The Stein-Dirichlet-Malliavin method

Laurent Decreusefond

THE STEIN-DIRICHLET-MALLIAVIN METHOD

L. DECREUSEFOND

Abstract. The Stein's method is a popular method used to derive upper-bounds of distances between probability distributions. It can be viewed, in certain of its formulations, as an avatar of the semi-group or of the smart-path method used commonly in Gaussian analysis. We show how this procedure can be enriched by Malliavin calculus leading to a functional approach valid in infinite dimensional spaces.

1. Introduction

Distances between probability or probability metrics is a very old topic since it is rich of a wide range of applications. As mathematical objects, it is natural to define a metric topology on spaces of probability measures. As modeling objects, it is natural to compare probability measures which appear in the mathematical representations of random phenomena. This topic has at least three facets: The diverse definitions of probability metrics which are tailored for each applications; the computations and comparisons of these different distances for the widest possible range of situations and at last, the applications which go from mathematical considerations like functional inequalities to more practical results of rate of convergence of stochastic algorithms. The Figure 1 shows a partial view of the different aspects of this subject.

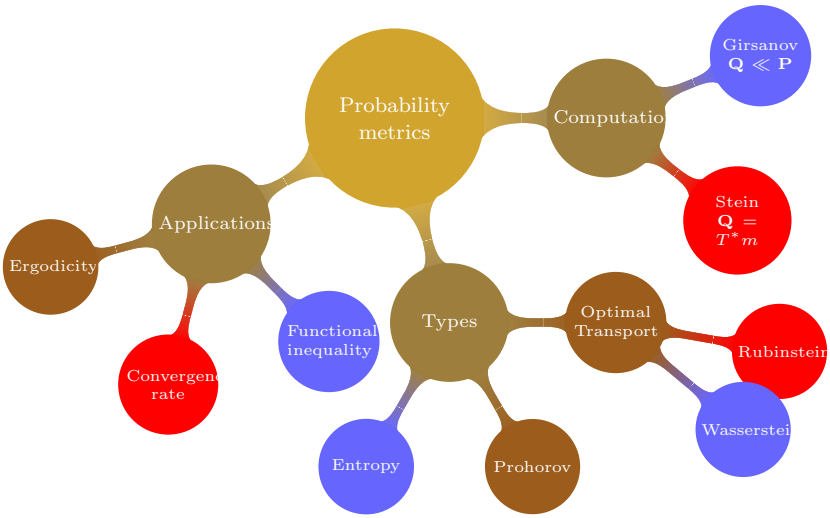

Figure 1. Mindmap

A few words are in order to explain the blue and red colors. For the computations of distances between measures μ and ν, we need to impose some relationships between these two measures. Absolute continuity is one very frequent type of relationships between two measures. The Radon-Nykodim theorem gives a precious tool to estimate divergence-like and Wasserstein distances (see for instance [15] for

such an application). One may also reverse the point of view: Given a positive function F, compare the μ and $\nu = F\,\mathrm{d}\mu$ to obtain some precious functional inequalities on F (see [1]). These results thus belong to the same *spirit* and are colored in blue. Another natural way to put a structure between two measures is to have a map which transforms a known measure into another one and to compare this transformed measure to a reference probability. This is exactly the framework in which the Stein's method performs well if we consider Kantorovitch-Rubinstein type distances (defined below). Typical applications of these form of distances are to give the convergence rates of celebrated theorem like CLT or Berry-Esseen Theorem or of random algorithms [25]. The links between these different points justify that they are all colored in red.

This paper is a rather informal introduction to the Stein-Dirichlet-Malliavin method (SDM for short henceforth). This is an extension of the classical Stein's method, enriched by the structure given by Dirichlet forms and Malliavin calculus. We hope that this new point of view will lead to more systematic proofs of convergence, extending their applicability. The price to pay is to master some new concepts from Malliavin calculus like the gradient and its associated adjoint. That is why we tried to maintain the technicalities at the lowest possible level, insisting more on the ideas at play.

We first show the different kinds of probability metrics that exist in the literature. We do not pretend to be exhaustive but aim to point out to the wide diversity of possible definitions. In Section 2, we establish the principles of the SDM method and show how it can be applied to the Poisson-Gaussian convergence. We then explain how to construct the necessary structures to extend this procedure to infinite dimensional spaces. In Section 4, Edgeworth expansions are obtained by iterating the previous procedure as often as desired.

2. Taxonomy of probability metrics

In what follows, all the probability measures are defined on Polish spaces denoted either by $\mathfrak{E}$ or $\mathfrak{F}$, whose borelian σ-fields is $\mathfrak{B}(\mathfrak{E})$, respectively $\mathfrak{B}(\mathfrak{F})$. There are several notions of metrics between probability measures. An interesting survey of the main variants and their mutual relationships can be found in [17]. Each of one is often adapted to a particular purpose. They can roughly and partly be classified in three types. The first one is the so-called Prokhorov distance.

$$\mathrm{Dist}_{\mathrm{Pro}}(\mathbf{P},\mathbf{Q}) = \inf\Big\{\epsilon>0, \mathbf{P}(A)\le \mathbf{Q}(A^\epsilon)+\epsilon \text{ for all } A\in\mathfrak{B}(\mathfrak{E})\Big\},$$

where A^ϵ is the ϵ-neighborhood of A defined by $A^\epsilon = \{y\in\mathfrak{E},\ \exists x\in A, d(x,y)\le\epsilon\}$. This distance is crucial as its associated topology is precisely the topology of the convergence in distribution, i.e. we have the following theorem which can be found in [13].

