Probabilistic Approximations

University of Helsinki
Department of mathematics and statistics
Spring 2015

The aim of this mini-course is to provide an introduction to the combination of two probabilistic techniques. First the Stein's method (1972). This is a collection of probabilistic techniques which allow to compare probability distributions by means of the properties of differential operators (for more information, see [4]). Second the Malliavin calculus (1973). It's an infinite dimensional differential calculus (for a detailed text, see the book [7]). Interestingly, the aforementioned techniques can be sweetly combined in order to provide CLTs for non-linear functionals of an infinite dimensional *isonormal Gaussian process*. As a substantial result, we will proof an astonishing discovery (this is of one the main objectives of the course) by Nualart-Peccati (2005)[1] known nowadays as the *fourth moment theorem*, stating that, for a sequence F_n of random variables living in a fixed *Wiener chaos* such that $\mathbb{E}(F_n^2) \to 1$, the sequence F_n converges in distribution towards a standard Gaussian distribution if and only if $\mathbb{E}(F_n^4) \to 3 (= \mathbb{E}(N^4)$, where $N \sim \mathscr{N}(0,1))$. This new and efficient methodology, i.e. combining the Malliavin calculus together with the Stein's method, in literatures, is called the *Malliavin-Stein* approach. For an exposition of this fertile line of research, one can consult the following constantly updated webpage:

`https://sites.google.com/site/malliavinstein/home`

for many applications of Malliavin-Stein approach, as well as for asymptotic results that are somehow connected with the fourth moment theorem. Moreover, the monograph [5] provides a quite detailed introduction to the topics that will be discussed in the course.

The plan of the course is the following. Lecture 1 : Stein's method, Gaussian measure, stochastic integration and chaotic decompositions, Malliavin calculus. Lecture 2 : combination of the Stein's method with the Malliavin calculus and CLTs on the Wiener chaos. Lecture 3 : applications, new directions (powerful Markov triplet approach [1, 2]) and generalizations (non-Gaussian target distributions [6, 3]) as well as some important open problems if time permits.

[1] Nualart, D., Peccati, G. (2005) Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* Volume 33, Number 1, 177-193.

Part I
Gaussian approximation

1 Introduction

Typical example. Take $W = \{W_t, t \geq 0\}$ a standard BM started from zero. This means that W is a centered Gaussian process such that $W_0 = 0$, W has continuous paths, and $\mathbb{E}(W_s W_t) = s \wedge t$ for every $t, s \geq 0$. A result by Julin (1979) says:

$$\int_0^1 \frac{W_t^2}{t^2} dt = \infty \quad \text{a.s.} \tag{1}$$

(Note that this is a property at around 0). Also, notice that for all $\varepsilon > 0$ we have

$$B_\varepsilon = \int_\varepsilon^t \frac{W_t^2}{t^2} dt < \infty$$

Remark 1. Define a new process $\hat{W}$ by $\hat{W}_0 = 0$ and $\hat{W}_u = uW_{1/u}$ for $u > 0$. It can be easily shown that $\hat{W}$ is a standard Brownian motion, and using the change of variable $u = 1/t$, it now follows that the property (1) is equivalent to the following statement:

$$\int_1^\infty \frac{W_t^2}{t^2} dt = \infty \quad \text{a.s.}$$

By direct computations, one can show that (check it!)

$$\mathbb{E}(B_\varepsilon) = -\log \varepsilon, \quad \mathrm{Var} B_\varepsilon \approx \sqrt{-4 \log \varepsilon}, \quad \text{as}\, \varepsilon \to 0.$$

By setting

$$\tilde{B}_\varepsilon = \frac{B_\varepsilon + \log \varepsilon}{\sqrt{-4 \log \varepsilon}}, \qquad \varepsilon \in (0, 1)$$

one can ask the following natural question:

Problem 1. *Prove that, as $\varepsilon \to 0$, we have $\tilde{B}_\varepsilon \overset{law}{\to} \mathscr{N}(0, 1)$.*

Later on, we will present two different solutions to the above problem. One, using the classical *method of moments/cumulant*, and second, using the techniques introduced in this course. It will turn out that using the second approach, we are not only able to give a fruitful solution to the above problem but also we can provide the following quantitative bound: there exist constants C_1 and C_2 such that

$$C_1(\sqrt{-\log \varepsilon})^{-1} \leq d_{\mathrm{Kol}}(\tilde{B}_\varepsilon, \mathcal{N}(0, 1)) \leq C_2(\sqrt{-\log \varepsilon})^{-1}.$$

2 Elements of Stein's method

The typical route is the following (a) Stein's lemma, then (b) develop a heuristic, followed by (c) an equation whose solutions (and properties thereof) will lead to bounds.

2.1 Moments/Cumulants

During the lectures, the notion of *cumulant* is sometimes used. Recall that, given a random variable Y with finite moments of all orders, i.e. $\mathbb{E}|Y|^r < \infty$ for all $r \geq 1$, and with characteristic function $\varphi_Y(t) := \mathbb{E}(e^{itY})$, $t \in \mathbb{R}$, one define the sequence of cumulants of Y, noted as $\{\kappa_r(Y) : r \geq 1\}$, as

$$\kappa_r(Y) = (-i)^r \frac{d^r}{dt^r} \log \varphi_Y(t)\Big|_{t=0}, \quad r \geq 1.$$

For instance,

$$\begin{aligned}
\kappa_1(Y) &= \mathbb{E}(Y) \\
\kappa_2(Y) &= \mathrm{Var}(Y) \\
\kappa_3(Y) &= \mathbb{E}(Y^3) - 3\mathbb{E}(Y^2)\mathbb{E}(Y) + 2\mathbb{E}(Y)^3 \\
\kappa_4(Y) &= \mathbb{E}(Y^4) - 4\mathbb{E}(Y)\mathbb{E}(Y^3) - 3\mathbb{E}(Y^2)^2 + 12\mathbb{E}(Y)^2\mathbb{E}(Y^2) - 6\mathbb{E}(Y)^4.
\end{aligned}$$

In particular, if $\mathbb{E}(Y) = 0$, then $\kappa_3(Y) = \mathbb{E}(Y^3)$ and $\kappa_4(Y) = \mathbb{E}(Y^4) - 3\mathbb{E}(Y^2)^2$. Recall that for a standard Gaussian random variable $N \sim \mathscr{N}(0,1)$, we have $\log \varphi_N(t) = -t^2/2$, and therefore $\kappa_1(N) = \mathbb{E}(N) = 0$, $\kappa_2(N) = \mathrm{Var}(N) = 1$, and $\kappa_r(N) = 0$ **for all** $r \geq 3$.

Remark 2. The following relation shows that moments can be recursively defined in terms of cumulants (and vice-versa): fix $r = 1, 2, \cdots$ and assume that $\mathbb{E}|Y|^{r+1} < \infty$, then

$$\mathbb{E}(Y^{r+1}) = \sum_{s=0}^{r} \binom{s}{r} \kappa_{s+1}(Y)\mathbb{E}(Y^{r-s}). \tag{2}$$

The reader is referred to [8, Chapter 3] for a proof of relation (2), as well as, for a self-contained presentation of more properties of cumulants and for several combinatorial characterizations.

Exercise 1. Let $N \sim \mathscr{N}(0,1)$. (a) Show that the moments sequence $\{m_r(N) := \mathbb{E}(N^r) : r \geq 1\}$ of N satisfies in the following recursion formula

$$m_{r+1}(N) = r m_{r-1}(N), \quad r \geq 1. \tag{3}$$

(b) Using induction and part (a) to prove that

$$m_r(N) = \begin{cases} (2k-1)!! & \text{if } r = 2k \\ 0 & \text{otherwise.} \end{cases}$$

where the notation double factorial $(2k-1)!! = (2k-1) \times (2k-3) \times \cdots \times 3 \times 1$.

The following lemma is a fundamental key to provide CLTs using the method of moments/cumulants.

Lemma 1. *The law of the random variable $N \sim \mathscr{N}(0,1)$ is determined by its moments/cumulants, i.e if X be a random variable such that $\mathbb{E}(X^r) = \mathbb{E}(N^r)$ [or equivalently $\kappa_r(X) = \kappa_r(N)$] for all $r \geq 1$, then $X \stackrel{law}{=} N$.*

Proof. Let law$(N) = \gamma$ and law$(X) = \nu$. Then, it is enough to show that their Fourier transforms are the same: $\int_{\mathbb{R}} e^{itx} \gamma(dx) = \int_{\mathbb{R}} e^{itx} \nu(dx)$, for every $t \in \mathbb{R}$. Since $m_r(N) = m_r(X)$ for all $r \geq 1$, using Taylor's formula, triangle inequality, the following elementary inequality

$$\Big| e^{itx} - \sum_{k=0}^{r} \frac{(itx)^k}{k!} \Big| \leq \frac{|tx|^{r+1}}{(r+1)!}$$

and Cauchy-Schwarz inequality to write

$$\begin{aligned} \Big| \int_{\mathbb{R}} e^{itx} \gamma(dx) - \int_{\mathbb{R}} e^{itx} \nu(dx) \Big| &\leq \int_{\mathbb{R}} \Big| e^{itx} - \sum_{k=0}^{r} \frac{(itx)^k}{k!} \Big| \gamma(dx) \\ &+ \int_{\mathbb{R}} \Big| e^{itx} - \sum_{k=0}^{r} \frac{(itx)^k}{k!} \Big| \nu(dx) \\ &\leq \Big(\int_{\mathbb{R}} \frac{|tx|^{2r+2}}{(r+1)!^2} \gamma(dx) \Big)^{\frac{1}{2}} \\ &+ \Big(\int_{\mathbb{R}} \frac{|tx|^{2r+2}}{(r+1)!^2} \nu(dx) \Big)^{\frac{1}{2}} \\ &= 2 \sqrt{\frac{|t|^{2r+2} m_{2r+2}(N)}{(r+1)!^2}}, \end{aligned}$$

for every $r \geq 1$. Now, using Stirling formula $r! \sim \sqrt{2\pi r}(\frac{r}{e})^r$ as $r \to \infty$, and Lemma 2.1, one can infer that

$$\lim_{r\to\infty} \frac{|t|^{2r+2} m_{2r+2}(N)}{(r+1)!^2} = 0.$$

□

b) The following lemma known as *Stein's lemma* provides a useful characterization of one-dimensional standard Gaussian distributions.

