Lecture 5

Integration by Parts and Its Applications

The following is an immediate consequence of Theorem 4.7 [page 51] and the chain rule [Lemma 1.7, p. 23].

Theorem 0.1. *For every* $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ *and* $\varphi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$*,*

$$\mathrm{Cov}(f\,,\varphi(f)) = \mathbb{E}\left[(D\varphi)(f) \times \langle Df\,,Df\rangle_{R_1}\right],$$

where $\langle\,\cdot\,,\cdot\rangle_{R_1}$ *was defined in* (4.13)*, page 51.*

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 $\varphi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$ and $X := f(Z)$ where $f \in C_0^1(\mathbb{P}_n)$. Then Theorem 0.1 reduces to the assertion that

$$\mathbb{E}\left[X\varphi(X)\right] = \mathbb{E}(X)\,\mathbb{E}[\varphi(X)] + \mathbb{E}\left[\varphi'(X)\langle DX\,,DX\rangle_{R_1}\right], \tag{5.1}$$

where I am writing $\varphi'(X)$ in place of the more appropriate notation, $(D\varphi)(X)$ for the sake of clarity. In this chapter we explore some of the consequence of these integration by parts results.

1. Concentration of Measure

As a first application of Theorem 0.1 we deduce the concentration of measure property of $\mathbb{P}_n$ 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 37] and

Chebyshev's inequality:

$$\mathrm{P}\left\{|f-\mathrm{E}(f)|>t\right\} \leqslant \frac{[\mathrm{Lip}(f)]^2}{t^2} \qquad \text{for all } t>0.$$

The following is a more precise estimate.

Theorem 1.1. *For every Lipschitz-continuous function $f:\mathbb{R}^n\to\mathbb{R}$,*

$$\mathrm{P}_n\left\{|f-\mathrm{E}f|\geqslant t\right\} \leqslant 2\exp\left(-\frac{t^2}{2[\mathrm{Lip}(f)]^2}\right) \qquad \textit{for all } t>0.$$

There are many proofs of this fact. Perhaps the most elegant is the following, which is due to Houdré et al, XXX.

Proof. 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 22], $f\in\mathbb{D}^{1,2}(\mathrm{P}_n)$ and $\|Df\|\leqslant 1$ a.s. Also, Mehler's formula implies that $\|P_tDf\|\leqslant 1$ for all $t\geqslant 0$, and hence $\|R_1Df\|\leqslant 1$ a.s. [Theorem 2.1, page 45]. Consequently,

$$\langle Df\,,Df\rangle_{R_1}\leqslant \|Df\|\,\|R_1Df\|\leqslant 1 \quad \text{a.s.}$$

Choose and fix some number $\lambda\geqslant 0$. We apply Corollary 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 compute antiderivatives. 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

$$\mathrm{P}_n\{f\geqslant t\}\leqslant \mathrm{e}^{-t^2/2} \qquad [t\geqslant 0].$$

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.

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* $t \geqslant 0$,

$$\mathbb{P}\left\{|M_n - \mathbb{E}(M_n)| \geqslant t\right\} \leqslant 2\exp\left(-\frac{t^2}{2\sigma^2}\right),$$

provided that $\sigma^2 := \max_{1\leqslant i\leqslant n} \mathrm{Var}(X_i) > 0$.

Remark 2.2. Frequently, $\sigma^2 \ll \mathbb{E}(M_n)$ when n is large. When this happens, Theorem 2.1 tells us that, with probability very close to one, $M_n \approx \mathbb{E}(M_n)$. One way to see this is to integrate by parts:

$$\mathrm{Var}(M_n) = 2\int_0^\infty t\mathbb{P}\left\{|M_n - \mathbb{E}(M_n)| > t\right\} \mathrm{d}t \leqslant 4\int_0^\infty t\mathrm{e}^{-t^2/(2\sigma^2)} \mathrm{d}t = 4\sigma^2.$$

[The constant 4 can be removed; see Example 2.9, page 38.] 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 $\mathbb{E}(M_n) = (1 + o(1))\sqrt{2\log n}$ as $n \to \infty$ thanks to Proposition 1.3, page 7. In this case, Borell's inequality yields

$$\mathbb{P}\left\{|M_n - \mathbb{E}(M_n)| \geqslant \sqrt{2\varepsilon \log n}\right\} \leqslant 2n^{-\varepsilon},$$

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 this case,

$$\lim_{n\to\infty} \frac{M_n}{\mathbb{E}(M_n)} = \lim_{n\to\infty} \frac{M_n}{\sqrt{2\log n}} = 1 \qquad \text{a.s.,} \tag{5.2}$$

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 $\mathbb{E}(M_n)$, 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 $Q = A^2$ where A is a symmetric $n \times n$ matrix. Consider the functions

$$f(x) := \max_{1\leqslant i\leqslant n} (Ax)_i \quad \text{and} \quad g(x) := \max_{1\leqslant i\leqslant n} |(Ax)_i| \qquad [x \in \mathbb{R}^n].$$

We have seen already that f and g are both Lipschitz continuous with $\mathrm{Lip}(f), \mathrm{Lip}(g) \leqslant \sigma^2$. Therefore, Theorem 1.1 implies that

$$\mathbb{P}_n \left\{ |f(Z) - \mathbb{E}[f(Z)]| \geqslant t \right\} \leqslant 2 \exp\left(-\frac{t^2}{2\sigma^2} \right) \qquad [t \geqslant 0],$$

and the very same holds also with $g(Z) - \mathbb{E}[g(Z)]$ in place of $f(Z) - \mathbb{E}[f(Z)]$. This proves the result since AZ has the same distribution as X, whence $f(Z) = \max_{i \leqslant n}[AZ]_i$ has the same distribution as $\max_{i \leqslant n} X_i$ and $g(Z)$ likewise has the same distribution as $\max_{i \leqslant n} |X_i|$. □

3. The S-K Model of Spin Glass at High Temperature

Let us pause and discuss an elegant solution of Talgrand 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, I will 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 $\mathbb{P}_n$, $\mathbb{E}_n$, 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 $(\Omega\,, \mathcal{F}\,, \mathbb{P})$.

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. And $1/\sqrt{n}$ is just a normalization factor.[1]

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,

$$\mathrm{P}_n^{(x)}(\sigma_1\,, \ldots, \sigma_n) := \frac{\mathrm{e}^{\beta H_n(\sigma; x)}}{\Pi_n(x)},$$

---

[1]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.

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,

$$\Pi_n(x) := \sum_{\sigma \in \{-1,1\}^n} e^{\beta H_n(\sigma;x)}.$$

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

$$\Pi_n(x) = 2^n \, \mathbb{E}\left[e^{\beta H_n(S;x)}\right],$$

where $S := (S_1, \ldots, S_n)$ for a system of i.i.d. random variables $S_1, \ldots, S_n$ with $\mathbb{P}\{S_1 = 1\} = \mathbb{P}\{S_1 = -1\} = 1/2$.

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, and opposite sign for most pairs with negative interactions. 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 [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 $\exp\{F_\beta(x)n(1+o(1))\}$, when n is large, where $F_\beta(x)$ is a number in $(0\,,\infty)$. Then the number $F_\beta(x)$ is called the *free energy* of the system. A general rule of thumb is that if the free energy exists then its value says some things about the system. In any case, if the free energy exists then it is

$$F_\beta(x) := \lim_{n\to\infty} \frac{1}{n} \log \Pi_n(x).$$

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 the free energy is equal to $\log(2) + \beta^2/4$ "for almost all interaction choices" when $\beta \in (-1\,,1)$. One way to make this precise is to consider $F_\beta(Z)$ where $Z := (Z_{i,j})_{1\leqslant i<j\leqslant n}$ is 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.

