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

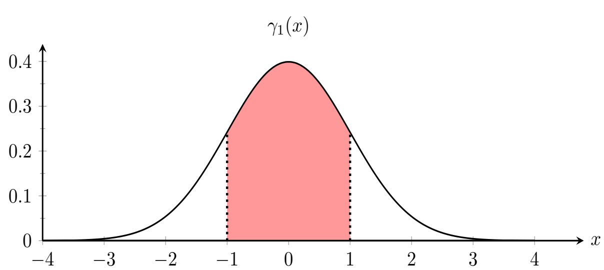

Figure 1.1. Area over $[-1\,,1]$ under the normal curve.

⟨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 $\varepsilon > 0$,

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

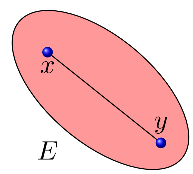

Figure 1.2. A bounded, convex set $E \subset \mathbb{R}^2$: $\overline{xy} \subseteq E \quad \forall x, y \in E$.

⟨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 $B_1, \ldots, B_{3^n}$.

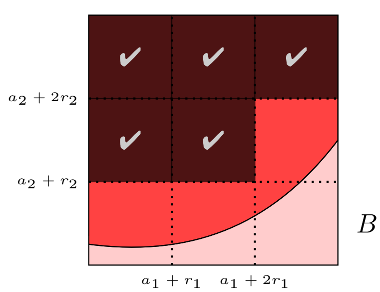

Figure 1.3. A subdivison of B. The region above the curved line belongs to the convex set E. The smaller, darker, checked boxes are those boxes in B that do not intersect ∂E.

⟨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° <u>and</u> 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.

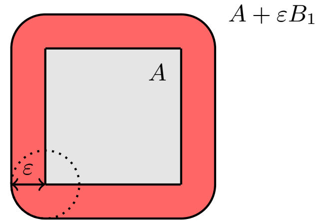

Figure 1.4. The ε-enlargement, $A + \varepsilon B_1$, of the inner square A.

⟨fig:Cvx:Set:2⟩

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 $A + B$.

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):

$$p_X(x) = \frac{1}{(2\pi)^{n/2}\,|\det\Gamma|^{1/2}} \exp\left\{-\tfrac{1}{2}(x-\mu)'\Gamma^{-1}(x-\mu)\right\}. \qquad (1.23)$$ 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 $(X_1, \ldots, X_n)$ and $(-X_1, \ldots, -X_n)$ have the same law, it follows from the parity of n that $m_n = -m_n$, whence

$$\mathrm{E}\left[\prod_{i=1}^n X_i\right] = 0 \quad \text{when } n \text{ is odd.}$$

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

$$\mathrm{E}[X_1 \cdots X_n] = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \, \mathrm{e}^{\frac{1}{2} t' \Gamma t} \bigg\rfloor_{t=0} .$$

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. $\square$

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^2_n 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 $\mathrm{E}(X_i\mid X_0)=\kappa_i X_0$ for every $i=1,\ldots,n$. (Hint: Find κ_i such that $X_i-\kappa_i X_0$ 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 $Y_i:=X_i-\kappa_i X_0$ 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

$$\mathrm{P}\{XY<0\}=\frac12-\frac1\pi\arcsin\rho=\frac1\pi\arccos\rho.$$

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

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

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