Lemma 2. (Stein's lemma) *For a real-valued random variable Y we have $Y \sim \mathscr{N}(0,1)$ if, and only if for every $f : \mathbb{R} \to \mathbb{R}$ such that $E|f'(N)| < \infty$, we have*

$$\mathbb{E}(f'(Y) - Yf(Y)) = 0. \tag{4}$$

Proof. The sufficient condition is trivial. For the other way, note that for all polynomials the relation (4) works. But this means that

$$\mathbb{E}(Y^{r+1}) = r\mathbb{E}(Y^{r-1}).$$

Now, use Exercise 2.1 and Lemma 1. Another way is to take f complex exponential and therefore determine the characteristic function of Y (do it!). □

Theorem 1. (The method of moments/cumulants) *Let F be a real-valued random variable whose law is determined by its moments/cumulants. Assume that $\{F_n\}_{n\geq 1}$ be a sequence of random variables in which each F_n has all moments/cumulants such that $\mathbb{E}(F_n^r) \to \mathbb{E}(F^r)$, for every $r \geq 1$. Then F_n converges in distribution towards F.*

b) Heuristic. Suppose that for "many" functions f we have

$$\mathbb{E}(f'(Y) - Yf(Y)) \approx 0.$$

Can we conclude that Y is close – in some sense – to N? This is, *a priori*, not clear since there are many ways to characterize N and not all lead to a nice theory of probabilistic approximation. We will consider a very strong measure of closeness in terms of the total variation (TV) distance.

2.2 Distances between probability measures

(i) *The Kolmogorov distance*: Let F and G be two $\mathbb{R}^d$, $(d \geq 1)$ valued random variables. Let

$$\mathscr{H}_{\text{Kol}} = \{h : \mathbb{R}^d \to \mathbb{R} \,:\, h(x_1, \cdots, x_d) = \Pi_{k=1}^d \mathbf{1}_{(-\infty, z_k]}(x_k), \text{ for some } z_1, \cdots, z_d \in \mathbb{R}\}.$$

The Kolmogorov distance between the laws of random variables F and G, noted as $d_{\text{Kol}}(F,G)$, define as

$$\begin{aligned} d_{\text{Kol}}(F,G) &= \sup_{h \in \mathscr{H}_{\text{Kol}}} \Big| \mathbb{E}(h(F)) - \mathbb{E}(h(G)) \Big| \\ &= \sup_{z_1, \cdots, z_d \in \mathbb{R}} \Big| \mathbb{P}(F \in (-\infty, z_1] \times \cdots \times (-\infty, z_d]) \\ &\qquad\qquad - \mathbb{P}(G \in (-\infty, z_1] \times \cdots \times (-\infty, z_d]) \Big|. \end{aligned}$$

In particular $(d = 1)$: $d_{\text{Kol}}(F,G) = \sup_{z \in \mathbb{R}} \Big| \mathbb{P}(F \leq z) - \mathbb{P}(G \leq z) \Big|$. Note that always $d_{\text{Kol}}(F,G) \leq d_{\text{TV}}(F,G)$.
(ii) *The total variation distance*:

$$\mathscr{H}_{\text{TV}} = \{h : \mathbb{R}^d \to \mathbb{R} \,:\, h = \mathbf{1}_B \text{ for some } B \in \mathcal{B}(\mathbb{R}^d)\}.$$

$$\begin{aligned} d_{\text{TV}}(F,G) &= \sup_{h \in \mathscr{H}_{\text{TV}}} \Big| \mathbb{E}(h(F)) - \mathbb{E}(h(G)) \Big| \\ &= \sup_{B \in \mathcal{B}(\mathbb{R}^d)} \Big| \mathbb{P}(F \in B) - \mathbb{P}(G \in B) \Big|. \end{aligned}$$

(iii) *The Wasserstein distance*:

$$\mathscr{H}_{\text{W}} = \{h : \mathbb{R}^d \to \mathbb{R} \,:\, \|h\|_{\text{lip}} \leq 1\}, \quad \|h\|_{\text{lip}} := \sup_{x \neq y \in \mathbb{R}^d} \frac{|h(x) - h(y)|}{\|x - y\|_{\mathbb{R}^d}}.$$

$$d_{\text{W}}(F,G) = \sup_{h \in \mathscr{H}_{\text{W}}} \Big| \mathbb{E}(h(F)) - \mathbb{E}(h(G)) \Big|.$$

Exercise 2. Let $d \geq 1$. Show that the topologies induced by three distance d_{Kol}, d_{TV} and d_{W} on the set of probability measures on $\mathbb{R}^d$ are strictly stronger than the topology of the convergence in distribution, i.e.

$$d_{\text{Kol,TV,W}}(F_n, F) \to 0 \implies F_n \overset{\text{law}}{\to} F.$$

Remark 3. The *Fortet–Mourier distance* (or *bounded Wasserstein distance*: $d_{\mathrm{FM}}(F,G) = \sup_{h\in\mathscr{H}_{\mathrm{FM}}} \Big|\mathbb{E}(h(F)) - \mathbb{E}(h(G))\Big|$, where $\mathscr{H}_{\mathrm{FM}} = \{h : \mathbb{R}^d \to \mathbb{R} : \|h\|_\infty + \|h\|_{\mathrm{lip}} \leq 1\}$. The d_{FM} distance metrizes the convergence in distribution, i.e.

$$d_{\mathrm{FM}}(F_n, F) \to 0 \iff F_n \overset{\text{law}}{\to} F.$$

c) Stein's equation for normal approximation. Let $N \sim \mathscr{N}(0,1)$. Consider a function $h : \mathbb{R} \to [0,1]$ so that $\mathbb{E}|h(N)| < \infty$. The Stein equation associated to the test function h is

$$f'(x) - xf(x) = h(x) - \mathbb{E}h(N) \tag{5}$$

which is taken to hold at all $x \in \mathbb{R}$. A solution is a function f_h whose derivative is a.e. defined and for which there exists a version which satisfies (5). In particular we always speak of f' in the weak sense. For a moment, assume that f_h is a solution of (5). Then, by taking expectation of both sides (5) (together with plugging in $x = Y$, where Y is a real-valued random variable):

$$\mathbb{E}h(Y) - \mathbb{E}h(N) = \mathbb{E}(f'(Y) - Yf(Y)).$$

Therefore, for any integrable random variable Y:

$$\sup_{h\in\mathscr{H}_{\mathrm{TV}}} \Big|\mathbb{E}h(Y) - \mathbb{E}h(N)\Big| = \sup_{f_h, h\in\mathscr{H}_{\mathrm{TV}}} \Big|\mathbb{E}(f_h'(Y) - Yf_h(Y))\Big|. \tag{6}$$

Note that the expression in the right hand side in above **does not** involved the target random variable N at all!

Proposition 1. *For every $c \in \mathbb{R}$, set*

$$f_{c,h}(x) = ce^{x^2/2} + e^{x^2/2}\int_{-\infty}^{x} (h(u) - \mathbb{E}h(N))e^{-u^2/2}du;$$

Then, $f_{c,h}$ is a solution of the Stein's equation. Moreover, the unique solution satisfying in $\lim_{n\to\infty} e^{-\frac{x^2}{2}} f(x) = 0$ is given by $f_h = f_{0,h}$, i.e. $c = 0$.

Proof. Note that, the Stein's equation can be written as

$$e^{\frac{x^2}{2}} \frac{d}{dx}\Big(e^{-\frac{x^2}{2}} f(x)\Big) = h(x) - \mathbb{E}h(N).$$

Now, take integral of both sides. For the second part, using dominated convergence theorem (DCT) we have

$$\lim_{|x|\to\infty} \int_{-\infty}^{x} (h(u) - \mathbb{E}h(N))e^{-u^2/2}du = 0.$$

Recall that $\mathbb{E}|h(N)| < \infty$. □

The gist of the method is that it will transform the study of a non-smooth object (the TV distance) in terms of smooth objects (the solutions $f_{0,h}$). This happens through the following lemma.

Lemma 3. *Let $h : \mathbb{R} \to [0,1]$. Then the solution f_h of the Stein's equation associated to h satisfying in*

$$\|f_h\|_\infty \leq \sqrt{\pi/2} \quad \textit{and} \quad \|f_h'\|_\infty \leq 2$$

(the Stein's magic factors).

Note that these bounds are uniform over the whole family h. An immediate consequence of Lemma 3 is the following.

Corollary 1. *Let Y be a real-valued random variable such that $\mathbb{E}|Y| < \infty$. Then*

$$d_{TV}(Y, N) \leq \sup_{f \in \mathscr{F}_{TV}} \Big|\mathbb{E}(f'(Y)) - \mathbb{E}(Yf(Y))\Big|,$$

where $\mathscr{F}_{TV} = \{f \,:\, \|f\|_\infty \leq \sqrt{\pi/2} \quad \textit{and} \quad \|f'\|_\infty \leq 2\}$.

Let to stress that we have explicitly transformed the non-smooth problem, the lhs of (6), into a smooth one, the rhs of (6). Moreover, the bounds in Lemma 3 are independent of the target Gaussian random variable N, and just depends on (a functional of) a nice Y! Once we have this, is it true that the rhs is easier to evaluate than the lhs? This is Stein's intuition and it turns out to be true in many different and complicated cases. There are several techniques for working out this quantity : exchangeable pairs (Stein, 1972); dependency graphs; zero-bias transforms (Goldstein - Reinert 1995). Here we are going to develop tools to evaluate the Stein bound when Y is a (sufficiently regular) functional of a infinite-dimensional Gaussian field (e.g. the Brownian motion, the fractional Brownian motion,...). The answer for this is through *Malliavin calculus.*

Proof. (of Lemma 3) First remark that $|h(u) - \mathbb{E}h(N)| \le 1$. Then we easily get (with Φ the Gaussian CDF)

$$\begin{aligned}|f_h(x)| &\le e^{x^2/2}\min\{\Phi(x), 1-\Phi(x)\}\\ &= e^{x^2/2}\int_{|x|}^{\infty} e^{-y^2/2}dy := S(x).\end{aligned}$$

A direct computation shows that S attains its maximum at $x = 0$, and also $S(0) = \sqrt{\pi/2}$. Hence, $|f_h| \le S(0) = \sqrt{\pi/2}$, and the first claim follows.