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 $n \geqslant 1 + (2L_\beta/\varepsilon)^2$,*

$$\mathbb{P}_n\left\{\left|\frac{\log \Pi_n(Z)}{n-1} - \left(\log 2 + \frac{\beta^2}{4}\right)\right| \leqslant \varepsilon\right\} \geqslant 1 - 2\exp\left(-\frac{\varepsilon^2(n-1)}{2\beta^2}\right).$$

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, $\log \Pi_n(Z) \approx \mathbb{E}[\log \Pi_n(Z)] \approx \log \mathbb{E}[\Pi_n(Z)]$ with high probability. Whereas it is believed that $\mathbb{E}[\log \Pi_n(Z)]$ is a great deal smaller than $\log \mathbb{E}[\Pi_n(Z)]$ when $|\beta| \gg 1$.[2] These approximations are useful for small values of β since $\mathbb{E}[\Pi_n(Z)]$ is easy to compute exactly. In fact,

Lemma 3.2. *For all $\beta \in \mathbb{R}$ and $n \geqslant 2$,*

$$\mathbb{E}\left[\Pi_n(Z)\right] = 2^n \exp\left(\frac{\beta^2(n-1)}{4}\right).$$

Proof. Since $\mathbb{E}\exp(\alpha Z_{1,1}) = \exp(\alpha^2/2)$ for all $\alpha \in \mathbb{R}$,

$$\mathbb{E}[\Pi_n(Z)] = \sum_{\sigma \in \{-1,1\}^n} \exp\left(\frac{\beta^2}{2n} \sum_{1 \leqslant i < j \leqslant n} [\sigma_i \sigma_j]^2\right),$$

which is equal to the expression of the lemma. □

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.[3]

[2]Of course, we always have $\mathbb{E}[\log \Pi_n(Z)] \leqslant \log \mathbb{E}[\Pi_n(Z)]$, by Jensen's inequality.

[3]Lemmas (3.2) and 3.3 teach us that, if $|\beta| < 1$, then

$$\mathbb{E}\left(|\Pi_n(Z)|^2\right) = O\left(|\mathbb{E}[\Pi_n(Z)]|^2\right) \qquad \text{as } n \to \infty.$$

This property fails to hold when $|\beta| \geqslant 1$. Indeed, (5.4) below, and the central limit theorem, together show that, for all $\beta \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{\mathbb{E}\left(|\Pi_n(Z)|^2\right)}{|\mathbb{E}[\Pi_n(Z)]|^2} = \mathbb{E}\left[e^{\beta Z_{1,1} Z_{1,2}}\right] = \mathbb{E}\left[e^{\beta^2 Z_{1,1}^2/2}\right],$$

which is infinite when $|\beta| \geqslant 1$.

Lemma 3.3. *If $-1 < \beta < 1$, then for all $n \geqslant 2$,*

$$\mathbb{E}\left(|\Pi_n(Z)|^2\right) \leqslant \frac{4^n}{\sqrt{1-\beta^2}} \exp\left(\frac{\beta^2(n-2)}{2}\right).$$

Proof. Let us write

$$\begin{aligned}\mathbb{E}\left(|\Pi_n(Z)|^2\right) &= \sum_{\sigma,\sigma'\in\{-1,1\}^n} \mathbb{E}\left[\exp\left(\frac{\beta}{\sqrt{n}} \sum_{1\leqslant i<j\leqslant n} (\sigma_i\sigma_j + \sigma_i'\sigma_j')Z_{i,j}\right)\right] \\ &= \sum_{\sigma,\sigma'\in\{-1,1\}^n} \exp\left(\frac{\beta^2}{2n} \sum_{1\leqslant i<j\leqslant n} [\sigma_i\sigma_j + \sigma_i'\sigma_j']^2\right).\end{aligned}$$

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,

$$\mathbb{E}\left(|\Pi_n(Z)|^2\right) = \mathrm{e}^{\beta^2(n-2)/2} \sum_{\sigma,\sigma'\in\{-1,1\}^n} \exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n \sigma_i\sigma_i'\right]^2\right). \tag{5.3}$$

If we bound $\sigma_i\sigma_i'$ by 1, then we will obtain the bound $\mathbb{E}(|\Pi_n(Z)|^2) \leqslant 4^n \exp\{\beta^2(2n-2)2\}$. The asserted bound of the proposition is a little better when n is large since $\exp\{\beta^2(2n-2)/2\} \gg \exp\{\beta^2(n-2)/2\}$ in that case. In order to deduce the better inequality, we proceed with a little more care.

Let $S_1', \ldots, S_n'$ be i.i.d., all independent of the Z's and the S's, and with $\mathbb{P}\{S_1' = 1\} = \mathbb{P}\{S_1' = -1\} = 1/2$. We can interpret (5.3) as

$$\begin{aligned}\mathbb{E}\left(|\Pi_n(Z)|^2\right) &= 4^n \mathrm{e}^{\beta^2(n-2)/2}\, \mathbb{E}\left[\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n S_iS_i'\right]^2\right)\right] \\ &= 4^n \mathrm{e}^{\beta^2(n-2)/2}\, \mathbb{E}\left[\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n S_i\right]^2\right)\right],\end{aligned}$$

because $S_1S_1', \ldots, S_nS_n'$ are i.i.d. with the same common distribution as S_1. By independence,

$$\mathbb{E}\left[\exp\left(\frac{\beta^2}{2n}\left[\sum_{i=1}^n S_i\right]^2\right)\right] = \mathbb{E}\left[\exp\left(\frac{Z_{1,1}\beta}{\sqrt{n}}\sum_{i=1}^n S_i\right)\right]. \tag{5.4}$$

This Khintchine-type trick of reintroducing the Gaussian variable $Z_{1,1}$ bypasses messy combinatorial arguments. Indeed, for all $z \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left(\frac{z\beta}{\sqrt{n}}\sum_{i=1}^n S_i\right)\right] = \left\{\mathbb{E}\left[\exp\left(\frac{z\beta}{\sqrt{n}}S_1\right)\right]\right\}^n = \{\cosh(z\beta)\}^n \leqslant \mathrm{e}^{nz^2\beta^2/2}.$$

Therefore, we condition on $Z_{1,1}$ first in order to see that the last term in (5.4) is at most $\mathbb{E}[(\beta^2 Z_{1,1}^2/2] = (1-\beta^2)^{-1/2}$, as long as $\beta \in (-1\,,1)$. □

And now we come to the next, very important, step of the proof: Concentration of measure!

Lemma 3.4. *If $|\beta| < 1$, then*

$$\mathrm{P}_n\left\{\left|\frac{\log \Pi_n(Z)}{n-1} - \mathbb{E}\left[\frac{\log \Pi_n(Z)}{n-1}\right]\right| > t\right\} \leqslant 2\exp\left(-\frac{t^2(n-1)}{\beta^2}\right),$$

for all $t > 0$ and $n \geqslant 2$.