Theorem 1. *A sequence $(\mathbf{P}_n,\, n\ge 1)$ of probability measures converges weakly to $\mathbf{P}$ if and only if $Dist_{Pro}(\mathbf{P}_n,\mathbf{P})$ tends to 0 as n goes to ∞.*

Unfortunately, this distance is hardly computable and that justifies the search for alternative and more tractable definitions. A vast category of probability metrics is represented by the f-divergence defined as follows.

Definition 1. *Let f be a convex function such that $f(1)=0$. Then, for two probability measures $\mathbf{P}$ and $\mathbf{Q}$ on a Polish space $\mathfrak{E}$,*

$$D_f(\mathbf{Q}\,|\,\mathbf{P}) = \begin{cases} \displaystyle\int_{\mathfrak{E}} f\left(\frac{d\mathbf{Q}}{d\mathbf{P}}\right)\, d\mathbf{P} & \text{if } \mathbf{Q}\ll\mathbf{P},\\ \infty & \text{otherwise.}\end{cases}$$

For instance, if we choose $f = t \ln t$, we obtain the Kullblack-Leibler distance. The Hellinger distance corresponds to the case where $f(t) = (\sqrt{t} - 1)^2$. Total variation between absolutely continuous measures boils down to take $f(t) = |t-1|$.

Another class of distances between measures can be obtained via optimal transportation theory. For general results about this theory, we refer to the books [24, 25, 29, 28].

Definition 2. *Let* $(\mathfrak{E}, \mathbf{P})$ *and* $(\mathfrak{F}, \mathbf{Q})$ *two Polish spaces equipped with a probability measure and* c *a semi-continuous function from* $\mathfrak{E} \times \mathfrak{F}$ *to* $\mathbf{R}^+ \cup \{\infty\}$*. The optimal-transportation problem or Monge-Kantorovitch problem MKP*$(\mathbf{P}, \mathbf{Q})$ *is to find*

$$\min_{\gamma \in \Sigma(\mathbf{P}, \mathbf{Q})} \int_{\mathfrak{E} \times \mathfrak{F}} c(x, y) \; d\gamma(x, y)$$

where $\Sigma(\mathbf{P}, \mathbf{Q})$ *denoted the space of probability measures on* $\mathfrak{E} \times \mathfrak{F}$ *with first marginal* $\mathbf{P}$ *and second marginal* $\mathbf{Q}$*.*

Said otherwise in a more probabilistic way, it amounts to find the coupling between $\mathbf{P}$ and $\mathbf{Q}$ which minimizes the cost, i.e. to construct on the same probability space, two random variables X and Y of respective distribution $\mathbf{P}$ and $\mathbf{Q}$ which minimizes $\mathbf{E}\left[c(X, Y)\right]$ among all the possible constructions. The usual cost functions are of the type $c(x, y) = \mathrm{dist}(x, y)^p$ where dist is a distance and p a positive real number. For the Euclidean distance and $p = 2$, we can construct the so-called Wasserstein distance by considering

$$W(\mathbf{P}, \mathbf{Q}) = \sqrt{\min_{\gamma \in \Sigma(\mathbf{P}, \mathbf{Q})} \int_{\mathbf{R}^d \times \mathbf{R}^d} |\, x - y \,|^{\,2} \; \mathrm{d}\gamma(x, y)}.$$

All the distances viewed so far are not unrelated as many functional inequalities do exist between all of them. Just to mention two examples, the Pinsker inequality states that the total variation distance is controlled by the Kullblack-Leibler distance.

$$D_{|t-1|}(\mathbf{P}, \mathbf{Q}) \leq \sqrt{\frac{1}{2} D_{t \ln t}(\mathbf{P}, \mathbf{Q})}.$$

On the other hand, the so-called HWI identity (see [28]) relates the relative entropy (H), the Wasserstein distance (W) and the Fischer information (I) as follows.

Theorem 2. *Let* $\mathbf{P}$ *and* $\mathbf{Q}$ *two probability measures on* $\mathbf{R}^n$ *such that* $\mathbf{P} = \exp(-V) \; dx$ *with* $\nabla^2 V \geq K Id_n$*. Then,*

$$D_{t \ln t}(\mathbf{P}, \mathbf{Q}) \leq W(\mathbf{P}, \mathbf{Q}) \sqrt{D_{\nabla |\ln t|^2}(\mathbf{P}, \mathbf{Q})} - \frac{K}{2} W(\mathbf{P}, \mathbf{Q})^2.$$

These examples are here only to give a glimpse of the vast subject of the relationship between all these notions of distances. However, this is not the true subject of the present paper. The theorem which justifies the sequel is known as Kantorovitch-Rubinstein theorem (see [13, 14]) and says the following.

