



HAL
open science

Malliavin and dirichlet structures for independent random variables

Laurent Decreusefond, H el ene Halconruy

► **To cite this version:**

Laurent Decreusefond, H el ene Halconruy. Malliavin and dirichlet structures for independent random variables. 2017. hal-01565240v1

HAL Id: hal-01565240

<https://hal.science/hal-01565240v1>

Preprint submitted on 19 Jul 2017 (v1), last revised 26 Jul 2018 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

MALLIAVIN AND DIRICHLET STRUCTURES FOR INDEPENDENT RANDOM VARIABLES

LAURENT DECREUSEFOND AND HÉLÈNE HALCONRUY

ABSTRACT. On any denumerable product of probability spaces, we construct a Malliavin gradient and then a divergence and a number operator. This yields a Dirichlet structure which can be shown to approach the usual structures for Poisson and Brownian processes. We obtain versions of almost all the classical functional inequalities in discrete settings which show that the Efron-Stein inequality can be interpreted as a Poincaré inequality or that Hoeffding decomposition of U -statistics can be interpreted as a chaos decomposition. We obtain a version of the Lyapounov central limit theorem for independent random variables without resorting to ad-hoc couplings, thus increasing the scope of the Stein method.

1. INTRODUCTION

There are two motivations to the present paper. After some years of development, the Malliavin calculus has reached a certain maturity. The most complete theories are for Gaussian processes (see for instance [22, 31]) and Poisson point processes (see for instance [1, 27]). When looking deeply at the main proofs, it becomes clear that the independence of increments plays a major role in the effectiveness of the concepts. At a very formal level, independence and stationarity of increments induce the martingale representation property which by induction entails the chaos decomposition, which is one way to develop Malliavin calculus for Poisson [23], Lévy processes [25] and Brownian motion. It thus motivates to investigate the simplest situation of all with independence: that of a family of independent, non necessarily identically distributed, random variables.

The second motivation comes from the Stein's method¹. The Stein method which was initially developed to quantify the rate of convergence in the Central Limit Theorem [29] and then for Poisson convergence [9], can be decomposed in three steps (see [13]). In the first step, we have to find a functional identity which characterizes the target distribution and solve

Key words and phrases. Dirichlet structure, Ewens distribution, log-Sobolev inequality, Lyapounov CLT, Malliavin calculus, Stein's method, Talagrand inequality.

¹Giving an exhaustive bibliography about Stein's method is somehow impossible (actually, MathSciNet refers more than 500 papers on this subject). The references given here are only entry points to the items alluded to.

implicitly or explicitly (as in the semi-group method) the so-called Stein's equation. It reduces the computation of the distance to the calculation of

$$\sup_{F \in \mathcal{H}} \left(\mathbf{E} [L_1 F(X)] + \mathbf{E} [L_2 F(X)] \right),$$

where \mathcal{H} is the set of test functions which depends on the distance we are considering, L_1 and L_2 are two functional operators and X is a random variable whose distribution we want to compare to the target distribution. For instance, if the target distribution is the Gaussian law on \mathbf{R} ,

$$L_1 F(x) = x f'(x) \text{ and } L_2 F(x) = -f''(x).$$

If the target distribution is the Poisson law of parameter λ ,

$$L_1 F(n) = n(F(n) - F(n-1)) \text{ and } L_2 F(n) = \lambda(F(n+1) - F(n)).$$

In the next step, we have to take into account how is defined X and transform $L_1 F$ such that it can be written as $-L_2 F + \text{remainder}$. This remainder is what gives the rate of convergence. To make the transformation of $L_1 F$, several approaches appeared along the years. One of the most popular approach (see for instance [4]) is to use exchangeable pairs: Construct a copy X' of X with good properties which gives another expression of $L_1 F$, suitable to a comparison with $L_2 F$. To be more specific, for the proof of the CLT, it is necessary to create an exchangeable pair (S, S') with $S = \sum_{i=1}^n X_i$. This is usually done by first, choosing uniformly an index $I \in \{1, \dots, n\}$ and then, replacing X_I with X' an independent copy of X_I , so that the couple $(S, S' = S - X_I + X')$ is an exchangeable pair. This means that

$$(1) \quad \mathbf{E} [F(S') | I = a; X_b, b \neq a] = \mathbf{E} [F(S) | X_b, b \neq a].$$

Actually, it is the right-hand-side of (1) which gave us some clue on how to proceed when dealing with functionals more general than the sum of random variables. An alternative to exchangeable pairs, is the size-biased [10] or zero biased [16] couplings, which again conveniently transform $L_1 F$. For Gaussian approximation, it amounts to find a distribution X^* such that

$$\mathbf{E} [L_1 F(X)] = \mathbf{E} [F''(X^*)].$$

Note that for S as above, one can choose $S^* = S'$. If the distribution of X^* is absolutely continuous with respect to that of X , with Radon derivative Λ , we obtain

$$\mathbf{E} [L_1 F(X)] = \mathbf{E} [F''(X) \Lambda(X)],$$

which means that we are reduced to estimate how far Λ is from the constant random variable equal to 1. This kind of identity, where the second order derivative is multiplied by a weight factor, is very similar to what can be obtained via integration by parts. Actually, Nourdin and Peccati (see [20]) showed that the transformation step can be advantageously made simple using integration by parts in the sense of Malliavin calculus. This works well only if there exists a Malliavin gradient on the space on which X is defined (see for instance [15]). That is to say, that up to now, this approach is restricted to functionals of Rademacher [21], Poisson [15, 26] or Gaussian random variables [24] or processes [11, 12]. Then, strangely enough, the first example of applications of the Stein's method which was the CLT, cannot be handled through this approach. On the one hand, exchangeable pairs

or size-biased coupling have the main drawback to have to be adapted to each particular version of X . On the other hand, Malliavin integration by parts are in some sense more automatic but we need to be provided with a Malliavin structure. So, in order to prove the CLT without exchangeable pairs or size-biased coupling, we need to construct a sort of gradient on a space of independent random variables.

The closest situation to our investigations is that of the Rademacher space, namely $\{-1, 1\}^{\mathbf{N}}$, equipped with the product probability $\otimes_{n \in \mathbf{N}} \mu_k$ where μ_k is a Bernoulli probability on $\{-1, 1\}$. The gradient on the Rademacher space (see [21, 27]) does exist but it requires, for its very definition to be meaningful,

$$D_k F(X_1, \dots, X_n) = \mathbf{E}[X_k F(X_1, \dots, X_n) | X_l, l \neq k],$$

that the random variables are real valued. In what follows, it must be made clear that all the random variables may leave on different spaces, which are only supposed to be Polish spaces. That means that in the definition of the gradient, we cannot use any algebraic property of the underlying spaces.

Since Malliavin calculus is agnostic to any time reference, we do not even assume that we have an order on the product space. It is not a major feature since a denumerable A is by definition in bijection with the set of natural integers and thus inherits of at least one order structure. However, this added degree of freedom appears to be useful (see the Clark decomposition of the number of fixed points of a random permutations in Section 4) and bears strong resemblance with the different filtrations which can be put on an abstract Wiener space, via the notion of resolution of the identity [30]. During the preparation of this work, we found strong reminiscences of our gradient with the map Δ , introduced in [6, 28] for the proof of the Efron-Stein inequality, defined by

$$\Delta_k F(X_1, \dots, X_n) = \mathbf{E}[F | X_1, \dots, X_k] - \mathbf{E}[F | X_1, \dots, X_{k-1}].$$

Actually, our point of view diverges from that of these works as we do not focus on a particular inequality but rather on the intrinsic properties of our newly defined gradient.

We would like to stress the fact that our Malliavin-Dirichlet structure gives a unified framework for many results scattered in the literature. We hope to give new insights on why these apparently disjointed results (Efron-Stein, exchangeable pairs, etc.) are in fact multiple sides of the same coin.

We proceed as follows. In Section 2, we define the gradient D and its adjoint δ , which we call divergence as it appears as the sum of *the partial derivatives*, as in \mathbf{R}^n . We establish a Clark representation formula of square integrable random variables and an Helmholtz decomposition of vector fields. Clark formula appears to reduce to the Hoeffding decomposition of U -statistics when applied to such functionals. We establish a log-Sobolev inequality, strongly reminding that obtained for Poisson process [32], together with a concentration inequality. Then, we define the number operator $L = \delta D$. It is the generator of a Markov process whose stationary

distribution is the tensor probability we started with. We show in Section 3 that we can retrieve the classical Dirichlet-Malliavin structures for Poisson processes and Brownian motion as limits of our structures. We borrow for that the idea of convergence of Dirichlet structures to [8]. The construction of random permutations in [18], which is similar in spirit to the so-called Feller coupling (see [2]), is an interesting situation to apply our results since this construction involves a cartesian product of distinct finite spaces. In Section 4, we derive the chaos decomposition of the number of fixed points of a random permutations under the Ewens distribution. To the price of an additional complexity, it is certainly possible to find such a decomposition for the number of k -cycles in a random permutation. We finish in Section 5 by a quantitative version of the Lyapounov central limit theorem for independent but not necessarily identically distributed random variables.

2. MALLIAVIN CALCULUS FOR INDEPENDENT RANDOM VARIABLES

For A a countable set, let $(E_a, a \in A)$ be a family of Polish spaces. For any $a \in A$, let \mathcal{E}_a and \mathbf{P}_a be respectively a σ -field and a probability measure defined on E_a . We consider the probability space $E_A = \prod_{a \in A} E_a$ equipped with the product σ -field $\mathcal{E}_A = \bigvee_{a \in A} \mathcal{E}_a$ and the tensor product measure $\mathbf{P} = \bigotimes_{a \in A} \mathbf{P}_a$. The coordinate random variables are denoted by $(X_a, a \in A)$. For any $B \subset A$, X_B denotes the random vector $(X_a, a \in B)$, defined on $E_B = \prod_{a \in B} E_a$ equipped with the probability $\mathbf{P}_B = \bigotimes_{a \in B} \mathbf{P}_a$. A process U is a measurable random variable defined on $(A \times E_A, \mathcal{P}(A) \otimes \mathcal{E}_A)$. We denote by $L^2(A \times E_A)$ the Hilbert space of processes which are square integrable with respect to the measure $\sum_{a \in A} \varepsilon_a \otimes \mathbf{P}_A$ (where ε_a is the Dirac measure at point a):

$$\|U\|_{L^2(A \times E_A)}^2 = \sum_{a \in A} \mathbf{E}[U_a^2] \quad \text{and} \quad \langle U, V \rangle_{L^2(A \times E_A)} = \sum_{a \in A} \mathbf{E}[U_a V_a].$$

Following the vocabulary of point processes theory [17], we have the following definition.

Definition 2.1. For $a \in A$, the *exvisible* σ -field \mathcal{G}_a is defined as $\mathcal{G}_a = \sigma(X_b, b \in A \setminus \{a\})$. A process $(U_a, a \in A)$ is said to be *exvisible* if for any $a \in A$, the random variable U_a is *exvisible*, i.e. $U_a \in \mathcal{G}_a$.

We now introduce the set of cylindrical functionals, denoted by \mathcal{S} , which is as usual, of key importance.

Definition 2.2. A random variable F is said to be *cylindrical* if there exist a finite subset $B \subset A$ and a function $F_B : E_B \rightarrow L^2(E_A)$ such that $F = F_B \circ r_B$, where r_B is the restriction operator:

$$\begin{aligned} r_B : E_A &\longrightarrow E_B \\ (x_a, a \in A) &\longmapsto (x_a, a \in B). \end{aligned}$$

This means that F only depends on the finite set of random variables $(X_a, a \in B)$.

It is clear that \mathcal{S} is dense in $L^2(E_A)$.

2.1. Gradient.

Definition 2.3 (Discrete gradient). *For $F \in \mathcal{S}$, DF is the process of $L^2(A \times E_A)$ defined by one of the following equivalent formulations: For all $a \in A$,*

$$\begin{aligned} D_a F(X_A) &= F(X_A) - \mathbf{E}[F(X_A) | \mathcal{G}_a] \\ &= F(X_A) - \int_{E_a} F(X_{A \setminus a}, x_a) d\mathbf{P}_a(x_a) \\ &= F(X_A) - \mathbf{E}'[F(X_{A \setminus a}, X'_a)] \end{aligned}$$

where X'_a is an independent copy of X_a .

For $F \in \mathcal{S}$, $(D_a F, a \in A)$ defines an element of $L^2(A \times E_A)$ since there exists a finite subset $B \subset A$ such that $F = F_B \circ r_B$. Thus, for every $a \notin B$, F is \mathcal{G}_a -measurable and then $D_a F = 0$. This implies that

$$\|DF\|_{L^2(A \times E_A)}^2 = \mathbf{E} \left[\sum_{a \in A} |D_a F|^2 \right] = \mathbf{E} \left[\sum_{a \in B} |D_a F|^2 \right] < \infty.$$

Definition 2.4. *Let $(g_a, a \in A)$ be an orthonormal basis of $l^2(A)$. The set of simple processes, denoted by $\mathcal{S}_0(l^2(A))$ is the set of random variables defined on $A \times E_A$ of the form*

$$U = \sum_{a \in B} U_a g_a,$$

for B a finite subset of A and such that U_a belongs to \mathcal{S} for any $a \in B$.