Proof. Consider the function $f(x) := \log \Pi_n(x)$ $[x \in \mathbb{R}^n \times \mathbb{R}^n]$. It is easy to see that f is linear, hence Lipschitz continuous. This is because

$$\begin{aligned}\frac{\partial}{\partial x_{i,j}}\Pi_n(x) &= 2^n\frac{\partial}{\partial x_{i,j}}\mathbb{E}\left[\mathrm{e}^{\beta H_n(S;x)}\right] = 2^n\beta\,\mathbb{E}\left[\mathrm{e}^{\beta H_n(S;x)}\cdot\frac{\partial}{\partial x_{i,j}}H_n(S\,;x)\right]\\ &= \frac{2^n\beta\sigma_i\sigma_j}{\sqrt{n}}\mathbb{E}\left[\mathrm{e}^{\beta H_n(S;x)}\right] = \frac{\beta\sigma_i\sigma_j}{\sqrt{n}}\Pi_n(x).\end{aligned}$$

Therefore, $\|(Df)(x)\|^2 = \beta^2 n^{-1}\sum_{1\leqslant i<j\leqslant n}\sigma_i^2\sigma_j^2 = \frac{1}{2}\beta^2(n-1)$. This shows that $\mathrm{Lip}(f) = \beta\sqrt{(n-1)/2}$, and Theorem 1.1 implies the result. □

Now we use the preceding concentration of measure estimate in order to estimate $\mathbb{E}[\log \Pi_n(Z)]$ 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 $\mathbb{E}[\log \Pi_n(Z)] \approx \log \mathbb{E}[\Pi_n(Z)]$.

Lemma 3.5. *For all $\beta \in (-1\,,1)$ there exists $K_\beta < \infty$ such that*

$$\frac{n\log 2}{n-1} + \frac{\beta^2}{4} - \frac{K_\beta}{\sqrt{n-1}} \leqslant \mathbb{E}\left[\frac{\log \Pi_n(Z)}{n-1}\right] \leqslant \frac{n\log 2}{n-1} + \frac{\beta^2}{4},$$

for all $n \geqslant 2$.

Proof. Recall the *Paley–Zygmund inequality* XXX: If $W \geqslant 0$ has two finite moments, then

$$\mathbb{P}\left\{W \geqslant \tfrac{1}{2}\,\mathbb{E}(W)\right\} \geqslant \frac{(\mathbb{E}[W])^2}{4\,\mathbb{E}(W^2)},$$

provided that $\mathbb{E}(W^2) > 0$. The proof is quick:

$$\begin{aligned}\mathbb{E}[W] &= \mathbb{E}\left[W; W \leqslant \tfrac{1}{2}\,\mathbb{E}(W)\right] + \mathbb{E}\left[W; W > \tfrac{1}{2}\,\mathbb{E}(W)\right]\\ &\leqslant \tfrac{1}{2}\,\mathbb{E}(W) + \sqrt{\mathbb{E}(W^2)\mathbb{P}\left\{W > \tfrac{1}{2}\,\mathbb{E}(W)\right\}},\end{aligned}$$

thanks to the Cauchy–Schwarz inequality.

The Paley–Zygmund inequality and Lemmas 3.2 and 3.3 together show us that

$$\mathbb{P}\left\{\log \Pi_n(Z) \geqslant \log\left(\tfrac{1}{2}\,\mathbb{E}\left[\Pi_n(Z)\right]\right)\right\} \geqslant \tfrac{1}{4}\sqrt{1-\beta^2}\,\mathrm{e}^{-\beta^2/2}.$$

If $t := \log(\frac{1}{2}\,\mathbb{E}(\Pi_n(Z))) - \mathbb{E}[\log \Pi_n(Z)] > 0$, then

$$\begin{aligned}\tfrac{1}{4}\sqrt{1-\beta^2}\,\mathrm{e}^{-\beta^2/2} &\leqslant \mathbb{P}\left\{\left|\log \Pi_n(Z) - \mathbb{E}\left[\log \Pi_n(Z)\right]\right| \geqslant t\right\}\\ &\leqslant 2\exp\left(-\frac{t^2}{\beta^2(n-1)}\right),\end{aligned}$$

thanks to Lemma 3.4. Thus,

$$t \leqslant \sqrt{n-1}\left[\frac{\beta^4}{2} + \frac{\beta^2}{2}\left|\log\left(\frac{1-\beta^2}{16}\right)\right|\right]^{1/2} := C_\beta\sqrt{n-1}.$$

And if $t \leqslant 0$ then certainly the preceding holds also. This proves that in any case,

$$\begin{aligned}\mathbb{E}\left[\log \Pi_n(Z)\right] &\geqslant \log \mathbb{E}\left[\Pi_n(Z)\right] - C_\beta\sqrt{n-1} - \log 2\\ &\geqslant \log \mathbb{E}\left[\Pi_n(Z)\right] - \left[C_\beta + \log 2\right]\sqrt{n-1},\end{aligned}$$

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 $\mathbb{E}[\log \Pi_n(Z)] \leqslant \log \mathbb{E}[\Pi_n(Z)]$, owing to Jensen's inequality. □

Proof of Theorem 3.1. Lemma 3.5 ensures that

$$\left|\mathbb{E}\left[\log \Pi_n(Z)\right] - \left(\log 2 + \frac{\beta^2}{4}\right)\right| \leqslant \frac{K_\beta}{\sqrt{n-1}} + \frac{\log 2}{n-1} \leqslant \frac{L_\beta}{2\sqrt{n-1}},$$

where $L_\beta := 2(K_\beta + \log 2)$. Therefore, Lemma 3.4 implies that

$$\mathbb{P}\left\{\left|\frac{\log \Pi_n(Z)}{n-1} - \left(\log 2 + \frac{\beta^2}{4}\right)\right| > t + \frac{L_\beta}{2\sqrt{n-1}}\right\} \leqslant 2\mathrm{e}^{-t^2(n-1)/\beta^2}.$$

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 $\mu_f := \mathbb{P}_n \circ f^{-1}$, defined via

$$\mu_f(A) := \mathbb{P}_n\,\{f \in A\} = \mathbb{P}_n\,\{x \in \mathbb{R}^n : f(x) \in A\} \qquad \text{for all } A \in \mathscr{B}(\mathbb{R}^n).$$

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!].

§1. A Simple Condition for Absolute Continuity. Recently, Nourdin and Peccati XXX have found a necessary and sufficient condition for the law of a random variable $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ 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.

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 $f = \mathbb{E}(f)$ a.s. thanks to Nash's Poincaré inequality [Proposition 2.4, page 36], and $\mu_f = \delta_{\mathbb{E}(f)}$ 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 $\varphi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$ [Example 1.6, page 22]. 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 23] 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 $\mathbb{P}_n\{f \in B\} = 0$. The Radon–Nikodým theorem does the rest. $\square$

§2. The Support of the Law. The Nourdin–Peccati 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 $\text{supp}(\mu_f)$ such that

$$\mu_f(A) = 0 \quad \text{for all Borel sets } A \text{ that do not intersect } \text{supp}(\mu_f).$$

Of course, $\text{supp}(\mu_f)$ is a closed subset of $\mathbb{R}$.

The main goal of this subsection is to verify the following XXX.

Theorem 4.2 (Fang, XXX). *If $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$, then $\text{supp}(\mu_f)$ is an interval.*

The proof hinges on the following elegant zero-one law.