Theorem 3. *For* $\mathbf{P}$ *and* $\mathbf{Q}$ *two probability measures on a Polish space* $\mathfrak{E}$*, consider the Monge-Kantorovitch problem for a cost function* c *which is a distance on* $\mathfrak{E}$*. Then, we have the following representation*

$$\min_{\gamma \in \Sigma(\mathbf{P}, \mathbf{Q})} \int_{\mathfrak{E} \times \mathfrak{F}} c(x, \, y) \; d\gamma(x, y) = \sup_{F \in \mathrm{Lip}_c(1)} \left(\mathbf{E}_{\mathbf{P}}\left[F\right] - \mathbf{E}_{\mathbf{Q}}\left[F\right] \right),$$

where $F \in \mathrm{Lip}_c(1)$ *means that* F *is* c*-Lipschitz continuous:* $|F(x) - F(y)| \leq c(x, y)$ *for all* $x, \, y \in E$*. The resulting distance between* $\mathbf{P}$ *and* $\mathbf{Q}$*, will be called henceforth the Kantorovitch-Rubinstein distance as in* [28].

This formulation of a distance motivates alternative definitions by changing the set of test functions. For instance, for $\mathcal{F} = \{\mathbf{1}_{(-\infty;\, x]},\, x \in R\}$,

$$\sup_{F\in\mathcal{F}} |\mathbf{E_P}\left[F\right] - \mathbf{E_Q}\left[F\right]|$$

is the total-variation distance. It turns out that Stein's method is particularly well suited to estimate such kind of distances as we shall see now.

3. Stein's method

Historically, the Stein's method for Gaussian distribution dates back to the seminal paper of Stein [27]. It was soon extended to the Poisson distribution in the paper of Chen [7]. It is then impossible to track all the extensions of this approach, made mainly by A. Barbour and his collaborators, to several other distributions like compound Poisson [5], Poisson point processes[30], stationary measure of birth-death process, even Brownian motion [2]. For a whole account of all this period, one may refer to the books [3, 4] and references therein. The main breakthrough came with the paper of Nourdin and Peccati [21], in which it is shown that combining Malliavin calculus and Stein's approach, one can obtain a rather simple proof of the striking *fourth moment theorem*, established earlier in [22]. This was the starting point of a bunch of articles with with a wide area of applications: rate of convergence in the central limit theorem, Berry-Esseen theorem, iterated-logarithm theorem, limit theorems on manifolds, etc.

3.1. **Dirichlet-Malliavin structure.** The procedure of the Stein's method can be abstracted within the setting of Dirichlet structures (for details, we refer to [6, 16, 20]). The subsequent explanations are at a very formal level since the hard part for this machinery to work is to find the convenient functional spaces for each case of applications.

The first idea underlying the Stein's method is to characterize the target measure by an algebraic equation: Find a functional operator L on $\mathcal{F}$ such that $\mathbf{E_Q}\left[LF\right] = 0$ for any F in $\mathcal{F}$ if and only if $\mathbf{Q} = \mathbf{P}$. It turns out that this functional operator L can be viewed as the (infinitesimal) generator of a Markovian semi-group, which we denote by $P = (P_t,\, t \geq 0)$ whose stationary measure is $\mathbf{P}$: The image measure of $\mathbf{P}$ by P_t is still $\mathbf{P}$ for any $t \geq 0$. Under some technical hypothesis, there exists a strong ergodic Markov process $X = (X(t),\, t \geq 0)$ of invariant measure $\mathbf{P}$ and of generator L. It must be noted that the knowledge of one of L, P or X is equivalent to the knowledge of the other two. Formally speaking, for any $x \in E$,

$$P_t f(x) = e^{tL} f(x),\ Lf(x) = \left.\frac{dP_t f(x)}{dt}\right|_{t=0},\ P_t f(x) = \mathbf{E}\left[f(X(t)) \,|\, X(0) = x\right].$$

One can also associate to X, the so-called Dirichlet form defined formally by

$$\mathcal{E}(F, G) = \mathbf{E_P}\left[LF\, G\right],$$

for any F and G sufficiently regular. As before, if we are given such a bilinear form $\mathcal{E}$, one can retrieve L by the following relationship: For any F, LF is the unique element H such that for any G, $\mathcal{E}(F,\, G) = \mathbf{E_P}\left[HG\right]$. This means that whichever of L, X, P or $\mathcal{E}$ we are given, the others are uniquely determined (the reader is referred to the particularly illuminating Diagram 2, page 36 of [20]).