Theorem 2.5 (Integration by parts). *Let $F \in \mathcal{S}$. For every simple process U ,*

$$(2) \quad \langle DF, U \rangle_{L^2(A \times E_A)} = \mathbf{E} \left[F \sum_{a \in A} D_a U_a \right].$$

Proof. The process $\text{trace}(DU) = (D_a U_a, a \in B)$ belongs to $L^2(A \times E_A)$: Using the Jensen inequality, we have

$$(3) \quad \|\text{trace}(DU)\|_{L^2(A \times E_A)}^2 = \mathbf{E} \left[\sum_{a \in B} |D_a U_a|^2 \right] \leq 2 \sum_{a \in B} \mathbf{E} [U_a^2] < \infty.$$

Moreover,

$$\begin{aligned} \langle DF, U \rangle_{L^2(A \times E_A)} &= \mathbf{E} \left[\sum_{a \in A} (F - \mathbf{E}[F | \mathcal{G}_a]) U_a \right] \\ &= \mathbf{E} \left[\sum_{a \in B} (F - \mathbf{E}[F | \mathcal{G}_a]) U_a \right] = \mathbf{E} \left[F \sum_{a \in B} (U_a - \mathbf{E}[U_a | \mathcal{G}_a]) \right], \end{aligned}$$

since the conditional expectation is a projection in $L^2(E_A)$. \square

Corollary 2.6. *The operator D is closable from $L^2(E_A)$ into $L^2(A \times E_A)$.*

Proof. Let $(F_n, n \geq 1)$ be a sequence of random variables defined on \mathcal{S} such that F_n converges to 0 in $L^2(E_A)$ and the sequence DF_n converges to η in $L^2(A \times E_A)$. Let U be a simple process. From the integration by parts formula (2)

$$\mathbf{E} \left[\sum_{a \in A} D_a F_n U_a \right] = \mathbf{E} \left[F_n \sum_{a \in A} D_a U_a \right]$$

where $\sum_{a \in A} D_a U_a \in L^2(E_A)$ in view of (3). Then,

$$\langle \eta, U \rangle_{L^2(A \times E_A)} = \lim_{n \rightarrow \infty} \mathbf{E} \left[F_n \sum_{a \in A} D_a U_a \right] = 0,$$

for any simple process U . It follows that $\eta = 0$ and then the operator D is closable from $L^2(E_A)$ to $L^2(A \times E_A)$. \square

We denote the domain of D in $L^2(E_A)$ by \mathbf{D} , the closure of the class of cylindrical functions with respect to the norm

$$\|F\|_{1,2} = \left(\|F\|_{L^2(E_A)}^2 + \|DF\|_{L^2(A \times E_A)}^2 \right)^{\frac{1}{2}}.$$

We could as well define p -norms corresponding to L^p integrability. However, for the current applications, the case $p = 2$ is sufficient and the apparent lack of hypercontractivity of the Ornstein-Uhlenbeck semi-group (see below Section 2.3) lessens the probable usage of other integrability order.

As A is countable, we can choose with no loss of generality an order on A , or equivalently say that $A = \mathbf{N}$ and define $\mathcal{F}_n = \sigma(X_k, k \leq n)$, for any $n \in A$. Define $\mathcal{F}_0 = \{\emptyset, E_A\}$.

Lemma 2.7. *If there exists a sequence $(F_n, n \geq 1)$ of elements of \mathbf{D} such that 1) F_n converges to F in $L^2(E_A)$ and 2) $\sup_n \|DF_n\|_{\mathbf{D}}$ is finite, then F belongs to \mathbf{D} and $DF = \lim_{n \rightarrow \infty} DF_n$ in \mathbf{D} .*

Proof. Since $\sup_n \|DF_n\|_{\mathbf{D}}$ is finite, there exists a subsequence which we still denote by $(DF_n, n \geq 1)$ weakly convergent in $L^2(A \times E_A)$ to some limit denoted by η . For $k > 0$, let n_k be such that $\|F_m - F\|_{L^2} < 1/k$ for $m \geq n_k$. The Mazur's Theorem implies that there exists a convex combination of elements of $(DF_m, m \geq n_k)$ such that

$$\left\| \sum_{i=1}^{M_k} \alpha_i^k DF_{m_i} - \eta \right\|_{L^2(A \times E_A)} < 1/k.$$

Moreover, since the α_i^k are positive and sums to 1,

$$\left\| \sum_{i=1}^{M_k} \alpha_i^k F_{m_i} - F \right\|_{L^2} \leq 1/k.$$

We have thus constructed a sequence

$$F^k = \sum_{i=1}^{M_k} \alpha_i^k F_{m_i}$$

such that F^k tends to F in L^2 and DF^k converges in $L^2(A \times E_A)$ to a limit. By the construction of \mathbf{D} , this means that F belongs to \mathbf{D} and that $DF = \eta$. \square

Lemma 2.8. *For any $F \in \mathbf{D}$, F is \mathcal{F}_k -measurable if and only if $D_n F = 0$ for any $n > k$. As a consequence, $DF = 0$ if and only if $F = \mathbf{E}[F]$.*

Proof. Let $k \in A$. Assume that $F \in \mathcal{F}_k$. Then, for every $n > k$, F is \mathcal{G}_n -measurable and $D_n F = 0$.

Let $F \in \mathbf{D}$ such that $D_n F = 0$ for every $n > k$. Then F is \mathcal{G}_n -measurable for any $n > k$. From the equality $\mathcal{F}_k = \bigcap_{n>k} \mathcal{G}_n$, it follows that F is \mathcal{F}_k -measurable. \square

Lemma 2.9. *For any $F \in \mathbf{D}$, for any $k \geq 1$, we have*

$$(4) \quad D_k \mathbf{E}[F | \mathcal{F}_k] = \mathbf{E}[D_k F | \mathcal{F}_k].$$

Proof. For any $k \geq 1$, $\mathcal{F}_k \cap \mathcal{G}_k = \mathcal{F}_{k-1}$, hence

$$D_k \mathbf{E}[F | \mathcal{F}_k] = \mathbf{E}[F | \mathcal{F}_k] - \mathbf{E}[F | \mathcal{F}_{k-1}] = \mathbf{E}[D_k F | \mathcal{F}_k].$$

The proof is thus complete. \square

Theorem 2.10 (Clark formula). *For $A = \mathbf{N}$ and $F \in \mathbf{D}$,*

$$F = \mathbf{E}[F] + \sum_{k=1}^{\infty} D_k \mathbf{E}[F | \mathcal{F}_k].$$

If A is finite and if there is no privileged order on A , we can write

$$F = \mathbf{E}[F] + \sum_{B \subset A} \binom{|A|}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} D_b \mathbf{E}[F | X_B].$$

Proof. Let F an \mathcal{F}_n -measurable random variable. It is clear that

$$F - \mathbf{E}[F] = \sum_{k=1}^n (\mathbf{E}[F | \mathcal{F}_k] - \mathbf{E}[F | \mathcal{F}_{k-1}]) = \sum_{k=1}^n D_k \mathbf{E}[F | \mathcal{F}_k].$$

For $F \in \mathbf{D}$, apply this identity to $F_n = \mathbf{E}[F | \mathcal{F}_n]$ to obtain

$$F_n - \mathbf{E}[F] = \sum_{k=1}^n D_k \mathbf{E}[F | \mathcal{F}_k].$$

Remark that for $l > k$, in view of Lemma 2.8,

$$(5) \quad \mathbf{E}[D_k \mathbf{E}[F | \mathcal{F}_k] D_l \mathbf{E}[F | \mathcal{F}_l]] = \mathbf{E}[D_l D_k \mathbf{E}[F | \mathcal{F}_k] \mathbf{E}[F | \mathcal{F}_l]] = 0,$$

since $D_k \mathbf{E}[F | \mathcal{F}_k]$ is \mathcal{F}_k -measurable. Hence, we get

$$\mathbf{E}[|F - \mathbf{E}[F]|^2] \geq \mathbf{E}[|F_n - \mathbf{E}[F]|^2] = \sum_{k=1}^n \mathbf{E}[D_k \mathbf{E}[F | \mathcal{F}_k]^2].$$

Thus, the sequence $(D_k \mathbf{E}[F | \mathcal{F}_k], k \geq 1)$ belongs to $l^2(\mathbf{N})$ and the result follows by a limiting procedure.

If A is finite, each bijection between A and $\{1, \dots, n\}$ defines an order on A . Hence, there are $|A|!$ possible filtrations. Each term of the form

$$D_{i_k} \mathbf{E}[F | X_{i_1}, \dots, X_{i_k}]$$

appears $(k-1)! (|A|-k)!$ times since the order of $X_{i_1}, \dots, X_{i_{k-1}}$ is irrelevant to the conditioning. The result follows by summation then renormalization of the identities obtained for each filtration. \square

The chaos decomposition is usually deduced from the Clark formula by iteration. If we apply Clark formula to $\mathbf{E}[F | \mathcal{F}_k]$, we get

$$D_k \mathbf{E}[F | \mathcal{F}_k] = \sum_{j=1}^{\infty} D_k D_j \mathbf{E}[F | \mathcal{F}_{j \wedge k}] = D_k \mathbf{E}[F | \mathcal{F}_k],$$

since $j > k$ implies $D_j \mathbf{E}[F | \mathcal{F}_k] = 0$ in view of Lemma 2.8 and the same holds when $k > j$ since it is easily seen that $D_j D_k = D_k D_j$. For $j = k$, simply remark that $D_k D_k = D_k$. Hence, it seems that we cannot go further this way to find a potential chaos decomposition.

For U -statistics, we can however simplify the Clark formula and retrieve the so-called Hoeffding decomposition.

Definition 2.11. For an integer m , let $h : \mathbb{R}^m \rightarrow \mathbb{R}$ be a symmetric function, and X_1, \dots, X_n , n random variables supposed to be independent and identically distributed. The U -statistics of degree m and kernel h is defined, for any $n \geq m$ by

$$U_n = U(X_1, \dots, X_n) = \binom{n}{m}^{-1} \sum_{A \in \binom{[n]}{m}} h(X_A)$$

where $\binom{[n]}{m}$ denotes the set of ordered subsets $A \subset [n] = \{1, \dots, n\}$, of cardinal m .

More generally, for a set B , $\binom{B}{m}$ denotes the set of subsets of B with m elements.

If $\mathbf{E}[|h(X_1, \dots, X_m)|]$ is finite, we define $h_m = h$ and for $1 \leq k \leq m-1$,

$$h_k(X_1, \dots, X_k) = \mathbf{E}[h(X_1, \dots, X_m) | X_1, \dots, X_k].$$

Let $\theta = \mathbf{E}[h(X_1, \dots, X_m)]$, consider $g_1(X_1) = h_1(X_1) - \theta$, and

$$g_k(X_1, \dots, X_k) = h_k(X_1, \dots, X_k) - \theta - \sum_{j=1}^{k-1} \sum_{B \in \binom{[k]}{j}} g_j(X_B),$$

for any $1 \leq k \leq m$. Since the variables X_1, \dots, X_n are independent and identically distributed, and the function h is symmetric, the equality

$$\mathbf{E}[h(X_{A \cup B}) | X_B] = \mathbf{E}[h(X_{C \cup B}) | X_B],$$

holds for any subsets A and C of $[n] \setminus B$, of cardinal $n - k$.

Theorem 2.12 (Hoeffding decomposition of U-statistics, [19]). *For any integer n , we have*

$$(6) \quad U_n = \theta + \sum_{k=1}^m H_n^{(k)}$$

where $H_n^{(k)}$ is the U-statistics based on kernel g_k , i.e. defined by

$$H_n^{(k)} = \binom{n}{k}^{-1} \sum_{B \subset [n], |B|=k} g_k(X_B).$$

Proof. Take care that in the argument of h , all the sets are considered as ordered: When we write $B \cup C$, we implicitly reorder its elements, for instance

$$h(X_{\{1,3\} \cup \{2\}}) = h(X_1, X_2, X_3).$$

Apply the Clark formula,

$$\begin{aligned} U_n - \theta &= \binom{n}{m}^{-1} \sum_{A \in [n], |A|=m} \sum_{B \subset A} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} D_b \mathbf{E}[h(X_A) | X_B] \\ &= \binom{n}{m}^{-1} \sum_{B \subset [n], |B|=m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{\substack{A \supset B \\ A \in [n], |A|=m}} D_b \mathbf{E}[h(X_A) | X_B] \\ &= \binom{n}{m}^{-1} \sum_{B \subset [n], |B|=m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{C \in ([n] \setminus B, m-|B|)} D_b \mathbf{E}[h(X_{B \cup C}) | X_B]. \end{aligned}$$