For the second bound, we can simply write out f_h', to get

$$\begin{aligned}|f_h'(x)| &= \left|h(x) - \mathbb{E}h(N) + xe^{x^2/2}\int_{-\infty}^{x}(h(u) - \mathbb{E}h(N))e^{-u^2/2}du\right|\\ &\le 1 + |x|e^{x^2/2}\int_{|x|}^{\infty} e^{-y^2/2}dy = 2.\end{aligned}$$

□

Exercise 3. Stein's bound for the Kolmogorov distance (a) For every $z \in \mathbb{R}$, write $f_z = f_{\mathbf{1}_{(-\infty,z]}}$, that is, f_z is the solution of the Stein's equation associated to the indicator function $h = \mathbf{1}_{(-\infty,z]}$. Also, Φ stands for the cumulative distribution function of a $\mathscr{N}(0,1)$ random variable. Show that

$$f_z(x) = \begin{cases}\sqrt{2\pi}e^{\frac{x^2}{2}}\Phi(x)[1-\Phi(z)] & \text{if } x \le z,\\ \sqrt{2\pi}e^{\frac{x^2}{2}}\Phi(z)[1-\Phi(x)] & \text{if } x \ge z.\end{cases}$$

(b) Prove that, for every $x \in \mathbb{R}$, $f_z(x) = f_{-z}(-x)$ (this implies that, in the estimates below, one can assume that $z \ge 0$ without loss of generality).
(c) Compute the derivative $\frac{d}{dx}[xf_z(x)]$, and deduce that the mapping $x \mapsto xf_z(x)$ is increasing.
(d) Show that $\lim_{x\to-\infty} xf_z(x) = \Phi(z) - 1$ and also that $\lim_{x\to+\infty} xf_z(x) = \Phi(z)$.
(e) Use part (a) to prove that

$$f_z'(x) = \begin{cases}[\sqrt{2\pi}xe^{\frac{x^2}{2}}\Phi(x) + 1][1-\Phi(z)] & \text{if } x < z,\\ [\sqrt{2\pi}xe^{\frac{x^2}{2}}(1-\Phi(x)) - 1]\Phi(z) & \text{if } x > z.\end{cases}$$

(f) Use part (e) in order to prove that

$$0 < f_z'(x) \le zf_z(x) + 1 - \Phi(z) < 1, \quad \text{if}\, x < z,$$

and

$$-1 < zf_z(x) - \Phi(z) \le f_z'(x) < 0, \quad \text{if}\, x > z$$

to deduce that $\|f_z'\|_\infty \le 1$.
(g) Use part (f) to show that $x \mapsto f_z(x)$ attains its maximum in $x = z$. Compute $f_z(z)$ and prove that $f_z(z) \le \frac{\sqrt{2\pi}}{4}$ for every $z \in \mathbb{R}$, to complete a proof of the following theorem.

Theorem 2. *Let $z \in \mathbb{R}$. Then the function f_z is such that $\|f_z\|_\infty \le \frac{\sqrt{2\pi}}{4}$ and $\|f_z'\|_\infty \le 1$. Therefore, for $N \sim \mathscr{N}(0,1)$, and for any integrable random variable F,*

$$d_{Kol}(F, N) \le \sup_{f \in \mathscr{F}_{Kol}} |\mathbb{E}[f'(F)] - \mathbb{E}[Ff(F)]|,$$

where $\mathscr{F}_{Kol} = \{f : \|f\|_\infty \le \frac{\sqrt{2\pi}}{4}, \|f'\|_\infty \le 1\}$.

Part II
Gaussian measures and chaos

Take $(\Omega, \mathcal{F}, \mathbb{P})$ an underlying probability space.

2.3 Definition and first properties

We first define Gaussian measures.

Definition 1. *Take $(A, \mathcal{A}, \mu)$ a measure space (Polish, i.e. metric, separable and complete) with μ positive, σ-finite and non-atomic measure. A Gaussian measure over $(A, \mathcal{A})$ with control μ is a centered Gaussian family*

$$G = \{G(B); \quad \mu(B) < \infty\}$$

such that

$$\mathbb{E}\left(G(B)G(C)\right) = \mu(B \cap C), \quad \mu(B) < \infty \, and \, \mu(C) < \infty.$$

A couple of remarks : (i) If $A = \mathbb{R}^+$ and μ is the Lebesgue measure then $W_t = G[0,t]$ is, up to continuity, a Brownian motion (because then $\mathbb{E}W_tW_s = \min(t,s)$); (ii) one can prove that if $\{B_i\}_{i\geq 1}$ is a sequence of **disjoint** sets such that $\mu(\bigcup_i B_i) < \infty$ then

$$G(\bigcup_i B_i) = \sum_i G(B_i)$$

with convergence in $L^2(\Omega)$.

Proposition 2. *G, in fact, exists.*

Proof. Take $\{e_i\}_{i\geq 1}$ an ONB of $L^2(\mu)$. Then for all $f \in L^2(\mu)$ we have $f = \sum \langle f, e_i\rangle e_i$ with the $L^2(\mu)$ scalar product. Next take $\{\xi_i\}_{i\geq 1}$ a sequence of i.i.d. $\mathscr{N}(0,1)$ random variables and construct

$$X(f) = \sum_{ige1} \xi_i \langle f, e_i\rangle.$$

Then $\{X(f); f \in L^2(\mu)\}$ is a centered Gaussian family such that $\mathbb{E}(X(f)X(g)) = \langle f, g\rangle$ (easy exercise through Parseval's identity, check it!). We are then ready to conclude, since

$$G(B) = X(\mathbf{1}_B), \quad \mu(B) < \infty$$

is a Gaussian measure with control measure μ. □

A final remark is that GM are not probability measures! More precisely:

Proposition 3. *The mapping*

$$B \mapsto G(B)(\omega)$$

is not a signed measure for a fixed ω.

Proof. Take a Borel set B with $\mu(B) < \infty$. Since μ is non-atomic, we observe that

$$\int_A \int_A \mathbf{1}_{B\times B}(x,y)\mathbf{1}_{x=y}\mu(dx)\mu(dy) = \mathrm{Diag}_\mu(B) = 0.$$

But, on the other hand side, one can easily show that

$$\int_A \int_A \mathbf{1}_{B\times B}(x,y)\mathbf{1}_{x=y}G(dx)G(dy) := \mathrm{Diag}_G(B) = \mu(B).$$

(Note that the integration wrt G is shaky but will be proven rigorously later on). Indeed here a standard way to construct $\mathrm{Diag}_G(B)$ is through

$$\lim_{n\to\infty}\sum_{i=1}^{k_n} G(B_i^{(n)})G(B_i^{(n)}) = \lim_n \sum G(B_i^{(n)})^2$$

where $\{B_i^{(n)}, i = 1, \ldots, k_n\}$ is a sequence of partitions of B such that $\sup \mu(B_i^{(n)}) \to 0$, and one can show that (check it!), for any partition,

$$\mathbb{E}\big(\sum G(B_i^{(n)})^2 - \mu(B)\big)^2 \to 0.$$

In other words G charges, in a nontrivial way, the diagonal and hence cannot be a signed measure. $\square$

Remark 4. In the case, when $A = \mathbb{R}^+$, μ is the Lebesgue measure, the statement $B \mapsto G(B)(\omega)$ is a singed measure on a set of positive probability will imply that the mapping $t \mapsto W_t := G[0,t]$ is of bounded variation on a set of positive probability in which is a contradiction with the following well-known fact that

$$\sum_{0\le t_i\le t} (W_{t_i^{(n)}} - W_{t_{i-1}^{(n)}})^2 \to t.$$

2.4 Single integrals

We want for any $f \in L^2(\mu)$ to define an object of the type

$$I_1(f) := \int_A f(x)G(dx)$$

To this end, we introduce a collection of "simple integrands"

$$\mathcal{E}(\mu) = \{f(x) = \sum_{j=1}^{N} c_j \mathbf{1}_{B_j},\ \mu(B_j) < \infty\}$$

which has the density property $\bar{\mathcal{E}}(\mu) = L^2(\mu)$. Then we can define, for any simple integrand f

$$I_1(f) := \sum_{j=1}^{N} c_j G(B_j).$$

With this in hand it is easy to show that, for all simple f, g we have

$$\mathbb{E}(I_1(f)I_1(g)) = \langle f, g\rangle_{L^2(\mu)}, \quad \mathbb{E}I_1(f) = 0.$$

Now, it is straightforward to extend to all functions $f \in L^2(\mu)$. Let $f \in L^2(\mu)$, then there exists a sequence of simple functions $\{f_n\}_{n\geq 1}$ such that $\|f_n - f\| \to 0$ and $\{I_1(f_n)\}$ is Cauchy in $L^2(\mathbb{P})$. One therefore sets

$$I_1(f) = \lim I_1(f_n)$$

the limit being taken in $L^2(\mathbb{P})$ and being independent of the choice of the sequence.

Remark 5. For all $f, g \in L^2(\mu)$ we have

$$\mathbb{E}I_1(f) = 0 \quad \text{and} \quad \mathbb{E}\left(I_1(f)I_1(g)\right) = \langle f, g\rangle_{L^2(\mu)}.$$

Definition 2 (Wiener, 1938)**.** The space $\mathcal{H}_1 = \{I_1(f) : f \in L^2(\mu)\}$ is the *first Wiener chaos of* G. Note that, since everything is obtained through centered Gaussian random variables, so $\mathcal{H}_1$ is a centered Gaussian family. It therefore cannot be suffice to describe all square integrable functionals measurable wrt G.

2.5 Multiple integrals

Let $p \geq 2$. Consider $L^2(\mu^p)$ the space of square integrable functions of p arguments on the space $(A^p, \mathcal{A}^p, \mu^p)$. We then define the simple functions

$$\begin{aligned}\mathcal{E}(\mu^p) =& \text{ Simple integrands}\\ =& \left\{ f = \sum_{i_1,\cdots,i_p=1}^{n} a_{i_1\cdots i_p}\mathbf{1}_{B_{i_1}} \otimes \cdots \otimes \mathbf{1}_{B_{i_p}} : B_{i_k} \cap B_{i_l} = \emptyset\, \forall\, k \neq l \text{ and } \mu(B_{i_j}) < \infty \right\},\end{aligned}$$

and **more importantly** the coefficients $a_{i_1\cdots i_p} = 0$ if any of two indices $i_1, \cdots, i_p$ are equal, i.e. $\exists\, k \neq l$ such that $i_k = i_l$.

For any $f \in \mathcal{E}(\mu^p)$ then we define the *multiple Wiener-Itô integral of order* p *of* $f \in \mathcal{E}(\mu^p)$ *w.r.t.* G through

$$I_p(f) = \sum_{i_1,\cdots,i_p=1}^{n} a_{i_1\cdots i_p} G(B_1)\dots G(B_d). \tag{7}$$

Now, the main properties are gathered in the following exercise.

Exercise 4. **(a)** Show that for $f \in \mathcal{E}(\mu^p)$, the definition of $I_p(f)$ does not depend on a particular representation of f.
(b) $I_p : \mathcal{E}(\mu^p) \to L^2(\Omega, \mathbb{P})$ is a linear map.
(c) $I_p(f) = I_p(\tilde{f})$, where $\tilde{f}$ is the *symmetrization* of f, i.e.

$$\tilde{f}(x_1, \cdots, x_p) = 1/p! \sum_{\sigma} f(x_{\sigma(1)}, \cdots, x_{\sigma(p)})$$

where the sum runs over all permutations σ of $\{1, \cdots, p\}$.

(d) For all $f \in \mathcal{E}(\mu^p)$ and $g \in \mathcal{E}(\mu^q)$ we have

$$\mathbb{E}(I_p(f)) = 0 \qquad \text{(centered)}$$

and

$$\mathbb{E}(I_p(f)I_q(g)) = \begin{cases} 0 & \text{if } p \neq q, \\ p!\langle \tilde{f}, \tilde{g} \rangle_{L^2(\mu^p)} & \end{cases} \qquad \text{(orthogonality / isometry)}$$

(e) $\overline{\mathcal{E}_{\text{sym}}(\mu^p)}^{\| \|_{L^2(\mu^p)}} = L^2_{\text{sym}}(\mu^p)$, where here "sym" stands for symmetrized functions. Hence, deduce that the mapping I_p can be continuously extended to $L^2_{\text{sym}}(\mu^p)$.

