$$

Because of Lemma 3.7, $\sum_{\mathfrak{m}'\subseteq\mathfrak{m}}(-1)^{|\mathfrak{m}'|} = (-1+1)^{|\mathfrak{m}|} = 0$. This shows that $\mathrm{E}[Y]=0$, and hence Y is orthogonal to $\mathcal{P}_0$, as was asserted.

Claim 2. $Y\in\mathcal{P}_n$.

In order to avoid trivialities, we assume that $n\geqslant 2$. Thanks to this and Claim 1, it remains to prove that Y is orthogonal to $\mathcal{P}_1,\ldots,\mathcal{P}_{n-1}$.

Choose and fix an integer $\ell = 1,\ldots,n-1$, and let $X_{n+1},\ldots,X_{n+\ell}$ denote any ℓ predescribed elements of $\{X_1\,,\ldots,X_n\}$. We plan to show that $\mathrm{E}[YX_{n+1}\ldots X_{n+\ell}]=0$. This will prove that Y is orthogonal to $\mathcal{P}_1,\ldots,\mathcal{P}_{n-1}$, and ends the proof of Claim 2.

We first apply the definition of Y, and then the Isserlis formula (Theorem 5.1, p. 20) for the [degenerate] Gaussian random variable $(X_1\,,\ldots,X_n\,,X_{n+1}\,,\ldots,X_{n+\ell})$, in order to see that

$$\begin{aligned}\mathrm{E}[YX_{n+1}\ldots X_{n+\ell}] &= \sum_{\mathfrak{m}\oplus\mathfrak{u}}(-1)^{|\mathfrak{m}|}\prod_{(i,j)\in\mathfrak{m}}\Gamma_{i,j}\,\mathrm{E}\left[X_{n+1}\ldots X_{n+\ell}\prod_{k\in\mathfrak{u}}X_k\right]\\ &= \sum_{\mathfrak{m}\oplus\mathfrak{u}}(-1)^{|\mathfrak{m}|}\prod_{(i,j)\in\mathfrak{m}}\Gamma_{i,j}\sum_{\substack{\text{perfect matchings}\\ \mathfrak{m}' \text{ of } \mathfrak{u}\cup\{n+1,\ldots,n+\ell\}}}\prod_{(i',j')\in\mathfrak{m}'}\Gamma_{i',j'},\end{aligned}$$

where, we recall, the final sum over perfect matchings $\mathfrak{m}'$ defaults to zero when there are no such perfect matchings; that is when $|\mathfrak{u}|+\ell$ is odd. And of course, $\Gamma_{i',j'} = \mathrm{Cov}(X_{i'},X_{j'})$ for $(i',j')\in\mathfrak{m}'$ even when one or both of i' and j' exceed n.

Now we rearrange the sum in order to see that

$$\mathrm{E}[YX_{n+1}\cdots X_{n+\ell}] = \sum_{\substack{\text{non-empty}\\ \text{perfect matchings}\\ \mathfrak{m}\text{ of }[n+\ell]}} \prod_{(i,j)\in\mathfrak{m}} \Gamma_{i,j} \sum_{\substack{\text{perfect matchings}\\ \mathfrak{m}'\text{ of }[n]:\\ \mathfrak{m}'\subseteq\mathfrak{m}}} (-1)^{|\mathfrak{m}'|},$$

which is zero by Lemma 3.7. This proves Claim 2.

Claim 3. Eq. (3.10) *holds; that is,* $Y = \pi_n(\prod_{i=1}^n X_i)$.

Recall that, for every $W \in L^2(\mathrm{Q}_n)$, $\pi_n(W)$ is the unique element of $\mathcal{P}_n$ that is orthogonal to $W - \pi_n(W)$. Since $Y \in \mathcal{P}_n$ (Claim 2), it remains to show that $\prod_{i=1}^n X_i - Y$ is orthogonal to Y. But this follows from Claim 2 since (3.10) implies that $\prod_{i=1}^n X_i - Y$ is a polynomial of degree $n-2$, and hence is an element of $\mathcal{P}_0 \cup \cdots \cup \mathcal{P}_{n-2}$. Indeed, the only term of degree $\geqslant n-1$ in Y is $\prod_{i=1}^n X_i$ itself, which corresponds to the partial matching in which every element is left unmatched. This proves Claim 3.

Claim 4. Eq. (3.11) *is valid.*

Claim 4 implies Wick's theorem, and follows by using (3.10) and similar combinatorial computations as above. Here are some of the details, written in short hand: Insert (3.10) into the right-hand side of (3.11) in order to deduce the identity,

$$\sum_{\substack{\mathfrak{m}\oplus\mathfrak{u}\\ \text{of }[n]}} \prod_{(i,j)\in\mathfrak{m}} \Gamma_{i,j}\pi_{|\mathfrak{u}|}\left(\prod_{k\in\mathfrak{u}} X_k\right) = \sum_{\substack{\mathfrak{m}\oplus\mathfrak{u}\\ \text{of }[n]}} \sum_{\substack{\mathfrak{m}'\oplus\mathfrak{u}'\\ \text{of }\mathfrak{u}}} (-1)^{|\mathfrak{m}'|} \prod_{(i,j)\in\mathfrak{m}} \Gamma_{i,j} \prod_{(i',j')\in\mathfrak{m}'} \Gamma_{i',j'} \prod_{k\in\mathfrak{u}'} X_k.$$

Note that the sum $\sum_{\mathfrak{m}\oplus\mathfrak{u}\text{ of }[n]}(\cdots)$ on the right-hand side is zero unless $|\mathfrak{u}|$ is even. Therefore, every non-zero term in the right-hand side's double summation corresponds to a partial matching of $[n]$ (namely to $(\mathfrak{m}\cup\mathfrak{m}')\oplus\mathfrak{u}'$), and the only term in the sum that does not depend on this partial matching (i.e., does not depend on $\mathfrak{m}\cup\mathfrak{m}'$ or $\mathfrak{u}'$) is $(-1)^{|\mathfrak{m}'|}$. Thus, by renaming the matched part as simply $\mathfrak{m}$ and the unmatched part as $\mathfrak{u}$, the right-hand side can be rewritten as

$$\sum_{\substack{\mathfrak{m}\oplus\mathfrak{u}\\ \text{of }[n]}} \prod_{(i,j)\in\mathfrak{m}} \Gamma_{i,j} \prod_{k\in\mathfrak{u}} X_k \sum_{\mathfrak{m}'\subseteq\mathfrak{m}} (-1)^{|\mathfrak{m}'|}.$$

As in the previous claims, $\sum_{\mathfrak{m}'\subseteq\mathfrak{m}}(-1)^{|\mathfrak{m}'|} = 0$ for all partial matchings $\mathfrak{m}\oplus\mathfrak{u}$ of $[n]$ except when $\mathfrak{m} = \varnothing$ when it trivially gives a value of one. Thus the entire sum collapses to $\prod_{j=1}^n X_j$, thereby completing the proof of (3.11). □

Problems

1. Prove that if we apply the Gram-Schmidt orthogonalization procedure – in Gauss space – to the monomials $\{1, x, x^2, x^3, \ldots\}$, then we obtain the Hermite polynomials $H_1, H_2, \ldots$.

⟨ex:Hermite:Taylor⟩ 2. Choose and fix an $w \in \mathbb{R}$ and define $f(z) = \exp(wz - w^2/2)$ for all $z \in \mathbb{R}$.
   (a) Verify that $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and calculate $\|f\|_{1,2}$.
   (b) Use integration by parts to show that $\mathrm{E}[f(Z)H_k] = w^k$.
   (c) Conclude that $H_k(x)$ is the coefficient of $w^k/k!$ in the Taylor series expansion of $w \mapsto \exp(wx - w^2/2)$. In other words, the Hermite polynomials are defined uniquely via the relation,
$$\exp\left(wx - \frac{w^2}{2}\right) = \sum_{k=0}^{\infty} \frac{w^k}{k!} H_k(x).$$

3. Verify that $H_n(0) = 0$ for all odd integers n, and
$$H_n(0) = (-1)^{n-1}(n-1)!! \quad \text{for all even integers } n,$$
where $k!! := k \times (k-2) \times (k-4) \times \cdots \times 1$ for all even integers k.

4. Recall the adjoint operators A_j of (2.4). Show that for $k \in \mathbb{Z}_+^n$
$$A_1^{k_1} \ldots A_n^{k_n} 1 = \mathcal{H}_k(Z),$$
where A_j^0 is the identity operator. Show that the order of the adjoint operators does not matter, so that if $q \in \{1, \ldots, n\}^p$ and if σ is a permutation of $\{1, 2, \ldots, p\}$ then $A_{q_1} \ldots A_{q_p} 1 = A_{q_{\sigma(1)}} \ldots A_{q_{\sigma(p)}} 1$.

⟨pbm:comb⟩ 5. Verify Lemma 3.7 and use it to derive the binomial theorem for $(x+y)^n$.

6. Derive the following "binomial theorem" for Hermite functions:
$$H_n(x+y) = \sum_{k=0}^{n} \binom{n}{k} y^k H_{n-k}(x) \quad \text{for all } x, y \in \mathbb{R} \text{ and } n \in \mathbb{Z}_+.$$
(Hint: See Problem 2.)

⟨pbm:HPF⟩ 7. In the following, consider the canonical probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathrm{P}_n)$, so that $\mathcal{P}_\ell$ denotes the linear subspace spanned by $\{\mathcal{H}_k : k \in \mathbb{Z}_+^n, |k| = \ell\}$ for every $\ell = 0, 1, \cdots$, and π_ℓ denotes projection onto $\mathcal{P}_\ell$.
   (a) Prove that, when $n = 1$, $\pi_m(Z^m) = H_m(Z)$ a.s. $[\mathrm{P}_1]$ for every $m \in \mathbb{N}$. (Hint: $Z^m - H_m(Z)$ is polynomial of degree $m-1$.)
   (b) Deduce (3.13) from the previous part.
   (c) Define a polynomial $P : \mathbb{R}^n \to \mathbb{R}$ by $P = \sum \alpha(m) Z_1^{m_1} \cdots Z_n^{m_n}$, where $\{\alpha(m)\}_{m \in \mathbb{Z}_+^n}$ are real numbers, the sum is taken over all $m \in \mathbb{Z}_+^n$ such that $|m| = \ell$, and $\ell \geqslant 1$ is integral. Prove the following generalization of (3.13): For every integer $\ell \geqslant 0$,
$$\pi_\ell(P) = \sum_{m \in \mathbb{Z}_+^n : |m| = \ell} \alpha(m) \mathcal{H}_m(Z) \qquad \text{a.s.}[\mathrm{P}_n].$$

8. Use Wick's formula (3.10) and the fact that $\pi_n(Z^n) = H_n(Z)$ to show that the Hermite polynomials can be written as
$$H_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \frac{n!}{2^k k!(n-2k)!} x^{n-2k},$$
for every even integer n and $x \in \mathbb{R}$.

9. Use Wick's formula (3.11) to show that for every even integer n and $x \in \mathbb{R}$,

$$x^n = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} \frac{(2k)!}{2^k k!} H_{n-2k}(x).$$

Then re-verify this formula using induction and (3.4) on page 40.

10. (**This problem requires some background in Itô calculus.**)
Let B be a standard Brownian motion, and define $B_n(t) := \int_0^t B_{n-1}(s)\, \mathrm{d}B(s)$ as an Itô integral for every integer $n \geqslant 1$, where $B_0(t) := 1$ for all $t \geqslant 0$. These are *multiple Itô integrals*; for example,

$$B_1(t) = B(t), \quad B_2(t) = \int_0^t \int_0^s \mathrm{d}B(r)\, \mathrm{d}B(s), \cdots$$
$$B_{n+1}(t) = \int_0^t \int_0^{s_1} \cdots \int_0^{s_n} \mathrm{d}B(s_{n+1}) \cdots \mathrm{d}B(s_2)\, \mathrm{d}B(s_1), \cdots .$$

Choose and fix some $\alpha \geqslant 0$, and define $X(t) := \sum_{n=0}^\infty \alpha^n B_n(t)$ for all $t \geqslant 0$.

(a) Verify that for every $T > 0$, the series converges in $L^2(\Omega)$, uniformly for $t \in [0\,, T]$. [Hint: Doob's maximal $L^2(\Omega)$ inequality.]

(b) Prove that X satisfies the Itô stochastic differential equation, $\mathrm{d}X(t) = \alpha X(t)\mathrm{d}B(t)$ subject to $X(0) = 1$. Conclude that

$$X(t) = \exp\left(\alpha B(t) - \frac{t\alpha^2}{2}\right) \qquad \text{for all } t \geqslant 0 \text{ a.s.}$$

(c) Compare (b) to Problem 2 in order to conclude that

$$H_n(B(1)) = n! \int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-1}} \mathrm{d}B(s_n) \cdots \mathrm{d}B(s_2)\, \mathrm{d}B(s_1) \quad \text{a.s.}$$

Because $B(1)$ has the same distribution as Z, the above gives a particular construction of $H_n(Z)$ using Brownian motion. This construction is part of a deep theory of Wiener XXX and Itô XXX. The exposition is due to McKean XXX.

11. Extend Problem 2 to $\mathbb{R}^n$ for all $n \geqslant 1$ by showing that

$$\exp\left(w \cdot x - \frac{\|w\|^2}{2}\right) = \sum_{k \in \mathbb{Z}_+^n} \frac{w^k}{k!}\, \mathcal{H}_k(x) \qquad \text{for every } w, x \in \mathbb{R}^n,$$

where $w^k := \prod_{i=1}^n w_i^{k_i}$.

⟨ex:Dk,2⟩ 12. Suppose that $f \in L^2(\mathrm{P}_n)$ satisfies

$$\sum_{k \in \mathbb{Z}_+^n} \frac{\|k\|^{2m}}{k!} \left|\mathrm{E}(f\mathcal{H}_k)\right|^2 < \infty \qquad \text{for some } m \in \mathbb{N}.$$

Prove that $f \in \mathbb{D}^{m,2}(\mathrm{P}_n)$, using the following steps:

(a) Expand f in terms of Hermite polynomials, and let f_ℓ denote the same sum but restricted to indices $k \in \mathbb{Z}_+^n$ that satisfy $\|k\| \leqslant \ell$. Prove that $f_\ell \in C_0^\infty(\mathrm{P}_n)$ and $\lim_{\ell\to\infty} f_\ell = f$ in $L^2(\mathrm{P}_n)$. Use this to deduce that it suffices to prove that $\{D_{i_1} \cdots D_{i_m} f_\ell\}_{\ell=1}^\infty$ is Cauchy in $L^2(\mathrm{P}_n)$ for every $i_1, \ldots, i_m \in \{1\,, \ldots, n\}$.

(b) Now find an expression for $D_{i_1} \cdots D_{i_m} f_\ell$ in terms of Hermite polynomials. [Hint: Examine the proof of Proposition 2.4.]

13. Show that if we are working in $L^2(\mathrm{P}_n)$ then the jth Wiener chaos $\mathcal{P}_j$ has dimension $\binom{j+n-1}{n-1}$.

⟨que:quadratic_form_Wick⟩ 14. Suppose we are working in $L^2(\mathrm{P}_n)$, so that $Z_1, \ldots, Z_n$ are iid standard normals. As usual write $Z = (Z_1, \ldots, Z_n)'$ for the random column vector consisting of these normals (recall $'$ means transpose). Let A be an $n \times n$ matrix of real numbers.

(a) Show that the quadratic form $Z'AZ$ is unchanged if A is replaced by its symmetrized version $(A + A')/2$.

(b) Thus assume from now on that $A = A'$; i.e., A is symmetric. Show that $\pi_2(Z'AZ) = Z'AZ - \mathrm{tr}(A)$, where $\mathrm{tr}(A)$ is the trace of the matrix A.

(c) More generally, show that if we are working under $L^2(\mathrm{Q}_n)$ where Q_n is the measure of the $\mathrm{N}_n(0\,,\Sigma)$ distribution, then $\pi_2(Z'AZ) = Z'AZ - \mathrm{tr}(A\Sigma)$.

(d) Finally, show the following higher degree version. For an integer $k \geqslant 2$ let K be a k-tensor, meaning that $K : [n]^k \to \mathbb{R}$, and for the polynomial of degree k given by

$$f = \sum_{q \in [n]^k} K_q Z_{q_1} \ldots Z_{q_k}$$

show that under $L^2(\mathrm{P}_n)$

$$\pi_k(f) = \sum_{q \in [n]^k} K_q \pi_k(Z_{q_1} \ldots Z_{q_k}) = \sum_{q \in [n]^k} K_q \mathcal{H}_{c(q)}(Z)$$

where $c(q) = (c_1(q), \ldots, c_n(q))$ and $c_i(q)$ is the number of times that i appears in the k-tuple q, for $i \in [n] = \{1, \ldots, n\}$. Further simplify this formula by showing that every such f can be represented by the symmetrized version $\tilde{K}$ of the k-tensor as

$$f = \sum_{q \in [n]^k} \tilde{K}_q Z_{q_1} \ldots Z_{q_k}$$

where $\tilde{K}$ is defined by

$$\tilde{K}_q = \frac{c(q)!}{k!} \sum_{\sigma} K_q.$$

Here σ is a permutation of $\{1, \ldots, k\}$ and $\sigma(q) = (q_{\sigma(1)}, \ldots, q_{\sigma(k)})$, and the coefficient of the sum is the inverse of a multinomial coefficient. Conclude that

$$\pi_k(f) = \sum_{q \in [n]^k} \tilde{K}_q \mathcal{H}_{c(q)}(Z) = \sum_{q \in [n]^{\odot k}} \frac{k!}{c(q)!} \tilde{K}_q \mathcal{H}_{c(q)}(Z),$$

where $[n]^{\odot k} = [n]^k / \sim$, and $\sim$ is the equivalence relation $q \sim q'$ iff $c(q) = c(q')$; i.e., each element of $[n]$ appears the same number of times in both q and q' and therefore one is just a permutation of the other.

15. Prove that the Poincaré inequality on $\mathbb{R}^n$ [Proposition 2.4] follows directly from the one-dimensional case [Corollary 1.6] and induction on the value of $n \geqslant 1$. This method is sometimes called "tensorization."

16. Let $f(x) := \max_{1 \leqslant i \leqslant n} x_i$ for all $x \in \mathbb{R}^n$, and prove that Proposition 2.9 improves Proposition 2.8. That is, prove that $\lambda_* \, \mathrm{E}(\|Df\|^2) \geqslant \max_{1 \leqslant i \leqslant n} \mathrm{Var}(X_i)$ for the present choice of f.

The following Problems 17–22 depend sequentially on one another. Throughout these problems, let us choose and fix some $\lambda > 0$, and let X have a Poisson distribution with $\mathrm{E}(X) = \lambda$. Also, let μ denote the distribution of X; that is, $\mu\{k\} = \mathrm{e}^{-\lambda}\lambda^k/k!$ for $k \in \mathbb{Z}_+$ and $\mu\{k\} = 0$ otherwise. Finally, define $C_0, C_1, C_2, \ldots$ canonically as the real-valued functions on $\mathbb{Z}_+$ that satisfy the following for all $x = 0, 1, 2, \ldots$ and $w > -1$:

$$\mathrm{e}^{-w\lambda}(1+w)^x = \sum_{k=0}^{\infty} \frac{w^k}{k!}\, C_k(x).$$

Many authors usually write $C_k^{(\lambda)}$ instead of C_k, and refers to C_k as the kth *monic Charlier polynomial* with parameter λ.

⟨ex:Poisson:1⟩ 17. Prove that $C_0(x) = 1$ and $C_k(0) = (-\lambda)^k$ for all $x, k \in \mathbb{Z}_+$.

18. Verify that $\mathrm{E}[\exp\{-w\lambda\}(1+w)^X] = 1$ for all $w > -1$. Conclude from this that $\mathrm{E}[C_k(X)] = 0$ for all $k \geqslant 1$.

19. Verify that

$$C_k(x) = \sum_{m=0}^{k\wedge x} \binom{k}{m}\binom{x}{m} m!(-\lambda)^{k-m} \quad \text{for all } x, k \in \mathbb{Z}_+,$$

and conclude that every C_k is a polynomial of degree at most k on the semigroup $\mathbb{Z}_+$.

20. Prove that the sequence $\left\{\sqrt{n!/\lambda^n}\, C_k\right\}_{k=0}^{\infty}$ is a complete orthonormal basis for $L^2(\mu)$. Conclude that for all $f, g \in L^2(\mu)$,

$$\mathrm{Cov}[f(X)\,, g(X)] = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \langle f\,, C_k\rangle_{L^2(\mu)} \langle g\,, C_k\rangle_{L^2(\mu)}.$$

[Hint: Consider the second moment of $\exp\{-w\lambda\}(1+w)^X$.]

⟨ex:Poisson:n-1⟩ 21. Define a linear operator $\mathscr{A}$ via the following:

$$(\mathscr{A}f)(x) := xf(x-1) - \lambda f(x) \qquad \text{for every } f : \mathbb{Z}_+ \to \mathbb{R} \text{ and } x \in \mathbb{Z}_+,$$

where $f(-1) := 0$. Show that $\mathscr{A}$ is a linear mapping from $L^2(\mu)$ to $L^2(\mu)$ and whose adjoint is $\mathscr{D}$, where $(\mathscr{D}f)(x) := \lambda\{f(x+1) - f(x)\}$. Then proceed to verify the following, steps which essentially show that the role of the pair $(\mathscr{D}\,, \mathscr{A})$ is the "Poisson space" analogue of the role of the pair $(D\,, A)$ in the Gauss space:

(a) Prove that $C_{k+1} = \mathscr{A}C_k$ for all $k \geqslant 0$. [Hint: Start with the derivative of $w \mapsto \mathrm{e}^{-w\lambda}(1+w)^x$.]
(b) Prove that $\mathscr{D}C_{k+1} = (k+1)C_k$ for all $k \in \mathbb{Z}_+$.
(c) Prove that $\mathrm{Var}[f(X)] \leqslant \lambda\,\mathrm{E}(|(\mathscr{D}f)(X)|^2)$ for every $f \in L^2(\mu)$.

⟨ex:Poisson:n⟩ 22. Use Problem 21 and the central limit theorem in order to find another proof of the Poincaré inequality for P_1 [Corollary 1.6].

Chapter 4

Heat Flow

1 The Ornstein–Uhlenbeck Operator

The Laplacian $\Delta := D \cdot D := \sum_{i=1}^n \partial^2_{i,i}$ is one of the central differential operators in the analysis of Lebesgue spaces. In a purely analytic sense this is because Δ is the dot product of D with the negative of its adjoint. The analogue of the Laplacian in Gauss space is the generalized differential operator

$$\mathcal{L} := -A \cdot D := -\sum_{j=1}^n A_j D_j \tag{4.1}$$ $\boxed{\text{L}}$

which is called the *Ornstein–Uhlenbeck operator* on the Gauss space. We can think of $\mathcal{L}$ in the form, $(\mathcal{L}\, g)(x) = \sum_{i=1}^n (D^2_{i,i} g)(x) - \sum_{i=1}^n x_i (D_i g)(x)$, or as random variables as

$$(\mathcal{L}\, g)(Z) = \sum_{i=1}^n (D^2_{i,i} g)(Z) - Z \cdot (Dg)(Z).$$

The preceding makes sense as an identity in $L^2(\mathrm{P}_n)$ whenever $g \in \mathbb{D}^{2,2}(\mathrm{P}_n)$ and $Z_i(D_i g)(Z)$ is in $L^2(\mathrm{P}_n)$ for every $1 \leqslant i \leqslant n$. And when $g \in C^2(\mathbb{R}^n)$, then

$$(\mathcal{L}\, g)(x) = (\Delta g)(x) - x \cdot (\nabla g)(x),$$

for every $x \in \mathbb{R}^n$.

Definition 1.1. The *domain of the definition* of $\mathcal{L}$ is

$$\mathrm{Dom}[\mathcal{L}] := \left\{ g \in \mathbb{D}^{2,2}(\mathrm{P}_n) : \mathcal{L}\, g \in L^2(\mathrm{P}_n) \right\} = \left\{ g \in \mathbb{D}^{2,2}(\mathrm{P}_n) : Z \cdot (Dg)(Z) \in L^2(\mathrm{P}_n) \right\}.$$

Since A is the adjoint of D in $L^2(\mathrm{P}_n)$, it follows immediately that the linear operator $\mathcal{L}$ is self adjoint; that is,

$$\mathrm{E}[f\, \mathcal{L}(g)] = \mathrm{E}[\mathcal{L}(f) g] \qquad \text{for every } f, g \in \mathrm{Dom}[\mathcal{L}].$$

Thus, it is helpful to know more about the domain of the operator $\mathcal{L}$. We will shortly identify the domain of $\mathcal{L}$ in a slightly different way than the definition above shows;

see (4.3) below. In the mean time, let us identify large, natural, subsets of $\text{Dom}[\mathcal{L}]$ as follows: If $g \in \mathbb{D}^{1,2q}(\mathrm{P}_n)$ for some $q > 1$, then by Hölder's inequality,

$$\mathrm{E}\left(|Z \cdot (Dg)(Z)|^2\right) \leqslant \left\{\mathrm{E}\left(\|Z\|^{2p}\right)\right\}^{1/p} \left\{\mathrm{E}\left(\|Dg\|^{2q}\right)\right\}^{1/q} \leqslant C_p \|g\|_{1,2q}^{2/q},$$

where $C_p^p := \mathrm{E}(\|Z\|^{2p}) < \infty$ and $p^{-1} + q^{-1} = 1$. Set $\alpha = 2q$ in order to see that

$$\mathbb{D}^{2,\alpha}(\mathrm{P}_n) \subset \mathbb{D}^{2,2}(\mathrm{P}_n) \cap \mathbb{D}^{1,\alpha}(\mathrm{P}_n) \subset \text{Dom}[\mathcal{L}] \quad \text{for all } \alpha > 2.$$

It is not difficult to see how $\mathcal{L}$ acts on Hermite polynomials. The following hashes out the details of that computation.

⟨lem:L:H⟩ **Lemma 1.2.** *$\mathcal{L}\mathcal{H}_k = -|k|\mathcal{H}_k$ for every $k \in \mathbb{Z}_+^n$, where $|k| := \sum_{i=1}^n k_i$.*

Proof. We apply (3.5) [p. 40] to see that $A_j D_j \mathcal{H}_k = k_j \mathcal{H}_k$ for all $k \in \mathbb{Z}_+^n$ and $1 \leqslant j \leqslant n$. Sum over j to finish. □

In other words, for every $k \in \mathbb{Z}_+^n$, the Hermite polynomial $\mathcal{H}_k$ is an eigenfunction of $\mathcal{L}$, with eigenvalue $-|k|$. Since

$$f = \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{E}(f\mathcal{H}_k)}{k!} \mathcal{H}_k \qquad \text{in } L^2(\mathrm{P}_n),$$

it follows readily from Lemma 1.2, and the fact that $\mathcal{L}$ is self adjoint on $L^2(\mathrm{P}_n)$, that for every $f \in \mathbb{D}^{2,2}(\mathrm{P}_n)$,

$$\mathcal{L} f = -\sum_{k \in \mathbb{Z}_+^n} \frac{|k|}{k!} \mathrm{E}(f\mathcal{H}_k)\mathcal{H}_k \qquad \text{in } L^2(\mathrm{P}_n). \tag{4.2}$$ L

Therefore, Theorem 2.1 ensures that

$$\text{Dom}[\mathcal{L}] = \left\{ f \in \mathbb{D}^{2,2}(\mathrm{P}_n) : \sum_{k \in \mathbb{Z}_+^n} \frac{|k|^2}{k!} |\mathrm{E}(f\mathcal{H}_k)|^2 < \infty \right\}. \tag{4.3}$$ Dom:L

We now define the heat flow operator P_t on $L^2(\mathrm{P}_n)$ that corresponds to the Ornstein-Uhlenbeck operator $\mathcal{L}$. There are many reasons for studying this operator, but as we shall soon see one particularly nice reason is that, as a family indexed by time, it "smoothly" interpolates between the random variable and its average. This allows one to produce non-trivial bounds on the variance of the random variable or on the covariance between different random variables. This produces, for example, a new proof of the Poincaré Inequality.

Definition 1.3. The family of heat flow operators is defined, for every $t \geqslant 0$, as

$$P_t f := P(t) f := \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{e}^{-|k|t}}{k!} \mathrm{E}(f\mathcal{H}_k)\mathcal{H}_k, \tag{4.4}$$ P(t)

where the identity holds in $L^2(\mathrm{P}_n)$.

We will now explore some of its properties. But first recall that a sequence $\{X_n\}_{n=1}^\infty$ of elements of $L^2(\mathrm{P}_n)$ is said to converge *weakly in* $L^2(\mathrm{P}_n)$ to $X \in L^2(\mathrm{P}_n)$ if $\lim_{n\to\infty} \mathrm{E}(X_nY) = \mathrm{E}(XY)$ for every $Y \in L^2(\mathrm{P}_n)$.

⟨pr:heat⟩ **Proposition 1.4.** *If $f \in \mathrm{Dom}[\mathcal{L}]$, then $P_t f \in \mathrm{Dom}[\mathcal{L}]$ for all $t > 0$. Moreover, $u(t) := P_t f$ is the unique $L^2(\mathrm{P}_n)$-valued solution to the generalized partial differential equation,*

$$\begin{bmatrix} \dfrac{\partial}{\partial t} u(t) = \mathcal{L}[u(t)], \quad \textit{for all } t > 0, \textit{ subject to} \\ u(0) = f, \end{bmatrix} \tag{4.5}$$ heat

where $(\partial/\partial t)u(t) := \lim_{\varepsilon \downarrow 0}[u(t+\varepsilon) - u(t)]/\varepsilon$ exists weakly in $L^2(\mathrm{P}_n)$.

Definition 1.5. The family $\{P_t\}_{t \geqslant 0}$ is called the *Ornstein–Uhlenbeck semigroup*, and the linear partial differential equation (4.5) is the *heat equation for the Ornstein–Uhlenbeck operator $\mathcal{L}$*.

Proof of Proposition 1.4. For every $g \in L^2(\mathrm{P}_n)$, $t \geqslant 0$, and $\varepsilon > 0$,

$$\begin{aligned} \mathrm{E}\left[g \cdot \frac{u(t+\varepsilon) - u(t)}{\varepsilon}\right] &= \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{e}^{-(t+\varepsilon)|k|} - \mathrm{e}^{-t|k|}}{k!} \mathrm{E}(f\mathcal{H}_k)\,\mathrm{E}(g\mathcal{H}_k) \\ &\to - \sum_{k \in \mathbb{Z}_+^n} \frac{|k|}{k!} \mathrm{E}(f\mathcal{H}_k)\,\mathrm{E}(g\mathcal{H}_k) \qquad \text{as } \varepsilon \downarrow 0, \end{aligned}$$

thanks to (4.3) and the dominated convergence theorem. In particular, the approximate derivative $\{u(t+\varepsilon) - u(t)\}/\varepsilon$ converges weakly in $L^2(\mathrm{P}_n)$ to

$$\frac{\partial}{\partial t} u(t) = \sum_{k \in \mathbb{Z}_+^n} \frac{|k|}{k!} \mathrm{E}(f\mathcal{H}_k)\mathcal{H}_k \qquad \text{as } \varepsilon \downarrow 0,$$

since $f \in \mathrm{Dom}[\mathcal{L}]$. Next we derive the [generalized] PDE (4.5).

By (4.4),

$$\mathrm{E}\left[P_t(f)\mathcal{H}_k\right] = \mathrm{e}^{-|k|t}\,\mathrm{E}\left[f\mathcal{H}_k\right] \qquad \text{for all } t \geqslant 0 \text{ and } k \in \mathbb{Z}_+^n.$$

Therefore,

$$\sum_{k \in \mathbb{Z}_+^n} \frac{|k|}{k!} \left|\mathrm{E}\left[P_t(f)\mathcal{H}_k\right]\right|^2 = \sum_{k \in \mathbb{Z}_+^n} \frac{|k|\mathrm{e}^{-2|k|t}}{k!} \left|\mathrm{E}\left[f\mathcal{H}_k\right]\right|^2 \leqslant \sum_{k \in \mathbb{Z}_+^n} \frac{|k|}{k!} \left|\mathrm{E}\left[f\mathcal{H}_k\right]\right|^2 < \infty.$$

It is possible to show that $P_t f \in \mathbb{D}^{2,2}(\mathrm{P}_n)$ whenever $t > 0$. In fact, $P_t f \in \mathbb{D}^{\infty,2}(\mathrm{P}_n)$ when $t > 0$; see Problem 12, page 52. This, the preceding display, and (4.3) together imply that $P_t f \in \mathrm{Dom}[\mathcal{L}]$ for all $t > 0$.

It is intuitively clear from (4.2) and (4.4) that $\partial P_t f/\partial t = \mathcal{L}\,P_t f$, when u solves (4.5). But since $P_t f$ and $\mathcal{L}\,f$ are not numbers, rather elements of $L^2(\mathrm{P}_n)$, let us write the details to be sure: We know that $u(t) \in L^2(\mathrm{P}_n)$ for every $t \geqslant 0$, and that

$$\mathrm{E}[gu(t)] = \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{e}^{-|k|t}}{k!} \mathrm{E}\left[f\mathcal{H}_k\right] \mathrm{E}\left[g\mathcal{H}_k\right],$$

for all $t \geqslant 0$ and $g \in L^2(\mathrm{P}_n)$. It is not hard to see that the time derivative operator commutes with the sum to yield

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{E}[gu(t)] &= - \sum_{k \in \mathbb{Z}_+^n} \frac{|k|\mathrm{e}^{-|k|t}}{k!} \mathrm{E}\left[f\mathcal{H}_k\right] \mathrm{E}\left[g\mathcal{H}_k\right] \\ &= \mathrm{E}\left[g\,\mathcal{L}[u(t)]\right] \qquad \text{for all } t \geqslant 0, \end{aligned}$$

since $\mathrm{E}(\mathcal{H}_k\, \mathcal{L}[u(t)]) = -|k|(k!)^{-1}\, \mathrm{E}[u(t)\mathcal{H}_k]$ for all $k \geqslant 0$, by (4.2). Thus, u solves the PDE (4.5).

If v is another $L^2(\mathrm{P}_n)$-valued solution to (4.5), then $\phi := u - v$ solves

$$\left[\begin{array}{l} \dfrac{\partial}{\partial t}\phi(t) = \mathcal{L}[\phi(t)], \quad \text{subject to} \\ \phi(0) = 0. \end{array}\right.$$

Project ϕ onto $\mathcal{H}_k$, where $k \in \mathbb{Z}_+^n$ is fixed, in order to find that

$$\frac{\mathrm{d}}{\mathrm{d}t}\, \mathrm{E}\left[\phi(t)\mathcal{H}_k\right] = -|k|\, \mathrm{E}\left[\phi(t)\mathcal{H}_k\right],$$

by (4.2). Since $\mathrm{E}[\phi(0)\mathcal{H}_k] = 0$, it follows that $\mathrm{E}[\phi(t)\mathcal{H}_k] = 0$ for all $t \geqslant 0$ and $k \in \mathbb{Z}_+^n$. The completeness of the Hermite polynomials [Theorem 2.1] ensures that $\phi(t) = 0$ for all $t \geqslant 0$. This implies the remaining uniqueness portion of the proposition. □

⟨OU:Semigrp⟩ **Proposition 1.6.** *The family $\{P_t\}_{t\geqslant 0}$ is a* <u>*symmetric Markov semigroup*</u> *on $L^2(\mathrm{P}_n)$. That is:*

1. *Each P_t is a linear operator from $L^2(\mathrm{P}_n)$ to $L^2(\mathrm{P}_n)$, and $P_t 1 = 1$;*
2. *$P_0 :=$ the identity map. That is, $P_0 f = f$ for all $f \in L^2(\mathrm{P}_n)$;*
3. *Each P_t is self-adjoint on $L^2(\mathrm{P}_n)$. That is,*
$$\mathrm{E}[gP_t(f)] = \mathrm{E}[P_t(g)f] \qquad \text{for all } f, g \in L^2(\mathrm{P}_n) \text{ and } t \geqslant 0;$$
4. *Each $P_t : L^2(\mathrm{P}_n) \to L^2(\mathrm{P}_n)$ is non expansive with constant one. That is,*
$$\mathrm{E}(|P_t f|^2) \leqslant \mathrm{E}(|f|^2) \qquad \text{for all } f \in L^2(\mathrm{P}_n) \text{ and } t \geqslant 0;$$
5. *$\{P_t\}_{\geqslant 0}$ is a semigroup of linear operators. That is,*
$$P_{t+s} = P_t P_s = P_s P_t \qquad \text{for all } s, t \geqslant 0.$$

Finally, P_n is <u>*invariant*</u> *for $\{P_t\}_{t\geqslant 0}$. That is,*

$$\mathrm{E}\left[P_t f\right] = \int P_t f \,\mathrm{d}\mathrm{P}_n = \int f \,\mathrm{d}\mathrm{P}_n = \mathrm{E}(f) = \lim_{s\uparrow\infty} P_s f \qquad \text{a.s. and in } L^2(\mathrm{P}_n).$$

Proof. Parts (1) and (2) are immediate consequences of the definition (4.4) of P_t. [For example, $P_t 1 = 1$ because $\mathcal{H}_0 = 1$.]

Part (3) follows since

$$\mathrm{E}\left[gP_t f\right] = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{e}^{-|k|t}}{k!}\, \mathrm{E}\left[f\mathcal{H}_k\right] \mathrm{E}\left[g\mathcal{H}_k\right],$$

which is clearly a symmetric form in $(f\,, g)$. Part (4) is a consequence of the following calculation.

$$\mathrm{E}\left(|P_t f|^2\right) = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{e}^{-2|k|t}}{k!} \left|\mathrm{E}\left[f\mathcal{H}_k\right]\right|^2 \leqslant \sum_{k\in\mathbb{Z}_+^n} \frac{1}{k!} \left|\mathrm{E}\left[f\mathcal{H}_k\right]\right|^2 = \mathrm{E}\left(|f|^2\right).$$

For part (5) we observe that $\mathrm{E}[P_s f\mathcal{H}_k] = \mathrm{e}^{-|k|s}\, \mathrm{E}[f\mathcal{H}_k]$ for all real numbers $s \geqslant 0$ and integral vectors $k \in \mathbb{Z}_+^n$. Therefore,

$$P_t\left[P_s f\right] = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{e}^{-|k|t}}{k!}\, \mathrm{E}\left[P_s(f)\mathcal{H}_k\right] \mathcal{H}_k = P_{t+s} f.$$

Since $P_{t+s} = P_{s+t}$, this shows also that $P_s P_t = P_t P_s$, and verifies (5).

In order to finish the proof we need to verify the invariance of P_n. First of all note that $1(x) := 1$ is in $L^2(\mathrm{P}_n)$. Therefore, $\mathrm{E}(P_t f) = \mathrm{E}[P_t(1)f] = \mathrm{E}(f)$ since f is self adjoint on $L^2(\mathrm{P}_n)$. Now we need to prove the convergence of $P_s f$ as $s \to \infty$. By (4.4),

$$P_t f = \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{e}^{-|k|t}}{k!} \mathrm{E}\left[f \mathcal{H}_k\right] \mathcal{H}_k \qquad \text{a.s.,} \tag{4.6}$$ P(t)f:H

where the convergence holds in $L^2(\mathrm{P}_n)$ (for each fixed t).

The Cauchy–Schwarz inequality yields $|\langle f\,, \mathcal{H}_k\rangle_{L^2(\mathrm{P}_n)}| \leqslant \|f\|_{L^2(\mathrm{P}_n)}$, valid for all $k \in \mathbb{Z}_+^n$. Therefore, the identity $\|\mathcal{H}_k\|_{L^2(\mathrm{P}_n)} = \sqrt{k!}$ and the Minkowski inequality together imply that

$$\left\| \sup_{t \geqslant 0} \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{e}^{-|k|t}}{k!} \left|\mathrm{E}\left[f \mathcal{H}_k\right] \mathcal{H}_k\right| \right\|_{L^2(\mathrm{P}_n)} \leqslant \|f\|_{L^2(\mathrm{P}_n)} \sum_{k \in \mathbb{Z}_+^n} \frac{1}{\sqrt{k!}} < \infty.$$

In particular, the sum in (4.6) also converges absolutely, uniformly in $t \geqslant 0$, with P_n-probability one. Consequently,

$$\lim_{t \uparrow \infty} P_t f = \sum_{k \in \mathbb{Z}_+^n} \lim_{t \to \infty} \frac{\mathrm{e}^{-|k|t}}{k!} \mathrm{E}\left[f \mathcal{H}_k\right] \mathcal{H}_k = \mathrm{E}\left[f \mathcal{H}_0\right],$$

almost surely. The final quantity is equal to $\mathrm{E}(f)$, as desired. □

2 Mehler's Formula

The heat equation (4.5) for the OU operator $\mathcal{L}$ is just the initial-value problem,

$$\begin{bmatrix} \dfrac{\partial}{\partial t} u(t\,,x) = (\Delta u)(t\,,x) - x \cdot (\nabla u)(t\,,x) & [t > 0,\ x \in \mathbb{R}^n], \\ u(0\,,x) = f(x) & [x \in \mathbb{R}^n], \end{bmatrix}$$

but written out in an infinite-dimensional manner. As such, it can be solved by other, more elementary, methods as well. We have taken this route in order to introduce the OU semigroup $\{P_t\}_{t \geqslant 0}$ and the associated OU operator $\mathcal{L}$. These objects will play a central role in Gaussian analysis, more so than does the heat equation for the Laplacian. Still, it might be good to know that every $L^2(\mathrm{P}_n)$-valued solution is also a classical solution when, for example, f is in $C_0^2(\mathrm{P}_n)$. Among many other things, this fact follows immediately from the following interesting formula and the dominated convergence theorem.

⟨Mehler⟩ **Theorem 2.1** (Mehler's Formula). *If $f \in L^2(\mathrm{P}_n)$ and $t \geqslant 0$, then*

$$(P_t f)(x) = \mathrm{E}\left[f\left(\mathrm{e}^{-t} x + \sqrt{1 - \mathrm{e}^{-2t}}\, Z\right)\right],$$

for almost every $x \in \mathbb{R}^n$.

?⟨rem:Mehler⟩? *Remark* 2.2. One of the many by-products of Mehler's formula is the fact that each mapping $f \mapsto P_t f$ is a Bochner integral; in particular, every P_t satisfies the "Cauchy–Schwarz inequality," which is stronger than the non-expansiveness of P_t:

$$|P_t f|^2 \leqslant P_t(f^2) \qquad \text{a.s. for all } t \geqslant 0 \text{ and } f \in L^2(\mathrm{P}_n).$$

Proof of Mehler's Formula. We prove the result for $n = 1$ since the notation is simpler in that case. The general case is proved by extending the argument directly to higher dimensions.

By density it suffices to prove the result for all $f \in C_0^\infty(\mathrm{P}_1)$. Define for such functions f and $t \geqslant 0$,

$$(T_t f)(x) := \mathrm{E}\left[f\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\, Z\right)\right],$$

for every $x \in \mathbb{R}$. Both sides are C^∞ functions in either variable t and x [dominated convergence]. Our goal is to prove that $T_t f = P_t f$ for all $t \geqslant 0$. This will complete the proof. Note that

$$\begin{aligned}(T_t f)(x) &= \int_{-\infty}^{\infty} f\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\, z\right)\gamma_1(z)\,\mathrm{d}z \\ &= \int_{-\infty}^{\infty} f(y)\gamma_1\left(\frac{y - \mathrm{e}^{-t}x}{\sqrt{1-\mathrm{e}^{-2t}}}\right)\mathrm{d}y.\end{aligned} \tag{4.7}$$ T_tf

Since $f \in C_0^\infty(\mathrm{P}_1)$, we can differentiate under the integral any number of times we want in order to see that $\partial(T_t f)/\partial t = \mathcal{L}(T_t f)$, after a few lines of calculus applied to the function γ_1. Since $T_0 f = f$, the uniqueness portion of Proposition 1.4 implies that $T_t f = P_t f$ for all $t \geqslant 0$. □

3 A Covariance Formula

One of the highlights of our analysis so far is that it leads to an explicit formula for $\mathrm{Cov}(f\,, g)$ for a large number of nice functions f and g. That formula will take a quite different form than the one in Corollary 2.3 (p. 43), and will have novel uses as a result. Before we discuss that formula, let us observe the following.

⟨lem:DP:PD⟩ **Lemma 3.1.** *For all $t \geqslant 0$ and $1 \leqslant j \leqslant n$,*

$$D_j P_t = \mathrm{e}^{-t} P_t D_j \qquad \text{and} \qquad A_j P_t = \mathrm{e}^{t} P_t A_j.$$

Consequently, $\mathcal{L}(P_t f) = P_t(\mathcal{L}\, f)$, also.

Proof. First consider the case that $n = 1$. In that case,

$$P_t f = \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-kt}}{k!}\,\mathrm{E}(f H_k) H_k,$$

for all $f \in L^2(\mathrm{P}_1)$. Therefore, whenever $f \in \mathbb{D}^{1,2}(\mathrm{P}_1)$,

$$\begin{aligned} DP_t f &= \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-kt}}{k!}\,\mathrm{E}(f H_k) DH_k = \sum_{k=0}^{\infty} \frac{k\mathrm{e}^{-kt}}{k!}\,\mathrm{E}(f H_k) H_{k-1} \\ &= \mathrm{e}^{-t}\sum_{k=0}^{\infty} \frac{\mathrm{e}^{-kt}}{k!}\,\mathrm{E}\left[f H_{k+1}\right] H_k,\end{aligned} \tag{4.8}$$ lala1

by (3.5) [page 40]. Similarly, $\mathrm{E}[D(f)H_k] = \mathrm{E}[fA(H_k)] = \mathrm{E}[fH_{k+1}]$ for all $k \geqslant 0$. Therefore,

$$P_t Df = \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-kt}}{k!}\,\mathrm{E}\left[f H_{k+1}\right] H_k.$$

Match this expression with (4.8) in order to see that $DP_t = \exp(-t)P_tD$ when $n = 1$. A similar argument shows that $AP_t = \exp(t)P_tA$ in this case as well.

When $n \geqslant 1$ and $f = f_1 \otimes \cdots \otimes f_n$ for $f_1, \ldots, f_n \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ (i.e. $f(x_1, \ldots, x_n) = f_1(x_1)\ldots f_n(x_n)$), we can check using obvious notation that

$$(D_jP_tf)(x) = \prod_{\substack{1\leqslant q\leqslant n\\ q\neq j}} \left(P_t^{(q)}f_q\right)(x_q) \times (D_jP_t^{(j)}f_j)(x_j) = \mathrm{e}^{-t}(P_tD_jf)(x),$$

by the one-dimensional part of the proof that we just developed. Since every $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ can be approximated arbitrarily well by functions of the form $f_1 \otimes \cdots \otimes f_n$, where $f_j \in \mathbb{D}^{1,2}(\mathrm{P}_1)$, it follows that $D_jP_t = \exp(-t)P_tD_j$ on $\mathbb{D}^{1,2}(\mathrm{P}_n)$.

Similarly, one proves that $A_jP_t = \exp(t)P_tA_j$ in general.

To finish note that

$$\mathcal{L}\,P_t = -\sum_{j=1}^n A_jD_jP_t = -\mathrm{e}^{-t}\sum_{j=1}^n A_jP_tD_j = \sum_{j=1}^n P_tA_jD_j = \mathcal{L}\,P_t.$$

This completes the proof. □

Lemma 3.1 has the following important corollary.

⟨pr:Cov⟩ **Proposition 3.2.** *For every $f, g \in \mathbb{D}^{1,2}(\mathrm{P}_n)$,*

$$\mathrm{Cov}(f\,, g) = \int_0^\infty \mathrm{e}^{-t}\,\mathrm{E}\left[(Df)\cdot(P_tDg)\right]\mathrm{d}t,$$

where $P_tDg = (P_tD_1g\,, \ldots, P_tD_ng)$.

Proof. Recall from Proposition 1.6 that $P_tg \to \mathrm{E}[g]$ in $L^2(\mathrm{P}_n)$ as $t \to \infty$, and $P_0g = g$. Therefore,

$$g(x) - \mathrm{E}[g] = -\int_0^\infty \frac{\partial}{\partial t}(P_tg)(x)\,\mathrm{d}t,$$

where the identity is understood to hold in $L^2(\mathrm{P}_n)$, and the integral converges in $L^2(\mathrm{P}_n)$ as well. That is, $\int_0^\infty(\partial/\partial t)(P_tg)\,\mathrm{d}x = \lim_{N\to\infty}\int_0^N(\partial/\partial t)(P_tg)\,\mathrm{d}x$ in $L^2(\mathrm{P}_n)$. Therefore, by Fubini's theorem,

$$\begin{aligned}\mathrm{Cov}(f\,, g) = \mathrm{E}\left[f(Z)(g(Z) - \mathrm{E}[g])\right] &= -\int_0^\infty \mathrm{E}\left[f(Z)\frac{\partial}{\partial t}(P_tg)(Z)\right]\mathrm{d}t\\ &= -\int_0^\infty \mathrm{E}\left[f(Z)(\mathcal{L}\,P_tg)(Z)\right]\mathrm{d}t,\end{aligned}$$

since P_tg solves the heat equation for the operator $\mathcal{L}$. Next we may observe that, since $\mathcal{L} = -A\cdot D$ and A_j is the adjoint to D_j,

$$\begin{aligned}\mathrm{E}\left[f(Z)(\mathcal{L}\,P_tg)(Z)\right] &= -\sum_{j=1}^n \mathrm{E}\left[(D_jf)(Z)(D_jP_tg)(Z)\right]\\ &= -\mathrm{e}^{-t}\sum_{j=1}^n \mathrm{E}\left[(D_jf)(Z)(P_tD_jg)(Z)\right].\end{aligned}$$

We have appealed to Lemma 3.1 in the second line. This concludes the proof. □

Let us conclude with a quick application of Proposition 3.2.

A Second Proof of the Poincaré Inequality. For every $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $t \geqslant 0$, the Cauchy-Schwarz inequality implies that

$$|\mathrm{E}\left[(Df)\cdot(P_t Df)\right]| \leqslant \left|\mathrm{E}\left(\|Df\|^2\right)\mathrm{E}\left(\|P_t Df\|^2\right)\right]^{1/2},$$

and the right-hand side is at most $\mathrm{E}(\|Df\|^2)$ since P_t is non-expanding on $L^2(\mathrm{P}_n)$ [Proposition 1.6]. The Poincaré inequality now follows from Proposition 3.2. □

4 The Resolvent of the OU Semigroup

The classical theory of linear semigroups tells us that it is frequently better to study a semigroup of linear operators via its "resolvent." In the present context, this leads us to the following.

Definition 4.1. The *resolvent* of the Ornstein–Uhlenbeck [or OU] semigroup $\{P_t\}_{t\geqslant 0}$ is the family $\{R_\lambda\}_{\lambda>0}$ of linear operators defined via

$$(R_\lambda f)(x) := \int_0^\infty \mathrm{e}^{-\lambda t}(P_t f)(x)\,\mathrm{d}t, \tag{4.9}$$ R

for all bounded and measurable functions $f : \mathbb{R}^n \to \mathbb{R}$ and all $\lambda > 0$.

Informally speaking, $R(\lambda) := \int_0^\infty \exp(-\lambda t)P(t)\,\mathrm{d}t$ defines the Laplace transform of the semigroup $\{P(t)\}_{t\geqslant 0}$, and knowing R should in principle be the same as knowing P. We will see soon that this is the case. But first let us define the resolvent not pointwise, as we just did, but as an element of the Hilbert space $L^2(\mathrm{P}_n)$.

According to Mehler's formula [Theorem 2.1], if f is bounded and measurable, then $P_t f$ is also; in fact, $\sup_x |(P_t f)(x)| \leqslant \sup_x |f(x)|$, whence the integral in (4.9) converges absolutely, uniformly in $x \in \mathbb{R}^n$. One can extend the domain of the definition of R_λ further by standard means. In fact, because P_t is non expensive on $L^2(\mathrm{P}_n)$ [Proposition 1.6],

$$\mathrm{E}\left(|P_t f|^2\right) \leqslant \mathrm{E}(f^2), \quad \text{whence} \quad \mathrm{E}\left(|R_\lambda f|^2\right) \leqslant \lambda^{-2}\,\mathrm{E}(f^2),$$

for all bounded functions $f \in L^2(\mathrm{P}_n)$ and every $t, \lambda > 0$. If $f \in L^2(\mathrm{P}_n)$ then we can find bounded functions $f_1, f_2, \ldots \in L^2(\mathrm{P}_n)$ such that $\mathrm{E}(|f_\ell - f|^2) \leqslant 2^{-\ell}$ for all $\ell \geqslant 1$, and hence the preceding inequality shows that

$$\mathrm{E}\left(|R_\lambda f_m - R_\lambda f_\ell|^2\right) \leqslant \lambda^{-2}\,\mathrm{E}\left(|f_m - f_\ell|^2\right) \leqslant \frac{2^{-\ell} + 2^{-m}}{\lambda^2},$$

for all $m, \ell \geqslant 1$. Therefore, $\ell \mapsto R_\lambda f_\ell$ is a Cauchy sequence in $L^2(\mathrm{P}_n)$ and hence $R_\lambda f := \lim_{\ell\to\infty} R_\lambda f_\ell$ is a well-defined limit in $L^2(\mathrm{P}_n)$. Since every P_t is non expansive on $L^2(\mathrm{P}_n)$, it follows similarly that (4.9) holds a.s. for all $f \in L^2(\mathrm{P}_n)$ and $\lambda > 0$. Let us pause and record these observations before we go further.

?⟨pr:R⟩? **Proposition 4.2.** *For every $\lambda > 0$, R_λ is a bounded continuous linear map from $L^2(\mathrm{P}_n)$ to $L^2(\mathrm{P}_n)$, with operator norm $\leqslant \lambda^{-2}$. Finally, (4.9) holds a.s. for all $f \in L^2(\mathrm{P}_n)$ and $\lambda > 0$, and*

$$R_\lambda f = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{E}(f\mathcal{H}_k)}{k!(\lambda + |k|)}\,\mathcal{H}_k \qquad \textit{a.s.}, \tag{4.10}$$ R:H

where the sum converges in $L^2(\mathrm{P}_n)$.

Proof. The only unproved part of the assertion is the representation (4.10) of $R_\lambda f$ in terms of Hermite polynomials.

If $f \in L^2(\mathrm{P}_n)$, then $R_\lambda f \in L^2(\mathrm{P}_n)$ for all $\lambda > 0$, and Theorem 2.1 ensures that

$$R_\lambda f = \sum_{k \in \mathbb{Z}_+^n} \frac{1}{k!} \mathrm{E}\left[(R_\lambda f)\mathcal{H}_k\right] \mathcal{H}_k.$$

By Fubini's theorem, (4.9), and (4.4),

$$\mathrm{E}\left[(R_\lambda f)\mathcal{H}_k\right] = \int_0^\infty \mathrm{e}^{-\lambda t}\, \mathrm{E}\left[(P_t f)\mathcal{H}_k\right] \mathrm{d}t = \int_0^\infty \mathrm{e}^{-(|k|+\lambda)t}\, \mathrm{E}(f\mathcal{H}_k)\mathrm{d}t = \frac{\mathrm{E}(f\mathcal{H}_k)}{\lambda + |k|},$$

for all $k \in \mathbb{Z}_+^n$, $t \geqslant 0$, and $\lambda > 0$. Multiply the preceding by $\mathcal{H}_k/k!$ and sum over $k \in \mathbb{Z}_+^n$ to finish. ☐

?⟨pr:RE⟩? **Proposition 4.3** (The Resolvent Equation)**.** *For all $f \in L^2(\mathrm{P}_n)$, and for every distinct pair $\alpha, \lambda > 0$,*

$$R_\lambda R_\alpha f = R_\alpha R_\lambda f = -\frac{R_\lambda f - R_\alpha f}{\lambda - \alpha} \qquad \textit{a.s.} \tag{4.11}$$ RE

Proof. We apply the Fubini theorem and (4.9) a few times back-to-back as follows: Almost surely,

$$\begin{aligned} R_\lambda R_\alpha f &= \int_0^\infty \mathrm{e}^{-\lambda t} P_t(R_\alpha f)\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-\lambda t} P_t \left(\int_0^\infty \mathrm{e}^{-\alpha s}\ P_s f\,\mathrm{d}s \right) \mathrm{d}t \\ &= \int_0^\infty \mathrm{e}^{-\lambda t}\,\mathrm{d}t \int_0^\infty \mathrm{e}^{-\alpha s}\,\mathrm{d}s\ P_{t+s} f = \int_0^\infty \mathrm{e}^{-(\lambda-\alpha)t}\,\mathrm{d}t \int_t^\infty \mathrm{e}^{-\alpha r}\mathrm{d}r\ P_r f \\ &= \int_0^\infty \mathrm{e}^{-\alpha r} P_r f\,\mathrm{d}r \int_0^r \mathrm{e}^{-(\lambda-\alpha)t}\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-\alpha r} P_r(f) \left(\frac{1 - \mathrm{e}^{-(\lambda-\alpha)r}}{\lambda - \alpha} \right) \mathrm{d}r. \end{aligned}$$

Reorganize the integral to finish. ☐

Eq. (4.11) is called the *resolvent equation*, and readily implies the following.

⟨co:RE⟩ **Corollary 4.4.** *For every $\lambda > 0$, R_λ maps $L^2(\mathrm{P}_n)$ bijectively onto its range*

$$R_\lambda\left(L^2(\mathrm{P}_n)\right) := \left\{ R_\lambda f : f \in L^2(\mathrm{P}_n) \right\}.$$

The preceding range does not depend on $\lambda > 0$. Moreover, the range is dense in $L^2(\mathrm{P}_n)$; in fact, $\lim_{\lambda\to\infty} \lambda R_\lambda f = f$ in $L^2(\mathrm{P}_n)$ for every $f \in L^2(\mathrm{P}_n)$.

Proof. First, we observe that $x \mapsto (R_\lambda f)(x)$ is a.s. equal to a continuous function for all $f \in L^2(\mathrm{P}_n)$ and $\lambda > 0$. This follows from Mehler's formula [Proposition 2.1, page 59] and the dominated convergence theorem. Therefore, we can always redefine it so that $R_\lambda f$ is continuous. In particular, if $R_\lambda f = 0$ a.s. for some $\lambda > 0$, then $R_\lambda f \equiv 0$ and hence $R_\alpha f \equiv 0$ for all $\alpha > 0$ thanks to the resolvent equation. The uniqueness theorem for Laplace transforms now shows that if $R_\lambda f = 0$ a.s. for some $\lambda > 0$ then $f = 0$ a.s. By linearity we find that if $R_\lambda f = R_\lambda g$ a.s. for some $f, g \in L^2(\mathrm{P}_n)$ and $\lambda > 0$, then $f = g$ a.s. Consequently, R_λ is a one-to-one and onto map from $L^2(\mathrm{P}_n)$ to its range $R_\lambda(L^2(\mathrm{P}_n))$.