It remains to prove that

$$(7) \quad \sum_{k=1}^m \binom{m}{k} H_n^{(k)} = \binom{n}{m}^{-1} \sum_{B \subset [n], |B| \leq m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{C \in ([n] \setminus B, m-|B|)} D_b \mathbf{E}[h(X_{B \cup C}) | X_B].$$

for any integer n . For $n = 1$, it is straightforward that

$$g_1(X_1) = h(X_1) - \theta = D_1 \mathbf{E}[h(X_1) | X_1].$$

Assume the existence of an integer n such that (7) holds for any set of cardinal n . In particular, for any $l \in [n + 1]$

$$\begin{aligned} &\sum_{k=1}^m \binom{m}{k} H_{A_l}^{(k)} \\ &= \binom{n}{m}^{-1} \sum_{B \subset [A_l], |B| \leq m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{C \in ([A_l] \setminus B, m-|B|)} D_b \mathbf{E}[h(X_{B \cup C}) | X_B], \end{aligned}$$

where $A_l = [n+1] \setminus \{l\}$. Let m such that $m \leq n$. Then,

$$\begin{aligned}
& \sum_{k=1}^m \binom{m}{k} H_{n+1}^{(k)} \\
&= \sum_{k=1}^m \binom{m}{k} \binom{n+1}{k}^{-1} \frac{1}{n+1-k} \sum_{l=1}^{n+1} \sum_{B \in ([A_l], k)} g_k(X_B) \\
&= \frac{1}{n+1} \sum_{l=1}^{n+1} \sum_{k=1}^m \binom{m}{k} \binom{n}{k}^{-1} \sum_{B \in ([A_l], k)} g_k(X_B) \\
&= \frac{1}{n+1} \sum_{l=1}^{n+1} \binom{n}{m}^{-1} \\
&\quad \times \sum_{B \subset [A_l], |A_l| \leq m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{C \in ([A_l] \setminus B, m-|B|)} D_b \mathbf{E} [h(X_{B \cup C}) | X_B] \\
&= \frac{n+1-m}{n+1} \binom{n}{m}^{-1} \\
&\quad \times \sum_{B \subset [n+1], |B| \leq m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{C \in ([n+1] \setminus B, m-|B|)} D_b \mathbf{E} [h(X_{B \cup C}) | X_B] \\
&= \binom{n+1}{m}^{-1} \\
&\quad \times \sum_{B \subset [n+1], |B| \leq m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} \sum_{C \in ([n+1] \setminus B, m-|B|)} D_b \mathbf{E} [h(X_{B \cup C}) | X_B],
\end{aligned}$$

where we have used in the first line that each subset B of $[n+1]$ of cardinal k appears in $n+1-k$ different subsets A_l (for $l \in [n+1] \setminus B$), and in the same way, in the penultimate line, that each subset $B \cup C$ of $[n+1]$ of cardinal m appears in $n+1-m$ different subsets A_l (for $l \in [n+1] \setminus B \cup C$). Eventually, the case $m = n+1$ follows from

$$\begin{aligned}
\sum_{k=1}^{n+1} \sum_{B \in ([n+1], k)} g_k(X_B) &= h(X_{[n+1]}) - \theta \\
&= \sum_{B \subset [n+1]} \binom{n+1}{|B|}^{-1} \frac{1}{|B|} \sum_{b \in B} D_b \mathbf{E} [h(X_{[n+1]}) | X_B],
\end{aligned}$$

by applying the Clark formula to h . \square

As mentioned in the Introduction, it may be useful to reverse the time arrow. Choose an order on A so that A can be seen as \mathbf{N} . Then, let

$$\mathcal{H}_n = \sigma\{X_k, k > n\}.$$

and for any $n \in \{0, \dots, N-1\}$,

$$\mathcal{H}_n^N = \mathcal{H}_n \cap \mathcal{F}_N \quad \text{and} \quad \mathcal{H}_k^N = \mathcal{F}_0 = \{\emptyset, E_A\} \text{ for } k \geq N.$$

Note that $\mathcal{H}_0^N = \mathcal{F}_N$ and as in Lemma 2.8, F is \mathcal{H}_k -measurable if and only if $D_n F = 0$ for any $n \leq k$.

Theorem 2.13. *For every F in \mathbf{D} ,*

$$(8) \quad F = \mathbf{E}[F] + \sum_{k \geq 1} D_k \mathbf{E}[F | \mathcal{H}_{k-1}].$$

Proof. Remark that

$$\begin{aligned} D_k \mathbf{E}[F | \mathcal{H}_{k-1}^N] &= \mathbf{E}[F | \mathcal{H}_{k-1}^N] - \mathbf{E}[F | \mathcal{H}_{k-1}^N \cap \mathcal{G}_k] \\ &= \mathbf{E}[F | \mathcal{H}_{k-1}^N] - \mathbf{E}[F | \mathcal{H}_k^N]. \end{aligned}$$

For $F \in \mathcal{F}_N$, since the successive terms collapse, we get

$$\begin{aligned} F - \mathbf{E}[F] &= \mathbf{E}[F | \mathcal{H}_0^N] - \mathbf{E}[F | \mathcal{H}_N^N] \\ &= \sum_{k=1}^N D_k \mathbf{E}[F | \mathcal{H}_{k-1}^N] = \sum_{k=1}^{\infty} D_k \mathbf{E}[F | \mathcal{H}_{k-1}^N], \end{aligned}$$

by the very definition of the gradient map. As in (5), we can show that for any N ,

$$\mathbf{E}[D_k \mathbf{E}[F | \mathcal{H}_{k-1}^N] D_l \mathbf{E}[F | \mathcal{H}_{l-1}^N]] = 0, \text{ for } k \neq l.$$

Consider $F_N = \mathbf{E}[F | \mathcal{F}_N]$ and proceed as in the proof of Lemma 2.10 to conclude. \square

In the present context, the next result is a Poincaré type inequality as it gives a bound for the variance of F in terms of the *oscillations* of F . In other context, it turns to be called the Efron-Stein inequality [6].

Corollary 2.14 (Poincaré or Efron-Stein inequality). *For any $F \in \mathbf{D}$,*

$$\text{var}(F) \leq \|DF\|_{L^2(A \times E_A)}^2.$$

Proof. According to (5) and (4), we have

$$\begin{aligned} \text{var}(F) &= \mathbf{E} \left[\left| \sum_{k \in A} D_k \mathbf{E}[F | \mathcal{F}_k] \right|^2 \right] \\ &= \mathbf{E} \left[\sum_{k \in A} |D_k \mathbf{E}[F | \mathcal{F}_k]|^2 \right] \\ &= \mathbf{E} \left[\sum_{k \in A} |\mathbf{E}[D_k F | \mathcal{F}_k]|^2 \right] \\ &\leq \mathbf{E} \left[\sum_{k \in A} \mathbf{E}[|D_k F|^2 | \mathcal{F}_k] \right] = \mathbf{E} \left[\sum_{k \in A} |D_k F|^2 \right], \end{aligned}$$

where the inequality follows from then Jensen inequality. \square

Another corollary of the Clark formula is the following covariance identity.

Theorem 2.15 (Covariance identity). *For any $F, G \in \mathbf{D}$,*

$$(9) \quad \text{cov}(F, G) = \mathbf{E} \left[\sum_{k \in A} D_k \mathbf{E}[F | \mathcal{F}_k] D_k G \right].$$

Proof. Let $F, G \in \mathbf{D}$, the Clark formula entails

$$\begin{aligned} \text{cov}(F, G) &= \mathbf{E} [(F - \mathbf{E}[F])(G - \mathbf{E}[G])] \\ &= \mathbf{E} \left[\sum_{k, l \in A} D_k \mathbf{E}[F | \mathcal{F}_k] D_l \mathbf{E}[G | \mathcal{F}_l] \right] \\ &= \mathbf{E} \left[\sum_{k \in A} D_k \mathbf{E}[F | \mathcal{F}_k] D_k \mathbf{E}[G | \mathcal{F}_k] \right] \\ &= \mathbf{E} \left[\sum_{k \in A} D_k F D_k \mathbf{E}[G | \mathcal{F}_k] \right] \end{aligned}$$

where we have used (5) in the third equality and the identity $D_k D_k = D_k$ in the last one. \square

Theorem 2.16 (Concentration inequality). *Let F for which there exists an order on A with*

$$M = \sup_{X \in E_A} \sum_{k=1}^{\infty} |D_k F(X)| \mathbf{E}[|D_k F(X)| | \mathcal{F}_k] < \infty.$$

Then, for any $x \geq 0$, we have

$$\mathbf{P}(F - \mathbf{E}[F] \geq x) \leq \exp\left(-\frac{x^2}{2M}\right).$$

Proof. Assume with no loss of generality that F is centered. Apply (9) to θF and $e^{\theta F}$,

$$\begin{aligned} \theta \left| \mathbf{E}[F e^{\theta F}] \right| &= \theta \left| \mathbf{E} \left[\sum_{k \in A} D_k F D_k \mathbf{E}[e^{\theta F} | \mathcal{F}_k] \right] \right| \\ &\leq \theta \sum_{k \in A} \mathbf{E} \left[|D_k F| \left| D_k \mathbf{E}[e^{\theta F} | \mathcal{F}_k] \right| \right]. \end{aligned}$$

Recall that

$$\begin{aligned} D_k \mathbf{E}[e^{\theta F} | \mathcal{F}_k] &= \mathbf{E}' \left[\mathbf{E}[e^{\theta F} | \mathcal{F}_k] - \mathbf{E}[e^{\theta F(X_{-k}, X'_k)} | \mathcal{F}_k] \right] \\ &= \mathbf{E} \left[\mathbf{E}'[e^{\theta F} - e^{\theta F(X_{-k}, X'_k)}] \mid \mathcal{F}_k \right] \\ &= \mathbf{E} \left[e^{\theta F} \mathbf{E}'[1 - e^{-\theta \Delta_k F}] \mid \mathcal{F}_k \right] \end{aligned}$$

where $\Delta_k F = F - F(X_{-k}, X'_k)$ so that $D_k F = \mathbf{E}'[\Delta_k F]$.

Since $(x \mapsto 1 - e^{-x})$ is concave, we get

$$D_k \mathbf{E}[e^{\theta F} | \mathcal{F}_k] \leq \mathbf{E} \left[e^{\theta F} (1 - e^{-\theta D_k F}) \mid \mathcal{F}_k \right] \leq \theta \mathbf{E} \left[e^{\theta F} |D_k F| \mid \mathcal{F}_k \right].$$

Thus,

$$\left| \mathbf{E} \left[F e^{\theta F} \right] \right| \leq \theta \mathbf{E} \left[e^{\theta F} \sum_{k=1}^{\infty} |D_k F| \mathbf{E} [|D_k F| | \mathcal{F}_k] \right] \leq M \theta \mathbf{E} \left[e^{\theta F} \right].$$

By Gronwall lemma, this implies that

$$\mathbf{E} \left[e^{\theta F} \right] \leq \exp\left(\frac{\theta^2}{2} M\right).$$

Hence,

$$\mathbf{P}(F - \mathbf{E}[F] \geq x) = \mathbf{P}(e^{\theta(F - \mathbf{E}[F])} \geq e^{\theta x}) \leq \exp(-\theta x + \frac{\theta^2}{2} M).$$

Optimize with respect to θ gives $\theta_{\text{opt}} = x/M$, hence the result. \square

In the Gaussian case, the concentration inequality is deduced from the logarithmic Sobolev inequality. This does not seem to be feasible in the present context. However, we still have an LSI identity.

Theorem 2.17 (Logarithmic Sobolev inequality). *Let a positive random variable $G \in L \log L(E_A)$. Then,*

$$(10) \quad \mathbf{E} [G \log G] - \mathbf{E} [G] \log \mathbf{E} [G] \leq \sum_{k \in A} \mathbf{E} \left[\frac{|D_k G|^2}{\mathbf{E} [G | \mathcal{G}_k]} \right].$$

Proof. We follow closely the proof of [32] for Poisson process. Let $G \in L^2(E_A)$ be a positive random variable such that $DG \in L^2(E_A \times A)$. For any integer N , define $G_N = \min(\max(\frac{1}{N}, G), N)$, for any k , $L_k = \mathbf{E} [G_N | \mathcal{F}_k]$ and $L_0 = \mathbf{E} [G_N]$. We have,

$$\begin{aligned} L_n \log L_n - L_0 \log L_0 &= \sum_{k=0}^{n-1} L_{k+1} \log L_{k+1} - L_k \log L_k \\ &= \sum_{k=0}^{n-1} \log L_k (L_{k+1} - L_k) + \sum_{k=0}^{n-1} L_{k+1} (\log L_{k+1} - \log L_k). \end{aligned}$$

Note that $(\log L_k (L_{k+1} - L_k), k \geq 0)$ and $(L_{k+1} - L_k, k \geq 0)$ are martingales, hence

$$\begin{aligned} &\mathbf{E} [L_n \log L_n - L_0 \log L_0] \\ &= \mathbf{E} \left[\sum_{k=0}^{n-1} L_{k+1} \log L_{k+1} - L_{k+1} \log L_k - L_{k+1} + L_k \right] \\ &= \mathbf{E} \left[\sum_{k=0}^{n-1} L_{k+1} \log L_{k+1} - L_k \log L_k - (\log L_k + 1)(L_{k+1} - L_k) \right] \\ &= \mathbf{E} \left[\sum_{k=0}^{n-1} \ell(L_k, L_{k+1} - L_k) \right], \end{aligned}$$

where the function ℓ is defined on $\Theta = \{(x, y) \in \mathbf{R}^2 : x > 0, x + y > 0\}$ by

$$\ell(x, y) = (x + y) \log(x + y) - x \log x - (\log x + 1)y.$$

Since ℓ is convex on Θ , it comes from the Jensen inequality for conditional expectations that

$$\begin{aligned}
\sum_{k=0}^{n-1} \mathbf{E}[\ell(L_k, L_{k+1} - L_k)] &= \sum_{k=0}^{n-1} \mathbf{E}[\ell(\mathbf{E}[G_n | \mathcal{F}_k], D_{k+1} \mathbf{E}[G_n | \mathcal{F}_{k+1}])] \\
&= \sum_{k=1}^n \mathbf{E}[\ell(\mathbf{E}[G_n | \mathcal{F}_{k-1}], \mathbf{E}[D_k G_n | \mathcal{F}_k])] \\
&\leq \sum_{k=1}^n \mathbf{E}[\mathbf{E}[\ell(\mathbf{E}[G_n | \mathcal{G}_k], D_k G_n) | \mathcal{F}_k]] \\
&= \sum_{k=1}^n \mathbf{E}[\ell(\mathbf{E}[G_n | \mathcal{G}_k], D_k G_n)] \\
&= \sum_{k=1}^{\infty} \mathbf{E}[\ell(\mathbf{E}[G_n | \mathcal{G}_k], D_k G_n)].
\end{aligned}$$