Now let $f \in L^2_{\text{sym}}(\mu^p)$; then there exists a sequence $\{f_n\} \subset \mathcal{E}_{\text{sym}}(\mu^p)$ such that $f_n \to f$ in $L^2(\mu^p)$. Therefore, we define

$$I_p(f) := \lim_{n\to\infty} I_p(f_n) \tag{8}$$

the limit being taken in $L^2(\mathbb{P})$ and being independent of the chosen sequence.

Definition 3. *The* Wiener chaos of order p associated with G, *denoted by* $\mathcal{H}_p$, *is defined as*

$$\mathcal{H}_p := v.s.\Big\{ I_p(f)\,; \quad f \in L^2_{sym}(\mu^p) \Big\}.$$

Moreover, the three properties (centered, isometry and orthogonality) extend to the whole class $\mathcal{H}_p$. Note that we write $I_0(c) = c; c \in \mathbb{R}$

Remark 6. *If $A = [0, T]$, μ is Lebesgue and $W_t = G([0, t])$ is a Brownian motion, then for symmetric $f \in L^2([0, T]^p)$*

$$I_p(f) = p! \int_0^T dW_{t_1} \int_0^{t_1} dW_{t_2} \dots \int_0^{t_{p-1}} dW_{t_p} f(t_1, \dots, t_p).$$

The random variables in Definition are fundamental: they allow for example to write any G-square integrable random variable as an infinite series (this will be treated later on). The following remark provides some crucial properties of random variables living in a Wiener chaos.

Remark 7. **(a)** Shigekawa [2] proves that if $F = \sum_{p=0}^{M} I_p(f_p)$ then the law of F has a density w.r.t. the Lebesgue measure.
(b) (Nelson, 1968) $\mathcal{H}_p$ is *hypercontractive*, i.e. $\forall q > 0$ there exists $C_{p,q} > 0$ such that for all $F \in \mathcal{H}_p$ we have

$$\mathbb{E}(|F|^q)^{1/q} \leq C_{p,q}\, \mathbb{E}(F^2)^{1/2}. \tag{9}$$

In particular all these L^p topologies are equivalent on the Wiener chaoses.

2.6 Multiplication formulae & Chaotic expansion

The problem. What is $I_p(f) \times I_q(g)$?

Definition 4 (Contraction)**.** *For $f \in L^2_{sym}(\mu^p)$ and $g \in L^2_{sym}(\mu^q)$ for $p, q \geq 1$ we define for all $r = 0, \ldots, \min(p,q)$*

$$f \otimes g(x_1, x_2, \ldots, , x_{p+q-2r})$$
$$= \int_{A^r} f(\underline{a_r}, x_1, \ldots, x_{p-r}) g(\underline{a_r}, x_{p-r+1}, \ldots, x_{p+q-2r}) \mu(da_1, \cdots, da_r).$$

Here $\underline{a_r} = (a_1, \cdots, a_r)$.

Example 1. *If $p = q = r$ then $f \otimes_p g = \langle f, g\rangle_{L^2(\mu^p)}$.*

Example 2. *If $r = 0$ then*

$$f \otimes_0 g(x_1, \ldots, x_{p+q}) = f \otimes g = f(x_1, \ldots, x_p) g(x_{p+1}, \ldots, x_{p+q}).$$

Example 3. *If $p = q = 2$ and $r = 1$ then*

$$f \otimes_1 g(x, y) = \int_A \mu(da) f(a, x) g(a, y).$$

[2] Shigekawa, I. *Derivatives of Wiener functionals and absolute continuity of induced measures.* J. Math. Kyoto Univ. 20 (1980), no. 2, 263-289.

Remark 8.

$$\|f \otimes_r g\|^2_{L^2(\mu^{p+q-2r})} \leq \|f\|^2_{L^2(\mu^p)} \|g\|^2_{L^2(\mu^q)} < \infty.$$

Note that if $p = q = r$ this is just the CS inequality.

Remark 9. *In general $f \otimes_r q$ is not symmetric, so we define the symmetrization*

$$\widetilde{f \otimes_r g}(x_1, \ldots, x_{p+q-2r}) = \sum_{\sigma \in \mathcal{S}_{p+q-2r}} \frac{f \otimes_r g(x_{\sigma(1)}, \ldots, x_{\sigma(p+q-2r)})}{(p+q-2r)!}.$$

Note that, in general,

$$\|\tilde{f}\|_{L^2(\mu^p)} \leq \|f\|_{L^2(\mu^p)},$$

i.e. symmetrization shrinks.

We are now ready to state a fundamental result.

Theorem 3 (Multiplication formulae)**.** *Take $f \in L^2_{sym}(\mu^p)$ and $g \in L^2_{sym}(\mu^q)$. Then*

$$I_p(f) \times I_q(g) = \sum_{r=0}^{\min(p,q)} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}\left(\widetilde{f \otimes_r g}\right).$$

We will proceed to a heuristic proof of this result. For a detailed proof consult [7].

Proof. First note how we have, at each step, been very careful to avoid the diagonals. Hence by our construction $I_p(f)$ can be seen as

$$I_p(f) = \int_A \cdots \int_A f(x_1, \ldots, x_p) \mathbf{1}_{\{x_i \neq x_j, i \neq j\}} G(dx_1) \ldots G(dx_p).$$

Then we have (as through Fubini)

$$\begin{aligned} &I_p(f) I_q(g) \\ &= \int_{A^{p+q}} \underbrace{f(x_1, \ldots, x_p)}_{\text{no diag}} \underbrace{g(y_1, \ldots, y_q)}_{\text{no diag}} G(dx_1) \ldots G(dx_p) G(dy_1) \ldots G(dy_p). \end{aligned}$$

While there are no diagonals in the first and second blocks, there are all possible mixed diagonals in the joint writing. Hence we need to take into account all these diagonals (whence the combinatorial coefficients in the statement, which count all possible diagonal sets of size r) and then “integrate out”, in other words we get

$$I_p(f)I_q(g)$$

$$= \sum_{r=0}^{\min(p,q)} r!\binom{p}{r}\binom{q}{r}\int_{A^{p-r}}\int_{A^{q-r}} \zeta G(dx_1)\dots G(dx_{p-r})G(dy_1)\dots G(dy_{q-r}),$$

with

$$\zeta = \left(\int_{A^r} f(\underline{a_r}, x_1, \dots, x_{p-r})g(\underline{a_r}, x_{p-r+1}, \dots, x_{p+q-2r})\mu^r(d\underline{a_r})\right).$$

Since $\mathrm{Diag}_G(da) = \mu(da)$, we get the proof.

□

2.7 Hermite polynomials and chaos

Definition 5. *We define the Hermite polynomials as the family of polynomials* $\{H_n; n \geq 0\}$ *such that* $H_0 \equiv 1$ *and, for all* $n \geq 1$,

$$H_n(x) = (-1)^n e^{x^2/2} d_x^n (e^{-x^2/2}).$$

The following exercise gather the important properties of the Hermite polynomials.

Exercise 5. Define the **divergence operator** δ on the space $\mathrm{Dom}(\delta) \subset L^2(\mathbb{R}, \gamma)$ as $\delta\varphi(x) := -\varphi'(x) + x\varphi = -e^{\frac{x^2}{2}}\frac{d}{dx}(e^{-\frac{x^2}{2}}\varphi(x))$. Let $p \geq 0$ be an integer. We define the pth **Hermite Polynomial** as $H_0 = 1$ and $H_p = \delta^p 1$, where here $\delta^p = \delta \circ \cdots \circ \delta$, p times.
(a) Show that $d\delta - \delta d =$ Identity, where $d = \frac{d}{dx}$, and moreover $\delta H_p = H_{p+1}$, $dH_p = pH_{p-1}$ and $(\delta + d)H_p = xH_p$.
(b) for any $p, q \geq 0$ show that

$$\int_R H_p(x)H_q(x)\gamma(dx) = \delta_{p,q}p!,$$

where here $\delta_{p,q}$ stands for the Kronecker delta.
(c) Show that the family $\{\frac{1}{\sqrt{p!}}H_p : p \geq 0\}$ is an orthonormal basis of $L^2(\mathbb{R}, \gamma)$.
(d) Define the **Ornstein-Uhlenbeck generator** $L\varphi(x) = -x\varphi'(x) + \varphi''(x)$. Show that $LH_p = -pH_p$.

In other words the multiple integrals are infinite dimensional versions of the Hermite polynomials.

Proposition 4. *For all $h \in L^2(\mu)$ such that $\|h\|_{L^2(\mu)} = 1$ we define*

$$h^{\otimes p}(x_1, \ldots, x_p) = \prod_{i=1}^{p} h(x_i) \in L^2_{sym}(\mu^p).$$

Then, for all $p \geq 1$, we have

$$I_p(h^{\otimes p}) = H_p(I_1(h)).$$

This is sometimes called the Wick product of order p of $I_1(h)$.

Proof. Trivial for $p = 1$. Proceed by induction and choose $p \geq 1$. Then note that

$$I_p(h^{\otimes p})I_1(h) = I_{p+1}(h^{\otimes p+1}) + pI_{p-1}(h^{\otimes p})$$

Whence, using the recursion,

$$I_{p+1}(h^{\otimes p+1}) = H_p(I_1(h))I_1(h) - pH_{p-1}(I_1(h))$$

Using the previous exercise we also know that

$$d_x H_p(x) = pH_{p-1}(x)$$

and thus

$$\begin{aligned} I_{p+1}(h^{\otimes p+1}) &= H_p(I_1(h))I_1(h) - H_p(I_1(h)) \\ &= \delta H_p(I_1(h)) \\ &= H_{p+1}(I_1(h)). \end{aligned}$$

□

Theorem 4. *[Chaotic representation] For all $F \in L^2(\sigma(G))$ there exists a unique $\{f_q; q \geq 1\}$ such that $f_q \in L^2_{sym}(\mu^q)$ and we have*

$$F = \mathbb{E}(F) + \sum_{q=1}^{\infty} I_q(f_q) \tag{10}$$

(the equality is in $L^2(\mathbb{P})$). In particular

$$\mathbb{E}(F^2) = \mathbb{E}(F)^2 + \sum_{q=1}^{\infty} q!\|f_q\|^2_{L^2(\mu^q)}.$$

Proof. We start with a few facts.

- Fact 1 : random variables of the type $I_1(h)$ with $\|h\|_{L^2(\mu)} = 1$ generate $\sigma(G)$.

- Fact 2 : for all λ the function $e^{i\lambda I_1(h)}$ can be approximated in $L^2(\mathbb{P})$ by complex linear combinations (through Taylor) of powers $I_1(h)^m$, $m \geq 1$.

- Fact 3 : If $X \in L^2(\sigma(G))$ is such that $\mathbb{E}(XI_1(h)^m) = 0$ for all h, m, then $\mathbb{E}(Xe^{i\lambda I_1(h)}) = 0$ for all λ, h implies that $X = 0$ almost surely.