Next, let us suppose that g is in the range of R_α; that is, $g = R_\alpha f$ for some $f \in L^2(\mathrm{P}_n)$. By the resolvent equation,

$$R_\lambda g = -\frac{R_\lambda f - g}{\lambda - \alpha} \quad \Rightarrow \quad g = (\lambda - \alpha)R_\lambda g - R_\lambda f = R_\lambda h,$$

for $h = (\lambda - \alpha)g - f$. This shows that g is in the range of R_λ, whence $R_\alpha(L^2(\mathrm{P}_n)) \subset R_\lambda(L^2(\mathrm{P}_n))$. Reverse the roles of α and λ to see that $R_\lambda(L^2(\mathrm{P}_n))$ does not depend on $\lambda > 0$.

Finally, we verify the density assertion. Let $f \in L^2(\mathrm{P}_n)$, and recall [Proposition 1.6, page 58] that $P_t f \to f$ in $L^2(\mathrm{P}_n)$ as $t \downarrow 0$. By this and the dominated convergence theorem,

$$\lambda R_\lambda f = \lambda \int_0^\infty \mathrm{e}^{-\lambda t} P_t f \, \mathrm{d}t = \int_0^\infty \mathrm{e}^{-s} P_{s/\lambda} f \, \mathrm{d}s \to f \qquad \text{in } L^2(\mathrm{P}_n) \text{ as } \lambda \uparrow \infty.$$

This implies that the range of the resolvent in dense in $L^2(\mathrm{P}_n)$ because it proves that for all $\varepsilon > 0$ there exists an elements of the range

$$R_1(L^2(\mathrm{P}_n)) = \bigcup_{\alpha > 0} R_\alpha(L^2(\mathrm{P}_n)),$$

namely $\lambda R_\lambda f = R_\lambda(\lambda f)$ for a sufficiently large λ, that is close to within ε of f in the $L^2(\mathrm{P}_n)$ norm. □

Corollary 4.4 tells us that we can in principle compute the entire semigroup $\{P_t\}_{t \geqslant 0}$ from the operator R_λ for a given $\lambda > 0$. And of course the converse is also true by (4.9). From now on we will consider $\lambda = 1$ only.

Definition 4.5. If $f \in L^2(P_n)$ then $R_1 f$ is called the *one-potential* of f. The linear operator R_1 is also known as the [one-] *potential operator*.

⟨lem:R⟩ **Lemma 4.6.** *R_1 is a non-expansive, self-adjoint linear operator on $L^2(\mathrm{P}_n)$.*

Proof. Linearity is obvious. We need to prove that for all $f, g \in L^2(\mathrm{P}_n)$:

1. $\mathrm{E}(|R_1 f|^2) \leqslant \mathrm{E}(f^2)$; and
2. $\mathrm{E}[g(R_1 f)] = \mathrm{E}[(R_1 g) f]$.

Both of these properties follow from the corresponding properties of the semigroup $\{P_t\}_{t \geqslant 0}$, and from (4.9). □

The potential operator arises naturally in a number of ways. For example, Proposition 3.2 can be recast in terms of the potential operator as follows:

⟨th:Cov:1⟩ **Theorem 4.7** (Houdré, Pérez-Abreu, and Surgailis, XXX)**.** *For every $f, g \in \mathbb{D}^{1,2}(\mathrm{P}_n)$,*

$$\mathrm{Cov}(f\,, g) = \mathrm{E}\left[\langle Df\,, Dg\rangle_{R_1}\right],$$

where

$$\langle p\,, q\rangle_{R_1} := p \cdot (R_1 q) \qquad \textit{for all } p, q \in L^2(\mathrm{P}_n \times \chi_n), \tag{4.12}$$ energy

and $R_1 q = (R_1 q_1\,, \ldots, R_1 q_n) = \int_0^\infty \exp(-t)(P_t q)\,\mathrm{d}t$.

The bilinear symmetric form $(f\,, g) \mapsto \mathrm{E}[\langle Df\,, Dg\rangle_{R_1}]$ is known as a *Dirichlet form*, and the integral $\mathrm{E}[\langle Df\,, Dg\rangle_{R_1}]$ is called the *Dirichlet energy* between f and g. Thus, Theorem 4.7 is another way to state that the covariance between the random variables $f(Z)$ and $g(Z)$ is the Dirichlet energy between the functions f and g.

Let us mention another property of the potential operator.

⟨pr:R1:L⟩ **Proposition 4.8** (Hille, XXX and Yoshida, XXX)**.** *$\mathrm{Dom}[\mathcal{L}] = R_1(L^2(\mathrm{P}_n)) \cap \mathbb{D}^{2,2}(\mathrm{P}_n)$ and*

$$\mathcal{L} f = f - R_1^{-1} f \qquad \textit{a.s. for all } f \in \mathrm{Dom}[\mathcal{L}].$$

Let I denote the identity operator on $L^2(\mathrm{P}_n)$; that is, $I(f) := f$ for all $f \in L^2(\mathrm{P}_n)$. Then, Proposition 4.8 essentially says that $\mathcal{L} = I - R_1^{-1}$ and hence also $R_1 = (I - \mathcal{L})^{-1}$.

Proof. First, choose and fix an arbitrary $f \in R_1(L^2(\mathrm{P}_n))$. Corollary 4.4 ensures that there exists a unique $g \in L^2(\mathrm{P}_n)$ such that $f = R_1 g$, equivalently $g = R_1^{-1} f$. Therefore, by (4.10) and the orthogonality of Hermite polynomials,

$$\mathrm{E}[f\mathcal{H}_k] = \mathrm{E}\left[(R_1 g)\mathcal{H}_k\right] = \frac{\mathrm{E}(g\mathcal{H}_k)}{1 + |k|} \quad \text{for all } k \in \mathbb{Z}^n.$$

It follows that

$$R_1^{-1} f = g = \sum_{k \in \mathbb{Z}^n_+} \frac{\mathrm{E}[g\mathcal{H}_k]}{k!}\,\mathcal{H}_k = \sum_{k \in \mathbb{Z}^n_+} \frac{1 + |k|}{k!}\,\mathrm{E}[f\mathcal{H}_k]\,\mathcal{H}_k. \tag{4.13}$$ R11

Conversely, the preceding infinite sum defines an element of $L^2(\mathrm{P}_n)$ as long as it converges in $L^2(\mathrm{P}_n)$, which can happen if and only if $\sum_{k \in \mathbb{Z}^n_+} |k|^2 |\,\mathrm{E}(f\mathcal{H}_k)|^2/k! < \infty$. We can summarize these remarks as follows:

$$R_1\left(L^2(\mathrm{P}_n)\right) = \left\{ f \in L^2(\mathrm{P}_n) : \sum_{k \in \mathbb{Z}^n_+} \frac{|k|^2}{k!}\,|\mathrm{E}(f\mathcal{H}_k)|^2 < \infty \right\}.$$

Therefore, (4.3) implies the first assertion of the theorem. Namely, that $R_1(L^2(\mathrm{P}_n)) \cap \mathbb{D}^{2,2}(\mathrm{P}_n) = \mathrm{Dom}[\mathcal{L}]$.

The identity $\mathcal{L} f = f - R_1^{-1} f$ is a consequence of (4.2) and (4.13). □

Problems

1. For every $a \in \mathbb{R}^n$ consider the function $f_a \in L^2(\mathrm{P}_n)$ defined by

$$f_a(x) = \exp\left(a \cdot x - \frac{\|a\|^2}{2}\right) \qquad \text{for all } x \in \mathbb{R}^n.$$

Verify that $P_t f_a = f_{a\exp(-t)}$, and conclude from this property that P_t is non-expansive on f_a and that $P_t f_a$ converges to $\mathrm{E}[f_a]$ as $t \to \infty$ both almost surely and in $L^2(\mathrm{P}_n)$.

2. Do the calculations to complete the proof of Theorem 2.1.

⟨ex:p⟩ 3. For all $x, y \in \mathbb{R}^n$ and $t, \lambda > 0$ define

$$p(t\,, x\,, y) := \gamma_n\left(\frac{y - \mathrm{e}^{-t}x}{\sqrt{1-\mathrm{e}^{-2t}}}\right) \quad \text{and} \quad r_\lambda(x\,, y) := \int_0^1 r^{1-\lambda}\gamma_n\left(\frac{y - rx}{\sqrt{1-r^2}}\right)\mathrm{d}r.$$

Prove that

$$(P_t f)(x) = \int_{\mathbb{R}^n} p(t\,, x\,, y) f(y)\,\mathrm{d}y \quad \text{and} \quad (R_\lambda f)(x) = \int_{\mathbb{R}^n} r_\lambda(x\,, y) f(y)\,\mathrm{d}y,$$

for all $t, \lambda > 0$, $x \in \mathbb{R}^n$, and measurable $f : \mathbb{R}^n \to \mathbb{R}_+$. [Hint: Consult (4.7) on page 60.] Prove also that p is the *fundamental solution* to the heat operator $\mathcal{H} := (\partial/\partial t) - \mathcal{L}$. That is, prove that for every $x \in \mathbb{R}^n$, the function $(t\,, y) \mapsto p(t\,, x\,, y)$ satisfies the linear partial differential equation,

$$\frac{\partial p(t\,, x\,, y)}{\partial t} = \sum_{i=1}^n \frac{\partial^2 p(t\,, x\,, y)}{\partial y_i^2} - \sum_{i=1}^n y_i \frac{\partial p(t\,, x\,, y)}{\partial y_i}$$

on $(0\,,\infty) \times \mathbb{R}^n \times \mathbb{R}^n$, subject to the following initial condition: As $t \downarrow 0$, $\int_{\mathbb{R}^n} p(t\,, x\,, y) f(y)\,\mathrm{d}y \to f(x)$ for all smooth, bounded functions $f : \mathbb{R}^n \to \mathbb{R}$.

4. Establish the following properties of the Ornstein–Uhlenbeck semigroup:
   (a) If $P_s f$ is bounded and continuous for some $s \geqslant 0$, then so is $P_{t+s} f$ for every $t \geqslant 0$.
   (b) If $f : \mathbb{R}^n \to \mathbb{R}$ is bounded and measurable, then $P_t f$ is bounded and continuous for every $t > 0$. [Hint: Consult Problem 3.]
   (c) If $f : \mathbb{R}^n \to \mathbb{R}$ is bounded and measurable, then $(P_{t+s} f)(x) = [P_t(P_s f)](x)$ for every $s, t \geqslant 0$ and $x \in \mathbb{R}^n$.

5. Prove that if $f \in C_0^\infty(\mathrm{P}_n)$, then

$$(R_1^{-1} f)(x) = f(x) + x \cdot (\nabla f)(x) - (\Delta f)(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

6. Extend Lemma 4.6 by showing that, for every $\lambda > 0$, the linear operator λR_λ is non-expansive, and self-adjoint on $L^2(\mathrm{P}_n)$.

7. Suppose $f : \mathbb{R}^n \to \mathbb{R}_+$ is measurable and bounded.
   (a) Prove that $R_\lambda f$ is bounded and continuous for every $\lambda > 0$.
   (b) A function $h : \mathbb{R}^n \to \mathbb{R}_+$ is said to be λ-*excessive* for some $\lambda > 0$ if $\exp\{-\lambda t\}(P_t h)(x) \uparrow h(x)$ as $t \downarrow 0$ for every $x \in \mathbb{R}^n$. Prove that $R_\lambda f$ is λ-excessive for every $\lambda > 0$.
   (c) Verify that non-negative, bounded elements of $\mathrm{Dom}[\mathcal{L}]$ are λ-excessive for every $\lambda > 0$.
   (d) Suppose $f, g \in \mathrm{Dom}[\mathcal{L}]$ satisfy $f \geqslant g$ almost everywhere on $\mathbb{R}^n$. Prove that $f \geqslant g$ pointwise.

8. Suppose $g \in \mathbb{D}^{1,2}(\mathrm{P}_n)$.
   (a) Prove that $\mathrm{E}[Z_j g(Z)] = \mathrm{E}[D_j g]$ for every $g \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $i = 1, \ldots, n$.
   (b) Use the above in order to prove that $\mathrm{E}[X_i g(X)] = \sum_{j=1}^n \Gamma_{i,j}\, \mathrm{E}[D_j g]$ whenever X is distributed as $\mathrm{N}_n(0\,, \Gamma)$ for a positive-definite matrix Γ.
   (c) Extend the preceding to the case that Γ is an arbitrary covariance matrix.
   (d) Use the preceding parts, together with induction, in order to give an alternative proof of the Isserlis formula (page 20).
9. (**This problem requires some background in Itô calculus.**)
   Here is a way of proving Mehler's formula [Proposition 2.1, page 59] using stochastic calculus: Let X_t be an Itô diffusion in $\mathbb{R}^n$ which satisfies the stochastic differential equation,

$$\mathrm{d}X_t = -X_t\, \mathrm{d}t + \sqrt{2}\, \mathrm{d}B_t \quad \text{subject to} \quad X_0 = x,$$

   where $\{B_t\}_{t \geqslant 0}$ denotes a standard, n-dimensional Brownian motion. The process $\{X_t\}_{t \geqslant 0}$ is called an *Ornstein–Uhlenbeck* process on $\mathbb{R}^n$.
   (a) Use Itô's formula to compute $\mathrm{d}(\mathrm{e}^t X_t)$. Derive an integral equation for X_t, and then show that at each time t, the random vector X_t has a normal distribution with mean $\mathrm{e}^{-t}x$ and covariance matrix $\sqrt{1 - \mathrm{e}^{-2t}}$ times the $n \times n$ identity matrix. Conclude from this that

$$\mathrm{E}\left[f\left(x\mathrm{e}^{-t} + \sqrt{1 - \mathrm{e}^{-2t}}\, Z\right)\right] = \mathrm{E}_x\left[f(X_t)\right],$$

   where the subscript x, under the second expectation E_x, is added there to remind us that $X_0 = x$.
   (b) Use Itô's formula and Lemma 1.2 to show that $\mathrm{E}_x[\mathcal{H}(X_t)] = \mathcal{H}(x)\mathrm{e}^{-|k|t}$.
   (c) Use the Hermite expansion of f [Corollary 2.2, p. 43] and part (b) in order to conclude that

$$\mathrm{E}_x\left[f(X_t)\right] = \sum_{k \in \mathbb{Z}_+^n} \frac{\mathrm{E}[f\mathcal{H}_k]}{k!} \mathrm{e}^{-|k|t} \mathcal{H}_k(x) = (P_t f)(x).$$

   Deduce Mehler's formula from this and part (a).

Chapter 5

Integration by Parts and Its Applications

The following is an immediate consequence of Theorem 4.7 [page 64] and the chain rule [Lemma 1.7, p. 31].

⟨th:IbP⟩ **Theorem 0.1.** *For every* $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ *and* $\varphi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$,

$$\mathrm{Cov}(f\,,\varphi(f)) = \mathrm{E}\left[(D\varphi)(f) \times \langle Df\,,Df\rangle_{R_1}\right],$$

where $\langle\,\cdot\,,\cdot\rangle_{R_1}$ *was defined in* (4.12)*, page 64.*

Theorems 4.7 and its corollary Theorem 0.1 are integration by parts formula on Gauss space. In order to see this claim more clearly, suppose for example that $\varphi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ and $X := f(Z)$ where $f \in C_0^1(\mathrm{P}_n)$. Then Theorem 0.1 reduces to the assertion that

$$\mathrm{E}\left[X\varphi(X)\right] = \mathrm{E}(X)\,\mathrm{E}[\varphi(X)] + \mathrm{E}\left[\varphi'(X)\langle DX\,,DX\rangle_{R_1}\right], \tag{5.1}$$ eq:IbP

where we are writing $\varphi'(X)$ in place of the more appropriate notation, $(D\varphi)(X)$ for the sake of clarity.[1] Thus, for example, we find that the "Dirichlet form," $\mathrm{E}[\langle DX\,,DX\rangle_{R_1}]$ is simply equal to the variance of $X \in \mathbb{D}^{1,2}(\mathrm{P}_n)$. In this chapter we explore some of the deeper consequences of these integration by parts results.

1 Concentration of Measure

Concentration of measure is an important and well-studied phenomenon that has broad applications in measure theory, combinatorics, probability., theoretical computer science, etc. In this book we only study a small but important subclass of the theory: The concentration phenomenon for Gaussian measures. The results of this section are stated in a finite-dimensional setting, although they are typically applied in very high – occasionally infinite – dimensions.

[1] To be sure, the "D" in DX refers to the Malliavin derivative with respect to P_n, where $\varphi' = D\varphi$ refers to the Malliavin derivative with respect to P_1.

As a first application of Theorem 0.1 we deduce the concentration of measure property of P_n that was alluded to in the first chapter. The claim is simply that with very high probability every Lipschitz-continuous function is very close to its mean, regardless of the value of the ambient dimension n. One can obtain a crude version of this assertion by appealing to the Poincaré inequality of Nash [Corollary 2.5, page 44] and Chebyshev's inequality:

$$\mathrm{P}\left\{|f - \mathrm{E}(f)| > t\right\} \leqslant \frac{[\mathrm{Lip}(f)]^2}{t^2} \qquad \text{for all } t > 0.$$

This bound is sometimes good enough. But the following is a much more precise estimate that has wider utility as well.

⟨th:CoM⟩ **Theorem 1.1.** *For every Lipschitz-continuous function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathrm{P}_n\left\{|f - \mathrm{E}\, f| \geqslant t\right\} \leqslant 2\exp\left(-\frac{t^2}{2[\mathrm{Lip}(f)]^2}\right) \qquad \textit{for all } t > 0.$$

Proof (Houdé et al XXX). Without loss of generality, we may assume that $\mathrm{E}(f) = 0$ and $\mathrm{Lip}(f) = 1$; else, we replace f by $[f - \mathrm{E}(f)]/\mathrm{Lip}(f)$ everywhere below.

According to Example 1.6 [page 30], $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $\|Df\| \leqslant 1$ a.s. Also, Mehler's formula implies that $\|P_t Df\| \leqslant 1$ for all $t \geqslant 0$, and hence $\|R_1 Df\| \leqslant 1$ a.s. [Theorem 2.1, page 59]. Consequently,

$$\langle Df\,, Df\rangle_{R_1} \leqslant \|Df\|\,\|R_1 Df\| \leqslant 1 \quad \text{a.s.}$$

Choose and fix some number $\lambda \geqslant 0$. We apply Theorem 0.1 with $\varphi(x) := \exp(\lambda x)$ to see that

$$\mathrm{E}\left[f\mathrm{e}^{\lambda f}\right] = \lambda\,\mathrm{E}\left[\mathrm{e}^{\lambda f}\langle Df\,, Df\rangle_{R_1}\right] \leqslant \lambda\,\mathrm{E}\left[\mathrm{e}^{\lambda f}\right].$$

In other words, the function $M(\lambda) := \mathrm{E}[\exp(\lambda f)]$ $[\lambda \geqslant 0]$ satisfies the differential inequality $M'(\lambda) \leqslant \lambda M(\lambda)$ for all $\lambda \geqslant 0$. It is easy to solve this differential inequality: Divide both sides by $M(\lambda)$ and integrate. Since $M(0) = 1$ it follows that

$$\mathrm{E}\,\mathrm{e}^{\lambda f} \leqslant \mathrm{e}^{\lambda^2/2} \qquad \text{for all } \lambda \geqslant 0.$$

By Chebyshev's inequality, if $t \geqslant 0$ then

$$\mathrm{P}_n\{f \geqslant t\} = \mathrm{P}\left\{\mathrm{e}^{\lambda f} \geqslant \mathrm{e}^{\lambda t}\right\} \leqslant \exp\left(-\lambda t + \frac{\lambda^2}{2}\right).$$

Optimize this $[\lambda := t]$ to find that $\mathrm{P}_n\{f \geqslant t\} \leqslant \mathrm{e}^{-t^2/2}$ for every $t \geqslant 0$. Finally, apply the same inequality to the function $-f$ in place of f to deduce the theorem. □

2 The Borell, Sudakov–Tsirelson Inequality

Theorem 1.1 itself has a number of noteworthy consequences. The next result is a particularly useful consequence, as well as a central example of a broader theorem that is generally known as *Borell's inequality*, and was discovered independently and around the same time by Borell XXX, and Sudakov and Tsirelson XXX.

⟨th:Borell⟩ **Theorem 2.1.** *Suppose* $X := (X_1, \ldots, X_n)$ *has a* $\mathrm{N}_n(0, Q)$ *distribution, where* Q *is positive semidefinite, and let* M_n *denote either* $\max_{1\leqslant i\leqslant n} X_i$ *or* $\max_{1\leqslant i\leqslant n} |X_i|$. *Then for all* $t \geqslant 0$,

$$\mathrm{P}\left\{|M_n - \mathrm{E}(M_n)| \geqslant t\right\} \leqslant 2\exp\left(-\frac{t^2}{2\sigma^2}\right),$$

provided that $\sigma^2 := \max_{1\leqslant i\leqslant n} \mathrm{Var}(X_i) > 0$.

Remark 2.2. Frequently, $\sigma^2 \ll \mathrm{E}(M_n)$ when n is large. When this happens, Theorem 2.1 tells us that $M_n \approx \mathrm{E}(M_n)$ with probability very close to one. One way to see this is to integrate by parts:

$$\mathrm{Var}(M_n) = 2\int_0^\infty t\mathrm{P}\left\{|M_n - \mathrm{E}(M_n)| > t\right\}\,\mathrm{d}t \leqslant 4\int_0^\infty t\mathrm{e}^{-t^2/(2\sigma^2)}\,\mathrm{d}t = 4\sigma^2.$$

[The constant 4 can be removed; see Proposition 2.9 on page 45.] For a more concrete illustration, consider the case that $X_1, \ldots, X_n$ are i.i.d. standard normal random variables. In this case, $\sigma^2 = 1$, whereas $\mathrm{E}(M_n) = \sqrt{2\log n} + o(1)$ as $n \to \infty$ thanks to Proposition 1.3, page 7. Thus, Borell's inequality yields that

$$\mathrm{P}\left\{|M_n - \mathrm{E}(M_n)| \geqslant \sqrt{2\varepsilon\log n}\right\} \leqslant 2n^{-\varepsilon},$$

for all $n \geqslant 1$ and $\varepsilon > 0$. We first pass to a subsequence $[n \leftrightarrow 2^n]$ and then use monotonicity and the Borel–Cantelli lemma, in a standard way, in order to deduce that, in the i.i.d. case,

$$\lim_{n\to\infty} \frac{M_n}{\mathrm{E}(M_n)} = \lim_{n\to\infty} \frac{M_n}{\sqrt{2\log n}} = 1 \qquad \text{a.s.,} \tag{5.2}$$ limsup

provided that we construct all of the X_i's on the same probability space. This is of course an elementary statement. It is included here to highlight the fact that, once we know $\mathrm{E}(M_n)$, we frequently need to know very little else in order to analyze the behavior of M_n.

Proof of Theorem 2.1. We can write $Q = S^2$ where S is a symmetric $n \times n$ matrix. Consider the functions

$$f(x) := \max_{1\leqslant i\leqslant n} (Sx)_i \quad \text{and} \quad g(x) := \max_{1\leqslant i\leqslant n} |(Sx)_i| \qquad [x \in \mathbb{R}^n].$$

The proof of [Proposition 2.9, p. 45] shows that f and g are both Lipschitz continuous with $\mathrm{Lip}(f), \mathrm{Lip}(g) \leqslant \sigma^2$ (since, for example, $\|(Df)(x)\|^2 \leqslant \sigma^2$ a.s.). Thus, Theorem 1.1 implies that

$$\mathrm{P}_n\left\{|f(Z) - \mathrm{E}[f(Z)]| \geqslant t\right\} \leqslant 2\exp\left(-\frac{t^2}{2\sigma^2}\right) \qquad [t \geqslant 0],$$

and the very same holds also with $g(Z) - \mathrm{E}[g(Z)]$ in place of $f(Z) - \mathrm{E}[f(Z)]$. This proves the result since SZ has the same distribution as X, whence $f(Z) = \max_{1\leqslant i\leqslant n}[SZ]_i$ has the same distribution as $\max_{1\leqslant i\leqslant n} X_i$, and likewise $g(Z)$ has the same distribution as $\max_{1\leqslant i\leqslant n} |X_i|$. $\square$

3 The S-K Model of Spin Glass

Let us pause and discuss an elegant use of these concentration inequalities, due to Talagrand XXX, to earlier physical predictions of a model in statistical mechanics XXX. In order to see how the following fits into the general scheme of science, we briefly mention the model that we are about to study.

Before we start let me state, once and for all, that we temporarily suspend the notation P_n, E_n, etc. that was used to denote the various objects that act on the Gauss space. In this section we work with the standard notation of probability theory, and on a suitable abstract probability space $(\Omega, \mathcal{F}, \mathrm{P})$.

Imagine n particles charged with unit charges. If the charge of particle i is $\sigma_i \in \{-1, 1\}$, then a simplest model for the total [magnetization] energy of the system is given by the so-called *Hamiltonian*,

$$H_n(\sigma; x) := \frac{1}{\sqrt{n}} \sum_{1\leqslant i<j\leqslant n} \sigma_i \sigma_j x_{i,j},$$

for every $n \geqslant 2$, $x \in \mathbb{R}^n \times \mathbb{R}^n$, and $\sigma \in \{-1, 1\}^n$. Since $\sigma_i \in \{-1, 1\}$, people refer to σ_i as the *spin* of particle i. Moreover, $x_{i,j}$ is a real number that gauges the strength of the interaction between particle i and particle j; this is an input into the model that we control. And $1/\sqrt{n}$ is just a normalization factor.[2]

A standard model of statistical mechanics for the probability distribution of the spins is the following: For every possible set of spins $(\sigma_1, \ldots, \sigma_n) \in \{-1, 1\}^n$, the probability $\mathrm{P}_n^{(x)}(\sigma_1, \ldots, \sigma_n)$ that the respective particle spins are $(\sigma_1, \ldots, \sigma_n)$ is proportional to $\exp\{\beta H_n(\sigma; x)\}$. That is,

$$\mathrm{P}_n^{(x)}(\sigma_1, \ldots, \sigma_n) := \frac{\mathrm{e}^{\beta H_n(\sigma;x)}}{\Pi_n(x)},$$

where $\beta \in \mathbb{R}$ is a parameter that is called *inverse temperature*, and $\Pi_n(x)$ is there to make sure that the probabilities add up to one. That is,

$$\Pi_n(x) := \sum_{\sigma\in\{-1,1\}^n} \mathrm{e}^{\beta H_n(\sigma;x)}. \tag{5.3}$$ Pi_n(Z)

We may, and will, think of Π_n as a function of the interactions $\{x_{i,j}\}$, in which case Π_n is called the *partition function* of the particle system. One can think of the partition function combinatorially—as above—or probabilistically as

$$\Pi_n(x) = 2^n \, \mathrm{E}\left[\mathrm{e}^{\beta H_n(S;x)}\right],$$

where $S := (S_1, \ldots, S_n)$ for a system of i.i.d. random variables $S_1, \ldots, S_n$ with $\mathrm{P}\{S_1 = 1\} = \mathrm{P}\{S_1 = -1\} = 1/2$.

Intuitively speaking, a given set $\{\sigma_1, \ldots, \sigma_n)$ of possible spins has a good chance of being realized iff $H_n(\sigma; x)$ is positive and large. And that ought to happen iff σ_i and σ_j have the same sign for most pairs (i, j) of particles that have positive interactions (i.e. $x_{i,j} > 0$), and opposite sign for most pairs with negative interactions (i.e. $x_{i,j} < 0$). The parameter β ensures the effect of the interaction on this probability: If $|\beta|$ is very small [high temperature], then the interactions matter less; and when $|\beta|$ is very large

[2]In physical terms, we are assuming that there is no external field, and that particles only have pairwise interactions; all higher-order interactions are negligible and hence suppressed.

[low temperature], then the interactions play an extremely important role. In all cases, the spins are highly correlated, except when $\beta = 0$. In the case that $\beta = 0$ [infinite temperature], the spins are i.i.d. [no interaction] and distributed as $S_1, \ldots, S_n$.

Suppose that the partition function behaves as $\exp\{\mathcal{F}_\beta n(1+o(1))\}$, when n is large, where $\mathcal{F}_\beta$ is a number in $(0\,,\infty)$. Then the number $\mathcal{F}_\beta$ is called the *free energy* of the system. A general rule of thumb is that if the free energy exists then its value describes the amount of energy in the system that can be converted to work at temperature $1/\beta$. In such a case, the system is called "extensive." In any case, if the free energy exists then its value is

$$\mathcal{F}_\beta := \lim_{n\to\infty} \frac{1}{n} \log \Pi_n(x).$$

It is possible to prove a carefully-stated version of the ansatz that "free energy exists for almost all choices of interaction terms $\{x_{i,j}\}$"; see Guerra XXX. This requires a relatively-simply, standard "subadditivity argument," though the details of the problem escaped many attempts for a long time until Guerra's work was published. And there are many conjectures about the value of free energy in various cases where $|\beta| \geqslant 1$.

A remarkable theorem of Talagrand XXX implies that if $\beta \in (-1\,,1)$, then

$$\mathcal{F}_\beta = \log(2) + \frac{\beta^2}{4},$$

"for almost all interaction choices." One way to make this precise is to consider the case that the $x_{i,j}$'s are replaced by $Z := (Z_{i,j})_{1\leqslant i<j\leqslant n}$, a system of $n(n-1)/2$ i.i.d. standard normal random variables. We can relabel the $Z_{i,j}$'s, so that they are labeled as a random $2n$-vector rather than the superdiagonal elements of a random $n \times n$ symmetric matrix. In this way we can apply the theory of Gauss space, and the following is a way to state Talagrand's theorem. The resulting *spin glass* model behind this is due to Sherrington and Kirkpatrick XXX. The following discussion follows the exposition of Ledoux XXX.

⟨th:SP⟩ **Theorem 3.1** (Talagrand, XXX)**.** *For every $\beta \in (-1\,,1)$ there exists a finite constant L_β such that for all $\varepsilon > 0$ and $n \geqslant 1 + (2L_\beta/\varepsilon)^2$,*

$$\mathrm{P}_n\left\{\left|\frac{\log \Pi_n(Z)}{n-1} - \left(\log 2 + \frac{\beta^2}{4}\right)\right| \leqslant \varepsilon\right\} \geqslant 1 - 2\exp\left(-\frac{\varepsilon^2(n-1)}{4\beta^2}\right).$$

This theorem addresses the *high-temperature* case where $|\beta|$ is small. The case that $|\beta| \geqslant 1$ is still relatively poorly understood. The difference between the two cases is mainly that when $|\beta|$ is small the interactions are relatively weak; whereas they are strong when $|\beta|$ is large. Mathematically, this manifests itself as follows: When $|\beta|$ is small, $\log \Pi_n(Z) \approx \mathrm{E}[\log \Pi_n(Z)] \approx \log \mathrm{E}[\Pi_n(Z)]$ with high probability. Whereas it is believed that $\mathrm{E}[\log \Pi_n(Z)]$ is a great deal smaller than $\log \mathrm{E}[\Pi_n(Z)]$ when $|\beta| \gg 1$.[3] These approximations are useful for small values of β since $\mathrm{E}[\Pi_n(Z)]$ is easy to compute exactly. In fact, $\mathrm{E}[\Pi_n(Z)]$ satisfies the following elegant identity, whose proof is deferred to the Problems.

⟨lem:SP1⟩ **Lemma 3.2.** *For all $\beta \in \mathbb{R}$ and $n \geqslant 2$,*

$$\mathrm{E}\left[\Pi_n(Z)\right] = 2^n \exp\left(\frac{\beta^2(n-1)}{4}\right).$$

[3] Of course, we always have $\mathrm{E}[\log \Pi_n(Z)] \leqslant \log \mathrm{E}[\Pi_n(Z)]$, by Jensen's inequality.

Taking logarithms of the first moment of $\Pi_n(Z)$ partially explains the statement of Theorem 3.1. Other moments of $\Pi_n(Z)$ can be harder to compute exactly, in a way that is useful. The following yields a useful bound for the second moment in the high-temperature regime.

⟨lem:SP2⟩ **Lemma 3.3.** *If $-1<\beta<1$, then for all $n\geqslant 2$,*

$$\mathrm{E}\left(|\Pi_n(Z)|^2\right)\leqslant\frac{4^n}{\sqrt{1-\beta^2}}\exp\left(\frac{\beta^2(n-2)}{2}\right).$$

Proof. Let us write

$$\begin{aligned}\mathrm{E}\left(|\Pi_n(Z)|^2\right)&=\sum_{\sigma,\sigma'\in\{-1,1\}^n}\mathrm{E}\left[\exp\left(\frac{\beta}{\sqrt{n}}\sum_{1\leqslant i<j\leqslant n}(\sigma_i\sigma_j+\sigma_i'\sigma_j')Z_{i,j}\right)\right]\\&=\sum_{\sigma,\sigma'\in\{-1,1\}^n}\exp\left(\frac{\beta^2}{2n}\sum_{1\leqslant i<j\leqslant n}[\sigma_i\sigma_j+\sigma_i'\sigma_j']^2\right).\end{aligned}$$

If $\sigma,\sigma'\in\{-1\,,1\}^n$, then

$$\begin{aligned}\sum_{1\leqslant i<j\leqslant n}[\sigma_i\sigma_j+\sigma_i'\sigma_j']^2&=2\sum_{1\leqslant i<j\leqslant n}\left[1+\sigma_i\sigma_j\sigma_i'\sigma_j'\right]=n(n-1)+2\sum_{i=1}^{n-1}\sigma_i\sigma_i'\sum_{j=i+1}^{n}\sigma_j\sigma_j'\\&=n(n-2)+\left[\sum_{i=1}^n\sigma_i\sigma_i'\right]^2.\end{aligned}$$

Therefore,

$$\mathrm{E}\left(|\Pi_n(Z)|^2\right)=\mathrm{e}^{\beta^2(n-2)/2}\sum_{\sigma,\sigma'\in\{-1,1\}^n}\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n\sigma_i\sigma_i'\right]^2\right).\tag{5.4}$$ LTLT

Since $|\sigma_i\sigma_i'|\leqslant 1$, this immediately yields $\mathrm{E}(|\Pi_n(Z)|^2)\leqslant 4^n\exp\{\beta^2(n-1)\}$, which is less precise than the inequality of the lemma when $n\gg 1$. In order to deduce the better inequality, we proceed with a little more care.

Let $S_1,S_1',\ldots,S_n,S_n'$ be i.i.d., with $\mathrm{P}\{S_1=\pm1\}=\mathrm{P}\{S_1'=\pm1\}=1/2$, all independent of the Z's. We may rewrite (5.4) as follows:

$$\begin{aligned}\mathrm{E}\left(|\Pi_n(Z)|^2\right)&=4^n\mathrm{e}^{\beta^2(n-2)/2}\,\mathrm{E}\left[\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n S_iS_i'\right]^2\right)\right]\\&=4^n\mathrm{e}^{\beta^2(n-2)/2}\,\mathrm{E}\left[\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n S_i\right]^2\right)\right],\end{aligned}$$

because $S_1S_1',\ldots,S_nS_n'$ are i.i.d. with the same common distribution as S_1. Note that the last two expectations are over the S and S' variables, the Z variables have already been integrated out. By independence,

$$\mathrm{E}\left[\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n S_i\right]^2\right)\right]=\mathrm{E}\left[\exp\left(\frac{Z_{1,1}\beta}{\sqrt{n}}\sum_{i=1}^n S_i\right)\right].\tag{5.5}$$ star

This Khintchine-type trick of reintroducing the Gaussian variable $Z_{1,1}$ bypasses messy combinatorial arguments. Indeed, for all $z \in \mathbb{R}$,

$$\mathrm{E}\left[\exp\left(\frac{z\beta}{\sqrt{n}}\sum_{i=1}^{n} S_i\right)\right] = \left\{\mathrm{E}\left[\exp\left(\frac{z\beta}{\sqrt{n}} S_1\right)\right]\right\}^n = \left\{\cosh\left(\frac{z\beta}{\sqrt{n}}\right)\right\}^n \leqslant \mathrm{e}^{z^2\beta^2/2}.$$

Therefore, we condition on $Z_{1,1}$ first in order to see that the last term in (5.5) is at most $\mathrm{E}[\exp(\beta^2 Z_{1,1}^2/2)] = (1-\beta^2)^{-1/2}$, as long as $\beta \in (-1\,,1)$. $\square$

Lemmas 3.2 and 3.3 teach us that, if $|\beta| < 1$, then

$$\mathrm{E}\left(|\Pi_n(Z)|^2\right) = O\left(|\mathrm{E}[\Pi_n(Z)]|^2\right) \qquad \text{as } n \to \infty.$$

This property fails to hold when $|\beta| \geqslant 1$. See Problem 5 below.

And now we come to the next, very important, step of the proof: Concentration of measure!

⟨lem:SP3⟩ **Lemma 3.4.** *If $|\beta| < 1$, then*

$$\mathrm{P}\left\{\left|\frac{\log \Pi_n(Z)}{n-1} - \mathrm{E}\left[\frac{\log \Pi_n(Z)}{n-1}\right]\right| > t\right\} \leqslant 2\exp\left(-\frac{t^2(n-1)}{\beta^2}\right),$$

for all $t > 0$ and $n \geqslant 2$.

Proof. Consider the function $f(x) := \log \Pi_n(x)$ $[x \in \mathbb{R}^n \times \mathbb{R}^n]$. We can easily compute the derivative of f and show that it is uniformly bounded, which immediately implies Lipschitz continuity. Indeed, for the derivative we have

$$\begin{aligned}\frac{\partial}{\partial x_{i,j}}\Pi_n(x) &= 2^n \frac{\partial}{\partial x_{i,j}}\,\mathrm{E}\left[\mathrm{e}^{\beta H_n(S;x)}\right] \\ &= 2^n\beta\,\mathrm{E}\left[\mathrm{e}^{\beta H_n(S;x)}\cdot\frac{\partial}{\partial x_{i,j}} H_n(S\,;x)\right] = \frac{2^n\beta}{\sqrt{n}}\,\mathrm{E}\left[\mathrm{e}^{\beta H_n(S;x)} S_i S_j\right].\end{aligned}$$

Since $|S_iS_j| = 1$, this gives the bound

$$\left|\frac{\partial}{\partial x_{i,j}}\log \Pi_n(x)\right| \leqslant \frac{\beta}{\sqrt{n}}.$$

Therefore, $\|(Df)(x)\|^2 \leqslant \beta^2 n^{-1}\sum_{1\leqslant i<j\leqslant n}\sigma_i^2\sigma_j^2 = \frac{1}{2}\beta^2(n-1)$. This shows that $\mathrm{Lip}(f) \leqslant \beta\sqrt{(n-1)/2}$, and Theorem 1.1 implies the result. $\square$

Now we use the preceding concentration of measure estimate in order to estimate $\mathrm{E}[\log \Pi_n(Z)]$ accurately for large n. As was mentioned earlier, the key idea is that when $|\beta|$ is small the model is *mean field*; in this case, this means that $\mathrm{E}[\log \Pi_n(Z)] \approx \log \mathrm{E}[\Pi_n(Z)]$.

⟨lem:SP4⟩ **Lemma 3.5.** *For all $\beta \in (-1\,,1)$ there exists $K_\beta < \infty$ such that*

$$\frac{n\log 2}{n-1} + \frac{\beta^2}{4} - \frac{K_\beta}{\sqrt{n-1}} \leqslant \mathrm{E}\left[\frac{\log \Pi_n(Z)}{n-1}\right] \leqslant \frac{n\log 2}{n-1} + \frac{\beta^2}{4},$$

for all $n \geqslant 2$.

Proof. Recall the *Paley–Zygmund inequality* XXX: If $W \geqslant 0$ has two finite moments, then

$$\mathrm{P}\left\{W \geqslant \tfrac{1}{2}\,\mathrm{E}(W)\right\} \geqslant \frac{(\mathrm{E}[W])^2}{4\,\mathrm{E}(W^2)}, \tag{5.6}$$ PZ

provided that $\mathrm{E}(W^2) > 0$.

The Paley–Zygmund inequality and Lemmas 3.2 and 3.3 together show us that

$$\mathrm{P}\left\{\log \Pi_n(Z) \geqslant \log\left(\tfrac{1}{2}\,\mathrm{E}\left[\Pi_n(Z)\right]\right)\right\} \geqslant \tfrac{1}{4}\sqrt{1-\beta^2}\,\mathrm{e}^{\beta^2/2}.$$

Now define $t := \log(\frac{1}{2}\,\mathrm{E}(\Pi_n(Z))) - \mathrm{E}[\log \Pi_n(Z)] > 0$, and note that by Jensen's inequality $t > -\log 2$. If $t \geqslant 0$ then

$$\begin{aligned}\tfrac{1}{4}\sqrt{1-\beta^2}\,\mathrm{e}^{\beta^2/2} &\leqslant \mathrm{P}\left\{\left|\log \Pi_n(Z) - \mathrm{E}\left[\log \Pi_n(Z)\right]\right| \geqslant t\right\}\\ &\leqslant 2\exp\left(-\frac{t^2}{\beta^2(n-1)}\right),\end{aligned}$$

thanks to concentration of measure (Lemma 3.4). Thus,

$$t \leqslant \sqrt{n-1}\left[\frac{\beta^4}{2} + \frac{\beta^2}{2}\left|\log\left(64(1-\beta^2)\right)\right|\right]^{1/2} := C_\beta\sqrt{n-1}.$$

And if $t \leqslant 0$ then certainly the preceding holds also. This proves that in any case,

$$\begin{aligned}\mathrm{E}\left[\log \Pi_n(Z)\right] &= \log(\tfrac{1}{2}\,\mathrm{E}(\Pi_n(Z))) - t\\ &\geqslant \log \mathrm{E}\left[\Pi_n(Z)\right] - C_\beta\sqrt{n-1} - \log 2\\ &\geqslant \log \mathrm{E}\left[\Pi_n(Z)\right] - \left[C_\beta + \log 2\right]\sqrt{n-1},\end{aligned}$$

since $n \geqslant 2$. Apply Lemma 3.2 to obtain the asserted lower bound with $K_\beta := C_\beta + \log 2$.

The upper bound is much simpler to prove, since $\mathrm{E}[\log \Pi_n(Z)] \leqslant \log \mathrm{E}[\Pi_n(Z)]$, owing to Jensen's inequality. □

Proof of Theorem 3.1. Lemma 3.5 ensures that

$$\left|\mathrm{E}\left[\frac{\log \Pi_n(Z)}{n-1}\right] - \left(\log 2 + \frac{\beta^2}{4}\right)\right| \leqslant \frac{K_\beta}{\sqrt{n-1}} + \frac{\log 2}{n-1} \leqslant \frac{L_\beta}{2\sqrt{n-1}},$$

where $L_\beta := 2(K_\beta + \log 2)$. Therefore, Lemma 3.4 implies that

$$\mathrm{P}\left\{\left|\frac{\log \Pi_n(Z)}{n-1} - \left(\log 2 + \frac{\beta^2}{4}\right)\right| > t + \frac{L_\beta}{2\sqrt{n-1}}\right\} \leqslant 2\mathrm{e}^{-t^2(n-1)/\beta^2}.$$

The above probability decreases further if we replace $L_\beta/2\sqrt{n-1}$ by t, provided that $t > L_\beta/2\sqrt{n-1}$. Let $\varepsilon := 2t$ to deduce the theorem. □

4 Absolute Continuity of the Law

Now we turn to a quite delicate consequence of integration by parts. Recall that the *distribution*, or *law*, of a random variable $f : \mathbb{R}^n \to \mathbb{R}$ is the Borel probability measure $\mu_f := \mathrm{P}_n \circ f^{-1}$, defined via

$$\mu_f(A) := \mathrm{P}_n\left\{f \in A\right\} = \mathrm{P}_n\left\{x \in \mathbb{R}^n : f(x) \in A\right\} \qquad \text{for all } A \in \mathcal{B}(\mathbb{R}^n).$$

In the remainder of this chapter we address the question of when μ_f is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$. Moreover, we will say a few things about the structure of the density,

$$p_f(x) := \frac{\mathrm{d}\mu_f(x)}{\mathrm{d}x} \qquad [x \in \mathbb{R}^n],$$

if and when it exists.

The existence of a density is not a trivial issue. For example, the random variable $f(x) \equiv 1$ does not have a density with respect to Lebesgue's measure; yet $f(x) = x_1$ does and its probability density is exactly γ_1 [sort this out!].

§4.1 A Simple Condition for Absolute Continuity

Recently, Nourdin and Viens XXX have found a necessary and sufficient condition for the law of a one-dimensional random variable $f \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ to have a density with respect to the Lebesgue measure on $\mathbb{R}$, together with a formula for the density p_f if and when it exists. Before we discuss this beautiful topic, let us present an easy-to-verify, quite elegant, sufficient condition for the existence of a density.

?⟨th:NZ⟩? **Theorem 4.1** (Nualart and Zakai, XXX)**.** *If $\|Df\| > 0$ a.s., then μ_f is absolutely continuous with respect to the Lebegue measure on $\mathbb{R}$.*

It is clear that we need some sort of non-degeneracy condition on Df. For instance, if $Df = 0$ a.s., then $f = \mathrm{E}(f)$ a.s. thanks to the Poincaré inequality [Proposition 2.4, page 43], and $\mu_f = \delta_{\mathrm{E}(f)}$ is not absolutely continuous.

Proof. Choose and fix an arbitrary bounded Borel set $B \subseteq \mathbb{R}$, and define

$$\varphi(t) := \int_{-\infty}^{t} \mathbb{1}_B(r)\,\mathrm{d}r \qquad [t \in \mathbb{R}].$$

Then, φ is Lipschitz continuous with $\mathrm{Lip}(\varphi) \leqslant 1$, and hence $\varphi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ [Example 1.6, page 30]. We can approximate $\mathbb{1}_B$ with a smooth function in order to see also that $D\varphi = \mathbb{1}_B$ a.s. Therefore, the chain rule of Malliavin calculus [Lemma 1.7, page 31] implies the almost-sure identity,

$$D(\varphi \circ f) = \mathbb{1}_B(f)D(f).$$

If, in addition, B were Lebesgue-null, then $\varphi \equiv 0$ and hence $D(\varphi \circ f) = 0$ a.s. Since $\|Df\| > 0$ a.s., it would then follow that $\mathbb{1}_B(f) = 0$ a.s., which is to say that $\mathrm{P}_n\{f \in B\} = 0$. The Radon–Nikodỳm theorem does the rest. □

§4.2 The Support of the Law

The Nourdin–Viens theory relies on a few well-known, earlier, facts about the support of the law of f XXX. Recall that the *support* of the measure μ_f is the smallest closed set $\mathrm{supp}(\mu_f)$ such that

$$\mu_f(A) = 0 \quad \text{for all Borel sets } A \text{ that do not intersect } \mathrm{supp}(\mu_f).$$

Of course, $\mathrm{supp}(\mu_f)$ is a closed subset of $\mathbb{R}$.

The main goal of this subsection is to verify the following XXX.

⟨th:supp⟩ **Theorem 4.2** (Fang, XXX). *If $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$, then $\mathrm{supp}(\mu_f)$ is an interval.*

Proof. We plan to prove that $\mathrm{supp}(\mu_f)$ is connected. Suppose to the contrary that there exist $-\infty < a < b < \infty$ such that $[a\,,b] \not\subseteq \mathrm{supp}(\mu_f)$, yet $\mathrm{supp}(\mu_f)$ intersects both $(-\infty\,,a]$ and $[b\,,\infty)$.

For every $\varepsilon \in (0\,,(b-a)/2)$ define

$$\varphi_\varepsilon(w) := \begin{cases} 1 & \text{if } w < a+\varepsilon,\\ \varepsilon^{-1}[-w+a+2\varepsilon] & \text{if } a+\varepsilon \leqslant w \leqslant a+2\varepsilon,\\ 0 & \text{if } w > a+2\varepsilon. \end{cases}$$

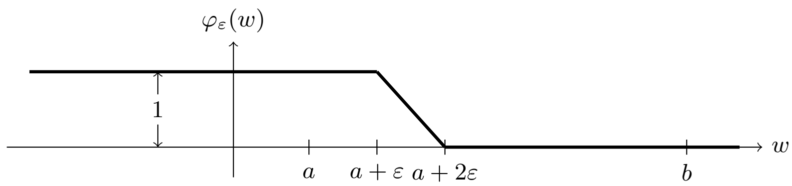

Figure 5.1. An approximation φ_ε to $\mathbb{1}_{(-\infty,a]}$.

Clearly, φ_ε is a Lipschitz-continuous function for every $\varepsilon > 0$, in fact piecewise linear, and $\mathrm{Lip}(\varphi_\varepsilon) = \varepsilon^{-1}$. The chain rule [Lemma 1.7, page 31] implies that $\varphi_\varepsilon(f) \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and

$$D(\varphi_\varepsilon \circ f) = \varphi'_\varepsilon(f)D(f) \qquad \text{a.s.},$$

where we write φ'_ε in place of the more precise $D\varphi_\varepsilon$ for typographical convenience.

By construction, $[a\,,b] \not\subseteq \mathrm{supp}(\mu_f)$ and φ'_ε vanishes [a.s.] off the closed interval $[a+\varepsilon\,,a+2\varepsilon] \subset (a\,,b)$. Therefore, $\varphi'_\varepsilon(f) = 0$ a.s., whence $D(\varphi_\varepsilon \circ f) = 0$ a.s. This and the Poincaré inequality together imply that

$$\varphi_\varepsilon \circ f = \mathrm{E}[\varphi_\varepsilon(f)] \qquad \text{a.s.}$$

Send $\varepsilon \to 0$ and appeal to the bounded convergence theorem in order to see that $\mathbb{1}_{\{f \leqslant a\}} = \mathrm{P}_n\{f \leqslant a\}$ a.s. In particular, $\mathrm{P}_n\{f \leqslant a\} = 0$ or 1 which implies in turn that $\mathrm{supp}(\mu_f)$ cannot intersect both $(-\infty\,,a]$ and $[b\,,\infty)$. This establishes the desired contradiction. □

§4.3 The Nourdin–Viens Formula

We now begin work toward developing the Nourdin–Viens formula.

Let us first recall a fact about conditional expectations. Let $X : \mathbb{R}^n \to \mathbb{R}$ denote an arbitrary random variable and $Y \in L^1(\mathrm{P}_n)$. Then there exists a Borel-measurable function $G_{Y|X} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathrm{E}(Y \mid X) = G_{Y|X}(X) \qquad \text{a.s.,;} \tag{5.7}$$ eq:E(Y|X)

see Problem 3. Note that the preceding depends only on the restriction of $G_{Y|X}$ to the support of the law of X. Thus, we can always define, without affecting the validity of the identity (5.7),

$$G_{Y|X} = \infty \qquad \text{off the support of the law of } X.$$

In particular, it follows that for every random variable $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ we can find a Borel–measurable function $\mathfrak{S}_f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that[4]

$$\mathrm{E}\left[\langle Df\,, Df\rangle_{R_1} \,\middle|\, f\right] = \mathfrak{S}_f \circ f \qquad \text{a.s.} \tag{5.8}$$ S_f

The recipe is $\mathfrak{S}_f(x) := G_{\langle Df\,, Df\rangle_{R_1}|f}(x)$ for all x in the support of the law μ_f of f, and $\mathfrak{S}_f(x) = \infty$ otherwise.

We may apply the integration-by-parts Theorem 0.1 with $\varphi(w) := w$ in order to that

$$\mathrm{Var}(f) = \mathrm{E}[\mathfrak{S}_f \circ f] \qquad \text{for all } f \in \mathbb{D}^{1,2}(\mathrm{P}_n).$$

In other words, $\mathfrak{S}_f \circ f = \mathfrak{S}_f(f(Z))$ is an "unbiased estimate" of the variance of f. The following suggests further that $\mathfrak{S}_f(f(Z))$ might be a good "variance estimator."

?⟨lem:S>0⟩? **Lemma 4.3.** *If $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ has mean zero, then $\mathfrak{S}_f \circ f \geqslant 0$ a.s.*

Proof. Let $\psi : \mathbb{R} \to \mathbb{R}_+$ be an arbitrary non-negative, bounded and measurable function. Define

$$\varphi(x) := \int_0^x \psi(y)\,\mathrm{d}y \qquad [x \in \mathbb{R}].$$

It is possible to check directly that $\varphi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ and $D\varphi = \psi$ a.s. Since $\mathrm{E}(f) = 0$, these facts and Theorem 0.1 together imply that

$$\mathrm{E}\left[f\varphi(f)\right] = \mathrm{E}\left[\psi(f)\langle Df\,, Df\rangle_{R_1}\right] = \mathrm{E}\left[\psi(f)(\mathfrak{S}_f \circ f)\right], \tag{5.9}$$ eq:Cov:2

thanks to integration by parts formula (5.1), the tower property of conditional expectations, and (5.8). Since $x\varphi(x) \geqslant 0$ for all $x \in \mathbb{R}$, the left-most term in (5.9) is non-negative, and hence $\mathrm{E}[\psi(f) \times (\mathfrak{S}_f \circ f)] \geqslant 0$ for all bounded and measurable scalar function $\psi \geqslant 0$. Choose and fix $\eta > 0$ and appeal to the preceding with

$$\psi(x) := \mathbb{1}_{(-\infty,-\eta)}(\mathfrak{S}_f(x)) \qquad [x \in \mathbb{R}^n],$$

in order to see that $\mathrm{P}\{\mathfrak{S}_f \circ f \leqslant -\eta\} = 0$ for all $\eta > 0$. This proves the remainder of the proposition. □

Thus we see that $\mathfrak{S}_f \circ f$ is always non-negative when $\mathrm{E}(f) = 0$. A remarkable theorem of Nourdin and Viens XXX asserts that strict inequality holds—that is $\mathfrak{S}_f \circ f > 0$ a.s.—if and only if μ_f is absolutely continuous. Moreoever, one can obtain a formula for the probability density of f when $\mathfrak{S}_f \circ f > 0$ a.s. The precise statement follows.

⟨th:NP:density⟩ **Theorem 4.4** (Nourdin and Viens, XXX)**.** *Suppose $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ satisfies $\mathrm{E}(f) = 0$. Then $\mu_f(\mathrm{d}x) \ll \mathrm{d}x$ if and only if $\mathfrak{S}_f \circ f > 0$ a.s. Moreover, when $\mathfrak{S}_f \circ f > 0$ a.s., the following defines a version of the probability density function of f:*

$$p_f(x) = \frac{\mathrm{E}(|f|)}{2\mathfrak{S}_f(x)} \exp\left(-\int_0^x \frac{z\,\mathrm{d}z}{\mathfrak{S}_f(z)}\right), \tag{5.10}$$ eq:NP:density

where $1/\infty := 0$.

Remark 4.5. Observe that $\mathrm{P}_n\{\mathfrak{S}_f \circ f > 0\} = 1$ iff $\mu_f\{\mathfrak{S}_f > 0\} = 1$.

[4]The space $\mathbb{R} \cup \{\infty\}$ is viewed as the usual one-point compactification of $\mathbb{R}$, endowed with the corresponding topology and Borel σ-algebra.

The proof of Theorem 4.4 is naturally broken into three separate parts, which we record as Propositions 4.6 through 4.8 below.

⟨pr:NP:1⟩ **Proposition 4.6.** *Let f be a mean-zero random variable in $\mathbb{D}^{1,2}(\mathrm{P}_n)$. If $\mathfrak{S}_f \circ f > 0$ a.s., then $\mu_f(\mathrm{d}x) \ll \mathrm{d}x$.*

Proof. Let $B \subset \mathbb{R}$ be an arbitrary Borel-measurable set, and define

$$\varphi(x) := \int_0^x \mathbb{1}_B(y)\,\mathrm{d}y \qquad [x \in \mathbb{R}].$$

Then $\varphi \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $D\varphi = \mathbb{1}_B$ a.s. The integration-by-parts Theorem 0.1 implies that

$$\mathrm{E}\left[f \int_0^f \mathbb{1}_B(y)\,\mathrm{d}y\right] = \mathrm{E}\left[1_B(f) \times (\mathfrak{S}_f \circ f)\right].$$

If B were Lebesgue null, then $\int_0^f \mathbb{1}_B(y)\,\mathrm{d}y = 0$ a.s., and hence $\mathrm{E}[\mathbb{1}_B(f) \times (\mathfrak{S}_f \circ f)] = 0$. Because we have assumed that $\mathfrak{S}_f \circ f > 0$ a.s., it follows that $\mathbb{1}_B(f) = 0$ a.s., equivalently, $\mathrm{P}_n\{f \in B\} = 0$. The Radon–Nikodỳm theorem does the rest. □

?⟨pr:NP:2⟩? **Proposition 4.7.** *If $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ has mean zero and satisfies $\mu_f(\mathrm{d}x) \ll \mathrm{d}x$, then $\mathfrak{S}_f \circ f > 0$ a.s.*

Proof. Let p_f denote the probability density of f; that is, p_f satisfies $\int_B p_f(x)\,\mathrm{d}x = \mathrm{P}_n\{f \in B\}$ for all Borel sets $B \subset \mathbb{R}$.

If $\psi \in C_c(\mathbb{R})$ (that is, if ψ is continuous with compact support), then $\varphi(x) := \int_{-\infty}^x \psi(y)\,\mathrm{d}y$ $[x \in \mathbb{R}]$ is bounded and $\varphi'(x) = \psi(x)$ for every $x \in \mathbb{R}$. We may integrate by parts in Gauss space [see (5.9)] in order to see that

$$\mathrm{E}\left[\psi(f) \times (\mathfrak{S}_f \circ f)\right] = \mathrm{E}\left[f \times \varphi(f)\right] = \int_{-\infty}^{\infty} y\varphi(y)p_f(y)\,\mathrm{d}y.$$

Since $\int_{-\infty}^{\infty} yp_f(y)\,\mathrm{d}y = 0$ and φ is bounded, we can integrate by parts—in Lebesgue space—in order to see that

$$\mathrm{E}\left[\psi(f) \times (\mathfrak{S}_f \circ f)\right] = \int_{-\infty}^{\infty} \psi(y)\left(\int_y^{\infty} zp_f(z)\,\mathrm{d}z\right)\mathrm{d}y. \tag{5.11}$$ eq:NP1

Now $\mathrm{P}_n\{p_f(f) = 0\} = \mu_f\{p_f = 0\} = \int_{\{p_f=0\}} p_f(a)\,\mathrm{d}a = 0$. Therefore, we can rewrite (5.11) as

$$\mathrm{E}\left[\psi(f) \times (\mathfrak{S}_f \circ f)\right] = \mathrm{E}\left[\psi(f) \times \frac{\int_f^{\infty} zp_f(z)\,\mathrm{d}z}{p_f(f)}\right],$$

for all continuous $\psi \in C_c(\mathbb{R})$. The preceding holds for all bounded and measurable functions $\psi : \mathbb{R} \to \mathbb{R}$ by density. Consequently,

$$\mathfrak{S}_f \circ f = \frac{\int_f^{\infty} zp_f(z)\,\mathrm{d}z}{p_f(f)} \qquad \text{a.s.} \tag{5.12}$$ S=Phi/f

It remains to prove that

$$\int_f^{\infty} zp_f(z)\,\mathrm{d}z > 0 \qquad \text{a.s.} \tag{5.13}$$ goal:NP

Thanks to Theorem 4.2, the law of f is supported in some closed interval $[\alpha\,, \beta]$ where $-\infty \leqslant \alpha \leqslant \beta \leqslant \infty$. And since f has mean zero, it follows that $\alpha < 0 < \beta$.