Within this framework, it is easy to see that the Stein-Dirichlet representation formula holds: For any bounded F,

$$\mathbf{E_Q}\left[F\right] - \mathbf{E_P}\left[F\right] = \mathbf{E_Q}\left[\int_0^\infty LP_t F\ \mathrm{d}t\right]. \tag{1}$$

This formula is also known as *the semi-group method* or *the smart-path formula* in the Stein's method literature. This means that we can write

$$\mathrm{dist}_{\mathcal{F}}(\mathbf{P},\mathbf{Q}) = \sup_{F\in\mathcal{F}} |\mathbf{E}_{\mathbf{P}}[F] - \mathbf{E}_{\mathbf{Q}}[F]| = \sup_{F\in\mathcal{F}} \left|\mathbf{E}_{\mathbf{Q}}\left[\int_0^\infty LP_tF\ \mathrm{d}t\right]\right|.$$

Instead of using coupling arguments to estimate this right-hand-side as usually done in the Stein's method, we use another functional operator which is the *gradient* in the sense of Malliavin. It is usually denoted by D and satisfies the identity $L = D^*D$ where D^* is the adjoint of D. This a square root of the symmetric operator L, but not all square-roots are interesting as we also need a nice commutation relationship between D and P. A few examples are the best way to illustrate what we mean.

3.2. **One dimensional examples.** If $\mathbf{P}$ denote the standard Gaussian measure on $\mathbf{R}$, then X is the Ornstein-Uhlenbeck process defined by

$$dX(t) = \sqrt{2}\ \mathrm{d}B(t) - X(t)\ \mathrm{d}t,\ X(0) = x,$$

where B is a standard one-dimensional Brownian motion. A straightforward application of the Itô formula gives the following expression of X:

$$X(t) = e^{-t}x + \sqrt{2}\int_0^t e^{-(t-s)}\ \mathrm{d}B(s).$$

It is then easy to see that $X(t) \sim \mathcal{N}(e^{-t}x,\ 1-e^{-2t})$, which, in turn, entails the Mehler representation formula:

$$P_tF(x) = \int_{\mathbf{R}} F(e^{-t}x + \sqrt{1-e^{-2t}}y)\ \mathrm{d}\mathbf{P}(y).$$

It follows by differentiation and integration by parts that for $F \in \mathcal{C}_b^2$,

$$LF(x) = xF'(x) - F''(x),\ \text{for all } x \in \mathbf{R}.$$

The *Malliavin* gradient is the usual derivative operator and standard computations show that

$$\int_{\mathbf{R}} DF(x)\,G(x)\ \mathrm{d}\mathbf{P}(x) = \int_{\mathbf{R}} F(x)(xG(x) - DG(x))\ \mathrm{d}\mathbf{P}(x),$$

hence that $D^*G(x) = xG(x) - DG(x)$ and $L = D^*D$. Moreover, we have $DP_tF(x) = e^{-t}P_tDF(x)$ which is the commutation relationship alluded above.

If $\mathbf{P}$ represents the Poisson measure on $\mathbf{N}$ of parameter λ, the process X can be viewed as the number of occupied servers in an M/M/∞ queue (see [11]), L is the corresponding generator:

$$LF(x) = \lambda(F(x+1) - F(x)) + x(F(x-1) - F(x)),\ \text{for all } x \in \mathbf{N},$$

with the convention that $0.F(-1) = 0$. The *gradient* is defined by

$$DF(x) = F(x+1) - F(x),$$

and we have $DP_tF = e^{-t}P_tDF$ (see [11, Theorem 11.16] or [12]). For the scalar product in $L^2(\mathbf{P})$, we have

$$\int_{\mathbf{N}} DF(x)\,G(x)\ \mathrm{d}\mathbf{P}(x) = \int_{\mathbf{N}} F(x)(\frac{x}{\lambda}G(x-1) - G(x))\ \mathrm{d}\mathbf{P}(x). \tag{2}$$

Hence,

$$D^*F(x) = \frac{x}{\lambda}G(x-1) - G(x) \text{ and } L = D^*D.$$

We now show how these constructions do articulate to give a new approach to the Stein's method.

It is well known that for Z_λ a Poisson random variable of parameter λ,

$$\hat{Z}_\lambda = \frac{Z_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{\lambda\to\infty} \mathcal{N}(0,\ 1) \text{ in distribution.}$$

We are going to use the Stein-Dirichlet-Malliavin method to evaluate the rate of convergence. We are in a situation where the target measure in defined $\mathbf{R}$ whereas the initial randomness comes from a probability measure on $\mathbf{N}$. The map T defined by

$$\begin{aligned} T\ :\ \mathfrak{E} = \mathbf{N} &\longrightarrow \mathfrak{F} = \mathbf{R} \\ n &\longmapsto \frac{n-\lambda}{\sqrt{\lambda}}, \end{aligned}$$

maps one space to the other and we are to evaluate the distance between $T^*\mathbf{Q}_\lambda$, the image measure of $\mathbf{Q}_\lambda$, the Poisson(λ) probability, by the map T and $\mathbf{P}$ the standard normal distribution on $\mathbf{R}$. This is a particular case of the general situation illustrated in Figure 2.