We know from [32] that for any integer k , $\ell(\mathbf{E}[G_n | \mathcal{G}_k], D_k G_n)$ converges increasingly to $\ell(\mathbf{E}[G | \mathcal{G}_k], D_k G)$ \mathbf{P} -a.s., hence by Fatou Lemma,

$$\mathbf{E}[G \log G] - \mathbf{E}[G] \log \mathbf{E}[G] \leq \sum_{k=1}^{\infty} \mathbf{E}[\ell(\mathbf{E}[G | \mathcal{G}_k], D_k G)].$$

Furthermore, for any $(x, y) \in \Theta$, $\ell(x, y) \leq |y|^2/x$, then,

$$\mathbf{E}[G \log G] - \mathbf{E}[G] \log \mathbf{E}[G] \leq \sum_{k=1}^{\infty} \mathbf{E}\left[\frac{|D_k G|^2}{\mathbf{E}[G | \mathcal{G}_k]}\right].$$

The proof is thus complete. \square

2.2. Divergence.

Definition 2.18 (Divergence). *Let*

$$\begin{aligned}
\text{Dom } \delta &= \left\{ U \in L^2(A \times E_A) : \right. \\
&\quad \left. \exists c > 0, \forall F \in \mathbf{D}, |\langle DF, U \rangle_{L^2(A \times E_A)}| \leq c \|F\|_{L^2(E_A)} \right\}.
\end{aligned}$$

For any U belonging to $\text{Dom } \delta$, δU is the element of $L^2(E_A)$ characterized by the following identity

$$\langle DF, U \rangle_{L^2(A \times E_A)} = \mathbf{E}[F \delta U], \text{ for all } F \in \mathbf{D}.$$

The integration by parts formula (2) entails that for every $U \in \text{Dom } \delta$,

$$\delta U = \sum_{a \in A} D_a U_a.$$

Definition 2.19. *The Hilbert space $\mathbf{D}(l^2(A))$ is the closure of $\mathcal{S}_0(l^2(A))$ with respect to the norm*

$$\|U\|_{\mathbf{D}(l^2(A))}^2 = \mathbf{E}\left[\sum_{a \in A} |U_a|^2\right] + \mathbf{E}\left[\sum_{a \in A} \sum_{b \in A} |D_a U_b|^2\right].$$

In particular, this means that the map $DU = (D_a U_b, a, b \in A)$ is Hilbert-Schmidt as a map from $L^2(A \times E_A)$ into itself. As a consequence, for two such maps DU and DV , the map $DU \circ DV$ is trace-class (see [33]) with

$$\text{trace}(DU \circ DV) = \sum_{a,b \in A} D_a U_b D_b V_a.$$

Theorem 2.20. *The space $\mathbf{D}(l^2(A))$ is included in $\text{Dom } \delta$. For any U, V belonging to $\mathbf{D}(l^2(A))$,*

$$(11) \quad \mathbf{E}[\delta U \delta V] = \mathbf{E}[\text{trace}(DU \circ DV)].$$

Proof. For U and V in $\mathcal{S}_0(l^2(A))$, from the integration by parts formula,

$$\begin{aligned} \mathbf{E}[\delta U \delta V] &= \langle D\delta(U), V \rangle_{L^2(A \times E_A)} \\ &= \mathbf{E} \left[\sum_{a \in A} D_a(\delta U) V_a \right] \\ &= \mathbf{E} \left[\sum_{(a,b) \in A^2} V_a D_a U_b \right] - \mathbf{E} \left[\sum_{(a,b) \in A^2} V_a \mathbf{E}[D_a U_b | \mathcal{G}_a] \right] \\ &= \mathbf{E} \left[\sum_{(a,b) \in A^2} V_a D_a U_b \right] - \mathbf{E} \left[\sum_{(a,b) \in A^2} \mathbf{E}[V_a | \mathcal{G}_a] D_a U_b \right] \\ &= \mathbf{E} \left[\sum_{(a,b) \in A^2} D_b V_a D_a U_b \right] \\ &= \mathbf{E}[\text{trace}(DU \circ DV)]. \end{aligned}$$

It follows that $\mathbf{E}[\delta U^2] \leq \|U\|_{\mathbf{D}(l^2(A))}^2$. Then, by density, $\mathbf{D}(l^2(A)) \subset \text{Dom } \delta$ and Eqn. (11) holds for U and V in $\text{Dom } \delta$. \square

Theorem 2.21 (Helmholtz decomposition). *Let $U \in \mathbf{D}(l^2(A))$. There exists a unique couple (φ, V) where $\varphi \in L^2(E_A)$ and $V \in L^2(A \times E_A)$ such that $\mathbf{E}[\varphi] = 0$, $\delta V = 0$ and*

$$U_a = D_a \varphi + V_a$$

for any $a \in A$.

Proof. We first prove the uniqueness. Let (φ, V) and (φ', V') two convenient couples. We have $D_a(\varphi - \varphi') = V'_a - V_a$ for any $a \in A$ and $\sum_{a \in A} D_a(V'_a - V_a) = 0$, hence

$$\begin{aligned} 0 &= \mathbf{E} \left[(\varphi - \varphi') \sum_{a \in A} D_a(V'_a - V_a) \right] = \mathbf{E} \left[\sum_{a \in A} D_a(\varphi - \varphi')(V'_a - V_a) \right] \\ &= \mathbf{E} \left[\sum_{a \in A} (V'_a - V_a)^2 \right]. \end{aligned}$$

This implies that $V = V'$ and $D(\varphi - \varphi') = 0$. The Clark formula (Theorem 2.10) entails that $0 = \mathbf{E}[\varphi - \varphi'] = \varphi - \varphi'$.

We now prove the existence. Since $\mathbf{E}[D_a\varphi | \mathcal{G}_a] = 0$, we can choose

$$V_a = \mathbf{E}[U_a | \mathcal{G}_a],$$

which implies $D_a\varphi = D_aU_a$, and guarantees $\delta V = 0$. Choose any ordering of the elements of A and remark that, in view of (5),

$$\begin{aligned} \mathbf{E} \left[\left(\sum_{k=1}^{\infty} \mathbf{E}[D_k U_k | \mathcal{F}_k] \right)^2 \right] &= \mathbf{E} \left[\left(\sum_{k=1}^{\infty} D_k \mathbf{E}[U_k | \mathcal{F}_k] \right)^2 \right] \\ &= \mathbf{E} \left[\sum_{k=1}^{\infty} \left(D_k \mathbf{E}[U_k | \mathcal{F}_k] \right)^2 \right] \leq \sum_{k=1}^{\infty} \mathbf{E}[|D_k U_k|^2] \leq \|U\|_{\mathbf{D}(L^2(A))}^2, \end{aligned}$$

hence

$$\varphi = \sum_{k=1}^{\infty} \mathbf{E}[D_k U_k | \mathcal{F}_k],$$

defines a square integrable random variable of null expectation, which satisfies the required property. \square

2.3. Ornstein-Uhlenbeck semi-group and generator. Having defined a gradient and a divergence, one may consider the Laplacian-like operator defined by $L = -\delta D$, which is also called the number operator in the settings of Gaussian Malliavin calculus.

Definition 2.22. *The number operator, denoted by L , is defined on its domain*

$$\text{Dom } L = \left\{ F \in L^2(E_A) : \mathbf{E} \left[\sum_{a \in A} |D_a F|^2 \right] < \infty \right\}$$

by

$$LF = -\delta DF = -\sum_{a \in A} D_a F.$$

The map L can be viewed as the generator of a symmetric Markov process X , which is ergodic, whose stationary probability is \mathbf{P}_A . Assume first that A is finite. Consider $(Z(t), t \geq 0)$ a Poisson process on the half-line of rate $|A|$, and the process $X(t) = (X_1(t), \dots, X_N(t), t \geq 0)$ which evolves according to the following rule: At a jump time of Z ,

- Choose randomly (with equiprobability) an index $a \in A$,
- Replace X_a by an independent random variable X'_a distributed according to \mathbf{P}_a .

For every $x \in E_A$, $a \in A$, set $x_{-a} = (x_1, \dots, x_{a-1}, x_{a+1}, \dots, x_{|A|})$. The generator of the Markov process X is clearly given by

$$|A| \sum_{a \in A} \frac{1}{|A|} \int_{E_a} \left(F(x_{-a}, x'_a) - F(x) \right) d\mathbf{P}_a(x'_a) = -\sum_{a \in A} D_a F(x).$$

The factor $|A|$ is due to the intensity of the Poisson process Z which jumps at rate $|A|$, the factor $|A|^{-1}$ is due to the uniform random choice of an index

$a \in A$. Thus, for a finite set A , L coincides with the generator of X . To extend this result to the case of a countable set, we apply the Hille-Yosida theorem:

Proposition 2.23 (Hille-Yosida). *A linear operator L on $L^2(E_A)$ is the generator of a strongly continuous contraction semigroup on $L^2(E_A)$ if and only if*

- (1) $\text{Dom } L$ is dense in $L^2(E_A)$.
- (2) L is dissipative i.e. for any $\lambda > 0, F \in \text{Dom } L$,
$$\|\lambda F - LF\|_{L^2(E_A)} \geq \lambda \|F\|_{L^2(E_A)}.$$
- (3) $\text{Im}(\lambda \text{Id} - L)$ dense in $L^2(E_A)$.

Proof. We know that $\mathcal{S} \subset \text{Dom } L$ and that \mathcal{S} is dense in $L^2(E_A)$, then so does $\text{Dom } L$.

Let $(A_n, n \geq 1)$ an increasing sequence of subsets of A such that $\cup_{n \geq 1} A_n = A$. For $F \in L^2(E_A)$, let $F_n = \mathbf{E}[F | \mathcal{F}_{A_n}]$. Since $(F_n, n \geq 1)$ is a square integrable martingale, F_n converges to F both almost-surely and in $L^2(E_A)$. For any $n \geq 1$, F_n depends only on X_{A_n} . Abusing the notation, we still denote by F_n its restriction to E_{A_n} so that we can consider $L_n F_n$ where L_n is defined as above on E_{A_n} . Moreover, according to Lemma 2.9, $D_a F_n = \mathbf{E}[D_a F | \mathcal{F}_{A_n}]$, hence

$$\begin{aligned} \lambda^2 \|F_n\|_{L^2(E_A)}^2 &\leq \|\lambda F_n - L_n F_n\|_{L^2(E_{A_n})}^2 = \mathbf{E} \left[\left(\lambda F_n - \sum_{a \in A} D_a F_n \right)^2 \right] \\ &= \mathbf{E} \left[\mathbf{E} \left[\lambda F - \sum_{a \in A} D_a F \mid \mathcal{F}_{A_n} \right]^2 \right] \xrightarrow{n \rightarrow \infty} \|\lambda F - LF\|_{L^2(E_A)}^2. \end{aligned}$$

Therefore, point (2) is satisfied.

Since A_n is finite, there exists $G_n \in L^2(E_{A_n})$ such that

$$\begin{aligned} F_n &= (\lambda \text{Id} - L_n) G_n(X_{A_n}) = \lambda G_n(X_{A_n}) - \sum_{a \in A_n} D_a G_n(X_{A_n}) \\ &= \lambda \tilde{G}_n(X_A) - \sum_{a \in A_n} D_a \tilde{G}_n(X_A) = \lambda \tilde{G}_n(X_A) - \sum_{a \in A} D_a \tilde{G}_n(X_A), \end{aligned}$$

where $\tilde{G}_n(X_A) = G_n(X_{A_n})$ depends only on the components whose index belongs to A_n . This means that F_n belongs to the range of $\lambda \text{Id} - L$ and we already know it converges in $L^2(E_A)$ to F . \square

Definition 2.24. *The Ornstein-Uhlenbeck semi-group of generator L , denoted $(P_t)_{t \geq 0}$, is defined on $L^2(E_A)$ by*

$$P_t F(x) = \mathbf{E}[F(X(t)) | X(0) = x]$$

for any $x \in E_A$.

From the sample-path construction of X , the next result is straightforward for A finite and can be obtained by a limiting process for A denumerable.