Define

$$\Phi(x) := \int_x^\infty z p_f(z)\,\mathrm{d}z \qquad [x \in \mathbb{R}]. \tag{5.14}$$ eq:NP:Phi

Since p_f is supported in $[\alpha\,,\beta]$, Φ is constant off $[\alpha\,,\beta]$; and the mean-zero property of f implies that $\Phi(\alpha+) = \Phi(\beta-) = 0$. Furthermore, Φ is a.e. differentiable and $\Phi'(x) = -xp_f(x)$ a.e., thanks to the Lebesgue differentiation theorem. Since $p_f > 0$ a.e. on $[\alpha\,,\beta]$, it follows that Φ is strictly increasing on $(\alpha\,,0]$ and strictly decreasing on $[0\,,\beta)$. As Φ vanishes at α and β, this proves that $\Phi(x) > 0$ for all $x \in (\alpha\,,\beta)$ whence $\Phi(f) > 0$ a.s. This implies (5.13) and completes the proof. □

⟨pr:NP:3⟩ **Proposition 4.8.** *If $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ has mean zero and $\mathfrak{S}_f \circ f > 0$ a.s., then the density function of f is given by* (5.10).

Proof. Recall the function Φ from (5.14). Then Φ is almost-everywhere differentiable [Lebesgue's theorem] with derivative $\Phi'(x) = -xp_f(x)$ a.e. At the same time, (5.12) implies that

$$\Phi(x) = \mathfrak{S}_f(x)p_f(x) \qquad \text{for almost all } x \in \mathrm{supp}(\mu_f). \tag{5.15}$$ Phi=Sp

It follows that

$$\frac{\Phi'(x)}{\Phi(x)} = -\frac{x}{\mathfrak{S}_f(x)} \qquad \text{for almost all } x \in \mathrm{supp}(\mu_f).$$

The preceding in fact holds also for every $x \not\in \mathrm{supp}(\mu_f)$ since $\Phi' = 0$ and $\mathfrak{S}_f = \infty$ off μ_f, provided that $1/\infty = 0$. We integrate the preceding to obtain

$$\Phi(x) = \Phi(0)\exp\left(-\int_0^x \frac{z\,\mathrm{d}z}{\mathfrak{S}_f(z)}\right) \qquad \text{for all } x \in \mathbb{R}^n. \tag{5.16}$$ eq:Phi

But $\Phi(0) = \int_0^\infty zp_f(z)\,\mathrm{d}z = \mathrm{E}(f\,;f > 0) = \frac12\,\mathrm{E}(|f|)$, because $\mathrm{E}(f) = 0$. Therefore, (5.15) implies the result. □

5 Aspects of the Nourdin–Peccati Theory

Recently, Ivan Nourdin and Giovanni Peccati XXX recognized a number of remarkable consequences of integration by parts [Theorem 4.7, page 64] that lie at the very heart of Gaussian analysis. Theorem 4.4 is only one such example. We will next decscribe a few other examples. Their monograph XXX contains a host of others.

§5.1 A Characterization of Normality

One of the remarkable consequences of the Nourdin–Peccati theory is that it characterizes when a non-degenerate random variable $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ has a mean-zero normal distribution.

⟨th:NP:Normality⟩ **Theorem 5.1** (Nourdin and Peccati, XXX)**.** *Suppose $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ satisfies $\mathrm{E}(f) = 0$. Then, the random variable f has a normal distribution iff $\mathfrak{S}_f \circ f$ is a constant a.s.*

Remark 5.2. The constancy condition on $\mathfrak{S}_f \circ f$ is equivalent to the condition that

$$\mathfrak{S}_f \circ f = \mathrm{E}[\mathfrak{S}_f \circ f] = \mathrm{E}\left[\langle Df\,,Df\rangle_{R_1}\right] = \mathrm{Var}(f) \quad \text{a.s.},$$

thanks to Theorem 0.1. Therefore, Theorem 5.1 is saying that f is a normally distributed if and only if its variance estimator $\mathfrak{S}_f \circ f$ is exact. For a stronger result see Example 5.9 below.

The proof of Theorem 5.1 rests on the following "heat kernel estimate," which is interesting in its own right.

⟨th:Heat:Kernel⟩ **Theorem 5.3.** *Suppose* $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ *has mean zero, and there exists a constant* $\sigma > 0$ *such that* $\mathfrak{S}_f \circ f \geqslant \sigma^2$ *a.s. Then* $\operatorname{supp}(\mu_f) = \mathbb{R}$, *and*

$$p_f(x) \geqslant \frac{\mathrm{E}(|f|)}{2\mathfrak{S}_f(x)} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{5.17}$$ eq:Sf(f):Sigma

for almost every $x \in \mathbb{R}$. *Suppose, in addition, that there exists a constant* $\Sigma < \infty$ *such that*

$$\mathfrak{S}_f \circ f \leqslant \Sigma^2 \qquad a.s. \tag{5.18}$$ eq:Sf(f):Sigma:1

Then, for almost every $x \in \mathbb{R}$,

$$\frac{\mathrm{E}(|f|)}{2\sigma^2} \exp\left(-\frac{x^2}{2\Sigma^2}\right) \geqslant p_f(x) \geqslant \frac{\mathrm{E}(|f|)}{2\Sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{5.19}$$ eq:Sf(f):Sigma:2

Remark 5.4. If f is Lipschitz continuous, then we have seen that $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ [Example 1.6, p. 30]. Furthermore, $\|Df\| \leqslant \mathrm{Lip}(f)$ a.s., whence $\|R_1 Df\| \leqslant \mathrm{Lip}(f)$ a.s., by the Mehler formula [Theorem 2.1, page 59]. Thus, whenever f is Lipschitz continuous, condition (5.18) holds with $\Sigma := \mathrm{Lip}(f)$.

Proof. Recall Φ from (5.14). According to (5.16),

$$\Phi(x) \geqslant \frac{1}{2}\,\mathrm{E}(|f|) \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad \text{for almost all } x \in \operatorname{supp}(\mu_f). \tag{5.20}$$ eq:Phi>Gauss

It follows from the fact that $\mathrm{E}(f) = 0$ that $\Phi(x) \to 0$ as x tends to the boundary of $\operatorname{supp}(\mu_f)$. Since $\operatorname{supp}(\mu_f)$ is an interval [Theorem 4.2, page 78], (5.20) shows that $\operatorname{supp}(\mu_f)$ must be unbounded. This proves that $\operatorname{supp}(\mu_f) = \mathbb{R}$. The inequality (5.17) follows from (5.20), and (5.19) follows readily from (5.17) and (5.18). □

Now we can verify Theorem 5.1.

Proof of Theorem 5.1. Suppose f has a normal distribution with mean zero and $\sigma^2 := \mathrm{Var}(f) > 0$. Since μ_f and the Lebesgue measure are mutually absolutely continuous with respect to one another, (5.12) ensures that

$$\mathfrak{S}_f(x) = \frac{\int_x^\infty z p_f(z)\,\mathrm{d}z}{p_f(x)} = \frac{\int_x^\infty z \exp(-z^2/(2\sigma^2))\,\mathrm{d}z}{\exp(-x^2/(2\sigma^2))} = \sigma^2,$$

for μ_f-a.e.—whence almost every—$x \in \mathbb{R}$. The converse follows from Theorems 4.4 and 5.3. □

We highlight some of the scope, as well as some of the limitations, of Theorem 4.4 by studying two elementary special cases. Problem 1 contains a third illustrative example of this kind.

Example 5.5 (A Linear Example). Consider the random variable $f(x) := a \cdot x$ $[x \in \mathbb{R}^n]$, where a is a non-zero constant n-vector. Equivalently, $f = a \cdot Z$, where Z is the standard-normal n-vector from (1.2) [page 3]. Then $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $Df = a$ a.s. Moreover, $\mathrm{E}(f) = a \cdot \mathrm{E}(Z) = 0$ and $\mathrm{Var}(f) = \|a\|^2 > 0$. Furthermore, Mehler's formula [Theorem 2.1, page 59] ensures that $P_t Df = a$ a.s., whence $R_1 Df = \int_0^\infty \mathrm{e}^{-t} a\,\mathrm{d}t = a$ a.s. It follows that $\mathfrak{S}_f \circ f = \|a\|^2$ a.s. Therefore, in the linear case, Theorem 5.1 reduces to the obvious statement that linear combinations of $Z_1, \ldots, Z_n$ are normally distributed.

Example 5.6 (A Quadratic Example). Consider the random variable $f(x) := \|x\|^2 - n$ $[x \in \mathbb{R}^n]$. Equivalently, $f = \|Z\|^2 - \mathrm{E}(\|Z\|^2)$. Clearly, $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$—in fact, $f \in C_0^\infty(\mathrm{P}_n)$—and $\mathrm{E}(f) = 0$ and $(Df)(x) = 2x$. Mehler's formula [Theorem 2.1, page 59] yields $(P_t Df)(x) = 2\mathrm{e}^{-t}x$ for almost all $x \in \mathbb{R}^n$. In particular,

$$(R_1 Df)(x) = \int_0^\infty \mathrm{e}^{-t}(P_t Df)(x)\,\mathrm{d}t = x \qquad \text{for almost all } x \in \mathbb{R}^n.$$

Since $(Df)(x) \cdot (R_1 Df)(x) = 2\|x\|^2 = 2f(x) + 2n$ a.s., it follows that $\mathfrak{S}_f \circ f = \mathrm{E}(2f + 2n \mid f) = 2f + 2n$ a.s. Equivalently, $\mathfrak{S}_f(z) = 2(z + n)$ a.s. for all $z \in \mathrm{supp}(\mu_f)$. Because $\mathrm{P}\{\mathfrak{S}_f = 0\} = \mathrm{P}\{Z = 0\} = 0$, Theorem 4.4 reduces to the statement that $\|Z\|^2 - n$ has a probability density $p_{\|Z\|^2-n}$, and

$$p_{\|Z\|^2-n}(x) = \frac{\mathrm{E}\left(\left|\|Z\|^2 - n\right|\right)}{4(x+n)} \exp\left(-\tfrac{1}{2}\int_0^x \frac{z}{z+n}\,\mathrm{d}z\right) = \frac{\mathrm{E}\left(\left|\|Z\|^2 - n\right|\right)\mathrm{e}^{-x/2}}{4n^{n/2}(x+n)^{1-(n/2)}},$$

for a.e. $x \in$ the support of the law of $\|Z\|^2 - n$. Equivalently,

$$p_{\|Z\|^2}(x) \propto \frac{\mathrm{e}^{-x/2}}{x^{1-(n/2)}} \quad \text{for a.e. } x \in \text{the support of the law of } \|Z\|^2.$$

From this we can see that Theorem 4.4 is consistent with the well-known fact that $\|Z\|^2$ has a χ^2_n distribution.

§5.2 Distance to Normality

Suppose $f \in \mathbb{D}^{1,2}(\mathrm{P}_n)$. Theorem 5.1 suggests that if $\langle Df\,, Df\rangle_{R_1} \approx \tau^2$ we can then expect the distribution of f to be approximately $\mathrm{N}(0\,,\tau^2)$. We might expect even more. Namely, suppose that $X = (X_1\,,\ldots,X_n)$ is a random vector such that $X_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ has mean zero and $\mathrm{Cov}(X_i\,,X_j) = Q_{i,j}$, with $\langle DX_i\,, DX_j\rangle_{R_1} \approx Q_{i,j}$ for all $1 \leqslant i,j \leqslant n$. Then we might expect the distribution of X might be approximately $\mathrm{N}_n(0, Q)$. This is indeed the case, as is shown by the theory of Nourdin, Peccati, and Reinert, XXX. We will work out the details first in the case that $Q = I$ is the $n \times n$ identity matrix.

⟨th:NPR:I⟩ **Theorem 5.7** (Nourdin, Peccati, and Reinert, XXX). *Consider a random vector* $X = (X_1, \ldots, X_n)$, *where* $X_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$, *and* $\mathrm{E}(X_i) = 0$. *Then, for every* $\Phi \in \mathbb{D}^{2,2}(\mathrm{P}_n)$,

$$\mathrm{E}\left[\Phi(X)\right] - \mathrm{E}\left[\Phi(Z)\right] = \mathrm{E}\left[\sum_{i,j=1}^n \left(R_2 D^2_{i,j}\Phi\right)(X) \times \left(\langle DX_i\,, DX_j\rangle_{R_1} - I_{i,j}\right)\right],$$

where R_2 *denotes the 2-potential of the OU semigroup. In particular,*

$$\left|\mathrm{E}[\Phi(X)] - \mathrm{E}[\Phi(Z)]\right| \leqslant K(\Phi)\sum_{i,j=1}^n \mathrm{E}\left(\left|\langle DX_i\,, DX_j\rangle_{R_1} - I_{i,j}\right|\right),$$

for all $\Phi : \mathbb{R}^n \to \mathbb{R}$ *that are bounded and have continuous and bounded mixed partial derivatives of order* $\leqslant 2$, *where* $K(\Phi) := \frac{1}{2}\sup_{x\in\mathbb{R}^n}\max_{1\leqslant i,j\leqslant n}|(D^2_{i,j}\Phi)(x)|$.

Proof. We need only prove the first assertion of the theorem; the second assertion follows readily from the first because of the elementary fact that whenever $|g(x)| \leqslant c$ for all $x \in \mathbb{R}^n$, $|(P_t g)(x)| \leqslant c$ for all t and hence $|(R_2 g)(x)| \leqslant c\int_0^\infty \exp(-2t)\,\mathrm{d}t = c/2$.

The theorem is a fact about the distribution of X, as compared with the distribution of Z. In the proof we wish to construct X and Z—on the same Gaussian probability space—so that they have the correct marginal distributions, but also are independent.

A natural way to achieve our coupling is to define, on $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$, two functions $\bar{X}$ and $\bar{Z}$, as follows: For all $\omega = (\omega_1, \ldots, \omega_{2n}) \in \mathbb{R}^{2n}$,

$$\bar{Z}(\omega) := Z(\omega_1, \ldots, \omega_n), \quad \text{and} \quad \bar{X}(\omega) := X(\omega_{n+1}, \ldots, \omega_{2n}).$$

Then:

1. $\bar{X}$ and $\bar{Z}$ are n-dimensional random vectors on the $2n$-dimensional Gauss space $(\mathbb{R}^{2n}, \mathcal{B}(\mathbb{R}^{2n}), \mathrm{P}_{2n})$;
2. The P_{2n}-distribution of $\bar{X}$ is the same as the P_n-distribution of X; and
3. The P_{2n}-distribution of $\bar{Z}$ is the same as the P_n-distribution of Z.

In this way, Theorem 5.7 can be restated as follows:

$$\begin{aligned} &\mathrm{E}_{2n}\left[\Phi(\bar{X})\right] - \mathrm{E}_{2n}\left[\Phi(\bar{Z})\right] \\ &\qquad = \mathrm{E}_{2n}\left[\sum_{i,j=1}^{n} \left(R_2 D_{i,j}^2 \Phi\right)(\bar{X}) \times \left(\langle D\bar{X}_i\,, D\bar{X}_j\rangle_{R_1} - I_{i,j}\right)\right], \end{aligned} \tag{5.21}$$

coupled0

where, we recall, $R_2 f := \int_0^\infty \mathrm{e}^{-2t} P_t f\,\mathrm{d}t$. We will prove this version of the theorem next.

We will use the same "Gaussian interpolation" trick that has been used a few times already; see, for example, the last statement of Proposition 1.6, p. 58. Note that with P_{2n}-probability one: $(P_0\Phi)(\bar{X}) = \Phi(\bar{X})$; and $(P_t\Phi)(\bar{X}) \to \mathrm{E}_{2n}[\Phi(\bar{Z})]$ as $t \to \infty$. Therefore, P_{2n}-a.s.,

$$\begin{aligned} \Phi(\bar{X}) - \mathrm{E}_{2n}[\Phi(\bar{Z})] &= -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}(P_t\Phi)(\bar{X})\,\mathrm{d}t \\ &= -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}_{2n}\left[\Phi\left(\mathrm{e}^{-t}\bar{X} + \sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\,\Big|\,\bar{X}\right]\mathrm{d}t, \end{aligned} \tag{5.22}$$

?pre:Q?

owing to Mehler's formula [Theorem 2.1, page 59]. We take expectations of both sides and apply the dominated convergence theorem, to interchange the derivative with the expectation, in order to find that

$$\begin{aligned} &\mathrm{E}_{2n}[\Phi(\bar{X})] - \mathrm{E}_{2n}[\Phi(\bar{Z})] \\ &= -\int_0^\infty \mathrm{E}_{2n}\left[\frac{\mathrm{d}}{\mathrm{d}t}\Phi\left(\mathrm{e}^{-t}\bar{X} + \sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\right]\mathrm{d}t \\ &= \sum_{i=1}^{n}\int_0^\infty \mathrm{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar{X} + \sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\left[\mathrm{e}^{-t}\bar{X}_i - \frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\bar{Z}_i\right]\right\}\mathrm{d}t. \end{aligned} \tag{5.23}$$

coupled1

Since $\mathrm{E}_{2n}(\bar{Z}_i) = 0$ Theorem 4.7 on page 64 implies that for all $G \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ and $1 \leqslant i \leqslant n$,[5]

$$\mathrm{E}_{2n}\left[G(\bar{Z})\bar{Z}_i\right] = \mathrm{E}_{2n}\left[\langle D(G\circ\bar{Z})\,, D\bar{Z}_i\rangle_{R_1}\right] = \mathrm{E}_{2n}\left[D_i(G\circ\bar{Z})\right].$$

[5]An alternative way to see this is to recall that $\bar{Z}_i = H_1(\bar{Z}_i) = A_i(H_0) = A_i(1)$.

Therefore, for every $x \in \mathbb{R}^n$ and $1 \leqslant i \leqslant n$,

$$\begin{aligned}
\mathrm{E}_{2n}&\left\{(D_i\Phi)\left(\mathrm{e}^{-t}x+\sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\bar Z_i\right\}\\
&=\frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\mathrm{E}_n\left\{D_i\left[(D_i\Phi)\left(\mathrm{e}^{-t}x+\sqrt{1-\mathrm{e}^{-2t}}\,\bullet\right)\right](\bar Z)\right\}\\
&=\mathrm{e}^{-2t}\,\mathrm{E}_{2n}\left[(D^2_{i,i}\Phi)\left(\mathrm{e}^{-t}x+\sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\right]\\
&=\mathrm{e}^{-2t}\left(P_tD^2_{i,i}\Phi\right)(x),
\end{aligned}$$

thanks first to the chain rule [Lemma 1.7, page 31], and then Mehler's formula [Theorem 2.1, page 59]. Since $\bar X$ and $\bar Z$ are independent, we can first condition on $\bar X = x$ and then integrate $[\mathrm{d}(\mathrm{P}_{2n}\circ \bar X^{-1})]$ to deduce from the preceding that

$$\begin{aligned}
\mathrm{E}_{2n}&\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar X+\sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\bar Z_i\right\}\\
&=\mathrm{e}^{-2t}\,\mathrm{E}_{2n}\left[\left(P_tD^2_{i,i}\Phi\right)(\bar X)\right]\\
&=\mathrm{e}^{-2t}\sum_{j=1}^{n}\mathrm{E}_{2n}\left[\left(P_tD^2_{i,j}\Phi\right)(\bar X)I_{i,j}\right].
\end{aligned}\tag{5.24}$$ coupled2

Similarly, because $\mathrm{E}_{2n}(\bar X_i)=0$ for all $1 \leqslant i \leqslant n$, we can write

$$\begin{aligned}
\mathrm{E}_{2n}\left[G(\bar X)\bar X_i\right]&=\mathrm{E}_{2n}\left[\left\langle D(G\circ\bar X)\,,D\bar X_i\right\rangle_{R_1}\right]=\sum_{k=1}^{2n}\mathrm{E}_{2n}\left[D_k(G\circ\bar X)\times\left(R_1D_k\bar X_i\right)\right]\\
&=\sum_{j=1}^{n}\sum_{k=1}^{2n}\mathrm{E}_{2n}\left[(D_jG)(\bar X)\times D_k(\bar X_j)\times\left(R_1D_k\bar X_i\right)\right]\\
&=\sum_{j=1}^{n}\mathrm{E}_{2n}\left[(D_jG)(\bar X)\times\left\langle D\bar X_j\,,D\bar X_i\right\rangle_{R_1}\right],
\end{aligned}$$

by the chain rule, and hence

$$\begin{aligned}
\mathrm{E}_{2n}&\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar X+\sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\mathrm{e}^{-2t}\bar X_i\right\}\\
&=\mathrm{e}^{-2t}\sum_{j=1}^{n}\mathrm{E}_{2n}\left[(D^2_{i,j}\Phi)\left(\mathrm{e}^{-t}\bar X+\sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\left\langle D\bar X_j\,,D\bar X_i\right\rangle_{R_1}\right]\\
&=\mathrm{e}^{-2t}\sum_{j=1}^{n}\mathrm{E}_{2n}\left[\left(P_tD^2_{i,j}\Phi\right)(\bar X)\left\langle D\bar X_j\,,D\bar X_i\right\rangle_{R_1}\right].
\end{aligned}\tag{5.25}$$ coupled3

We now merely combine (5.23), (5.24), and (5.25) in order to deduce (5.21) and hence the theorem. □

Theorem 5.7 has a useful extension in which one compares the distribution of a smooth mean-zero random variable X to that of an arbitrary mean-zero normal random variable. That is, we consider $\mathrm{E}[\Phi(X)]-\mathrm{E}[\Phi(Q^{1/2}Z)]$, where Q is a symmetric, positive definite matrix that is not necessarily the identity matrix. Consider the linear operators $\{P^Q_t\}_{t\geqslant 0}$ defined as

$$(P^Q_tf)(x):=\mathrm{E}\left[f\left(\mathrm{e}^{-t}x+\sqrt{1-\mathrm{e}^{-2t}}\,Q^{1/2}Z\right)\right].$$

It is not hard to check that the preceding defines a semigroup $\{P_t^Q\}_{t\geqslant 0}$ of linear operators that solve a heat equation of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t^Q = \mathcal{L}^Q\, P_t^Q \text{ for } t>0, \tag{5.26}$$ LQ

subject to $P_0^Q =$ the identity operator. Here $\mathcal{L}^Q$ is a differential operator, much like $\mathcal{L}$, but with coefficients that come from Q. Also there is a corresponding resolvent $R_\lambda^Q := \int_0^\infty \exp(-\lambda t)P_t^Q\,\mathrm{d}t$, etc. Now we begin with the following variation on (5.23): Define

$$\Psi(t) := \mathrm{E}_{2n}\left[\Phi\left(\mathrm{e}^{-t}\bar{X} + \sqrt{1-\mathrm{e}^{-2t}}\,Q^{1/2}\bar{Z}\right)\right] = \mathrm{E}\left[\left(P_t^Q\Phi\right)(X)\right],$$

and notice that $\Psi\in\mathbb{D}^{1,2}(\mathrm{P}_1)$, $\Psi(0)=\mathrm{E}_{2n}[\Phi(\bar{X})]$, and $\lim_{t\to\infty}\Psi(t) = \mathrm{E}_{2n}[\Phi(Q^{1/2}\bar{Z})]$. Therefore,

$$\begin{aligned}
&\mathrm{E}_{2n}[\Phi(\bar{X})] - \mathrm{E}_{2n}[\Phi(Q^{1/2}\bar{Z})]\\
&= \Psi(0) - \lim_{t\to\infty}\Psi(t) = -\int_0^\infty \Psi'(t)\,\mathrm{d}t\\
&= \sum_{i=1}^n\int_0^\infty \mathrm{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar{X}+\sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\left[\mathrm{e}^{-t}\bar{X}_i - \frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,Q^{1/2}\bar{Z}_i\right]\right\}\mathrm{d}t.
\end{aligned}$$

Now we translate the proof of Theorem 5.7 in order to obtain the following important generalization.

⟨th:NPR⟩ **Theorem 5.8** (Nourdin, Peccati, and Reinert, XXX). *Consider a random vector $X=(X_1,\ldots,X_n)$, where $X_i\in\mathbb{D}^{1,2}(\mathrm{P}_n)$, and $\mathrm{E}(X_i)=0$. Then, for every $\Phi\in\mathbb{D}^{2,2}(\mathrm{P}_n)$ and for all $n\times n$ covariance matrices Q,*

$$\mathrm{E}\left[\Phi(X)\right] - \mathrm{E}\left[\Phi(Q^{1/2}Z)\right] = \mathrm{E}\left[\sum_{i,j=1}^n\left(R_2^Q D^2_{i,j}\Phi\right)(X)\times\left(\langle DX_i\,,DX_j\rangle_{R_1^Q} - Q_{i,j}\right)\right],$$

where R_2^Q denotes the 2-potential of the semigroup $\{P_t^Q\}_{t\geqslant 0}$ and

$$\langle Df\,,Dg\rangle_{R_1^Q}(x) := (Df)(x)\cdot(R_1^Q Dg)(x)\qquad a.s.$$

In particular,

$$\left|\mathrm{E}[\Phi(X)] - \mathrm{E}[\Phi(Q^{1/2}Z)]\right| \leqslant K(\Phi)\sum_{i,j=1}^n\mathrm{E}\left(\left|\langle DX_i\,,DX_j\rangle_{R_1^Q} - Q_{i,j}\right|\right),$$

for all $\Phi:\mathbb{R}^n\to\mathbb{R}$ that are bounded and have continuous and bounded mixed partial derivatives of order $\leqslant 2$.

⟨ex:GTS⟩ *Example* 5.9. If $\langle DX_i\,,DX_j\rangle_{R_1^Q} = Q_{i,j}$ a.s. for all $1\leqslant i,j\leqslant n$, then Theorem 5.8 ensures that X has a $\mathrm{N}_n(0\,,Q)$ distribution. Conversely, suppose that X has a $\mathrm{N}_n(0\,,Q)$ distribution. Recall that X has the same distribution as $W:=SZ$, where S is the [symmetric] square root of Q. Of course,

$$(D_kW_i)(x) = \frac{\partial}{\partial x_k}\left([SZ]_i\,(x)\right) = \frac{\partial}{\partial x_k}\sum_{l=1}^n S_{l,i}x_l = S_{k,i},$$

for all $x \in \mathbb{R}^n$. Therefore, the fact that $R_1^Q 1 = 1$ implies that for all $1 \leqslant i, j \leqslant n$,

$$\langle DW_i\,, DW_j\rangle_{R_1^Q} = \sum_{k=1}^n S_{i,k}S_{k,j} = Q_{i,j} \qquad \text{a.s.}$$

Consequently, every centered random vector X such that $X_i \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ for all i has a $\mathrm{N}_n(0\,,Q)$ distribution iff $\langle DX_i\,, DX_j\rangle_{R_1^Q} = Q_{i,j}$ a.s.

The following is an immediate consequence of Theorem 5.8, and provides an important starting point for proving convergence in distribution to normality in the analysis of Nourdin and Peccati (**?**, Theorem 5.3.1, p. 102).

?⟨ex:NPR⟩? *Example* 5.10. Suppose $X^{(1)}, X^{(2)}, \ldots \in \mathbb{D}^{1,2}(\mathrm{P}_n)$, all have mean vector $0 \in \mathbb{R}^n$, and for all $1 \leqslant i, j \leqslant n$,

$$\lim_{L\to\infty} \left\langle DX_i^{(L)}\,, DX_j^{(L)}\right\rangle_{R_1^Q} = Q_{i,j} \quad \text{in } L^1(\mathrm{P}_n).$$

Then, $X^{(L)}$ converges in distribution to $\mathrm{N}_n(0\,,Q)$ as $L \to \infty$.

Theorem 5.8 has other connections to results about asymptotic normality as well. The following shows how Theorem 5.8 is related to the classical CLT, for instance.

?⟨ex:NPR:1⟩? *Example* 5.11. Let $n \geqslant 2$, and suppose $\phi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ satisfies $\mathrm{E}_1(\phi) = 0$ and $\mathrm{Var}_1(\phi) = \sigma^2 < \infty$. Define

$$X_1 := \frac{1}{\sqrt{n}}\sum_{k=1}^n \phi(Z_k) \quad \text{and} \quad X_\ell := 0 \text{ for } 2 \leqslant \ell \leqslant n.$$

By the chain rule [Lemma 1.7, page 31],

$$(D_kX_1)(x) = \frac{\phi'(x_k)}{\sqrt{n}} \quad \text{and} \quad (D_kX_\ell)(x) = 0 \text{ for } 2 \leqslant \ell \leqslant n,$$

almost surely for every $1 \leqslant k \leqslant n$. We are writing ϕ' in place of the more cumbersome $D\phi$, as we have done before. In any case, we can see that, with probability one: $\langle DX_i\,, DX_j\rangle_{R_1^Q} = 0$ unless $i = j = 1$; and

$$\langle DX_1\,, DX_1\rangle_{R_1^Q} = \frac{1}{n}\sum_{k=1}^n \phi'(Z_k)(R_1^Q\phi')(Z_k).$$

Define $Y_k := \phi'(Z_k)(R_1^Q\phi')(Z_k)$, and observe that $Y_1, \cdots, Y_n$ are i.i.d. with

$$\mathrm{E}_n(Y_1) = \mathrm{E}_1\left[\langle D\phi\,, D\phi\rangle_{R_1^Q}\right] = \mathrm{Var}(\phi) = \sigma^2,$$

thanks to integration by parts [see the proof of Theorem 4.7, page 64]. Therefore, Khintchine's form of the weak law of large numbers implies that $\lim_{n\to\infty}\langle DX_1\,, DX_1\rangle_{R_1^Q} = \sigma^2$ in $L^1(\mathrm{P}_n)$. In particular, we can deduce from Theorem 5.8 that for every $\Phi \in C_c^2(\mathbb{R})$,

$$\left|\mathrm{E}\left[\Phi\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \phi(Z_k)\right)\right] - \mathrm{E}[\Phi(\sigma Z_1)]\right| \leqslant C\,\mathrm{E}\left(\left|\sigma^2 - \langle DX_1\,, DX_1\rangle_{R_1^Q}\right|\right)$$
$$\to 0 \qquad \text{as } n \to \infty,$$

where $C := \frac{1}{2}\sup_{x\in\mathbb{R}^n}\max_{1\leqslant i,j\leqslant n}|(D^2_{i,j}\Phi)(x)|$. That is, Theorem 5.8 and Khintchine's weak law of large numbers together imply the classical central limit theorem for sums of the form $n^{-1/2}\sum_{k=1}^n \phi(Z_k)$, where $\phi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$ has mean zero and finite variance.[6] Moreover, we can see from the preceding how to estimate the rate of convergence of the distribution of $n^{-1/2}\sum_{k=1}^n \phi(Z_k)$ to $\mathrm{N}(0\,,\sigma^2)$ in terms of the rate of convergence in Khintchine's weak law of large numbers. The latter is a very well-studied topic; see, for example XXX.

§5.3 Slepian's Inequality

Slepian's inequality is a useful comparison principle that can sometimes be used to estimate probabilities, or expectations, that are difficult to compute exactly. There are many variations of this inequality. Here is the original one that is actually due to D. Slepian.

⟨th:Slepian⟩ **Theorem 5.12** (Slepian, XXX). *Let X and Y be two mean-zero Gaussian random vectors on $\mathbb{R}^n$. Suppose that for every $1 \leqslant i, j \leqslant n$:*

1. $\mathrm{Var}(X_i) = \mathrm{Var}(Y_i)$; *and*
2. $\mathrm{Cov}(X_i\,, X_j) \leqslant \mathrm{Cov}(Y_i\,, Y_j)$.

Then for all $a_1, \ldots, a_n \in \mathbb{R}$,

$$\mathrm{P}\left\{X_i \leqslant a_i \ \ \forall\, 1 \leqslant i \leqslant n\right\} \leqslant \mathrm{P}\left\{Y_i \leqslant a_i \ \ \forall\, 1 \leqslant i \leqslant n\right\}.$$

In particular, $\mathrm{P}\left\{\max_{1\leqslant i\leqslant n} X_i \geqslant a\right\} \geqslant \mathrm{P}\left\{\max_{1\leqslant i\leqslant n} Y_i \geqslant a\right\}$ *for all $a \in \mathbb{R}$.*

The following is an immediate consequence of Theorem 5.12 and integration by parts. It states that less correlated Gaussian vectors tend to take on larger values.

⟨co:Slepian⟩ **Corollary 5.13.** *Under the assumptions of Theorem 5.12,*

$$\mathrm{E}\left[\max_{1\leqslant i\leqslant n} X_i\right] \geqslant \mathrm{E}\left[\max_{1\leqslant i\leqslant n} Y_i\right].$$

Proof. By integration by parts,

$$\mathrm{E}(W) = \int_0^\infty \mathrm{P}\{W > a\}\,\mathrm{d}a - \int_{-\infty}^0 \left(1 - \mathrm{P}\{W > a\}\right)\mathrm{d}a,$$

for all $W \in L^1(\mathrm{P})$. We apply this once with $W := \max_{i\leqslant n} X_i$ and once with $W := \max_{i\leqslant n} Y_i$, and then appeal to Theorem 5.12 to compare the two formulas. □

One can frequently use Corollary 5.13 in order to estimate the size of the expectation of the maximum of a Gaussian sequence. The following example highlights a simple example of the technique that is typically used.

[6] One can recast the classical CLT as the statement that the distribution of $n^{-1/2}\sum_{k=1}^n \phi(Z_k)$ is asymptotically normal for all $\phi \in L^2(\mathrm{P}_1)$ with $\mathrm{E}_1(\phi) = 0$. The present formulation is slightly weaker since we need the additional smoothness condition that $\phi \in \mathbb{D}^{1,2}(\mathrm{P}_1)$. It is possible to obtain the general form from the weaker one by an approximation argument.

⟨ex:Slepian⟩ *Example* 5.14. Suppose $X = (X_1\,, \ldots, X_n)$ is a Gaussian random vector with $\mathrm{E}(X_i) = 0$, $\mathrm{Var}(X_i) = 1$, and $\mathrm{Cov}(X_i\,, X_j) \leqslant 1-\varepsilon$ for some $\varepsilon \in (0\,,1]$. Let Z_0 be a standard normal random variable, independent of Z, and define

$$Y_i := \sqrt{1-\varepsilon}\, Z_0 + \sqrt{\varepsilon}\, Z_i.$$

Then clearly, $\mathrm{E}(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1$, and $\mathrm{Cov}(Y_i\,, Y_j) = 1-\varepsilon$ when $i \neq j$. Slepian's inequality implies that $\mathrm{E}[\max_{i\leqslant n} X_i] \geqslant \mathrm{E}[\max_{i\leqslant n} Y_i]$. Since $\max_{i\leqslant n} Y_i = \sqrt{1-\varepsilon}\, Z_0 + \sqrt{\varepsilon}\, \max_{i\leqslant n} Z_i$, we find from Proposition 1.3 [page 7] that

$$\mathrm{E}\left[\max_{1\leqslant i\leqslant n} X_i\right] \geqslant \mathrm{E}\left[\max_{1\leqslant i\leqslant n} Y_i\right] = \sqrt{\varepsilon}\, \mathrm{E}\left[\max_{1\leqslant i\leqslant n} Z_i\right] = (1+o(1))\sqrt{2\varepsilon\log n},$$

as $n\to\infty$. This is sharp, up to a constant. In fact, the same proof as in the i.i.d. case shows us the following: For *any* sequence $X_1, \ldots, X_n$ of mean-zero, variance-one Gaussian random variables,

$$\mathrm{E}\left[\max_{1\leqslant i\leqslant n} X_i\right] \leqslant (1+o(1))\sqrt{2\log n} \quad \text{as } n\to\infty.$$

[For this, one does not even need to know that $(X_1\,, \ldots, X_n)$ has a multivariate normal distribution.]

Example 5.15. We proceed as we did in the previous example and note that if $W := (W_1, \ldots, W_n)$ is a Gaussian random vector with $\mathrm{E}(W_i) = 0$, $\mathrm{Var}(W_i) = 1$, and $\mathrm{Cov}(W_i\,, W_j) \geqslant -1+\delta$ for some $\delta\in(0\,,1]$, then

$$\mathrm{E}\left[\max_{1\leqslant i\leqslant n} W_i\right] \leqslant (1+o(1))\sqrt{2\delta\log n} \qquad \text{as } n\to\infty.$$

Proof of Theorem 5.12. Let Q^X and Q^Y denote the respective covariance matrices of X and Y and let S^X and S^Y denote the respective square roots of Q^X and Q^Y.

Without loss of generality, we assume that X and Y are defined on the same Gauss space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\,, \mathrm{P}_n)$, and defined as $X = S^X Z$ and $Y = S^Y Z$. Since $\langle DX_i\,, DX_j\rangle_{R_1} = Q^X_{i,j}$, Theorem 5.8 shows that for all $\Phi\in\mathbb{D}^{1,2}(\mathrm{P}_n)$

$$\mathrm{E}\left[\Phi(X)\right] - \mathrm{E}\left[\Phi(Y)\right] = \sum_{i,j=1}^n \mathrm{E}\left[\left(R_2^{Q^Y} D^2_{i,j}\Phi\right)(X)\right] \times \left(Q^X_{i,j} - Q^Y_{i,j}\right).$$

Suppose, in addition, that $D^2_{i,j}\Phi \geqslant 0$ a.s. when $i\neq j$. Because $Q^X_{i,j} \leqslant Q^Y_{i,j}$ and $Q^X_{i,i} = Q^Y_{i,i}$, it follows that $\mathrm{E}[\Phi(X)] \leqslant \mathrm{E}[\Phi(Y)]$. In particular,

$$\mathrm{E}\left[\prod_{i=1}^n \varphi_i(X_i)\right] \leqslant \mathrm{E}\left[\prod_{i=1}^n \varphi_i(Y_i)\right],$$

whenever $\varphi_1, \ldots, \varphi_n \in C^2_0(\mathrm{P}_1)$ are non increasing and non negative. Approximate every $\mathbb{1}_{(-\infty,a_i]}$ by a non-increasing function $\varphi_i \in C^2_0(\mathrm{P}_n)$ to finish. □

The following inequality of Fernique XXX refines Slepian's inequality in a certain direction.

⟨th:Fernique⟩ **Theorem 5.16** (Fernique, XXX). *Let X and Y be two mean-zero Gaussian random vectors on $\mathbb{R}^n$. Suppose that for every $1 \leqslant i, j \leqslant n$:*

$$\mathrm{E}\left(|X_i - X_j|^2\right) \geqslant \mathrm{E}\left(|Y_i - Y_j|^2\right). \tag{5.27}$$ cond:Fernique

Then, $\mathrm{P}\{\max_{1\leqslant i\leqslant n} X_i \geqslant a\} \geqslant \mathrm{P}\{\max_{1\leqslant i\leqslant n} Y_i \geqslant a\}$ *for all $a \in \mathbb{R}$. In particular,* $\mathrm{E}[\max_{1\leqslant i\leqslant n} X_i] \geqslant \mathrm{E}[\max_{1\leqslant i\leqslant n} Y_i]$.

If, in addition, $\mathrm{Var}(X_i) = \mathrm{Var}(Y_i)$ for all $1 \leqslant i \leqslant n$, then condition (5.27) reduces to the covariance condition of Slepian's inequality. Therefore, we can view Fernique's inequality as an improvement of Slepian's inequality to the setting of non-stationary Gaussian random vectors. The proof itself is a variation on the proof of Theorem 5.12, but the variation is non trivial and involves many computations. The idea is, as before, to show that, for $\Phi(x) := \max_{1\leqslant i\leqslant n} x_i$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}\left[\Phi\left(\mathrm{e}^{-t}X + \sqrt{1-\mathrm{e}^{-2t}}\,Y\right)\right] \geqslant 0,$$

whence $\mathrm{E}[\Phi(X)] \leqslant \mathrm{E}[\Phi(Y)]$. You can find the numerous details, for example, in Ledoux and Talagrand XXX and Marcus and Rosen XXX.

Problems

⟨ex:log-normal⟩ 1. Consider the case $n = 1$, and use the Nourdin–Peccati formula (Theorem 4.4) to derive the *lognormal density*; that is, the probability density function of the random variable $f = \exp(Z)$. In your computation of $R_1(Df)$, you will have to use the error function,

$$\Phi(x) = \int_{-\infty}^{x} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}}\,\mathrm{d}u \qquad [x \in \mathbb{R}],$$

and recall that you have to center f before you can apply the formula.

2. Verify the Paley–Zygmund inequality (5.6), p. 76.

⟨pbm:E(Y|X)⟩ 3. Let X and Y be two random variables, defined both on the same abstract probability space $(\Omega\,, \mathcal{F}, \mathrm{P})$. Suppose X takes values in $\mathbb{R}^n$ and Y is real-valued and integrable [P]. Then prove that there exists a Borel-measurable function $G_{Y|X} : \mathbb{R}^n \to \mathbb{R}$ such that $\mathrm{E}(Y \mid X) = G_{Y\,|\,X}(X)$ a.s.

4. Prove Lemma 3.2, p. 73.

⟨ex:SK⟩ 5. Let $\Pi_n(Z)$ be defined via (5.3). Prove that if $|\beta| \geqslant 1$, then

$$\lim_{n\to\infty} \frac{\mathrm{E}\left(|\Pi_n(Z)|^2\right)}{|\mathrm{E}\,[\Pi_n(Z)]|^2} = \infty.$$

6. Prove that $\mathbb{1}_A \in \mathbb{D}^{1,2}(\mathrm{P}_n)$ iff $\mathrm{P}_n(A) = 0$ or 1 for every Borel set $A \subseteq \mathbb{R}^n$. This is due to Sekiguchi and Shiota XXX. (Hint: Start by checking that $D(\mathbb{1}_A) = 2\mathbb{1}_A D(\mathbb{1}_A)$ a.s. because $\mathbb{1}_A = \mathbb{1}_A^2$.)

7. Compute explicitly the linear operator $\mathcal{L}^Q$ in (5.26).

8. Suppose that $(X\,, Y)$ is distributed as $\mathrm{N}_2(0\,, \Sigma)$, where $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\Sigma_{1,2} = \Sigma_{2,1} = \rho$, where $-1 \leqslant \rho \leqslant 1$ is a fixed number.
   (a) Verify that one can construct a version of $(X\,, Y)$ on the Gauss space $(\mathbb{R}^2\,, \mathcal{B}(\mathbb{R}^2)\,, \mathrm{P}_2)$ as follows: $X := Z_1$ and $Y := \rho Z_1 + \sqrt{1-\rho^2}\,Z_2$.
   (b) Use the preceding to compute $\mathrm{E}\exp\{wX + vY\}$ for all $w, v \in \mathbb{R}$. Conclude from your calculation that

$$\mathrm{E}\,[H_k(X)H_m(Y)] = \begin{cases} \dfrac{\rho^k}{k!} & \text{if } k = m,\\ 0 & \text{if } k \neq m.\end{cases}$$

⟨pbm:CLT:stat⟩ 9. Let $X := \{X_i\}_{i=1}^{\infty}$ be a stochastic process, defined on a probability space $(\Omega\,, \mathcal{F}, \mathrm{Q})$. Suppose that, for every integer $N \geqslant 1$, $(X_1\,, \ldots, X_N)$ has a multivariate normal distribution with mean vector zero and $\mathrm{Cov}(X_i\,, X_j) = \rho(|i-j|)$ for all $i, j = 1, \ldots, N$, for a function $\rho : \mathbb{Z}_+ \to [-1\,, 1]$ that satisfies $\rho(0) = 1$. Then, X is said to be a *centered, stationary Gaussian process with correlation function ρ*. It is known that such a stochastic process exists if and only if ρ is positive semi-definite; see XXX. Choose and fix a function $\phi \in L^2(\mathrm{P}_1)$, and suppose X exists. For simplicity also assume that $\rho(i) \geqslant 0$ for all $i \geqslant 1$.
   (a) Prove that if $\sum_{i=1}^{\infty} \rho(i) < \infty$, then the following CLT-type variance condition holds:

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N}\,\mathrm{Var}\left(\sum_{i=1}^{N} \phi(X_i)\right) \qquad \text{exists and is finite.}$$

   Compute σ^2.

[Hint: Since each X_i has the same law as Z_1, we can expand $\phi(X_i)$ in terms of Hermite polynomials.]

(b) Prove that $\lim_{N\to\infty}\rho(N) = 0$ if and only if the following law of large numbers holds:

$$\frac{1}{N}\sum_{i=1}^{N}\phi(X_i) \to \mathrm{E}[\phi(X_1)] \qquad \text{in } L^2(\mathrm{Q}) \text{ as } N\to\infty.$$

10. Let $X = \{X_i\}_{i=1}^{\infty}$ be a centered, stationary Gaussian process, as defined in Problem 9. Suppose in addition that $\mathrm{Cov}(X_0\,, X_i) \geqslant 0$ for every $i\in\mathbb{N}$.

(a) Prove that $\mu_{n+m}(a) \geqslant \mu_n(a)\mu_m(a)$ for all $n, m \in \mathbb{Z}_+$ and $a\in\mathbb{R}$, where

$$\mu_n(a) := \mathrm{P}\left\{\max_{1\leqslant i\leqslant n} X_i \leqslant a\right\} \qquad \text{for all } n\geqslant 1 \text{ and } a\in\mathbb{R}.$$

(b) A sequence $\{y_i\}_{i=1}^{\infty}$ is said to be *subadditive* if $y_{n+m}\leqslant y_n + y_m$ for every $n, m\in\mathbb{N}$. Prove that if $\{y_i\}_{i=1}^{\infty}$ is subadditive then $\lim_{n\to\infty}(y_n/n)$ exists and is finite. Do this by showing that

$$\lim_{n\to\infty}\frac{y_n}{n} = \inf_{n\geqslant 1}\frac{y_n}{n}.$$

This is due to M. Fekete XXX.

(c) Conclude that for every $a\in\mathbb{R}$ there exists an extended real number $\theta(a)\in(0\,,\infty]$ such that $\mu_n(a) = \exp\{-\theta(a)n + o(n)\}$ as $n\to\infty$.

Chapter 6

ce_Stochastic_Integrals)?

Four Moment Theorems

Consider the following simple question: suppose that A is a symmetric $n \times n$ matrix with $\operatorname{tr}(A) = 0$. Then, by Wick's formula, the quadratic form $Z'AZ$ is in the second Wiener chaos (see Exercise 14, Chapter 3). Moreover, $\operatorname{tr}(A) = 0$ means that A has both positive and negative eigenvalues and so the quadratic form $Z'AZ$ takes on all possible values of $\mathbb{R}$ (excluding the trivial case $A = 0$). Is it possible that the random variable $Z'AZ$ is normally distributed?

The intuitive answer is likely "no", since normality is preserved by linear transformations but typically not quadratic or higher order ones. This intuition turns out to be correct, but it is actually more difficult to prove than one might expect. After all, how can one rule out that there is no such matrix A with this property?

In this chapter we will answer this question for random variables in $L^2(\mathrm{P}_n)$ that live within a fixed Wiener chaos. Such random variables often appear naturally in statistics, for example the sample standard deviation

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2,$$

where $\bar{Z} = (Z_1 + \ldots + Z_n)/n$ is the sample mean, has mean one and $S^2 - 1$ is in the second Wiener chaos. It turns out that the squared norm of the gradient of such random variables has a nice expression, which can be used to give a simple test for normality of the original variable in terms of its fourth moment. In short, a random variable living inside a fixed chaos is normal iff its fourth moment is equal to three times the square of its second moment, and if the latter holds then it in fact lives inside the first chaos. Such results are called *four moment theorems*.

1 Random Variables Living in a Fixed Chaos

We begin working in $L^2(\mathrm{P}_n)$ so that the Hermite polynomials $\mathcal{H}_k/\sqrt{k!}, k \in \mathbb{Z}_+^n$ form an orthonormal basis of the space. Random variables within the pth Wiener chaos are exactly the linear combinations of $\mathcal{H}_k$ with $|k| = k_1 + \ldots + k_n = p$. We define

$\mathcal{S}_p^n = \{k \in \mathbb{Z}_+^n : |k| = p\}$ and a map $I_p : \mathbb{R}^{\mathcal{S}_p^n} \to \mathbb{R}$ by

$$I_p(a) = \sum_{k \in \mathcal{S}_p^n} \binom{p}{k} a_k \mathcal{H}_k. \tag{6.1}$$

eq:Ip_defn

Then I_p is a linear map whose range is precisely the pth Wiener chaos of $L^2(\mathrm{P}_n)$, and as a map onto this range it is invertible. Recall that the multinomial coefficients are given by

$$\binom{p}{k} = \frac{p!}{k_1! k_2! \dots k_n!}$$

and count the number of ways that p distinct objects can be put into n groups with k_1 objects in the first group, k_2 in the second group, etc. These multinomial factors turn out to be a convenient normalization, and we will later see that they are quite natural. By the orthogonality of the Hermite polynomials it is straightforward to compute that for $a, b \in \mathbb{R}^{\mathcal{S}_p^n}$

$$\mathrm{E}[I_p(a) I_p(b)] = p! \sum_{k \in \mathcal{S}_p^n} \binom{p}{k} a_k b_k =: p! \langle a, b \rangle_{\mathcal{S}_p^n}, \tag{6.2}$$

eq:HS_isometry

where latter equality defines the inner product. Similarly, if $a \in \mathbb{R}^{\mathcal{S}_p^n}, b \in \mathbb{R}^{\mathcal{S}_q^n}$ with $p \neq q$ then $\mathrm{E}[I_p(a) I_q(b)] = 0$. By the differentiation properties of Hermite polynomials it follows that the gradient of $I_p(a)$ is

$$\begin{aligned} D I_p(a) &= \sum_{k \in \mathcal{S}_p^n} \binom{p}{k} a_k \sum_{i=1}^n k_i \mathcal{H}_{k - e_i} e_i \\ &= p \sum_{i=1}^n \sum_{k \in \mathcal{S}_p^n} \binom{p-1}{k - e_i} a_k \mathcal{H}_{k - e_i} e_i \\ &= p \sum_{i=1}^n \sum_{l \in \mathcal{S}_{p-1}^n} \binom{p-1}{l} a_{l + e_i} \mathcal{H}_l e_i = p \sum_{i=1}^n I_{p-1}(a_{\cdot + e_i}) e_i, \end{aligned} \tag{6.3}$$

eqn:DIp_formula

where e_i is the ith standard basis vector in $\mathbb{R}^n$. Note that we are adopting the convention that $H_k = 0$ for $k < 0$, and therefore $\mathcal{H}_k = 0$ if any $k_i < 0$. The notation $a_{\cdot + e_i}$ refers to the function on $\mathcal{S}_{p-1}^n$ defined by adding e_i to each index to obtain a value from a. Note that the last expression clearly shows that each component of the gradient is an element of the $(p-1)$st chaos, as one would expect. Now from this formula for the gradient we can compute its squared norm as

$$||D I_p(a)||_{\mathbb{R}^n}^2 = p^2 \sum_{i=1}^n I_{p-1}(a_{\cdot + e_i})^2. \tag{6.4}$$

eqn:DIp_squared_norm_form

This real-valued random variable is a polynomial in the Z_i of degree no more than $2p - 2$. Hence it is in $L^2(\mathrm{P}_n)$ and so has a Wiener chaos decomposition that allows for it to be uniquely written as an orthogonal sum of polynomials of degree $2p-2, 2p-3, 2p-4$, etc. The claim is that the structure of $I_p(a)$ allows for the decomposition to be computed explicitly. Essentially it requires nothing more than computing the Wiener chaos decomposition of $H_k(x) H_l(x)$ for arbitrary $k, l \in \mathbb{N}$. We do this in the next section and then use it to expand out the squared norm of the gradient in terms of Hermite polynomials. This leads to a nice expression for $\mathrm{Var}(||DI_p(a)||^2_{\mathbb{R}^n})$, which turns out to be related to the excess kurtosis $\mathrm{E}[I_p(a)^4] - 3\,\mathrm{E}[I_p(a)^2]^2$ in the following way:

thm:chaos_variance_bound⟩ **Theorem 1.1.** *For $p \geqslant 2$ and $a \in \mathbb{R}^{\mathcal{S}_p^n}$ we have*

$$\mathrm{Var}\left(\frac{1}{p}||DI_p(a)||^2_{\mathbb{R}^n}\right) \leqslant \frac{p-1}{3p}\left(\mathrm{E}[I_p(a)^4] - 3\,\mathrm{E}[I_p(a)^2]^2\right) \leqslant (p-1)\,\mathrm{Var}\left(\frac{1}{p}||DI_p(a)||^2_{\mathbb{R}^n}\right).$$

This statement quite elegantly shows the equivalence of the excess kurtosis with the variance of the squared norm of the gradient. The fact that $I_p(a)$ is in a fixed chaos turns out to play a key role in deriving these bounds, but note that they hold for arbitrarily large p. It is worth observing that they also trivially hold for $p = 1$ (all three terms are zero) since in that case $DI_1(a)$ is almost surely a constant and the normality of $I_1(a)$ implies that the excess kurtosis is zero. In the $p \geqslant 2$ case these bounds immediately lead to the following non-trivial conclusion.

on_fixed_chaos_normality⟩ **Corollary 1.2.** *For $p \geqslant 2$ and $a \in \mathbb{R}^{\mathcal{S}_p^n}$ different from zero, $\mathrm{E}[I_p(a)^4] > 3\,\mathrm{E}[I_p(a)^2]^2$ and hence $I_p(a)$ is not normally distributed.*

Proof. By Theorem 1.1, $\mathrm{E}[I_p(a)^4] = 3\,\mathrm{E}[I_p(a)^2]^2$ iff $||DI_p(a)||^2_{\mathbb{R}^n}||$ is constant almost surely. By equation (6.3) every component of $DI_p(a)$ is a polynomial of degree $p-1$ in the Z_i, from which it follows that $||DI_p(a)||^2_{\mathbb{R}^n}$ is a polynomial of degree $2p-2$. Since $2p-2 \geqslant 2$ this polynomial cannot be almost surely constant, and since $\mathrm{E}[I_p(a)^4] < 3\,\mathrm{E}[I_p(a)^2]^2$ is impossible by Theorem 1.1 the only option is $\mathrm{E}[I_p(a)^4] > 3\,\mathrm{E}[I_p(a)^2]^2$. Since the fourth moment of a mean zero Gaussian is three times the square of its second moment, this rules out that $I_p(a)$ is normally distributed. ☐

The next few sections will be geared towards proving Theorem 1.1. The idea is to compute the variance of $||DI_p(a)||^2_{\mathbb{R}^n}$ by expanding it into its Wiener chaos decomposition and then using the orthogonality of the Hermite polynomials. Similar ideas help us to compute $\mathrm{E}[I_p(a)^4]$. We first perform the computations in a very hands on, combinatorial way, and then in later parts we show how they can be repeated quite elegantly using the language of tensors. Tensor language is more useful in the infinite-dimensional setting with the isonormal representation of the Gaussian space, but in the finite-dimensional case it involves nothing more than standard multi-linear algebra. The tensor language also explains how the I_p operators can be thought of as discrete stochastic integrals and are the natural adjoints to the higher order differentiation operators D^k on Gauss space. Before going into the lengthy calculations we point out that there is one term in the Wiener chaos expansion of $||DI_p(a)||^2$ that can be easily computed: the mean.

⟨lem:mean_of_DIp⟩ **Lemma 1.3.** *For $a \in \mathcal{S}_p^n$*

$$\mathrm{E}\left[||DI_p(a)||^2_{\mathbb{R}^n}\right] = p\,\mathrm{E}[I_p(a)^2].$$

Proof. By formulas (6.4) and (6.2) we have

$$\mathrm{E}[||DI_p(a)||^2_{\mathbb{R}^n}] = p^2\sum_{i=1}^n \mathrm{E}[I_p(a_{\cdot+e_i})^2] = p^2\sum_{i=1}^n (p-1)!\sum_{l\in\mathcal{S}_{p-1}^n}\binom{p-1}{l}a^2_{l+e_i}.$$

We also apply (6.2) to $\mathrm{E}[I_p(a)^2]$ to obtain

$$\mathrm{E}[I_p(a)^2] = p! \sum_{k \in \mathcal{S}_p^n} \binom{p}{k} a_k^2.$$

Now use the multinomial identity

$$\binom{p}{k} = \sum_{i=1}^{n} \binom{p-1}{k-e_i}, \tag{6.5}$$ eq:multinom_id

to rewrite the above as

$$\begin{aligned} \mathrm{E}[I_p(a)^2] &= p! \sum_{i=1}^{n} \sum_{k \in \mathcal{S}_p^n} \binom{p-1}{k-e_i} a_k^2 \\ &= p \sum_{i=1}^{n} (p-1)! \sum_{l \in \mathcal{S}_{p-1}^n} \binom{p-1}{l} a_{l+e_i}^2 = p \sum_{i=1}^{n} \mathrm{E}[I_{p-1}(a_{\cdot+e_i})^2]. \end{aligned}$$

This completes the proof. □

2 Product Formula for Hermite Polynomials

⟨prop:Hermite_product⟩ **Proposition 2.1.** *Fix $k, l \in \mathbb{Z}_+^n$. Then*

$$\mathcal{H}_k(x)\mathcal{H}_l(x) = \sum_{j \in [0, k\wedge l]} j! \prod_{i=1}^{n} \binom{k_i}{j_i} \binom{l_i}{j_i} \mathcal{H}_{k+l-2j}(x),$$

where $k\wedge l = (k_1\wedge l_1, \dots, k_n \wedge l_n)$ and $[0, k\wedge l] = \{j \in \mathbb{Z}_+^n : 0 \leqslant j_i \leqslant k_i \wedge l_i, i = 1, \dots, n\}$.