Proposition 4.3 (Sekiguchi and Shiota, XXX). *Let A be a Borel set in $\mathbb{R}^n$. Then, $\mathbb{1}_A \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ iff $\mathbb{P}_n(A) = 0$ or 1.*

Proof. If $\mathbb{1}_A \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ then by the chain rule [Lemma 1.7, page 23],

$$D(\mathbb{1}_A) = D(\mathbb{1}_A^2) = 2\mathbb{1}_A D(\mathbb{1}_A) \qquad \text{a.s.}$$

We consider the function $D(\mathbb{1}_A)$ separately on and off A in order to see that $D(\mathbb{1}_A) = 0$ a.s. Therefore, the Poincaré inequality [Proposition 2.4, page 36] implies that $\text{Var}(\mathbb{1}_A) = 0$ whence $\mathbb{1}_A = \mathbb{P}_n(A)$ a.s. This proves one direction of the theorem. The other direction is trivial: If $\mathbb{P}_n(A) = 0$ or 1 then $\mathbb{1}_A$ is a constant [as an element of $L^2(\mathbb{P}_n)$] whence is in $\mathbb{D}^{1,2}(\mathbb{P}_n)$ since all constants are in $\mathbb{D}^{1,2}(\mathbb{P}_n)$, manifestly. $\square$

Proof of Theorem 4.2. We plan to prove that $\text{supp}(\mu_f)$ is connected. Suppose to the contrary that there exist $-\infty < a < b < \infty$ such that $[a\,,b]$ is not in $\text{supp}(\mu_f)$, yet $\text{supp}(\mu_f)$ intersects both $(-\infty\,,a]$ and $[b\,,\infty)$. For every $\varepsilon \in (0\,,(b-a)/2)$ define

$$\varphi_\varepsilon(w) := \begin{cases} 1 & \text{if } w < a+\varepsilon, \\ \varepsilon^{-1}[-w+a+2\varepsilon] & \text{if } a+\varepsilon \leqslant w \leqslant a+2\varepsilon, \\ 0 & \text{if } w > a+2\varepsilon. \end{cases}$$

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 23] implies that $\varphi_\varepsilon(f) \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ and

$$D(\varphi_\varepsilon \circ f) = \varphi'_\varepsilon(f)D(f) \qquad \text{a.s.,}$$

where I am writing φ'_ε in place of the more precise $D\varphi_\varepsilon$ for typographical convenience.

By construction, $[a\,,b] \not\subset \mathrm{supp}(\mu_f)$ and φ'_ε vanishes [a.s.] off the 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 $\varphi_\varepsilon \circ f$ is a.s. a constant. Therefore, so is $\psi(x) := \mathbb{1}_{\{f \leqslant a\}}(x)$ $[x \in \mathbb{R}^n]$. The zero-one law [Proposition 4.3] then implies that $\mathbb{P}_n\{f \leqslant a\} = 0$ or 1. In particular, $\mathrm{supp}(\mu_f)$ cannot intersect both $(-\infty\,,a]$ and $[b\,,\infty)$. This establishes the desired contradiction. □

§3. The Nourdin–Peccati Formula. We now begin work toward developing the Nourdin–Peccati formula.

Let us first recall a fact about conditional expectations. Let $X : \mathbb{R}^n \to \mathbb{R}$ denote an arbitrary random variable and $Y \in L^1(\mathbb{P}_n)$. Then there exists a Borel-measurable function $G_Y : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}(Y \mid X) = G_{Y|X}(X)$ a.s. In particular, it follows that for every random variable $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ we can find a Borel–measurable function $\mathfrak{S}_f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbb{E}\left[\langle Df\,,Df\rangle_{R_1} \,\middle|\, f\right] = \mathfrak{S}_f(f) \qquad \text{a.s.} \tag{5.5}$$

The recipe is, $\mathfrak{S}_f(x) := G_{\langle Df\,,Df\rangle_{R_1}|f}(x)$.

We may apply the integration-by-parts Theorem 0.1 with $\varphi(w) := w$ in order to that

$$\mathrm{Var}(f) = \mathbb{E}[\mathfrak{S}_f(f)] \qquad \text{for all } f \in \mathbb{D}^{1,2}(\mathbb{P}_n).$$

In other words, $\mathfrak{S}_f(f)$ is an "unbiased estimate" of the variance of f. The following suggests further that $\mathfrak{S}_f$ might be a good "variance estimator."

Lemma 4.4. *If $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ has mean zero, then $\mathfrak{S}_f(f) \geqslant 0$ 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 $\varphi \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ and $D\varphi = \psi$ a.s. Since $\mathbb{E}(f) = 0$, these facts and Theorem 0.1 together imply that

$$\mathbb{E}\left[f\varphi(f)\right] = \mathbb{E}\left[\psi(f)\langle Df\,,Df\rangle_{R_1}\right] = \mathbb{E}\left[\psi(f)\mathfrak{S}_f(f)\right], \tag{5.6}$$

thanks to the tower property of conditional expectations. Since $x\varphi(x) \geqslant 0$ for all $x \in \mathbb{R}$, the left-most term in (5.6) is equal to $\mathbb{E}[f \times \varphi(f)] \geqslant 0$, and hence

$$\mathbb{E}\left[\psi(f) \times \mathfrak{S}_f(f)\right] \geqslant 0,$$

for all bounded and measurable scalar function $\psi \geqslant 0$. Choose and fix $\eta > 0$ and appeal to the preceding with

$$\psi(x) := \mathbb{1}_{(-\infty,-\eta)}(\mathfrak{S}_f(x)) \qquad [x \in \mathbb{R}],$$

in order to see that $\mathrm{P}\{\mathfrak{S}_f(f) \leqslant -\eta\} = 0$ for all $\eta > 0$. This proves the remainder of the proposition. □

Thus we see that $\mathfrak{S}_f(f)$ is always non-negative when $\mathbb{E}(f) = 0$. A remarkable theorem of Nourdin and Peccati XXX asserts that strict inequality holds—that is $\mathfrak{S}_f(f) > 0$ a.s.—if and only if μ_f is absolutely continuous. Moreoever, one can obtain a formula for the probability density of f when $S_f(f) > 0$ a.s. The precise statement follows.

Theorem 4.5 (Nourdin and Peccati, XXX). *Suppose $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ satisfies $\mathbb{E}(f) = 0$. Then $\mu_f(\mathrm{d}x) \ll \mathrm{d}x$ if and only if $\mathfrak{S}_f(f) > 0$ a.s. Moreover, when $\mathfrak{S}_f(f) > 0$ a.s., the probability density function of f is*

$$p_f(x) = \frac{\mathbb{E}(|f|)}{2\mathfrak{S}_f(x)} \exp\left(-\int_0^x \frac{z\,\mathrm{d}z}{\mathfrak{S}_f(z)}\right), \tag{5.7}$$

for almost every $x \in \operatorname{supp}(\mu_f)$.

Remark 4.6. Observe that $\mathbb{P}_n\{\mathfrak{S}_f(f) > 0\} = 1$ iff $\mu_f\{\mathfrak{S}_f > 0\} = 1$.

The proof of Theorem 4.5 is naturally broken into three separate parts, which we record as Propositions 4.7 through 4.9 below.

Proposition 4.7. *Let f be a mean-zero random variable in $\mathbb{D}^{1,2}(\mathbb{P}_n)$. If $\mathfrak{S}_f(f) > 0$ 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 $\varphi \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ and $D\varphi = \mathbb{1}_B$ a.s. The integration-by-parts Theorem 0.1 implies that

$$\mathbb{E}\left[f \int_0^f \mathbb{1}_B(y)\,\mathrm{d}y\right] = \mathbb{E}\left[1_B(f) \times \mathfrak{S}_f(f)\right].$$

If B were Lebesgue null, then $\int_0^f \mathbb{1}_B(y)\,\mathrm{d}y = 0$ a.s., and hence $\mathbb{E}[\mathbb{1}_B(f) \times \mathfrak{S}_f(f)] = 0$. Because we have assumed that $\mathfrak{S}_f(f) > 0$ a.s., it follows that $\mathbb{1}_B(f) = 0$ a.s., equivalently, $\mathbb{P}_n\{f \in B\} = 0$. The Radon–Nikodỳm theorem does the rest. □

Proposition 4.8. *If $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ has mean zero and satisfies $\mu_f(\mathrm{d}x) \ll \mathrm{d}x$, then $\mathfrak{S}_f(f) > 0$ a.s.*

Proof. Let p_f denote the probability density of f; that is, $\int_B p_f(x)\,\mathrm{d}x = \mathbb{P}_n\{f \in B\}$ for all Borel sets $B \subset \mathbb{R}$.