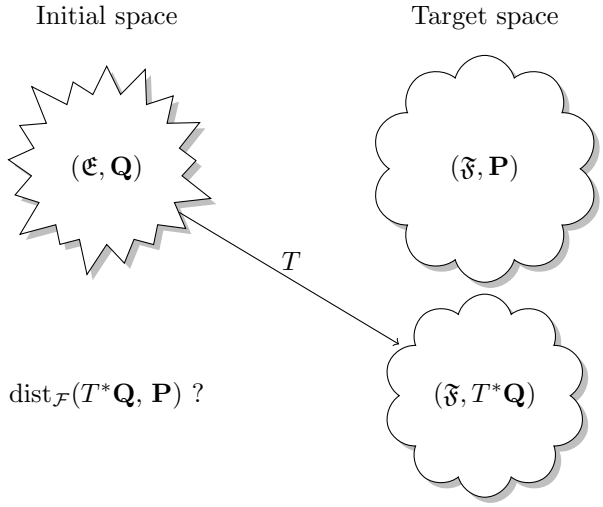

FIGURE 2. Comparison between a measure $\mathbf{P}$ and $T^*\mathbf{Q}$.

In view of (1), we have to estimate

$$\sup_{F\in\mathcal{F}} \int_0^\infty \int_{\mathbf{R}} x.(P_tF)'(x) - (P_tF)''(x)\ \mathrm{d}T^*\mathbf{Q}_\lambda(x)\ \mathrm{d}t,$$

where is the Ornstein-Uhlenbeck semi-group given by the Mehler formula above and $\mathcal{F}$ is a functional space to be conveniently chosen. According to the definition of T, the quantity to maximize is equal to

$$\mathbf{E}\left[\int_0^\infty \hat{Z}_\lambda.(P_tF)'(\hat{Z}_\lambda) - (P_tF)''(\hat{Z}_\lambda)\ \mathrm{d}t\right].$$

Applying (2) to $G = 1$ and $F \circ T$, we get

$$\sqrt{\lambda}\ \mathbf{E}\left[F(\hat{Z}_\lambda + \frac{1}{\sqrt{\lambda}}) - F(\hat{Z}_\lambda)\right] = \mathbf{E}\left[\hat{Z}_\lambda\, F(\hat{Z}_\lambda)\right].$$

Hence,

$$\mathbf{E}\left[\hat{Z}_\lambda.(P_tF)'(\hat{Z}_\lambda)\right] = \sqrt{\lambda}\mathbf{E}\left[(P_tF)'(\hat{Z}_\lambda + \frac{1}{\sqrt{\lambda}}) - (P_tF)'(\hat{Z}_\lambda)\right]. \tag{3}$$

For any $t > 0$, the regularizing properties of P_t entails that $P_t F$ is thrice differentiable. Hence,

(4)
$$(P_tF)'(\hat{Z}_\lambda+\frac{1}{\sqrt{\lambda}})-(P_tF)'(\hat{Z}_\lambda) = \frac{1}{\sqrt{\lambda}}(P_tF)''(\hat{Z}_\lambda)+\frac{1}{\lambda}\int_0^1 (1-r)(P_tF)^{(3)}(\hat{Z}_\lambda+\frac{r}{\sqrt{\lambda}})\ \mathrm{d}r.$$

And then, a miracle occurs: The term involving the second order derivative vanishes and we are lead to maximize

$$\frac{1}{\sqrt{\lambda}}\ \mathbf{E}\left[\int_0^\infty \int_0^1 (1-r)(P_tF)^{(3)}(\hat{Z}_\lambda + \frac{r}{\sqrt{\lambda}})\ \mathrm{d}r\ \mathrm{d}t\right] \tag{5}$$

for F over $\mathcal{F}$. There is now a delicate point. If F is in $\mathcal{C}_b^1$, we already mentioned that

$$(P_tF)'(x) = e^{-t}P_t(F')(x).$$

Furthermore, by integration by parts with respect to the Gaussian measure, it is easy to see that

$$(P_tF)^{(k)}(x) = \left(\frac{e^{-t}}{\sqrt{1-e^{-2t}}}\right)^k \int_{\mathbf{R}} F(e^{-t}x + \sqrt{1-e^{-2t}}y)\, y^k\ \mathrm{d}\mathbf{P}(y),$$

whenever F is bounded, for any $k \geq 1$. At first glance, it seems easy to bound (5) by using the previous formula for $k = 3$. Unfortunately, the term $\exp(-kt)(1-\exp(-2t))^{-k/2}$ is integrable over $[0,+\infty)$ only for $k = 1$. Hence, we must choose $\mathcal{F} = \{F \in \mathcal{C}_b^2, \|F\|_{\mathcal{C}_b^2} \leq 1\}$ and then we have

$$\begin{aligned}\left|(P_tF)^{(3)}(x)\right| &= \left|\frac{e^{-3t}}{\sqrt{1-e^{-2t}}}\int_{\mathbf{R}} F^{(2)}(e^{-t}x + \sqrt{1-e^{-2t}}y)\, y\ \mathrm{d}\mathbf{P}(y)\right| \\ &\leq \frac{e^{-3t}}{\sqrt{1-e^{-2t}}}\|F^{(2)}\|_\infty\ \int_{\mathbf{R}} |y|\ \mathrm{d}\mathbf{P}(y).\end{aligned}$$