Proof. Since the multi-dimensional Hermite polynomials are products of the one-dimensional Hermite polynomials it is enough to study the $n = 1$ case. Recall that $H_k(Z) = \pi_k(Z^k)$, and therefore it is equivalent to show that

$$\pi_k(Z^k)\pi_l(Z^l) = \sum_{j=0}^{k\wedge l} j! \binom{k}{j} \binom{l}{j} \pi_{k+l-2j}(Z^{k+l-2j}).$$

Wick's formula explains this nicely. We have one group consisting of Z repeated k times and a second group consisting of Z repeated l times. Then we can have any number of matchings j where $0 \leqslant j \leqslant k \wedge l$, and to have precisely j matchings we choose j elements from the first group, j from the second, and then can match within those two groups exactly $j!$ ways. This explains the combinatorial terms in the sum, and when there is exactly j matchings there are $k - j + l - j$ unmatched Z's whose product we then project into the corresponding chaos. This completes the proof for the $n = 1$ case, and now for the $n > 1$ case simply expand out the identity

$$\mathcal{H}_k\mathcal{H}_l = \prod_{i=1}^{n} H_{k_i} H_{l_i}.$$

□

Proof. Since the multi-dimensional Hermite polynomials are products of the one-dimensional ones it is enough to study the $n = 1$ case. First note that $H_k(x)H_l(x)$ is a polynomial of degree $k + l$, thus it can be uniquely decomposed into a sum

$$H_k(x)H_l(x) = \sum_{m=0}^{k+l} a_m H_m(x)$$

for some constants a_m that we will now compute. Of course we already know $a_{k+l} = 1$ since H_k and H_l are both monic, but the following argument will take care of that case as well. Differentiate i times and use the differentiation formula for Hermite polynomials to obtain

$$D^i(H_k H_l) = \sum_{m=0}^{k+l} a_m \frac{m!}{(m-i)!} H_{m-i}$$

with the convention that $H_{-p} = 0$ for $p > 0$. Since all Hermite polynomials but $H_0 = 1$ are mean zero with respect to the Gaussian measure P_1, it follows that

$$a_m = \frac{1}{m!}\,\mathrm{E}[D^m(H_k H_l)(Z)]$$

By the Leibniz rule we have

$$D^m(H_k H_l) = \sum_{i=0}^{m} \binom{m}{i} D^i H_k D^{m-i} H_l = \sum_{i=0}^{m} \binom{m}{i} \frac{k!}{(k-i)!} H_{k-i} \frac{l!}{(l-m+i)!} H_{l-m+i}.$$

Now by the orthogonality of the Hermite polynomials it follows that the expected value of the above is zero unless $k - i = l - m + i$, or $2i = k - l + m$. Thus $k - l + m$ must be even, which is equivalent to $k - l$ and m having the same parity and is in turn equivalent to $k + l$ and m having the same parity. Thus we can write $m = k + l - 2j$ for $j \geqslant 0$. Since $0 \leqslant i \leqslant m$ and $2i = k - l + m$ this also requires $-m \leqslant k - l \leqslant m$, which in terms of j requires $j \leqslant k$ and $j \leqslant l$. Thus for $0 \leqslant j \leqslant k \wedge l$ we have $i = k - j$, $k - i = j$, and $l - m + i = j$, so that

$$\mathrm{E}[D^{k+l-2j}(H_k H_l)(Z)] = \binom{k+l-2j}{k-j} \frac{k!}{j!} \frac{l!}{j!} j!,$$

with the last $j!$ term in the product following from $\mathrm{E}[H_p^2] = p!$ for $p \geqslant 0$. Therefore we obtain

$$a_{k+l-2j} = \frac{k!l!}{j!(k-j)!(l-j)!} = j! \binom{k}{j} \binom{l}{j},$$

with $0 \leqslant j \leqslant k \wedge l$. This leads to the $n = 1$ identity

$$H_k(x)H_l(x) = \sum_{j=0}^{k \wedge l} j! \binom{k}{l} \binom{j}{l} H_{k+l-2j}(x).$$

Now for $n > 1$ use this identity and expand out the defining relation

$$\mathcal{H}_k \mathcal{H}_l = \prod_{i=1}^{n} H_{k_i} H_{l_i}.$$

$\square$

Note that the restriction on j in the summation formula ensures that the inputs to the binomial coefficients are valid. If we adopt the convention that all binomial or multinomial coefficients are zero if the inputs are not valid then we can drop the restriction on j entirely, which turns out to be useful in later computations. This allows us to write

$$\mathcal{H}_k\mathcal{H}_l = \sum_{j\in\mathbb{Z}_+^n} j!\prod_{i=1}^n \binom{k_i}{j_i}\binom{l_i}{j_i}\mathcal{H}_{k+l-2j} = \sum_{r=0}^{|k|\wedge|l|}\sum_{j\in\mathcal{S}_r^n} j!\prod_{i=1}^n \binom{k_i}{j_i}\binom{l_i}{j_i}\mathcal{H}_{k+l-2j}.$$

The second equality groups the terms according to the order of the Wiener chaos they appear in. Using these formulas we come to the following more general product formula.

⟨defn:contraction_one⟩ **Definition 2.2.** Let $a \in \mathbb{R}^{\mathcal{S}_p^n}$ and $b \in \mathbb{R}^{\mathcal{S}_q^n}$ and $0 \leqslant r \leqslant p\wedge q$. Define $a \star_r b \in \mathcal{S}_{p+q-2r}$ by

$$\binom{p+q-2r}{k}(a\star_r b)_k = \sum_{\substack{m\in\mathcal{S}_r^n\\ l\in\mathcal{S}_{p-r}^n}} \binom{r}{m}\binom{p-r}{l}\binom{q-r}{k-l} a_{l+m}b_{k-l+m},$$

for $k \in \mathcal{S}_{p+q-2r}^n$. In the case $p = q = r$ the operation produces the number

$$(a\star_p b) = \sum_{m\in\mathcal{S}_p^n}\binom{p}{m}a_m b_m = \langle a, b\rangle_{\mathcal{S}_p^n},$$

the latter by (6.2).

This definition produces the following compact formula for the product of discrete stochastic integrals.

:discrete_stoch_integral_product_orig⟩ **Proposition 2.3.** *Let $a \in \mathbb{R}^{\mathcal{S}_p^n}$ and $b \in \mathbb{R}^{\mathcal{S}_q^n}$. Then*

$$I_p(a)I_q(b) = \sum_{r=0}^{p\wedge q} r!\binom{p}{r}\binom{q}{r} I_{p+q-2r}(a\star_r b).$$

Proof. We simply use the product formula of Proposition 2.1 to expand out the stochastic integrals. This gives

$$\begin{aligned}
I_p(a)I_q(b) &= \sum_{k\in\mathcal{S}_p^n}\sum_{l\in\mathcal{S}_q^n}\binom{p}{k}\binom{q}{l}a_k b_l \mathcal{H}_k\mathcal{H}_l\\
&= \sum_{k,l}\sum_{r=0}^{p\wedge q}\sum_{j\in\mathcal{S}_r^n} a_k b_l\binom{p}{k}\binom{q}{l} j!\prod_{i=1}^n\binom{k_i}{j_i}\binom{l_i}{j_i}\mathcal{H}_{k+l-2j}\\
&= \sum_{r=0}^{p\wedge q}\sum_{j\in\mathcal{S}_r^n}\sum_{\substack{u\in\mathcal{S}_{p-r}^n\\ v\in\mathcal{S}_{q-r}^n}} a_{u+j}b_{v+j}\binom{p}{u+j}\binom{q}{v+j} j!\prod_{i=1}^n\binom{u_i+j_i}{j_i}\binom{v_i+j_i}{j_i}\mathcal{H}_{u+v},
\end{aligned}$$

where in the last equality we used the substitution $u = k - j$, $v = l - j$. The definition of the multinomial coefficients implies that

$$\binom{p}{u+j}\prod_{i=1}^n\binom{u_i+j_i}{j_i} = \frac{p!}{(p-r)!}\frac{1}{j!}\binom{p-r}{u},$$

with a similar expression with q and v replacing p and u, respectively, leading to

$$j!\binom{p}{u+j}\binom{q}{v+j}\prod_{i=1}^{n}\binom{u_i+j_i}{j_i}\binom{v_i+j_i}{j_i} = r!\binom{p}{r}\binom{q}{r}\binom{r}{m}\binom{p-r}{u}\binom{q-r}{v}.$$

Substituting this into the above produces the identity

$$I_p(a)I_q(b) = \sum_{r=0}^{p\wedge q} r!\binom{p}{r}\binom{q}{r}\sum_{j\in\mathcal{S}_r^n}\binom{r}{m}\sum_{\substack{u\in\mathcal{S}_{p-r}^n\\ v\in\mathcal{S}_{q-r}^n}} a_{u+j}b_{v+j}\binom{p}{u+j}\binom{q}{v+j}\mathcal{H}_{u+v}.$$

Now simply redefine k and l by $k=u+v$, $l=u$, so that $k\in\mathcal{S}_{p+q-2r}$ and $l\in\mathcal{S}_{p-r}^n$, and change variables in the last sum to obtain

$$I_p(a)I_q(b) = \sum_{r=0}^{p\wedge q} r!\binom{p}{r}\binom{q}{r}\sum_{k\in\mathcal{S}_{p+q-2r}^n}\left[\sum_{\substack{m\in\mathcal{S}_r^n\\ l\in\mathcal{S}_{p-r}^n}}\binom{r}{m}\binom{p-r}{l}\binom{q-r}{k-l}a_{l+m}b_{k-l+m}\right]\mathcal{H}_k.$$

To complete the proof simply compare this last equality with Definition 2.2 and the definition (6.1) of the discrete stochastic integral. □

An immediate corollary of the product formula is an expression for the variance of $||DI_p(a)||^2_{\mathbb{R}^n}$.

⟨prop:DIp_variance⟩ **Proposition 2.4.** *For $p\geqslant 1$ and $a\in\mathbb{R}^{\mathcal{S}_p^n}$*

$$\operatorname{Var}\left(\frac{1}{p}||DI_p(a)||^2_{\mathbb{R}^n}\right) = \frac{1}{p^2}\sum_{r=1}^{p-1} r^2(r!)^2\binom{p}{r}^4(2p-2r)!\,||a\star_r a||^2_{\mathcal{S}_{2p-2r}^n},$$

where the latter norm is defined by (6.2).

Proof. By formula (6.4) and Proposition 2.3 we have

$$\begin{aligned}||DI_p(a)||^2_{\mathbb{R}^n} = p^2\sum_{i=1}^n I_{p-1}(a_{\cdot+e_i})^2 &= p^2\sum_{i=1}^n\sum_{r=0}^{p-1} r!\binom{p-1}{r}^2 I_{2p-2-2r}(a_{\cdot+e_i}\star_r a_{\cdot+e_i})\\ &= p^2\sum_{r=0}^{p-1} r!\binom{p-1}{r}^2 I_{2p-2-2r}\left(\sum_{i=1}^n a_{\cdot+e_i}\star_r a_{\cdot+e_i}\right).\end{aligned}$$

The last equality is by linearity of the stochastic integral. The $r=p-1$ term is

$$p^2\sum_{i=1}^n(p-1)!\sum_{l\in\mathcal{S}_{p-1}^n}\binom{p-1}{l}a_{l+e_i}^2 = \mathrm{E}[||DI_p(a)||^2_{\mathbb{R}^n}],$$

which follows from (6.2) (and was already proved in Lemma 1.3). Thus the $r=p-1$ term is a constant and hence irrelevant for the variance, and since the remaining terms are mean zero and uncorrelated for different r we obtain

$$\operatorname{Var}(||DI_p(a)||^2_{\mathbb{R}^n}) = p^4\sum_{r=0}^{p-2}(r!)^2\binom{p-1}{r}^4(2p-2-2r)!\left|\left|\sum_{i=1}^n a_{\cdot+e_i}\star_r a_{\cdot+e_i}\right|\right|^2_{\mathcal{S}_{2p-2-2r}^n},$$

the latter following by (6.2). Now shift the indexing of r by one and combine the p^4 term with the binomial coefficient to obtain

$$\mathrm{Var}(||DI_p(a)||^2_{\mathbb{R}^n}) = \sum_{r=1}^{p-1} r^2 (r!)^2 \binom{p}{r}^4 (2p-2r)! \left|\left| \sum_{i=1}^{n} a_{\cdot+e_i} \star_{r-1} a_{\cdot+e_i} \right|\right|^2_{\mathcal{S}^n_{2p-2r}}.$$

Finally, we only need to show that

$$\sum_{i=1}^{n} a_{\cdot+e_i} \star_{r-1} a_{\cdot+e_i} = a \star_r a.$$

This is a straightforward consequence of Definition 2.2 for the $\star$ operation and the multinomial identity (6.5). □

Using the product formula for Hermite polynomials we also get a similar expression for the excess kurtosis of a discrete stochastic integral.

⟨prop:Ip_kurtosis⟩ **Proposition 2.5.** *For $p \geqslant 1$ and $a \in \mathbb{R}^{\mathcal{S}^n_p}$*

$$\mathrm{E}[I_p(a)^4] - 3\,\mathrm{E}[I_p(a)^2]^2 = \frac{3}{p} \sum_{r=1}^{p-1} r(r!)^2 \binom{p}{r}^4 (2p-2r)!\, ||a \star_r a||^2_{\mathcal{S}^n_{2p-2r}}$$

Proof. Expand out $I_p(a)^4$ using the defining formula (6.3) for $I_p(a)$ and observe that because of the orthogonality of the Hermite polynomials the only terms that are not mean zero are those in which there are two pairs of distinct indices or all four indices are the same. Thus

$$\mathrm{E}[I_p(a)^4] = 3 \sum_{\substack{k,l \in \mathcal{S}^n_p \\ k \neq l}} \binom{p}{k}^2 \binom{p}{l}^2 a_k^2 a_l^2\, \mathrm{E}[\mathcal{H}_k^2 \mathcal{H}_l^2] + \sum_{k \in \mathcal{S}^n_p} \binom{p}{k}^4 a_k^4\, \mathrm{E}[\mathcal{H}_k^4].$$

On the first term the factor of 3 is because there are three distinct ways to choose which of the four indices forms the matching. !!!NEED TO FINISH!!! □

The proof of Theorem 1.1 is now a straightforward consequence of the last two propositions.

Proof of Theorem 1.1. Theorem 1.1 follows by direct comparison of the formulas in Propositions 2.4 and 2.5. Since the ratio of each term in the variance to the corresponding term in the excess kurtosis is $r/3q$, the left hand inequality follows from $r \leqslant p-1$. Similarly, since the ratio of each term in $\kappa_4(I_p(a))(q-1)/3q$ ($\kappa_4(I_p(a))$ being the excess kurtosis) to the corresponding term in the variance is $(q-1)/r$, the right hand inequality follows from $r \geqslant 1$. □

3 Tensorization

In this section we redefine and reprove the four moment theorem using the language of tensors. This is a more natural formulation in the infinite-dimensional setting and also highlights the role that discrete stochastic integrals play as adjoints to the higher order Malliavan derivatives. The proof is also less combinatorial than in the previous section and is of interest for that reason alone. Recall that we use the shorthand notation $[n] = \{1, \dots, n\}$, and for integers $p \geqslant 1$ we define the space of p-tensors by

$$(\mathbb{R}^n)^{\otimes p} = \{K : [n]^p \to \mathbb{R}\}.$$

We redefine I_p as an operator $I_p : (\mathbb{R}^n)^{\otimes p} \to L^2(\mathrm{P}_n)$ on p-tensors by

$$I_p(K) := \sum_{q\in[n]^p} K_q \pi_p(Z_{q_1} \dots Z_{q_p}) \tag{6.6}$$ eq:Ip_tensors

and we will show how it is consistent with definition (6.1). By definition (6.6) $I_p(K)$ is also an element of the pth Wiener chaos, but the sum has redundancy because the terms $\pi_p(Z_{q_1} \dots Z_{q_p})$ only depend on the number of times that each $i \in [n]$ appears within q. To deal with this we make the following definition.

?⟨def:counts_vector⟩? **Definition 3.1.** For $q \in [n]^p$ and $i \in [n]$, let $c_i(q) = |\{l : q_l = i\}|$ be the number of times that i appears in q, and $c(q) = (c_1(q), \dots, c_n(q)) \in \mathbb{Z}_+^n$ be the counts vector of q.

Note that $q \in [n]^p$ implies $c(q) \in \mathcal{S}_p^n$, but the mapping $c : [n]^p \to \mathcal{S}_p^n$ is not invertible. However $c(q) = c(q')$ iff q and q' are permutations of one another, i.e. there exists a permutation σ of $[p]$ such that $\sigma(q) := (q_{\sigma(1)}, \dots, q_{\sigma(p)}) = q'$. In particular, the number of elements $q' \in [n]^p$ with $c(q) = c(q')$ is given by the multinomial coefficient

$$\binom{p}{c(q)}.$$

Now in $L^2(\mathrm{P}_n)$ we have the identity

$$\pi_p(Z_{q_1} \dots Z_{q_p}) = \mathcal{H}_{c(q)}(Z),$$

therefore (6.6) reduces to

$$I_p(K) = \sum_{q\in[n]^p} K_q \mathcal{H}_{c(q)}(Z) = \sum_{k\in\mathcal{S}_p^n} \mathcal{H}_k(Z) \sum_{\substack{q\in[n]^p:\\ c(q)=k}} K_q. \tag{6.7}$$ eq:Ip_other_formulation

Hence by defining $a(K) \in \mathcal{S}_p^n$ by

$$\binom{p}{k} a(K)_k := \sum_{\substack{q\in[n]^p:\\ c(q)=k}} K_q$$

we see that $I_p(a(K)) = I_p(K)$. Conversely, an $a \in \mathcal{S}_p^n$ can be associated with several p-tensors that produce the same stochastic integral, but the most natural one is the symmetric p-tensor $K(a)$ given by

$$K(a)_q = a_{c(q)}.$$

By defining $K(a)$ in this way we see that (6.7) implies $I_p(K(a)) = I_p(a)$, thus proving the equivalence of definitions (6.6) and (6.1). This also explains why we defined (6.1) with the multinomial coefficients embedded into the summation. Equation (6.7) also shows that it is enough to restrict the definition of I_p as an operator on p-tensors to the symmetric p-tensors. The space of symmetric p-tensors is denoted by

$$(\mathbb{R}^n)^{\odot p} = \{K : [n]^p \to \mathbb{R} \text{ such that } K_q = K_{\sigma(q)} \text{ for all permutations } \sigma \text{ of } [p]\}.$$

All p-tensors $K \in (\mathbb{R}^n)^{\otimes p}$ can be naturally symmeterized into an element $\tilde{K} \in (\mathbb{R}^n)^{\odot p}$ by the map

$$\tilde{K}_q = \binom{p}{c(q)}^{-1} \sum_{q' : c(q') = c(q)} K_q,$$

and thus from (6.7) we see that $I_p(K) = I_p(\tilde{K})$ for all p-tensors K. As an operator on symmetric p-tensors a natural way to write it is

$$I_p(K) = \sum_{q \in [n]^{\odot p}} \binom{p}{c(q)} K_q \mathcal{H}_{c(q)}(Z) \tag{6.8}$$

eq:Ip_symmetric

where $[n]^{\odot p}$ is the ordered subset of $[n]^p$ given by

$$[n]^{\odot p} = \{q \in [n]^p : q_1 \leqslant q_2 \leqslant \ldots \leqslant q_p\}.$$

Equation (6.8) holds by representing all elements in $[n]^p$ by their unique ordered version, which is sufficient for symmetric p-tensors. From this representation it is also straightforward to prove the isometry formula

$$\mathrm{E}[I_p(K) I_q(L)] = \begin{cases} p!(K \cdot L), & p = q \\ 0, & p \neq q \end{cases} \tag{6.9}$$

eq:Ip_tensor_isometry

Equation (6.6) and the first equality of (6.7) partially explain why we regard $I_p(K)$ as a stochastic integral of order p. The tensor K is a function on $[n]^p$ and the terms $\mathcal{H}_{c(q)}(Z)$ are random weights on $[n]^p$ that are orthogonal in the $L^2(\mathrm{P}_n)$ space, and we can regard the sum as an integral of the function against these weights. The second equality of (6.7) is the general fact that integration of functions on product spaces is equivalent to the integration of their symmeterizations.

Furthermore, just as when we have a function of several variables on a product space we can choose not to integrate out some of the variables, we can do the same with these summations. That is, for $1 \leqslant m \leqslant p$ we can also define the operator I_{p-m} acting on $(\mathbb{R}^n)^{\odot p}$ as the summation over the last $p-m$ variables of K. What remains is a function of m variables, so that $I_{p-m} : (\mathbb{R}^n)^{\odot p} \to (\mathbb{R}^n)^{\odot m}$. For $r \in [n]^m$ the formal definition is

$$I_{p-m}(K)_r := \sum_{s \in [n]^{p-m}} K_{r \oplus s} \mathcal{H}_{c(s)}(Z), \tag{6.10}$$

eq:Ip_partial

where $r \oplus s$ is the obvious concatenation of r and s into an element of $[n]^p$. Note that $I_{p-m}(K)$ is a random element of $(\mathbb{R}^n)^{\odot m}$ and that each component is an element of the $(p-m)$th Wiener chaos, just as $I_p(K)$ is a random element of $\mathbb{R}$ that is in the pth Wiener chaos. Also note that this definition relies on K being symmetric so that the result is independent of which $p-m$ components of $[n]^p$ are summed over. If $K \in (\mathbb{R}^n)^{\otimes p}$ then $I_{p-m}(K)$ is defined only after symmeterization, i.e. by $I_{p-m}(K) := I_{p-m}(\tilde{K})$. An alternative way to write (6.10) coordinatewise is

$$I_{p-m}(K)_r = I_{p-m}(K_{r \oplus \bullet}) \tag{6.11}$$

eq:Ip_partial_coordinatew

where $K_{r\oplus\bullet}$ is the $(p-m)$-tensor $s \mapsto K_{r\oplus s}$. Alternatively, instead of writing the formula coordinatewise we can write the tensor formula

$$\begin{aligned} I_{p-m}(K) &= \sum_{r\in[n]^m} \sum_{s\in[n]^{p-m}} K_{r\oplus s}\mathcal{H}_{c(s)}\, e_{r_1} \otimes \ldots \otimes e_{r_m} \\ &= \sum_{r\in[n]^m} I_{p-m}(K_{r\oplus\bullet}) e_{r_1} \otimes \ldots \otimes e_{r_m}, \end{aligned} \tag{6.12}$$

?eq:Ip_partial_as_tensor?

The latter equality makes it clear that $I_{p-m}(K)$ is a random m-tensor. Using the symmetry of K we can reduce the inner sum to its ordered version, meaning we can also write $I_{p-m}(K)$ as

$$I_{p-m}(K) = \sum_{r\in[n]^m} \sum_{s\in[n]^{\odot(p-m)}} \binom{p-m}{c(s)} K_{r\oplus s}\mathcal{H}_{c(s)} e_{r_1} \otimes \ldots \otimes e_{r_m}. \tag{6.13}$$

eq:Ip_partial_inner_ordered

With the definitions set we now prove some useful properties of stochastic integrals.

⟨prop:Ip_properties⟩ **Proposition 3.2.** *The following hold:*

⟨prop:Ip_properties:1⟩ *(a) for $K \in (\mathbb{R}^n)^{\odot p}$ and $0 \leqslant m \leqslant p$*

$$D^m I_p(K) = \frac{p!}{(p-m)!} I_{p-m}(K),$$

⟨prop:Ip_properties:2⟩ *(b) for $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ and $v^{\otimes p}$ defined by $v_q^{\otimes p} = v_{q_1} \ldots v_{q_p} = v_1^{c_1(q)} \ldots v_n^{c_n(q)}$*

$$I_p(v^{\otimes p}) = H_p(v'Z),$$

⟨prop:Ip_properties:3⟩ *(c) for every random variable $f \in L^2(\mathrm{P}_n)$ there exists a sequence of p-tensors $K^p(f) \in (\mathbb{R}^n)^{\odot p}$ such that*

$$f = \mathrm{E}[f] + \sum_{p=1}^{\infty} \frac{1}{p!} I_p(K^p(f)),$$

⟨prop:Ip_properties:4⟩ *(d) for every $f \in \mathbb{D}^{p,2}$ and $K \in (\mathbb{R}^n)^{\otimes p}$*

$$\mathrm{E}[D^p f \cdot K] = \mathrm{E}[f I_p(K)]$$

where on the left hand side $\cdot$ refers to the Hilbert-Schmidt inner product.

⟨prop:Ip_properties:5⟩ *(e)* ***Stroock's formula:*** *if $f \in \mathbb{D}^{\infty,2}$ then*

$$f = \mathrm{E}[f] + \sum_{p=1}^{\infty} \frac{1}{p!} I_p(E[D^p f]).$$

Part (d) says that the stochastic integral $I_p(K)$ is the adjoint to the Malliavin derivative $D^p f$. This is yet another natural reason for regarding $I_p(K)$ as a multiple stochastic integral.

Proof. For part (a) we first prove the $m = 1$ case. This follows from the differentiation formulas for the Hermite polynomials, which give

$$\begin{aligned}
DI_p(K) &= \sum_{q\in[n]^p} K_q \sum_{i=1}^n c_i(q)\mathcal{H}_{c(q)-e_i}e_i \\
&= \sum_{i=1}^n \sum_{q\in[n]^{\odot p}} \binom{p}{c(q)} c_i(q) K_q \mathcal{H}_{c(q)-e_i}e_i \\
&= \sum_{i=1}^n \sum_{q\in[n]^{\odot p}} \mathbb{1}\{c_i(q)>0\}\, p \binom{p-1}{c(q)-e_i} K_q \mathcal{H}_{c(q)-e_i}e_i \\
&= p\sum_{i=1}^n \sum_{q\in[n]^{\odot(p-1)}} \binom{p-1}{c(q)} K_{i\oplus q}\mathcal{H}_{c(q)}e_i \\
&= pI_{p-1}(K).
\end{aligned}$$

The second equality uses the symmetry of K to reduce the sum to the ordered chamber $[n]^{\odot p}$. The third equality is straightforward algebra for the multinomial coefficients, while the fourth uses that every element of $[n]^{\odot p}$ that contains an i can be uniquely written as $i \oplus q$, where q is an ordered element of $[n]^{\odot(p-1)}$. The last equality uses (6.13). This completes the proof for $m = 1$, and the proof for $m > 1$ follows by induction.

For part (b) first expand out the term $(v'Z)^p$ as

$$(v'Z)^p = (v_1Z_1 + \ldots + v_nZ_n)^p = \sum_{q\in[n]^p} v_{q_1}\ldots v_{q_p} Z_{q_1}\ldots Z_{q_p}.$$

Now recall (3.13) which says that for centered Gaussians X we have $H_p(X) = \pi_p(X^p)$. Apply π_p to both sides to obtain

$$H_p(v'Z) = \sum_{q\in[n]^p} v_{q_1}\ldots v_{q_p}\pi_p(Z_{q_1}\ldots Z_{q_p}) = I_p(v^{\otimes p}).$$

Part (c) is simply a restatement of the Wiener chaos decomposition of Chapter 3, Corollary 2.2 that

$$f = \sum_{k\in\mathbb{Z}_+^n} \frac{\mathrm{E}[f\mathcal{H}_k]}{k!}\mathcal{H}_k.$$

For $p \geqslant 1$ and $q \in [n]^p$ set $K^p(f)_q = \mathrm{E}[f\mathcal{H}_{c(q)}]$. Then each $K^p(f)$ is a symmetric p-tensor and

$$\frac{1}{p!}I_p(K^p) = \frac{1}{p!}\sum_{q\in[n]^p}\mathrm{E}[f\mathcal{H}_{c(q)}]\mathcal{H}_{c(q)} = \frac{1}{p!}\sum_{k\in\mathcal{S}_p^n}\binom{p}{k}\mathrm{E}[f\mathcal{H}_k]\mathcal{H}_k = \sum_{k\in\mathcal{S}_p^n}\frac{\mathrm{E}[f\mathcal{H}_k]}{k!}\mathcal{H}_k.$$

For part (d), note that the recursion formula $H_{k+1}(x) = xH_k(x) - kH_{k-1}(x) = xH_k(x) - H_k'(x) = (x-\partial)H_k(x)$ for the Hermite polynomials implies that

$$A_{q_1}A_{q_2}\ldots A_{q_p}1 = \mathcal{H}_{c(q)},$$

where $A_i = Z_i - D_i$ for $i \in [n]$. Therefore by linearity we can rewrite $I_p(K)$ as

$$I_p(K) = \sum_{q\in[n]^p} A_{q_1}\ldots A_{q_p}K_q = A^p\cdot K,$$

where A^p is the formal p-tensor whose q-th coordinate is $A_{q_1} \dots A_{q_p}$ and $\cdot$ is the Hilbert-Schmidt inner product. Consequently, by applying the integration-by-parts formula (2.3) multiple times we obtain

$$\mathrm{E}[fI_p(K)] = \sum_{q\in[n]^p} \mathrm{E}[fA_{q_1}\dots A_{q_p}K_q] = \sum_{q\in[n]^p} \mathrm{E}[(D_{q_p}\dots D_{q_1}f)K_q] = \mathrm{E}[D^p f\cdot K].$$

Note that the last equality uses equality of mixed partials to reverse the order of the derivatives.

Stroock's formula of part (e) is now a consequence of parts (c) and (d). Define $K^p(f)_q = \mathrm{E}[f\mathcal{H}_{c(q)}]$ as in the proof of part (c). As in part (d), write this as $K^p(f)_q = \mathrm{E}[fA_{q_1}\dots A_{q_p}1] = \mathrm{E}[f(A^p1)_q]$, or more succintly $K^p(f) = \mathrm{E}[fA^p1]$. Thus integration-by-parts gives $K^p(f) = \mathrm{E}[fA^p1] = \mathrm{E}[D^pf1] = \mathrm{E}[D^pf]$, so the statement of part (c) completes the proof. □

It is worth making note of the notation used in the proof of part (d). It says that $I_p(K)$ can be rewritten as the tensor inner product,

$$I_p(K) = A^p\cdot K$$

and then the statement of part (d) can be rephrased as

$$\mathrm{E}[D^pf\cdot K] = \mathrm{E}[fA^p\cdot K],$$

This formulation very explicitly shows that I_p is an adjoint operator to D^p. The formal p-tensor A^p can also be used to represent the tensor of Hermite polynomials, namely

$$(A^p1)_q = \mathcal{H}_{c(q)} = \pi_p(Z_{q_1}\dots Z_{q_p}).$$

In tensor notation this identity can be rephrased as $A^p1 = \pi_p(Z^{\otimes p})$. Using this notation we also have the following relation.

⟨cor:Ip_anti_derivative⟩ **Corollary 3.3.** *Let K be a symmetric p-tensor and $0\leqslant m\leqslant p$. Then $A^m\cdot I_{p-m}(K) = I_p(K)$, and in particular $A^m\cdot D^mI_p(K) = p!/(p-m)!I_p(K)$.*

Proof. By definition of A^m and $I_{p-m}(K)$ we have

$$\begin{aligned} A^m\cdot I_{p-m}(K) = \sum_{q\in[n]^m} A^m_q I_{p-m}(K)_q &= \sum_{q\in[n]^m}\sum_{s\in[n]^{p-m}} K_{q\oplus s}A^m_q\mathcal{H}_{c(s)}(Z)\\ &= \sum_{q\in[n]^m}\sum_{s\in[n]^{p-m}} K_{q\oplus s}A_{q_1}\dots A_{q_m}A_{s_1}\dots A_{s_{p-m}}1\\ &= \sum_{q\in[n]^m}\sum_{s\in[n]^{p-m}} K_{q\oplus s}A^p_{q\oplus s}1\\ &= \sum_{q\in[n]^m}\sum_{s\in[n]^{p-m}} K_{q\oplus s}\mathcal{H}_{q\oplus s}(Z)\\ &= I_p(K). \end{aligned}$$

The second property follows from Proposition 3.2, part (a). □

Stroock's formula can be used to give a straightforward, calculus based proof of Proposition 2.3 for the product of stochastic integrals that does not rely on Wick's formula. Indeed, for a p-tensor K and a q-tensor L it is clear that $I_p(K)$ and $I_q(L)$ each being polynomials implies that $I_p(K)I_q(L)$ is in $\mathbb{D}^{\infty,2}$. Computing higher order derivatives of $I_p(K)I_q(L)$ is done with a Leibniz rule, although some care must be taken since it is a tensor version of the rule that is needed. We derive this alternate proof of the product formula below in Proposition 3.6, but first we review the necessary tensor operations, including the natural contraction operator $\otimes_r$ that replaces the $\star_r$ operation of Definition 2.2.

Definition 3.4. Given a p-tensor L and a q-tensor K and $0 \leqslant r \leqslant p \wedge q$, their r contraction is the $p - r + q - r = p + q - 2r$ tensor defined coordinate-wise by

$$(K \otimes_r L)_{a\oplus b} = \sum_{l\in[n]^r} K_{l\oplus a} L_{l\oplus b},$$

where $a \in [n]^{p-r}$ and $b \in [n]^{q-r}$ and $\oplus$ is the concatenation operator on integer vectors. As a tensor we can also write the contraction as

$$K \otimes_r L = \sum_{\substack{a\in[n]^{p-r}\\ b\in[n]^{q-r}}} ((K_{\bullet\oplus a}) \cdot (L_{\bullet\oplus b}))\,(e_{a_1} \otimes \ldots \otimes e_{a_{p-r}} \otimes e_{b_1} \otimes \ldots \otimes e_{b_{q-r}}),$$

and for K and L symmetric we may rewrite this equality by using the identity $(K_{\bullet\oplus a})\cdot(L_{\bullet\oplus b}) = (K_{a\oplus\bullet}) \cdot (L_{b\oplus\bullet})$.

In the case $r = 0$ we usually write $\otimes_0 = \otimes$ for shorthand. Note that $K \otimes L$ is an "outer product" of the two tensors with entries

$$(K \otimes L)_{a\oplus b} = K_a L_b$$

for $a \in [n]^p, b \in [n]^q$. Clearly then $\otimes$ is not a commutative operator, and in general $\otimes_r$ is not either. If K and L are both p-tensors then

$$K \otimes_p L = \sum_{a\in[n]^p} K_a L_a = K \cdot L,$$

where $\cdot$ is the Hilbert-Schmidt inner product. With this notation $Z^{\otimes p}$ is the random p-tensor with entires $Z^{\otimes p}_q = Z_{q_1} \ldots Z_{q_p}$, and if we apply π_p coordinate-wise we get $\pi_p(Z^{\otimes p}_q) = \mathcal{H}_{c(q)}(Z)$. This leads to the suggestive notation

$$I_p(K) = \sum_{q\in[n]^p} K_q \pi_p(Z^{\otimes p}_q) = \pi_p(K \cdot Z^{\otimes p}) = \pi_p(K \otimes_p Z^{\otimes p}).$$

Now if we think of D as a 1-tensor with entries D_i and random variables F as 0-tensors, then its gradient $DF = D\otimes F$ is a 1-tensor. More generally, if K is a random p-tensor then $DK = D \otimes K$ is the $(p+1)$-tensor whose entries are the partial derivatives of the entries of K. More generally, thinking of D^q as a q-tensor and K as a random p-tensor then $D^qK = D^q \otimes K$ is the random $(p+q)$-tensor whose entries are the mixed partial derivatives of the entries of K.

Note that even if K and L are symmetric tensors then their contraction $K \otimes_r L$ may not be. In fact, $K \otimes_r K$ may not even be symmetric. Therefore after computing the contraction one may symmeterize it, for which we use the notation

$$K \widetilde{\otimes}_r L := \widetilde{K \otimes_r L}.$$

When K and L are symmetric tensors the contraction can be defined with some extra freedom. If σ is a permutation of $[p]$ and η is a permutation of $[q]$, then by symmetry we have

$$(K \otimes_r L)_{a\oplus b} = \sum_{l\in[n]^r} K_{l\oplus a} L_{l\oplus b} = \sum_{l\in[n]^r} K_{\sigma(l\oplus a)} L_{\eta(l\oplus b)}.$$

That is to say, the symmetry of K and L allows us to arbitrarily choose which indices we sum over and which we keep fixed, and the choice can be made separately for both K and L.

From these definitions follows an isometry formula for discrete stochastic integrals.

:stoch_integral_isometry⟩ **Proposition 3.5.** *Let K be a symmetric p-tensor and L be a symmetric q-tensor and $0 \leqslant m \leqslant p$, $0 \leqslant l \leqslant q$. Then for $0 \leqslant r \leqslant p \wedge q$, $\mathrm{E}[I_{p-m}(K) \otimes_r I_{q-l}(K)]$ is a $(m+l-2r)$-tensor equal to*

$$\mathrm{E}[I_{p-m}(K) \otimes_r I_{q-l}(L)] = \begin{cases} (p-m)!(K \otimes_{p-m+r} L), & p-m = q-l \\ 0, & p-m \neq q-l \end{cases}.$$

Proof. Using formula (6.11) gives, for $a \in [n]^{m-r}$ and $b \in [n]^{l-r}$,

$$\begin{aligned}(I_{p-m}(K) \otimes_r I_{q-l}(L))_{a\oplus b} &= \sum_{s\in[n]^r} I_{p-m}(K)_{s\oplus a} I_{q-l}(L)_{s\oplus b} \\ &= \sum_{s\in[n]^r} I_{p-m}(K_{s\oplus a\oplus\bullet}) I_{q-l}(L_{s\oplus b\oplus\bullet}).\end{aligned}$$

Now using the isometry (6.9), since $p-m = q-l$ we have

$$\begin{aligned}\mathrm{E}[(I_{p-m}(K) \otimes_r I_{q-l}(L))_{a\oplus b}] &= (p-m)! \sum_{s\in[n]^r} (K_{s\oplus a\oplus\bullet}) \cdot (L_{s\oplus b\oplus\bullet}) \\ &= (p-m)!\,(K \otimes_{p-m+r} L)_{a\oplus b}.\end{aligned}$$

The last equality follows simply by unravelling the definition of the tensor product and the contraction. ☐

With these definitions in hand we now turn our attention to the following formula for the product of stochastic integrals.

e_stoch_integral_product⟩ **Proposition 3.6.** *Let K be a symmetric p-tensor and L be a symmetric q-tensor and $0 \leqslant m \leqslant p, 0 \leqslant l \leqslant q$. Then for $0 \leqslant s \leqslant p \wedge q$*

$$I_{p-m}(K) \otimes_s I_{q-l}(L) = \sum_{r=0}^{(p-m)\wedge(q-l)} r! \binom{p-m}{r} \binom{q-l}{r} I_{p+q-m-l-2r}(K \otimes_{r+s} L).$$

Note that the case $m = l = s = 0$ is an identity of random variables and is the exact analogue of the earlier Proposition 2.3. The case $m = l = s = 1$ is also an identity of random variables and will be used shortly to compute the Wiener chaos expansion of the squared norm of the gradient $DI_p(K)$. In general, Proposition 3.6 is a statement of the equality of two random $m + l - 2s$ tensors.

Proof. We first prove the statement in the case $m = l = s = 0$ and then build from there. We use Stroock's formula, part (e) of Proposition 3.2. Since $I_p(K)I_q(L)$ is a polynomial of degree $p+q$ Stroock's formula implies that

$$I_p(K)I_q(L) = \mathrm{E}[I_p(K)I_q(L)] + \sum_{k=1}^{p+q} \frac{1}{k!} I_k(E[D^k(I_p(K)I_q(L))]), \tag{6.14}$$

eqn:Stroock_for_Ip

although we will soon see that for many k these terms are zero. The higher order derivative of a product is calculated by the Leibniz rule: if $f(Z)$ and $g(Z)$ are in $\mathbb{D}^{k,2}$ then

$$D^k(fg) = \sum_{j=0}^{k} \binom{k}{j} D^j f \otimes D^{k-j} g.$$

Note that each term in the sum is a k-tensor, as one would expect. Specializing Leibniz's rule to the product of $I_p(K)$ and $I_q(L)$ and using the differentiation formula of Proposition 3.2, part (a), we obtain

$$D^k(I_p(K)I_q(L)) = \sum_{j=0}^{k\wedge p\wedge q} \binom{k}{j} \frac{p!}{(p-j)!} \frac{q!}{(q-k+j)!} I_{p-j}(K) \otimes I_{q-k+j}(L).$$

Now we take expected values. By Proposition 3.5 the only term that is not mean zero is for j such that $p - j = q - k + j$. If this happens then Proposition 3.5 implies

$$\mathrm{E}[D^k(I_p(K)I_q(L))] = \binom{k}{j} \frac{p!}{(p-j)!} \frac{q!}{(p-j)!} (p-j)!(K \otimes_{p-j} L). \tag{6.15}$$

eqn:expectation_Dk

Now $p - j = q - k + j$ is equivalent to $2j = p - q + k$, which in turn implies that $p - q$ and k must have the same parity and thus $p + q$ and k must also have the same parity. Thus there is (exactly) one term in $D^k(I_p(K)I_q(L))$ with non-zero expectation if and only if $k = p + q - 2r$ for some $0 \leqslant r \leqslant p + q$. But since also $0 \leqslant j \leqslant k$, by inserting $j = 0$ and $j = k$ into $2j = p - q + k$ we conclude $-k \leqslant p - q \leqslant k$. Inserting $k = p + q - 2r$ into both sides yields $r \leqslant p$ and $r \leqslant q$. Thus there is a non-zero expectation only when $k = p + q - 2r$ with $0 \leqslant r \leqslant p \wedge q$. We use this to rewrite the product formula (6.14) as

$$I_p(K)I_q(L) = \sum_{r=0}^{p\wedge q} \frac{1}{(p+q-2r)!} I_{p+q-2r}(E[D^{p+q-2r}(I_p(K)I_q(L))]).$$

Note that we were able to remove the expectation term because it only entered in if $p = q$ in the first place, and if $p = q$ then the $r = p \wedge q = p = q$ case handles the expectation. Now $k = p+q-2r$ and $2j = p-q+k$ imply that $j = p-r$, and inserting these expressions into (6.15) this is equivalent to

$$\begin{aligned} I_p(K)I_q(L) &= \sum_{r=0}^{p\wedge q} \frac{1}{(p+q-2r)!} \binom{p+q-2r}{p-r} \frac{p!}{r!} \frac{q!}{r!} r! I_{p+q-2r}(K \otimes_r L) \\ &= \sum_{r=0}^{p\wedge q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(K \otimes_r L). \end{aligned}$$

This completes the proof of the $m = l = s = 0$ case. The proof of the general case is now built upon this one. Indeed, since $I_{p-m}(K)$ is an m-tensor and $I_{q-l}(K)$ is an l-tensor, for $a \in [n]^{m-s}$ and $b \in [n]^{l-s}$ we have

$$\begin{aligned}(I_{p-m}(K) \oplus_s I_{q-l}(L))_{a\oplus b} &= \sum_{u\in[n]^s} I_{p-m}(K)_{u\oplus a} I_{q-l}(L)_{u\oplus b} \\ &= \sum_{u\in[n]^s} I_{p-m}(K_{u\oplus a\oplus\bullet}) I_{q-l}(L_{u\oplus b\oplus\bullet}).\end{aligned}$$

The second equality is by equation 6.11 that defines the multiple stochastic integrals coordinatewise. Now $K_{u\oplus a\oplus\bullet}$ is a $(p-m)$-tensor for each fixed u and a, while $L_{u\oplus b\oplus\bullet}$ is a $(q-l)$-tensor for each fixed u and b. Therefore the previous "$m = l = s = 0$" case of the proposition can be applied, giving

$$\begin{aligned}&I_{p-m}(K_{u\oplus a\oplus\bullet}) I_{q-l}(L_{u\oplus b\oplus\bullet}) \\ &\qquad = \sum_{r=0}^{(p-m)\wedge(q-l)} r! \binom{p-m}{r} \binom{q-l}{r} I_{p+q-m-l-2r}(K_{u\oplus a\oplus\bullet} \oplus_r L_{u\oplus b\oplus\bullet}).\end{aligned}$$

Now by linearity of the stochastic integral, in order to complete the proof it is enough to show that

$$\sum_{u\in[n]^s} K_{u\oplus a\oplus\bullet} \oplus_r L_{u\oplus b\oplus\bullet} = (K \otimes_{r+s} L)_{a\oplus b}.$$

This is a straightforward consequence of the definition of the contraction of tensors, which completes the proof. □

With the product formula in hand we can now reprove Propositions 2.4 and 2.5 using the language of tensors. Although the statements are virtually unchanged we include them for the sake of completeness. The proofs, however, are less combinatorial and more straightforward from an analytic point of view.

Proposition 3.7. *For $p \geqslant 1$ and a p-tensor K we have*

$$\operatorname{Var}\left(\frac{1}{p}||DI_p(K)||^2_{\mathbb{R}^n}\right) = \frac{1}{p^2}\sum_{r=1}^{p-1} r^2(r!)^2 \binom{p}{r}^4 (2p-2r)!||(K \otimes_r K)||^2,$$

and for the kurtosis

$$\mathrm{E}[I_p(K)^4] - 3\,\mathrm{E}[I_p(K)^2]^2 = \frac{3}{p}\sum_{r=1}^{p-1} r(r!)^2 \binom{p}{r}^4 (2p-2r)!||(K \otimes_r K)||^2.$$

Proof. For the variance formula begin with $DI_p(K) = pI_{p-1}(K)$, which is a 1-tensor. Its squared norm is therefore equal to its 1-contraction with itself, leading to the identity

$$\begin{aligned}||DI_p(K)||^2_{\mathbb{R}^n} &= p^2(I_{p-1}(K) \otimes_1 I_{p-1}(K)) \\ &= p^2 \sum_{r=0}^{p-1} r! \binom{p-1}{r}^2 I_{2p-2-2r}(K \otimes_{r+1} K) \\ &= \sum_{r=1}^{p} r(r!) \binom{p}{r}^2 I_{2p-2r}(K \otimes_r K),\end{aligned} \tag{6.16}$$

eq:DIp_expansion

with the second equality following by Proposition 3.6. The $r = p$ term is simply

$$p(p!)(K \otimes_p K) = p(p!)(K \cdot K) = p^2 \,\mathrm{E}[I_{p-1}(K) \otimes_1 I_{p-1}(K)] = \mathrm{E}\left[||DI_p(K)||^2_{\mathbb{R}^n}\right].$$

The second equality is from Proposition 3.5, while the third is from $DI_p(K) = pI_{p-1}(K)$. Thus the $r = p$ term is constant and hence irrelevant for the variance, while the remaining terms are all mean zero since they are stochastic integrals. Now since the terms for different r are uncorrelated, and by the isometry (6.9), we obtain

$$\mathrm{Var}\left(\frac{1}{p}||DI_p(K)||^2_{\mathbb{R}^n}\right) = \frac{1}{p^2}\sum_{r=1}^{p-1} r^2(r!)^2 \binom{p}{r}^4 (2p-2r)!||K \otimes_r K||^2,$$

which completes the proof for the variance formula. For the kurtosis formula use the relation $A \cdot DI_p(k) = pI_p(K)$ from Corollary 3.3 and the integration by parts formula from Chapter 2, (2.3) to compute

$$\begin{aligned}
\mathrm{E}[I_p(K)^4] = \frac{1}{p}\,\mathrm{E}[(A \cdot DI_p(K))I_p(K)^3] &= \frac{1}{p}\sum_{i=1}^{n} \mathrm{E}[D_i I_p(K)^3 D_i I_p(K)] \\
&= \frac{3}{p}\sum_{i=1}^{n} \mathrm{E}[3I_p(K)^2 (D_i I_p(K))^2] \\
&= \frac{3}{p}\,\mathrm{E}[I_p(K)^2 ||DI_p(K)||^2_{\mathbb{R}^n}].
\end{aligned}$$

Now use Proposition 3.6 to expand $I_p(K)^2$ and (6.16) to expand $||DI_p(K)||^2_{\mathbb{R}^n}$, and use the isometry (6.9) to compute that the expectation of their product is

$$\mathrm{E}[I_p(K)^4] = \frac{3}{p}\sum_{r=1}^{p} r(r!)^2 \binom{p}{r}^4 (2p-2r)!||K \otimes_r K||^2.$$

Similarly, using Proposition 3.6 to compute $\mathrm{E}[I_p(K)^2]^2$ (equivalently use (6.9)) gives

$$\mathrm{E}[I_p(K)^2]^2 = (p!)^2 (K \cdot K)^2 = (p!)^2 ||K \otimes_p K||^2.$$

Combining these two equations we obtain

$$\mathrm{E}[I_p(K)^4] - 3E[I_p(K)^2]^2 = \frac{3}{p}\sum_{r=1}^{p-1} (r!)^2 \binom{p}{r}^4 (2p-2r)!||K \otimes_r K||^2,$$

as claimed. □

From here the proofs of Theorem 1.1 and Corollary 1.2 are exactly the same, leading to the conclusion that if $p > 1$ then no element of the pth Wiener chaos is normally distributed. However, as we will see in the next section, this does not proclude that there is a sequence of elements in the pth Wiener chaos whose distribution converges to the normal one. This is in fact possible, and the ideas of the fourth moment theorem can be extended to give a sufficient condition for when it is true.

4 Convergence to Normality

5 Extensions to Correlated Gaussians

Problems

1. Show that $S^2 - 1$ is in the second Wiener chaos, where
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2.$$
Since it is already a quadratic you only need to show that it is orthogonal to all linear functions of the Z_i. Alternatively you can use Exercise 14 of Chapter 3 by observing that
$$S^2 = \frac{1}{n-1} Z'(I - n^{-1}\mathbf{1}\mathbf{1}')Z$$
where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ is the vector of all ones.
2. Prove that for arbitrary $k \in \mathbb{Z}_+^n$ and $i \in \{1, \dots, n\}$ the Wiener chaos decomposition of $\mathcal{H}_k \mathcal{H}_{e_i}$ is $\mathcal{H}_k \mathcal{H}_{e_i} = \mathcal{H}_{k+e_i} + k_i \mathcal{H}_{k-e_i}$. Check that this agrees with Proposition 2.1.
3. Derive the combinatorial proof of the multinomial identity
$$\binom{p}{k} = \sum_{i=1}^{n} \binom{p-1}{k-e_i}$$
for $k \in \mathcal{S}_p^n$. You might want to start off by recalling the combinatorial proof of the binomial identity
$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$
4. Verify that the statement of either Proposition 2.3 or Proposition 3.6 implies the statement of Proposition 2.1.
5. Use equation (6.8) to prove the isometry formula (6.9).
6. Verify that 2-tensors are just $n \times n$ matrices, and that if K is a 2-tensor then its symmeterization $\tilde{K}$ is just $\tilde{K} = (K + K')/2$, where $'$ denotes transpose. Also check that if K and L are both 2-tensors then $\otimes_1$ is just matrix multiplication with $K \otimes_1 L = KL'$.
7. To be finished still: find a sequence of $n \times n$ matrices A_k with $\operatorname{tr}(A_k) = 0$ such that $Z'A_kZ$ converges in law to a normal distribution as $k \to \infty$. Try the $n = 2$ case first and see if you can deduce a pattern.

Part II

The Infinite-Dimensional Theory

Chapter 7

Gaussian Processes

1 Basic Notions

Let T be a set, and $X := \{X_t\}_{t\in T}$ a stochastic process, defined on a suitable probability space $(\Omega, \mathcal{F}, \mathrm{P})$, that is indexed by T.

Definition 1.1. We say that X is a *Gaussian process indexed by* T when $(X_{t_1}, \dots, X_{t_n})$ is a Gaussian random vector for every $t_1, \dots, t_n \in T$ and $n \geqslant 1$. The distribution of X—that is the Borel measure $\mathbb{R}^T \ni A \mapsto \mu(A) := \mathrm{P}\{X \in A\}$—is called a *Gaussian measure*.

Lemma 1.2. *Suppose* $X := (X_1, \dots, X_n)$ *is a Gaussian random vector. If we set* $T := \{1, \dots, n\}$, *then the stochastic process* $\{X_t\}_{t\in T}$ *is a Gaussian process. Conversely, if* $\{X_t\}_{t\in T}$ *is a Gaussian process, then* $(X_1, \dots, X_n)$ *is a Gaussian random vector.*

The proof is left as exercise.

Definition 1.3. If X is a Gaussian process indexed by T, then we define $\mu(t) := \mathrm{E}(X_t)$ $[t \in T]$ and $C(s\,,t) := \mathrm{Cov}(X_s\,, X_t)$ for all $s, t \in T$. The functions μ and C are called the *mean* and *covariance* functions of X respectively.

⟨lem:pos:def⟩ **Lemma 1.4.** *A symmetric* $n \times n$ *real matrix* C *is the covariance of some Gaussian random vector if and only if* C *is positive semidefinite. The latter property means that*

$$z'Cz = \sum_{i=1}^n \sum_{j=1}^n z_i z_j C_{i,j} \geqslant 0 \qquad \text{for all } z_1, \dots, z_n \in \mathbb{R}.$$

Proof. Consult any textbook on multivariate normal distributions. □

Corollary 1.5. *A function* $C : T \times T \to \mathbb{R}$ *is the covariance function of some* T*-indexed Gaussian process if and only if* $(C(t_i\,, t_j))_{1\leqslant i,j\leqslant n}$ *is a positive semidefinite matrix for all* $t_1, \dots, t_n \in T$.

Definition 1.6. From now on we will say that a function $C : T \times T \to \mathbb{R}$ is *positive semidefinite* when $(C(t_i\,, t_j))_{1\leqslant i,j\leqslant n}$ is a positive semidefinite matrix for all $t_1, \dots, t_n \in T$.

Note that we understand the structure of every Gaussian process by looking only at finitely-many Gaussian random variables at a time. As a result, the theory of Gaussian processes does not depend *a priori* on the topological structure of the indexing set T. In this sense, the theory of Gaussian processes is quite different from Markov processes, martingales, etc. In those theories, it is essential that T is a totally-ordered set [such as $\mathbb{R}$ or $\mathbb{R}_+$], for example. Here, T can in principle be any set. Still, it can happen that X has particularly-nice structure when T is Euclidean, or more generally, has some nice group structure. We anticipate this possibility and introduce the following.

?⟨def:stationary⟩? **Definition 1.7.** Suppose T is an abelian group and $\{X_t\}_{t\in T}$ a Gaussian process indexed by T. Then we use the additive notation for T, and say that X is *stationary* when $(X_{t_1}, \ldots, X_{t_k})$ and $(X_{s+t_1}, \ldots, X_{s+t_k})$ have the same law for all $s, t_1, \ldots, t_k \in T$.

The following is a simple result but still worth recording.

⟨lem:stationary⟩ **Lemma 1.8.** *Let T be an abelian group and let $X := \{X_t\}_{t\in T}$ denote a T-indexed Gaussian process with mean function m and covariance function C. Then X is stationary if and only if m and C are "translation invariant." That means that*

$$m(s+t) = m(t) \quad \textit{and} \quad C(t_1\,, t_2) = C(s+t_1\,, s+t_2) \textit{ for all } s, t, t_1, t_2 \in T,$$

using additive notation for the group operations on T.

?⟨rem:stationary⟩? *Remark* 1.9. In other words, Lemma 1.8 says that if X is stationary, then its mean function is a constant and its covariance function is a function of the difference of its variables; i.e.,

$$m(t) = m(0) \quad \text{and} \quad C(t_1\,, t_2) = C(t_1 - t_2\,, 0) \text{ for all } s, t, t_1, t_2 \in T,$$

still using additive notation for the group operations on T.

2 Examples of Gaussian Processes

§2.1 Gaussian Random Polynomials

⟨subsec:Gauss:Poly⟩ Let $Z_0, \ldots, Z_n$ denote $q+1$ i.i.d. standard normal random variables, where $q \geqslant 0$ is an integer, and consider the *Gaussian random polynomial* $X := \{X_t\}_{t\in\mathbb{R}}$ defined by

$$X_t := Z_0 + Z_1 t + Z_2 t^2 + \cdots + Z_q t^q \qquad \text{for all } t \in \mathbb{R}.$$

Then, X is manifestly a mean-zero Gaussian process with covariance function

$$C(s\,, t) = 1 + st + s^2t^2 + \cdots + s^q t^q \qquad \text{for all } s, t \in \mathbb{R}.$$

Clearly, $t \mapsto X_t$ is a.s. C^∞ and all of its derivatives are themselves Gaussian random polynomials. For example,

$$\begin{aligned} X'_t &:= \frac{\mathrm{d}X_t}{\mathrm{d}t} = Z_1 + 2Z_2 t + 3Z_3 t^2 + \cdots + qZ_q t^{q-1}, \\ X''_t &:= \frac{\mathrm{d}^2 X_t}{\mathrm{d}t^2} = 2Z_2 + 6Z_3 t + \cdots + q(q-1)Z_q t^{q-2}, \end{aligned}$$

etc. for all $t \in \mathbb{R}$.

§2.2 Brownian Motion

By *Brownian motion* X, we mean a Gaussian process, indexed by $\mathbb{R}_+ := [0\,,\infty)$, with mean function 0 and covariance function

$$C(s\,,t) := \min(s\,,t) \qquad [s,t \geqslant 0].$$

In order to justify this definition, it suffices to prove that C is a positive semidefinite function on $T \times T = \mathbb{R}_+^2$. Suppose $z_1, \ldots, z_n \in \mathbb{R}$ and $t_1, \ldots, t_n \geqslant 0$. Then,

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n z_i z_j C(t_i\,,t_j) &= \sum_{i=1}^n \sum_{j=1}^n z_i z_j \int_0^\infty \mathbb{1}_{[0,t_i]}(s)\mathbb{1}_{[0,t_j]}(s)\,\mathrm{d}s \\ &= \int_0^\infty \left|\sum_{i=1}^n z_i \mathbb{1}_{[0,t_i]}(s)\,\mathrm{d}s\right|^2 \geqslant 0.\end{aligned}$$

Therefore, Brownian motion exists.

§2.3 The Brownian Bridge

A *Brownian bridge* is a mean-zero Gaussian process, indexed by $[0\,,1]$, and with covariance

$$C(s\,,t) = \min(s\,,t) - st \qquad [0 \leqslant s,t \leqslant 1]. \tag{7.1}$$ Cov:BB

The most elegant proof of existence, that I am aware of, is due to J. L. Doob: Let B be a Brownian motion, and define

$$X_t := B_t - tB_1 \quad [0 \leqslant t \leqslant 1].$$

Then, $X := \{X_t\}_{0\leqslant t\leqslant 1}$ is a mean-zero Gaussian process that is indexed by $[0\,,1]$ and has the covariance function of (7.1).

§2.4 The Ornstein–Uhlenbeck Process

⟨subsec:OU⟩ An *Ornstein–Uhlenbeck process* is a stationary Gaussian process X indexed by $\mathbb{R}$ with mean function 0 and covariance

$$C(s\,,t) = \mathrm{e}^{-|t-s|} \qquad [s,t \in \mathbb{R}]. \tag{7.2}$$ Cov:OU

This is our first example of a stationary Gaussian process; stationarity itself is justified by Lemma 1.8. It remains to prove that C is a positive semidefinite function. The proof rests on the following well-known formula:[1]

$$\mathrm{e}^{-|x|} = \frac{1}{\pi}\int_{-\infty}^\infty \frac{\mathrm{e}^{ixa}}{1+a^2}\,\mathrm{d}a \qquad [x \in \mathbb{R}]. \tag{7.3}$$ FT:Cauchy

Thanks to (7.3),

$$\begin{aligned}\sum_{j=1}^n \sum_{k=1}^n z_j z_k C(t_j\,,t_k) &= \frac{1}{\pi}\int_{-\infty}^\infty \frac{\mathrm{d}a}{1+a^2} \sum_{j=1}^n \sum_{k=1}^n z_j z_k \mathrm{e}^{ia(t_j-t_k)} \\ &= \frac{1}{\pi}\int_{-\infty}^\infty \frac{\mathrm{d}a}{1+a^2}\left|\sum_{j=1}^n z_j \mathrm{e}^{iat_j}\right|^2 \geqslant 0,\end{aligned}$$

for every $t_1, \ldots, t_n \in \mathbb{R}$ and $z_1, \ldots, z_n \in \mathbb{R}$.