Theorem 2.25 (Mehler formula). *For $a \in A$, $x_a \in E_A$ and $t > 0$, let $X_a(x_a, t)$ the random variable defined by*

$$X_a(x_a, t) = \begin{cases} x_a & \text{with probability } (1 - e^{-t}), \\ X'_a & \text{with probability } e^{-t}, \end{cases}$$

where X'_a is a \mathbf{P}_a -distributed random variable independent from everything else. In other words, if $P_a^{x_a, t}$ denotes the distribution of $X_a(x_a, t)$, $P_a^{x_a, t}$ is a convex combination of ε_{x_a} and \mathbf{P}_a :

$$P_a^{x_a, t} = (1 - e^{-t}) \varepsilon_{x_a} + e^{-t} \mathbf{P}_a.$$

For any $x \in E_A$, any $t > 0$,

$$P_t F(x) = \int_{E_A} F(y) \otimes_{a \in A} d\mathbf{P}_a^{x_a, t}(y_a).$$

It follows easily that $(P_t, t \geq 0)$ is ergodic and stationary:

$$\lim_{t \rightarrow \infty} P_t F(x) = \int_{E_A} F d\mathbf{P}_a \text{ and } X(0) \stackrel{\text{law}}{=} \mathbf{P}_a \implies X(t) \stackrel{\text{law}}{=} \mathbf{P}_a.$$

Theorem 2.26. *Let $F \in L^2(E_A)$. For every $a \in A$, $x \in E_A$,*

$$(12) \quad D_a P_t F(x) = e^{-t} P_t D_a F(x)$$

Proof. For A finite, denote by Z_a the Poisson process of intensity 1 which represents the time at which the a -th component is modified in the dynamics of X . Let $\tau_a = \inf\{t \geq 0, Z_a(t) \neq Z_a(0)\}$ and remark that τ_a is exponentially distributed with parameter 1, hence

$$\begin{aligned} \mathbf{E}[F(X(t)) \mathbf{1}_{t \geq \tau_a} \mid X(0) = x] &= (1 - e^{-t}) \mathbf{E} \left[\int_{E_a} F(X_{-a}(t), x'_a) d\mathbf{P}_a(x'_a) \mid X(0) = x \right] \\ &= (1 - e^{-t}) \mathbf{E}[\mathbf{E}[F(X(t)) \mid \mathcal{G}_a] \mid X(0) = x] \\ &= \mathbf{E}[\mathbf{E}[F(X(t)) \mid \mathcal{G}_a] \mathbf{1}_{t \geq \tau_a} \mid X(0) = x]. \end{aligned}$$

Then,

$$\begin{aligned} D_a P_t F(x) &= P_t F(x) - \mathbf{E}[P_t F(x) \mid \mathcal{G}_a] \\ &= \mathbf{E}[(F(X(t)) - \mathbf{E}[F(X(t)) \mid \mathcal{G}_a]) \mathbf{1}_{t < \tau_a} \mid X(0) = x] \\ &\quad + \mathbf{E}[(F(X(t)) - \mathbf{E}[F(X(t)) \mid \mathcal{G}_a]) \mathbf{1}_{t \geq \tau_a} \mid X(0) = x] \\ &= e^{-t} P_t D_a F(x). \end{aligned}$$

For A infinite, let $(A_n, n \geq 1)$ an increasing sequence of finite subsets of A such that $\cup_{n \geq 1} A_n = A$. For $F \in L^2(E_A)$, let $F_n = \mathbf{E}[F \mid \mathcal{F}_{A_n}]$. Since P is a contraction semi-group, for any t , $P_t F_n$ tends to $P_t F$ in $L^2(E_A)$ as n goes to infinity. From the Mehler formula, we know that $P_t F_n = P_t^n F_n$ where P^n is the semi-group associated to A_n , hence

$$(13) \quad D_a P_t F_n = D_a P_t^n F_n = e^{-t} P_t^n D_a F_n.$$

Moreover,

$$\begin{aligned}
\mathbf{E} \left[\sum_{a \in A_n} |D_a P_t F_n|^2 \right] &= e^{-2t} \sum_{a \in A_n} \mathbf{E} [|P_t D_a F_n|^2] \\
&\leq e^{-2t} \sum_{a \in A_n} \mathbf{E} [|D_a F_n|^2] \\
&= e^{-2t} \sum_{a \in A_n} \mathbf{E} [|\mathbf{E} [D_a F | \mathcal{F}_{A_n}]|^2] \\
&\leq e^{-2t} \sum_{a \in A_n} \mathbf{E} [|D_a F|^2] \\
&\leq e^{-2t} \|DF\|_{\mathbf{D}}^2.
\end{aligned}$$

According to Lemma [2.7], this means that $P_t F$ belongs to \mathbf{D} . Let n go to infinity in (13) yields (12). \square

From (12), can be deduced another covariance identity.

Theorem 2.27. *For any $F, G \in \mathbf{D}$,*

$$(14) \quad \text{cov}(F, G) = \mathbf{E} \left[\sum_{k \in A} D_k F \int_0^\infty e^{-t} P_t \mathbf{E} [D_k G | \mathcal{F}_k] dt \right].$$

Proof. Let $F, G \in L^2(E_A)$.

$$\begin{aligned}
\text{cov}(F, G) &= \mathbf{E} \left[\sum_{k \in A} D_k \mathbf{E} [F | \mathcal{F}_k] D_k \mathbf{E} [G | \mathcal{F}_k] \right] \\
&= \mathbf{E} \left[\sum_{k \in A} D_k \mathbf{E} [F | \mathcal{F}_k] \left(- \int_0^\infty L P_t \mathbf{E} [G | \mathcal{F}_k] dt \right) \right] \\
&= \int_0^\infty \mathbf{E} \left[\sum_{k \in A} D_k \mathbf{E} [F | \mathcal{F}_k] \left(\sum_{l \in A} D_l P_t \mathbf{E} [G | \mathcal{F}_k] dt \right) \right] \\
&= \int_0^\infty e^{-t} \mathbf{E} \left[\sum_{k \in A} D_k F P_t D_k \mathbf{E} [G | \mathcal{F}_k] \right] dt
\end{aligned}$$

when we have used the orthogonality of the sum, (12) and the \mathcal{F}_k -measurability of $P_t D_k \mathbf{E} [G | \mathcal{F}_k]$ to get the last equality. \square

3. DIRICHLET STRUCTURES

For the definitions and properties of Dirichlet calculus, we refer to the first chapter of [7]. On (E_A, \mathbf{P}_A) , we have already implicitly built a Dirichlet structure, i.e. a Markov process X , a semi-group P and a generator L (see subsection 2.3). It remains to define the Dirichlet form \mathcal{E}_A such that $\mathcal{E}_A(F) = \mathbf{E} [F L F]$ for any sufficiently regular functional F .

Definition 3.1. For $F \in \mathbf{D}$, define

$$\mathcal{E}_A(F) = \mathbf{E} \left[\sum_{a \in A} |D_a F|^2 \right] = \|DF\|_{L^2(A \times E_A)}^2.$$

The integration by parts formula means that this form is closed. Since we do not assume any property on E_a for any $a \in A$ and since we do not seem to have a product rule formula for the gradient, we cannot assert more properties for \mathcal{E}_A . However, following [8], we now show that we can reconstruct the usual gradient structures on Poisson and Wiener spaces as well chosen limits of our construction. For these two situations, we have a Polish space W , equipped with \mathcal{B} its borelean σ -field and a probability measure \mathbf{P} . There also exists a Dirichlet form \mathcal{E} defined on a set of functionals \mathbf{D} . Let (E_N, \mathcal{A}_N) be a sequence of Polish spaces, all equipped with a probability measure \mathbf{P}_N and their own Dirichlet form \mathcal{E}_N , defined on \mathbf{D}_N . Consider maps U_N from E_N into W such that $(U_N)_* \mathbf{P}_N$, the pullback measure of \mathbf{P}_N by U_N , converges in distribution to \mathbf{P} . We assume that for any $F \in \mathbf{D}$, the map $F \circ U_N$ belongs to \mathbf{D}_N . The image Dirichlet structure is defined as follows. For any $F \in \mathbf{D}$,

$$\mathcal{E}^{U_N}(F) = \mathcal{E}_N(F \circ U_N).$$

We adapt the following definition from [8].

Definition 3.2. With the previous notations, we say that $((U_N)_* \mathbf{P}_N, N \geq 1)$ converges as a Dirichlet distribution whenever for any $F \in \text{Lip} \cap \mathbf{D}$,

$$\lim_{N \rightarrow \infty} \mathcal{E}^{U_N}(F) = \mathcal{E}(F).$$

3.1. Poisson point process. Let \mathbb{Y} be a compact Polish space and $\mathfrak{N}_{\mathbb{Y}}$ be the set of weighted configurations, i.e. the set of locally finite, integer valued measures on \mathbb{Y} . Such a measure is of the form

$$\omega = \sum_{n=1}^{\infty} p_n \varepsilon_{\zeta_n},$$

where $(\zeta_n, n \geq 1)$ is a set of distinct points in \mathbb{Y} with no accumulation point, $(p_n, n \geq 1)$ any sequence of positive integers. The topology on $\mathfrak{N}_{\mathbb{Y}}$ is defined by the semi-norms

$$p_f(\omega) = \left| \sum_{n=1}^{\infty} p_n f(\zeta_n) \right|,$$

when f runs through the set of continuous functions on \mathbb{Y} . It is known (see for instance [17]) that $\mathfrak{N}_{\mathbb{Y}}$ is then a Polish space for this topology. For some finite measure \mathbf{M} on \mathbb{Y} , we put on $\mathfrak{N}_{\mathbb{Y}}$, the probability measure \mathbf{P} such that the canonical process is a Poisson point process of control measure \mathbf{M} , which we consider without loss of generality, to have total mass $\mathbf{M}(\mathbb{Y}) = 1$.

On $\mathfrak{N}_{\mathbb{Y}}$, it is customary to consider the difference gradient (see [14, 23, 27]): For any $x \in \mathbb{Y}$, any $\omega \in \mathfrak{N}_{\mathbb{Y}}$,

$$D_x F(\omega) = F(\omega + \varepsilon_x) - F(\omega).$$

Set

$$\mathbf{D}_P = \left\{ F : \mathfrak{N}_Y \rightarrow \mathbf{R} \text{ such that } \mathbf{E} \left[\int_Y |D_x F|^2 d\mathbf{M}(y) \right] < \infty \right\}$$

and for any $F \in \mathbf{D}_P$,

$$(15) \quad \mathcal{E}(F) = \mathbf{E} \left[\int_Y |D_x F|^2 d\mathbf{M}(y) \right].$$

To see the Poisson point process as a Dirichlet limit, the idea is to partition the set Y into N parts, C_1^N, \dots, C_N^N such that $\mathbf{M}(C_k^N) = p_k^N$ and then for each $k \in \{1, \dots, N\}$, take a point ζ_k^N into C_k^N so that the Poisson point process ω on Y with intensity measure \mathbf{M} is approximated by

$$\omega^N = \sum_{k=1}^N \omega(C_k^N) \varepsilon_{\zeta_k^N}.$$

We denote by \mathbf{P}_N the distribution of ω^N . By computing its Laplace transform, it is clear that \mathbf{P}_N converges in distribution to \mathbf{P} . It remains to see this convergence holds in the Dirichlet sense for the sequence of Dirichlet structures induced by our approach for independent random variables.

Let $(\zeta_k^N, k = 1, \dots, N)$ (respectively $(p_k^N, k = 1, \dots, N)$) be a triangular array of points in Y (respectively of non-negative numbers) such that the following two properties hold: 1) the p_k^N 's tends to 0 uniformly:

$$(16) \quad p^N = \sup_{k \leq N} p_k^N = O\left(\frac{1}{N}\right);$$

2) the ζ_k^N 's are sufficiently well spread so that we have convergence of Riemann sums: For any continuous and \mathbf{M} -integrable function $f : Y \rightarrow \mathbf{R}$, we have

$$(17) \quad \sum_{k=1}^N f(\zeta_k^N) p_k^N \xrightarrow{N \rightarrow \infty} \int f(x) d\mathbf{M}(x).$$

Take $f = 1$ implies that $\sum_k p_k^N$ tends to 1 as N goes to infinity.

For any N and any $k \in \{1, \dots, N\}$, let μ_k^N be the Poisson distribution on \mathbf{N} , of parameter p_k^N . In this situation, let $E_N = \mathbf{N}^N$ with $\mu^N = \otimes_{k=1}^N \mu_k^N$. That means we have independent random variables M_1^N, \dots, M_N^N , where M_k^N follows a Poisson distribution of parameter p_k^N for any $k \in \{1, \dots, N\}$. We turn these independent random variables into a point process by the map U_N defined as

$$U_N : \mathbf{N}^N \longrightarrow \mathfrak{N}_Y$$

$$(m_1, \dots, m_N) \longmapsto \sum_{k=1}^N m_k \varepsilon_{\zeta_k^N}.$$

Lemma 3.3. *For any $F \in \mathbf{D}_P$,*

$$(18) \quad \mathcal{E}^{U_N}(F) = \sum_{m=1}^N \sum_{\ell=0}^{\infty} \mathbf{E} \left[\left(\sum_{\tau=0}^{\infty} \left(F(\omega_{(m)}^N + \ell \varepsilon_{\zeta_m^N}) - F(\omega_{(m)}^N + \tau \varepsilon_{\zeta_m^N}) \right) \mu_m^N(\tau) \right)^2 \right] \mu_m^N(\ell),$$

where $\omega_{(m)}^N = \sum_{k \neq m} M_k^N \varepsilon_{\zeta_k^N}$.

Proof. According its very definition,

$$\mathcal{E}^{U_N}(F) = \sum_{m=1}^N \mathbf{E} \left[\left(F(\omega_{(m)}^N + M_m^N \varepsilon_{\zeta_m^N}) - \sum_{\tau=0}^{\infty} F(\omega_{(m)}^N + \tau \varepsilon_{\zeta_m^N}) \mu_m^N(\tau) \right)^2 \right].$$

The result follows by conditioning with respect to M_m^N , whose law is μ_m^N . \square

Since the vague topology on $\mathfrak{N}_{\mathbb{Y}}$ is metrizable, one could define Lipschitz functions with respect to this distance. However, this turns out to be not sufficient for the convergence to hold.