If $\psi \in C_c(\mathbb{R})$, then $\varphi(x) := \int_{-\infty}^x \psi(y)\,\mathrm{d}y[x \in \mathbb{R}]$ is bounded and $\varphi'(x) = \psi(x)$. We may integrate by parts in Gauss space [Theorem 0.1] in order to see that

$$\mathbb{E}\left[\psi(f) \times \mathfrak{S}_f(f)\right] = \mathbb{E}\left[f \times \varphi(f)\right] = \int_{-\infty}^{\infty} y\varphi(y)p_f(y)\,\mathrm{d}y.$$

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

$$\mathbb{E}\left[\psi(f) \times \mathfrak{S}_f(f)\right] = \int_{-\infty}^{\infty} \psi(y)\left(\int_y^{\infty} zp_f(z)\,\mathrm{d}z\right)\mathrm{d}y. \tag{5.8}$$

Now $\mathbb{P}_n\{p_f(f) = 0\} = \mu_f\{p_f = 0\} = \int_{\{p_f=0\}} p_f(a)\,\mathrm{d}a = 0$. Therefore, we can rewrite (5.8) as

$$\mathbb{E}\left[\psi(f) \times \mathfrak{S}_f(f)\right] = \mathbb{E}\left[\psi(f) \times \frac{\int_f^{\infty} zp_f(z)\,\mathrm{d}z}{p_f(f)}\right],$$

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,

$$\mathfrak{S}_f(f) = \frac{\int_f^{\infty} zp_f(z)\,\mathrm{d}z}{p_f(f)} \qquad \text{a.s.} \tag{5.9}$$

It remains to prove that

$$\int_f^{\infty} zp_f(z)\,\mathrm{d}z > 0 \qquad \text{a.s.} \tag{5.10}$$

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

$$\Phi(x) := \int_x^{\infty} zp_f(z)\,\mathrm{d}z \qquad [x \in \mathbb{R}]. \tag{5.11}$$

Since p_f is supported in $[\alpha\,,\beta]$, Φ is constant off $[\alpha\,,\beta]$ and $\Phi(\alpha+) = \Phi(\beta-) = 0$. 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 (5.10) and completes the proof. □

Proposition 4.9. *If $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ has mean zero and $\mathfrak{S}_f(f) > 0$ a.s., then the density function of f is given by* (5.7).

Proof. Recall the function Φ from (5.11). Then Φ is almost-everywhere differentiable [Lebesgue's theorem] with density $\Phi'(x) = -xp_f(x)$ a.e. At the same time, (5.9) implies that

$$\Phi(x) = \mathfrak{S}_f(x)p_f(x) \qquad \text{for almost all } x \in \text{supp}(\mu_f). \tag{5.12}$$

It follows that

$$\frac{\Phi'(x)}{\Phi(x)} = -\frac{x}{\mathfrak{S}_f(x)} \qquad \text{for almost all } x \in \text{supp}(\mu_f).$$

Since $0 \in \text{supp}(\mu_f)$, we integrate the preceding to obtain

$$\Phi(x) = \Phi(0)\exp\left(-\int_0^x \frac{z\,\mathrm{d}z}{\mathfrak{S}_f(z)}\right) \qquad \text{for all } x \in \text{supp}(\mu_f). \tag{5.13}$$

But $\Phi(0) = \int_0^\infty zp_f(z)\,\mathrm{d}z = \mathbb{E}(f\,;f > 0) = \frac{1}{2}\mathbb{E}(|f|)$, because $\mathbb{E}(f) = 0$. Therefore, (5.12) implies the theorem. □

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 51] that lie at the very heart of Gaussian analysis. Theorem 4.5 is only one such example. I will next decscribe a few other examples. Their monograph XXX contains a host of others.

§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 $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ has a mean-zero normal distribution.

Theorem 5.1 (Nourdin and Peccati, XXX). *Suppose $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ satisfies $\mathbb{E}(f) = 0$. Then the random variable f has a normal distribution iff $\mathfrak{S}_f(f)$ is a constant a.s.*

Remark 5.2. The constancy condition on $\mathfrak{S}_f(f)$ is equivalent to the condition that

$$\mathfrak{S}_f(f) = \mathbb{E}[\mathfrak{S}_f(f)] = \mathbb{E}\left[\langle Df\,, Df\rangle_{R_1}\right] = \text{Var}(f) \quad \text{a.s.},$$

thanks to Theorem 0.1. Therefore, Theorem 5.1 is saying that f is a normally distributed if and only if its variance estimator $\mathfrak{S}_f(f)$ is exact.

Remark 5.3. For a stronger result see Example 5.10 below.

The proof of Theorem 5.1 rests on the following "heat kernel estimate," which is interesting in its own right.

Theorem 5.4. *Suppose $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ has mean zero, and there exists a constant $\sigma > 0$ such that $\mathfrak{S}_f(f) \geqslant \sigma^2$ a.s. Then $\mathrm{supp}(\mu_f) = \mathbb{R}$, and*

$$p_f(x) \geqslant \frac{\mathbb{E}(|f|)}{2\mathfrak{S}_f(x)} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{5.14}$$

for almost every $x \in \mathbb{R}$. Suppose, in addition, that there exists a constant $\Sigma < \infty$ such that

$$\mathfrak{S}_f(f) \leqslant \Sigma^2 \qquad a.s. \tag{5.15}$$

Then, for almost every $x \in \mathbb{R}$,

$$\frac{\mathbb{E}(|f|)}{2\sigma^2} \exp\left(-\frac{x^2}{2\Sigma^2}\right) \geqslant p_f(x) \geqslant \frac{\mathbb{E}(|f|)}{2\Sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{5.16}$$

Remark 5.5. If f is Lipschitz continuous, then we have seen that $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ [Example 1.6, p. 22]. 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 45]. Thus, whenever f is Lipschitz continuous, condition (5.15) holds with $\Sigma := \mathrm{Lip}(f)$.

Proof. Recall Φ from (5.11). According to (5.13),

$$\Phi(x) \geqslant \frac{1}{2}\,\mathbb{E}(|f|) \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad \text{for all } x \in \mathrm{supp}(\mu_f). \tag{5.17}$$

It follows from the fact that $\mathbb{E}(f) = 0$ that $\Phi(x) \to 0$ as x tends to the boundary of $\mathrm{supp}(\mu_f)$. Since $\mathrm{supp}(\mu_f)$ is an interval [Theorem 4.2, page 63], (5.17) shows that $\mathrm{supp}(\mu_f)$ must be unbounded. This proves that $\mathrm{supp}(\mu_f) = \mathbb{R}$. The inequality (5.14) follows from (5.17), and (5.16) follows from (5.14) readily. $\square$

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, (5.9) 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 4.5 and 5.4. $\square$

We highlight some of the scope, as well as some of the limitations, of Theorem 4.5 by studying two elementary special cases.