As a consequence

$$\overline{v.s.}^{L^2(\sigma(G))}\{I_1(h)^m\,;\, \|h\|_{L^2(\mu)} = 1 \,\text{and}\, \ m \geq 1\} = L^2(\sigma(G)).$$

Hence, all we need to do is to show the theorem for random variables of the type $I_1(h)^m$, i.e. we need to show that every $F = I_1(h)^m$ admits a representation (10). But we already now that there exist $C_{q,m}$, some real constants, such that

$$\begin{aligned} I_1(h)^m &= \sum_{q=0}^{m} C_{q,m} H_q(I_1(h)), \\ &= \sum_{q=0}^{m} C_{q,m} I_q(h^{\otimes q}). \end{aligned}$$

□

3 Elements of Malliavin Calculus

We work within the framework of a Gaussian measure G with the control measure μ having the suitable properties. We associate to all $F \in L^2(\sigma(G))$ an expansion $F = \mathbb{E}(F) + \sum_{q\geq 1} I_q(f_q)$.

3.1 The derivative operator D

We take

$$\mathrm{dom}(D) := \left\{ F \in L^2(\sigma(G)) \,:\, \sum_{q=1}^{\infty} qq!\|f_q\|^2_{L^2(\mu^q)} < \infty \right\}$$

For $F \in \text{dom}(D)$ we define

$$D_t F = \sum_{q=1}^{\infty} q I_{q-1}(f_q(t, \bullet)),\ t \in A$$

where $\bullet$ indicates that we integrate over the $(q-1)$ remaining variables. We can then see that

$$\begin{aligned}
\mathbb{E}\left[\int_A (D_t F)^2 \mu(dt)\right] &= \int_A \mu(dt) E\left[\left(\sum_{q=1}^{\infty} q I_{q-1}(f_q(t, \bullet))\right)^2\right] \\
&= \int_A \mu(dt) \sum_{q=1}^{\infty} q^2 (q-1) \|f_q(t, \bullet)\|^2 \\
&= \sum_{q=1}^{\infty} q q! \|f_q\|^2_{L^2(\mu^q)} < \infty.
\end{aligned}$$

First note that for $f_q = h^{\otimes q}$ with $\|h\| = 1$, then $I_q(f_q) = H_q(I_1(h))$ and

$$\begin{aligned}
D_t I_q(f_q) &= q I_{q-1}(f_q(t, \bullet)) = q I_{q-1}(h^{\otimes q-1}) h(t) = q H_{q-1}(I_1(h)) h(t) \\
&= H_q'(I_1(h)) h(t). \qquad (11)
\end{aligned}$$

In particular $D_t I_1(h) = h(t)$. Using this fact together with several approximation arguments one can prove the following *chain rules.*

Proposition 5. *[Chain rule 1]Let $h_1, \ldots, h_d \in L^2(\mu)$ and take $f : \mathbb{R}^d \to \mathbb{R} \in C_b^1$. Now define $F = f(I_1(h_1), \ldots, I_1(h_d))$. Then $F \in \text{dom}(D)$, and*

$$D_t F = \sum_{j=1}^{d} \partial_{x_j} f(I_1(h_1), \ldots, I_1(h_d)) h_j(t).$$

The requirement that the functions be C_b^1 (differentiable with bounded derivatives) is too stringent, and can be replaced by polynomial tail behavior.

Proposition 6. *[Chain rule 2] Take $F \in \text{dom}(D)$ and $f : \mathbb{R} \to \mathbb{R} \in C_b^1$. Then*

$$D_t f(F) = f'(F) D_t F.$$

Note that nowhere do we suppose that F have a density; we could end up sometimes with random variables defined a.e. and for which $f'(F)$ is only defined almost everywhere. Assuming that F has a density one can go a step further.

Proposition 7. *[Chain rule 3] Take* $F \in \operatorname{dom}(D)$ *and* $f : \mathbb{R} \to \mathbb{R}$ *Lipschitz (in particular absolutely continuous and a.e. differentiable). Suppose moreover that* F *has a density (wrt Lebesgue measure). Then*

$$D_t f(F) = f'(F) D_t F.$$

Working upwards we can also show the last chain rule, which will in particular allow us to work with polynomials (and hence compute moments).

Proposition 8. *[Chain rule 4] If* $F = \sum_{j=0}^{M} I_j(f_j)$ *be a finite sum of multiple integrals (in particular having density), then*

$$D_t p(F) = p'(F) D_t F$$

for every polynomial $p : \mathbb{R} \to \mathbb{R}$.

3.2 Generator of the Ornstein-Uhlenbeck semigroup L

We take

$$\operatorname{dom}(L) := \left\{ F \in L^2(\sigma(G)) \;:\; \sum_{q=1}^{\infty} q^2 q! \|f_q\|^2_{L^2(\mu^q)} < \infty \right\}.$$

For all $F \in \operatorname{dom}(L)$ we define

$$LF = -\sum_{q=1}^{\infty} q I_q(f_q).$$

For every $F \in L^2(\sigma(G))$ we also define

$$L^{-1}F = -\sum_{q=1}^{\infty} \frac{1}{q} I_q(f_q).$$

This is a pseudo-inverse of the operator L, because

$$LL^{-1}F = F - \mathbb{E}(F).$$

Note that $L^{-1}(F) \in \operatorname{dom}(D)$ and $\operatorname{dom}(L)$ always, because this is just the chaotic expansion of a r.v. whose kernels are I_q/q which can be safely multiplied by q and q^2.

Proposition 9. **[Malliavin integration by parts]** *Assume that $F, G \in L^2(\sigma(G))$ with $\mathbb{E}(F) = 0$ and $G \in \mathrm{dom}(D)$. Then*

$$\mathbb{E}(FG) = \mathbb{E}(\langle DG, -DL^{-1}F\rangle_{L^2(\mu)}).$$

Proof. Again by density arguments we just prove it for $F = I_q(f)$ and $G = I_p(g)$. But then

$$\mathbb{E}(FG) = \delta_{p,q} q! \langle f, g\rangle_{L^2(\mu^q)}$$

and

$$F_t G = pI_{p-1}(g(t, \bullet)).$$

Also we have

$$L^{-1}F = -\frac{1}{q} I_q(f) \text{ and } -D_t L^{-1}F = I_{q-1}(f(t, \bullet))$$

so that, taking expectations, we get

$$\begin{aligned}\mathbb{E}(\langle DG, -DL^{-1}F\rangle_{L^2(\mu)}) &= \int_A \mu(dt)\mathbb{E}[pI_{p-1}(g(t, \bullet))I_{q-1}(f(t, \bullet))] \\ &= \delta_{p,q} p \int_A \mu(dt)(p-1)! \int_{A^{p-1}} g(t, \bar{x}_{p-1}) f(t, \bar{x}_{p-1}) d\mu^{p-1} \\ &= \delta_{p,q} p! \langle f, g\rangle_{L^2(\mu^p)}.\end{aligned}$$

□

Corollary 2. *Assume that $\mathbb{E}(F) = 0$ for $F \in \mathrm{dom}(D)$. Also assume that f is such that the chain rule applies. Then*

$$\begin{aligned}\mathbb{E}(Ff(F)) &= \mathbb{E}(\langle DF, -DL^{-1}F\rangle_{L^2(\mu)}) \\ &= \mathbb{E}(f'(F)\langle DF, -DL^{-1}F\rangle_{L^2(\mu)})\end{aligned}$$

Corollary 3. *Take $F = I_q(f)$ and $n \geq 1$. Then*

$$\begin{aligned}\mathbb{E}(F^{n+1}) = \mathbb{E}(FF^n) &= n\mathbb{E}(F^{n-1}\langle DF, -DL^{-1}F\rangle_{L^2(\mu)}) \\ &= \frac{n}{q}\mathbb{E}(F^{n-1}\|DF\|^2_{L^2(\mu)}),\end{aligned}$$

where $L^{-1}F = -\frac{1}{q}F$. In particular, we have

$$\mathbb{E}(F^4) = \frac{3}{q}\mathbb{E}\left(F^2\|DF\|^2_{L^2(\mu)}\right).$$

Part III
Stein meets Malliavin

Via the Stein's approach, we have already seen that for any integrable random variable $F \in L^1(\mathbb{P})$ we have

$$d_{TV}(F, \mathscr{N}(0,1)) \leq \sup_{f \in \mathscr{F}_{\text{TV}}} \left| \mathbb{E}(Ff(F) - f'(F)) \right|$$

with $\mathscr{F}_{\text{TV}} = \{\|f\| \leq \sqrt{\pi/2}, \|f'\| \leq 2\}$. As was noted before, the supremum is annoying. The following theorem shows that, in the Gaussian framework, things are extremely favorable. Hereon, we assume that G is a Gaussian random measure over $(A, \mathcal{A}, \mu)$, and all random variables are measurable functionals of G.

Theorem 5. *Let $F \in \text{dom}(D)$ with $\mathbb{E}(F) = 0$. Assume that F has a density (to use Proposition 7). Then*

$$d_{TV}(F, \mathscr{N}(0,1)) \leq 2\mathbb{E} \left| 1 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)} \right|.$$

We have thus reduces the problem of controlling the TV distance to the computation of an expectation!

Proof. For every $f \in \mathscr{F}_{\text{TV}}$, and using Proposition 7 we have

$$\begin{aligned} \left| \mathbb{E}(Ff(F) - f'(F)) \right| &= \left| \mathbb{E}(f'(F)(\langle DF, -DL^{-1}F \rangle_{L^2(\mu)} - 1)) \right| \\ &\leq 2\mathbb{E} \left| (\langle DF, -DL^{-1}F \rangle_{L^2(\mu)} - 1) \right|. \end{aligned}$$

Note that to obtain the last inequality in above, we use the fact that $\|f'\|_\infty \leq 2$ for every $f \in \mathscr{F}_{\text{TV}}$. □

Corollary 4. *If $F = I_1(h)$, i.e. $F \sim \mathscr{N}(0, \|h\|^2)$ then*

$$\langle DF, -DL^{-1}F \rangle_{L^2(\mu)} = \|h\|^2.$$

Hence

$$d_{TV}(I_1(h), \mathscr{N}(0,1)) \leq 2 \left| 1 - \|h\|^2 \right|.$$

Corollary 5. *If $F = I_p(h)$ for some $f \in L^2_{sym}(\mu^p)$, then $\langle DF, -DL^{-1}F \rangle_{L^2(\mu)} = \frac{1}{p}\|DF\|^2$, and hence*

$$\begin{aligned} d_{TV}(F, \mathscr{N}(0,1)) &\leq 2\mathbb{E} \left| 1 - \frac{1}{p}\|DF\|^2 \right| \\ &\leq 2\sqrt{\mathbb{E}\left(1 - \frac{1}{p}\|DF\|^2\right)^2}. \end{aligned}$$