Definition 3.4. *A function $F : \mathfrak{N}_{\mathbb{Y}} \rightarrow \mathbf{R}$ is said to be TV – Lip if F is continuous for the vague topology and if for any $\omega, \eta \in \mathfrak{N}_{\mathbb{Y}}$,*

$$|F(\omega) - F(\eta)| \leq \text{dist}_{TV}(\omega, \eta),$$

where dist_{TV} represents the distance in total variation between two point measures, i.e. the number of distinct points counted with multiplicity.

Theorem 3.5. *For any $F \in \text{TV – Lip} \cap \mathbf{D}_P$, with the notations of Lemma [3.3] and (15),*

$$\mathcal{E}^{U_N}(F) \xrightarrow{N \rightarrow \infty} \mathcal{E}(F).$$

Proof. Starting from (18), the terms with $\tau = 0$ can be decomposed as

$$e^{-2p_m^N} \sum_{m=1}^N \mathbf{E} \left[\left(F(\omega_{(m)}^N + \varepsilon_{\zeta_m^N}) - F(\omega_{(m)}^N) \right)^2 \right] \mu_m^N(1) + R_0^N.$$

Since F belongs to TV – Lip,

$$R_0^N \leq \sum_{m=1}^N \sum_{\ell=2}^{\infty} \ell^2 \mu_m^N(\ell) \leq c_1 N (p^N)^2 \mathbf{E} [(\text{Poisson}(p^N) + 2)^2] \leq c_2 N (p^N)^2,$$

where the c_1 and c_2 are irrelevant constants. As Np^N is bounded, R_0^N goes to 0 as N grows to infinity. For the very same reasons, the sum of the terms of (18) with $\tau \geq 1$ converge to 0, thus

$$\lim_{N \rightarrow \infty} \mathcal{E}^{U_N}(F) = \lim_{N \rightarrow \infty} \sum_{m=1}^N e^{-2p_m^N} \mathbf{E} \left[\left(F(\omega_{(m)}^N + \varepsilon_{\zeta_m^N}) - F(\omega_{(m)}^N) \right)^2 \right] p_m^N.$$

Consider now the space $\mathfrak{N}_{\mathbb{Y}}^{\zeta} = \mathfrak{N}_{\mathbb{Y}} \times \{\zeta_k^N, k = 1, \dots, N\}$ with the product topology and probability measure $\tilde{\mathbf{P}}_N = \mathbf{P}_N \otimes \sum_k p_m^N \varepsilon_{\zeta_k^N}$. Let

$$\psi : \mathfrak{N}_{\mathbb{Y}} \times \{\zeta_k^N, k = 1, \dots, N\} \longrightarrow E$$

$$(\omega, \zeta) \longmapsto \left(F(\omega - (\omega(\zeta) - 1)\varepsilon_{\zeta}) - F(\omega - \omega(\zeta)\varepsilon_{\zeta}) \right)^2.$$

Then, we can write

$$\sum_{m=1}^N \mathbf{E} \left[\left(F(\omega_{(m)}^N + \varepsilon_{\zeta_m^N}) - F(\omega_{(m)}^N) \right)^2 \right] p_m^N = \int_{\mathfrak{N}_{\mathbb{Y}}^{\zeta}} \psi(\omega, \eta) d\tilde{\mathbf{P}}_N(\omega, \zeta).$$

Under $\tilde{\mathbf{P}}_N$, the random variables ω and ζ are independent. Equation (17) means that the marginal distribution of ζ tends to \mathbf{M} (assumed to be a probability measure at the very beginning of this construction). Moreover, we already know that \mathbf{P}_N converges in distribution to \mathbf{P} . Hence, $\tilde{\mathbf{P}}_N$ tends to $\mathbf{P} \otimes \mathbf{M}$ as N goes to infinity. Since F is in $\text{TV} - \text{Lip}$, ψ is continuous and bounded, hence the result. \square

3.2. Donsker theorem. For details on Gaussian Malliavin calculus, we refer to [22, 31]. We now consider \mathbf{P} as the Wiener measure on $W = \mathcal{C}_0([0, 1]; \mathbf{R})$. Let $(h_k, k \geq 1)$ be an orthonormal basis of the Cameron-Martin space H ,

$$H = \{f : [0, 1] \rightarrow \mathbf{R}, \exists \dot{f} \in L^2 \text{ with } f(t) = \int_0^t \dot{f}(s) ds\} \text{ and } \|f\|_H = \|\dot{f}\|_{L^2}.$$

A function $F : W \rightarrow \mathbf{R}$ is said to be cylindrical if it is of the form

$$F = f(\delta_B v_1, \dots, \delta_B v_n),$$

where v_1, \dots, v_n belong to H , $\delta_B v$ is the Wiener integral of v and f belongs to the Schwartz space $\mathcal{S}(\mathbf{R}^n)$. For $h \in H$,

$$\nabla_h F(w) = \sum_{k=1}^n \frac{\partial f}{\partial x_k}(\delta_B v_1, \dots, \delta_B v_n) h_k.$$

The map ∇ is closable from $L^2(W; \mathbf{R})$ to $L^2(W; H)$. Thus, it is meaningful to define \mathbf{D}_B as the closure of cylindrical functions for the norm

$$\|F\|_{1,2} = \|F\|_2 + \|\nabla F\|_{L^2(W; H)}.$$

Definition 3.6. A function $F : W \rightarrow \mathbf{R}$ is said to be H-C^1 if

- for almost all $\omega \in W$, $h \longmapsto F(\omega + h)$ is a continuous function on H ,
- for almost all $\omega \in W$, $h \longmapsto F(\omega + h)$ is continuously Fréchet differentiable and this Fréchet derivative is continuous from H into $\mathbf{R} \otimes H$.

We still denote by ∇F the element of H such that

$$\left. \frac{d}{d\tau} F(\omega + \tau h) \right|_{\tau=0} = \langle \nabla F(\omega), h \rangle_H.$$

For $N \geq 1$, let

$$e_k^N(t) = \sqrt{N} \mathbf{1}_{[(k-1)/N, k/N)}(t) \text{ and } h_k^N(t) = \int_0^t e_k^N(s) ds.$$

The family $(h_k^N, k \geq 1)$ is then orthonormal in H . For $(M_k, k \geq 1)$ a sequence of independent identically distributed random variables, centered with unit variance, the random walk

$$\omega^N(t) = \sum_{k=1}^N M_k h_k^N(t), \text{ for all } t \in [0, 1],$$

is known to converge in distribution in W to \mathbf{P} . Let $E_N = \mathbf{R}^N$ equipped with the product measure $\mathbf{P}_N = \otimes_{k=1}^N \nu$ where ν is the standard Gaussian measure on \mathbf{R} . We define the map U_N as follows:

$$\begin{aligned} U_N : E_N &\longrightarrow W \\ m = (m_1, \dots, m_N) &\longmapsto \sum_{k=1}^N m_k h_k^N. \end{aligned}$$

We can then restate the Donsker theorem by saying that $(U_N)_* \mathbf{P}_N$ converges to \mathbf{P} . It follows from our definition that:

Lemma 3.7. *For any $F \in L^2(W; \mathbf{R})$,*

$$\mathcal{E}^{U_N}(F) = \sum_{k=1}^N \mathbf{E} \left[\left(F(\omega^N) - \mathbf{E}' \left[F(\omega_{(k)}^N + M'_k h_k^N) \right] \right)^2 \right],$$

where $\omega_{(k)}^N = \omega^N - M_k h_k^N$ and M'_k is an independent copy of M_k . The expectation is taken on the product space \mathbf{R}^{N+1} equipped with the measure $\mathbf{P}_N \otimes \nu$.

The definition of Lipschitz function we use here is the following:

Definition 3.8. *A function $F : W \rightarrow \mathbf{R}$ is said to be Lipschitz if it is $\mathbf{H-C}^1$ and for almost all $\omega \in W$,*

$$|\langle \nabla F, h \rangle| \leq \|\dot{h}\|_{L^1}.$$

In particular since $e_k^N \geq 0$, this implies that

$$|\langle \nabla F, h_k^N \rangle| \leq h_k^N(1) - h_k^N(0) = \frac{1}{\sqrt{N}}.$$

For $F \in \mathbf{D}_B \cap \mathbf{H-C}^1$, we have

$$(19) \quad F(\omega + h) - F(\omega) = \langle \nabla F(\omega), h \rangle_H + \|\dot{h}\|_{L^1} \varepsilon(\omega, h),$$

where $\varepsilon(\omega, h)$ is bounded and goes to 0 in L^2 , uniformly with as $\|\dot{h}\|_{L^1}$ tends to 0.

Theorem 3.9. *For any $F \in \mathbf{D}_B \cap \mathbf{H-C}^1$,*

$$\mathcal{E}^{U_N}(F) \xrightarrow{N \rightarrow \infty} \mathbf{E} [\|\nabla F\|_H^2] = \mathcal{E}(F).$$

Proof. For $F \in \mathbf{D}_B \cap \mathbf{H}\text{-C}^1$, in view of (19), we have

$$\begin{aligned} F(\omega^N) - F(\omega_{(k)}^N + M'_k h_k^N) \\ = (M_k - M'_k) \langle \nabla F(\omega_{(k)}^N), h_k^N \rangle_H + \frac{|M_k - M'_k|}{\sqrt{N}} \varepsilon(\omega_{(k)}^N, h_k^N). \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{k=1}^N \mathbf{E} \left[\left(F(\omega^N) - \mathbf{E}' \left[F(\omega_{(k)}^N + M'_k h_k^N) \right] \right)^2 \right] \\ = \sum_{k=1}^N \mathbf{E} \left[\left(M_k \langle \nabla F(\omega_{(k)}^N), h_k^N \rangle_H + \mathbf{E}' \left[\frac{|M_k - M'_k|}{\sqrt{N}} \varepsilon(\omega_{(k)}^N, h_k^N) \right] \right)^2 \right] \\ = \sum_{k=1}^N \mathbf{E} \left[\langle \nabla F(\omega_{(k)}^N), h_k^N \rangle_H^2 \right] + \text{Rem}, \end{aligned}$$

and

$$\text{Rem} \leq \frac{c}{N} \sum_{k=1}^N \mathbf{E} \left[\varepsilon(\omega_{(k)}^N, h_k^N)^2 \right] \xrightarrow{N \rightarrow \infty} 0,$$

by the Césaro theorem. It follows that $\mathcal{E}^{U_N}(F)$ has the same limit as

$$\sum_{k=1}^N \mathbf{E} \left[\langle \nabla F(\omega_{(k)}^N), h_k^N \rangle_H^2 \right].$$

As N goes to infinity, we add more and more terms to the random walk, so that the influence of one particular term becomes negligible. The following result is well known [8, Proposition 3]: For any k , for any bounded ψ and φ ,

$$\mathbf{E} [\psi(M_k) \varphi(\omega^N)] \xrightarrow{N \rightarrow \infty} \mathbf{E} [\psi(M_k)] \mathbf{E} [\varphi(\omega)].$$

Since $\|\nabla F\|_H$ belongs to L^∞ and $\|h_k^N\|_\infty$ tends to 0, this entails that for any k ,

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbf{E} \left[\langle \nabla F(\omega_{(k)}^N), h_k^N \rangle_H^2 \right] &= \lim_{N \rightarrow \infty} \mathbf{E} \left[\langle \nabla F(\omega^N), h_k^N \rangle_H^2 \right] \\ &= \lim_{N \rightarrow \infty} \mathbf{E} \left[\|\pi_{V_N} \nabla F(\omega^N)\|_H^2 \right], \end{aligned}$$

where π_{V_N} is the orthogonal projection in H onto $\text{span}\{h_k^N, k = 1, \dots, N\}$. We conclude by dominated convergence. \square

4. APPLICATIONS TO PERMUTATIONS

For every integer N , denote by \mathfrak{S}_N the space of permutations on $\{1, \dots, N\}$. We always identify \mathfrak{S}_N as the subgroup of \mathfrak{S}_{N+1} stabilizing the element $N+1$. For every $k \in \{1, \dots, N\}$, define $\mathcal{J}_k = \{1, \dots, k\}$ and

$$\mathcal{J} = \mathcal{J}_1 \times \mathcal{J}_2 \times \dots \times \mathcal{J}_N.$$

The coordinate map from \mathcal{J} to \mathcal{J}_k is denoted by I_k . Following [18], we have

Theorem 4.1. *There exists a natural bijection Γ between \mathcal{J} and \mathfrak{S}_N .*

Proof. To a sequence (i_1, \dots, i_N) where $i_k \in \mathcal{J}_k$, we associate the permutation

$$\Gamma(i_1, \dots, i_N) = (N, i_N) \circ (N-1, i_{N-1}) \dots \circ (2, i_2).$$

where (i, j) denotes the transposition between the two elements i and j .

To an element $\sigma_N \in \mathfrak{S}_N$, we associate $i_N = \sigma_N(N)$. Then, N is a fixed point of $\sigma_{N-1} = (N, i_N) \circ \sigma_N$, hence it can be identified as an element σ_{N-1} of \mathfrak{S}_{N-1} . Then, $i_{N-1} = \sigma_{N-1}(N-1)$ and so on for decreasing indices.