[1] In other words, if Y has a standard Cauchy distribution on the line, then its characteristic function is $\mathrm{E}\exp(ixY) = \exp(-|x|)$.

§2.5 Brownian Sheet

An N-parameter *Brownian sheet* X is a Gaussian process, indexed by $\mathbb{R}^N_+ := [0\,,\infty)^N$, whose mean function is zero and covariance function is

$$C(\boldsymbol{s}\,,\boldsymbol{t}) = \prod_{j=1}^n \min(s^j\,,t^j) \qquad [\boldsymbol{s} := (s^1,\ldots,s^N), \boldsymbol{t} := (t^1,\ldots,t^N) \in \mathbb{R}^N_+].$$

Clearly, a 1-parameter Brownian sheet is Brownian motion; in that case, the existence problem has been addressed. In general, we may argue as follows: For all $z_1,\ldots,z_n \in \mathbb{R}$ and $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n \in \mathbb{R}^N_+$,

$$\begin{aligned}\sum_{j=1}^n\sum_{k=1}^n z_jz_k \prod_{\ell=1}^N \min(s^\ell_j\,,s^\ell_k) &= \sum_{j=1}^n\sum_{k=1}^n z_jz_k \prod_{\ell=1}^N \int_0^\infty \mathbb{1}_{[0,s^\ell_j]}(r)\mathbb{1}_{[0,s^\ell_k]}(r)\,\mathrm{d}r\\ &= \sum_{j=1}^n\sum_{k=1}^n z_jz_k \int_{\mathbb{R}^N_+} \prod_{\ell=1}^N \mathbb{1}_{[0,s^\ell_j]}(r^\ell)\mathbb{1}_{[0,s^\ell_k]}(r^\ell)\,\mathrm{d}r.\end{aligned}$$

Thus, we find that

$$\sum_{j=1}^n\sum_{k=1}^n z_jz_kC(\boldsymbol{s}_j\,,\boldsymbol{s}_k) = \int_{\mathbb{R}^N_+} \left|\sum_{j=1}^n \prod_{\ell=1}^N z_j^{1/N}\mathbb{1}_{[0,s^\ell_j]}(r^\ell)\right|^2 \mathrm{d}r \geqslant 0,$$

for all $z_1,\ldots,z_n \in \mathbb{R}$ and $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n \in \mathbb{R}^N_+$. This proves that the Brownian sheet exists.

§2.6 Fractional Brownian Motion

⟨subsec:fBm⟩ A *fractional Brownian motion* [or fBm] is a Gaussian process indexed by $\mathbb{R}_+$ that has mean function 0, $X_0 := 0$, and covariance function given by

$$\mathrm{E}(|X_t - X_s|^2) = |t-s|^{2\alpha} \qquad [s,t \geqslant 0], \tag{7.4}$$ Var:fBm

for some constant $\alpha > 0$. The constant α is called any one the *parameter*, *Hurst parameter*, *index*, or *Hurst index* of X.

Note that (7.4) indeed yields the covariance function of X: Since $\mathrm{Var}(X_t) = \mathrm{E}(|X_t - X_0|^2) = t^{2\alpha}$, it follows that

$$|t-s|^{2\alpha} = \mathrm{E}\left(X_t^2 + X_s^2 - 2X_sX_t\right) = t^{2\alpha} + s^{2\alpha} - 2\mathrm{Cov}(X_s\,,X_t).$$

Therefore,

$$\mathrm{Cov}(X_s\,,X_t) = \frac{t^{2\alpha}+s^{2\alpha}-|t-s|^{2\alpha}}{2} \qquad [s,t\geqslant 0]. \tag{7.5}$$ Cov:fBm

Direct inspection shows that (7.5) does not define a positive-definite function C when $\alpha \leqslant 0$. This is why we have limited ourselves to the case that $\alpha > 0$.

Note that an fBm with Hurst index $\alpha = 1/2$ is a Brownian motion. The reason is the following elementary identity:

$$\frac{t+s-|t-s|}{2} = \min(s\,,t) \qquad [s,t\geqslant 0],$$

which can be verified by considering the cases $s \geqslant t$ and $t \geqslant s$ separately.

The more interesting "if" portion of the following is due to Mandelbrot and Van Ness (1968).

?⟨th:fBm:exists⟩? **Theorem 2.1.** *An fBm with Hurst index α exists if and only if $\alpha \leqslant 1$.*

Fractional Brownian motion with Hurst index $\alpha = 1$ is a trivial process in the following sense: Let N be a standard normal random variable, and define $X_t := tN$. Then, $X := \{X_t\}_{t \geqslant 0}$ is fBm with index $\alpha = 1$. For this reason, many experts do not refer to the $\alpha = 1$ case as fractional Brownian motion, and reserve the teminology fBm for the case that $\alpha \in (0\,,1)$. Also, fractional Brownian motion with Hurst index $\alpha = 1/2$ is Brownian motion.

Proof. First we examine the case that $\alpha < 1$. Our goal is to prove that

$$C(s\,,t) := \frac{t^{2\alpha} + s^{2\alpha} - |t-s|^{2\alpha}}{2}$$

is a covariance function.

Consider the function

$$\Phi(t\,,r) := (t-r)_+^{\alpha-(1/2)} - (-r)_+^{\alpha-(1/2)},$$

defined for all $t \geqslant 0$ and $r \in \mathbb{R}$, where $a_+ := \max(a\,,0)$ for all $a \in \mathbb{R}$. Direct inspection yields that $\int_{-\infty}^{\infty} [\Phi(t\,,r)]^2\,\mathrm{d}r < \infty$, since $\alpha < 1$, and in fact a second computation on the side yields

$$\int_{-\infty}^{\infty} \Phi(t\,,r)\Phi(s\,,r)\,\mathrm{d}r = \kappa C(s\,,t) \qquad \text{for all } s, t \geqslant 0, \tag{7.6}$$ Phi2

where κ is a positive and finite constant that depends only on α. In particular,

$$\begin{aligned}\sum_{i=1}^n \sum_{j=2}^n z_i z_j C(t_i\,,t_j) &= \frac{1}{\kappa} \sum_{i=1}^n \sum_{j=2}^n z_i z_j \int_{-\infty}^{\infty} \Phi(t_i\,,r)\Phi(t_j\,,r)\,\mathrm{d}r \\ &= \frac{1}{\kappa} \int_{-\infty}^{\infty} \left[\sum_{i=1}^n z_i \Phi(t_i\,,r)\right]^2 \mathrm{d}r \geqslant 0.\end{aligned}$$

This proves the Theorem in the case that $\alpha < 1$. We have seen already that theorem holds [easily] when $\alpha = 1$. Therefore, we now consider $\alpha > 1$, and strive to prove that fBm does not exist in this case.

The proof hinges on a technical fact which we state without proof; a proof can be found in the next chapter and its exercises (see Problem XXX, p. XXX). Recall that $\bar{Y}$ is a *modification* of Y when $\mathrm{P}\{Y_t = \bar{Y}_t\} = 1$ for all t.

⟨pr:KCT:Gauss⟩ **Proposition 2.2.** *Let $Y := \{Y_t\}_{t \in [0,\tau]}$ denote a Gaussian process indexed by $T := [0\,,\tau]$, where $\tau > 0$ is a fixed constant. Suppose there exists a finite constant C and a constant $\eta > 0$ such that*

$$\mathrm{E}\left(|Y_t - Y_s|^2\right) \leqslant C|t-s|^\eta \qquad \textit{for all } 0 \leqslant s, t \leqslant \tau. \tag{7.7}$$?cond:KCT:Gauss?

Then Y has a Hölder-continuous modification $\bar{Y}$. Moreover, for every non-random constant $\rho \in (0\,,\eta/2)$,

$$\sup_{0 \leqslant s \neq t \leqslant \tau} \frac{|\bar{Y}_t - \bar{Y}_s|}{|t-s|^\rho} < \infty \qquad \textit{almost surely.}$$

We use Proposition 2.2 in the following way: Suppose to the contrary that there existed an fBm X with Hurst parameter $\alpha > 1$. By Proposition 2.2, X would have a continuous modification $\bar{X}$ such that for all $\rho \in (0\,,\alpha)$ and $\tau > 0$,

$$V(\tau) := \sup_{0\leqslant s\neq t\leqslant \tau} \frac{|\bar{X}_t - \bar{X}_s|}{|t-s|^\rho} < \infty \qquad \text{almost surely.}$$

Choose $\rho \in (1\,,\alpha)$ and observe that

$$\left|\bar{X}_t - \bar{X}_s\right| \leqslant V(\tau)|t-s|^\rho \qquad \text{for all } s,t \in [0\,,\tau],$$

almost surely for all $\tau > 0$. Divide both side by $|t-s|$ and let $s \to t$ in order to see that $\bar{X}$ is differentiable and its derivative is zero everywhere, a.s. Since $\bar{X}_0 = X_0 = 0$ a.s., it then follows that $\bar{X}_t = 0$ a.s. for all $t \geqslant 0$. In particular, $\mathrm{P}\{X_t = 0\} = 1$ for all $t \geqslant 0$. Since the variance of X_t is supposed to be $t^{2\alpha}$, we are led to a contradiction. $\square$

§2.7 Isonormal Processes, White Noise, and Wiener Integrals

⟨subsec:WN⟩ Let $\mathbb{H}$ be a complex Hilbert space with norm $\|\ldots\|_{\mathbb{H}}$ and corresponding inner product $\langle\cdot\,,\cdot\rangle_{\mathbb{H}}$.

Definition 2.3. The *isonormal process* indexed by $T = \mathbb{H}$ is a Gaussian process $\{\xi(h)\}_{h\in\mathbb{H}}$, indexed by $\mathbb{H}$, with mean function 0 and covariance function,

$$C(h_1\,,h_2) = \langle h_1\,,h_2\rangle_{\mathbb{H}} \qquad [h_1,h_2 \in \mathbb{H}].$$

The proof of existence is fairly elementary: For all $z_1,\ldots,z_n \in \mathbb{R}$ and $h_1,\ldots,h_n \in \mathbb{H}$,

$$\begin{aligned}\sum_{j=1}^n\sum_{k=1}^n z_jz_kC(h_j\,,h_k) &= \sum_{j=1}^n\sum_{k=1}^n z_jz_k\,\langle h_j\,,h_k\rangle_{\mathbb{H}}\\ &= \left\langle \sum_{j=1}^n z_jh_j\,,\sum_{k=1}^n z_kh_k\right\rangle_{\mathbb{H}} = \left\|\sum_{j=1}^n z_jh_j\right\|_{\mathbb{H}}^2,\end{aligned}$$

which is clearly $\geqslant 0$.

The following simple result is one of the centerpieces of this section, and plays an important role in the sequel.

⟨lem:WN:Lin⟩ **Lemma 2.4.** *For every $a_1,\ldots,a_m \in \mathbb{R}$ and $h_1,\cdots,h_m \in \mathbb{H}$,*

$$\xi\left(\sum_{j=1}^m a_jh_j\right) = \sum_{j=1}^m a_j\xi(h_j) \qquad a.s.$$

Proof. We plan to prove that: (a) For all $a \in \mathbb{R}$ and $h \in \mathbb{H}$,

$$\xi(ah) = a\xi(h) \qquad \text{a.s.;} \tag{7.8}$$ WN:Lin1

and (b) For all $h_1,h_2 \in \mathbb{H}$,

$$\xi(h_1+h_2) = \xi(h_1)+\xi(h_2) \qquad \text{a.s.} \tag{7.9}$$ WN:Lin2

Together, (7.8) and (7.9) imply the lemma with $m = 2$; the general case follows from this case, after we apply induction. Let us prove (7.8) then:

$$\begin{aligned}\mathrm{E}\left(|\xi(ah) - a\xi(h)|^2\right) &= \mathrm{E}\left(|\xi(ah)|^2\right) + a^2\,\mathrm{E}\left(|\xi(h)|^2\right) - 2a\mathrm{Cov}\left(\xi(ah)\,,\xi(h)\right)\\ &= \|ah\|_{\mathbb{H}}^2 + a^2\|h\|_{\mathbb{H}}^2 - 2a\langle ah\,,h\rangle_{\mathbb{H}} = 0.\end{aligned}$$

This proves (7.8). As regards (7.9), we note that

$$\begin{aligned}&\mathrm{E}\left(|\xi(h_1+h_2) - \xi(h_1) - \xi(h_2)|^2\right)\\ &= \mathrm{E}\left(|\xi(h_1+h_2)|^2\right) + \mathrm{E}\left(|\xi(h_1)+\xi(h_2)|^2\right) - 2\mathrm{Cov}\left(\xi(h_1+h_2)\,,\xi(h_1)+\xi(h_2)\right)\\ &= \|h_1+h_2\|_{\mathbb{H}}^2 + \|h_1\|_{\mathbb{H}}^2 + \|h_2\|_{\mathbb{H}}^2 + 2\langle h_1\,,h_2\rangle_{\mathbb{H}}\\ &\qquad\qquad - 2\left[\langle h_1+h_2\,,h_1\rangle_{\mathbb{H}} + \langle h_1+h_2\,,h_2\rangle_{\mathbb{H}}\right]\\ &= \|h_1+h_2\|_{\mathbb{H}}^2 - 2\langle h_1\,,h_2\rangle_{\mathbb{H}} - \|h_1\|_{\mathbb{H}}^2 - \|h_2\|_{\mathbb{H}}^2,\end{aligned}$$

which is zero, thanks to the Pythagorean rule on $\mathbb{H}$. This proves (7.9) an hence the lemma. □

Lemma 2.4 can be rewritten in the following essentially-equivalent form.

⟨th:Wiener⟩ **Theorem 2.5** (Wiener). *The map $\xi : \mathbb{H} \to L^2(\Omega\,,\mathcal{F},\mathrm{P}) := L^2(\mathrm{P})$ is a linear Hilbert-space isometry.*

Theorem 2.5 justifies the reason for calling ξ an *iso-normal*—sometimes also an *iso-gaussian*—process.

Very often, the Hilbert space $\mathbb{H}$ is an L^2-space itself; say, $\mathbb{H} = L^2(\mu) := L^2(A\,,\mathcal{A}\,,\mu)$. Then, we can think of $\xi(h)$ as an $L^2(\mathrm{P})$-valued integral of $h \in \mathbb{H}$. In such a case, we sometimes adopt an integral notation; namely,

$$\int h(x)\,\xi(\mathrm{d}x) := \int h\,\mathrm{d}\xi := \xi(h).$$

This operation has all but one of the properties of integrals: The triangle inequality does not hold.[2]

Definition 2.6. The isonormal process $\{\xi(h)\}_{h\in L^2(\mu)}$ is sometimes referred to as *white noise* with *intensity measure* μ. The random variable $\int h\,\mathrm{d}\xi$ is called the *Wiener integral* of $h \in \mathbb{H} = L^2(\mu)$. One also defines definite Wiener integrals as follows: For all $h \in L^2(\mu)$ and $E \in \mathcal{A}$,

$$\int_E h(x)\,\xi(\mathrm{d}x) := \int_E h\,\mathrm{d}\xi := \xi(h\mathbb{1}_E).$$

This is a rational definition since $\|h\mathbb{1}_E\|_{L^2(\mu)} \leqslant \|h\|_{L^2(\mu)} < \infty$.

An important property of white noise is that, since it is a Hilbert-space isometry, it maps orthogonal elements of $\mathbb{H}$ to orthogonal elements of $L^2(\mathrm{P})$. In other words:

$$\mathrm{E}[\xi(h_1)\xi(h_2)] = 0 \quad \text{if and only if} \quad (h_1\,,h_2)_{\mathbb{H}} = 0.$$

Because $(\xi(h_1)\,,\xi(h_2))$ is a Gaussian random vector of uncorrelated coordindates, we find that

$$\xi(h_1) \text{ and } \xi(h_2) \text{ are independent if and only if } (h_1\,,h_2)_{\mathbb{H}} = 0.$$

The following is a ready consequence of this rationale.

[2] In fact, $|\xi(h)| \geqslant 0$ a.s., whereas $\xi(|h|)$ is negative with probability $1/2$.

⟨pr:uncorr:indep⟩ **Proposition 2.7.** *If $\mathbb{H}_1, \mathbb{H}_2, \ldots$ are orthogonal subspaces of $\mathbb{H}$, then*

$$\{\xi(h)\}_{h\in\mathbb{H}_i} \qquad i = 1, 2, \ldots$$

are independent Gaussian processes.

The following highlights the strength of the preceding result.

⟨pr:KL⟩ **Proposition 2.8.** *Let $\{\psi_i\}_{i=1}^{\infty}$ be a complete orthonormal basis for $\mathbb{H}$. Then, we can find a sequence of i.i.d. standard normal random variables $X_1, X_2, \ldots$ such that*

$$\xi(h) = \sum_{j=1}^{\infty} \langle h\,, \psi_j\rangle_{\mathbb{H}}\, X_j,$$

where the sum converges in $L^2(\mathrm{P})$.

Remark 2.9. Let $\mathbb{H} := \mathbb{R}^N$ in the usual way, and $\mu :=$ the counting measure on $\{1\,, \ldots, N\}$. Then, we can think of $\mathbb{H}$ as $L^2(\mu)$. In this case, Proposition 2.8 yields a 1-1 identification of the corresponding white noise ξ with the i.i.d. sequence $\{X_i\}_{i=1}^N$. Therefore, in the setting of Proposition 2.8, some people refer to a sequence of i.i.d. standard normal random variables as white noise. See Example 2.11 below for more details.

Proof. Thanks to Proposition 2.7, $X_j := \xi(\psi_j)$ defines an i.i.d. sequence of standard normal random variables. According to the Riesz–Fischer theorem

$$h = \sum_{j=1}^{\infty} c_j\psi_j \qquad \text{for every } h \in \mathbb{H},$$

where the sum converges in $\mathbb{H}$. Therefore, Theorem 2.5 ensures that

$$\xi(h) = \sum_{j=1}^{\infty} c_j\xi(\psi_j) = \sum_{j=1}^{\infty} c_jX_j \qquad \text{for every } h \in \mathbb{H},$$

where the sum converges in $L^2(\mathrm{P})$. We have implicitly used the following ready consequence of Wiener's isometry [Theorem 2.5]: If $h_n \to h$ in $\mathbb{H}$ then $\xi(h_n) \to \xi(h)$ in $L^2(\mathrm{P})$. It might help to recall that the reason is simply that $\|\xi(h_n - h)\|_{L^2(\mathrm{P})} = \|h_n - h\|_{\mathbb{H}}$. □

Next we work out a few examples of Hilbert spaces that arise in the literature.

Example 2.10 (Zero-Dimensional Hilbert Spaces). We can identify $\mathbb{H} = \{0\}$ with a Hilbert space in a canonical way. In this case, white noise indexed by $\mathbb{H}$ is just a normal random variable with mean zero and variance 0 [i.e., $\xi(0) := 0$].

⟨ex:fin:dim:H⟩ *Example* 2.11 (Finite-Dimensional Hilbert Spaces). Choose and fix an integer $n \geqslant 1$. The space $\mathbb{H} := \mathbb{R}^n$ is a real Hilbert space with inner product $(a\,, b)_{\mathbb{H}} := \sum_{j=1}^n a_jb_j$ and norm $\|a\|_{\mathbb{H}}^2 := \sum_{j=1}^n a_j^2$. Let ξ denote white noise indexed by $\mathbb{H} = \mathbb{R}^n$ and define a random vector $X := (X_1\,, \ldots, X_n)$ via

$$X_j := \xi(\boldsymbol{e}_j) \qquad j = 1, 2, \ldots, n,$$

where $\boldsymbol{e}_1 := (1\,,0\,,\ldots,0)',\ldots,\boldsymbol{e}_n := (0\,,\ldots 0\,,1)'$ denote the usual orthonormal basis elements of $\mathbb{R}^n$. According to Proposition 2.8 and its proof, $X_1,\ldots,X_n$ are i.i.d. standard normal random variables and for every n-vector $a := (a_1\,,\ldots,a_n)$,

$$\xi(a) = \sum_{j=1}^n a_j X_j = a'X. \tag{7.10}$$ MVN

Now consider m points $a_1,\ldots,a_m \in \mathbb{R}^n$ and define

$$Y := \begin{pmatrix} \xi(a_1) \\ \vdots \\ \xi(a_m) \end{pmatrix}.$$

Then Y is a mean-zero Gaussian random vector with covariance matrix $A'A$ where A is an $m \times m$ matrix whose jth column is a_j. Then we can apply (7.10) to see that $Y = A'X$. In other words, every multivariate normal random vector with mean vector $\mathbf{0}$ and covariance matrix $A'A$ can be written as a linear combination $A'X$ of i.i.d. standard normals.

Example 2.12 (Lebesgue Spaces). Consider the usual Lebesgue space $\mathbb{H} := L^2(\mathbb{R}_+)$. Since $\mathbb{1}_{[0,t]} \in L^2(\mathbb{R}_+)$ for all $t \geqslant 0$, we can define a mean-zero Gaussian process $B := \{B_t\}_{t\geqslant 0}$ by setting

$$B_t := \xi(\mathbb{1}_{[0,t]}) = \int_0^t \mathrm{d}\xi. \tag{7.11}$$ B:xi

Then, B is a Brownian motion because

$$\mathrm{E}[B_s B_t] = \left\langle \mathbb{1}_{[0,t]}\,, \mathbb{1}_{[0,s]} \right\rangle_{L^2(\mathbb{R}_+)} = \min(s\,,t).$$

Since $\mathrm{E}(|B_t - B_s|^2) = |t-s|$, Kolmgorov's continuity theorem [Proposition 2.2] shows that B has a continuous modification $\bar{B}$. Of course, $\bar{B}$ is also a Brownian motion, but it has continuous trajectories [Wiener's Brownian motion]. Some authors intepret (7.11) somewhat loosely and present white noise as the derivative of Brownian motion. This viewpoint can be made rigorous in the following way: White noise is the weak derivative of Brownian motion, in the sense of distribution theory. We will not delve into this matter further though.

Let us close this example by mentioning, to the reader who knows Wiener and Itô's theories of stochastic integration against Brownian motion, that the Wiener integral $\int_0^\infty \varphi_s\,\mathrm{d}B_s$ of a non-random function $\varphi \in L^2(\mathbb{R}_+)$ is the same object as $\int_0^\infty \varphi\,\mathrm{d}\xi = \xi(\varphi)$ here. Indeed, it suffices to prove this assertion when $\varphi_s = \mathbb{1}_{[0,t]}(s)$ for some fixed number $t > 0$. But then the assertion is just our definition (7.11) of the Brownian motion B.

Example 2.13 (Lebesgue Spaces, Continued). Here is a fairly general receipe for constructing mean-zero Gaussian processes from white noise: Suppose we could write

$$C(s\,,t) = \int K(s\,,r)K(t\,,r)\,\mu(\mathrm{d}r) \qquad [s,t \in T],$$

where μ is a locally-finite measure on some measure space $(A\,,\mathcal{A})$, and $K : A \times T \to \mathbb{R}$ is a function such that $K(t\,,\bullet) \in L^2(\mu)$ for all $t \in T$. Then, the receipe is this: Let ξ be white noise on $\mathbb{H} := L^2(\mu)$, and define

$$X_t := \int K(t\,,r)\,\xi(\mathrm{d}r) \qquad [t \in T].$$

Then, $X := \{X_t\}_{t\in T}$ defines a mean-zero T-indexed Gaussian process with covariance function C. Here are some examples of how we can use this idea to build mean-zero Gaussian processes from white noise.

1. Let $A := \mathbb{R}_+$, $\mu :=$ Lebesgue measure, and $K(t\,,r) := \mathbb{1}_{[0,t]}(r)$. These choices lead us to the same white-noise construction of Brownian motion as the previous example.
2. Given a number $\alpha \in (0\,,1)$, let ξ be a white noise on $\mathbb{H} := L^2(\mathbb{R})$. Because of (7.6) and our general discussion, earlier in this example, we find that

$$X_t := \frac{1}{\kappa}\int_{\mathbb{R}} \left[(t-r)_+^{\alpha-(1/2)} - (-r)_+^{\alpha-(1/2)}\right] \xi(\mathrm{d}r) \qquad [t \geqslant 0]$$

defines an fBm with Hurst index α.

3. For a more interesting example, consider the covariance function of the Ornstein–Uhlenbeck process whose covariance function is, we recall,

$$C(s\,,t) = \mathrm{e}^{-|t-s|} \qquad [s,t \geqslant 0].$$

Define

$$\mu(\mathrm{d}a) := \frac{1}{\pi}\frac{\mathrm{d}a}{1+a^2} \qquad [-\infty < a < \infty].$$

According to (7.3), and thanks to symmetry,

$$\begin{aligned} C(s\,,t) &= \int \mathrm{e}^{i(t-s)r}\,\mu(\mathrm{d}r) = \int \cos(tr - sr)\,\mu(\mathrm{d}r) \\ &= \int \cos(tr)\cos(sr)\,\mu(\mathrm{d}r) + \int \sin(tr)\sin(sr)\,\mu(\mathrm{d}r). \end{aligned}$$

Now we follow our general discussion, let ξ and ξ' are two independent white noises on $L^2(\mu)$, and then define

$$X_t := \int \cos(tr)\,\xi(\mathrm{d}r) - \int \sin(tr)\,\xi'(\mathrm{d}r) \qquad [t \geqslant 0].$$

Then, $X := \{X_t\}_{t\geqslant 0}$ is an Ornstein–Uhlenbeck process.[3]

[3] One could just as easily put a plus sign in place of the minus sign here. The rationale for this particular way of writing is that if we study the "complex-valued white noise" $\zeta := \xi + i\xi'$, where ξ' is an independent copy of ξ, then $X_t = \mathrm{Re}\int \exp(itr)\,\zeta(\mathrm{d}r)$. A fully-rigorous discussion requires facts about "complex-valued" Gaussian processes, which we will not develop here.

Problems

1. Let $\{X_t\}_{t\geqslant 0}$ denote a 1-dimensional Brownian motion, and define $Y_t := tX_{1/t}$ for $t > 0$ and $Y_0 = 0$. Prove that $\{Y_t\}_{t\geqslant 0}$ and $\{-X_t\}_{t\geqslant 0}$ are both Brownian motions.
2. Let $\{X_t\}_{t\geqslant 0}$ denote a Brownian motion, and define $X^\circ_t := X_t - tX_1$ for all $t \in [0\,,1]$, so that the process X is a Brownian bridge.
   (a) Verify that X and X° are jointly Gaussian, and that $\mathrm{Cov}(X^\circ_t\,, X_1) = 0$ for all $0 \leqslant t \leqslant 1$. Conclude that X_1 and $\{X^\circ_t\}_{t\in[0,1]}$ are independent.
   (b) Consider the space $C[0\,,1]$ of all real-valued continuous functions on $[0\,,1]$, endowed with the sup norm, the corresponding topology, and associated Borel σ-algebra. For every $\varepsilon > 0$ define a probability measure P_ε on $C[0\,,1]$ by setting
   $$P_\varepsilon(A) := \mathrm{P}(X \in A \mid -\varepsilon < X_1 < \varepsilon).$$
   Prove that P_ε converges weakly to the law of X°. That is, prove that for all bounded and continuous functions $f : C[0\,,1] \to \mathbb{R}$,
   $$\lim_{\varepsilon\to 0}\int f\,\mathrm{d}P_\varepsilon = \lim_{\varepsilon\to 0}\mathrm{E}[f(X) \mid -\varepsilon < X_1 < \varepsilon] = \mathrm{E}[f(X^\circ)].$$
   (c) Use the preceding to argue informally that "Brownian bridge is Brownian motion, conditioned to be zero at time 1." This is informal since the event $\{X_t = 0\}$ is P-null.
3. Let $X_1, X_2, \ldots$ be i.i.d. random variables on $\mathbb{R}$.
   (a) Suppose in addition that $\mathrm{E}(X_1) = 0$ and $\mathrm{Var}(X_1) = 1$. Define
   $$S_n(t) := n^{-1/2}\sum_{1\leqslant j\leqslant nt} X_j \qquad \text{for every } n \in \mathbb{N} \text{ and } 0 \leqslant t \leqslant 1.$$
   Prove that for every $0 \leqslant t_1 < \cdots < t_k \leqslant 1$, fixed, the random variable $(S_n(t_1)\,, \ldots, S_n(t_k))$ converges weakly to $(B_{t_1}\,, \ldots, B_{t_k})$ as $n \to \infty$, where B is a Brownian motion.
   (b) Suppose instead that X_1 is distributed uniformly on $[0\,,1]$. Define
   $$F_n(t) := n^{-1/2}\sum_{j=1}^n \left(\mathbb{1}_{\{X_j\leqslant t\}} - t\right) \qquad \text{for every } n \in \mathbb{N} \text{ and } 0 \leqslant t \leqslant 1.$$
   Then prove that for every $0 \leqslant t_1 < \cdots < t_k \leqslant 1$, fixed, the random variable $(S_n(t_1)\,, \ldots, S_n(t_k))$ converges weakly to $(B^\circ_{t_1}\,, \ldots, B^\circ_{t_k})$ as $n \to \infty$, where B is a Brownian bridge on $[0\,,1]$.
   (c) Finally suppose only that $t \mapsto F(t) := \mathrm{P}\{X_1 \leqslant t\}$ is continuous. Define
   $$F_n(t) := n^{-1/2}\sum_{j=1}^n \left(\mathbb{1}_{\{X_j\leqslant F(t)\}} - F(t)\right) \qquad \text{for every } n \in \mathbb{N} \text{ and } t \in \mathbb{R}.$$
   Then prove that for all real numbers $\leqslant t_1 < \cdots < t_k$, fixed, the random variable $(S_n(t_1)\,, \ldots, S_n(t_k))$ converges weakly to $(B^\circ_{F(t_1)}\,, \ldots, B^\circ_{F(t_k)})$ as $n \to \infty$, where B is a Brownian bridge on $[0\,,1]$.

Throughout these problems, $X := \{X_s\}_{s\geqslant 0}$ denotes an fBm with index $\alpha \in (0\,,2)$. Recall that X is Brownian motion when $\alpha = 1$.

⟨Pbm:fbM:inc⟩ 4. Prove that for all $0 < s < t < u < v$,

$$\operatorname{Cov}(X_t - X_s\,, X_v - X_v)\begin{cases} < 0 & \text{if } 0 < \alpha < 1/2,\\ = 0 & \text{if } \alpha = 1/2,\\ > 0 & \text{if } 1/2 < \alpha < 1.\end{cases}$$

Conclude from this and Slepian's inequality (Theorem 5.12, p. 88) that if $0 < \alpha < 1/2$, then for all $\lambda > 0$,

$$\mathrm{P}\left\{X_{(j+1)/n} - X_{j/n} \geqslant \frac{\lambda}{n^\alpha} \text{ for all } j = 0, \ldots, n-1\right\} \geqslant \left(\mathrm{P}\left\{X_1 \geqslant \lambda\right\}\right)^n.$$

Formulate and establish similarly-flavored inequalities in the cases that $\alpha = 1/2$ and $1/2 < \alpha < 1$.

⟨Pbm:fBm⟩ 5. Define

$$Y_t := \exp(-ct)X_{\exp(ct/\alpha)} \qquad \text{for all } t \in \mathbb{R},$$

where $c \in \mathbb{R}$ is fixed but otherwise arbitrary. Prove that $Y := \{Y_t\}_{t\in\mathbb{R}}$ is a stationary, mean-zero Gaussian process that is indexed by $\mathbb{R}$, and compute its covariance function. Conclude from your computation that if $X := \{X_s\}_{s\geqslant 0}$ is a Brownian motion then $Y_t = \exp(-t)X_{\exp(2t)}$ $[t \in \mathbb{R}]$ defines an Ornstein–Uhlenbeck process with covariance given by (7.2).

6. Observe from Proposition 2.2 that $t \mapsto X_t$ almost surely has a continuous modification. By adopting that modification if necessary we may and will assume that the fBm X has continuous trajectories.
   (a) Prove that the random variable $(X_{t+h} - X_t)/h$ has the same distribution as $h^{\alpha-1}X_1$ for every $t \geqslant 0$ and $h > 0$.
   (b) Use the preceding to prove that

$$\mathrm{P}\left\{\limsup_{h\downarrow 0} \frac{X_{t+h} - X_t}{h} = \infty\right\} = 1 \qquad \text{for every } t \geqslant 0.$$

   That is, X is a.s. not differentiable at any given point. In fact, it can be shown that X is a.s. nowhere differentiable; see XXX.

7. Prove that fBm satisfies the following "interpolatory–extrapolatory formula": For every $s, t > 0$,

$$\mathrm{E}\left(X_{st} \mid X_t\right) = \left(\frac{s^{2\alpha} + 1 - |s-1|^{2\alpha}}{2}\right) X_t \qquad \text{a.s.}$$

This is due to Mandelbrot and van Ness XXX.

Chapter 8

Regularity Theory

In this chapter we treat two questions about Gaussian processes simultanteously. Namely, "when is a Gaussian process continuous?"; and "when is a Gaussian process bounded"? We begin by discussing a sufficient condition for questions such as continuity and boundedness. That condition is based on a very general principle about abstract stochastic processes, and involves the notion of metric entropy. Later on in the chapter, we also discuss differentiability questions.

1 Metric Entropy

Let $(T\,,d)$ be a non-empty, compact metric space and define $B(t\,,r)$ to be the closed d-ball of radius $r>0$ about $t\in T$; that is,

$$B(t\,,r) := \{s\in T:\ d(s\,,t)\leqslant r\} \qquad [t\in T,\ r>0].$$

By default, for every $\varepsilon>0$ there exists an integer $n(\varepsilon)\geqslant 1$ and points $t_1,\ldots,t_{n(\varepsilon)}\in T$ such that $T=\cup_{j=1}^{n(\varepsilon)} B(t_j\,,\varepsilon)$.

?⟨def:N_T⟩? **Definition 1.1.** We write $N_T(\varepsilon)$ for the smallest such integer $n(\varepsilon)$. The function N_T is called the *metric entropy* of $(T\,,d)$.

The function $N_T:(0\,,\infty)\to\mathbb{Z}_+$ is non-increasing, and the behavior of $N_T(\varepsilon)$ for $\varepsilon\approx 0$ quantifies the "size" of the compact set T. For example, we can see easily that T is a finite set if and only if $\lim_{\varepsilon\downarrow 0} N_T(\varepsilon)<\infty$. And if $\lim_{\varepsilon\downarrow 0} N_T(\varepsilon)<\infty$, then the rate at which $N_T(\varepsilon)$ diverges can yield information about the geometry of T. The following is one way in which this statement can be quantified.

?⟨def:dim_M⟩? **Definition 1.2.** The *Minkowski dimension* of $(T\,,d)$ is defined as

$$\dim_{\mathrm{M}}(T) := \limsup_{\varepsilon\downarrow 0}\frac{\log N_T(\varepsilon)}{\log(1/\varepsilon)}.$$

Some authors refer to $\dim_{\mathrm{M}}(T)$ as the "fractal dimension" of T. See Mandelbrot XXX, for instance.

The behavior of N_T near zero describes the "size" of the set T. In order to understand this behavior, we generally proceed in two stages: We obtain an upper bound for $N_T(\varepsilon)$ as $\varepsilon \downarrow 0$; and then separately find a lower bound for $N_T(\varepsilon)$ as $\varepsilon \downarrow 0$. Upper bounds are often not hard to find: Since $N_T(\varepsilon)$ is a minimal covering number by balls of radius ε, we merely need to find a reasonable ε-covering of T. The following elementary example highlights this idea.

⟨ex:[0,1]⟩ *Example* 1.3. Suppose that $T = [0\,,1]$ and $d(s\,,t) = |s-t|$ for every $s, t \in T$. Then,

$$B(t\,,r) = [(t-r)\vee 0\,,(t+r)\wedge 1] \qquad \text{for all } t \in T \text{ and } r > 0.$$

If $n \geqslant 1$ is an integer then the $(1/n)$-balls of the form $[j/n\,,(j+1)/n]$ (for $j = 0, \ldots, n-1$) form a cover of T. Because there are n such balls, it follows from the minimality of metric entropy that $N_T(1/n) \leqslant n$ for all $n \geqslant 1$. If $\varepsilon \in (0\,,1)$ then we can find a unique integer $n \geqslant 1$ such that $1/(n+1) \leqslant \varepsilon < 1/n$, and hence

$$N_T(\varepsilon) \leqslant N_T\left(\frac{1}{n+1}\right) \leqslant n+1 < \frac{1}{\varepsilon} + 1.$$

It follows readily from this also that $\dim_{\mathrm{M}}(T) \leqslant 1$.

To summarize what we have learned so far, we find upper bounds for N_T by finding *one* [hopefully efficient] cover of T. By contrast, it can be tricky to directly find lower bounds for N_T since we can establish $N_T(\varepsilon) \geqslant m$ only after we verify that *every* ε-cover of T is comprised of m or more balls. In order to circumvent this difficulty we may resort to a related covering notion.

⟨def:P_T⟩ **Definition 1.4.** Choose and fix some $\varepsilon > 0$. We say that $t_1, \ldots, t_m \in T$ is an *ε-packing of T* if $d(t_i\,,t_j) > \varepsilon$ whenever $1 \leqslant i \neq j \leqslant m$. Let $P_T(\varepsilon)$ denote the largest integer $m \geqslant 1$ for which there exists an ε-packing of T. The function P_T is the *[Kolmogorov] capacity* of the pseudo-metric space $(T\,,d)$.

Actually, Kolmogorov capacity and metric entropy are not very different, as the following shows.

⟨lem:N:C⟩ **Lemma 1.5.** *$N_T(\varepsilon) \leqslant P_T(\varepsilon) \leqslant N_T(\varepsilon/2)$ for every $\varepsilon > 0$.*

Proof. If $P_T(\varepsilon) = m$, then we can find $t_1, \ldots, t_m \in T$ such that: (i) $d(t_i\,,t_j) > \varepsilon$ when $i \neq j$; and (ii) $\min_{i \leqslant m} d(t_i\,,t) \leqslant \varepsilon$ for all $t \in T$ thanks to the maximality of P_T. Among other things, this shows that $t_1, \ldots, t_m$ is an ε-covering of T. Since $N_T(\varepsilon)$ is the minimum size of all ε-coverings of T, it follows that $N_T(\varepsilon) \leqslant m = P_T(\varepsilon)$.

Conversely, suppose we can find $t_1, \ldots, t_\nu \in T$ such that $\cup_{i=1}^{\nu} B_d(t_i\,,\varepsilon/2) = T$. If s_1 and s_2 are two points in T such that $d(s_1\,,s_2) > \varepsilon$, then s_1 and s_2 cannot be in the same ball $B_d(t_i\,,\varepsilon/2)$ for any $1 \leqslant i \leqslant \nu$. In particular, $P_T(\varepsilon) \leqslant \nu$. The minimum such ν is of course $N_T(\varepsilon/2)$. $\square$

Lemma 1.5 shows that in order to find a good lower bound for $N_T(\varepsilon)$, we need to find only *one* good 2ε-packing of T. Let us illustrate this idea in the context of our simple Example 1.3.

⟨ex:[0,1]:1⟩ *Example* 1.6. Suppose that $T = [0\,,1]$, $d(s\,,t) = |s-t|$ for every $s, t \in T$, and $n \geqslant 2$ is integral. Since $\{2j/n\}_{j=0}^{n/2}$ is a $(1/n)$-packing of T, it follows that $P_T(1/n) \geqslant 1 + (n/2)$ and hence $N_T(1/(2n)) \geqslant 1 + (n/2)$ by Lemma 1.5. If $\varepsilon \in (0\,,1/2)$ then we can find a unique integer $n \geqslant 1$ such that $1/(n+1) \leqslant 2\varepsilon < 1/n$. Thus,

$$N_T(\varepsilon) \geqslant N_T\left(\frac{1}{2n}\right) \geqslant 1 + \frac{n}{2} \geqslant \frac{1}{4\varepsilon} + \frac{1}{2}.$$

This and Example 1.3 together imply that $N_T(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$ at sharp rate ε^{-1}. In particular, it follows from this also that $\dim_{\mathrm{M}}(T) = 1$.

Next are a few more examples.

⟨ex:dimM:1⟩ *Example* 1.7. 1. Consider the set $T := \prod_{i=1}^n [a_i\,, b_i]$ where $a_i < b_i$ are real numbers. We endow T with the Euclidean metric, $d(s\,,t) := \|s-t\|$ for all $s, t \in T$. Then it is not hard to see that $\dim_{\rm M}(T) = n$. In other words, the Minkowski dimension of T agrees with any reasonable topological notion of dimension for T.

2. Let T denote the standard ternary Cantor set in $\mathbb{R}_+$, and endow T with the Euclidean metric, $d(s\,,t) := |t-s|$ for all $s, t \in T$. Then it is possible to verify that $\dim_{\rm M}(T) = \log 2/\log 3$, which ought to be a familiar computation to you.
3. If T is a finite set then $\dim_{\rm M}(T) = 0$. However, there are countable compact spaces that have non-zero Minkowski dimension. For instance, consider the set $T := \{1\,, {}^1\!/_2\,, {}^1\!/_3\,, {}^1\!/_4\,, \ldots\} \cup \{0\}$, endowed with the Euclidean metric, $d(s\,,t) := |t-s|$ for all $s, t \in T$. Then one can prove that $\dim_{\rm M}(T) = {}^1\!/_2$.
4. There are also many metric spaces of infinite Minkowski dimension. An example is the space T of all continuous, real-valued functions on $[0\,,1]$, endowed with the usual metric,
$$d(s\,,t) := \sup_{0\leqslant y\leqslant 1} |s(y) - t(y)| \qquad \text{for all } s, t \in T.$$
Or one can consider $T = L^p[0\,,1]$ for any $1 \leqslant p \leqslant \infty$, endowed with $d(s\,,t) := \|s-t\|_{L^p[0,1]}$ for all $s, t \in L^p[0\,,1]$.

The main result of this section is a careful version of the assertion that a T-indexed stochastic process is continuous if the index set T is "not too big," as understood, in one fashion or another, via the behavior of N_T near zero.

Now let $\{X_t\}_{t\in T}$ be a real-valued stochastic process, indexed by a set T, where $(T\,,d)$ is a metric space. Define
$$\Psi(u) := \sup_{s,t\in T} \mathrm{P}\left\{|X_t - X_s| > d(s\,,t)u\right\} \qquad [u > 0]. \tag{8.1}$$ Psi

Also, introduce the "tail" functions $\{\mathcal{T}_p\}_{p\geqslant 1}$ as follows.
$$\mathcal{T}_p(\lambda) := p\int_0^\infty u^{p-1}\left(\lambda\Psi(u) \wedge 1\right) \mathrm{d}u \qquad [\lambda > 0, p \geqslant 1]. \tag{8.2}$$ T

The goal of this section is to prove the following result about increments of general T-indexed stochastic processes. In the next section we will work out examples that highlight some of the uses of such a theorem.

⟨th:entropy⟩ **Theorem 1.8.** *If $S \subset T$ is finite, then for all $p \geqslant 1$,*
$$\mathrm{E}\left[\max_{\substack{s,t\in S:\\ d(s,t)\leqslant\delta}} |X_t - X_s|^p\right] \leqslant \left\{16\int_0^{\delta/4} \left[\mathcal{T}_p\left([N_S(r)]^2\right)\right]^{1/p} \mathrm{d}r\right\}^p,$$
for every $0 < \delta \leqslant \triangle(T)$, where $\triangle(S) := \sup_{s,t\in S} d(s\,,t)$ denotes the d-diameter of S.

There are many variations on Theorem 1.8 XXX. Though this formulation of Theorem 1.8 is particularly elegant, and might even have a new aspect, the essence of the proof can be traced back to an unpublished manuscript of Kolmogorov, with non-trivial extensions due to Preston XXX, Fernique XXX, and particularly Dudley XXX, after whom this type of theorem is sometimes named. The argument rests on the following simple *a priori* estimate.

⟨lem:entropy:1⟩ **Lemma 1.9.** *Let $\Theta \subset T \times T$ be a finite set of cardinality $|\Theta|$. Then,*

$$\mathrm{E}\left[\max_{(s,t)\in\Theta} |X_t - X_s|^p\right] \leqslant \mathcal{T}_p(|\Theta|) \cdot \max_{(s,t)\in\Theta} [d(s\,,t)]^p \qquad \textit{for all } p \geqslant 1.$$

Proof. For every $u > 0$,

$$\mathrm{P}\left\{\max_{(s,t)\in\Theta}\left|\frac{X_t - X_s}{d(s\,,t)}\right| \geqslant u\right\} \leqslant \sum_{(s,t)\in\Theta} \mathrm{P}\left\{\left|\frac{X_t - X_s}{d(s\,,t)}\right| \geqslant u\right\} \leqslant |\Theta|\Psi(u) \wedge 1.$$

Integrate $[pu^{p-1}\,\mathrm{d}u]$ to see that

$$\mathrm{E}\left(\max_{(s,t)\in\Theta}\left|\frac{X_t - X_s}{d(s\,,t)}\right|^p\right) \leqslant \mathcal{T}_p(|\Theta|).$$

This implies the lemma. □

Next we apply Lemma 1.12 to improve itself.

⟨lem:entropy⟩ **Lemma 1.10.** *If T is a finite set, then for every $p \geqslant 1$,*

$$\max_{t_0\in T} \mathrm{E}\left[\max_{t\in T} |X_t - X_{t_0}|^p\right] \leqslant \left\{8\int_0^{\Delta(T)/4} [\mathcal{T}_p(N_T(r))]^{1/p}\,\mathrm{d}r\right\}^p.$$

Remark 1.11. Choose and fix $t_0 \in T$ and reduce attention to the values of X over $B(t_0\,,\varepsilon)$ where $\varepsilon \in (0\,,\frac12\Delta(T)]$ is fixed. Then, Lemma 1.10 immediately implies the following "local" version of Theorem 1.8:

$$\max_{t_0\in T} \mathrm{E}\left[\max_{t\in B(t_0,\varepsilon)} |X_t - X_{t_0}|^p\right] \leqslant \left\{8\int_0^{\varepsilon/4} [\mathcal{T}_p(N_T(r))]^{1/p}\,\mathrm{d}r\right\}^p.$$

Proof of Lemma 1.10. Recall the capacity function P_T of $(T\,,d)$ from Definition 1.4, and define

$$\varepsilon_n := 2^{-n}\Delta(T) \quad \text{and} \quad K_n := P_T(\varepsilon_n) \qquad [n \geqslant 0].$$

One can see readily that $1 = K_0 \leqslant K_1 \leqslant K_2 \leqslant \dots$.

The definition of Kolmogorov capacity ensures that for every integer $n \geqslant 0$ we can find a finite set $T_n \subset T$ such that:

- $|T_n| = K_n$;
- $d(u\,,v) > \varepsilon_n$ for all distinct pairs of points $u, v \in T_n$;
- $\inf_{s\in T_n} d(s\,,t) \leqslant \varepsilon_n$ for all $t \in T$; and
- There exists an integer $M = M(T\,,d) \geqslant 1$ such that $T_n = T$ for all $n \geqslant M$.

For every $n \geqslant 0$ let π_n denote the projection of T onto T_n; more precisely, $\pi_n(t)$ denotes the point in T_n that is closest to t for every $t \in T$. If there are many such points then we break the ties in some arbitrary fashion. Since T_0 is a singleton we can write it as $T_0 = \{t_0\}$ and observe that $\pi_0(t) = t_0$ for all $t \in T$. Also, observe that $t_0 \in T$ can be chosen in a completely arbitrary manner, without altering any of the preceding statements.

Since $T_n = T$ for all $n \geqslant M$ it follows that $\pi_n(t) = t$ for every $n \geqslant M$. Thus, to every $t \in T$ we can associate a "chain" $\{t_i\}_{i=0}^\infty$ of points as follows: Set $t_n = \pi_M(t) = t$ for all $n \geqslant M$, and then recursively define $t_{i-1} = \pi_{i-1}(t_i)$ for all $i = M, \dots, 1$. This sequence ends with t_0—the unique element of T_0—and therefore, $X_t - X_{t_0} =$

$\sum_{i=0}^{\infty}(X_{t_{i+1}} - X_{t_i})$. Clearly, all of the summands vanish after the M-th term. In any case, it follows that

$$|X_t - X_{t_0}| \leqslant \sum_{i=0}^{\infty} \max_{u \in T_{i+1}} \left|X_u - X_{\pi_i(u)}\right|,$$

uniformly for all $t \in T$. Because the right-hand side is independent of t, we apply Lemma 1.12 and Minkowski's inequality in order to find that

$$\left\|\max_{t\in T} |X_t - X_{t_0}|\right\|_p \leqslant \sum_{i=0}^{\infty} \{\mathcal{T}_p(|T_{i+1}|)\}^{1/p}\, \varepsilon_i = \sum_{i=0}^{\infty} \{\mathcal{T}_p\left(P_T(\varepsilon_{i+1})\right)\}^{1/p}\, \varepsilon_i,$$

valid because $|T_j| = P_T(\varepsilon_j)$ and $\varepsilon_j = 2^{-j}\Delta(T)$. Since $\varepsilon_i = 4(\varepsilon_{i+1} - \varepsilon_{i+2})$ for every $i \geqslant 0$, we can then write

$$\begin{aligned}\left\|\max_{t\in T} |X_t - X_{t_0}|\right\|_p &\leqslant 4\sum_{i=0}^{\infty}\int_{\varepsilon_{i+2}}^{\varepsilon_{i+1}} \{\mathcal{T}_p\left(P_T(\varepsilon_{i+1})\right)\}^{1/p}\, \mathrm{d}r \leqslant 4\sum_{i=0}^{\infty}\int_{\varepsilon_{i+2}}^{\varepsilon_{i+1}} \{\mathcal{T}_p\left(P_T(r)\right)\}^{1/p}\, \mathrm{d}r \\ &= 4\int_0^{\Delta(T)/2} \{\mathcal{T}_p\left(P_T(r)\right)\}^{1/p}\, \mathrm{d}r \leqslant 4\int_0^{\Delta(T)/2} \{\mathcal{T}_p\left(N_T(r/2)\right)\}^{1/p}\, \mathrm{d}r;\end{aligned}$$

see Lemma 1.5. Because $t_0 \in T$ is arbitrary, this and a change of variables together yield the lemma. □

Next, we apply Lemma 1.10 to improve itself.

⟨lem:entropy:1⟩ **Lemma 1.12.** *If T is a finite set, then for every $p \geqslant 1$,*

$$\mathrm{E}\left[\max_{s,t\in T} |X_t - X_s|^p\right] \leqslant \left\{16\int_0^{\Delta(T)/4} \left[\mathcal{T}_p\left(|N_T(r)|^2\right)\right]^{1/p}\, \mathrm{d}r\right\}^p.$$

Proof. The proof hinges on "tensorization."

Define $\tilde{T} := T \times T$, and endow it with "product distance,"

$$\tilde{d}\left((s\,,t)\,,(s'\,,t')\right) := d(s\,,s') \vee d(t\,,t') \qquad \text{for every } s,t,t',t' \in T. \tag{8.3}$$ tilde:d

Every ε-ball in $\tilde{T}$ has the form $B(s\,,\varepsilon) \times B(t\,,\varepsilon)$ where $s,t \in T$. In particular, if the balls $B_1,\ldots,B_m$ form an ε-cover for $(T\,,d)$, then certainly the balls $\{B_i \times B_j\}_{i,j=1}^m$ form an ε-cover for $(\tilde{T}\,,\tilde{d})$. In this way, we can relate the metric entropy of $\tilde{T}$ to that of T as follows:

$$N_{\tilde{T}}(\varepsilon) \leqslant [N_T(\varepsilon)]^2 \qquad \text{for every } \varepsilon > 0. \tag{8.4}$$ NN2

Consider the stochastic process $\tilde{X}$, indexed by $\tilde{T}$, as follows:

$$\tilde{X}_{(s,t)} := X_t - X_s \qquad \text{for every } (s\,,t) \in \tilde{T}.$$

We may combine (8.4) and Lemma 1.10 (applied to $\tilde{X}$ in place of X) in order to see that for every $p \geqslant 1$ and $\tilde{t}_0 \in \tilde{T}$,

$$\mathrm{E}\left[\max_{(s,t)\in\tilde{T}} \left|\tilde{X}_{(s,t)} - \tilde{X}_{\tilde{t}_0}\right|^p\right] \leqslant \left\{8\int_0^{\Delta(\tilde{T})/4} \left[\tilde{\mathcal{T}}_p\left(|N_T(r)|^2\right)\right]^{1/p}\, \mathrm{d}r\right\}^p, \tag{8.5}$$ ENT

where $\Delta(\tilde{T})$ denotes the $\tilde{d}$-diameter of $\tilde{T}$, and

$$\tilde{\mathcal{T}}_p(\lambda) := p\int_0^\infty u^{p-1}\left(\lambda\tilde{\Psi}(u)\wedge 1\right)\mathrm{d}u \qquad [\lambda>0, p\geqslant 1],$$

where

$$\tilde{\Psi}(u) := \sup_{(s,t),(s',t')\in\tilde{T}} \mathrm{P}\left\{\left|\tilde{X}_{(s,t)}-\tilde{X}_{(s',t')}\right| > \tilde{d}\left((s\,,t)\,,(s',t')\right)u\right\} \qquad [u>0].$$

The definition of $\tilde{d}$ ensures that

$$\Delta(\tilde{T}) = \Delta(T). \tag{8.6}$$ ENT:1

Next we estimate $\tilde{\Psi}$ as follows:

$$\begin{aligned}
\tilde{\Psi}(u) &\leqslant \sup_{(s,t),(s',t')\in\tilde{T}} \mathrm{P}\left\{|X_s-X_{s'}|+|X_t-X_{t'}| > \tilde{d}\left((s\,,t)\,,(s',t')\right)u\right\}\\
&\leqslant \sup_{(s,t),(s',t')\in\tilde{T}} \mathrm{P}\left\{|X_s-X_{s'}| > \tfrac12\tilde{d}\left((s\,,t)\,,(s',t')\right)u\right\}\\
&\qquad\qquad + \sup_{(s,t),(s',t')\in\tilde{T}} \mathrm{P}\left\{|X_t-X_{t'}| > \tfrac12\tilde{d}\left((s\,,t)\,,(s',t')\right)u\right\};
\end{aligned}$$

this is true simply because if $a,b,c\geqslant 0$ satisfy $a+b>c$ then either $a>c/2$ or $b>c/2$, or both. Consequently, the definition of $\tilde{d}$ implies that for every $u>0$,

$$\tilde{\Psi}(u) \leqslant 2\sup_{s,s'\in T}\mathrm{P}\left\{|X_s-X_{s'}| > \tfrac12 d(s\,,s')u\right\} = 2\Psi(u/2).$$

This inequality, in turn, implies that for every $\lambda>0$ and $p\geqslant 1$,

$$\begin{aligned}
\tilde{\mathcal{T}}_p(\lambda) &\leqslant p\int_0^\infty u^{p-1}\left(2\lambda\Psi(u/2)\wedge 1\right)\mathrm{d}u \leqslant 2p\int_0^\infty u^{p-1}\left(\lambda\Psi(u/2)\wedge 1\right)\mathrm{d}u\\
&= 2^p\mathcal{T}_p(\lambda).
\end{aligned}$$

Combine this with (8.7) and (8.6) in order to find that for every $p\geqslant 1$ and $\tilde{t}_0\in\tilde{T}$,

$$\mathrm{E}\left[\max_{(s,t)\in\tilde{T}}\left|\tilde{X}_{(s,t)}-\tilde{X}_{\tilde{t}_0}\right|^p\right] \leqslant \left\{16\int_0^{\Delta(T)/4}\left[\mathcal{T}_p\left(|N_T(r)|^2\right)\right]^{1/p}\,\mathrm{d}r\right\}^p, \tag{8.7}$$ ENT

To simplify this, choose and fix some $t_0\in T$ and set $\tilde{t}_0 := (t_0\,,t_0)$. Because $\tilde{X}_{\tilde{t}_0}=0$, the lemma follows. □

We are ready to prove the main portion of Theorem 1.8; that is, when T is arbitrary.

Proof of Theorem 1.8. Lemma 1.12 and its proof (see in particular (8.4)) together imply that whenever $U\subseteq S\times S$ and $p\geqslant 1$,

$$\mathrm{E}\left[\max_{(s,t)\in U}|X_t-X_s|^p\right] \leqslant \left\{16\int_0^{\Delta(U)/4}\left[\mathcal{T}_p\left(|N_S(r)|^2\right)\right]^{1/p}\,\mathrm{d}r\right\}^p,$$

where $\Delta(U)$ denotes the diameter of U in the distance $\tilde{d}$ defined in (8.3). We may apply this fact with

$$U := \{(s\,,t)\in S\times S:\ d(s\,,t)\leqslant\delta\}.$$

Because $\Delta(U)\leqslant\delta$, Theorem 1.8 follows. □

2 Continuity Theorems

§2.1 Continuity and modifications

Among other things, Theorem 1.8 and its variations can be used to sometimes show that a stochastic process $\{X_t\}_{t\in T}$ can be constructed in a nice way, where as before $(T\,,d)$ is a metric space. The following is a slight abstraction of a notion that we encountered earlier (see Proposition 2.2 on page 119).

Definition 2.1. Let $X := \{X_t\}_{t\in T}$ and $Y := \{Y_t\}_{t\in T}$ be two stochastic processes. We say that X is a *modification*—sometimes also called a *version*—of Y if $\mathrm{P}\{X_t = Y_t\} = 1$ for all $t \in T$.

Of course, if X is a version of Y, then in turn Y is a version of X as well. What the preceding really says is that X and Y have the same *finite-dimensional distributions* in the sense that

$$\mathrm{P}\left\{X_{t_1} \in A_1\,, \ldots, X_{t_k} \in A_k\right\} = \mathrm{P}\left\{Y_{t_1} \in A_1\,, \ldots, Y_{t_k} \in A_k\right\},$$

for all Borel sets $A_1, \ldots, A_k \subset \mathbb{R}$ and all $t_1, \ldots, t_k \in T$. In particular, all computable probabilities for X are the same as their counterparts for the process Y. In this sense, if X and Y are modifications of one other, then they are "stochastically indistinguishable."

?⟨th:entropy:1⟩? **Theorem 2.2.** *Let $X := \{X_t\}_{t\in T}$ be a stochastic process and suppose there exists a real number $p \geqslant 1$ and a separable $S \subseteq T$ such that*

$$\int_0^{\triangle(S)} \left[\mathcal{T}_p\left(\left[N_S(r)\right]^2\right)\right]^{1/p} \mathrm{d}r < \infty, \tag{8.8}$$ cond:entropy:1

for some $p \geqslant 1$. Then, $\{X\}_{t\in S}$ has a continuous version $\{Y_t\}_{t\in S}$ which satisfies

$$\mathrm{E}\left(\sup_{\substack{s,t\in S:\\ d(s,t)\leqslant\delta}} |Y_t - Y_s|^p\right) \leqslant \left\{16\int_0^{\delta/4} \left[\mathcal{T}_p\left(\left[N_S(r)\right]^2\right)\right]^{1/p} \mathrm{d}r\right\}^p,$$

for every $0 < \delta \leqslant \triangle(S)$.