Therefore, for a sequence $\{F_n = I_p(h_n)\}_{n\geq 1}$ *of multiple integrals of a fixed order* $p \geq 2$ *such that* $\mathbb{E}(F_n)^2 \to 1$*, we have*

$$\|DF_n\|^2_{L^2(\mu)} \xrightarrow{L^2(\mathbb{P})} p \quad \Longrightarrow \quad F_n \overset{law}{\to} \mathscr{N}(0,1).$$

Corollary 6. *If* $F = I_p(h)$ *for some* $f \in L^2_{sym}(\mu^p)$*, and* $\mathbb{E}(F^2) = \sigma^2$*. Then*

$$d_{TV}(F, \mathscr{N}(0,\sigma^2)) \leq \frac{2}{\sigma^2}\sqrt{\mathbb{E}\left(\sigma^2 - \frac{1}{p}\|DF\|^2\right)^2}. \tag{12}$$

Hence,

$$d_{TV}(F, \mathscr{N}(0,1)) \leq 2|1-\sigma^2| + \frac{2}{\sigma^2}\sqrt{Var\left(\frac{1}{p}\|DF\|^2\right)}. \tag{13}$$

Moreover, for a sequence $\{F_n = I_p(h_n)\}_{n\geq 1}$ *of multiple integrals of a fixed order* $p \geq 2$ *such that* $\mathbb{E}(F_n^2) \to \sigma^2 > 0$*, we have*

$$\|DF_n\|^2_{L^2(\mu)} \xrightarrow{L^2(\mathbb{P})} p \times \sigma^2 \quad \Longrightarrow \quad F_n \overset{law}{\to} \mathscr{N}(0,\sigma^2).$$

Proof. For the claim (12), use the facts that if $N \sim \mathscr{N}(0,\sigma^2)$, then $\frac{N}{\sigma} \sim \mathscr{N}(0,1)$ together with

$$d_{\mathrm{TV}}(F, \mathscr{N}(0,\sigma^2)) = d_{\mathrm{TV}}(\frac{F}{\sigma}, \frac{N}{\sigma}).$$

Now, just left to apply Corollary 5. For the claim (13), use the triangular inequality, Corollary 4 and the relation (12). Note that when $\mathbb{E}(F^2) = \sigma^2$, then $\mathbb{E}(\frac{1}{p}\|DF\|) = \mathbb{E}(F^2)$, by using the Malliavin integration by part formula. □

Now, we are ready to state that approximating random variables in a fixed Wiener chaos by Gaussian is a nontrivial enterprise.

Theorem 6 (Nourdin–Peccati (2009) [3])**.** *Let* $p \geq 2$ *and* $f \in L^2_{sym}(\mu^p) \neq 0$*. Take* $F = I_p(f)$*. Then*

$$\begin{aligned} d_{TV}(F, \mathscr{N}(0,1)) &\leq 2\left|1 - \mathbb{E}(F^2)\right| + 2\sqrt{Var\left(\frac{1}{p}\|DF\|^2\right)} \\ &\leq 2\left|1 - \mathbb{E}(F^2)\right| + 2\sqrt{\frac{p-1}{3p}}\sqrt{\mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2}. \end{aligned} \tag{14}$$

[3] Nourdin, I., Peccati, G. (2009) Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 45, no. 1-2, 75-118

Note that $\sqrt{\frac{p-1}{3p}} \leq \frac{2}{\sqrt{3}}$ *and is thus independent of* p.

In particular in order to have a CLT in a fixed Wiener chaos it suffices to control the fourth moments (whereas before this discovery, one had to show convergence of all moments!).

Corollary 7 (Nualart–Peccati (2005))**.** *Assume that* $F_n = I_p(f_n)$ *for* $p \geq 2$ *such that* $\mathbb{E}(F_n^2) \to 1$*. Then*

$$F_n \to \mathscr{N}(0,1)$$

in TV distance (and in particular, in distribution) if and only if

$$\mathbb{E}(F_n^4) \to 3 = \mathbb{E}(\mathscr{N}(0,1)^4).$$

Remark 10. *The "If" part of Corollary 7 is a consequence of* (14)*. On the other hand if* $\mathbb{E}(F_n^2) \to 1$ *and* $F_n \to \mathscr{N}(0,1)$ *in distribution, then for all* $r \geq 2$ *the sequence* $\{\mathbb{E}(|F_n|^r)\}$ *is bounded by hypercontractivity and thus, for all* $r \geq 3$*, we have*

$$\mathbb{E}(F_n^r) \to \mathbb{E}(\mathscr{N}(0,1)^r).$$

Proof of Theorem 6. We aim to prove that

$$\mathrm{Var}(\frac{1}{p}\|DF\|^2) \leq \frac{p-1}{3p}\left\{\mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2\right\}.$$

We are going to use the formula

$$\mathbb{E}(F^4) = \frac{3}{p}\mathbb{E}(F^2\|DF\|^2_{L^2(\mu)}).$$

The whole proof relies on the derivation of the chaotic decompositions of the rv's of interest. Now, for $F = I_p(f)$ recall that $D_t F = pI_{p-1}(f(t,\cdot))$ to write (using product formula)

$$\begin{aligned}
\frac{1}{p}\|DF\|^2 &= p\int_A \mu(dt)(I_{p-1}(f(t,\cdot)))^2 \\
&= p\int_A \mu(dt)\sum_{r=0}^{p-1} r!\binom{p-1}{r}^2 I_{2(p-1)-2r}(f(t,\cdot)\widetilde{\otimes_r} f(t,\cdot)) \\
&= p\sum_{r=0}^{p-1} r!\binom{p-1}{r}^2 I_{2p-2(r+1)}(f\widetilde{\otimes_{r+1}} f)
\end{aligned}$$

where to obtain this we have used a stochastic integral version of Fubini's theorem. Pursuing with a change of summation variables we deduce

$$\begin{aligned}\frac{1}{p}\|DF\|^2 &= p\sum_{r=1}^{p}(r-1)!\binom{p-1}{r-1}^2 I_{2p-2r}(\widetilde{f\otimes_r f})\\ &= p!\|f\|^2 + p\sum_{r=1}^{p-1}(r-1)!\binom{p-1}{r-1}^2 I_{2p-2r}(\widetilde{f\otimes_r f}).\end{aligned}$$

Note that $p!\|f\|^2 = \mathbb{E}(F^2)$, and we get

$$\frac{1}{p}\|DF\|^2 = \mathbb{E}(F^2) + p\sum_{r=1}^{p-1}(r-1)!\binom{p-1}{r-1}^2 I_{2p-2r}(\widetilde{f\otimes_r f}).$$

Since we have already shown that $\frac{1}{p}\mathbb{E}(\|DF\|^2) = \mathbb{E}(F^2)$ we deduce, by orthogonality of Wiener chaoses, our first estimate

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{p}\|DF\|^2\right) &= \sum_{r=1}^{p-1} p^2(r-1)!^2\binom{p-1}{r-1}^4(2p-2r)!\|\widetilde{f\otimes_r f}\|^2\\ &= \frac{1}{p^2}\sum_{r=1}^{p-1} r^2 r!^2\binom{p}{r}^4(2p-2r)!\|\widetilde{f\otimes_r f}\|^2.\end{aligned} \tag{15}$$

Now we also have

$$F^2 = \sum_{r=0}^{p} r!\binom{p}{r}^2 I_{2p-2r}(\widetilde{f\otimes_r f}) = p!\|f\|^2 + \sum_{r=0}^{p-1} r!\binom{p}{r}^2 I_{2p-2r}(\widetilde{f\otimes_r f}). \tag{16}$$

We can now compute the fourth moment

$$\begin{aligned}\mathbb{E}(F^4) &= 3\mathbb{E}(F^2\times\frac{1}{p}\|DF\|^2)\\ &= 3\mathbb{E}(F^2)^2 + 3\sum_{r=1}^{p-1} pr!(r-1)!\binom{p}{r}^2\binom{p-1}{r-1}^2(2p-2r)!\|\widetilde{f\otimes_r f}\|^2.\end{aligned} \tag{17}$$

Therefore,

$$
\begin{aligned}
\mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2 &= 3p\sum_{r=1}^{p-1} r!(r-1)!\binom{p}{r}^2\binom{p-1}{r-1}^2(2p-2r)!\|\widetilde{f\otimes_r f}\|^2 \\
&= \frac{3}{p}\sum_{r=1}^{p-1} rr!^2\binom{p}{r}^4(2p-2r)!\|\widetilde{f\otimes_r f}\|^2.
\end{aligned}
\tag{18}
$$

Comparing the sums in (15) and (18) we recover the desired inequality

$$
\mathrm{Var}\left(\frac{1}{p}\|DF\|^2\right) \leq \frac{p-1}{3p}\left\{\mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2\right\}. \tag{19}
$$

Note that the following estimate is also in order:

$$
\frac{p-1}{3p}\left\{\mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2\right\} \leq (p-1)\mathrm{Var}\left(\frac{1}{p}\|DF\|^2\right).
$$

□

Remark 11. The proof of Theorem 6 reveals that for any multiple integral $F = I_p(f)$ of order $p \geq 2$ the following surprising fact takes place:

$$
\kappa_4(F) = \mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2 > 0.
$$

Recall that when $p = 1$, then $\kappa_4(F) = 0$. What about $\kappa_6(F) \geq 0$? This question is still totally open in its full generality! (see [2] for more detailed discussion).

Remark 12. If one instead using the clever Malliavin relation

$$
\mathbb{E}(F^4) = \frac{3}{p}\mathbb{E}(F^2\|DF\|^2_{L^2(\mu)})
$$

by expanding F^4 on Wiener chaoses to compute $\mathbb{E}(F^4)$ what we will end up with (take into account 16)

$$
\begin{aligned}
\mathbb{E}(F^4) = \mathbb{E}(F^2\times F^2) &= \sum_{r=0}^{p} r!^2\binom{p}{r}^4(2p-2r)!\|\widetilde{f\otimes_r f}\|^2 \\
&= p!^2\|f\|^4 + (2p)!\|\widetilde{f\otimes_0 f}\|^2 + \sum_{r=1}^{p-1} r!^2\binom{p}{r}^4(2p-2r)!\|\widetilde{f\otimes_r f}\|^2 \\
&= \mathbb{E}(F^2)^2 + (2p)!\|\widetilde{f\otimes_0 f}\|^2 + \sum_{r=1}^{p-1} r!^2\binom{p}{r}^4(2p-2r)!\|\widetilde{f\otimes_r f}\|^2.
\end{aligned}
$$

Hence, we have the presence of the norm of the zero-contraction $\|\widetilde{f \otimes_0 f}\|^2$, which in fact never appears in the expansion $\text{Var}\left(\frac{1}{p}\|DF\|^2\right)$. Therefore, using this approach, one needs to represent the norm of the zero-contraction $\|\widetilde{f \otimes_0 f}\|^2$ in terms of the norm of other non-zero contraction to be able to do comparison. Fortunatly, this can be done and is the message of the next exercise. Here, we highlight that the appearance of norms (inner products) of zero-contractions involving the kernel f is in fact one of the main obstacles in front towards generalization of the Malliavin-Stein approach for non-Gaussian approximation using product formula as the main tools. A typical example is when the target distribution is of the form $N_1 \times N_2$ and $N_1, N_2 \sim \mathscr{N}(0,1)$ are independent. More precisely, we want to understand the possibility of convergence

$$I_p(f_n) \stackrel{\text{law}}{\to} N_1 \times N_2$$

in terms of convergence of finitely many moments/cumulants. Let's stress that also the lack of the Stein's method for the target distribution $N_1 \times N_2$! So, one needs to take a different paths. Hopefully, this we can consider in more details in the last week of the course.