It is then clear that Γ is one-to-one and onto. \square

In [18], Γ is described by the following rule: Start with permutation $\sigma_1 = (1)$, if at the N -th step of the algorithm, we have $i_N = N$ then the current permutation is extended by leaving N fixed, otherwise, N is inserted in σ_{N-1} just before i_N in the cycle of this element. This construction is reminiscent of the Chinese restaurant process (see [2]) where i_N is placed immediately after N . An alternative construction of permutations is known as the Feller coupling (see [2]). In our notations, it is given by

$$\sigma_1 = (1); \quad \sigma_N = \sigma_{N-1} \circ (\sigma_{N-1}^{-1}(i_N), N).$$

Definition 4.2 (Ewens distribution). *For some $t \in \mathbf{R}^+$, for any k , consider the measure \mathbf{P}_k defined on \mathcal{J}_k by*

$$\mathbf{P}_k(\{j\}) = \begin{cases} \frac{1}{t+k-1} & \text{if } j \neq k, \\ \frac{t}{t+k-1} & \text{for } j = k. \end{cases}$$

Under the distribution $\mathbf{P} = \otimes_k \mathbf{P}_k$, the random variables $(I_k, k = 1, \dots, N)$ are independent with law given by $\mathbf{P}(I_k = j) = \mathbf{P}_k(\{j\})$, for any k .

The Ewens distribution of parameter t on \mathfrak{S}_N , denoted by \mathbf{P}^t , is the pull-back of \mathbf{P} by the map Γ .

A moment of thought shows that a new cycle begins in the first construction for each index where $i_k = k$. Moreover, it can be shown that

Theorem 4.3 (see [18]). *For any $\sigma \in \mathfrak{S}_N$,*

$$\mathbf{P}^t(\{\sigma\}) = \frac{t^{cyc(\sigma)}}{(t+1)(t+2) \times \dots \times (t+N-1)},$$

where $cyc(\sigma)$ is the number of cycles of σ .

For any F , a measurable function on \mathfrak{S}_N , we have the following diagram

$$\begin{array}{ccc}
(\mathcal{I}, \otimes_{k=1}^N \mathbf{P}_k) & & \\
\Gamma \downarrow & \searrow \tilde{F} = F \circ \Gamma & \\
(\mathfrak{S}_N, \mathbf{P}^t) & \xrightarrow{F} & \mathbf{R}
\end{array}$$

We denote by $i = (i_1, \dots, i_N)$ a generic element of \mathcal{J} and by $\sigma = \Gamma(i)$.

Let $C_1(\sigma)$ denote the number of fixed points of the permutation σ and $\tilde{C}_1 = C_1 \circ \Gamma$. For any $k \in \mathcal{J}_N$, the random variable $U_k(\sigma)$ is the indicator of the event (k is a fixed point of σ) and let $\tilde{U}_k^N = U_k \circ \Gamma$. The Clark formula with reverse filtration shows that we can write \tilde{U}_k^N as a sum of centered orthogonal random variables as in the Hoeffding decomposition of U-statistics (see Theorem 2.12).

Theorem 4.4. *For any $k \in \{1, \dots, N\}$,*

$$(20) \quad \tilde{U}_k = \mathbf{1}_{(I_k=k)} \mathbf{1}_{(I_m \neq k, m \in \{k+1, \dots, N\})}.$$

and under \mathbf{P}^t , \tilde{U}_k^N is Bernoulli distributed with parameter $tp_k \alpha_k$, where for any $k \in \{1, \dots, N\}$,

$$p_k = \frac{1}{t+k-1} \text{ and } \alpha_k = \prod_{j=k+1}^N \frac{j-1}{t+j-1}.$$

Moreover,

$$\begin{aligned}
\tilde{U}_k^N &= tp_k \alpha_k + \left(\mathbf{1}_{(I_k=k)} - tp_k \right) \prod_{m=k+1}^N \mathbf{1}_{(I_m \neq k)} \\
&\quad - tp_k \sum_{j=1}^{N-k-1} \frac{t+k-1}{t+k+j-2} \left(\mathbf{1}_{(I_{k+j}=k)} - p_{k+j} \right) \prod_{l=j+1}^{N-k} \mathbf{1}_{(I_{k+l} \neq k)}.
\end{aligned}$$

Proof. By the previous construction, for

$$i = (i_1, \dots, i_N) \in (I_k = k) \cap \bigcap_{m=k+1}^N (I_m \neq k),$$

the permutation $\sigma = \Gamma(i)$ admits k as a fixed point. Hence,

$$\left\{ (I_k = k) \cap \bigcap_{m=k+1}^N (I_m \neq k) \right\} \subset (\tilde{U}_k^N = 1).$$

As both events have cardinal $(N-1)!$, they do coincide. The values of p_k and α_k are easily computed since the random variables $(I_m, k \leq m \leq N)$

are independent. According to Theorem 2.13,

$$\begin{aligned}\tilde{U}_k^N &= \mathbf{E} \left[\tilde{U}_k^N \right] + \sum_{l=1}^N D_l \mathbf{E} \left[\tilde{U}_k | \mathcal{H}_{l-1} \right] \\ &= \mathbf{E} \left[\tilde{U}_k^N \right] + \sum_{l=1}^N \mathbf{E} \left[\tilde{U}_k^N | \mathcal{H}_{l-1} \right] - \mathbf{E} \left[\tilde{U}_k^N | \mathcal{H}_l \right].\end{aligned}$$

Since $\tilde{U}_k^N \in \mathcal{H}_{k-1}$, $D_l \mathbf{E} \left[\tilde{U}_k | \mathcal{H}_{l-1} \right] = 0$ for $l < k$. For $l = k$, we get

$$\begin{aligned}\mathbf{E} \left[\mathbf{1}_{(I_k=k)} \prod_{m=k+1}^N \mathbf{1}_{(I_m \neq k)} | I_k, I_{k+1}, \dots \right] \\ - \mathbf{E} \left[\mathbf{1}_{(I_k=k)} \prod_{m=k+1}^N \mathbf{1}_{(I_m \neq k)} | I_{k+1}, I_{k+2}, \dots \right] \\ = \left(\mathbf{1}_{(I_k=k)} - \mathbf{P}_k(\{k\}) \right) \prod_{m=k+1}^N \mathbf{1}_{(I_m \neq k)}.\end{aligned}$$

For $l = k + 1$,

$$\begin{aligned}\mathbf{E} \left[\mathbf{1}_{(I_k=k)} \prod_{m=k+1}^N \mathbf{1}_{(I_m \neq k)} | I_{k+1}, I_{k+2}, \dots \right] \\ - \mathbf{E} \left[\mathbf{1}_{(I_k=k)} \prod_{m=k+1}^N \mathbf{1}_{(I_m \neq k)} | I_{k+2}, I_{k+3}, \dots \right] \\ = tp_k \left(\mathbf{1}_{(I_{k+1} \neq k)} - \mathbf{P}_{k+1}(\{k\}^c) \right) \prod_{m=k+2}^N \mathbf{1}_{(I_m \neq k)} \\ = -tp_k \left(\mathbf{1}_{(I_{k+1}=k)} - \mathbf{P}_{k+1}(\{k\}) \right) \prod_{m=k+2}^N \mathbf{1}_{(I_m \neq k)}.\end{aligned}$$

The subsequent terms are handled similarly and the result follows. \square

Since

$$\tilde{C}_1 = \sum_{k=1}^N \tilde{U}_k^N,$$

we retrieve the result of [3]:

$$\mathbf{E} \left[\tilde{C}_1 \right] = \frac{tN}{t + N - 1}.$$

and the chaos decomposition of \tilde{C}_1 can be easily deduced from the previous theorem.

Theorem 4.5. *We can write*

$$\begin{aligned}\tilde{C}_1 &= t \left(1 - \frac{t-1}{N+t-1} \right) + \sum_{l=1}^N D_l \tilde{U}_l^N + \sum_{l=2}^N \frac{t}{t+l-2} D_l \left(\sum_{k=1}^{l-1} \prod_{m=l}^N \mathbf{1}_{(I_m \neq k)} \right) \\ &= t \left(1 - \frac{t-1}{N+t-1} \right) + \sum_{l=1}^N \left(\mathbf{1}_{(I_l=l)} - \frac{t}{t+l-1} \right) \prod_{m=l+1}^N \mathbf{1}_{(I_m \neq l)} \\ &\quad - \sum_{l=2}^{N-1} \frac{t}{t+l-2} \sum_{k=1}^{l-1} \left(\mathbf{1}_{(I_l=k)} - \frac{1}{t+l-1} \right) \prod_{m=l+1}^N \mathbf{1}_{(I_m \neq k)}.\end{aligned}$$

Proof. By the very definition of \tilde{C}_1 , we have

$$(21) \quad \tilde{C}_1 = \mathbf{E} [\tilde{C}_1] + \sum_{k=1}^N \sum_{l=k}^N D_l \mathbf{E} [\tilde{U}_k^N | \mathcal{H}_{l-1}].$$

For $k = l$, $\mathbf{E} [\tilde{U}_k^N | \mathcal{H}_{l-1}] = \tilde{U}_k^N$ and for $l > k$,

$$\begin{aligned}\mathbf{E} [\tilde{U}_k^N | \mathcal{H}_{l-1}] &= \frac{t}{t+k-1} \left(1 - \frac{1}{t+k} \right) \cdots \left(1 - \frac{1}{t+l-2} \right) \prod_{m=l}^N \mathbf{1}_{(I_m \neq k)} \\ &= \frac{t}{t+l-2} \prod_{m=l}^N \mathbf{1}_{(I_m \neq k)}.\end{aligned}$$

It is straightforward that $l > k$,

$$\begin{aligned}D_l \left(\prod_{m=l}^N \mathbf{1}_{(I_m \neq k)} \right) &= \left(\mathbf{1}_{(I_l \neq k)} - \left(1 - \frac{1}{t+l-1} \right) \right) \prod_{m=l+1}^N \mathbf{1}_{(I_m \neq k)} \\ &= - \left(\mathbf{1}_{(I_l=k)} - \frac{1}{t+l-1} \right) \prod_{m=l+1}^N \mathbf{1}_{(I_m \neq k)}.\end{aligned}$$

The result then follows by direct computations. \square

This decomposition can be used to compute the variance of \tilde{C}_1 .

Theorem 4.6. *For any $t \in \mathbf{R}$, we get*

$$\text{var}[\tilde{C}_1] = \frac{Nt}{t+N-1} \left(\frac{t}{t+N-1} + 1 - \frac{2t^2}{N} \sum_{k=1}^N \frac{1}{t+k-1} \right).$$

Proof. Recall that for $j \neq l$, $D_l \mathbf{E} [\tilde{U}_k^N | \mathcal{H}_{l-1}]$ and $D_j \mathbf{E} [\tilde{U}_m^N | \mathcal{H}_{j-1}]$ are orthogonal in L^2 . In view of (21), according to the integration by parts formula,

we have

$$\begin{aligned}
\text{var} [\tilde{C}_1] &= \sum_{k=1}^N \sum_{m=1}^N \sum_{l=k}^N \sum_{j=m}^N \mathbf{E} \left[D_l \mathbf{E} \left[\tilde{U}_k^N \mid \mathcal{H}_{l-1} \right] D_j \mathbf{E} \left[\tilde{U}_m^N \mid \mathcal{H}_{j-1} \right] \right] \\
&= \sum_{k=1}^N \sum_{m=1}^N \sum_{l=k \vee m}^N \mathbf{E} \left[D_l \mathbf{E} \left[\tilde{U}_k^N \mid \mathcal{H}_{l-1} \right] D_l \mathbf{E} \left[\tilde{U}_m^N \mid \mathcal{H}_{l-1} \right] \right] \\
&= 2 \sum_{k=1}^N \sum_{m=k+1}^N \sum_{l=m}^N \mathbf{E} \left[U_k^N D_l \mathbf{E} \left[\tilde{U}_m^N \mid \mathcal{H}_{l-1} \right] \right] \\
&\quad + \mathbf{E} \left[\sum_{k=1}^N \sum_{l=k}^N \tilde{U}_k^N D_l \mathbf{E} \left[\tilde{U}_k^N \mid \mathcal{H}_{l-1} \right] \right].
\end{aligned}$$

Then, for $l \geq m > k$,

$$\begin{aligned}
&\mathbf{E} \left[U_k^N D_l \mathbf{E} \left[\tilde{U}_m^N \mid \mathcal{H}_{l-1} \right] \right] \\
&= -\frac{t}{t+l-2} \mathbf{E} \left[\mathbf{1}_{(I_k=k)} \prod_{p=k+1}^N \mathbf{1}_{(I_p \neq k)} \left(\mathbf{1}_{(I_l=m)} - \frac{1}{t+l-1} \right) \prod_{j=l+1}^N \mathbf{1}_{(I_j \neq m)} \right] \\
&= -\frac{t \mathbf{P}_k(\{k\})}{t+l-2} \left(\mathbf{P}_l(\{m\}) - \frac{1}{t+l-1} \right) \mathbf{E} \left[\prod_{p=k+1}^{l-1} \mathbf{1}_{(I_p \neq k)} \right] \mathbf{E} \left[\prod_{p=l+1}^N \mathbf{1}_{(I_p \notin \{k,m\})} \right] \\
&= 0,
\end{aligned}$$

since, for any $l \geq m > k$

$$\mathbf{E} \left[\mathbf{1}_{(I_l=m)} \mathbf{1}_{(I_l \neq k)} \right] = \mathbf{E} \left[\mathbf{1}_{(I_l=m)} \right] = \mathbf{P}_l(\{m\}) = \frac{1}{t+l-1}.$$