Proof. First of all, let us observe that (8.8) implies that S is totally bounded and hence bounded; that is, $\Delta(S) < \infty$ throughout.

Theorem 1.8 immediately implies that

$$\mathrm{E}\left(\sup_{\substack{s,t\in S':\\ d(s,t)\leqslant\delta}} |X_t - X_s|^p\right) \leqslant \left\{16\int_0^{\delta/4} \left[\mathcal{T}_p\left(\left[N_S(r)\right]^2\right)\right]^{1/p} \mathrm{d}r\right\}^p,$$

uniformly for all $0 < \delta \leqslant \triangle(S)$ and all finite sets $S' \subset S$. Thanks to (8.8) and the monotone convergence theorem, the preceding continues to hold if $S' \subseteq S$ is countable. We may apply the preceding to the case that S' is dense in S.

The preceding display, (8.8), and Fatou's lemma together imply that $\{X_t\}_{t\in S'}$ is continuous a.s. Define

$$Y_s = \liminf_{\substack{t\to s:\\ t\in S'}} X_t \qquad \text{for all } s \in S.$$

Because X is a.s. continuous on S', the stochastic process Y is a.s. continuous on S; moreover,

$$\mathrm{P}\{Y_s = X_s\} = 1 \qquad \text{for all } s \in S'. \tag{8.9}$$ PYY

We claim that X is continuous in probability on S; that is, X_t converges to X_s in probability as $t \to s$ for all $s \in S$. If so, then it follows readily from the a.s. continuity of Y and (8.9) that $\mathrm{P}\{Y_s = X_s\} = 1$ for all $s \in S$; that is, the continuous process Y is a modification of X. This would complete the proof. But it is easy to see that X is continuous in probability on S; in fact, it follows immediately from Theorem 1.8 that for every $s, t \in T$,

$$\mathrm{E}\left(|X_t - X_s|^p\right) \leqslant \left\{16 \int_0^{d(s,t)/4} \left[\mathcal{T}_p\left([N_S(r)]^2\right)\right]^{1/p} \mathrm{d}r\right\}^p,$$

which goes to zero as $s \to t$ thanks to (8.8) and the dominated convergence theorem. Consequently, $\mathrm{P}\{|X_t - X_s| > \varepsilon\} \leqslant \varepsilon^{-p}\,\mathrm{E}(|X_t - X_s|^p) \to 0$ as $s \to t$ in T for every $\varepsilon > 0$. This completes the proof. □

§2.2 Application to Gaussian Processes

Let $X := \{X_t\}_{t \in T}$ denote a mean-zero Gaussian process, indexed by an arbitrary set T. Define

$$d(s\,,t) := \sqrt{\mathrm{E}\left(|X_t - X_s|^2\right)} \qquad [s, t \in T]. \tag{8.10}$$ d

It is easy to see that $d(s\,,t) \leqslant d(s\,,u) + d(u\,,t)$ for all $s, t, u \in T$, and $d(s\,,t) = d(t\,,s)$. That is, d is a pseudo-metric on T. Let us write $s \sim t$ if $d(s\,,t) = 0$, and $[t] := \{s \in T : s \sim t\}$. Clearly, $\sim$ is an equivalence relation on T and $[t] \in T/\sim$ denotes the equivalence class of $t \in T$.

We can define $\bar{X}_{[t]}$ for all $[t] \in T/\sim$ as $\bar{X}_{[t]} := X_s$ for any and every $s \in [t]$. Then it follows that $\bar{X}$ is a mean-zero Gaussian process, indexed by $T/\sim := \{[t] : t \in T\}$, and with the same "finite-dimensional distributions" as X. In this way we can assume without loss of generality that $(T\,, d)$ is a metric space; otherwise we study $\bar{X}$ in place of X, using the same methods. This is harmless as X is continuous if and only if $\bar{X}$ is.

With the preceding in mind, we can now see that Theorem 1.8 and its consequences imply sufficient conditions for X to have a continuous modification. In order to identify the details, we first develop two estimates on functions Ψ and $\mathcal{T}_1$ defined respectively in (8.1) and (8.2).

⟨lem:Psi:Gauss⟩ **Lemma 2.3.** *$\Psi(u) < \exp(-u^2/2)$ for all $u > 0$.*

Proof. The usual proof of this sort of fact yields twice the stated upper bound. The argument for this slight improvement is even easier, and borrowed from Khoshnevisan, XXX. Observe that for all $u \in \mathbb{R}$,

$$\begin{aligned}
\mathrm{P}\{Z_1 > u\} &= \frac{1}{\sqrt{2\pi}} \int_u^\infty \mathrm{e}^{-x^2/2}\,\mathrm{d}x = \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \int_u^\infty \exp\left(-\frac{(x-u)(x+u)}{2}\right) \mathrm{d}x \\
&= \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{y(y+2u)}{2}\right) \mathrm{d}y.
\end{aligned}$$

If $u > 0$, then $y + 2u > y$ whence follows $\mathrm{P}\{Z_1 > u\} < \frac{1}{2}\exp(-u^2/2)$. This proves the lemma because $\Psi(u) = \mathrm{P}\{|Z_1| > u\} = 2\mathrm{P}\{Z_1 > u\}$. □

Once we have an estimate for Ψ, we can bound $\mathcal{T}_p$ for all $p \geqslant 1$. The following does that in the case that $p = 1$.

⟨lem:T_1:Gauss⟩ **Lemma 2.4.** $\mathcal{T}_1(\lambda) \leqslant 4\sqrt{\log(\lambda \vee 2)}$ *for all* $\lambda > 0$.

Proof. If $0 < \lambda \leqslant 2$, then by Lemma 2.3,

$$\mathcal{T}_1(\lambda) \leqslant \int_0^\infty \exp(-u^2/2)\,\mathrm{d}u = \sqrt{\pi/2} < 4\sqrt{\log(\lambda \vee 2)}.$$

Else if $\lambda > 2$, then we write

$$\begin{aligned}\mathcal{T}_1(\lambda) &\leqslant \sqrt{2\log\lambda} + \lambda\int_{\sqrt{2\log\lambda}}^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = \sqrt{2\log\lambda} + \sqrt{2\pi}\,\lambda\mathrm{P}\left\{U > \sqrt{2\log\lambda}\right\}\\ &< \sqrt{2\log\lambda} + \sqrt{\frac{\pi}{2}} < \left(\sqrt{2}+\sqrt{3}\right)\sqrt{\log(\lambda\vee 2)} < 4\sqrt{\log(\lambda \vee 2)},\end{aligned}$$

using the fact that $\mathrm{P}\{Z_1 > u\} \leqslant \frac{1}{2}\exp(-u^2/2)$ for all $u > 0$ [see proof of Lemma 2.3]. This has the desired consequence. □

We can now appeal to the previous lemma and Theorem 1.8 in order to deduce the following, which is essentially due to Dudley XXX.

⟨th:Dudley⟩ **Theorem 2.5** (Dudley, XXX)**.** *If* $(T\,, d)$ *is separable and*

$$\int_0^{\triangle(T)/4} \sqrt{\log N_T(r)}\,\mathrm{d}r < \infty, \tag{8.11}$$ cond:cty

then X *has a continuous modification* Y*, and for all* $\delta > 0$*,*

$$\mathrm{E}\left(\sup_{\substack{s,t\in T:\\ d(s,t)\leqslant\delta}} |Y_t - Y_s|\right) \leqslant 64\sqrt{2}\int_0^{\delta/4}\sqrt{\log N_T(r)}\,\mathrm{d}r.$$

Remark 2.6. (8.11) holds iff $\int_0^\infty \sqrt{\log N_T(r)}\,\mathrm{d}r < \infty$; see Problem 1 below.

Proof. Theorem 1.8 and Lemma 2.4 together imply that under condition (8.11), the process X has a continuous modification Y such that

$$\mathrm{E}\left(\sup_{\substack{s,t\in T:\\ d(s,t)\leqslant\delta}} |Y_t - Y_s|\right) \leqslant 64\sqrt{2}\int_0^{\delta/4}\sqrt{\log(N_T(r)\vee 2)}\,\mathrm{d}r,$$

for all $0 < \delta \leqslant \Delta(T)$. Define r_0 to be the largest number $r > 0$ such that $N_T(r) = 1$. If there is no such r then $r_0 := 0$. Since $\log N_T(r) = 0$ whenever $r > r_0$ we can inspect the metric entropy integral according to whether or not $\theta < r_0$ in order to see that

$$\int_0^\theta \sqrt{\log N_T(r)}\,\mathrm{d}r = \int_0^\theta \sqrt{\log(N_T(r\vee 2)}\,\mathrm{d}r,$$

for all $\theta > 0$. This completes the proof. □

Let us also mention the following.

⟨co:Dudley⟩ **Corollary 2.7.** *For all denumerable sets $S \subset T$,*

$$\mathrm{E}\left[\max_{t\in S} X_t\right] \leqslant 64\sqrt{2}\int_0^\infty \sqrt{\log N_S(r)}\,\mathrm{d}r.$$

Thus, $\sup_{t\in S}|X_t|$ is finite [a.s.] if and only if $\mathrm{E}[\sup_{t\in S} X_t] < \infty$.

Proof. Without loss of generality, we may and will assume that $S = T$ is finite. Choose and fix an arbitrary $t_0 \in T$ and use the fact that $\mathrm{E}(X_{t_0}) = 0$ to write $\mathrm{E}[\max_{t\in T} X_t] = \mathrm{E}[\max_{t\in T}(X_t - X_{t_0})]$. Therefore, Dudley's Theorem (Theorem 2.5) implies that

$$\mathrm{E}\left[\max_{t\in S} X_t\right] \leqslant 64\sqrt{2}\int_0^\infty \sqrt{\log N_S(r)}\,\mathrm{d}r,$$

which is more than enough to yield the inequality of the corollary. Similar considerations (see Problem XXX) prove that

$$\mathrm{E}\left[\max_{t\in S}|X_t|^2\right] \leqslant \text{const}\cdot\left\{\int_0^\infty \sqrt{\log N_S(r)}\,\mathrm{d}r\right\}^2, \tag{8.12}$$ Dud2

where the implied constant is "universal"; that is, it does not depend on any of the parameters of the problem. We will use this fact momentarily.

For the remainder, let $S \subset T$ be a countable set [if T is finite, then there is nothing to prove]. By the Borell, Sudakov–Tsirelson inequality [Theorem 2.1, page 71],

$$\mathrm{P}\left\{\left|\max_{t\in U}|X_t| - \mathrm{E}\left[\max_{t\in U}|X_t|\right]\right| > z\right\} \leqslant 2\exp\left(-\frac{z^2}{2\sup_{t\in S}\mathrm{Var}(X_t)}\right) \qquad [z > 0],$$

for every finite $U \subset S$. Because $\sup_{t\in S}\mathrm{Var}(X_t) < \infty$ by (8.12), the preceding inequality implies that $\sup_{t\in S}|X_t| < \infty$ a.s. iff $\mathrm{E}[\sup_{t\in S}|X_t|] < \infty$.

Finally, because X and $-X$ have the same law,

$$\mathrm{E}\left[\sup_{t\in S} X_t\right] \leqslant \mathrm{E}\left[\sup_{t\in S}|X_t|\right] \leqslant 2\,\mathrm{E}\left[\sup_{t\in S} X_t\right].$$

Therefore, the corollary follows. □

We conclude this section by inspecting a classical condition (see for example XXX) for the continuity of a "stationary Gaussian process."

Example 2.8. Suppose $T = [0\,,1]$ and $\{X_t\}_{0\leqslant t\leqslant 1}$ is a stationary Gaussian process with $\mathrm{E}[X_t] = 0$ and $\mathrm{E}[X_tX_s] = \varrho(|t-s|)$ for a symmetric, strictly decreasing and continuous function $\varrho : \mathbb{R}_+ \to \mathbb{R}$ such that $\varrho(0) = 1$.[1] Because

$$d(s\,,t) = \sqrt{2\,(1 - \varrho(|s-t|))} \qquad [0 \leqslant s, t \leqslant 1],$$

it follows that: (a) $\triangle(T) \leqslant \sqrt{2}$; (b) $(T\,,d)$ is compact; and (c) Every ball in $(T\,,d)$ is also a Euclidean ball. In fact,

$$B(t\,,r) = \left\{s \geqslant 0 :\, |s-t| \leqslant \varrho^{-1}\left(1 - \frac{r^2}{2}\right)\right\} \qquad [0 \leqslant t \leqslant 1,\ 0 < r \leqslant \sqrt{2}].$$

[1]The linear Ornstein–Uhlenbeck process [p. 117] is an example of such a process with $\varrho(z) = \exp(-z)$.

From this, and Dudley's theorem, we can conclude that

$$N_{[0,1]}(\varepsilon) = \frac{(1+o(1))}{\varrho^{-1}\left(1 - \frac{1}{2}\varepsilon^2\right)} \qquad \text{as } \varepsilon \downarrow 0.$$

This leads to the following sufficient condition for the continuity of the process X,

$$\int_{0+} \sqrt{\log\left(\frac{1}{\varrho^{-1}\left(1-\frac{1}{2}r^2\right)}\right)}\,\mathrm{d}r < \infty.$$

As it turns out, this is also a necessary condition for both continuity and boundedness; see Theorem 3.5 on page 140 below.

§2.3 An Infinite-Dimensional Example

Among other things, Dudley's theorem has found many applications in the theory of empirical processes and its connections to machine learning, etc. The following example is the sort that arises naturally in empirical-process theory (see, Dudley XXX for instance).

Let $\mathbb{H} := L^2[0\,,1]$ and consider white noise $X := \{X(h)\}_{h\in\mathbb{H}}$. Recall that X is a mean-zero Gaussian process with

$$\mathrm{Cov}[X(f)\,, X(g)] = \int_0^1 f(x)g(x)\,\mathrm{d}x \qquad [f, g \in \mathbb{H}].$$

It is easy to see that the random function X is unbounded on $\mathbb{H}$. For instance, choose and fix an arbitrary orthonormal family $\{\phi_k\}_{k=1}^\infty$ in $\mathbb{H}$—such as $\phi_k(x) = \sin(2\pi kx)$ for all $k \geqslant 0$—and note that $X(\phi_1), X(\phi_2), \ldots$ are i.i.d. $\mathrm{N}(0\,,1)$ random variables, and hence are unbounded. In fact, we have seen already that

$$\limsup_{k\to\infty} \frac{X(\phi_k)}{\sqrt{2\log k}} = 1 \qquad \text{a.s;}$$

see (5.2) on page 71. Still, there are many infinite-dimensional subsets $T \subset \mathbb{H}$ for which $\{X(f)\}_{f\in T}$ defines a bounded random function. The following furnishes one such example.

Proposition 2.9. *Let T denote the collection of all Lipschitz-continuous functions $f : [0\,,1] \to \mathbb{R}$ such that $f(0) = 0$ and $\mathrm{Lip}(f) \leqslant 1$. Then, $\{X(f)\}_{f\in T}$ has a bounded version.*

Proof. Since T is not a closed subspace of $\mathbb{H}$, it is helpful to instead metrize T via $\varrho(f\,,g) := \sup_{0\leqslant x\leqslant 1} |f(x) - g(x)|$ for $f, g \in T$. We will prove that $\{X(f)\}_{f\in T}$ has a version Y that is continuous when T is metrized by ϱ rather than the courser metric

$$d(f\,,g) := \sqrt{\mathrm{E}(|X(f) - X(g)|^2)} = \|f - g\|_{\mathbb{H}}$$

of $\mathbb{H}$. This enterprise will immediately yield the a.s.-boundedness of $f \mapsto Y(f)$, for example, as well as the measurability of $\sup_{f\in T} Y(f)$, $\sup_{f\in T}|Y(f)|$, etc.

The Arzela–Ascoli theorem XXX ensures that $(T\,,\varrho)$ is compact; i.e., T is closed and $N_T(\varepsilon) < \infty$ for all $\varepsilon > 0$. We wish to understand the behavior of $N_T(\varepsilon)$ near $\varepsilon = 0$.

For every integer $n \geqslant 0$ define T_n to be the collection of all continuous, piecewise linear, functions $g : [0\,,1] \to \mathbb{R}$ such that:

1. g is linear on $[j/n\,,(j+1)/n]$ for every $0 \leqslant j \leqslant n-1$;
2. $|g((j+1)/n) - g(j/n)| = 1/n$.

It is easy to see that:

(i) For every $f \in T$ there exists $g \in T_n$ such that $\varrho(f\,,g) \leqslant n^{-1}$; and

(ii) $|T_n| \leqslant 2^n$.

Thus, $N_T(1/n) \leqslant 2^n$, whence $\log N_T(1/n) \leqslant n \log 2$. For all $\varepsilon \in (0\,,1)$ we can find an integer $n \geqslant 1$ such that $(n+1)^{-1} < \varepsilon \leqslant n^{-1}$. For this choice of n, we find that $\log N_T(\varepsilon) \leqslant \log N_T(1/(n+1)) \leqslant (n+1)\log 2 \leqslant 2 + 2\varepsilon^{-1}$. In particular,

$$\sqrt{\log(N_T(\varepsilon) \vee \mathrm{e})} \leqslant \frac{2}{\sqrt{\varepsilon}} \vee 1 \qquad [0 < \varepsilon < 1].$$

The preceding defines an integrable function of $\varepsilon \in (0\,,1)$. Since $\triangle(T\,,\varrho) = 1$ and $[d(f\,,g)]^2 = \mathrm{E}(|X(f) - X(g)|^2) = \int_0^1 |f(x) - g(x)|^2\,\mathrm{d}x \leqslant [\varrho(f\,,g)]^2$ for all $f, g \in T$, it follows from Theorem 2.5 that X has a continuous version Y which satisfies

$$\mathrm{E}\left(\sup_{\substack{f,g\in T\\ \varrho(f,g)\leqslant\delta}} |Y(f) - Y(g)|\right) \leqslant \text{const} \cdot \int_0^{\delta/4} \left(\frac{2}{\sqrt{\varepsilon}} \vee 1\right) \mathrm{d}\varepsilon \leqslant \text{const} \cdot \sqrt{\delta},$$

for all $\delta \in (0\,,1)$. We have seen already that T is compact. Therefore, the continuity of Y implies its uniform boundedness on T. □

3 Lower Bounds

We continue using the notation of the preceding subsection. In particular, $X := \{X_t\}_{t\in T}$ denotes a mean-zero Gaussian process with canonical distance $d(s\,,t) := \sqrt{\mathrm{E}(|X_t - X_s|^2)}$, and we assume that $(T\,,d)$ is a compact metric space. In this section we discuss some useful lower bounds for $\mathrm{E}[\max_{t\in T} X_t]$, for example when T is countable.

§3.1 Sudakov Minorization

Recall that $Z_1, \ldots, Z_n$ are i.i.d. with a $\mathrm{N}(0\,,1)$ distribution, and define

$$\mu(n) := \mathrm{E}\left[\max_{1\leqslant i\leqslant n} Z_i\right] \qquad [n \geqslant 1].$$

⟨lem:M⟩ **Lemma 3.1.** *There exist positive and finite constants K, L such that*

$$K\sqrt{\log n} \leqslant \mu(n) \leqslant L\sqrt{\log n} \qquad \textit{for all } n \geqslant 1.$$

Proof. The Lemma holds trivially when $n = 1$; we concentrate on $n \geqslant 2$. Since $\mu(1) = 0$, $\mu(n) = (1 + o(1))\sqrt{2\log n}$ as $n \to \infty$ [Proposition 1.3, page 7], and μ is increasing, it suffices to prove that $\mu(2) > 0$. But $\mu(2)$ is equal to $\mathrm{E}[\max(Z_1\,,Z_2)] = \mathrm{E}[\max(Z_1\,,Z_2) - Z_1] = \mathrm{E}[\max(0\,,Z_2 - Z_1)]$. It follows easily from this that $\mu(2) > 0$.[2] □

[2] In fact, because $Z_2 - Z_1$ has a $\mathrm{N}(0\,,2)$ distribution, $Z_2 - Z_1$ is independent of $\mathrm{sign}(Z_2 - Z_1)$, and hence $\mu(2) = \frac{1}{2}\,\mathrm{E}(|Z_2 - Z_2|) = \pi^{-1/2}$.

⟨rem:M⟩ *Remark* 3.2. The numerical values of K and L are not very good. For instance, the best-possible choice for K is

$$\inf_{n\geqslant 2}\frac{\mu(n)}{\sqrt{\log n}} \leqslant \frac{\mu(2)}{\sqrt{\log 2}} = \frac{1}{\sqrt{\pi\log 2}} < 0.7,$$

which is smaller than the limiting value, $\lim_{n\to\infty}\mu(n)/\sqrt{\log n} = \sqrt{2}$.

Lemma 3.1 will now be used in order to establish a useful lower bound for $\mathrm{E}[\sup_{t\in T} X_t]$.

⟨pr:Sudakov⟩ **Proposition 3.3** (Sudakov, XXX)**.** *Choose and fix some $\varepsilon > 0$, and let A be a subset of T with the property that $d(s\,,t) \geqslant \varepsilon$ whenever $s,t\in A$. Then,*

$$\mathrm{E}\left[\max_{t\in A} X_t\right] \geqslant \varepsilon\mu(|A|) \geqslant K\varepsilon\sqrt{\log(|A|)},$$

where K is the constant of Lemma 3.1 and $|\cdots|$ denotes cardinality.

Proof. In the case that $\mathrm{Var}(X_1) = \cdots = \mathrm{Var}(X_n)$, this theorem is just a restatement of Example 5.14 [page 89]. The general case is handled the same way, but requires an appeal to Fernique's inequality (Theorem 5.16, page 90) instead of Slepian's (Theorem 5.12, page 88). We will work out the details once more in order to gel the underlying ideas.

Without loss of generality, we can—and will—assume that $|A| \geqslant 2$; else the statement of the theorem is the tautology that $0 \geqslant 0$.

Define, for all $t \in A$,

$$Y_t := \varepsilon Z_t + \xi,$$

where ξ and the Z_t's are all independent $\mathrm{N}(0\,,1)$ random variable. Clearly, $(Y_t)_{t\in A}$ has a mean-zero multivariate normal distribution, and

$$\mathrm{E}\left(|Y_t - Y_s|^2\right) = \varepsilon^2 \geqslant [d(s\,,t)]^2 = \mathrm{E}\left(|X_s - X_t|^2\right).$$

Therefore, Fernique's inequality [Theorem 5.16, page 90] yields

$$\mathrm{E}\left[\max_{t\in A} X_t\right] \geqslant \mathrm{E}\left[\max_{t\in A} Y_t\right] = \varepsilon\,\mathrm{E}\left[\max_{t\in A} Z_t\right] = \varepsilon\mu(|A|),$$

by the definition of μ. □

We can summarize the results of this section as follows.

⟨pr:Sudakov:1⟩ **Proposition 3.4** (Sudakov Minorization)**.** *We always have*

$$\begin{aligned}\sup_{\substack{S\subset T:\\ S\ \text{finite}}} \mathrm{E}\left[\max_{t\in S} X_t\right] &\geqslant \sup_{0<\varepsilon<\triangle(T)} \varepsilon\mu\left(P_T(\varepsilon)\right) \geqslant \sup_{0<\varepsilon<\triangle(T)} \varepsilon\mu\left(N_T(\varepsilon)\right)\\ &\geqslant K \sup_{0<\varepsilon<\triangle(T)} \varepsilon\sqrt{\log N_T(\varepsilon)} \geqslant K\limsup_{\varepsilon\to 0}\varepsilon\sqrt{\log N_T(\varepsilon)},\end{aligned}$$

where K is the constant of Lemma 3.1.

§3.2 Fernique's Theorem

Sudakov minorization [Proposition 3.4] tells us that if

$$\limsup_{\varepsilon\to 0} \varepsilon\sqrt{\log N_T(\varepsilon)} = \infty, \tag{8.13}$$ Sudakov:minor

then X does not have continuous trajectories, whereas Dudley's theorem [Theorem 2.5] implies that the metric entropy condition

$$\int_0^\infty \sqrt{\log N_T(r)}\,\mathrm{d}r < \infty \tag{8.14}$$ Dudley:major

is sufficient for the continuity of X.

Sudakov's condition (8.13) and Dudley's condition (8.14) are related. Indeed,

$$\begin{aligned}\int_0^\infty \sqrt{\log N_T(r)}\,\mathrm{d}r &= \int_0^{\Delta(T)} \sqrt{\log N_T(r)}\,\mathrm{d}r \qquad \text{(see Problem 1)}\\ &= \sum_{n=0}^\infty \int_{2^{-n-1}\Delta(T)}^{2^{-n}\Delta(T)} \sqrt{\log N_T(r)}\,\mathrm{d}r\\ &\geqslant \frac{\Delta(T)}{2}\sum_{n=0}^\infty 2^{-n}\sqrt{\log N_T\left(2^{-n}\Delta(T)\right)}.\end{aligned}$$

Therefore, (8.14) fails to hold whenever (8.13) holds. But the converse is not always true; see, for example, XXX. As it turns out, this situation can arise because neither condition is always sharp. The sharp condition is a "majorizing-measure condition," which you can find in Talagrand's landmark paper XXX; see also XXX. We will not discuss majorizing measures in this course primarily because it is exceedingly difficult to compute them in concrete settings. In addition, the Sudakov and Dudley theorems have broad utility XXX that extend beyond the particular applications that we have in mind for these lectures.

Still, it would be a pity to say nothing about the beautiful general theory. As a compromise, we will state and prove Fernique's theorem which states that, for stationary Gaussian processes, the Dudley condition is necessary as well as sufficient. Recall that if $(T\,,d)$ is a compact abelian group, then we say that $X := \{X_t\}_{t\in T}$ is *stationary* if $d(s\,,t) = \sqrt{\mathrm{E}(|X_t - X_s|^2)}$ is a function of $s-t$, equivalently, $t-s$, where we are using the additive notation for the group T in order to be be concise.

⟨th:Fernique:1⟩ **Theorem 3.5** (Fernique, XXX)**.** *Let $X := \{X_t\}_{t\in T}$ be a stationary, mean-zero Gaussian process, where $(T\,,d)$ is a metric abelian group; in particular, $d(s\,,t) = d(s-t\,,0)$ for all $s,t\in T$. Then, for all denumerable sets $S\subset T$,*

$$\mathrm{E}\left[\max_{s\in S} X_s\right] \geqslant \frac{K^2}{16}\int_0^\infty \sqrt{\log N_S(\varepsilon)}\,\mathrm{d}\varepsilon,$$

where K is the constant of Lemma 3.1.

⟨rem:cont⟩ *Remark* 3.6. One can write the final assertion of Theorem 3.5 by choosing S to be a dense subset of a compact set T in order to see that

$$\sup_{\substack{S\subset T\\ S \text{ denumerable}}} \mathrm{E}\left[\max_{s\in S} X_s\right] \geqslant \frac{K^2}{16}\int_0^\infty \sqrt{\log N_T(\varepsilon)}\,\mathrm{d}\varepsilon.$$

It follows from this observation and Theorems 2.5 and 3.5 that a stationary, mean-zero Gaussian process $X := \{X_t\}_{t\in T}$ is continuous iff it is bounded iff $\int_0^\infty \sqrt{\log N_T(\varepsilon)}\,\mathrm{d}\varepsilon < \infty$. And, barring measure-theoretic details, we have

$$\frac{K^2}{16}\int_0^\infty \sqrt{\log N_T(r)}\,\mathrm{d}r \leqslant \mathrm{E}\left[\sup_{t\in T} X_t\right] \leqslant 64\sqrt{2}\int_0^\infty \sqrt{\log N_T(r)}\,\mathrm{d}r.$$

One can prove Fernique's theorem, fairly readily, using the following improvement of Sudakov minorization [Proposition 3.4], due to Talagrand XXX.

⟨pr:Talagrand⟩ **Proposition 3.7** (Talagrand, XXX). *Let $X := \{X_t\}_{t\in T}$ be any mean-zero Gaussian process, indexed by a general compact $(T\,,d)$, and consider $A \subset T$, a non-empty ε-packing of T for some $\varepsilon > 0$. Then,*

$$\begin{aligned}\mathrm{E}\left[\max_{t\in A} X_t\right] &\geqslant \tfrac{1}{2}\varepsilon\mu(|A|) + \min_{s\in A}\mathrm{E}\left(\max_{t\in B(s,K\varepsilon/8)} X_t\right)\\ &\geqslant \frac{K}{2}\varepsilon\sqrt{\log|A|} + \min_{s\in A}\mathrm{E}\left(\max_{t\in B(s,K\varepsilon/8)} X_t\right),\end{aligned}$$

where K was defined in Lemma 3.1.

Proof. We will present essentially the original proof of Talagrand XXX, since it is straightforward. Marcus and Rosen XXX have devised a clever argument which yields slightly better constants, but their argument is more involved.

By considering $\{X_t\}_{t\in A}$, it suffices to consider the case that $T = A$ and $d(s\,,t) \geqslant \varepsilon$ for all $s,t \in T$. If $|T| = 1$, then the proposition states that $0 \geqslant 0$. Therefore, we may consider only the case that $|T| \geqslant 2$.

Define, for all $t \in T$ and $r > 0$,

$$Y_t(r) := \max_{s\in B(t,\,r)}(X_s - X_t) = \max_{s\in B(t,\,r)} X_s - X_t.$$

Since $\max_{s\in B(t,\,r)}\mathrm{E}(|X_s - X_t|^2) \leqslant r^2$, the Borel, Sudakov–Tsirelson inequality (Theorem 2.1, page 71) implies that

$$\max_{t\in T}\mathrm{P}\left\{|Y_t(r) - \mathrm{E}[Y_t(r)]| \geqslant \lambda\right\} \leqslant 2\exp\left(-\frac{\lambda^2}{2r^2}\right),$$

for all $\lambda > 0$. In particular,

$$\mathrm{P}\left\{\max_{t\in T}|Y_t(r) - \mathrm{E}[Y_t(r)]| \geqslant \lambda\right\} \leqslant 2|T|\exp\left(-\frac{\lambda^2}{2r^2}\right)\wedge 1,$$

and hence,

$$\begin{aligned}\mathrm{E}\left(\max_{t\in T}|Y_t(r) - \mathrm{E}[Y_t(r)]|\right) &\leqslant \int_0^\infty\left[2|T|\exp\left(-\frac{\lambda^2}{2r^2}\right)\wedge 1\right]\mathrm{d}\lambda\\ &= r\sqrt{2\log(2|T|)} + 2|T|\int_{r\sqrt{2\log(2|T|)}}^\infty \exp\left(-\frac{\lambda^2}{2r^2}\right)\mathrm{d}\lambda\\ &= r\sqrt{2\log(2|T|)} + 2r|T|\sqrt{2\pi}\,\mathrm{P}\left\{U > \sqrt{2\log(2|T|)}\right\},\end{aligned}$$

where U has a $\mathrm{N}(0\,,1)$ distribution. Define

$$V := \max_{t\in T}|Y_t(r) - \mathrm{E}[Y_t(r)]|\,.$$

Since $\mathrm{P}\{U > u\} \leqslant \frac{1}{2}\exp(-u^2/2)$ [see proof of Lemma 2.3], the preceding and the triangle inequality together yield

$$\mathrm{E}(V) \leqslant r\sqrt{2\log(2|T|)} + r\sqrt{\frac{\pi}{2}} \leqslant 4r\sqrt{\log|T|},$$

thanks to the assumption that $|T| \geqslant 2$. Since $Y_t(r) \geqslant \mathrm{E}[Y_t(r)] - V$, the definition of $Y_t(r)$ yields

$$\max_{s\in B(t,r)} X_s \geqslant X_t + \mathrm{E}\left[\max_{s\in B(t,r)} X_s\right] - V \quad \text{a.s. for all } t\in T \text{ and } r>0.$$

Maximize over all $t \in T$ and take expectations to find that

$$\begin{aligned}\mathrm{E}\left[\max_{t\in T}\max_{s\in B(t,r)} X_s\right] &\geqslant \mathrm{E}\left[\max_{t\in T} X_t\right] + \min_{t\in T}\mathrm{E}\left[\max_{s\in B(t,r)} X_s\right] - 4r\sqrt{\log|T|}\\ &\geqslant \varepsilon\mu(|T|) + \min_{t\in T}\mathrm{E}\left[\max_{s\in B(t,r)} X_s\right] - \frac{4r}{K}\mu(|T|).\end{aligned}$$

We have appealed to Sudakov's inequality [Proposition 3.3] to bound $\mathrm{E}[\max_{t\in T} X_t]$ and Lemma 3.1 to bound $4r\sqrt{\log|T|}$. Set $r := K\varepsilon/8$ to deduce the lemma. □

⟨lem:Fernique:comb⟩ **Lemma 3.8.** *Suppose (T, d) is a compact abelian group, and $d(s,t) = d(s-t,0)$ for all $s,t \in T$. Then, $N_T(\alpha) \leqslant N_T(\beta)\cdot N_{B(t_0,\beta)}(\alpha)$ for all $0<\alpha<\beta$ such that $B(t_0,\beta)\subset T$.*

Proof. We have made this sort of calculation earlier in (8.4) on page 131. We adapt those ideas to the present setting next.

Observe that, regardless of the respective values of $\beta > \alpha > 0$, the stationarity of X implies that $N_{B(t,\beta)}(\alpha)$ does not depend on $t \in T$. Let $K := N_T(\beta)$ so that we can find $t_1,\ldots,t_K \in T$ such that the balls $B(t_i,\beta)$ cover T. We can cover every ball $B(t_i,\beta)$ by $L := N_{B(t_1,\beta)}(\alpha)$-many balls of radius α. Therefore, we can certainly cover T with KL-many balls of radius α. By the minimality property of the covering number $N_T(\alpha)$, this implies that $N_T(\alpha) \leqslant KL$, which is the lemma. □

Proof of Theorem 3.5. Throughout the proof, we may [and will] assume without loss of generality that $S = T$ is a countable set.

Let K denote the constant of Lemma 3.1 and recall that $K < 1$ [Remark 3.2], and define

$$R := \frac{8}{K} > 1.$$

Choose and fix some $t_0 \in T$ and observe that $T = B(t_0, R^m)$, where $m \in \mathbb{Z}$ is the unique integer defined so that

$$R^m \geqslant \triangle(T) > R^{m-1}.$$

With this observation in mind, we let $\varepsilon := R^{m-1}$ and deduce from Talagrand's inequality [Proposition 3.7] that

$$\mathrm{E}\left[\max_{t\in B(t_0,R^m)} X_t\right] \geqslant \frac{K}{2}R^{m-1}\sqrt{\log P_T\left(R^{m-1}\right)} + \min_{s\in T}\mathrm{E}\left[\max_{t\in B(s,R^{m-1})} X_s\right].$$

By stationarity, $\mathrm{E}[\max_{B(s,\varepsilon)} X]$ does not depend on $s \in T$. Therefore, Lemma 1.5 implies that

$$\begin{aligned}\mathrm{E}\left[\max_{t\in B(t_0,R^m)} X_t\right] &\geqslant \frac{K}{2}R^{m-1}\sqrt{\log N_T\left(R^{m-1}\right)} + \mathrm{E}\left[\max_{t\in B(t_0,R^{m-1})} X_s\right]\\ &= \frac{K}{2}R^{m-1}\left[\sqrt{\log N_T\left(R^{m-1}\right)} - \sqrt{\log N_T\left(R^{m}\right)}\right] + \mathrm{E}\left[\max_{t\in B(t_0,R^{m-1})} X_s\right],\end{aligned}$$

since $N_T(R^m) = 1$. We are set up nicely to carry out an induction argument: Appeal to Talagrand's inequality [Proposition 3.7] once again—this time with $\varepsilon := R^{m-2}$—in order to see that

$$\begin{aligned}\mathrm{E}\left[\max_{t\in B(t_0,R^{m-1})} X_s\right] &\geqslant \frac{K}{2}R^{m-2}\sqrt{\log N_{B(t_0,R^{m-1})}\left(R^{m-2}\right)} + \mathrm{E}\left[\max_{t\in B(t_0,R^{m-2})} X_s\right]\\ &\geqslant \frac{K}{2}R^{m-2}\sqrt{\log N_T\left(R^{m-2}\right) - \log N_T\left(R^{m-1}\right)} + \mathrm{E}\left[\max_{t\in B(t_0,R^{m-2})} X_s\right],\end{aligned}$$

by Lemma 3.8. The concavity of $f(x) := \sqrt{x}$ implies that $\sqrt{\log a - \log b} \geqslant \sqrt{\log a} - \sqrt{\log b}$ for all $a > b \geqslant 1$. Define

$$J_k := \mathrm{E}\left[\max_{s\in B(t_0,\, R^k)} X_s\right] \qquad [k \in \mathbb{Z}].$$

We apply the preceding, inductively, in order to find that

$$\begin{aligned}&\mathrm{E}\left[\max_{s\in T} X_s\right] = J_m\\ &\geqslant \frac{K}{2}\sum_{j=0}^{1} R^{m-j-1}\left[\sqrt{\log N_T\left(R^{m-j-1}\right)} - \sqrt{\log N_T\left(R^{m-j}\right)}\right] + J_{m-1}\\ &\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\ &\geqslant \frac{K}{2}\sum_{j=0}^{L} R^{m-j-1}\left[\sqrt{\log N_T\left(R^{m-j-1}\right)} - \sqrt{\log N_T\left(R^{m-j}\right)}\right] + J_{m-L-1},\end{aligned}$$

for every integer $L \geqslant 1$. Because T is a finite set, we can choose the integer $L \geqslant 1$ large enough to ensure that $B(t_0\,, R^{m-L-1}) = \{t_0\}$. Since $J_{m-L-1} = 0$ and $\log N_T(R^{m-j}) = 0$ for all $j > L$, we can deduce the following:

$$\begin{aligned}\mathrm{E}\left[\max_{s\in T} X_s\right] &\geqslant \frac{K}{2}\sum_{j=0}^{\infty} R^{m-j-1}\left[\sqrt{\log N_T\left(R^{m-j-1}\right)} - \sqrt{\log N_T\left(R^{m-j}\right)}\right]\\ &\geqslant \frac{K}{2}\sum_{j=0}^{\infty}\left[R^{m-j-1} - R^{m+j+2}\right]\sqrt{\log N_T\left(R^{m-j-1}\right)}\\ &= \frac{K}{2R}\sum_{j=0}^{\infty}\int_{R^{m-j-1}}^{R^{m-j}}\sqrt{\log N_T\left(R^{m-j-1}\right)}\,\mathrm{d}\varepsilon\\ &\geqslant \frac{K}{2R}\int_0^{R^m}\sqrt{\log N_T(\varepsilon)}\,\mathrm{d}\varepsilon = \frac{K}{2R}\int_0^{\infty}\sqrt{\log N_T(\varepsilon)}\,\mathrm{d}\varepsilon,\end{aligned}$$

since $R^m \geqslant \triangle(T)$; see Problem 1. Plug in the above the numerical value of $R = 8/K$ to finish. $\square$

4 Stationary Processes and Differentiability

Let $X := \{X_t\}_{t\in T}$ be a Gaussian process. When T has a nice manifold structure, one can go beyond asking questions about continuity to those about smoothness. The situation is particularly simple to understand when $T = \mathbb{R}$ and X is stationary, in which case we may assume that

$$\mathrm{E}(X_0) = 0 \quad \text{and} \quad \mathrm{Var}(X_0) = 1, \tag{8.15}$$ std:norm

without incurring any loss in generality.[3]

§4.1 A Necessary Condition for Differentiability

?⟨subsec:diff⟩? Because continuity is a sufficient condition for differentiability, we assume throughout that X has continuous trajectories a.s. Of course, as we have seen in Theorems 2.5 and 3.5 (pp. 135 and 140, respectively), a necessary and sufficient condition for continuity is the finiteness of the metric entropy integral for X; see also Remark 3.6. But one can do much better. Indeed, if X were continuously differentiable a.s., then

$$X'_t = \lim_{s\to t} \frac{X_t - X_s}{t-s}$$

would exist and be finite a.s. for every $t \in \mathbb{R}$. Since limits of Gaussians are Gaussian (XXX), it would follow readily that $X' := \{X'_t\}_{t\in\mathbb{R}}$ must be a stationary Gaussian process. In particular, for every $z \in \mathbb{R}$,

$$\begin{aligned}\mathrm{E}\exp\left(izX'_0\right) &= \lim_{t\to 0} \mathrm{E}\exp\left(iz\,\frac{X_t - X_0}{t}\right)\\ &= \lim_{t\to 0}\exp\left(-\frac{z^2}{2}\,\mathrm{Var}\left[\frac{X_t-X_0}{t}\right]\right).\end{aligned} \tag{8.16}$$ Eexp(izX')

Let us write

$$\rho(t) := \mathrm{E}(X_tX_0) \qquad \text{for all } t\in\mathbb{R}.$$

According to Lemma 1.8,

$$\mathrm{E}(X_sX_t) = \rho(t-s) = \rho(s-t) \qquad \text{for all } s,t\in\mathbb{R}.$$

Thus, ρ is symmetric, and we might think of ρ as the "covariance function" of X. In particular, ρ is also "positive definite" in the sense that

$$\sum_{j=1}^n\sum_{k=1}^n z_jz_k\rho(t_j-t_k) \geqslant 0 \qquad \text{for all } z_1,\ldots,z_n,t_1,\ldots,t_n\in\mathbb{R}.$$

See Lemma 1.4.

Since $\mathrm{Var}(X_t - X_0) = 2[1-\rho(t)]$, whence it follows from (8.16) that

$$\mathrm{E}\exp\left(izX'_0\right) = \exp\left(-z^2\lim_{t\to 0}\left[\frac{1-\rho(t)}{t^2}\right]\right) \qquad \text{for every } z\in\mathbb{R}, \tag{8.17}$$ Eexp(izX'):1

[3]This is because, when $\mathrm{Var}(X_0) > 0$, it follows that $t \mapsto \{X_t - \mathrm{E}(X_0)\}/\sqrt{\mathrm{Var}(X_0)}$ is differentiable if and only if $t \mapsto X_t$ is differentiable. And if $\mathrm{Var}(X_t) = 0$ for one hence all $t\in\mathbb{R}$, then $X_t = \mathrm{E}(X_t) = 0$ for all $t\in\mathbb{R}$, and differentiability questions are trivial.

a part of the assertion of this computation being that the latter limit exists. If that limit were infinity, then the left-hand side would be zero for every $z \neq 0$. This cannot be, since the left-hand side is 1 when $z = 0$ and is a continuous function of z by the dominated convergence theorem. Thus, it follows that if X were continuously differentiable a.s., then certainly we would have the following:

$$\lim_{t\to 0} \frac{1-\rho(t)}{t^2} \quad \text{exists and is finite.} \tag{8.18}$$ cond:rho''

By (8.15), $|\rho(h)| \leqslant 1$ and so the limit is also nonnegative. Also, it follows from (8.17) and the Gaussian nature of X' that

$$\mathrm{Var}(X_0') = \lim_{t\to 0} \mathrm{Var}\left(\frac{X_t - X_0}{t}\right) = 2 \lim_{t\to 0} \frac{1-\rho(t)}{t^2}. \tag{8.19}$$ Var(X'):1

⟨pr:rho:lip⟩ **Proposition 4.1.** *Condition* (8.18) *implies that ρ is uniformly Lipschitz continuous.*

Proof. Because $\rho(t) - \rho(s) = \mathrm{E}[X_0(X_t - X_s)]$, we may apply stationarity, (8.15), and the Cauchy–Schwarz inequality in order to see that

$$|\rho(t) - \rho(s)| \leqslant \sqrt{\mathrm{E}\left(|X_t - X_s|^2\right)} = \sqrt{\mathrm{E}\left(|X_{|t-s|} - X_0|^2\right)} = \sqrt{2[1-\rho(t-s)]},$$

for all $s, t \in \mathbb{R}$. Therefore,

$$\sup_{\substack{s,t\in\mathbb{R}\\ 0<|t-s|<\delta}} \frac{|\rho(t)-\rho(s)|}{|t-s|} \leqslant \sqrt{2 \sup_{0<h<\delta} \frac{1-\rho(h)}{h^2}}$$

is finite for all $\delta > 0$ sufficiently small. The preceding is also manifestly bounded above by $2\delta^{-1}$, whence it follows that ρ is uniformly Lipschitz continuous. □

We will soon see that in fact (8.18) implies that $\rho \in C^2(\mathbb{R})$. To this end, let us first recall **Herglotz's theorem:** *Every continuous positive-definite function on $\mathbb{R}$ is the Fourier transform of a finite Borel measure.* See XXX. Because $\rho(0) = 1$ and ρ is real valued, it follows from Herglotz's theorem and Proposition 4.1 that there exists a Borel probability measure ν on $\mathbb{R}$ such that

$$\rho(t) = \int_{-\infty}^{\infty} \cos(tx)\,\nu(\mathrm{d}x) \qquad \text{for all } t \in \mathbb{R}. \tag{8.20}$$ spectral:rep

The measure ν is called the *spectral measure* of X.

We can revisit (8.19) and observe that we can write $\mathrm{Var}(X_t')$ in terms of the spectral measure as follows:

$$\mathrm{Var}(X_0') = 2 \lim_{t\to 0} \int_{-\infty}^{\infty} \left(\frac{1-\cos(tx)}{t^2}\right) \nu(\mathrm{d}x). \tag{8.21}$$ Var(X'):2

Since the integrand is $\geqslant 0$, Fatou's lemma implies that $\mathrm{Var}(X_0') \geqslant \int_{-\infty}^{\infty} x^2\,\nu(\mathrm{d}x)$, and hence

$$\int_{-\infty}^{\infty} x^2\,\nu(\mathrm{d}x) < \infty.$$

Therefore, the spectral representation (8.20) of ρ, and the dominated convergence theorem, together imply that $\rho \in C^2(\mathbb{R})$, and

$$\rho'(t) = -\int_{-\infty}^{\infty} x \sin(tx)\,\nu(\mathrm{d}x) \quad \text{and} \quad \rho''(t) = -\int_{-\infty}^{\infty} x^2 \cos(tx)\,\nu(\mathrm{d}x), \tag{8.22}$$ rho'rho''

for all $t \in \mathbb{R}$. Because $0 \leqslant 1 - \cos\theta \leqslant \theta^2/2$ for all $\theta \in \mathbb{R}$, and since ν has two finite moments, the dominated convergence theorem implies that the limit and the integral can be exchanged in (8.21), which leads us to the following:

$$\mathrm{Var}(X_0') = 2\int_{-\infty}^{\infty} x^2\,\nu(\mathrm{d}x) = -\rho''(0). \tag{8.23}$$ Var(X')

Because X' is a stationary Gaussian process, in order to characterize the law of X', it remains to find $\mathrm{Cov}(X_0'\,, X_s')$ for every $s \in \mathbb{R}$ [Lemma 1.8]. One adapts the argument that led to (8.17) in order to see that

$$\begin{aligned}\mathrm{Cov}(X_0'\,, X_s') &= \lim_{t\to 0}\mathrm{Cov}\left(\frac{X_t - X_0}{t}\,, \frac{X_{s+t} - X_s}{t}\right)\\ &= -\lim_{t\to 0}\frac{\rho(s+t) + \rho(s-t) - 2\rho(s)}{t^2}\\ &= -\rho''(s),\end{aligned} \tag{8.24}$$ Cov(X0Xs)

after yet another application of Taylor's expansion. In particular, if X is continuously differentiable, then X' is necessarily a stationary, mean-zero Gaussian process with the above covariance, and hence (by Fernique's theorem, Theorem 3.5; see also Remark 3.6), has a finite metric entropy integral $\int_0^\infty \sqrt{\log N_T(\varepsilon)}\,\mathrm{d}\varepsilon$ for every closed and bounded interval $T \subset \mathbb{R}$, where the metric entropy N_T is computed with respect to the metric,

$$d(s\,,t) := \sqrt{\mathrm{E}\left(|X_t' - X_s'|^2\right)} = \sqrt{-2\left[\rho''(0) - \rho''(t-s)\right]}, \tag{8.25}$$ X':d:stationary

defined for all $s, t \in \mathbb{R}$. Since X' is stationary, it is in fact enough to consider the preceding metric entropy integral for $T = [0\,,1]$. Let us recap the efforts of this section as the following theorem.

⟨th:X':1⟩ **Theorem 4.2** (Cambanis XXX, Doob XXX)**.** *If X is an a.s. continuously-differentiable stationary Gaussian process that satisfies* (8.15)*, then $\rho \in C^2(\mathbb{R})$ and X' is a mean-zero, stationary Gaussian process with covariance given by* (8.24)*. Moreover, the metric entropy integral $\int_0^\infty \sqrt{\log N_{[0,1]}(\varepsilon)}\,\mathrm{d}\varepsilon$ is finite for the distance given by* (8.25)*.*

In the next section we will prove that Theorem 4.2 is in fact sharp.

§4.2 A Sufficient Condition for Differentiability

In this section we establish the following converse to Theorem 4.2.

⟨th:X':2⟩ **Theorem 4.3** (Cambanis XXX, Doob XXX)**.** *Suppose X is a stationary Gaussian process that satisfies* (8.15)*. Suppose also that $\rho \in C^2(\mathbb{R})$ and $\int_0^\infty \sqrt{\log N_{[0,1]}(\varepsilon)}\,\mathrm{d}\varepsilon < \infty$, where $N_{[0,1]}$ denotes the metric entropy of $[0\,,1]$ with respect to the metric d of* (8.25)*. Then, X is a.s. continuously differentiable.*

Together, Theorems 4.2 and 4.3 then imply the following:

⟨co:X'⟩ **Corollary 4.4** (Cambanis XXX, Doob XXX)**.** *Suppose X is a stationary Gaussian process that satisfies* (8.15)*. Then,*

$$\mathrm{P}\{X \in C^1(\mathbb{R})\} = 1$$

if and only if $\rho \in C^2(\mathbb{R})$ and $\int_0^\infty \sqrt{\log N_{[0,1]}(\varepsilon)}\,\mathrm{d}\varepsilon < \infty$, where $N_{[0,1]}$ denotes the metric entropy of $[0\,,1]$ with respect to the metric d of (8.25)*.*

We conclude this chapter by proving Theorem 4.3.

Proof of Theorem 4.3. First of all, let us compute directly to find that for all $t \in \mathbb{R}$ and $\varepsilon, \delta > 0$,

$$\begin{aligned}
&\mathrm{E}\left(\left|\frac{X_{t+\varepsilon} - X_t}{\varepsilon} - \frac{X_{t+\delta} - X_t}{\delta}\right|^2\right) \\
&= \mathrm{Var}\left(\frac{X_\varepsilon - X_0}{\varepsilon}\right) + \mathrm{Var}\left(\frac{X_\delta - X_0}{\delta}\right) - 2\,\mathrm{Cov}\left(\frac{X_\varepsilon - X_0}{\varepsilon}, \frac{X_\delta - X_0}{\delta}\right) \\
&= 2\frac{1-\rho(\varepsilon)}{\varepsilon^2} + 2\frac{1-\rho(\delta)}{\delta^2} - 2\frac{\rho(\varepsilon-\delta) - \rho(\varepsilon) - \rho(\delta) + 1}{\varepsilon\delta}.
\end{aligned}$$

A Taylor expansion reveals that $\rho(h) = 1 + \frac{1}{2}h^2\rho''(0) + o(1)$ as $h \to 0$, which forces the preceding quantities to tend to zero as $\varepsilon, \delta \to 0$. In other words, the condition $\rho \in C^2(\mathbb{R})$ implies that $h \mapsto h^{-1}(X_{t+h} - X_t)$ is Cauchy in $L^2(\mathrm{P})$ for every $t \in \mathbb{R}$. Define

$$X'_t := \lim_{h\to 0} \frac{X_{t+h} - X_t}{h} \qquad \text{for every } t \in \mathbb{R},$$

where the limit takes place in $L^2(\mathrm{P})$. Because there are null sets involved, the preceding is not enough to justify the differentiability of X, but it comes close. Next we close the remaining technical gaps in order to show that X' is indeed the derivative of X a.s.

Since limits of Gaussians are themselves Gaussian, it follows that $X' := \{X'_t\}_{t\in\mathbb{R}}$ is a mean-zero, stationary Gaussian process, and its covariance is described by the formula (8.24), as in the previous section.

The finiteness of the metric entropy integral $\int_0^{\Delta([0,1])} \sqrt{\log N_{[0,1]}(\varepsilon)}\,\mathrm{d}\varepsilon$ shows that $\{X'_t\}_{t\in[0,1]}$ is continuous a.s. [up to a modification]. By stationarity, this proves that X' is continuous. Therefore, Fubini's theorem shows that for all nonrandom functions $\varphi \in C^1_c(\mathbb{R})$,

$$\left\| \int_{-\infty}^{\infty} \varphi(t) \left[\frac{X_{t+h} - X_t}{h} - X'_t\right] \mathrm{d}t \right\|_{L^2(\mathrm{P})} \leqslant \int_{-\infty}^{\infty} |\varphi(t)| \left\| \frac{X_{t+h} - X_t}{h} - X'_t \right\|_{L^2(\mathrm{P})} \mathrm{d}t.$$

For every $h, \delta > 0$, $\{h^{-1}(X_{t+h} - X_t) - \delta^{-1}(X_{t+\delta} - X_t)\}_{t\in\mathbb{R}}$ is a mean-zero stationary Gaussian process. Let $\delta \to 0$ to see that so is $\{h^{-1}(X_{t+h} - X_t) - X'_t\}_{t\in\mathbb{R}}$. Consequently,

$$\begin{aligned}
&\limsup_{h\to 0} \left\| \int_{-\infty}^{\infty} \varphi(t) \left[\frac{X_{t+h} - X_t}{h} - X'_t\right] \mathrm{d}t \right\|_{L^2(\mathrm{P})} \\
&\qquad\qquad \leqslant \lim_{h\to 0} \left\| \frac{X_h - X_0}{h} - X'_0 \right\|_{L^2(\mathrm{P})} \int_{-\infty}^{\infty} |\varphi(t)|\,\mathrm{d}t = 0.
\end{aligned}$$

That is,

$$\lim_{h\to 0} \int_{-\infty}^{\infty} \varphi(t) \frac{X_{t+h} - X_t}{h}\,\mathrm{d}t = \int_{-\infty}^{\infty} \varphi(t) X'_t\,\mathrm{d}t \qquad \text{in } L^2(\mathrm{P}), \tag{8.26}$$ intint:1

for every $\varphi \in C^1_c(\mathbb{R})$. Clearly,

$$\mathrm{P}\left\{ \int_{-\infty}^{\infty} \varphi(t) \frac{X_{t+h} - X_t}{h}\,\mathrm{d}t = \int_{-\infty}^{\infty} X_t \frac{\varphi(t-h) - \varphi(t)}{h}\,\mathrm{d}t \right\} = 1,$$

owing to a change of variables. Because of (8.15) and Minkowski's inequality,

$$\limsup_{h\to 0}\left\|\int_{-\infty}^{\infty} X_t \frac{\varphi(t-h)-\varphi(t)}{h}\,\mathrm{d}t + \int_{-\infty}^{\infty} X_t\varphi'(t)\,\mathrm{d}t\right\|_{L^2(\mathrm{P})} \leqslant \lim_{h\to 0}\int_{-\infty}^{\infty}\left|\varphi'(t) - \frac{\varphi(t)-\varphi(t-h)}{h}\right|\mathrm{d}t = 0.$$

That is,

$$\lim_{h\to 0}\int_{-\infty}^{\infty} X_t \frac{\varphi(t-h)-\varphi(t)}{h}\,\mathrm{d}t = -\int_{-\infty}^{\infty}\varphi'(t)X_t\,\mathrm{d}t \qquad \text{in } L^2(\mathrm{P}),$$

for every $\varphi \in C_c^1(\mathbb{R})$. Therefore, it follows from (8.26) that

$$\mathrm{P}\left\{\int_{-\infty}^{\infty}\varphi(t)X_t'\,\mathrm{d}t = -\int_{-\infty}^{\infty}\varphi'(t)X_t\,\mathrm{d}t\right\} = 1,$$

for every $\varphi \in C_c^1(\mathbb{R})$. A standard approximation argument can now be used to show that $\mathrm{P}\{\int_0^a X_t'\,\mathrm{d}t = X_a - X_0\} = 1$ for every $a \in \mathbb{R}$ and in particular the following holds with probability one:

$$\int_0^a X_t'\,\mathrm{d}t = X_a - X_0, \tag{8.27}$$ intint:2

simultaneously for every rational number a. Since both sides of (8.27) describe a.s.-continuous functions of a, it follows that (8.27) holds for all $a \in \mathbb{R}$ off a single P-null set. This completes the proof. □

Problems

Unless it is stated to the contrary, throughout these problems, $(T\,,d)$ denotes a metric space, N_T denotes the correspoding metric entropy, and

$$\mathcal{I}(\delta) := \int_0^\delta \sqrt{\log N_T(r)}\,\mathrm{d}r \qquad \text{and} \quad \mathcal{J}(\delta) := \int_0^\delta \sqrt{\log(N_T(r) \vee 2)}\,\mathrm{d}r,$$

for all $\delta > 0$. Also, let $\Delta(T)$ denote the d-diameter of T, as before.

⟨Pbm:I⟩

1. Verify that $\mathcal{I}(\Delta(T)) = \int_0^\infty \sqrt{\log N_T(r)}\,\mathrm{d}r$, and also prove that

$$\mathcal{I}(\delta) \leqslant \mathcal{I}(\delta/2) \leqslant 2\mathcal{I}(\delta),$$

for all $\delta \in (0\,,\Delta(T)]$.