Example 5.6 (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 $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ and $Df = a$ a.s. Moreover, $\mathbb{E}(f) = a \cdot \mathbb{E}(Z) = 0$ and $\mathrm{Var}(f) = \|a\|^2 > 0$. Furthermore, Mehler's formula [Theorem 2.1, page 45] 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 $\mathfrak{S}_f(f) = \|a\|^2$ 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 5.7 (A Quadratic Example)**.** Consider the random variable $f(x) := \|x\|^2 - n$ $[x \in \mathbb{R}^n]$. Equivalently, $f = \|Z\|^2 - \mathbb{E}(\|Z\|^2)$. Clearly, $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$—in fact, $f \in C_0^\infty(\mathbb{P}_n)$—and $\mathbb{E}(f) = 0$ and $(Df)(x) = 2x$. Mehler's formula [Theorem 2.1, page 45] 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 $\mathfrak{S}_f(f) = \mathbb{E}(2f + 2n \,|\, f) = 2f + 2n$ a.s. Equivalently, $\mathfrak{S}_f(z) = 2(z + n)$ a.s. for all $z \in \mathrm{supp}(\mu_f)$. Because $\mathbb{P}\{\mathfrak{S}_f = 0\} = \mathbb{P}\{Z = 0\} = 0$, Theorem 4.5 reduces to the statement that $\|Z\|^2 - n$ has a probability density $p_{\|Z\|^2-n}$, and

$$p_{\|Z\|^2-n}(x) = \frac{\mathbb{E}\left(\left|\|Z\|^2 - n\right|\right)}{4(x+n)} \exp\left(-\frac{1}{2}\int_0^x \frac{z}{z+n}\,\mathrm{d}z\right) = \frac{\mathbb{E}\left(\left|\|Z\|^2 - n\right|\right)\mathrm{e}^{-x/2}}{4n^{n/2}(x+n)^{1-(n/2)}},$$

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 4.5 is consistent with the well-known fact that $\|Z\|^2$ has a χ_n^2 distribution.

§2. Distance to Normality. Suppose $f \in \mathbb{D}^{1,2}(\mathbb{P}_n)$. 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 $X_i \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ 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 close to the distribution of Z. This is indeed the case, as is shown by the theory of Nourdin, Peccati, and Reinert, XXX. I will work out the details first in the case that $Q = I$ is the $n \times n$ identity matrix.

Theorem 5.8 (Nourdin, Peccati, and Reinert, XXX)**.** *Consider a random vector $X = (X_1, \ldots, X_n)$, where $X_i \in \mathbb{D}^{1,2}(\mathbb{P}_n)$, and $\mathbb{E}(X_i) = 0$. Then, for* *every* $\Phi \in \mathbb{D}^{2,2}(\mathbb{P}_n)$,

$$\mathrm{E}\left[\Phi(X)\right] - \mathrm{E}\left[\Phi(Z)\right] = \mathrm{E}\left[\sum_{i,j=1}^{n}\left(R_2 D^2_{i,j}\Phi\right)(X) \times \left(I_{i,j} - \langle DX_i\,, DX_j\rangle_{R_1}\right)\right],$$

where R_2 *denotes the 2-potential of the OU semigroup. In particular,*

$$\begin{aligned}&|\mathrm{E}[\Phi(X)] - \mathrm{E}[\Phi(Z)]|\\ &\qquad\leqslant \tfrac{1}{2}\sup_{x\in\mathbb{R}^n}\max_{1\leqslant i,j\leqslant n}\left|(D^2_{i,j}\Phi)(x)\right| \cdot \sum_{i,j=1}^{n}\mathrm{E}\left(\left|I_{i,j} - \langle DX_i\,, DX_j\rangle_{R_1}\right|\right),\end{aligned}$$

for all $\Phi : \mathbb{R}^n \to \mathbb{R}$ *that are bounded and have continuous and bounded mixed partial derivatives of order* $\leqslant 2$.

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) Both $\bar{X}$ and $\bar{Z}$ are n-dimensional random vectors on the Gauss space $(\mathbb{R}^{2n}, \mathscr{B}(\mathbb{R}^{2n})\,, \mathbb{P}_{2n})$;

(2) The $\mathbb{P}_{2n}$-distribution of $\bar{X}$ is the same as the $\mathbb{P}_n$-distribution of X; and

(3) The $\mathbb{P}_{2n}$-distribution of $\bar{Z}$ is the same as the $\mathbb{P}_n$-distribution of Z.

In this way, Theorem 5.8 can be restated as follows:

$$\begin{aligned}&\mathrm{E}_{2n}\left[\Phi(\bar{X})\right] - \mathrm{E}_{2n}\left[\Phi(\bar{Z})\right]\\ &\qquad= \mathrm{E}_{2n}\left[\sum_{i,j=1}^{n}\left(R_2 D^2_{i,j}\Phi\right)(\bar{X}) \times \left(I_{i,j} - \langle D\bar{X}_i\,, D\bar{X}_j\rangle_{R_1}\right)\right],\end{aligned}\tag{5.18}$$

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 XXX. Note that with $\mathbb{P}_{2n}$-probability one: $(P_0\Phi)(\bar X) = \Phi(\bar X)$; and $(P_t\Phi)(\bar X) \to \mathbb{E}_{2n}[\Phi(\bar Z)]$ as $t \to \infty$. Therefore, $\mathbb{P}_{2n}$-a.s.,

$$\begin{aligned}\Phi(\bar X) - \mathbb{E}_{2n}[\Phi(\bar Z)] &= -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}(P_t\Phi)(\bar X)\,\mathrm{d}t \\ &= -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}_{2n}\left[\Phi\left(\mathrm{e}^{-t}\bar X + \sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\,\middle|\,\bar X\right]\mathrm{d}t,\end{aligned} \tag{5.19}$$

owing to Mehler's formula [Theorem 2.1, page 45]. We take expectations of both sides and apply the dominated convergence theorem, to interchange the derivative with the expectation, in order to find that

$$\begin{aligned}&\mathbb{E}_{2n}[\Phi(\bar X)] - \mathbb{E}_{2n}[\Phi(\bar Z)] \\ &= -\int_0^\infty \mathbb{E}_{2n}\left[\frac{\mathrm{d}}{\mathrm{d}t}\Phi\left(\mathrm{e}^{-t}\bar X + \sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\right]\mathrm{d}t \\ &= -\sum_{i=1}^n \int_0^\infty \mathbb{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar X + \sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\left[-\mathrm{e}^{-t}\bar X_i + \frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\bar Z_i\right]\right\}\mathrm{d}t.\end{aligned} \tag{5.20}$$

Since $\mathbb{E}_{2n}(\bar Z_i) = 0$, Theorem 4.7 implies that for all $G \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ and $1 \leqslant i \leqslant n$,

$$\mathbb{E}_{2n}\left[G(\bar Z)\bar Z_i\right] = \mathbb{E}_{2n}\left[\langle D(G\circ\bar Z)\,,D\bar Z_i\rangle_{R_1}\right] = \mathbb{E}_{2n}\left[D_i(G\circ\bar Z)\right] = \mathbb{E}_{2n}\left[(D_iG)(\bar Z)\right].$$

Therefore, for every $x \in \mathbb{R}^n$ and $1 \leqslant i \leqslant n$,

$$\begin{aligned}&\mathbb{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\bar Z_i\right\} \\ &\qquad = \frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\mathbb{E}_n\left\{D_i\left[(D_i\Phi)\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\,\bullet\right)\right](\bar Z)\right\} \\ &\qquad = \mathrm{e}^{-2t}\,\mathbb{E}_{2n}\left[(D^2_{i,i}\Phi)\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\right] \\ &\qquad = \mathrm{e}^{-2t}\left(P_tD^2_{i,i}\Phi\right)(x),\end{aligned}$$

thanks first to the chain rule [Lemma 1.7, page 23], and then Mehler's formula [Theorem 2.1, page 45]. Since $\bar X$ and $\bar Z$ are independent, we can first condition on $\bar X = x$ and then integrate $[\mathrm{d}(\mathbb{P}_n\circ\bar X^{-1})]$ to deduce from the preceding that

$$\begin{aligned}&\mathbb{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar X + \sqrt{1-\mathrm{e}^{-2t}}\,\bar Z\right)\frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,\bar Z_i\right\} \\ &\qquad = \mathrm{e}^{-2t}\sum_{j=1}^n \mathbb{E}_{2n}\left[\left(P_tD^2_{i,j}\Phi\right)(\bar X)I_{i,j}\right].\end{aligned} \tag{5.21}$$