Furthermore, for $l > k$,

$$\begin{aligned}
&\mathbf{E} \left[\tilde{U}_k^N D_l \mathbf{E} \left[\tilde{U}_k^N \mid \mathcal{H}_{l-1} \right] \right] \\
&= -\frac{t}{t+l-2} \mathbf{E} \left[\mathbf{1}_{(I_k=k)} \prod_{p=k+1}^N \mathbf{1}_{(I_p \neq k)} \left(\mathbf{1}_{(I_l=k)} - \frac{1}{t+l-1} \right) \prod_{p=l+1}^N \mathbf{1}_{(I_p \neq k)} \right] \\
&= \frac{t}{(t+l-1)(t+l-2)} \mathbf{P}_k(\{k\}) \mathbf{E} \left[\prod_{p=k+1}^N \mathbf{1}_{(I_p \neq k)} \right] \\
&= \frac{t^2}{(t+l-1)(t+l-2)(t+N-1)},
\end{aligned}$$

as $\prod_{p=k+1}^N \mathbf{1}_{(I_p \neq k)} \mathbf{1}_{(I_l = k)} = 0$, for $l > k$. Finally, for $l = k$, we get

$$\begin{aligned} & \mathbf{E} \left[\tilde{U}_k^N D_t \mathbf{E} \left[\tilde{U}_k^N \mid \mathcal{H}_{l-1} \right] \right] \\ &= \mathbf{E} \left[\mathbf{1}_{(I_k = k)} \prod_{p=k+1}^N \mathbf{1}_{(I_p \neq k)} \left(\mathbf{1}_{(I_k = k)} - \frac{t}{t+k-1} \right) \prod_{p=k+1}^N \mathbf{1}_{(I_p \neq k)} \right] \\ &= \left(\frac{t}{t+k-1} - \frac{t^2}{(t+k-1)^2} \right) \frac{t+k-1}{t+N-1} \\ &= \frac{t(k-1)}{(t+k-1)(t+N-1)}. \end{aligned}$$

It follows that

$$\begin{aligned} & \text{var} [\tilde{C}_1] \\ &= \frac{t^2}{t+N-1} \sum_{k=1}^N \sum_{l=k+1}^N \frac{1}{(t+l-1)(t+l-2)} + \frac{t}{t+N-1} \sum_{k=1}^N \frac{k-1}{t+k-1} \\ &= \frac{t}{t+N-1} \left(\frac{Nt}{t+N-1} + N - 2t^2 \sum_{k=1}^N \frac{1}{t+k-1} \right), \end{aligned}$$

so that we retrieve

$$\text{var} [\tilde{C}_1] \xrightarrow{N \rightarrow \infty} t,$$

as can be expected from the Poisson limit. \square

5. LYAPOUNOV CENTRAL LIMIT THEOREM

To the best of our knowledge, there is no quantitative version of the Lyapounov CLT [5] for independent but not identically distributed random variables. The main reason being that the usual coupling mentioned in the introduction is no longer effective when the random variables do not have the same distribution. We now show how to use our framework to get such a theorem.

The Wasserstein (or Kolmogorov-Rubinstein) distance between μ and ν , two probability measures on \mathbf{R} is defined by

$$\text{dist}(\mu, \nu) = \sup_{F \in \text{Lip}_1} \int_{\mathbf{R}} F d\mu - \int_{\mathbf{R}} F d\nu.$$

The Stein's method [4, Lemma 2.3] says that when μ is the standard Gaussian measure,

$$\text{dist}(\mu, \nu) = \sup_{F \in \mathcal{L}} \int_{\mathbf{R}} (xF(x) - F'(x)) d\nu(x),$$

where \mathcal{L} is the set of \mathcal{C}^2 functions such that $\|F''\|_{\infty} \leq 2$.

Theorem 5.1. *Let $(X_n, n \geq 1)$ be a sequence of thrice integrable, independent random variables. Denote*

$$\sigma_n^2 = \text{var}(X_n), s_n^2 = \sum_{j=1}^n \sigma_j^2 \text{ and } Y_n = \frac{1}{s_n} \sum_{j=1}^n (X_j - \mathbf{E}[X_j]).$$

Then,

$$\text{dist}(Y_n, \mathcal{N}(0, 1)) \leq \frac{6}{s_n^3} \sum_{j=1}^n \mathbf{E}[|X_j - \mathbf{E}[X_j]|^3].$$

Proof. With no loss of generality, we assume that $\mathbf{E}[X_n] = 0$ for any $n \geq 1$. For the sake of notation, we introduce

$$Y_n^{-k} = \frac{1}{s_n} \sum_{\substack{j=1 \\ j \neq k}}^n X_j = Y_n - \frac{X_k}{s_n}.$$

For any j , X_j is centered and independent from \mathcal{G}_j hence $D_j X_j = X_j$. Then,

$$\begin{aligned} \mathbf{E}[Y_n F(Y_n)] &= \frac{1}{s_n} \sum_{j=1}^n \mathbf{E}[D_j X_j F(Y_n)] \\ &= \frac{1}{s_n} \sum_{j=1}^n \mathbf{E}[X_j D_j F(Y_n)] \\ &= \frac{1}{s_n} \sum_{j=1}^n \mathbf{E}\left[X_j \left(F(Y_n^{-j} + \frac{X_j}{s_n}) - F(Y_n^{-j} + \frac{X'_j}{s_n})\right)\right], \end{aligned}$$

where X'_j is an independent copy of X_j and the expectation is computed on the product space. The Taylor expansion with integral remainder can be interpreted as

$$\begin{aligned} \mathbf{E}[Y_n F(Y_n)] &= \frac{1}{s_n^2} \sum_{j=1}^n \mathbf{E}[X_j (X_j - X'_j) F'(Y_n^{-j})] \\ &+ \frac{1}{2s_n^3} \sum_{j=1}^n \mathbf{E}\left[X_j (X_j - X'_j)^2 F''(Y_n^{-j} + \frac{X_j}{s_n} + \Theta(\frac{X_j}{s_n} - \frac{X'_j}{s_n}))\right] := A_1 + A_2, \end{aligned}$$

where Θ is a $[0, 1]$ -uniformly distributed random variable independent of the other variables. Since $\|F''\|_\infty \leq 2$, it is easily seen that

$$A_2 \leq \frac{1}{s_n^3} \sum_{j=1}^n \mathbf{E}[|X_j| (X_j - X'_j)^2] \leq \frac{4}{s_n^3} \sum_{j=1}^n \mathbf{E}[|X_j|^3].$$

Furthermore,

$$A_1 = \frac{1}{s_n^2} \sum_{j=1}^n \sigma_j^2 \mathbf{E}[F'(Y_n^{-j})].$$

Thus, since $F' \in \mathcal{L}$ is 2-Lipschitz continuous,

$$\begin{aligned} |A_1 - \mathbf{E} [F'(Y_n)]| &= \frac{1}{s_n^2} \left| \sum_{j=1}^n \sigma_j^2 \left(\mathbf{E} [F'(Y_n^{-j}) - F'(Y_n)] \right) \right| \\ &\leq \frac{2}{s_n^3} \sum_{j=1}^n \sigma_j^2 \mathbf{E} [|X_j|] \leq \frac{2}{s_n^3} \sum_{j=1}^n \mathbf{E} [|X_j|^3]. \end{aligned}$$

In brief,

$$|\mathbf{E} [Y_n F(Y_n) - F'(Y_n)]| \leq \frac{6}{s_n^3} \sum_{j=1}^n \mathbf{E} [|X_j|^3],$$

and the result follows. \square

REFERENCES

- [1] S. Albeverio, Y. G. Kondratiev, and M. Röckner, *Analysis and geometry on configuration spaces*, J. Funct. Anal. **154** (1998), no. 2, 444–500.
- [2] R. Arratia, A. D. Barbour, and S. Tavaré, *Poisson process approximations for the Ewens sampling formula*, Ann. Appl. Probab. **2** (1992), no. 3, 519–535.
- [3] ———, *Logarithmic combinatorial structures: a probabilistic approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2003.
- [4] A. D. Barbour and L. H. Y. Chen, *An introduction to Stein's method*, Lecture Notes Series, vol. 4, National University of Singapore, 2005.
- [5] P. Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication.
- [6] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, DOI: 10.1093/acprof:oso/9780199535255.001.0001.
- [7] N. Bouleau and F. Hirsch, *Dirichlet forms and analysis on Wiener space*, De Gruyter Studies in Mathematics, 1992.
- [8] N. Bouleau, *Théorème de Donsker et formes de Dirichlet*, Bulletin des sciences mathématiques **129** (2005), 63–73.
- [9] L. H. Y. Chen, *On the convergence of Poisson binomial to Poisson distributions*, Ann. Probability **2** (1974), no. 1, 178–180.
- [10] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao, *Normal approximation by Stein's method*, Probability and its Applications (New York), Springer, Heidelberg, 2011.
- [11] L. Coutin and L. Decreasefond, *Stein's method for Brownian approximations*, Communications on stochastic analysis **7** (2013), no. 5, 349–372.
- [12] ———, *Higher order expansions via Stein's method*, Communications on Stochastic Analysis **8** (2014), no. 2, 155–168.
- [13] L. Decreasefond, *The Stein-Dirichlet-Malliavin method*, ESAIM: Proceedings (2015), 11.
- [14] L. Decreasefond and I. Flint, *Moment formulae for general point processes*, Journal of Functional Analysis **267** (2014), no. 2, 452 – 476.
- [15] L. Decreasefond, M. Schulte, and C. Thäle, *Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry*, Annals of probability **44** (2015), no. 3, 2147–2197.
- [16] L. Goldstein and G. Reinert, *Stein's method and the zero bias transformation with application to simple random sampling*, Ann. Appl. Probab. **7** (1997), no. 4, 935–952.
- [17] O. Kallenberg, *Random measures*, 3rd ed., Academic Press, 1983.
- [18] S. Kerov, G. Olshanski, and A. Vershik, *Harmonic analysis on the infinite symmetric group*, Inventiones mathematicae **158** (2004), no. 3, 551–642.

- [19] V. S. Koroljuk and Yu. V. Borovskich, *Theory of U-statistics*, Mathematics and its Applications, vol. 273, Kluwer Academic Publishers Group, Dordrecht, 1994, Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.
- [20] I. Nourdin and G. Peccati, *Normal approximations with Malliavin calculus: From Stein's method to universality*, Cambridge University Press, 2012.
- [21] I. Nourdin, G. Peccati, and G. Reinert, *Stein's method and stochastic analysis of Rademacher functionals*, Electronic Journal of Probability **15** (2010), no. 55, 1703–1742.
- [22] D. Nualart, *The Malliavin calculus and related topics*, Springer-Verlag, 1995.
- [23] D. Nualart and J. Vives, *Anticipative calculus for the Poisson process based on the Fock space*, Séminaire de probabilités XXIV, Springer-Verlag, 1988, pp. 154–165.
- [24] D. Nualart and G. Peccati, *Central limit theorems for sequences of multiple stochastic integrals*, Ann. Probab. **33** (2005), no. 1, 177–193.
- [25] D. Nualart and W. Schoutens, *Chaotic and predictable representations for Lévy processes*, Stochastic Processes and their Applications **90** (2000), no. 1, 109 – 122.
- [26] G. Peccati, J.L. Solé, M.S. Taqqu, and F. Utzet, *Stein's method and normal approximation of Poisson functionals*, Annals of Probability **38** (2010), no. 2, 443–478.
- [27] N. Privault, *Stochastic analysis in discrete and continuous settings with normal martingales*, Lecture Notes in Mathematics, vol. 1982, Springer-Verlag, Berlin, 2009.
- [28] W. T. Rhee and M. Talagrand, *Martingale inequalities and the jackknife estimate of variance*, Statist. Probab. Lett. **4** (1986), no. 1, 5–6.
- [29] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory (Berkeley, Calif.), University of California Press, 1972, pp. 583–602.
- [30] A. S. Üstünel and M. Zakai, *The construction of filtrations on abstract Wiener space*, J. Funct. Anal. **143** (1997), no. 1, 10–32.
- [31] A.S. Üstünel and M. Zakai, *Transformations of measure on Wiener spaces*, Monographs in Mathematics, Springer-Verlag, 2000.
- [32] L. Wu, *A new modified logarithmic Sobolev inequality for Poisson point processes and several applications*, Probability Theory and Related Fields **118** (2000), no. 3, 427–438.
- [33] K. Yosida, *Functional analysis*, Classics in Mathematics, Springer-Verlag, Berlin, 1995.

LTCI, TELECOM PARISTECH, UNIVERSITÉ PARIS-SACLAY, 75013, PARIS, FRANCE

E-mail address: Laurent.Decreusefond@mines-telecom.fr

LTCI, TELECOM PARISTECH, UNIVERSITÉ PARIS-SACLAY, 75013, PARIS, FRANCE

LABORATOIRE ISTI, ESME SUDRIA, PARIS, 75015, FRANCE

E-mail address: helene.halconruy@mines-telecom.fr