2. Let $T = [0\,,1]$ and $d(s\,,t) := |t-s|$ for all $s,t \in T$. Improve the metric entropy estimates of Examples 1.3 and 1.6 by showing that $\varepsilon N_T(\varepsilon) \to 1$ as $\varepsilon \downarrow 0$.
3. Verify the claims of Example 1.7.
4. Suppose $T = \{1\,,\ldots,n\}$ where $n \geqslant 2$ is integral, and suppose $\{X_i\}_{i\in T}$ is a collection of i.i.d. standard-normal random variables. Let $M_n := \max_{1\leqslant i\leqslant n} X_i$.
   (a) Use Corollary 2.7 to prove that

$$A\sqrt{\log n} \leqslant \mathcal{I}(\Delta(T)) \leqslant \mathcal{J}(\Delta(T)) \leqslant B\sqrt{\log n},$$

   where $A, B > 0$ are real numbers that do not depend on n.
   (b) Conclude that $C\sqrt{\log n} \leqslant \mathrm{E}(M_n) \leqslant D\sqrt{\log n}$ for all $n \geqslant 1$ where $C, D > 0$ are real numbers that do not depend on n.
   (c) Compare with Proposition 1.3 (p. 7), and discuss the efficiency of Theorems 1.8 and 5.16 in the present setting.
5. Extend Lemma 2.4 by proving that for every $p \geqslant 1$,

$$c_p := \sup_{\lambda>0} \frac{\mathcal{T}_p(\lambda)}{[\log(\lambda \vee 2)]^{p/2}} < \infty.$$

   (a) Use your result in order to deduce (8.12) from Theorem 2.5.
   (b) The preceding problem came about in the proof of Corollary 2.7 because we needed to prove that $\sup_{t\in S}\mathrm{Var}(X_t) < \infty$. For a less heavy-handed approach prove directly that

$$\sup_{t\in S}\mathrm{Var}(X_t) \leqslant \Delta(S) + \inf_{s\in S}\mathrm{Var}(X_s).$$

   Then prove that (8.14) implies that $\Delta(S) < \infty$, which is tacitly assumed throughout any way.
6. Let $X := \{X_t\}_{t\in T}$ be a mean-zero Gaussian process where $(T\,,d)$ is a separable metric space. Suppose also that X is a.s. continuous and satisfies $\mathrm{E}(\sup_{t\in T} X_t) < \infty$ and $\mathrm{Var}(X_s) > 0$ for some $s \in T$. Prove that

$$\lim_{y\to\infty} \frac{1}{y^2}\log \mathrm{P}\left\{\sup_{t\in T}|X_t| > y\right\} = -\frac{1}{2\sup_{t\in T}\mathrm{Var}(X_t)}.$$

   This is due to Landau and Shepp XXX. (Hint: See Theorem 2.1, p. 71.)

7. On one hand, use Dudley's theorem [Theorem 2.5, p. 135] to verify that the examples in §2.1 through §2.6 of the previous chapter [pp. 116–118] are all continuous Gaussian processes. [Do not appeal to Kolmogorov's continuity theorem as a simpler alternative.] On the other hand, prove also that the isonormal process of §2.7 [p. 120] is not continuous a.s.
8. Carefully verify all three lines in (8.24).

⟨pbm:KCT⟩

9. Let $\{X_t\}_{t\in T}$ be a Gaussian process and suppose that $(T\,,d)$ is separable and $0 < \dim_{\mathrm{M}}(T) < \infty$, where d is defined in (8.10).
   (a) Prove that X has a continuous modification Y that satisfies the following for all $\delta \in (0\,,1)$:

$$\mathrm{E}\left(\sup_{\substack{s,t\in T:\\ d(s,t)\leqslant\delta}} |Y_t - Y_s|\right) \leqslant A\delta\sqrt{\log(1/\delta)},$$

   where $A = A(T\,,d)$ is a positive real number.
   (b) Conclude that

$$\Lambda(\alpha) := \mathrm{E}\left(\sup_{\substack{s,t\in T:\\ d(s,t)>0}} \frac{|Y_t - Y_s|}{[d(s\,,t)]^\alpha}\right) < \infty \quad \text{whenever } 0 < \alpha < \frac{1}{\dim_{\mathrm{M}}(T)}.$$

   (Hint: Start by checking that

$$\Lambda(\alpha) \leqslant \sum_{n=0}^\infty \mathrm{E}\left(\sup_{\substack{s,t\in T:\\ 2^{-n-1}\Delta(T)\leqslant d(s,t)<2^{-n}\Delta(T)}} \frac{|Y_t - Y_s|}{[d(s\,,t)]^\alpha}\right),$$

   for every $\alpha > 0$.)
   (c) Deduce the Kolmogorov continuity theorem (Proposition 2.2, p. 119) for Gaussian processes from the preceding parts.

⟨Pbm:KCT:1⟩

10. (Problem 9, continued) Let $\{X_t\}_{t\geqslant 0}$ denote an fBm with Hurst index $\alpha \in (0\,,1)$.
   (a) Prove that there exist real numbers $0 < c_1 \leqslant c_2$ such that

$$c_1\delta^\alpha\sqrt{\log(1/\delta)} \leqslant \mathrm{E}\left(\sup_{\substack{0\leqslant s,t\leqslant 1:\\ |t-s|\leqslant\delta}} |X_t - X_s|\right) \leqslant c_2\delta^\alpha\sqrt{\log(1/\delta)},$$

   for all $\delta \in (0\,,1)$. (Hint: See Problem 5, p. 126.)
   (b) For every $\kappa > 0$ let C^κ denote the collection of all Hölder-continuous, real-valued functions on $[0\,,1]$ with Hölder exponent κ. That is, $h \in C^\kappa$ iff

$$\|h\|_{C^\kappa} := \sup_{0\leqslant s<t\leqslant 1} \frac{|h(t) - h(s)|}{|t-s|^\kappa} < \infty.$$

   Conclude from the preceding that, for every $\kappa > 0$,

$$\mathrm{P}\{X \in C^\kappa\} = \begin{cases} 1 & \text{if } \kappa < \alpha,\\ 0 & \text{if } \kappa > \alpha.\end{cases}$$

   (c) It can be shown that $\mathrm{P}\{X \in C^\alpha\} = 0$ as well; see XXX. Verify this in the case that $\alpha \in (0\,,1/2]$, using Problem 4, p. 126.

11. (Problem 10, continued) Let $\{X_t\}_{t\geqslant 0}$ denote an fBm with index $\alpha \in (0\,,1)$. Prove that there exist real numbers $c_1, c_2 > 0$ such that

$$c_1\delta^\alpha \leqslant \mathrm{E}\left(\sup_{0\leqslant t\leqslant\delta} X_t\right) \leqslant c_2\delta^\alpha \qquad \text{for every } \delta \in (0\,,1).$$

12. Let $X := \{X_t\}_{t\in\mathbb{R}}$ be a stationary Gaussian process that satisfies (8.15) and let $\rho(t) := \mathrm{Cov}(X_s\,, X_t)$ for all $t \in \mathbb{R}$.
    (a) Prove that, if $\rho \in C^{2n}(\mathbb{R})$ for some integer $n \geqslant 1$, then $-\mathrm{d}^{2n}\rho/\mathrm{d}t^{2n}$ is positive definite; that is, prove that

    $$\sum_{j=1}^N\sum_{k=1}^N z_j z_k \frac{\mathrm{d}^{2n}}{\mathrm{d}t^{2n}}\rho(t_j - t_k) \leqslant 0 \qquad \text{for all } z_1,\ldots,z_N, t_1,\ldots,t_N \in \mathbb{R}.$$

    (b) Construct an example of a covariance function ρ such that $\rho(0) = 1$ and $\rho \in C^\infty(\mathbb{R})$.
13. Prove that if $Y := \{Y_t\}_{t\in\mathbb{R}}$ is a stationary mean-zero Gaussian process, then Y is continuous [up to a modification] if and only if the Gaussian process $\{Y_t\}_{t\in[0,1]}$ is continuous [up to a modification]. Apply this with $Y := X'$ in Theorem 4.2 to complete the proof of the necessity of the finiteness condition for $\mathcal{I}(\Delta([0\,,1]))$.

⟨pbm:X':Holder⟩ 14. Suppose $X := \{X_t\}_{t\in\mathbb{R}}$ is a stationary Gaussian process that satisfies (8.15). Suppose also that $\rho(t) := \mathrm{Cov}(X_t\,, X_0)$ is twice continuously differentiable on $\mathbb{R}$ and its second derivative ρ'' is Hölder continuous at zero. That is, suppose there exist $t_0, \theta > 0$ such that $1 - \rho''(t) \leqslant C|t|^\theta$ whenever $|t| \leqslant t_0$. Then prove that $\mathrm{P}\{X \in C^1(\mathbb{R})$ and X' is Hölder continuous$\} = 1$.
15. (Problem 14, continued) Suppose $\rho \in C^{2n}(\mathbb{R})$ for some integer $n \geqslant 1$ and the nth derivative $\mathrm{d}^n\rho/\mathrm{d}t^n$ is Hölder continuous at zero. Prove that $\mathrm{P}\{X \in C^n(\mathbb{R})$ and $\mathrm{d}^nX/\mathrm{d}t^n$ is Hölder continuous$\} = 1$.
16. Let X be a stationary Gaussian process that satisfies (8.15) and choose and fix an integer $n \geqslant 1$. Find a necessary and sufficient condition for $\mathrm{P}\{X \in C^n(\mathbb{R})\} = 1$, stated solely in terms of the covariance function ρ.

?⟨pbm:X:X'⟩? 17. Suppose X is a stationary Gaussian process that satisfies (8.15), and assume that $X \in C^1(\mathbb{R})$ a.s.
    (a) Prove that the collection of all random variables of the form either X_t or X_t' is a mean zero Gaussian process.
    (b) Compute $\mathrm{Cov}(X_t\,, X_s')$ for every $s, t \geqslant 0$. Use your formula to conclude that X_t and X_t' are independent random variables for every fixed choice of $t \in \mathbb{R}$. Prove also that the Gaussian process X and X' are not independent.
    (c) Is the last independence property still valid if we drop the stationarity assumption for X? Prove or construct a counterexample.

Chapter 9

Level Sets

1 Banach's Theorem, and Some Applications

The principal aim of this chapter is to study the level sets of Gaussian processes. To see why this might be interesting, let us briefly consider the following early example, to which we shall return later in this chapter in greater depth.

Let $q > 1$ be a fixed integer and consider the random polynomial,

$$X_t := Z_0 + Z_1 t + \cdots + Z_q t^q \qquad \text{for all } t \in \mathbb{R}. \tag{9.1}$$ rdm:poly

where $Z_0, Z_1, \ldots$ are independent random variables, each distributed as $\mathrm{N}(0\,,1)$. One can think of X as a "typical" polynomial. With this interpretation in mind, it might be interesting to know about the expected number of real zeros,

$$\mathcal{N}^{(q)} := \mathrm{E}(\#\{t \in \mathbb{R} :\ X_t = 0\}),$$

of the random polynomial X. Does $\mathcal{N}^{(q)}$ grow as $q \to \infty$? If so, then how fast? Where do we think the real zeros of X are? Etc. Answers will come in due time. For now, suffice it to say that the answers lie in a part of classical function theory that surrounds "Banach's indicatrix theorem." Thus, we begin with that result.

§1.1 Banach's Indicatrix Theorem

For every continuous function $f : \mathbb{R} \to \mathbb{R}$ and real numbers $a < b$ and y, define

$$N_{[a,b]}(f\,, y) := \#\, \{x \in [a\,, b] :\ f(x) = y\}\,, \tag{9.2}$$ N(f)

where "$N_{[a,b]}(f\,, y) = \infty$" is permitted. We frequently write $N_{[a,b]}(f)$ for the function $y \mapsto N_{[a,b]}(f\,, y)$, and refer to $N_{[a,b]}(f)$ as the *counting function* – or *indicatrix* – of f. For a graphical representation see Figure 9.1. That figure depicts a function f together with its counting function $N_{[a,b]}(f)$ drawn reflected and rotated in order to make the construction clear.

As was suggsted in the preamble to this chapter, we will be interested mainly in analyzing $N_{[a,b]}(f)$ when f is a realization of a nice Gaussian process. But before we study random functions, let us say a few things about a beautiful classical theorem, due to Stephan Banach, that studies non-random functions f. Our forthcoming discussion will culminate in the following elegant theorem.

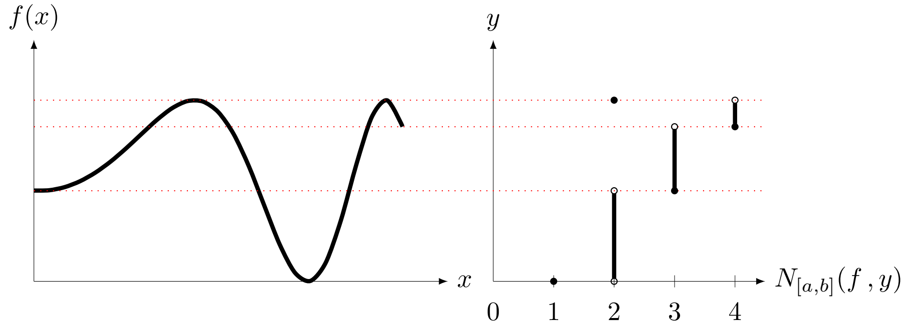

Figure 9.1. The graphs of a function f and its counting function $N_{[a,b]}(f)$

⟨fig:f:N(f)⟩

⟨th:Banach⟩ **Theorem 1.1** (Banach XXX)**.** *If $f \in C^1(\mathbb{R})$, then for all real numbers $a < b$ and all continuous functions $\Psi : \mathbb{R} \to \mathbb{R}$,*

$$\int_{-\infty}^{\infty} \Psi(y) N_{[a,b]}(f\,, y)\, \mathrm{d}y = \int_a^b \Psi(f(x))|f'(x)|\, \mathrm{d}x.$$

We now begin work toward proving Theorem 1.1. This undertaking requires some notions, from elementary function theory, which we recall and develop next.

Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function on $\mathbb{R}$ and $a < b$ two real numbers. Let P be a partition of $[a\,, b]$ with $n+1$ elements $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$, and define

$$S(P\,, f) := \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)|.$$

The *total variation* of f on $[a\,, b]$ is defined as

$$V_{[a,b]}(f) := \sup S(P\,, f),$$

where the supremum is taken over all such partitions P of the interval $[a\,, b]$. If $V_{[a,b]}(f) < \infty$ then we say that f has *bounded variation* on $[a\,, b]$. If f has bounded variation on $[a\,, b]$ for every two real numbers $a < b$, then we say that f has *bounded variation*.

⟨pr:V(f)⟩ **Proposition 1.2.** *Every C^1 function on $\mathbb{R}$ has bounded variation. Moreover,*

$$V_{[a,b]}(f) = \int_a^b |f'(x)|\, \mathrm{d}x, \tag{9.3}$$ eq:V(f)

for all real numbers $a < b$ and $f \in C^1(\mathbb{R})$.

Proof. Choose and fix real numbers $a < b$ and $f \in C^1(\mathbb{R})$. It suffices to verify the identity (9.3).

Choose and fix a real number $\varepsilon > 0$. Let us also choose and fix a partition P of $[a\,, b]$ with nodes $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ such that

$$V_{[a,b]}(f) - \varepsilon \leqslant S(P\,, f) \leqslant V_{[a,b]}(f). \tag{9.4}$$ VSV

If we add finitely-many additional points to P, then we obtain a new partition Q of $[a\,,b]$ which manifestly satisfies $V_{[a,b]}(f)-\varepsilon \leqslant S(P\,,f)\leqslant S(Q\,,f)\leqslant V_{[a,b]}(f)$. Such a partition Q is sometimes called a *refinement* of P.

We can assume, without loss in generality, that P has been refined enough already so that

$$\sup_{x,y\in[x_i,x_{i+1}]}|f'(x)-f'(y)|\leqslant\varepsilon \qquad \text{for all } i=0,\ldots,n-1. \tag{9.5}$$ f'-f'

For otherwise, the continuity of f' allows us to replace P by a refinement of P that satisfies (9.5). In this case, we can observe that for all $i=0,\ldots,n-1$,

$$\begin{aligned}\int_{x_i}^{x_{i+1}}|f'(x)|\,\mathrm{d}x &\leqslant \big(|f'(x_i)|+\varepsilon\big)\,(x_{i+1}-x_i)=\left|\int_{x_i}^{x_{i+1}}f'(x_i)\,\mathrm{d}x\right|+\varepsilon(x_{i+1}-x_i)\\ &\leqslant \left|\int_{x_i}^{x_{i+1}}f'(x)\,\mathrm{d}x\right|+2\varepsilon(x_{i+1}-x_i)\\ &=|f(x_{i+1})-f(x_i)|+2\varepsilon(x_{i+1}-x_i),\end{aligned}$$

and a matching lower bound also holds provided that we replace ε everywhere by $-\varepsilon$. This shows that, for all $i=0,\ldots,n-1$,

$$\left|\int_{x_i}^{x_{i+1}}|f'(x)|\,\mathrm{d}x-|f(x_{i+1})-f(x_i)|\right|\leqslant 2\varepsilon(x_{i+1}-x_i).$$

Sum over i in order to deduce from (9.4) that

$$\left|\int_a^b|f'(x)|\,\mathrm{d}x-S(P\,,f)\right|\leqslant 2\varepsilon(b-a)\ \Rightarrow\ \left|\int_a^b|f'(x)|\,\mathrm{d}x-V_{[a,b]}(f)\right|\leqslant\varepsilon(2b-2a+1),$$

as long as P is refined enough to ensure (9.5). This does the job because $\varepsilon>0$ can be made to be as small as we want. □

We have delved into the theory of functions of bounded variation because it has a connection to Banach's indicatrix. The next proposition makes clear the relationship between the indicatrix of a nice function f and its total variation; see also Proposition 1.2.

⟨pr:Banach⟩ **Proposition 1.3.** *If $f\in C^1(\mathbb{R})$, then for all real numbers $a<b$,*

$$\int_{-\infty}^{\infty}N_{[a,b]}(f\,,y)\,\mathrm{d}y=\int_a^b|f'(x)|\,\mathrm{d}x. \tag{9.6}$$ eq:Banach

Proposition 1.3 is a deceptively-subtle statement. For example, it has the non-trivial corollary that $N_{[a,b]}(f)<\infty$ almost everywhere for every $f\in C^1(\mathbb{R})$. For a second example, we may observe that, whereas the right-hand side of (9.6) is a Riemann integral, the left-hand side is not; though it is in general a Lebesgue integral. In order to see why, consider the case that f is a constant—say c—on $[a\,,b]$. In that case, $N_{[a,b]}(f\,,y)=0$ for all $y\neq c$ and $N_{[a,b]}(f\,,c)=\infty$. This suffices to show that $N_{[a,b]}(f)$ is not Riemann integrable in this case. In fact, the Riemann-integrability of the indicatrix is delicate enough to merit a more detailed discussion. We do that in the following form next.

⟨lem:N:meas⟩ **Lemma 1.4.** *$N_{[a,b]}(f)$ is Borel measurable for every two real numbers $a<b$ provided only that f is continuous.*

Remark 1.5. Lemma 1.4 does not claim that the Lebesgue integral $\int_{-\infty}^{\infty} N_{[a,b]}(f\,, y)\,\mathrm{d}y$ is finite. Only that it is well defined.

Proof. Define for every integer $n \geqslant 0$,

$$I_{k,n} := \left[a + \frac{k(b-a)}{2^n}\,, a + \frac{(k+1)(b-a)}{2^n}\right] \quad \text{for all } k = 0, \ldots, 2^n - 1, \tag{9.7}$$ Ikn

and

$$\mathcal{N}^{(n)}_{[a,b]}(f\,, y) := \sum_{k=0}^{2^n-1} \mathbb{1}_{f(I_{k,n})}(y) \qquad \text{for each } y \in \mathbb{R},$$

where $f(I_{k,n}) := \{f(x) : x \in I_{k,n}\}$ is, as usual, the image of $I_{k,n}$ under the map f. Because f is continuous, every image $f(I_{k,n})$ is in fact a closed interval. Therefore, $\mathcal{N}^{(n)}_{[a,b]}(f)$ is an elementary function in the sense of Lebesgue's integration theory.

Next, we may observe that $\mathcal{N}^{(n)}_{[a,b]}(f\,, y) \leqslant \mathcal{N}^{(n+1)}_{[a,b]}(f\,, y)$ for every $y \in \mathbb{R}$ and $n \geqslant 0$. Therefore,

$$\mathcal{N}^{(\infty)}_{[a,b]}(f\,, y) := \lim_{n\to\infty} \mathcal{N}^{(n)}_{[a,b]}(f\,, y)$$

exists pointwise, and hence defines a Borel-measurable function. It remains to prove that $\mathcal{N}^{(\infty)}_{[a,b]}(f) = N_{[a,b]}(f)$. Because $\mathcal{N}^{(n)}_{[a,b]}(f) \leqslant N_{[a,b]}(f)$ pointwise, we can see immediately that $\mathcal{N}^{(\infty)}_{[a,b]}(f) \leqslant N_{[a,b]}(f)$ pointwise. We plan to prove that the converse inequality also holds.

Choose and fix an arbitrary $y \in \mathbb{R}$ and let L denote any non-negative integer that satisfies $L \leqslant N_{[a,b]}(f\,, y)$. Fix such an L. There exist L distinct points $x_1, \ldots, x_L \in [a\,, b]$ such that $f(x_i) = y$ for all $i = 1, \ldots, L$. If we choose $n \geqslant 1$ large enough then $x_1, \ldots, x_L$ fall in L different dyadic intervals among $I^{(n)}_1, \ldots, I^{(n)}_{2^n-1}$. In particular, $L \leqslant \mathcal{N}^{(n)}_{[a,b]}(f\,, y)$ and hence $L \leqslant \mathcal{N}^{(\infty)}_{[a,b]}(f\,, y)$. Let L saturate $N_{[a,b]}(f\,, y)$ to see that $N_{[a,b]}(f\,, y) \leqslant \mathcal{N}^{(\infty)}_{[a,b]}(f\,, y)$ and complete the proof. □

With measurability under way, we can establish Proposition 1.3.

Proof of Proposition 1.3. We continue to use the notation and construction of the proof of Lemma 1.4, and observe that

$$\int_{-\infty}^{\infty} \mathcal{N}^{(n)}_{[a,b]}(f\,, y)\,\mathrm{d}y = \sum_{k=0}^{2^n-1}\left(\sup_{I_{k,n}} f - \inf_{I_{k,n}} f\right) = \sum_{k=0}^{2^n-1}\left|\sup_{I_{k,n}} f - \inf_{I_{k,n}} f\right|. \tag{9.8}$$ Nsupsup

Because f is continuous, the suprema and infima are all attained, and hence

$$\int_{-\infty}^{\infty} \mathcal{N}^{(n)}_{[a,b]}(f\,, y)\,\mathrm{d}y \leqslant V_{[a,b]}(f).$$

Let $n \to \infty$ and recall from the proof of Lemma 1.4 that $\mathcal{N}^{(n)}(f)$ converges upward to $N_{[a,b]}(f)$ pointwise as $n \to \infty$. Therefore, the monotone convergence theorem yields $\int_{-\infty}^{\infty} N_{[a,b]}(f\,, y)\,\mathrm{d}y \leqslant V_{[a,b]}(f) = \int_a^b |f'(x)|\,\mathrm{d}x$, thanks to Proposition 1.2. We work to establish the converse inequality now.

Choose and fix an arbitrary $\varepsilon > 0$ and note that there exists $\delta > 0$ such that

$$|f(u) - f(v)| \geqslant |f'(v)|(u - v) - \varepsilon(u - v).$$

simultaneously for all $u, v \in [a\,,b]$ that satisfy $u > v$ and $|u - v| < \delta$. Now, if I is a closed interval in $[a\,,b]$ then $\sup_I f - \inf_I f = \sup_{u,v\in I} |f(u) - f(v)|$. These observations and (9.8) together imply that

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{N}^{(n)}_{[a,b]}(f\,,y)\,\mathrm{d}y &\geqslant \sum_{k=0}^{2^n-1} \left| f\left(a + \frac{(k+1)(b-a)}{2^n}\right) - f\left(a + \frac{k(b-a)}{2^n}\right)\right| \\ &\geqslant \frac{b-a}{2^n} \sum_{k=0}^{2^n-1} \left| f'\left(a + \frac{k(b-a)}{2^n}\right)\right| - \varepsilon(b-a).\end{aligned}$$

Because the sum on the right-most side is the Riemann-sum approximation of the total variation $\int_a^b |f'(x)|\,\mathrm{d}x$ of f, we can let $n \to \infty$ to find that $\int_{-\infty}^{\infty} N_{[a,b]}(f\,,y)\,\mathrm{d}y \geqslant \int_a^b |f'(x)|\,\mathrm{d}x - \varepsilon(b-a)$. This completes the proof because $\varepsilon > 0$ is arbitrary. $\square$

We are now in position to prove Theorem 1.1.

Proof of Theorem 1.1. Throughout, we choose and fix two real numbers $a < b$.

Recall the dyadic intervals $I_{k,n}$'s from (9.7). Because Ψ and $\Psi \circ f$ are respectively uniformly continuous on $f([a\,,b])$ and $[a\,,b]$, it follows that for every $\varepsilon > 0$ there exists $n_0(\varepsilon) > 0$ such that whenever $n > n_0(\varepsilon)$,

$$\sup_{f(I_{k,n})} |\Psi(u) - \Psi(v)| \leqslant \varepsilon \quad\text{and}\quad \sup_{s,t\in I_{k,n}} |\Psi(f(s)) - \Psi(f(t))| \leqslant \varepsilon,$$

simultaneously for all $k = 0, \ldots, 2^n - 1$. Therefore, if $n > n_0(\varepsilon)$ then

$$\begin{aligned}&\left| \int_{f(I_{k,n})} \Psi(y) N_{[a,b]}(f\,,y)\,\mathrm{d}y - \Psi\left(f\left(k2^{-n}\right)\right) \int_{f(I_{k,n})} N_{[a,b]}(f\,,y)\,\mathrm{d}y \right| \\ &\qquad\qquad\leqslant \varepsilon \int_{f(I_{k,n})} N_{[a,b]}(f\,,y)\,\mathrm{d}y,\end{aligned}$$

for all $k = 0, \ldots, 2^n - 1$. Because $N_{[a,b]}(f) = N_{I_{k,n}}(f)$ on $f(I_{k,n})$, it follows that

$$\int_{f(I_{k,n})} N_{[a,b]}(f\,,y)\,\mathrm{d}y = \int_{f(I_{k,n})} N_{I_{k,n}}(f\,,y)\,\mathrm{d}y = \int_{I_{k,n}} |f'(x)|\,\mathrm{d}x;$$

see Proposition 1.3. Therefore, if $n > n_0(\varepsilon)$ then

$$\left| \int_{f(I_{k,n})} \Psi(y) N_{[a,b]}(f\,,y)\,\mathrm{d}y - \Psi\left(f\left(k2^{-n}\right)\right) \int_{I_{k,n}} |f'(x)|\,\mathrm{d}x \right| \leqslant \varepsilon \int_{I_{k,n}} |f'(x)|\,\mathrm{d}x,$$

for all $k = 0, \ldots, 2^n - 1$. Similarly,

$$\left| \Psi\left(f\left(k2^{-n}\right)\right) \int_{I_{k,n}} |f'(x)|\,\mathrm{d}x - \int_{I_{k,n}} \Psi(f(x))|f'(x)|\,\mathrm{d}x \right| \leqslant \varepsilon \int_{I_{k,n}} |f'(x)|\,\mathrm{d}x,$$

for all $k = 0, \ldots, 2^n - 1$. Therefore, by the triangle inequality,

$$\left| \int_{f(I_{k,n})} \Psi(y) N_{[a,b]}(f\,,y)\,\mathrm{d}y - \int_{I_{k,n}} \Psi(f(x))|f'(x)|\,\mathrm{d}x \right| \leqslant 2\varepsilon \int_{I_{k,n}} |f'(x)|\,\mathrm{d}x,$$

for all $k = 0, \ldots, 2^n - 1$, provided only that $n > n_0(\varepsilon)$. Sum over all these k's to see that, as long as $n > n_0(\varepsilon)$, as have

$$\left| \int_{f([a,b])} \Psi(y) N_{[a,b]}(f\,, y)\,\mathrm{d}y - \int_a^b \Psi(f(x))|f'(x)|\,\mathrm{d}x \right| \leqslant 2\varepsilon \int_a^b |f'(x)|\,\mathrm{d}x.$$

The theorem follows from this because: (i) $\varepsilon > 0$ is arbitrary; and (ii) $N_{[a,b]}(f\,, y) = 0$ when $y \notin f([a\,, b])$. □

§1.2 Rice's Formula

Let $X := \{X_t\}_{t\in\mathbb{R}}$ denote a mean-zero, stationary Gaussian process with

$$\mathrm{Var}(X_0) = 1 \qquad \text{and} \qquad \rho(s) = \mathrm{E}(X_0 X_s) \qquad \text{for all } s \in \mathbb{R}. \tag{9.9}$$ Rice:Var:Corr

Suppose further that

$$\mathrm{P}\{X \in C^1(\mathbb{R})\} = 1. \tag{9.10}$$ [1] Rice:C1

Recall that the *level set* of X at a point $y \in \mathbb{R}$ is $X^{-1}(\{y\}) := \{t \in \mathbb{R} : X_t = y\}$. A natural measure of the size of the level set $X^{-1}(\{y\})$ is of course the indicatrix $N_{[a,b]}(X)$ of X, measured for all $a < b$; see (9.2) on page 153. For example, Proposition 1.3 on page 155, and a little bit of measure theory, together imply that there exists a Lebesgue-null set $\mathcal{Y} \subset \mathbb{R}$ such that a.s.,

$$N_{[a,b]}(X\,, y) < \infty \quad \text{for all } y \notin \mathcal{Y} \text{ and real numbers } a < b.$$

One of the main goals of this section is to provide additional detail on the distribution of the indicatrix of X. Specifically, we plan to prove the following.

⟨th:Rice⟩ **Theorem 1.6** (Rice's Formula)**.** *Suppose $X := \{X_t\}_{t\in\mathbb{R}}$ is a mean-zero stationary Gaussian process that satisfies* (9.9) *and* (9.10)*. Then,*

$$\mathrm{E}\left[N_{[a,b]}(X\,, y)\right] = \frac{(b-a)\sqrt{|\rho''(0)|}}{\pi} \mathrm{e}^{-y^2/2}, \tag{9.11}$$?eq:Rice?

for all real numbers $a < b$ and y.

We begin the proof with the following important technical result. In somewhat imprecise words, this result states that, with probability one, the extreme points of X can only take truly-random values. A more precise statement follows.

⟨lem:Bulinskaya⟩ **Lemma 1.7** (Bulinskaya XXX)**.** *For all $y \in \mathbb{R}$,*

$$\mathrm{P}\left\{\exists t \in \mathbb{R} \text{ such that } X_t = y \text{ and } X'_t = 0\right\} = 0.$$

Proof. It is easy to see that

$$E_N := \left\{\omega : \exists t \in [N\,, N+1) \text{ such that } X_t(\omega) = y \text{ and } X'_t(\omega) = 0\right\} \tag{9.12}$$ E:Bulinskaya

is measurable for every $N \in \mathbb{Z}$. We plan to prove that $\mathrm{P}(E_0) = 0$. This will do the job since $\mathrm{P}\{\exists t \in \mathbb{R} : X_t = y \text{ and } X'_t = 0\} \leqslant \sum_{N\in\mathbb{Z}} \mathrm{P}(E_N)$, which will be zero since every summand is equal to $\mathrm{P}(E_0)$ by stationarity.

[1] Corollary 4.4 on page 146 contains a necessary and sufficient condition for condition (9.10) solely in terms of the correlation function of X.

Let us choose and fix an arbitrary $\varepsilon > 0$. Define for all integers $\ell, m, n \geqslant 1$ and $k = 0, \ldots, n-1$, the events

$$F_{k,n} := \left\{ \omega : X_t(\omega) = y \text{ and } X'_t(\omega) = 0 \text{ for some } \frac{k}{n} \leqslant t < \frac{k+1}{n} \right\},$$

$$G_{\ell,n} := \left\{ \omega : \mu_n(\omega) > \frac{1}{\ell} \right\} \quad \text{where} \quad \mu_n := \sup_{\substack{0 \leqslant u,v < 1 \\ |u-v| < 1/n}} \left| X'_u - X'_v \right|.$$

Then,

$$\mathrm{P}(E_0) \leqslant \sum_{k=0}^{n-1} \mathrm{P}\left(F_{k,n} \cap G^c_{\ell,n} \right) + \mathrm{P}(G_{\ell,n}), \tag{9.13}$$ P(E)P(E)

for all integers $\ell, n \geqslant 1$. Now we choose $\ell = \ell(n)$ as follows: Because of the a.s.-continuity of the process X', $\mu_n \to 0$ as $n \to \infty$, almost surely. Therefore, there exists $n_0(\varepsilon)$ such that for every $n \geqslant n_0(\varepsilon)$ we can find an integer

$$\ell = \ell(n) > 1/\varepsilon \quad \text{such that} \quad \mathrm{P}(G_{\ell(n),n}) \leqslant \varepsilon. \tag{9.14}$$ P(F)

This takes care of the final term in (9.13). We work to estimate the first term on the remainder of the right-hand side of (9.13) next.

Off a single P-null set, the following holds for every $t \in [k/n\,, (k+1)/n)$:

$$X_{k/n} = X_t - \int_{k/n}^t \left(X'_s - X'_t \right) \mathrm{d}s - X'_t \left(t - \frac{k}{n} \right).$$

Now,

$$\sup_{t \in [k/n, (k+1)/n]} \left| \int_{k/n}^t \left(X'_s - X'_t \right) \mathrm{d}s \right| \leqslant \frac{\mu_n}{n} \qquad \text{almost surely on } G^c_{\ell(n),n}.$$

Therefore,

$$F_{k,n} \cap G^c_{\ell(n),n} \subset \left\{ \omega : \left| X_{k/n}(\omega) - y \right| < \frac{1}{n\ell(n)} \right\}.$$

Because the probability density γ_1 of a $\mathrm{N}(0\,,1)$ is bounded above by $(2\pi)^{-1/2} < \frac{1}{2}$, it follows that $\mathrm{P}\{|X_{k/n} - y| < r\} = \int_{-r}^r \gamma_1(y-x)\,\mathrm{d}x < r$ for every $r > 0$. This yields

$$\mathrm{P}\left(F_{k,n} \cap G^c_{\ell(n),n} \right) < \frac{1}{n\ell(n)} \leqslant \frac{\varepsilon}{n},$$

valid for all $n > n_0(\varepsilon)$. It now follows from (9.13) and (9.14) that

$$\mathrm{P}(E_0) \leqslant 2\varepsilon \qquad \text{for all } \varepsilon > 0 \text{ and integers } n > n_0(\varepsilon).$$

Let $n \to \infty$, and then $\varepsilon \to 0$, in this order, to deduce the lemma. □

Next we present a slightly-different representation of the indicatrix of X. Before we do that, we need to introduce some notation.

For every two real numbers $a < b$ define

$$t_{k,n} = t_{n,k}(a\,,b) := \frac{k(b-a)}{2^n} \qquad \text{for } k = 0, \ldots, 2^n - 1 \text{ and } n \geqslant 0. \tag{9.15}$$ eq:tnk

If $f:\mathbb{R}\to\mathbb{R}$ and $y\in\mathbb{R}$, then we also let

$$\sigma_{k,n}(f\,,y) := \begin{cases} 1 & \text{if } [f(t_{k,n})-y]\,[f(t_{k+1,n})-y]<0,\\ 0 & \text{otherwise.}\end{cases}$$

Thus, $\sigma_{k,n}(f\,,y)=1$ if and only if one of $f(t_{k,n})$ and $f(t_{k+1,n})$ lies strictly above y and the other lies strictly below y. Define

$$K^{(n)}_{[a,b]}(f\,,y) := \sum_{k=0}^{2^n-1}\sigma_{k,n}(f\,,y) \qquad \text{for all } y\in\mathbb{R}.$$

The following shows that the quantity $K^{(n)}_{[a,b]}(f\,,y)$ frequently converges upward to $N_{[a,b]}(f\,,y)$ as $n\to\infty$. The reason that the following is helpful is that it is much simpler to analyze $K^{(n)}_{[a,b]}(f)$ than it is to analyze $N_{[a,b]}(f)$ because the former depends only on the values of f at the $t_{k,n}$'s, whereas the latter depends on the values of f at points that can be harder to pinpoint in general.

⟨lem:K⟩ **Lemma 1.8** (Grenander XXX)**.** *Choose and fix real numbers $a<b$ and y, and some $f\in C^1(\mathbb{R})$. Suppose that: (a) There is no $t\in\mathbb{R}$ such that $f(t)=y$ and $f'(t)=0$; and (b) $f(t_{n,k})\neq y$ for every $k=0,\ldots,2^n-1$ and $n\geqslant 0$, where $t_{n,k}=t_{n,k}(a\,,b)$ was defined in* (9.15)*. Then*

$$K^{(\infty)}_{[a,b]}(f\,,y) = \lim_{n\to\infty}K^{(n)}_{[a,b]}(f\,,y)$$

exists and is equal to $N_{[a,b]}(f\,,y)$ for every $y\in\mathbb{R}$.

Proof. We adapt the proof of Lemma 1.4 to the present setting.

Because

$$K^{(n)}_{[a,b]}(f\,,y)\leqslant K^{(n+1)}_{[a,b]}(f\,,y) \qquad \text{for all } n\geqslant 0 \text{ and } y\in\mathbb{R},$$

the positive integer $K^{(n)}_{[a,b]}(f\,,y)$ converges upward to a limit $K^{(\infty)}_{[a,b]}(f\,,y)$ for every $y\in\mathbb{R}$. If $\sigma_{k,n}(f\,,y)=1$ then the mean-value property of the function f ensures that $f(t)=y$ for some $t\in[t_{k,n}\,,t_{k+1,n}]$. This proves that $K^{(n)}_{[a,b]}(f\,,y)\leqslant N_{[a,b]}(f\,,y)$. Let $n\to\infty$ to see that $K^{(\infty)}_{[a,b]}(f\,,y)\leqslant N_{[a,b]}(f\,,y)$ pointwise. In order to prove the converse bound we need to introduce some notation.

We say that f *traverses through* $t\in\mathbb{R}$ *at* $y\in\mathbb{R}$ if there exists $\varepsilon=\varepsilon(t)>0$ such that $[f(u)-y][f(v)-y]<0$ for every $t-\varepsilon<u<v<t+\varepsilon$. See Figure 9.2. Because f is C^1, whenever $f(t)=y$ then either f traverses through y at t, or $f(t)=y$ and $f'(t)=0$, which cannot happen as per the assumption of the lemma. Therefore, $N_{[a,b]}(f\,,y)$ is equal to the total number of times f traverses y at some point in $[a\,,b]$. Suppose m is an integer such that $m\leqslant N_{[a,b]}(f\,,y)$. Then, there are at least m distinct points $z_1,\ldots,z_m\in[a\,,b]$ at which f traverses y. Because of this and the fact that $f(t_{j,\ell})\neq y$ for all j and ℓ it follows that, for all n large enough, the z_i's fall in distinct dyadic intervals of the form $[t_{k,n}\,,t_{k+1,n}]$ and hence

$$m\leqslant K^{(n)}_{[a,b]}(f\,,y)\leqslant K^{(\infty)}_{[a,b]}(f\,,y),$$

for all $m\leqslant N_{[a,b]}(f\,,y)$. This implies that $K^{(\infty)}_{[a,b]}(f\,,y)\geqslant N_{[a,b]}(f\,,y)$ and completes the proof. □

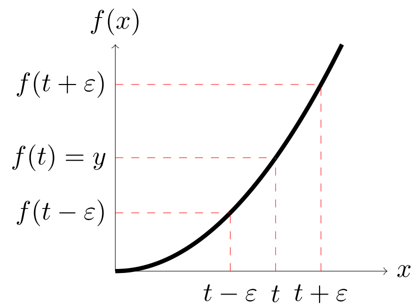

Figure 9.2. f traverses through y at t.

⟨fig:traverse⟩

The final step of proof of Theorem 1.6 is the following elegant limit result about our stationary Gaussian process X.

⟨lem:P(X<y<X)⟩ **Lemma 1.9** (Lindgren et al XXX)**.** *For every* $y \in \mathbb{R}$*,*

$$\lim_{t\to 0}\frac{1}{t}\mathrm{P}\{X_0 > y > X_t\} = \frac{\sqrt{|\rho''(0)|}}{2\pi}\,\mathrm{e}^{-y^2/2}.$$

Proof. The random vector $(X_0\,, X_t)$ has a $\mathrm{N}_2(0\,,\Gamma(t))$ distribution where $\Gamma_{1,1}(t) = \Gamma_{2,2}(t) = 1$ and $\Gamma_{1,2}(t) = \Gamma_{2,1}(t) = \rho(t)$. A quick computation of covariances show that X_0 and

$$Y_t := X_t - \rho(t)X_0$$

are independent, and Y_t has a mean-zero normal distribution with variance $\mathrm{Var}(Y_t) = [1-\rho(t)]^2$. Also, $\rho(0) = 1$ implies that there exists $t_0 \neq 0$ such that $\rho(t) > 0$ for all $t \in (-t_0\,, t_0)$. Therefore, for all $t \in (-t_0\,, t_0)$,

$$\begin{aligned}\mathrm{P}\{X_0 > y > X_t\} &= \mathrm{P}\left\{X_0 > y > \rho(t)X_0 + Y_t\right\}\\ &= \frac{1}{\sqrt{2\pi}}\,\mathrm{E}\left(\int_y^{(y-Y_t)/\rho(t)} \mathrm{e}^{-w^2/2}\,\mathrm{d}w;\ y < \frac{y-Y_t}{\rho(t)}\right).\end{aligned}$$

where $z_+ := \max(z\,,0)$, as is customary. Since $\rho'(0) = 0$ and $\rho''(0) < 0$—see (8.22)—a Taylor expansion shows that $\rho(t) = 1 - \frac{1}{2}t^2|\rho''(0)| + o(t^2)$ as $t \to \infty$, and hence

$$\frac{y-Y_t}{\rho(t)} = y - tX_0' + o(t) \qquad \text{a.s. and in } L^2(\mathrm{P}) \text{ as } t \to 0.$$

Therefore, the dominated convergence theorem yields

$$\lim_{t\to\infty}\frac{1}{t}\mathrm{P}\{X_0 > y > X_t\} = \frac{\mathrm{e}^{-y^2/2}}{\sqrt{2\pi}}\,\mathrm{E}\left(X_0';\ X_0' \geqslant 0\right) = \frac{\mathrm{e}^{-y^2/2}}{2\sqrt{2\pi}}\,\mathrm{E}\left(|X_0'|\right).$$

If W has a mean-zero normal distribution then $\mathrm{E}(|W|) = \sqrt{2\,\mathrm{Var}(W)/\pi}$. The result follows since X_0' has a $\mathrm{N}(0\,,|\rho''(0)|)$ distribution; see (8.23) on page 146. $\square$

We are in position to prove Theorem 1.6.

Proof of Theorem 1.6 (Grenander XXX). Choose and fix three real numbers $a < b$ and y. Because $\mathrm{P}\{X_t = y\} = 0$ for every $t \in \mathbb{R}$,

$$\mathrm{P}\left\{X_{t_{n,k}} = y \text{ for some } k = 0, \ldots, 2^n - 1 \text{ and } n \geqslant 0\right\} = 0,$$

where $t_{n,k}$ was defined in (9.15). This and Lemma 1.7 permit the use of Lemma 1.8, with $f \equiv X$, in order to see that

$$\mathrm{E}\left[N_{[a,b]}(X\,, y)\right] = \mathrm{E}\left[K^{(\infty)}_{[a,b]}(X\,, y)\right] = \lim_{n\to\infty} \mathrm{E}\left[K^{(n)}_{[a,b]}(X\,, y)\right],$$

where we have used the monotone convergence theorem to justify the last identity. Now, the stationarity of X implies that

$$\begin{aligned}\mathrm{E}\left[K^{(n)}_{[a,b]}(X\,, y)\right] &= 2^n \mathrm{P}\left\{[X_0 - y]\left[X_{(b-a)/2^n} - y\right] < 0\right\} \\ &= 2^{n+1}\mathrm{P}\left\{X_0 < y < X_{(b-a)/2^n}\right\},\end{aligned}$$

owing to the fact that X and $-X$ have the same law. Thus, Lemma 1.9 implies the theorem. □

§1.3 Kac's Theorem

We conclude this section by discussing a second application of Banach's theorem (Theorem 1.1). This application concerns the first set of questions that were asked at the beginning of this chapter. Namely, we would like to say a few things about the level sets of the Gaussian random polynomial in (9.1) (see p. 153). More precisely, we plan to prove the following result of M. Kac.

⟨th:Kac⟩ **Theorem 1.10** (Kac, XXX). *Let X be the Gaussian random polynomial of* (9.1), *where $q \geqslant 1$ is a non- random integer. Then, for every two real numbers $a < b$,*

$$\mathrm{E}\left[N_{[a,b]}(X\,, 0)\right] = \frac{1}{\pi}\int_a^b \sqrt{\frac{1}{(t^2-1)^2} - \frac{(q+1)^2 t^{2q}}{(t^{2q+2}-1)^2}}\,\mathrm{d}t. \tag{9.16}$$ Kac:E(N)

Originally, Kac XXX discovered an equivalent, though slightly different, representation of the formula (9.16). Edelman and Kostlan XXX found the formula the way it is stated here, and used it to expand on Kac's asymptotic evaluation of the expected number of real zeros of X as $q \to \infty$. In order to introduce Kac's asymptotic evaluation, let us write

$$N_{(-\infty,\infty)}(X\,, 0) := \sup_{a<b} N_{[a,b]}(X\,, 0)$$

for the total number of the real-valued zeros of X. Then we have the following.

⟨co:th:Kac⟩ **Corollary 1.11** (Kac, XXX). $\mathrm{E}[N_{(-\infty,\infty)}(X\,,0)] \sim (2/\pi)\log q$ *as* $q \to \infty$.

Thus, we see that the expected number of real zeros of X tends logarithmically to infinity as $q \to \infty$. It shall follow also from Theorem 1.10, Corollary 1.11, and (9.23) below that we can expect all but a bounded number of the zeros of X to lie in arbitrarily-small open neighborhoods of ± 1. These form fairly definitive answers to our earlier questions that followed (9.1); see page 153.

Theorem 1.10 and Corollary 1.11 both rely heavily on Banach's indicatrix of the random polynomial X (see Proposition 1.3, p. 155). As a first step toward clarifying this let us observe that Theorem 1.1 and a standard approximation argument from integration theory together yield the following.

⟨co:Banach⟩ **Corollary 1.12.** *If $f \in C^1(\mathbb{R})$ and $a < b$ and $c < d$ are real numbers, then*

$$\int_c^d N_{[a,b]}(f\,,y)\,\mathrm{d}y = \int_a^b \mathbb{1}_{[c,d]}(f(x))|f'(x)|\,\mathrm{d}x.$$

Therefore, it follows from Lebesgue's differentiation theorem (XXX) that, for every $f \in C^1(R)$ and all bounded intervals $[a\,,b]$,

$$N_{[a,b]}(f\,,y) = \lim_{\varepsilon\downarrow 0}\int_a^b \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon}|f'(x)|\,\mathrm{d}x, \tag{9.17}$$ eq:Kac

for almost every $y \in \mathbb{R}$. It is easy to see that, for every $y \in \mathbb{R}$, the probability measure $A \mapsto \int_A \varepsilon^{-1}\mathbb{1}_{[y,y+\varepsilon]}(a)\,\mathrm{d}a$ converges weakly to the point mass δ_y as $\varepsilon \downarrow 0$. Therefore, we may think of (9.17) as a rigorous interpretation of the "non-rigorous formula,"

$$N_{[a,b]}(f\,,y) = \int_a^b \delta_y(f(x))|f'(x)|\,\mathrm{d}x, \tag{9.18}$$ Banach:N:delta

valid for almost every $y \in \mathbb{R}$.

Simple examples show that (9.17), whence informally also (9.18), cannot possibly hold for every y; consider for example $f \equiv 1$ and $y = 1$. The following corollary of Banach's theorem shows that, nevertheless, (9.17) often holds pointwise for most values of y.

⟨co:Kac⟩ **Corollary 1.13** (Kac XXX)**.** *Choose and fix some $f \in C^1(\mathbb{R})$ and a closed interval $[a\,,b] \subset \mathbb{R}$. If, in addition, $\mathcal{Z}_f(a\,,b) := \{x \in [a\,,b] : f'(x) = 0\}$ is a finite set, then the limit in* (9.17) *holds for every $y \notin \{f(a)\,,f(b)\} \cup \mathcal{Z}_f(a\,,b)$.*

For variants, see Problems XXX.

Proof. Because $\mathcal{Z}_f(a\,,b)$ is finite, we can find a real number $\varepsilon_y > 0$ such that $f'(x) \neq 0$ for all $x \in [y\,,y+\varepsilon_y]$. An elementary argument shows that the jumps of $N_{[a,b]}(f)$ necessarily lie in the finite set $\{f(a)\,,f(b)\} \cup \mathcal{Z}_f(a\,,b)$; Figure 9.1 can serve as a visual aid to formulating such an argument. Instead of proving this we note that, as a result, if $y \notin \{f(a)\,,f(b)\} \cup \mathcal{Z}_f(a\,,b)$ and $\varepsilon \in (0\,,\varepsilon_y)$, then

$$N_{[a,b]}(f\,,x) = N_{[a,b]}(f\,,y) \qquad \text{for all } x \in (y\,,y+\varepsilon_y).$$

Consequently, Corollary 1.12 to Banach's theorem ensures that

$$N_{[a,b]}(f\,,y) = \frac{1}{\varepsilon}\int_y^{y+\varepsilon} N_{[a,b]}(f\,,x)\,\mathrm{d}x = \int_a^b \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon}|f'(x)|\,\mathrm{d}x,$$

for all $\varepsilon \in (0\,,\varepsilon_y)$. Let $\varepsilon \downarrow 0$ to finish. □

The proof of Theorem 1.10 requires a final technical result which we state next.

⟨lem:Kac⟩ **Lemma 1.14.** *If $f : \mathbb{R} \to \mathbb{R}$ is a polynomial of degree at most $q \geqslant 1$, then*

$$\sup_{\varepsilon>0}\sup_{y\in\mathbb{R}}\int_{-\infty}^{\infty} \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon}|f'(x)|\,\mathrm{d}x \leqslant q.$$

Proof. Define $T_0 := -\infty$ and then iteratively let $T_{k+1} := \inf\{x > T_k : f'(x) = 0\}$ for all $k \geqslant 0$, where $\inf \varnothing := \infty$. Define

$$\mathcal{F} := \{j \geqslant 0 : T_j < \infty\},$$

and notice that

$$\int_{-\infty}^{\infty} \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon} |f'(x)| \, \mathrm{d}x = \sum_{j \in \mathcal{F}} \int_{T_j}^{T_{j+1}} \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon} |f'(x)| \, \mathrm{d}x. \tag{9.19}$$ kac:int:sum

The mean-value property of f' ensures that f is strictly monotone in (T_j, T_{j+1}) for every $j \in \mathcal{F}$. Therefore, we may change variables in order to see that

$$\int_{T_j}^{T_{j+1}} \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon} |f'(x)| \, \mathrm{d}x = \left| \int_{f(T_j)}^{f(T_{j+1})} \frac{\mathbb{1}_{[y,y+\varepsilon]}(a)}{\varepsilon} \, \mathrm{d}a \right| \qquad \text{for all } j \in \mathcal{F},$$

where $f(-\infty) := \lim_{a \to -\infty} f(a)$ and $f(\infty) := \lim_{a \to \infty} f(a)$. In particular,

$$\int_{T_j}^{T_{j+1}} \frac{\mathbb{1}_{[y,y+\varepsilon]}(f(x))}{\varepsilon} |f'(x)| \, \mathrm{d}x \leqslant \int_{-\infty}^{\infty} \frac{\mathbb{1}_{[y,y+\varepsilon]}(a)}{\varepsilon} \, \mathrm{d}a = 1,$$

for every $y \in \mathbb{R}$, $\varepsilon > 0$ and $j \in \mathcal{F}$. Because f' is a polynomial of degree $\leqslant q-1$, the fundamental theorem of algebra implies that f' has no more than $q-1$ zeros, including real ones. This in turn implies that the cardinality of $\mathcal{F}$ is at most q. Thus, the lemma follows from (9.19) and the previously-displayed inequality. □

Armed with the preceding, we are in position to prove Kac's theorem.

Proof of Theorem 1.10. Because X' is a.s. a polynomial of degree $q-1$, the random set $\mathcal{Z}_X(a, b)$ has at most $q-1$ elements. Moreover, elementary considerations show that

$$\mathrm{P}\{y \notin \{X_a, X_b\} \cup \mathcal{Z}_X(a, b)\} = 1 \qquad \text{for every } y \in \mathbb{R}.$$

Therefore, Corollary 1.13 implies readily implies that for every $y \in \mathbb{R}$,

$$N_{[a,b]}(X, y) = \lim_{\varepsilon \downarrow 0} \int_a^b \frac{\mathbb{1}_{[y,y+\varepsilon]}(X_t)}{\varepsilon} |X_t'| \, \mathrm{d}t, \tag{9.20}$$ N:Kac

almost surely. Furthermore, Lemma 1.14 and the dominated convergence theorem together imply that (9.20) holds also in $L^p(\mathrm{P})$ for every $p \geqslant 1$. Thus,

$$\begin{aligned}
\mathrm{E}\left[N_{[a,b]}(X, y)\right] &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_a^b \mathrm{E}\left(|X_t'|;\ y < X_t < y + \varepsilon\right) \mathrm{d}t \\
&= \lim_{\varepsilon \downarrow 0} \int_a^b \mathrm{E}\left(|X_t'| \;\middle|\; y < X_t < y + \varepsilon\right) \frac{\mathrm{P}\{y < X_t < y + \varepsilon\}}{\varepsilon} \, \mathrm{d}t.
\end{aligned}$$

Evidently, $\mathrm{P}\{y < X_t < y + \varepsilon\}/\varepsilon$ tends to the probability density of X_t at y as $\varepsilon \downarrow 0$. Because

$$\frac{\mathrm{P}\{y < X_t < y + \varepsilon\}}{\varepsilon} \leqslant \mathrm{P}\{X_t > y\} \leqslant \exp\left(-\frac{y^2}{2\,\mathrm{Var}(X_t)}\right) \qquad \text{for all } y \in \mathbb{R},$$

the preceding and Problem 7 together imply that

$$\mathrm{E}\left[N_{[a,b]}(X, y)\right] = \int_a^b \mathrm{E}\left(|X_t'| \;\middle|\; X_t = y\right) \frac{1}{\sqrt{2\pi\,\mathrm{Var}(X_t)}} \exp\left(-\frac{y^2}{2\,\mathrm{Var}(X_t)}\right) \mathrm{d}t.$$

In particular, we may set $y = 0$ to see that

$$\mathrm{E}\left[N_{[a,b]}(X\,,0)\right] = \int_a^b \frac{\mathrm{E}\left(|X_t'| \mid X_t = 0\right)}{\sqrt{2\pi\,\mathrm{Var}(X_t)}}\,\mathrm{d}t.$$

It is not hard to check directly (see Problem 7 below) that

$$\mathrm{E}\left(|X_t'| \mid X_t = 0\right) = \sqrt{\frac{2}{\pi}\left[\mathrm{Var}(X_t') - \frac{[\mathrm{Cov}(X_t'\,,X_t)]^2}{\mathrm{Var}(X_t)}\right]}.$$

Therefore,

$$\mathrm{E}\left[N_{[a,b]}(X\,,0)\right] = \frac{1}{\pi}\int_a^b \sqrt{\frac{\mathrm{Var}(X_t')\,\mathrm{Var}(X_t) - [\mathrm{Cov}(X_t\,,X_t')]^2}{[\mathrm{Var}(X_t)]^2}}\,\mathrm{d}t, \qquad (9.21)\ \boxed{\texttt{pre:Kac:E(N)}}$$

which yields (9.16) after a direct computation; see Problem 8. □

We conclude this section with a proof of Kac's asymptotic evaluation of the expected number of real zeros of a random polynomial.

Proof of Corollary 1.11. Let $a \downarrow -\infty$ and $b \uparrow \infty$ and appeal to Theorem 1.10 an the monotone convergence theorem in order to see that

$$\mathrm{E}\left[N_{(-\infty,\infty)}(X\,,0)\right] = \frac{2}{\pi}\int_0^\infty \sqrt{\frac{1}{(t^2-1)^2} - \frac{(q+1)^2 t^{2q}}{(t^{2q+2}-1)^2}}\,\mathrm{d}t. \qquad (9.22)\ \boxed{\texttt{eq:Kac0:eq}}$$

Let $\mathcal{F}_q(t)$ denote the integrand of the last integral. Choose and fix a number $\delta \in (0\,,1)$ and observe that

$$\int_0^{1-\delta} \mathcal{F}_q(t)\,\mathrm{d}t + \int_{1+\delta}^\infty \mathcal{F}_q(t)\,\mathrm{d}t \leqslant \int_{1+\delta}^\infty \frac{\mathrm{d}t}{t^2-1} + \int_0^{1-\delta}\frac{\mathrm{d}t}{1-t^2} < \infty. \qquad (9.23)\ \boxed{\texttt{eq:Kac-1:eq}}$$

A direct computation reveals that

$$\int_{1-\delta}^1 \mathcal{F}_q(t)\,\mathrm{d}t = \int_0^{q\delta}\sqrt{\frac{1}{s^2\,|2-(s/q)|^2} - \frac{|1+(1/q)|^2\,|1-(s/q)|^{2q}}{\left(1-|1-(s/q)|^{2q+2}\right)^2}}\,\mathrm{d}s.$$

Let $\mathcal{H}_q(s)$ denote the integrand of the last integral. Because

$$\lim_{q\to\infty}\int_0^1 \mathcal{H}_q(s)\,\mathrm{d}s = \int_0^1\sqrt{\frac{1}{4s^2} - \frac{\mathrm{e}^{-2s}}{(1-\mathrm{e}^{-2s})^2}}\,\mathrm{d}s < \infty,$$

it follows readily from (9.23) that

$$\int_{1-\delta}^1 \mathcal{F}_q(t)\,\mathrm{d}t = O(1) + \int_1^{q\delta}\mathcal{H}_q(s)\,\mathrm{d}s \qquad \text{as } q\to\infty. \qquad (9.24)\ \boxed{\texttt{eq:Kac1:eq}}$$

Elementary computations show that $\mathcal{H}_q(s) = (2s)^{-1} + o(1)$ as $q \to \infty$, uniformly for all $s \in [1\,,q\delta]$. Consequently, as $q \to \infty$,

$$\int_{1-\delta}^1 \mathcal{F}_q(t)\,\mathrm{d}t = \tfrac{1}{2}\log q + o(1), \qquad (9.25)\ \boxed{\texttt{eq:Kac2:eq}}$$

by (9.24). A similar analysis yields $\int_1^{1+\delta}\mathcal{F}_q(t)\,\mathrm{d}t = \frac{1}{2}\log q + o(1)$ as $q \to \infty$. The corollary follows from this, (9.22), (9.23), and (9.25). □

2 Brownian Local Time

One can continue to ask about the indicatrix of Gaussian processes in the case that the process in question does not have C^1 sample trajectories. In this section we highlight aspects of this theory in the special case that the Gaussian process is Brownian motion. While it is possible to study far more general non-C^1 Gaussian processes than Brownian motion, the presentation here is greatly simplified thanks to the independent-increments property of Brownian motion.