Plugging this inequality into (5), we get

$$\begin{aligned}&\sup_{\|F\|_{\mathcal{C}_b^2}\leq 1}\left|\mathbf{E}\left[F(\hat{Z}_\lambda)\right] - \int F\ \mathrm{d}\mathbf{P}\right| \\ &\qquad \leq \frac{1}{\sqrt{\lambda}}\int_0^1 (1-r)\ \mathrm{d}r\ \int_0^\infty \frac{e^{-3t}}{\sqrt{1-e^{-2t}}}\ dt\ \int_{\mathbf{R}} |y|\ \mathrm{d}\mathbf{P}(y) = \frac{\sqrt{\pi}}{4\sqrt{2}}\ \frac{1}{\sqrt{\lambda}}.\end{aligned} \tag{6}$$

Hence we have established the rate of convergence for the Kantorovitch-Rubinstein distance associated to $\mathcal{F} = \{F \in \mathcal{C}_b^2, \|F\|_{\mathcal{C}_b^2} \leq 1\}$. In dimension 1, for Gaussian approximation, we could have used $LF(x) = xF(x) - F'(x)$ as a characterizing operator and thus used only 1-Lipschitz functions with a slightly different constant in front of the λ^{-1} factor, namely

$$\sup_{F\in\mathrm{Lip}(1)}\left|\mathbf{E}\left[F(\hat{Z}_\lambda)\right] - \int F\ \mathrm{d}\mathbf{P}\right| \leq \frac{1}{\sqrt{2\pi}}\ \frac{1}{\lambda}.$$

Note that this upper-bound is better than the bound obtained by the classical Stein's method where $(2\pi)^{-1/2}$ is replaced by 1. However, this line of thought is not applicable to higher dimensions.

More generally, the recipe of the Stein-Dirichlet-Malliavin method is the following.

- Characterize the target measure as the stationary distribution of an ergodic Markov process,
- Construct the two Dirichlet-Malliavin structure on both initial and target spaces,
- Perform an integration by parts on the initial space (see (3)),

- Replace the gradient on the initial space by a function of the gradient on the target space (this is done here by the Taylor formula (4)), at the price of additional terms to be controlled,
- Finish the computations in the target space using the commuting relationship : $DP_t = e^{-t}P_t D$.

3.3. **Higher dimensions.** This procedure can be generalized to any dimension provided that we have Dirichlet-Malliavin structures on both the initial and the target spaces. For the Gaussian measure in dimension d, the generator is given by

$$LF(x) = x.DF(x) - \Delta F(x), \text{ for all } x \in \mathbf{R}^d, \tag{7}$$

where D is the usual gradient in $\mathbf{R}^d$ and Δ is the Laplacian operator. The Mehler formula stays formally the same with an integral over $\mathbf{R}^d$ instead of $\mathbf{R}$ and X is the $\mathbf{R}^d$-valued process composed of d independent copies of the one dimensional Ornstein-Uhlenbeck process. The Malliavin gradient is still the usual gradient and the commutation relationship between D and P_t is easily seen to hold again. We can then retrieve the results of [23].

Real difficulties arise when we try to generalize this approach to infinite dimensional spaces like the Wiener space. It is tempting to define L formally as in (7), replacing the Laplacian by the trace of $D \circ D$. Unfortunately, for this trace term to exist, we need to restrict the space $\mathcal{F}$ of test functions and to choose conveniently the space $\mathfrak{F}$. There are actually two papers which address this problem. In both of them [8, 26], despite apparent dissimilarities, we end by considering $\mathfrak{F}$ a Hilbert space with a Gaussian measure.

Let us show how it works on an example. For N_λ a Poisson process on $\mathbf{R}^+$ of intensity λ, it is known that

$$\hat{N}_\lambda(t) = \frac{N_\lambda(t) - \lambda t}{\sqrt{\lambda}} \xrightarrow{\lambda\to\infty} B(t) \text{ in distribution,}$$

where B is a standard Brownian motion and the convergence is understood to hold in $\mathbb{D}$, the Skorohod space of rcll functions. To compare the two distributions implies to find a common Hilbert space which supports both the distribution of B and $\hat{N}_\lambda$. In principle, any Sobolev-like space should do. In [8], we chose the so-called Besov-Liouville space $I^{\beta,2}$ for $\beta < 1/2$ defined by

$$I^{\beta,2} = \{f, \exists \dot{f} \in L^2([0,1]) \text{ such that } f(x) = \frac{1}{\Gamma(\beta)} \int_0^x (x-y)^{\beta-1} \dot{f}(y)\, \mathrm{d}y\}.$$