Similarly, because $\mathbb{E}_{2n}(\bar{X}_i) = 0$ for all $1 \leqslant i \leqslant n$, we can write

$$\begin{aligned}\mathbb{E}_{2n}\left[G(\bar{X})\bar{X}_i\right] &= \mathbb{E}_{2n}\left[\left\langle D(G\circ\bar{X})\,,D\bar{X}_i\right\rangle_{R_1}\right] = \sum_{k=1}^{2n}\mathbb{E}_{2n}\left[D_k(G\circ\bar{X})\times\left(R_1D_k\bar{X}_i\right)\right]\\ &= \sum_{j=1}^{n}\sum_{k=1}^{2n}\mathbb{E}_{2n}\left[(D_jG)(\bar{X})\times D_k(X_j)\times\left(R_1D_k\bar{X}_i\right)\right]\\ &= \sum_{j=1}^{n}\mathbb{E}_{2n}\left[(D_jG)(\bar{X})\times\left\langle D\bar{X}_j\,,D\bar{X}_i\right\rangle_{R_1}\right],\end{aligned}$$

by the chain rule, and hence

$$\begin{aligned}&\mathbb{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar{X}+\sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\mathrm{e}^{-2t}\bar{X}_i\right\}\\ &\quad= \mathrm{e}^{-2t}\sum_{j=1}^{n}\mathbb{E}_{2n}\left[(D^2_{i,j}\Phi)\left(\mathrm{e}^{-t}\bar{X}+\sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\left\langle D\bar{X}_j\,,D\bar{X}_i\right\rangle_{R_1}\right]\\ &\quad= \mathrm{e}^{-2t}\sum_{j=1}^{n}\mathbb{E}_{n}\left[\left(P_tD^2_{i,j}\Phi\right)(\bar{X})\left\langle D\bar{X}_j\,,D\bar{X}_i\right\rangle_{R_1}\right].\end{aligned}\tag{5.22}$$

We now merely combine (5.20), (5.21), and (5.22) in order to deduce (5.18) and hence the theorem. □

Theorem 5.8 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 $\mathbb{E}[\Phi(X)] - \mathbb{E}[\Phi(Q^{1/2}Z)]$, 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

$$(P^Q_tf)(x) := \mathbb{E}\left[f\left(\mathrm{e}^{-t}x+\sqrt{1-\mathrm{e}^{-2t}}\,Q^{1/2}Z\right)\right].$$

It is not hard to check that the preceding defines a semigroup $\{P^Q_t\}_{t\geqslant 0}$ of linear operators that solve a heat equation of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}P^Q_t = \mathcal{L}^Q P^Q_t \text{ for } t > 0,$$

subject to P^Q_0 = 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 $R^Q_\lambda := \int_0^\infty \exp(-\lambda t)P^Q_t\,\mathrm{d}t$, etc. Now we begin with the following variation on (5.20): Define

$$\Psi(t) := \mathbb{E}_{2n}\left[\Phi\left(\mathrm{e}^{-t}\bar{X}+\sqrt{1-\mathrm{e}^{-2t}}\,Q^{1/2}\bar{Z}\right)\right] = \mathbb{E}\left[\left(P^Q_t\Phi\right)(X)\right],$$

and notice that $\Psi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$, $\Psi(0) = \mathbb{E}_{2n}[\Phi(\bar{X})]$, and $\lim_{t\to\infty}\Psi(t) = \mathbb{E}_{2n}[\Phi(Q^{1/2}\bar{Z})]$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{2n}[\Phi(\bar{X})] - \mathbb{E}_{2n}[\Phi(Q^{1/2}\bar{Z})] \\
&= \lim_{t\to\infty}\Psi(t) - \Psi(0) = \int_0^\infty \Psi'(t)\,\mathrm{d}t \\
&= \sum_{i=1}^n \int_0^\infty \mathbb{E}_{2n}\left\{(D_i\Phi)\left(\mathrm{e}^{-t}\bar{X} + \sqrt{1-\mathrm{e}^{-2t}}\,\bar{Z}\right)\left[-\mathrm{e}^{-t}\bar{X}_i + \frac{\mathrm{e}^{-2t}}{\sqrt{1-\mathrm{e}^{-2t}}}\,Q^{1/2}\bar{Z}_i\right]\right\}\mathrm{d}t.
\end{aligned}$$

Now we translate the proof of Theorem 5.8 in order to obtain the following important generalization.

Theorem 5.9 (Nourdin, Peccati, and Reinert, XXX). *Consider a random vector $X = (X_1, \ldots, X_n)$, where $X_i \in \mathbb{D}^{1,2}(\mathbb{P}_n)$, and $\mathbb{E}(X_i) = 0$. Then, for every $\Phi \in \mathbb{D}^{2,2}(\mathbb{P}_n)$ and for all $n \times n$ covariance matrices Q,*

$$\mathbb{E}\left[\Phi(X)\right] - \mathbb{E}\left[\Phi(Q^{1/2}Z)\right] = \mathbb{E}\left[\sum_{i,j=1}^n \left(R_2^Q D_{i,j}^2\Phi\right)(X) \times \left(Q_{i,j} - \langle DX_i\,, DX_j\rangle_{R_1^Q}\right)\right],$$

where R_2^Q denotes the 2-potential of the semigroup $\{P_t^Q\}_{t\geqslant 0}$ and

$$\langle Df\,, Dg\rangle_{R_1^Q}(x) := (Df)(x)\cdot(R_1^Q Dg)(x) \qquad \text{a.s.}$$

In particular,

$$\left|\mathbb{E}[\Phi(X)] - \mathbb{E}[\Phi(Q^{1/2}Z)]\right| \leqslant \tfrac{1}{2}\sup_{x\in\mathbb{R}^n}\max_{1\leqslant i,j\leqslant n}\left|(D_{i,j}^2\Phi)(x)\right| \cdot \sum_{i,j=1}^n \mathbb{E}\left(\left|Q_{i,j} - \langle DX_i\,, DX_j\rangle_{R_1^Q}\right|\right),$$

for all $\Phi : \mathbb{R}^n \to \mathbb{R}$ that are bounded and have continuous and bounded mixed partial derivatives of order $\leqslant 2$.

Example 5.10. 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 5.9 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 $W := AZ$, where A is the [symmetric] square root of Q. Of course,

$$(D_k W_i)(x) = \frac{\partial}{\partial x_k}\left([AZ]_i\,(x)\right) = \frac{\partial}{\partial x_k}\sum_{l=1}^n A_{l,i}x_l = A_{k,i},$$

for all $x \in \mathbb{R}^n$. Therefore, the fact that $R_1^Q 1 = 1$ implies that for all $1 \leqslant i,j \leqslant n$,

$$\langle DX_i\,, DX_j\rangle_{R_1^Q} = \sum_{k=1}^n A_{i,k}A_{k,j} = Q_{i,j} \qquad \text{a.s.}$$

Consequently, every centered random vector X such that $X_i \in \mathbb{D}^{1,2}(\mathbb{P}_n)$ 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 5.9, 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).

Example 5.11. Suppose $X^{(1)}, X^{(2)}, \ldots \in \mathbb{D}^{1,2}(\mathbb{P}_n)$, all have mean vector $0 \in \mathbb{R}^n$, and for all $1 \leqslant i, j \leqslant n$,

$$\lim_{L\to\infty} \left\langle DX_i^{(L)}\,,\, DX_j^{(L)}\right\rangle_{R_1^Q} = Q_{i,j} \quad \text{in } L^1(\mathbb{P}_n).$$

Then, $X^{(L)}$ converges in distribution to $\mathrm{N}_n(0\,,Q)$ as $L\to\infty$.

Theorem 5.9 has other connections to results about asymptotic normality as well. The following shows how Theorem 5.9 is related to the classical CLT, for instance.