Exercise 6. ***(a)*** *Show that*

$$(2p)!\|\widetilde{f \otimes_0 f}\|^2 = 2(p!)^2\|f\|^4 + p!^2 \sum_{r=1}^{p-1} \binom{p}{r}^2 \|f \otimes_r f\|^2.$$

(b) *Use part (a) to show that*

$$\mathbb{E}(F^4) - 3\mathbb{E}(F^2)^2 = p!^2 \sum_{r=1}^{p-1} \binom{p}{r} \left\{ \|f \otimes_r f\|^2 + \binom{2p-2r}{p-r} \|\widetilde{f \otimes_r f}\|^2 \right\}.$$

Corollary 8. *Assume that $\{F_n = I_p(f_n)\}_{n \geq 1}$ is a sequence of multiple integrals of fixed order $p \geq 2$ such that $\mathbb{E}(F_n^2) \to 1$. Then the following statements are equivalent.*

(a) $F_n \stackrel{law}{\to} \mathscr{N}(0,1)$.

(b) $\mathbb{E}(F_n^4) \to 3 = \mathbb{E}(N^4)$, *where* $N \sim \mathscr{N}(0,1)$.

(c) $\|DF_n\|^2 \stackrel{L^2(\mathbb{P})}{\to} p$.

(d) $\|\widetilde{f_n \otimes_r f_n}\|_{L^2(\mu^{2p-2r})} \to 0$ *for all* $1 \leq r \leq p-1$.

(e) $\|f_n \otimes_r f_n\|_{L^2(\mu^{2p-2r})} \to 0$ *for all* $1 \le r \le p-1$.

Proof. We are just left to show the implication $(e) \Rightarrow (a)$. Statement (e), and $\|f_n \widetilde{\otimes}_r f_n\|_{L^2(\mu^{2p-2r})} \le \|f_n \otimes_r f_n\|_{L^2(\mu^{2p-2r})}$, together with Exercise 6, part (b), imply that in fact $\kappa_4(F_n) \to 0$. So, we are done. □

Remark 13. *The computations in the proof show that there exists a constant* $c := c(p) > 0$ *only depending on* p *for which*

$$\begin{aligned} d_{TV}(I_p(f), \mathscr{N}(0,1)) &\le c(p) \max_{r=1,\ldots,p-1} \left\{ \|f \widetilde{\otimes}_r f\|_{L^2(\mu^{2p-2r})} \right\} \\ &\le c(p) \max_{r=1,\ldots,p-1} \left\{ \|f \otimes_r f\|_{L^2(\mu^{2p-2r})} \right\} \end{aligned} \tag{20}$$

The estimate in (20) *is that which is most used in practical situations since it is easier to estimate contractions rather than moments.*

Problem 2. *Can the maximum in* (20) *attain in, at least, one running index* r*? maybe middle running index ? A positive answer to this problem decrease drastically the complexity of checking CLTs on Wiener chaos.*

Remark 14. We have already seen that for a sequence $\{F_n = I_p(f_n)\}_{n\ge1}$ the convergences $\mathbb{E}(F_n^2) \to 1$ and $\mathbb{E}(F_n^4) \to 3 = \mathbb{E}(N^4)$ imply F_n converges in distribution towards $N \sim \mathscr{N}(0,1)$. It is a natural question to ask that why the fourth moment? In other words, do the other even moments do the same job as the fourth moment ? For example, do the convergences $\mathbb{E}(F_n^2) \to 1$ and $\mathbb{E}(F_n^6) \to 15 = \mathbb{E}(N^6)$ imply the convergence in distribution towards $\mathscr{N}(0,1)$? In fact, it was an open problem in this modern domain of probabilistic approximations since appearance of the Nualart–Peccati fourth moment theorem in 2005. It can be easily seen that if one expands F^6 (where $F = I_p(f)$) on Wiener chaoses to use product formula, then immediately very involved expressions such as inner products of three copies of the kernel f appears and makes this approach completely hopeless. Recently, in [2], we create a novel approach using Markov triplet (one of the topics we will briefly study during the last week of the course) to give a positive answer to this rather long-standing open problem. In fact, we prove that for every $r \ge 2$

$$\mathbb{E}(F_n^2) \to 1 \,\&\, \mathbb{E}(F^{2r}) \to \mathbb{E}(N^{2r}) = (2r-1)!! \quad \Longleftrightarrow \quad F_n \overset{\text{law}}{\to} \mathscr{N}(0,1). \tag{21}$$

Another interesting question to ask here is: can one replace the convergence of the second moment in (21) with convergence of some other even moments? For example,

$$\mathbb{E}(F_n^6) \to \mathbb{E}(N^6) \quad \& \quad \mathbb{E}(F_n^8) \to \mathbb{E}(N^8) \quad \overset{?}{\Longrightarrow} \quad F_n \overset{\text{law}}{\to} \mathscr{N}(0,1). \quad (22)$$

Using the techniques developed in [2], we could also give a positive answer to (22). However, we failed to prove it for convergences of the eighth and tenth moments. Theses observation lead us to the following interesting conjecture (any idea?).

Conjecture 1. *Let $\{F_n = I_p(f_n)\}_{n\geq 1}$, and $p \geq 2$. Then for every $r \neq s$*

$$\mathbb{E}(F_n^{2r}) \to \mathbb{E}(N^{2r}) = (2s-1)!! \quad \& \quad \mathbb{E}(F^{2s}) \to \mathbb{E}(N^{2s}) = (2s-1)!! \\ \Longleftrightarrow \quad F_n \overset{law}{\to} \mathscr{N}(0,1). \quad (23)$$

3.3 MULTIDIMENSIONAL CASE

Let $d \geq 2$, and fix d natural numbers $1 \leq p_1 \leq p_2 \leq \cdots \leq p_d$. Consider a sequence of d-dimensional random vectors of the form

$$F_n = (F_n^1, \cdots, F_n^d) = (I_{p_1}(f_n^1), \cdots, I_{p_d}(f_n^d)). \quad (24)$$

Our aim in this section is to prove the following multidimensional version of the fourth moment theorem due to Peccati–Tudor (2005).

Theorem 7. *Let F_n be a sequence of d-dimensional random vectors of the form (25) such that*

$$\lim_{n\to\infty} \mathbb{E}(F_n^i \times F_n^j) = \delta_{ij}, \qquad 1 \leq i,j \leq d.$$

Then the following statements are equivalent.

(a) $F_n^i \overset{law}{\to} \mathscr{N}(0,1)$, *for all* $1 \leq i \leq d$.

(b) $\mathbb{E}(F_n^i)^4 \to 3$, *for all* $1 \leq i \leq d$.

(c) $\|DF_n^i\|^2 \overset{L^2(\mathbb{P})}{\to} p_i$, *for all* $1 \leq i \leq d$.

(d) $\|f_n^i \otimes_r f_n^i\|_{L^2(\mu^{2p_i-2r})} \to 0$, *for all* $1 \leq i \leq d$, *and* $r = 1, \cdots, p_i - 1$.

(e) $F_n \overset{law}{\to} \mathscr{N}_d(0, I_d)$.

In other words, Theorem 7 tells us that the component-wise convergence to Gaussian distributions implies the joint convergence of the vector to the multidimensional Gaussian. There are different ways to prove Theorem 7. Here, we follow the path was developed by Nualart & Ortiz-Latorre. Their strategy mainly consists of showing that the characteristic function of any adherence value in distribution satisfies in the same ordinary differential equation as the characteristic function of the d-dimensional Gaussian random variable. The advantage of their approach compare to multidimensional Stein's method (which is more involved compare to one dimensional version) is its simplicity and as drawback this approach is not quantitative, and hence one can not provide any rate of convergence. The interested reader can consult [5, Chapter 6] for a proof of Theorem 7 using multidimensional Stein's method. The token of the main part of Theorem 7 can be decoded using the following lemma in which the Malliavin derivative matrix

$$\Gamma_n = (\Gamma_n^{i,j})_{1\leq i,j\leq d} = \big(\langle DF_n^i, DF_n^j\rangle_{L^2(\mu)}\big)_{1\leq i,j\leq d}$$

plays an essential role. The Malliavin derivative matrix Γ is in the core of the studies of regularities of laws of random vectors (see [7, Chapter 2]). In the next lemma, we will use again heavily the specific structure of the underlying random variables.

Lemma 4. *Let*

$$F_n = (F_n^1, \cdots, F_n^d) = (I_{p_1}(f_n^1), \cdots, I_{p_d}(f_n^d)) \tag{25}$$

such that for every $1 \leq i,j \leq d$, $\mathbb{E}(F_n^i \times F_n^j) \to \delta_{i,j}$. *Then*

$$\|DF_n^i\|^2 \overset{L^2(\mathbb{P})}{\to} p_i \quad \Longrightarrow \quad \Gamma_n^{i,j} = \langle DF_n^i, DF_n^j\rangle_{L^2(\mu)} \overset{L^2(\mathbb{P})}{\to} \sqrt{p_i p_j}\,\delta_{i,j}.$$

Proof. We need to show that for any $i<j$ (and so $p_i \leq p_j$) we have

$$\lim_{n\to\infty} \mathbb{E}\left(\langle DF_n^i, DF_n^j\rangle^2_{L^2(\mu)}\right) = 0.$$

Using exercise 2, part (a), we know that

$$\begin{aligned}\mathbb{E}\left(\langle DF_n^i, DF_n^j\rangle^2\right)_{L^2(\mu)} &= \sum_{r=1}^{p_i} \frac{(p_i!p_j!)^2}{((p_i-r)!(p_j-r)!(r-1)!)^2}\|f_n^i \widetilde{\otimes}_r f_n^j\|^2\\ &\leq \sum_{r=1}^{p_i} \frac{(p_i!p_j!)^2}{((p_i-r)!(p_j-r)!(r-1)!)^2}\|f_n^i \otimes_r f_n^j\|^2\end{aligned}$$