It is a Hilbert space when equipped with the scalar-product $\langle f, g\rangle_{\beta,2} = \langle \dot{f}, \dot{g}\rangle_{L^2}$. The Wiener measure on this space, denoted by μ_β, is defined by

$$\mathbf{E}_{\mu_\beta}\left[\exp(i\langle \eta, \omega\rangle_{\beta,2})\right] = \exp(-\frac{1}{2}\langle V_\beta \eta, \eta\rangle_{\beta,2}).$$

where

$$I^\beta_{0^+} f(x) = \frac{1}{\Gamma(\beta)} \int_0^x (x-y)^{\beta-1} \dot{f}(y)\, \mathrm{d}y,\ I^\beta_{1^-} f(x) = \frac{1}{\Gamma(\beta)} \int_x^1 (y-x)^{\beta-1} \dot{f}(y)\, \mathrm{d}y$$
$$\text{and } V_\beta = I^\beta_{0^+} \circ I^{1-\beta}_{0^+} \circ I^{1-\beta}_{1^-} \circ I^{-\beta}_{0^+}.$$

The Ornstein-Uhlenbeck semi-group on $(I^{\beta,2}, \mu_\beta)$ is defined for any $F \in L^2(I^{\beta,2}, \mu_\beta)$ by

$$P^\beta_t F(u) := \int_{I^{\beta,2}} F(e^{-t}u + \sqrt{1-e^{-2t}}\, v)\, \mathrm{d}\mu_\beta(v).$$

The gradient is the Fréchet gradient on $I^{\beta,2}$ and all the other properties still holds formally as in finite dimension.

As initial space, we consider $\mathfrak{E} = \mathfrak{N}$, the space of locally finite configurations on $\mathbf{R}^+$ equipped with the vague topology. The measure $\mathbf{Q}_\lambda$ is such that the canonical process, denoted by N_λ, is a Poisson process of intensity λ, for details we refer to [8]. On the initial space, we actually only need to know the gradient and an integration by parts formula. Here, we take

$$D_x F(N_\lambda) = F(N_\lambda + \delta_x) - F(N_\lambda),$$

where $N_\lambda + \delta_x$ is the configuration N_λ with an additional atom at location x. The well-known Campbell-Mecke formula ([18, 19]) is equivalent to say that

$$\mathbf{E}_{\mathbf{Q}_\lambda}\left[F \int_0^1 G_\tau(\, \mathrm{d}N_\lambda(\tau) - \lambda \, \mathrm{d}\tau)\right] = \lambda \, \mathbf{E}_{\mathbf{Q}_\lambda}\left[\int_0^1 D_\tau F \; G_\tau \, \mathrm{d}\tau\right],$$

for G a deterministic process. The map T is defined by

$$\begin{aligned} T \; : \; \mathfrak{N} &\longrightarrow I^{\beta,2} \\ N &\longmapsto (t \mapsto \frac{N(t) - \lambda t}{\sqrt{\lambda}}). \end{aligned}$$

Proceeding exactly along the same lines as before, one can show that there exists $c_\beta > 0$ such that

$$\sup_{\|F\|_{\mathcal{C}_b^2(I^{\beta,2};\,\mathbf{R})} \leq 1} \left|\mathbf{E}_{\mathbf{Q}_\lambda}[F] - \mathbf{E}_{\mu_\beta}[F]\right| \leq \frac{c_\beta}{\sqrt{\lambda}}, \tag{8}$$

where $\mathcal{C}_b^2(I^{\beta,2}; \mathbf{R})$ is the set of twice Fréchet differentiable functionals on $I^{\beta,2}$, with bounded differentials. This is the generalization we could expect of (6).

Other examples of the application of this procedure, involving other functional spaces, can be found in the papers [8, 12]. A similar approach with Malliavin calculus replaced by a coupling argument appears in [10].

4. Edgeworth expansion

The Stein's method as developed here can be iterated to obtain Edgeworth expansions. We now want to precise the expansion obtained in (6). For, we go one step further in the Taylor formula (4):

$$\psi(\hat{Z}_\lambda + 1/\sqrt{\lambda}) - \psi(\hat{Z}_\lambda) = \frac{1}{\sqrt{\lambda}}\psi'(\hat{Z}_\lambda) + \frac{1}{2\lambda}\psi''(\hat{Z}_\lambda) + \frac{1}{6\lambda^{3/2}}\psi^{(3)}(\hat{Z} + \theta/\sqrt{\lambda}).$$

Hence,

$$\begin{aligned} &\mathbf{E}\left[\hat{Z}_\lambda D P_t F(\hat{Z}_\lambda) - D^{(2)} P_t F(\hat{Z}_\lambda)\right] \\ &\qquad = \frac{1}{2\sqrt{\lambda}}\mathbf{E}\left[D^{(3)} P_t F(\hat{Z}_\lambda)\right] + \frac{1}{6\lambda}\mathbf{E}\left[D^{(4)} P_t F(\hat{Z} + \theta/\sqrt{\lambda})\right]. \end{aligned} \tag{9}$$