Example 5.12. Let $n \geqslant 2$, and suppose $\phi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$ 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 23],

$$(D_k X_1)(x) = \frac{\phi'(x_k)}{\sqrt{n}} \quad \text{and} \quad (D_k X_\ell)(x) = 0 \text{ for } 2 \leqslant \ell \leqslant n,$$

almost surely for every $1 \leqslant k \leqslant n$. I am writing ϕ' in place of the more cumbersome $D\phi$, as I 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 51]. 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 $L^1(\mathbb{P}_n)$. In particular, we can deduce from Theorem 5.9 that for every $\Phi \in C_c^2(\mathbb{R})$,

$$\left|\mathbb{E}\left[\Phi\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\phi(Z_k)\right)\right] - \mathbb{E}[\Phi(\sigma Z_1)]\right| \leqslant C\,\mathbb{E}\left(\left|\sigma^2 - \langle DX_1\,,DX_1\rangle_{R_1^Q}\right|\right)$$
$$\to 0 \qquad \text{as } n \to \infty,$$

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 5.9 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 $\phi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$ has mean zero and finite variance.[4] 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.

§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.

Theorem 5.13 (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 $a_1,\ldots,a_n \in \mathbb{R}$,

$$\mathbb{P}\left\{X_i \leqslant a_i\ {}^\forall 1 \leqslant i \leqslant n\right\} \leqslant \mathbb{P}\left\{Y_i \leqslant a_i\ {}^\forall 1 \leqslant i \leqslant n\right\}.$$

In particular, $\mathbb{P}\left\{\max_{1\leqslant i\leqslant n} X_i \geqslant a\right\} \geqslant \mathbb{P}\left\{\max_{1\leqslant i\leqslant n} Y_i \geqslant a\right\}$ for all $a \in \mathbb{R}$.

The following is an immediate consequence of Theorem 5.13 and integration by parts. It states that less correlated Gaussian vectors tend to take on larger values.

Corollary 5.14. *Under the assumptions of Theorem 5.13,*

$$\mathbb{E}\left[\max_{1\leqslant i\leqslant n} X_i\right] \geqslant \mathbb{E}\left[\max_{1\leqslant i\leqslant n} Y_i\right].$$

[4] 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 $\phi \in L^2(\mathbb{P}_1)$ with $\mathbb{E}_1(\phi) = 0$. The present formulation is slightly weaker since we need the additional smoothness condition that $\phi \in \mathbb{D}^{1,2}(\mathbb{P}_1)$. It is possible to obtain the general form from the weaker one by an approximation argument.

Proof. By integration by parts,

$$\mathbb{E}(W) = \int_0^\infty \mathbb{P}\{W > a\}\,\mathrm{d}a - \int_{-\infty}^0 (1 - \mathbb{P}\{W > a\})\,\mathrm{d}a,$$

for all $W \in L^1(\mathbb{P})$. 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 5.13 to compare the two formulas. □

One can frequently use Corollary 5.14 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.

Example 5.15. Suppose $X = (X_1, \ldots, X_n)$ is a Gaussian random vector with $\mathbb{E}(X_i) = 0$, $\mathrm{Var}(X_i) = 1$, 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, $\mathbb{E}(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1$, and $\mathrm{Cov}(Y_i, Y_j) = 1 - \varepsilon$ when $i \neq j$. Slepian's inequality implies that $\mathbb{E}[\max_{i\leqslant n} X_i] \geqslant \mathbb{E}[\max_{i\leqslant n} Y_i]$. 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

$$\mathbb{E}\left[\max_{1\leqslant i\leqslant n} X_i\right] \geqslant \mathbb{E}\left[\max_{1\leqslant i\leqslant n} Y_i\right] = \sqrt{\varepsilon}\,\mathbb{E}\left[\max_{1\leqslant i\leqslant n} Z_i\right] = (1 + o(1))\sqrt{2\varepsilon\log n},$$

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,

$$\mathbb{E}\left[\max_{1\leqslant i\leqslant n} X_i\right] \leqslant (1 + o(1))\sqrt{2\log n} \quad \text{as } n \to \infty.$$

[For this, one does not even need to know that $(X_1, \ldots, X_n)$ has a multivariate normal distribution.]

Example 5.16. 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 $\mathbb{E}(W_i) = 0$, $\mathrm{Var}(W_i) = 1$, and $\mathrm{Cov}(W_i, W_j) \geqslant -1 + \delta$ for some $\delta \in (0, 1]$, then

$$\mathbb{E}\left[\max_{1\leqslant i\leqslant n} W_i\right] \leqslant (1 + o(1))\sqrt{2\delta\log n} \qquad \text{as } n \to \infty.$$

Proof of Theorem 5.13. Let Q^X and Q^Y denote the respective covariance matrices of X and Y and let A and B 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 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_n)$, and defined as $X = AZ$ and $Y = BZ$. Since $\langle DX_i, DX_j\rangle_{R_1} = Q^X_{i,j}$, Theorem 5.9 shows that for all $\Phi \in \mathbb{D}^{1,2}(\mathbb{P}_n)$

$$\mathbb{E}[\Phi(X)] - \mathbb{E}[\Phi(Y)] = \sum_{i,j=1}^{n} \mathbb{E}\left[\left(R_2^{Q^Y} D^2_{i,j}\Phi\right)(X)\right] \times \left(Q^Y_{i,j} - Q^X_{i,j}\right).$$

Suppose, in addition, that $D^2_{i,j}\Phi \leqslant 0$ 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 $\mathbb{E}[\Phi(X)] \leqslant \mathbb{E}[\Phi(Y)]$. In particular,

$$\mathbb{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] \leqslant \mathbb{E}\left[\prod_{i=1}^{n} \varphi_i(Y_i)\right],$$

whenever $\varphi_1, \ldots, \varphi_n \in C^2_0(\mathbb{P}_1)$ are non increasing. Approximate every $\mathbb{1}_{(-\infty, a_i]}$ by a non-increasing function $\varphi_i \in C^2_0(\mathbb{P}_n)$ to finish. □

The following inequality of Fernique XXX refines Slepian's inequality in a certain direction.

Theorem 5.17 (Fernique, XXX)**.** *Let X and Y be two mean-zero Gaussian random vectors on $\mathbb{R}^n$. Suppose that for every $1 \leqslant i, j \leqslant n$:*

$$\mathbb{E}\left(|X_i - X_j|^2\right) \geqslant \mathbb{E}\left(|Y_i - Y_j|^2\right). \tag{5.23}$$

Then, $\mathbb{P}\{\max_{1\leqslant i\leqslant n} X_i \geqslant a\} \geqslant \mathbb{P}\{\max_{1\leqslant i\leqslant n} Y_i \geqslant a\}$ for all $a \in \mathbb{R}$. In particular, $\mathbb{E}[\max_{1\leqslant i\leqslant n} X_i] \geqslant \mathbb{E}[\max_{1\leqslant i\leqslant n} Y_i]$.

If, in addition, $\mathrm{Var}(X_i) = \mathrm{Var}(Y_i)$ for all $1 \leqslant i \leqslant n$, then condition (5.23) reduces to the covariance condition of Slepian's inequality. Therefore, you should 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 5.13, but the variation is non trivial and involves many computations. The idea is, as before, to show that, for $\Phi(x) := \max_{1\leqslant i\leqslant n} x_i$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\Phi\left(\mathrm{e}^{-t}X + \sqrt{1-\mathrm{e}^{-2t}}\,Y\right)\right] \leqslant 0,$$

whence $\mathbb{E}[\Phi(X)] \geqslant \mathbb{E}[\Phi(Y)]$. You can find the details, for example, in Ledoux and Talagrand XXX and Marcus and Rosen XXX.