So, we are left to show that $\|f_n^i \otimes_r f_n^j\|^2 \to 0$ for all $1 \le r \le p_i$. Using the very definition of the contraction, Fubini's theorem, and Cauchy–Schwarz inequality, we can write

$$\begin{aligned}\|f_n^i \otimes_r f_n^j\|^2 &= \langle f_n^i \otimes_r f_n^j, f_n^i \otimes_r f_n^j, \rangle_{L^2(\mu^{p_i+p_j-2r})}\\ &= \langle f_n^i \otimes_{p_i-r} f_n^i, f_n^j \otimes_{p_j-r} f_r^j \rangle_{L^2(\mu^{2r})}\\ &\le \|f_n^i \otimes_{p_i-r} f_n^i\| \times \|f_n^j \otimes_{p_j-r} f_n^j\|.\end{aligned} \tag{26}$$

Case (a): if $r = p_i = p_j$, then $\|f_n^i \otimes_r f_n^j\|^2 = \Big(\mathbb{E}(F_n^i \times F_n^j)\Big)^2 \to 0$ by assumption. Case (b): if $1 \le r \le p_i - 1$, then assumption $\|DF_n^i\|^2 \overset{L^2(\mathbb{P})}{\to} p_i$ implies that $F_n^i \overset{\text{law}}{\to} \mathscr{N}(0,1)$ and therefore $\|f_n^i \otimes_r f_n^i\|^2 \to 0$ for all $1 \le r \le p_i - 1$. Hence, the right hand side inequality (26) tends to zero. Case (c): if $r = p_i < p_j$. In this case, the right hand side of (26) takes the form

$$\begin{aligned}\|f_n^i \otimes_{p_i-r} f_n^i\| \times \|f_n^j \otimes_{p_j-r} f_n^j\| &= \|f_n^i\|^2 \times \|f_n^j \otimes_{p_j-r} f_n^j\|\\ &= \mathbb{E}(F_n^i)^2 \times \|f_n^j \otimes_{p_j-r} f_n^j\| \to 0,\end{aligned}$$

Because $\mathbb{E}(F_n^i)^2 \to 1$ and so bounded and $\|f_n^j \otimes_{p_j-r} f_n^j\| \to 0$. □

Proof. Proof of Theorem 7. It is enough to prove the implication $(c) \Rightarrow (e)$. Since $\mathbb{E}(F_n^i \times F_n^j) \to \delta_{i,j}$, for $i = j$, this implies that $\sup_{n\ge1} \mathbb{E}(F_n^i)^2 < +\infty$. Therefore, the sequence $\{F_n\}_{n\ge1}$ is tight, and so it is enough to show that the limit of any convergence in distribution subsequence $\{F_{n_k}\}_{n\ge1}$ is in fact $\mathscr{N}_d(0, I_d)$. To this end, assume that $F_{n_k} \overset{\text{law}}{\to} F_\infty$, as $k \to \infty$ for some random vector $F_\infty = (F_\infty^1, \cdots, F_\infty^d)$. By our assumptions, first we have that $F_\infty^i \in L^2(\Omega)$ for all $1 \le i \le d$, and moreover $\mathbb{E}(F_\infty^i \times F_\infty^j) = 0$ if $i \ne j$. Now, let's denote the characteristic function $\varphi_n(t) = \mathbb{E}(e^{i\langle t, F_n\rangle_{\mathbb{R}^d}})$ for $t \in \mathbb{R}^d$. Then $\varphi_{n_k}(t) \to \varphi(t)$ for any t, where φ_∞ is the characteristic function of F_∞. Note that the fact that $F_\infty^j \in L^2(\Omega)$ implies that the partial derivatives $\frac{\partial}{\partial t_j}\varphi_\infty = i\mathbb{E}(F_\infty^j e^{i\langle t, F_\infty\rangle_{\mathbb{R}^d}})$ are well defined. Now, continuous mapping theorem tells us that

$$F_{n_k} e^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}} \overset{\text{law}}{\to} F_\infty^j e^{i\langle t, F_\infty\rangle_{\mathbb{R}^d}}. \tag{27}$$

Note that the sequence in the left hand side of (27) is bounded in $L^2(\Omega)$ and so uniformly integrable. Hence, for all $1 \le j \le d$, and $t \in \mathbb{R}^d$, as $k \to \infty$:

$$\frac{\partial}{\partial t_j}\varphi_{n_k}(t) = i\mathbb{E}(F_{n_k}^j e^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}}) \to \frac{\partial}{\partial t_j}\varphi_\infty = i\mathbb{E}(F_\infty^j e^{i\langle t, F_\infty\rangle_{\mathbb{R}^d}}). \tag{28}$$

On the other hand side, using integration by part formula: (note that $\mathbb{E}(F^j_{n_k}) = 0$)

$$\begin{aligned}
\mathbb{E}\left(F^j_{n_k} e^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}}\right) &= \mathbb{E}\left(\langle De^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}}, -DL^{-1}F^j_{n_k}\rangle\right) \\
&= -\frac{i}{p_j}\sum_{l=1}^{d} t_l\, \mathbb{E}\left(e^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}} \mathbb{E}(\langle DF^l_{n_k}, DF^j_{n_k}\rangle)\right) \\
&= -\frac{i}{p_j}\sum_{l=1}^{d} t_l\, \mathbb{E}\left(e^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}} \Gamma^{l,j}_{n_k}\right).
\end{aligned}$$

Therefore

$$\frac{\partial}{\partial t_j}\varphi_{n_k}(t) = -\frac{1}{p_j}\sum_{l=1}^{d} t_l\, \mathbb{E}\left(e^{i\langle t, F_{n_k}\rangle_{\mathbb{R}^d}} \Gamma^{l,j}_{n_k}\right).$$

Hence, Lemma 4 implies that the charactrisitic function φ_∞ satisfies in equation

$$\frac{\partial}{\partial t_j}\varphi_\infty(t) = -t_j \varphi_\infty(t),$$

for all $j = 1, \cdots, d$ and $t \in \mathbb{R}^d$. Therefore, the only possibility for φ_∞ is to be the characteristic function of $\mathscr{N}_d(0, I_d)$. □

We finish this section with the following very general result.

Theorem 8. *Let $\{F_n\}_{n\geq 1}$ be a square-integrable sequence with the following chaos decompositions: for every $n \geq 1$*

$$F_n = \sum_{p=1}^{\infty} I_p(f_{n,p}). \tag{29}$$

In addition, assume the following:

(a) for all $p \geq 1$, we have $p!\|f_{n,p}\|^2 \to \sigma_p^2$.

(b) $\sum_{p\geq 1}\sigma_p^2 < +\infty$.

(c) for all $p \geq 2$ and every $r = 1, \cdots, p-1$, we have $\|f_{n,p} \otimes_r f_{n,p}\| \to 0$, as $n \to \infty$.

(d)

$$\lim_{N\to\infty}\sup_{n\geq 1}\sum_{p=N+1}^{\infty} p!\,\|f_{n,p}\|^2 = 0.$$

Then we have $F_n \overset{law}{\to} \mathscr{N}(0,\sigma^2)$.

Proof. For all $N \geq 1$, set

$$\begin{aligned} F_{n,N} &= \sum_{p=1}^{N} I_p(f_{n,p}) \\ G_N &\sim \mathscr{N}(0,\sigma_1^2+\cdots+\sigma_N^2) \\ G &\sim \mathscr{N}(0,\sigma^2). \end{aligned}$$

Therefore, for any $t \in \mathbb{R}$:

$$\begin{aligned} \left|\mathbb{E}(e^{itF_n}) - \mathbb{E}(e^{itG})\right| &\leq \left|\mathbb{E}(e^{itF_n}) - \mathbb{E}(e^{itF_{n,N}})\right| \\ &\quad + \left|\mathbb{E}(e^{itF_{n,N}}) - \mathbb{E}(e^{itG_N})\right| \\ &\quad + \left|\mathbb{E}(e^{itG_N}) - \mathbb{E}(e^{itG})\right| := a_{n,N} + b_{n,N} + c_N. \end{aligned}$$

Note that

$$c_N = \left|e^{-\frac{t^2}{2}(\sigma_1^2+\cdots+\sigma_N^2)} - e^{-\frac{t^2}{2}\sigma^2}\right| \leq \frac{t^2}{2}\left|\sigma^2 - \sum_{i=1}^{N}\sigma_i^2\right| \to 0,$$

as $N\to\infty$, because of assumption (b). Moreover,

$$\begin{aligned} \sup_{n\geq 1} a_{n,N} &= \sup_{n\geq 1}\left|\mathbb{E}(e^{itF_n}) - \mathbb{E}(e^{itF_{n,N}})\right| \\ &\leq |t|\sup_{n\geq 1}\mathbb{E}|F_n - F_{n,N}| \leq |t|\sqrt{\sup_{n\geq 1}\mathbb{E}(F_n - F_{n,N})^2} \\ &\leq |t|\sqrt{\sup_{n\geq 1}\sum_{p\geq N+1}\sigma_p^2} \to 0, \end{aligned}$$

by assumption (d). Hence, for $\varepsilon > 0$, choose N large enough so that $\sup_{n\geq 1} a_{n,N} \leq \varepsilon/3$ and $c_N \leq \varepsilon/3$. Also, according to Peccati–Tudor multidimensional version of the fourth moment theorem, we have in fact that, as $n \to \infty$,

$$(I_1(f_{n,1}),\cdots,I_N(f_{n,N}))\overset{\text{law}}{\to}\mathscr{N}_N\big(0,\mathrm{diag}(\sigma_1^2,\cdots,\sigma_N^2)\big).$$

Therefore, $F_{n,N}=\sum_{p=1}^N I_p(f_{n,p})\overset{\text{law}}{\to}\mathscr{N}(0,\sigma_1^2+\cdots+\sigma_N^2)$. Hence, $b_{n,N}\le\varepsilon/3$ if n is large enough. □

References

[1] Azmoodeh, E., Campese, S., Poly, G. (2014) Fourth Moment Theorems for Markov Diffusion Generators. *J. Funct. Anal.* 266, 4, 2341-2359.

[2] Azmoodeh, E., Malicet, D., Mijoule, G., Poly, G. (2014) Generalization of the Nualart-Peccati criterion. *To appear in: The Annals of Probability.*

[3] Azmoodeh, E., Peccati, G., Poly, G. (2014). Convergence towards linear combinations of chi-squared random variables: a Malliavin-based approach. *To appear in: Séminaire de Probabilités* (Special volume in memory of Marc Yor)

[4] Chen, L. H. Y., Goldstein, L., Shao, Qi-Man (2011) *Normal approximation by Stein's method.* Probability and its Applications. Springer.

[5] Nourdin, I., Peccati, G. (2012) *Normal Approximations Using Malliavin Calculus: from Stein's Method to Universality.* Cambridge Tracts in Mathematics. Cambridge University.

[6] Nourdin, I., Peccati, G. (2009) Noncentral convergence of multiple integrals. *Ann. Probab.* 37, no. 4, 1412-1426.

[7] Nualart D. (2006) The Malliavin Calculus and Related Topics. *Probability and its Application.* Springer.

[8] Peccati, G., Taqqu, M. (2011) Wiener Chaos: Moments, Cumulants and Diagrams. A survey with computer implementation. *Bocconi University Press.* Springer.