§2.1 A First Computation

We can try to apply Banach's theorem and its ramifications by applying it not to Brownian motion directly (since Brownian motion is not continuously differentiable), but to a piecewise-linear Gaussian process which approximates Brownian motion very well. To construct a typical example of such an approximating process, define for every integer $n \geqslant 1$,

$$B_n(t) := (nt - \lfloor nt \rfloor) \left\{ B\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - B\left(\frac{\lfloor nt \rfloor}{n}\right) \right\} + B\left(\frac{\lfloor nt \rfloor}{n}\right), \tag{9.26}$$ def:B_n(t)

for all $t \geqslant 0$, where $B := \{B(t)\}_{t \geqslant 0}$ denotes standard Brownian motion. Equivalently, we define B_n by setting $B_n(j/n) = B(j/n)$ for all integers $j \geqslant 0$, and by then linearly interpolating these values to obtain $B_n(t)$ for other values of $t \geqslant 0$. Figure 9.3 shows a simulation of the graph of a Brownian motion B (in blue), superimposed by the trajectory of the corresponding process B_n for a fixed value of n.

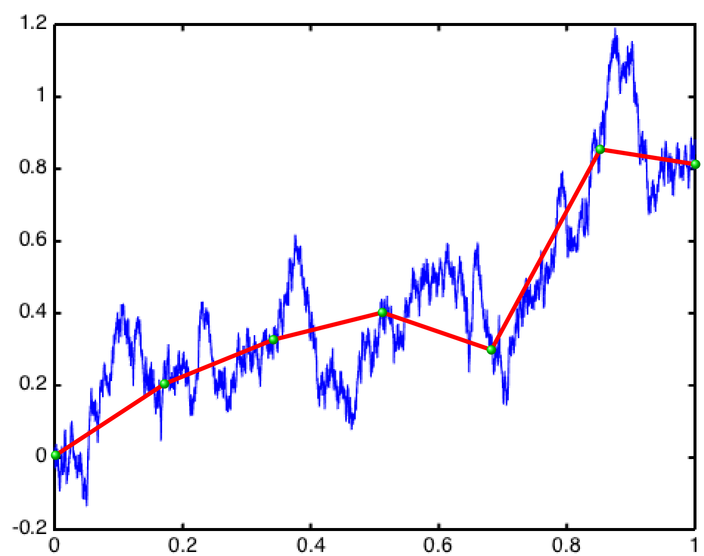

Figure 9.3. The blue line if Brownian motion B; the red line is the linear interpolant B_n. The x-axis denotes time t.

⟨fig:B_n(t)⟩

Since B_n is piecewise linear, its indicatrix is finite. Because B is continuous, it follows that for every $T>0$, $\sup_{t\in[0,T]}|B_n(t)-B(t)|\to 0$ almost surely as $n\to\infty$. Thus, one might hope to define an "indicatrix" for B by studying $\lim_{n\to\infty} N_{[0,T]}(B_n\,,y)$. The following shows that this quest is hopeless because we expect $N_{[0,T]}(B_n\,,0)$ to blow up as $n\to\infty$. This agrees with the following elementary property of Brownian motion $\mathrm{P}\{N_{[0,T]}(B\,,0)=\infty\}=1$ for all $T>0$. For simplicity, we shall study the case that $T=1$.

⟨pr:BM:N⟩ **Proposition 2.1.** *Let B_n be defined as above. Then,*

$$\mathrm{E}\left[N_{[0,1]}(B_n\,,0)\right]\sim\frac{2}{\pi}\sqrt{n}\qquad \textit{as } n\to\infty.$$

The proof requires a simple calculation with Brownian motion.

⟨lem:BM:abc⟩ **Lemma 2.2.** *If $a<b<c$, then*

$$\mathrm{E}\left(|B(c)-B(a)|\mid B(b)=0\right)=\sqrt{\frac{2}{\pi}\left(c-a-\frac{(b-a)^2}{b}\right)}\qquad \textit{a.s.}$$

Proof. Let $\alpha:=(b-a)/b$ and note that, for this choice of α, the random variables $B(c)-B(a)-\alpha B(b)$ and $B(b)$ are uncorrelated and hence independent. It follows from this calculation that

$$\mathrm{E}\left(|B(c)-B(a)\mid B(b)=0\right)=\mathrm{E}\left(|B(c)-B(a)-\alpha B(b)|\right)\qquad\text{a.s.}$$

If X has a mean-zero normal distribution, then an elementary computation shows that $\mathrm{E}(|X|)=\sqrt{(2/\pi)\operatorname{Var}(X)}$. Apply this identity with $X=B(c)-B(a)-\alpha B(b)$ to deduce the result. □

We now return to our immediate goal and prove Proposition 2.1.

Proof of Proposition 2.1. Thanks to the definition (9.26) of the process B_n, we compute directly to find that $\operatorname{Var}[B_n(t)]=n^{-1}(nt-\lfloor nt\rfloor)^2+n^{-1}\lfloor nt\rfloor$. In particular, it follows that

$$\sup_{t>0}\left|\operatorname{Var}\left[B_n(t)\right]-t\right|\leqslant\frac{2}{n}. \tag{9.27}$$ Var(B)-t

Just as was proved in (9.20) for random Gaussian polynomials, one can prove that

$$\begin{aligned}N_{[0,1]}(B_n\,,0)&=\lim_{\varepsilon\downarrow 0}\int_0^1\frac{\mathbb{1}_{[0,\varepsilon]}(B_n(t))}{\varepsilon}|B_n'(t)|\,\mathrm{d}t\\&=2^n\lim_{\varepsilon\downarrow 0}\int_0^1\frac{\mathbb{1}_{[0,\varepsilon]}(B_n(t))}{\varepsilon}\left|B\left(\frac{\lfloor nt\rfloor+1}{n}\right)-B\left(\frac{\lfloor nt\rfloor}{n}\right)\right|\mathrm{d}t.\end{aligned} \tag{9.28}$$ BM:Kac

Therefore, (9.27) and Lemma 2.2 together imply that

$$\begin{aligned}\mathrm{E}\left[N_{[0,1]}(B_n\,,0)\right]&=n\int_0^1\mathrm{E}\left(\left|B\left(\frac{\lfloor nt\rfloor+1}{n}\right)-B\left(\frac{\lfloor nt\rfloor}{n}\right)\right|\,\middle|\,B_n(t)=0\right)\frac{\mathrm{d}t}{\sqrt{2\pi\operatorname{Var}[B_n(t)]}}\\&=\frac{n}{\pi}\int_0^1\sqrt{\frac{1}{n}-\frac{\left(t-\frac{\lfloor nt\rfloor}{n}\right)^2}{t}}\,\frac{\mathrm{d}t}{\sqrt{\operatorname{Var}[B_n(t)]}}\\&\sim\frac{\sqrt{n}}{\pi}\int_0^1\sqrt{1-\frac{\left(t-\frac{\lfloor nt\rfloor}{n}\right)^2}{nt}}\,\frac{\mathrm{d}t}{\sqrt{t}}\qquad\text{as } n\to\infty.\end{aligned}$$

This proves the result readily. □

§2.2 Occupation Densities

We have argued in the previous section that one cannot expect a Brownian motion B to have a finite, well-defined indicatrix-type function in an obvious sense. Still, if B did have a reasonable indicatrix, then it might be also reasonable to imagine that, just as it was true in Banach's theorem,

$$\int_{-\infty}^{\infty} \Psi(y) N_{[0,T]}(B\,,y)\,\mathrm{d}y = \int_0^T \Psi(B(s))|B'(s)|\,\mathrm{d}s \qquad \text{a.s.,} \tag{9.29}$$

eq:BM:Banach:none

for every $T > 0$ and bounded and measurable function $\Psi : \mathbb{R} \to \mathbb{R}$. Proposition 2.1 gives an indication of why the above cannot be true. And, intuitively speaking, the preceding cannot hold because Brownian motion is nowhere-differentiable a.s.; see XXX. See also Problem 13. However, as it turns out, the preceding display has a Brownian-motion analogue, provided that we formally replace the right-hand side by $\int_0^T \Psi(B(s))|B'(s)|^2\,\mathrm{d}s$. There is a sense in which "$|B'(s)|^2\,\mathrm{d}s = \mathrm{d}s$." [This can be formalized by the statement that Brownian motion has finite quadratic variation; see Problem 4 on page 23]. In that case, the proper replacement of the indicatrix of B is an interesting object, called *Brownian local time*, which is the centerpiece of this section.

With the preceding in mind, define for every $T > 0$ and Borel set $A \subset \mathbb{R}$,

$$\mu_T(A) := \int_0^T \mathbb{1}_A(B(s))\,\mathrm{d}s. \tag{9.30}$$

eq:BM:mu_T

Evidently, every μ_T is a.s. a finite Borel measure on $\mathbb{R}$. Because $\mu_T(A)$ is the Lebesgue measure of the time spent in the set A by Brownian motion before time T, μ_T is referred to as the *occupation measure* of Brownian motion on the time interval $[0\,,T]$.

One can define μ_T, equivalently, via the following:

$$\int_0^T \Psi(B(s))\,\mathrm{d}s = \int_{-\infty}^{\infty} \Psi(x)\,\mu_T(\mathrm{d}x), \tag{9.31}$$

eq:BM:ODF:mu

a.s. for every $T > 0$ and bounded and measurable function $\Psi : \mathbb{R} \to \mathbb{R}$. Indeed, (9.31) certainly implies (9.30) upon setting $\Psi := \mathbb{1}_A$. And the converse implication is valid because integration theory teaches us that the general form of (9.31) follows provided that (9.31) is valid for all functions Ψ of the form $\mathbb{1}_A$.

Note that the left-hand side of (9.31) is equal to the right-hand side of (9.29) but with $|B'(s)|\,\mathrm{d}s$ replaced by the intuitively-pleasing expression $|B'(s)|^2\,\mathrm{d}s$ "=" $\mathrm{d}s$. The following is the main result of this section, and interprets the left-hand side of (9.29).

⟨th:BM:ODF⟩ **Theorem 2.3** (Lévy XXX). $\mathrm{P}\{\mu_T \ll \mathrm{Leb}\} = 1$ *for every* $T > 0$, *where* Leb *denotes the standard one-dimensional Lebesgue measure. Write* $L_T(x) := \mu_T(\mathrm{d}x)/\mathrm{d}x$ *for the Radon-Nikodým derivative. Then,* $\mathrm{P}\{L_T \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})\} = 1$ *for every* $T > 0$, *and*

$$\int_0^T \Psi(B(s))\,\mathrm{d}s = \int_{-\infty}^{\infty} \Psi(x) L_T(x)\,\mathrm{d}x, \tag{9.32}$$

eq:BM:ODF

a.s. for every $T > 0$ *and bounded and measurable function* $\Psi : \mathbb{R} \to \mathbb{R}$.

Among other things, Theorem 2.3 — see (9.32) — says that (9.29) becomes a true identity provided that: (1) We remove the formal expression $|B'(s)|$ from the right-hand side; and (2) Replace the indicatrix on the left-hand side by the function L_T. Thus, in a sense, the role of $N_{[0,T]}(B\,,y)$ is played by $L_T(y)$ in the theory of Brownian motion. The random process $y \mapsto L_T(y)$ is called the *local time of B at $y \in \mathbb{R}$ before time T*, and the identity (9.32) is called the *occupation density formula* for the Brownian motion B.

Proof (Kahane XXX). Choose and fix some $T > 0$. Since μ_T, as defined in (9.30), is a bona fide finite measure a.s., it has a well-defined Fourier transform a.s., and the latter can be computed as follows:

$$\widehat{\mu}_T(z) := \int_{-\infty}^{\infty} \mathrm{e}^{izx}\mu_T(\mathrm{d}x) = \int_0^T \mathrm{e}^{izB(s)}\,\mathrm{d}s \qquad \text{a.s. for every } z \in \mathbb{R}.$$

We may apply Fubini's theorem to find that for all $z \in \mathbb{R}$,

$$\begin{aligned}\mathrm{E}\left(|\widehat{\mu}_T(z)|^2\right) &= \int_0^T \mathrm{d}s \int_0^T \mathrm{d}t\ \mathrm{E}\left(\mathrm{e}^{iz\{B(t)-B(s)\}}\right) = \int_0^T \mathrm{d}s \int_0^T \mathrm{d}t\ \mathrm{e}^{-z^2|t-s|/2}\\ &\leqslant T^2 \wedge \frac{4T}{z^2},\end{aligned}$$

after a few computations. A second appeal to Fubini's theorem now yields

$$\mathrm{E}\left(\|\widehat{\mu}_T\|^2_{L^2(\mathbb{R})}\right) \leqslant \int_{-\infty}^{\infty}\left(T^2 \wedge \frac{4T}{z^2}\right)\mathrm{d}z < \infty.$$

In particular, we can conclude from this that $\widehat{\mu}_T \in L^2(\mathbb{R})$ a.s. According to the Plancherel theorem, the Fourier transform is a linear isometry from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$. Consequently, it follows that the random measure μ_T is a.s. a function $L_T \in L^2(\mathbb{R})$ in the sense of distribution theory. Equivalently, there a.s. exists a random element $L_T \in L^2(\mathbb{R}_+)$ such that $\mu_T(\mathrm{d}x) = L_T(x)\,\mathrm{d}x$, as measures. Since $\mu_T(\mathbb{R}) = T$, it follows also that $L_T \in L^1(\mathbb{R})$ a.s., and in fact $\|L_T\|_{L^1(\mathbb{R})} = T$. In particular, we can deduce (9.32) from this and the Radon–Nikodým theorem, and conclude the proof. □

§2.3 Regularity of Local Times

Theorem 2.3 and the Lebesgue density theorem XXX together show that with probability one,

$$L_T(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}\int_0^T \mathbb{1}_{[x,x+\varepsilon]}(B(s))\,\mathrm{d}s, \tag{9.33}$$ eq:BM:LT:approx

for all but a null-set of x's. Now apply Fubini's theorem to see that we can reverse the order of the logical quantifiers in the preceding. That is, there exists a null set of x's off which (9.33) holds a.s. This is the Brownian-motion analogue of the Kac formula (9.17) on page 163.[2]

Define for all $\varepsilon > 0$ and $x \in \mathbb{R}$,

$$\varphi_\varepsilon(x) := \varepsilon^{-1}\mathbb{1}_{[0,\varepsilon]}(x).$$

[2] Also, the intuitive, somewhat-informal statement (9.18) has the following analogue for Brownian motion: $L_T(x) = \int_0^T \delta_x(B(s))\,\mathrm{d}s$.

The quantities that appear in the right-hand side of (9.33) can be expressed as follows:

$$\begin{aligned}\int_0^T \varphi_\varepsilon(x+B(s))\,\mathrm{d}s &= \int_{-\infty}^{\infty} \varphi_\varepsilon(x+y)\,\mu_T(\mathrm{d}y) = \int_{-\infty}^{\infty} \varphi_\varepsilon(x+y) L_T(y)\,\mathrm{d}y \qquad \text{[see (9.32)]}\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-ixz}\widehat{\varphi}_\varepsilon(z)\overline{\widehat{\mu}_T(z)}\,\mathrm{d}z\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}z\int_0^T \mathrm{d}s\ \mathrm{e}^{-ixz-izB(s)}\left(\frac{1-\mathrm{e}^{i\varepsilon z}}{i\varepsilon z}\right).\end{aligned}$$

owing to Parseval's identity. In other words, we have shown that there exists a Lebesgue-null set $\mathcal{N}_T \subset \mathbb{R}$ such that $\lim_{\varepsilon\downarrow 0} L_T^{(\varepsilon)}(x) = L_T(x)$ a.s. for every $x \not\in \mathcal{N}_T$, where

$$\begin{aligned}L_T^{(\varepsilon)}(x) &= \frac{1}{\varepsilon}\int_0^T \mathbb{1}_{[x,x+\varepsilon]}(B(s))\,\mathrm{d}s\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}z\int_0^T\mathrm{d}s\ \mathrm{e}^{-ixz-izB(s)}\left(\frac{1-\mathrm{e}^{i\varepsilon z}}{i\varepsilon z}\right),\end{aligned} \tag{9.34}$$

eq:BM:LT:Fourier

for all $x \in \mathbb{R}$ and $T, \varepsilon > 0$. Since the only key property of L_T is that it is the Radon–Nikodým derivative of μ_T with respect to Lebesgue measure, we can change L_T on a Lebesgue-null set without altering any of the salient properties of L_T, as outlined in Theorem 2.3. Thus, we now redefine

$$L_T(x) := \liminf_{\varepsilon\downarrow 0} L_T^{(\varepsilon)}(x) \qquad \text{for all } T > 0 \text{ and } x \in \mathbb{R}.$$

Let us remark once again that, for this redefined version of Brownian local time, we continue to have, for every $T > 0$ fixed, the a.s. identity $L_T(x)\,\mathrm{d}x = \mu_T(\mathrm{d}x)$ (as measures). Moreover, it still continues to be the case that for every $T > 0$, $L_T(x) = \lim_{\varepsilon\downarrow 0} L_T^{(\varepsilon)}(x)$ a.s. for all $x \not\in \mathcal{N}_T$. An advantage of working with the newly-redefined notion of local times is that they satisfy the following inequality.

⟨pr:BM:L(x)-L(y)⟩ **Proposition 2.4.** *For every $\delta \in (0\,,1/2)$ and $T > 0$ there exists a real number $K_{n,T,\delta}$ such that*

$$\mathrm{E}\left(|L_T(x) - L_T(y)|^{2n}\right) \leqslant K_{n,T,\delta}|x-y|^{2n\delta},$$

simultaneously for all $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$.

Proof. Because of Fatou's lemma, it suffices to prove the proposition with $L_T(x) - L_T(y)$ replaced by $L_T^{(\varepsilon)}(x) - L_T^{(\varepsilon)}(y)$, where $\varepsilon > 0$ is now held fixed. With this aim in mind, choose and fix $T, \varepsilon > 0$, $x, y \in \mathbb{R}$, and $n \in \mathbb{N}$ and write

$$\mathcal{D} := \mathrm{E}\left(\left|L_T^{(\varepsilon)}(x) - L_T^{(\varepsilon)}(y)\right|^{2n}\right),$$

to simplify the typography. Thanks to (9.34),

$$\begin{aligned}\mathcal{D} &= \frac{1}{(2\pi)^{2n}}\int_{\mathbb{R}^{2n}}\mathrm{d}z\int_{[0,T]^{2n}}\mathrm{d}s\ \prod_{\ell=1}^{2n}\left(1-\mathrm{e}^{-i(x-y)z_\ell}\right)\prod_{j=1}^{2n}\left(\frac{1-\mathrm{e}^{i\varepsilon z_j}}{i\varepsilon z_j}\right)\mathrm{E}\left(\prod_{\ell=1}^{2n}\mathrm{e}^{-iyz_\ell - iz_\ell B(s_\ell)}\right)\\ &\leqslant \frac{|x-y|^{2n\delta}}{\pi^{2n}}\int_{\mathbb{R}^{2n}}\prod_{\ell=1}^{2n}|z_\ell|^\delta\,\mathrm{d}z\int_{[0,T]^{2n}}\mathrm{d}s\ \left|\mathrm{E}\left(\prod_{\ell=1}^{2n}\mathrm{e}^{-iz_\ell B(s_\ell)}\right)\right|,\end{aligned}$$

using the elementary fact that $|1-\exp(i\theta)| \leqslant |\theta| \wedge 2|\theta|^\delta$ for all $\theta \in \mathbb{R}$ and $\delta \in (0\,,1)$. This and a simple calculation (see Problem 12 on p. 180) together yield

$$\mathcal{D} \leqslant \frac{(2n)!|x-y|^{2n\delta}}{\pi^{2n}} \int_{\mathbb{R}^{2n}} \prod_{\ell=1}^{2n} |z_\ell|^\delta \, \mathrm{d}z \int_{\Delta_n(T)} \mathrm{d}s \; \mathrm{e}^{-\frac12 \sum_{k=1}^{2n}(s_k-s_{k-1})(z_k+\cdots+z_{2n})^2},$$

where $s_0 := 0$ and

$$\Delta_n(T) := \{s \in \mathbb{R}^{2n} :\ 0 \leqslant s_1 < \cdots < s_{2n} < T\}. \tag{9.35}$$ eq:Delta

Change variables and set $w_{2n+1} := 0$ in order to see that

$$\begin{aligned}\mathcal{D} &\leqslant \frac{(2n)!|x-y|^{2n\delta}}{\pi^{2n}} \int_{\Delta_n(T)} \mathrm{d}s \int_{\mathbb{R}^{2n}} \mathrm{d}w \prod_{\ell=1}^{2n} |w_\ell - w_{\ell+1}|^\delta \mathrm{e}^{-\frac12 \sum_{k=1}^{2n}(s_k-s_{k-1})w_k^2}\\ &\leqslant \frac{(2n)!|x-y|^{2n\delta}}{\pi^{2n}} \int_{\Delta_n(T)} \mathrm{d}s \int_{\mathbb{R}^{2n}} \mathrm{d}w \prod_{\ell=1}^{2n} \left(|w_\ell|^\delta + |w_{\ell+1}|^\delta\right) \mathrm{e}^{-\frac12 \sum_{k=1}^{2n}(s_k-s_{k-1})w_k^2}.\end{aligned}$$

If $a_1,\ldots,a_m \geqslant 1 =: a_{m+1}$, then

$$\prod_{\ell=1}^m (a_\ell + a_{\ell+1}) \leqslant 2^m \prod_{\ell=1}^m \max(a_\ell\,, a_{\ell+1}) \leqslant 2^m \sum_\Gamma \prod_{\ell=1}^m a_{\Gamma(\ell)},$$

where the sum is taken over all mappings Γ that map $\{1\,,\ldots,m\}$ to $\{1\,,\ldots,m+1\}$ such that $\Gamma(\ell) \in \{\ell\,,\ell+1\}$ for all $\ell = 1,\ldots,m$ (remembering all the time that $a_{m+1}=1$). Consequently,

$$\begin{aligned}\mathcal{D} &\leqslant \frac{(2n)!16^n|x-y|^{2n\delta}}{\pi^{2n}} \int_{\Delta_n(T)} \mathrm{d}s \prod_{\ell=1}^{2n} \int_{-\infty}^\infty \mathrm{d}q \left(1+|q|^{2\delta}\right) \mathrm{e}^{-\frac12 (s_k-s_{k-1})q^2}\\ &= \frac{(2n)!32^n A_\delta^{2n}|x-y|^{2n\delta}}{\pi^{2n}} \int_{\Delta_n(T)} \left[1+\left(\frac{2}{s_k-s_{k-1}}\right)^\alpha\right] \frac{\mathrm{d}s}{\sqrt{s_k-s_{k-1}}},\end{aligned}$$

where $A_\delta := \int_{-\infty}^\infty (1+|v|^{2\delta})\exp(-v^2/2)\,\mathrm{d}v$. Because the above integral is finite — see Problem 11 — this completes the proof. □

Proposition 2.4 and the Kolmogorov continuity theorem (Theorem **??**, page **??**) together imply that, for every $T>0$, $x \mapsto L_T(x)$ has a continuous modification. Once again, let us redefine L_T, this time as this continuous version. It still holds that for each fixed $T>0$, $L_T(x)\,\mathrm{d}x = \mu_T(\mathrm{d}x)$ almost surely (as measures). But this modification of our local times has nicer properties than its predecessor. Here is one such property that is worth mentioning: Because the right-hand side of (9.33) is $\varepsilon^{-1}\int_x^{x+\varepsilon} L_T(y)\,\mathrm{d}y$ — see (9.32) — it follows from the continuity of L_T and the fundamental theorem of calculus that, with probability one, (9.33) holds simultaneously for every $x \in \mathbb{R}$ (of course, for the new modification of the local times; notice the strengthened order of the quantifiers!). Next we seek to improve our definition of local times yet again by trying to find a modification of $L_T(x)$ that is continuous in both T and x. To this end, the following will suffice.

⟨pr:BM:L(T)-L(S)⟩ **Proposition 2.5.** *For every integer $m \geqslant 1$ there exists a number $K_m > 0$ such that*

$$\mathrm{E}\left(|L_T(x) - L_S(x)|^m\right) \leqslant K_m(T-S)^{m/2},$$

simultaneously for all $T > S > 0$ and $x \in \mathbb{R}$.

Proof. Thanks to (9.34), for every $x \in \mathbb{R}$, $\varepsilon > 0$, and $T > S > 0$,

$$\mathcal{D} := \mathcal{D}(T\,, S\,, x\,, \varepsilon) := L_T^{(\varepsilon)}(x) - L_S^{(\varepsilon)}(x) = \varepsilon^{-1}\int_S^T \mathbb{1}_{[x,x+\varepsilon]}(B(s))\,\mathrm{d}s.$$

Therefore,

$$\mathrm{E}\left(|\mathcal{D}|^m\right) = \frac{m!}{\varepsilon^m}\int_{0<s_1<\cdots<s_m<T-S} \mathrm{P}\left(\bigcap_{i=1}^m \{B(S+s_i) \in [x\,, x+\varepsilon]\}\right)\mathrm{d}s.$$

If $0 < s_1 < \cdots < s_m < T - S$, then the independence of the increments of B ensures that the above integrand is not more than

$$\mathrm{P}\left\{B(S+s_1) \in [x\,, x+\varepsilon]\right\}\prod_{i=1}^{m-1}\mathrm{P}\left\{B(s_{i+1}-s_i) \in [-\varepsilon\,, \varepsilon]\right\}.$$

Since the probability density function of $B(r)$ is bounded above uniformly by $(2\pi r)^{-1/2}$, it follows that

$$\begin{aligned}
\mathrm{E}\left(|\mathcal{D}|^m\right) &\leqslant \frac{m!}{\varepsilon^m}\int_{0<s_1<\cdots<s_m<T-S}\frac{\varepsilon}{\sqrt{2\pi(S+s_1)}}\prod_{i=1}^{m-1}\frac{2\varepsilon}{\sqrt{2\pi(s_{i+1}-s_i)}}\,\mathrm{d}s\\
&\leqslant \frac{m!2^{(m/2)-1}}{\pi^{m/2}}\int_{0<s_1<\cdots<s_m<T-S}\frac{\mathrm{d}s}{\sqrt{s_1(s_2-s_1)\cdots(s_m-s_{m-1})}}\\
&\leqslant \frac{m!2^{(m/2)-1}}{\pi^{m/2}}\left(\int_0^{T-S}\frac{\mathrm{d}r}{\sqrt{r}}\right)^m,
\end{aligned}$$

which is clearly equal to a finite constant K_m times $(T-S)^{m/2}$. Let $\varepsilon \downarrow 0$ and appeal to Fatou's lemma to finish. □

Owing to Propositions 2.4 and 2.5, the Kolmogorov continuity theorem now implies the following:

⟨th:Trotter⟩ **Theorem 2.6** (Trotter, XXX)**.** *The two-parameter process* $\mathbb{R}_+ \times \mathbb{R} \ni (T\,, x) \mapsto L_T(x)$ *has a continuous modification. In fact, for all* $\alpha, \delta \in (0\,, 1/2)$ *and* $M, T > 0$ *there exists a random variable* $V = V(\alpha\,, \delta\,, M\,, T)$ *such that* $\mathrm{E}(|V|^k) < \infty$ *for all* $k \geqslant 2$*, and with probability one,*

$$|L_t(x) - L_s(y)| \leqslant V\left\{(t-s)^\delta + (y-x)^\alpha\right\},$$

uniformly for all $0 < s < t < T$ *and* $-M < x < y < M$*.*

From now on, when we refer to "Brownian local times," we will mean the jointly-continuous version of the local times that is furnished by Trotter's theorem (Theorem 2.6).

3 The Zero Set of Brownian Motion

It is easy to see that, by virtue of its definition, the indicatrix of a function $f \in C^1$ describes the size of the level sets of f. For example, $N_{[0,T]}(f\,, 0)$ is by default the cardinality of the zero set of f in $[0\,, T]$.

The goal of this section is to prove that the Brownian local times describe, in a similar though more subtle manner, the "size" of the level sets of Brownian motion B. The subtletly comes about since "size" is now measured using Hausdorff dimension – rather than cardinality – which is the topic that we begin with.

§3.1 Hausdorff Dimension

Choose and fix a real number $s \geqslant 0$ and a set $A \subset \mathbb{R}$, and define

$$\mathcal{H}_s^\varepsilon(A) := \inf \sum_{j=1}^{\infty} (2r_j)^s,$$

where the infimum is taken over all countable closed covers $\{[x_j - r_j\,, x_j + r_j]\}_{j=1}^{\infty}$ of A such that $2r_j \leqslant \varepsilon$. Because $\varepsilon \mapsto \mathcal{H}_s^\varepsilon(A)$ is nonincreasing,

$$\mathcal{H}_s(A) := \lim_{\varepsilon \downarrow 0} \mathcal{H}_s^\varepsilon(A)$$

exists. The set function $\mathcal{H}_s$ maps subsets of $\mathbb{R}$ into the extended interval $[0\,, \infty]$.

It is easy to verify the following.

⟨lem:H_s⟩ **Lemma 3.1.** *For every fixed $s > 0$ and $A, B \subset \mathbb{R}$, the set function $\mathcal{H}_s$ satisfies the following:*

1. $\mathcal{H}_s(\varnothing) = 0$;
2. $\mathcal{H}_s(\cup_{i=1}^{\infty} A_i) \leqslant \sum_{i=1}^{\infty} \mathcal{H}_s(A_i)$ *for all* $A_1, A_2, \ldots \subset \mathbb{R}$;
3. $\mathcal{H}_s(A \cup B) = \mathcal{H}_s(A) + \mathcal{H}_s(B)$, *provided additionally that* $A \cap B = \varnothing$;
4. $\mathcal{H}_s(xA + y) = |x|^s \mathcal{H}_s(A)$ *for every* $x, y \in \mathbb{R}$.

Thus we see that $\mathcal{H}_s$ is a Carathéodory outer measure for every $s > 0$. This observation motivates the following.

Definition 3.2. $\mathcal{H}_s$ is called the *s-dimensional Hausdorff measure on* $\mathbb{R}$.

It is, in fact, possible to prove that every $\mathcal{H}_s$ is a bona fide Borel measure on $\mathbb{R}$.

⟨th:H_s⟩ **Theorem 3.3.** *For every fixed $s > 0$, the restriction of $\mathcal{H}_s$ to the Borel subsets of $\mathbb{R}$ is a measure. That is, part 2 of Lemma 3.1 can be extended to the assertion that $\mathcal{H}_s$ is countably additive on the Borel subsets of $\mathbb{R}$.*

The proof uses a classical method, and proceeds by proving that every Borel set is "measurable" for the outer measure $\mathcal{H}_s$. Because we will not need this fact we shall skip the proof and simply refer to XXX for details. Instead, let us observe that whenever $0 < \varepsilon < 1$, $0 < s < t$, and $A \subset \mathbb{R}$, and if $\{[x_j - r_j\,, x_j + r_j]\}_{j=1}^{\infty}$ is a countable closed cover of A such that $2r_j < \varepsilon$ for every $j \geqslant 1$, then

$$\mathcal{H}_s^\varepsilon(A) \leqslant \sum_{j=1}^{\infty} (2r_j)^t \leqslant \varepsilon^{t-s} \sum_{j=1}^{\infty} (2r_j)^s.$$

Optimize over all such closed covers of A, and then send $\varepsilon \downarrow 0$ in order to find that $\mathcal{H}_t(A) = 0$ as soon as $0 < s < t$ and $\mathcal{H}_s(A) < \infty$. Equivalently, if $\mathcal{H}_t(A) > 0$ and $0 < s < t$, then $\mathcal{H}_s(A) = \infty$. In other words,

$$\dim_{\mathrm{H}}(A) := \sup\{s > 0 :\ \mathcal{H}_s(A) = \infty\} = \inf\{s > 0 :\ \mathcal{H}_s(A) = 0\}, \tag{9.36}$$ eq:dimh

for every $A \subset \mathbb{R}$.

Definition 3.4. We call $\dim_{\mathrm{H}}(A)$ the *Hausdorff dimension* of $A \subset \mathbb{R}$.

We can readily deduce many of the properties of Hausdorff dimension from Lemma 3.1 and Theorem 3.3. The following is easy to prove. We include the proof as a way to introduce some of the ideas that are frequently used to compute and/or estimate the Hausdorff dimension of a given set.

⟨lem:dimh⟩ **Lemma 3.5.** *For every* $A, A_1, A_2, \ldots \subset \mathbb{R}$*:*

1. $\dim_{\rm H}(\varnothing) = 0$ *and* $\dim_{\rm H}([0\,,1]) = 1$*;*
2. $\dim_{\rm H}(xA + y) = \dim_{\rm H}(A)$ *for every* $x, y \in \mathbb{R}$*;*
3. $\dim_{\rm H}(A_1 \cap A_2) \leqslant \dim_{\rm H}(A_1)$*;*
4. $\dim_{\rm H}(\cup_{i=1}^{\infty} A_i) = \sup_{i \geqslant 1} \dim_{\rm H}(A_i)$.

Proof. It is particularly easy to prove parts 3 and 2: Because every countable closed cover of A_1 is also a countable closed cover of $A_1 \cap A_2$, it follows that $\mathcal{H}_s^{\varepsilon}(A_1 \cap A_2) \leqslant \mathcal{H}_s^{\varepsilon}(A_1)$ for all $\varepsilon > 0$. Let ε tend to zero to deduce part 3 of the lemma. And part 2 follows from the fact that $\mathcal{H}_s(xA + y) < \infty$ if and only if $\mathcal{H}_s(A) < \infty$; see Lemma 3.1. Next we prove part 1.

Since $\mathcal{H}_s(\varnothing) = 0$, we can see immediately that $\dim_{\rm H}(\varnothing) = 0$. To complete the proof of part 1 it remains to show that $\dim_{\rm H}([0\,,1]) = 1$. With this aim in mind, choose and fix an integer $N > 1$ and let $C_j := [j/N\,, (j+1)/N]$ for all $j = 0, \ldots, N-1$. Since $\{C_j\}_{j=1}^{\infty}$ is a finite closed cover of $[0\,,1]$, it follows that

$$\mathcal{H}_s^{1/N}(A) \leqslant \sum_{j=1}^{N} |C_j|^s = N^{1-s} \qquad \text{for all } s > 0, \tag{9.37}$$ `eq:H:UB`

where $|C_j|$ denotes the Lebesgue measure of C_j. We can let $N \to \infty$ to see that $\mathcal{H}_s([0\,,1]) = 0$—whence $\dim_{\rm H}([0\,,1]) \leqslant s$—whenever $s > 1$. Thus, $\dim_{\rm H}([0\,,1]) \leqslant 1$. We complete the proof of part 1 by verifying that $\dim_{\rm H}([0\,,1]) \geqslant 1$. Choose and fix $\varepsilon \in (0\,,1)$ and let $\{B_j\}_{j=1}^{\infty}$ denote a countable closed cover of $[0\,,1]$ such that $|B_j| \leqslant \varepsilon$ for all $j \geqslant 1$. Then, by the countable subadditivity of Lebesgue's measure,

$$\sum_{j=1}^{\infty} |B_j| \geqslant \left|\cup_{j=1}^{\infty} B_j\right| = |[0\,,1]| = 1. \tag{9.38}$$ `pre:Frostman`

Because the B_js are arbitrary, this proves that $\mathcal{H}_1^{\varepsilon}([0\,,1]) \geqslant 1$ for all $\varepsilon > 0$, and hence $\mathcal{H}_1([0\,,1]) > 0$. It follows from this and the definition of Hausdorff dimension that $\dim_{\rm H}([0\,,1]) \geqslant 1$, which completes the proof of part 1.

We conclude the proof of the lemma by establishing part 4. Before we move in this direction, let us first observe that

$$\dim_{\rm H}(A) \geqslant 0 \qquad \text{for every } A \subset \mathbb{R}, \tag{9.39}$$ `eq:dim>0`

thanks to the already-proven parts 1 and 3 of the lemma. Now we verify part 4 and complete the proof of the lemma.

It remains to demonstrate that

$$\dim_{\rm H}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sup_{i \geqslant 1} \dim_{\rm H}(A_i), \tag{9.40}$$ `eq:dimcup:le:supdim`

since the reverse inequality follows from part 3. If the right-hand side of (9.40) is infinite then there is nothing to prove; see (9.39). Therefore, we may assume without loss in generality that $\sup_{i \geqslant 1} \dim_{\rm H}(A_i) < \infty$. Choose and fix an arbitrary real number

$s > \sup_{i\geqslant 1}\dim_{\rm H}(A_i)$. By (9.36), $\mathcal{H}_s(A_i)=0$ for all $i\geqslant 1$, which in turn implies that $\mathcal{H}_s(\cup_{i=1}^\infty A_i)=0$; see part 2 of Lemma 3.1. Apply (9.36) once more to see that

$$\dim_{\rm H}\left(\bigcup_{i=1}^\infty A_i\right)\leqslant s \qquad \text{for every } s>\sup_{i\geqslant 1}\dim_{\rm H}(A_i),$$

which implies (9.40), whence also completes the proof of the lemma. □

As part of the proof of Lemma 3.5 we learned that a natural way to estimate $\dim_{\rm H}(A)$ from above is to find an "efficient" countable closed cover $\{C_j\}_{j=1}^\infty$ of A such that $|C_j|\leqslant\varepsilon$ for all $j\geqslant 1$, and then use the tautological bound, $\mathcal{H}_s^\varepsilon(A)\leqslant\sum_{j=1}^\infty|C_j|^s$; see (9.37). The more tricky method—see (9.38)—that was used to bound $\dim_{\rm H}(A)$ from below has an extension that is worthy of special mention.

⟨lem:Frostman⟩ **Lemma 3.6** (Frostman XXX)**.** *Suppose $A\subset\mathbb{R}$ is compact and there exists a probability measure μ that is supported in A and satisfies the following for some $s,\eta,c>0$:*

$$\sup_{x\in\mathbb{R}}\mu([x\,,x+\varepsilon])\leqslant c\varepsilon^s \qquad \textit{for all } \varepsilon\in(0\,,\eta).$$

Then, $\dim_{\rm H}(A)\geqslant s$.

Proof. Choose and fix $\varepsilon\in(0\,,\eta)$, and let $\{B_j\}_{j=1}^\infty$ be an otherwise arbitrary countable closed cover of A that satisfies $|B_j|\leqslant\varepsilon$ for all $j\geqslant 1$. Then, by the condition on μ,

$$\sum_{j=1}^\infty|B_j|^s\geqslant c^{-1}\sum_{j=1}^\infty\mu(B_j)\geqslant c^{-1}\mu\left(\bigcup_{j=1}^\infty B_j\right)=c^{-1}\mu(A)=c^{-1}.$$

Compare with (9.38). Optimize over $\{B_j\}_{j=1}^\infty$ to see that $\mathcal{H}_s^\varepsilon(A)\geqslant c^{-1}$, whence also $\mathcal{H}_s(A)\geqslant c^{-1}>0$. □

§3.2 The Ternary Cantor Set

Every $x\in[0\,,1]$ can be written as $x=\sum_{j=1}^\infty x_j3^{-j}$ where the ternary digits $x_1,x_2,\ldots$ of x are 0, 1, or 2. The *ternary Cantor set* $\mathscr{C}$ can be defined as the collection of all points in $[0\,,1]$ whose ternary digits are in $\{0\,,2\}$. To be sure,

$$\mathscr{C}:=\{0\leqslant x\leqslant 1:\ x_j\in\{0\,,2\}\text{ for all } j\geqslant 1\}\,.$$

For every integer $n\geqslant 1$ define

$$\mathscr{C}_n:=\{0\leqslant x\leqslant 1:\ x_j\in\{0\,,2\}\text{ for all } j=1,\ldots,n\}\,.$$

Then clearly $\mathscr{C}_1\supset\mathscr{C}_2\supset\cdots$ and $\mathscr{C}=\cap_{n=1}^\infty\mathscr{C}_n$.

Elementary properties of geometric series imply that $x_1=0$ if and only if $x\in[0\,,1/3]$, and $x_1=2$ if and only if $x\in[2/3\,,1]$. Therefore,

$$\mathscr{C}_1=\left[0\,,\tfrac13\right]\cup\left[\tfrac23\,,1\right].$$

Similarly,

$$\mathscr{C}_2=\left[0\,,\tfrac19\right]\cup\left[\tfrac29\,,\tfrac13\right]\cup\left[\tfrac23\,,\tfrac79\right]\cup\left[\tfrac89\,,1\right],$$

etc. In general, $\mathscr{C}_n$ is the disjoint union of 2^n closed intervals $B_{1,n},\ldots,B_{2^n,n}$, all with length 3^{-n}. Moreover, we can obtain $\mathscr{C}_{n+1}$ from $\mathscr{C}_n$ by dividing every $B_{j,n}$ into three

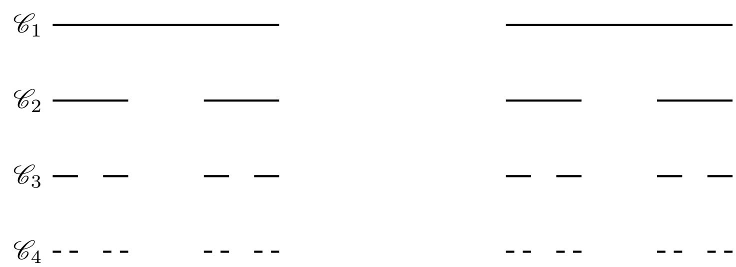

Figure 9.4. The first four stages of construction of $\mathscr{C}$

⟨fig:Cantor⟩

equal parts and jettisoning the middle part (see Figure 9.4). For this reason, $\mathscr{C}$ is also sometimes known as the *middle-thirds Cantor set.*

Because $\{B_{j,n}\}_{j=1}^{2^n}$ is a finite closed cover of $\mathscr{C}$, it follows readily that

$$\mathcal{H}_s^{2^{-n}}(\mathscr{C}) \leqslant \sum_{j=1}^{2^n} |B_{j,n}|^s = (2/3^s)^n .$$

If $s > \log 2/\log 3$, then the preceding tends to zero as $n \to \infty$. This implies that $\mathcal{H}_s(\mathscr{C}) = 0$ for all $s > \log 2/\log 3$, and hence $\dim_{\mathrm{H}}(\mathscr{C}) \leqslant s$. Let $s \downarrow \log 2/\log 3$ to see that $\dim_{\mathrm{H}}(\mathscr{C}) \leqslant \log 2/\log 3$. The following celebrated theorem of Cantor shows that this inequality is in fact an identity.

Theorem 3.7 (Cantor XXX)**.** $\dim_{\mathrm{H}}(\mathscr{C}) = \log 2/\log 3$.

Proof. Let $X_1, X_2, \ldots$ be independent random variables, all defined on a common, suitable, probability space $(\Omega, \mathcal{F}, \mathrm{P})$, such that $\mathrm{P}\{X_j = 0\} = \mathrm{P}\{X_j = 2\} = 1/2$. Then,

$$X := \sum_{j=1}^{\infty} X_j 3^{-j} \tag{9.41}$$ eq:CantorLeb:X

is a well-defined random variables that satisfies $\mathrm{P}\{X \in \mathscr{C}\} = 1$.

Let $\mu := \mathrm{P} \circ X^{-1}$ denote the distribution of X, and recall that $\mathscr{C} = \cap_{n=1}^{\infty} \mathscr{C}_n$ where $\mathscr{C}_n$ is comprised of 2^n closed, disjoint intervals $B_{1,n}, \ldots, B_{2^n,n}$ each with length 3^{-n}. A little thought shows that $\mu(B_{j,n})$ does not depend on j (because the X_js are exchangeable). Since

$$\sum_{j=1}^{2^n} \mu(B_{j,n}) = \mu\left(\bigcup_{j=1}^{2^n} B_{j,n}\right) = \mu(\mathscr{C}_n) = \mu(\mathscr{C}) = 1,$$

this proves that

$$\mu(B_{j,n}) = 2^{-n} \qquad \text{for all } j = 1, \ldots, 2^n \text{ and } n \geqslant 1. \tag{9.42}$$ eq:mu(B)

If $I \subset [0,1]$ is an interval of length 3^{-n}, then $I \cap \mathscr{C}$ intersects at most one $B_{j,n}$, and hence $\mu(I) = \mu(I \cap \mathscr{C}) \leqslant \mu(B_{j,n}) = 2^{-n}$ thanks to (9.42). In other words,

$$\sup_{x \in [0,1]} \mu\left(\left[x, x + 3^{-n}\right]\right) \leqslant 2^{-n} \qquad \text{for all } n \geqslant 1.$$

Every $\varepsilon \in (0\,,1)$ can be sandwiched between 3^{-n-1} and 3^{-n}, where $n = n(\varepsilon) \geqslant 1$ is an integer. Therefore, set $\rho := \log 2/\log 3$ to see that

$$\sup_{x\in[0,1]} \mu\left([x\,,x+\varepsilon]\right) \leqslant \sup_{x\in[0,1]} \mu\left(\left[x\,,x+3^{-n}\right]\right) \leqslant 2^{-n} = 3^{-n\rho} \leqslant (3\varepsilon)^{\rho}.$$

Because this is valid for every $\varepsilon \in (0\,,1)$, and since μ is supported in $\mathscr{C}$, Frostman's lemma (Lemma 3.9) implies that $\dim_{\mathrm{H}}(\mathscr{C}) \geqslant \rho = \log 2/\log 3$. This completes the proof since we established the reverse inequality (that is, $\dim_{\mathrm{H}}(\mathscr{C}) \leqslant \log 2/\log 3$) above the statement of the theorem. □

Define a random variable X by (9.41), and let $F(t) := \mathrm{P}\{X \leqslant t\}$ describe the cumulative distribution function of X. Then, F grows only on the Cantor set $\mathscr{C}$. Because $|\mathscr{C}_n| = (2/3)^n$ and $\mathscr{C} = \cap_{n=1}^{\infty}\mathscr{C}_n$, the Cantor set $\mathscr{C}$ has zero Lebesgue measure. This shows that the monotone function F grows only on a set of zero Lebesgue measure (and Hausdorff dimension $\log 2/\log 3$). The function F is called the *Cantor–Lebesgue function*. It is also known as the *devil's staircase* because of the shape of its graph (see Figure 9.5).

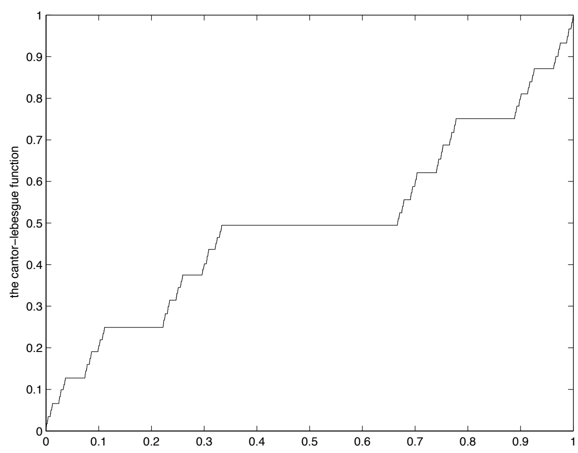

Figure 9.5. A plot of the Cantor–Lebesgue function F. The monotone function F is flat on, and only on, the relative complement $[0\,,1]\setminus\mathscr{C}$ of the Cantor set.

⟨fig:Cantor:Lebesgue⟩

§3.3 Brownian Motion

⟨th:BM:zero:set:Levy⟩ **Theorem 3.8.** *With probability one,* $\dim_{\mathrm{H}} B^{-1}\{0\} = 1/2$ *for every* $j \geqslant 1$.

Recall that the *support* of a locally-finite measure μ on $\mathbb{R}$ is the largest closed set $\mathrm{supp}(\mu)$ such that $\mu(\mathbb{R}\setminus\mathrm{supp}(\mu)) = 0$.

⟨lem:Frostman⟩ **Lemma 3.9** (Frostman's lemma)**.** *Let $A \subset \mathbb{R}$ be fixed. Suppose there exists a probability measure μ such that* $\mu(\overline{A^c}) = 0$

Lemma 3.10. *For all $t > 0$ and* $\eta \in (0\,,1/\mathrm{e})$,

$$\mathrm{P}\left\{B(s) = 0 \textit{ for some } s \in [t\,,t+\eta]\right\} \leqslant \left(\sqrt{\frac{2}{\pi t}} + 4\right)\sqrt{\eta\log(1/\eta)}.$$

Proof. Choose and fix an arbitrary $\eta \in (0\,,1/\mathrm{e})$ and $t > 0$, and observe that for every $\varepsilon > 0$,

$$\begin{aligned}\mathrm{P}\left\{B(s)=0 \text{ for some } s\in[t\,,t+\eta]\right\} &\leqslant \mathrm{P}\left\{|B(t)|\leqslant\varepsilon\right\} + \mathrm{P}\left\{\sup_{r\in[t,t+\eta]}|B(r)-B(t)|>\varepsilon\right\}\\ &\leqslant \frac{2\varepsilon}{\sqrt{2\pi t}} + \mathrm{P}\left\{\sup_{r\in[0,\eta]}|B(r)|>\varepsilon\right\}\end{aligned}$$

The last probability is not greater than $2\mathrm{P}\{\sup_{r\in[0,\eta]} B(r) > \varepsilon\} \leqslant 4\exp\{-\varepsilon^2/(2\eta)\}$, by a standard estimate on the maximum of Brownian motion. Hence,

$$\mathrm{P}\left\{B(s)=0 \text{ for some } s\in[t\,,t+\eta]\right\} \leqslant \inf_{\varepsilon>0}\left[\frac{\varepsilon}{\sqrt{\pi t/2}} + 4\exp\left(-\frac{\varepsilon^2}{2\eta}\right)\right].$$

If we plug in $\varepsilon := \sqrt{\eta\log(1/\eta)}$, then we obtain the upper bound

$$\mathrm{P}\left\{B(s)=0 \text{ for some } s\in[t\,,t+\eta]\right\} \leqslant \frac{\sqrt{\eta\log(1/\eta)}}{\sqrt{\pi t/2}} + 4\sqrt{\eta},$$

which implies the lemma since $\log(1/\eta) > 1$. □

Proof of Theorem 3.8.

$$Z := \{t\in[0\,,1]:\ B(t)=0\}.$$

$$\lambda((s\,,t]) := L_t(0) - L_s(0) \qquad \text{for all } 0<s<t<1.$$

$$\sup_{t\in[0,1]}\lambda((t\,,t+\varepsilon)) = O(\varepsilon^\delta) \qquad \text{as } \varepsilon\downarrow 0, \text{ a.s.}$$

This and Frostman's lemma together prove that $\dim_{\mathrm{H}} Z \geqslant \delta$ a.s. for every $\delta \in (0\,,1/2)$, and hence $\dim_{\mathrm{H}} Z \geqslant 1/2$ a.s.

$$\mathcal{H}_{\delta,2^{-n}}(Z) \leqslant 2^{-\delta n}\sum_{j=0}^{2^n-1}\mathbb{1}_{\{Z\cap[j2^{-n},(j+1)2^{-n}]\neq\varnothing\}}.$$

$$\begin{aligned}\mathrm{E}\left[\mathcal{H}_{\delta,2^{-n}}(Z)\right] &\leqslant 2^{-\delta n}\sum_{j=0}^{2^n-1}\mathrm{P}\left\{Z\cap\left[j2^{-n},(j+1)2^{-n}\right]\neq\varnothing\right\}\\ &= 2^{-\delta n}\sum_{j=0}^{2^n-1}\mathrm{P}\left\{B(t)=0 \text{ for some } t\in\left[j2^{-n},(j+1)2^{-n}\right]\right\}.\end{aligned}$$

$$\mathrm{E}\left[\mathcal{H}_{\delta,2^{-n}}(Z)\right] \leqslant 2^{-\delta n}\sum_{j=0}^{2^n-1}\sum_{k=n}^{\infty}\mathrm{P}\left\{B\left(k2^{-n}\right)B\left((k+1)2^{-n}\right)<0\right\}$$

□

Problems

1. Prove carefully Corollary 1.12. Note that the corollary is valid for *every* $c < d$ and not *almost every* such pair.
2. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is *piecewise* C^1 if f is continuous and for every two real numbers $a < b$ there exist finitely-many points $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ such that the restriction of f to $(x_i\,, x_{i+1})$ is C^1 for every $i = 0, \ldots, n-1$. Prove that Theorem 1.1 and Corollary 1.1 remain valid if f is assumed only to be piecewise C^1.
3. Prove that E_N, as defined by (9.12), is measurable.

?⟨pbm:Xcont->Ccont⟩?

4. Let $(T\,, d)$ be a compact metric space, and suppose $X := \{X_t\}_{t\in T}$ is a mean-zero Gaussian process indexed by T. Note that d is not necessarily the metric that was defined using X in (8.10), p. 134. Prove that if $t \mapsto X_t$ is almost surely continuous, then $(s\,, t) \mapsto \mathrm{Cov}(X_t\,, X_s)$ is a continuous function on $T \times T$.

⟨pbm:LD:stationary⟩

5. Suppose $X := \{X_t\}_{t\in\mathbb{R}}$ is a mean-zero stationary Gaussian process such that $\mathrm{P}\{X \in C^1(\mathbb{R})\} = 1$ and $\mathrm{Var}(X_0) = 1$. Let $\rho(t) := \mathrm{Cov}(X_t\,, X_0)$ for all $t \in \mathbb{R}$ and prove that, for all $T > 1$ and $\delta > 0$,
$$\mathrm{P}\left\{\sup_{t\in[0,T]} X_t > y\right\} \leqslant \frac{T\sqrt{|\rho''(0)|}}{\pi}\mathrm{e}^{-y^2/2} \qquad \text{for every } y > 0.$$
(Hint: If $\sup_{t\in[0,T]} X_t(\omega) > y$ then certainly $N_{[0,T]}(X\,, y)(\omega) \geqslant 1$.)
6. Use the result of Problem 5 in order to prove that:
   (a) $\limsup_{t\to\infty} X_t/\sqrt{2\log t} \leqslant 1$ a.s.; and
   (b) $\lim_{y\to\infty} y^{-2} \log \mathrm{P}\{\sup_{t\in[0,T]} X_t > y\} = -1/2$.

⟨pbm:joint:Gauss:cond⟩

7. Suppose $(X\,, Y)$ has a bivariate normal distribution with $\mathrm{E}(X) = \mathrm{E}(Y) = 0$, $\mathrm{Var}(X) = \sigma_X^2 > 0$, $\mathrm{Var}(Y) = \sigma_Y^2 > 0$, and $\rho = \mathrm{Corr}(X\,, Y) \in (-1\,, 1)$. Prove that
$$\lim_{\varepsilon\downarrow 0} \mathrm{E}\left(|X| \mid 0 < Y < \varepsilon\right) = \mathrm{E}\left(|X| \mid Y = 0\right) = \sqrt{\frac{2}{\pi}\,\sigma_X^2(1-\rho^2)},$$
and that $\sup_{\varepsilon\in(0,1)} \mathrm{E}(|X| \mid 0 < Y < \varepsilon) < \infty$.
(Hint: Start by finding a constant α such that $X - \alpha Y$ is independent of Y.)

⟨pbm:rdm:poly⟩

8. Let $X := \{X_t\}_{t\in\mathbb{R}}$ be the random Gaussian polynomial of (9.1) on page 153, and let $X_t' := \mathrm{d}X_t/\mathrm{d}t$. Prove that $\mathrm{Cov}(X_s'\,, X_t) = (\partial/\partial s)\,\mathrm{Cov}(X_s\,, X_t)$ for all $s, t \in \mathbb{R}$, and similarly that $\mathrm{Cov}(X_s'\,, X_t') = (\partial/\partial t)\,\mathrm{Cov}(X_s'\,, X_t)$.
   (a) Use these computations to verify the following formulas: For all $t \neq 1$,
$$\begin{aligned}
\mathrm{Var}(X_t) &= \frac{t^{2q+2}-1}{t^2-1},\\
\mathrm{Cov}(X_t'\,, X_t) &= \frac{t\left(qt^{2q+2}-(q+1)t^{2q}+1\right)}{(t^2-1)^2},\\
\mathrm{Var}(X_t') &= \frac{t^{2q+2}-t^2-1+t^{2q}\left(qt^2-q-1\right)^2}{(t^2-1)^3}.
\end{aligned}$$
   (b) Use the preceding to verify the details of the computation that begins from (9.21) (see p. 165) and leads to (9.16) (see p. 162).
9. Let B denote a standard Brownian motion, and $\{L_T\}_{T>0}$ its local time process.
   (a) Prove that a.s., $L_T(x) = 0$ for all $|x| > \sup_{t\in[0,T]} |B_t|$ and all $T > 0$.
   (b) Prove that $\mathrm{P}\{\sup_{t\in[0,T]} |B_t| > \lambda\} \leqslant 2\exp\{-\lambda^2/(2T)\}$ for all $\lambda, T > 0$.

(c) Use the preceding to improve Trotter's theorem (Theorem 2.6) to the following: For all $\alpha, \delta \in (0\,,1/2)$ and $T > 0$ there exists a random variable $V = V(\alpha\,,\delta\,,T)$ such that $\mathrm{E}(|V|^k) < \infty$ for all $k \geqslant 2$, and with probability one,

$$|L_t(x) - L_s(y)| \leqslant V\left\{(t-s)^\delta + (y-x)^\alpha\right\},$$

uniformly for all $0 < s < t < T$ and $-\infty < x < y < \infty$.

10. Verify the details of (9.28) (p. 167).

⟨pbm:BM:multiple:int⟩ 11. Choose and fix some $T > 0$. Recall $\Delta_n(T)$ from (9.35) and define, for all $\alpha > 0$,

$$\mathcal{I}_n(\alpha\,,T) := \int_{\Delta_n(T)} \frac{\mathrm{d}s}{s_1^\alpha (s_2 - s_1)^\alpha \cdots (s_{2n} - s_{2n-1})^\alpha}.$$

Prove that $\mathcal{I}_n(\alpha\,,T) < \infty$ if and only if $\alpha < 1$. For a greater challenge, suppose that $0 < \alpha < 1$, and compute $\mathcal{I}_n(\alpha\,,T)$. (Hint: Induction!)

⟨pbm:BM:CHF⟩ 12. Let B denote a standard one-dimensional Brownian motion. Prove that, for all $z_1, \ldots, z_m \in \mathbb{R}$ and $0 =: t_0 < t_1 < \cdots < t_m$,

$$\mathrm{E}\left(\prod_{j=1}^m \mathrm{e}^{iz_j B(t_j)}\right) = \exp\left\{-\frac{1}{2}\sum_{k=1}^m (t_k - t_{k-1})\,(z_k + \cdots + z_m)^2\right\}.$$

⟨pbm:BM:N:0-1⟩ 13. Prove that $\mathrm{P}\{N_{[0,T]}(B\,,y) \in \{0\,,\infty\}\} = 1$ for every $T > 0$ and $y \in \mathbb{R}$, where B denotes standard one-dimensional Brownian motion. You may use, without proof, the *strong Markov property of* B in the following (weak) form. Choose and fix some $a \in \mathbb{R}$, and let $\tau_a := \inf\{s > 0 : B(s) = a\}$. Then, $\tau_a < \infty$ a.s., and $\{B(\tau_a + t) - B(\tau_a)\}_{t \geqslant 0}$ is a standard Brownian motion. See XXX for details.

Bibliography