If F is thrice differentiable with bounded derivatives then $P_t F$ is four times differentiable, hence the last term of (9) is bounded by $\lambda^{-1}\frac{e^{-4t}}{\sqrt{1-e^{-2t}}}\|F^{(3)}\|_\infty/6$. Moreover, applying (6) to $DP_t F$ shows that

$$\mathbf{E}\left[D^{(3)} P_t F(\hat{Z}_\lambda)\right] = \mathbf{E}_{\mathbf{P}}\left[D^{(3)} P_t F\right] + O(\lambda^{-1/2}).$$

Combining the last two results, we obtain that for F thrice differentiable

$$\mathbf{E}\left[F(\hat{Z}_\lambda)\right] - \mathbf{E}_{\mathbf{P}}[F] = \frac{1}{2\sqrt{\lambda}}\mathbf{E}_{\mathbf{P}}\left[\int_0^\infty D^{(2)} P_t F \, \mathrm{d}t\right] + O(\lambda^{-1}).$$

This line of thought can be pursued at any order provided that F is assumed to have sufficient regularity and we get an Edgeworth expansion up to any power of $\lambda^{-1/2}$. Using the properties of Hermite polynomials, this leads to the expansion:

$$\mathbf{E}\left[F(\hat{Z}_\lambda)\right] - \mathbf{E}_{\mathbf{P}}\left[F\right] = \frac{1}{6\sqrt{\lambda}}\mathbf{E}_{\mathbf{P}}\left[F\mathcal{H}_3\right] + O(\lambda^{-1}),$$

where $\mathcal{H}_n$ is the n-th Hermite polynomials. In [9], we generalized this approach to the Poisson process-Brownian motion convergence established in (8).

5. Conclusion

We showed how the Stein's method can be abstracted in the framework of Dirichlet forms and Malliavin calculus. This gives raise to a new method of proof which can be applied to infinite dimensional spaces and iterated to get Edgeworth expansions. One open question is to apply this approach to other limiting processes like stable or max-stable processes, Brownian bridges, etc.

References

[1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[2] A. D. Barbour. Stein's method for diffusion approximations. *Probab. Theory Related Fields*, 84(3):297–322, 1990.

[3] A. D. Barbour, T. C. Brown, and A. Xia. Point processes in time and Stein's method. *Stochastics Stochastics Rep.*, 65(1-2):127–151, 1998.

[4] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, 1992. Oxford Science Publications.

[5] A. D. Barbour and M. Maansson. Compound Poisson process approximation. *Ann. Probab.*, 30(3):1492–1537, 2002.

[6] N. Bouleau and F. Hirsch. *Dirichlet Forms and Analysis on Wiener Space*. De Gruyter Studies in Mathematics, 1992.

[7] L. H. Y. Chen. On the convergence of Poisson binomial to Poisson distributions. *Ann. Probability*, 2(1):178–180, 1974.

[8] L. Coutin and L. Decreusefond. Stein's method for Brownian approximations. *Communications on stochastic analysis*, 7(5):349–372, 2013.

[9] L. Coutin and L. Decreusefond. Higher order expansions via Stein's method. *Communications on Stochastic Analysis*, May 2014.

[10] Schumacher D. and Stucki K. Gibbs point process approximation: total variation bounds using Steins method. *Ann. Probab.*, 42(5):1911–1951, 2014.

[11] L. Decreusefond and P. Moyal. *Stochastic modeling and analysis of telecom networks*. ISTE Ltd and John Wiley & Sons Inc, 2012.

[12] L. Decreusefond, M. Schulte, and C. Thäle. Functional Poisson approximation in Rubinstein distance. Annals of Probability, to appear, 2015.

[13] R. M. Dudley. *Real analysis and probability*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[14] X. Fernique. Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, volume 850 of *Lecture Notes in Math.*, pages 6–10. Springer, Berlin, 1981.

[15] D. Feyel and A. S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.

[16] M. Fukushima. *Dirichlet forms and Markov processes*, volume 23 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1980.

[17] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

[18] O. Kallenberg. *Random Measures*. Academic Press, 3rd edition, 1983.

[19] A. F. Karr. *Point processes and their statistical inference*, volume 7 of *Probability: Pure and Applied*. Marcel Dekker Inc., New York, second edition, 1991.

[20] Z. M. Ma and M. Röckner. *Introduction to the theory of (nonsymmetric) Dirichlet forms*. Universitext. Springer-Verlag, Berlin, 1992.

[21] I. Nourdin and G; Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009.

[22] David Nualart and Giovanni Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005.
[23] Giovanni Peccati and Cengbo Zheng. Multi-dimensional Gaussian fluctuations on the Poisson space. *Electron. J. Probab.*, 15:no. 48, 1487–1527, 2010.
[24] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.
[25] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. II*. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.
[26] H.-H. Shih. On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces. *Journal of Functional Analysis*, 261(5):1236 – 1283, 2011.
[27] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, 1972.
[28] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[29] C. Villani. *Optimal transport, old and new*. Lectures Notes in Mathematics. Springer Verlag, New York, 2007.
[30] A. Xia. Poisson approximation, compensators and coupling. *Stochastic Anal. Appl.*, 18(1):159–177, 2000.

Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France