Accepted Manuscript

Malliavin and Dirichlet structures for independent random variables

Laurent Decreusefond, Hélène Halconruy

PII: S0304-4149(18)30389-2
DOI: https://doi.org/10.1016/j.spa.2018.07.019
Reference: SPA 3360

To appear in: *Stochastic Processes and their Applications*

Received date : 20 July 2017
Revised date : 26 July 2018
Accepted date : 26 July 2018

Please cite this article as: L. Decreusefond, H. Halconruy, Malliavin and Dirichlet structures for independent random variables, *Stochastic Processes and their Applications* (2018), https://doi.org/10.1016/j.spa.2018.07.019

Malliavin and Dirichlet structures for independent random variables

Laurent Decreusefond[a], Hélène Halconruy[a,b]

[a] LTCI, Telecom ParisTech, Université Paris-Saclay, 75013, Paris, France
[b] ESME Sudria, Paris, 75015, France

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 the Hoeffding decomposition of U-statistics can be interpreted as an avatar of the Clark representation formula. Thanks to our framework, we obtain a bound for the distance between the distribution of any functional of independent variables and the Gaussian and Gamma distributions.

Keywords: Dirichlet structure, Ewens distribution, log-Sobolev inequality, Malliavin calculus, Stein's method, Talagrand inequality
2000 MSC: 60H07

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 [30, 41]) and Poisson point processes (see for instance [1, 36]). 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 [31], Lévy processes [33] 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.

Email addresses: `Laurent.Decreusefond@mines-telecom.fr` (Laurent Decreusefond), `helene.halconruy@mines-telecom.fr` (Hélène Halconruy)

The second motivation comes from Stein's method[1]. The Stein method which was initially developed to quantify the rate of convergence in the Central Limit Theorem [39] 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 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

$$\sup_{F\in\mathcal{F}}\Big(\mathbf{E}\left[L_1F(X)\right]+\mathbf{E}\left[L_2F(X)\right]\Big),$$

where $\mathcal{F}$ is the class of functions solutions of the Stein equation, which is related to the set of test functions $\mathcal{H}$ induced by 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 $\mathbf{R}$,

$$L_1F(x)=xF'(x) \text{ and } L_2F(x)=-F''(x).$$

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

$$L_1F(n)=n\left(F(n)-F(n-1)\right) \text{ and } L_2F(n)=\lambda(F(n+1)-F(n)).$$

In the next step, we have to take into account how X is defined and transform L_1F such that it can be written as $-L_2F+$remainder. This remainder is what gives the rate of convergence. To make the transformation of L_1F, several approaches appeared along the years. One of the most popular approach (see for instance [5]) is to use exchangeable pairs: Construct a copy X' of X with good properties which gives another expression of L_1F, suitable to a comparison with L_2F. 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 $I\in\{1,\cdots,n\}$ 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

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

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 [19] couplings, which again conveniently transform L_1F. For Gaussian approximation, it amounts to find a distribution X^* such that

$$\mathbf{E}\left[L_1F(X)\right]=\mathbf{E}\left[F''(X^*)\right].$$

[1] Giving 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.

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

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

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 reminiscent to what can be obtained via integration by parts. Actually, Nourdin and Peccati (see [26]) 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 [27], Poisson [15, 34] or Gaussian random variables [32] 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.

The closest situation to our investigations is that of the Rademacher space, namely $\{-1,1\}^{\mathbf{N}}$, equipped with the product probability $\otimes_{k\in\mathbf{N}}\mu_k$ where μ_k is a Bernoulli probability on $\{-1,1\}$.

The gradient on the Rademacher space (see [27, 36]) is usually defined as

$$\begin{aligned}\hat{D}_k F(X_1,\cdots,X_n) &= \mathbf{E}\left[X_k\,F(X_1,\cdots,X_n)\,|\,X_l,\,l\neq k\right]\\ &= \mathbf{P}(X_k=1)\,F(X_1,\cdots,+1,\cdots,X_n)\\ &\qquad - \mathbf{P}(X_k=-1)\,F(X_1,\cdots,-1,\cdots,X_n), \qquad (2)\end{aligned}$$

where the ± 1 are put in the k-th coordinate. It requires, for its very definition to be meaningful, either that the random variables are real valued or that they only have two possible outcomes. 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. Some of our applications does concern random variables with finite number of outcomes but it does not seem straightforward to devise what should be the weights, replacing $\mathbf{P}(X_k = 1)$ and $-\mathbf{P}(X_k = -1)$. Furthermore, many applications, notably those revolving around functional identities, rely not directly on the gradient D but rather on the operator number $L = -\delta D$ where δ is the adjoint, in a sense to be defined later. It turns out that for the Rademacher space, the operators $\hat{L} = -\hat{\delta}\hat{D}$ defined according to (2) and L defined in Definition 2.2 do coincide. Our framework then fully generalizes what is known about Rademacher 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 5) 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 [40]. During the preparation of this work, we found strong reminiscences of our gradient with the map Δ, introduced in [6, 38] for the proof of the Efron-Stein inequality, defined by

$$\Delta_k F(X_1, \cdots, X_n) = \mathbf{E}\left[F \mid X_1, \cdots, X_k\right] - \mathbf{E}\left[F \mid X_1, \cdots, X_{k-1}\right].$$

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 processes [43], 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 4 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 [23], which is similar in spirit to the so-called Feller coupling (see [3]), is an interesting situation to apply our results since this construction involves a cartesian product of distinct finite spaces. In Section 5, we present several applications of our results. In subsection 5.1, we derive the chaos decomposition of the number of fixed points of a random permutations under the Ewens distribution. This yields an exact expression for the variance of this random variable. 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. In subection 5.2, we give an analog to Theorem 3.1 of [25, 34], which is a general bound of the Kolmogorov Rubinstein distance to a Gaussian or Gamma distribution, in terms of our gradient D. We apply this to a degenerate U-statistics of order 2.

2. Malliavin calculus for independent random variables

Let A be an at most denumerable set equipped with the counting measure:

$$L^2(A) = \left\{ u : A \to \mathbf{R}, \sum_{a\in A} |u_a|^2 < \infty \right\} \text{ and } \langle u, v\rangle_{L^2(A)} = \sum_{a\in A} u_a v_a.$$

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 $\mathcal{E}_A = \underset{a\in A}{\vee} \mathcal{E}_a$ and the tensor product measure $\mathbf{P} = \underset{a\in A}{\otimes} \mathbf{P}_a$.
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 $\mathbf{P}_B = \underset{a\in B}{\otimes} \mathbf{P}_a$.
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):

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

Our presentation follows closely the usual construction of Malliavin calculus.

Definition 2.1. *A random variable F is said to be cylindrical if there exist a finite subset $B \subset A$ and a function $F_B : E_B \longrightarrow L^2(E_A)$ 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 $L^2(E_A)$.

The very first tool to be considered is the discrete gradient, whose form has been motivated in the introduction.

We first define the gradient of cylindrical functionals, for there is no question of integrability and then extend the domain of the gradient to a larger set of functionals by a limiting procedure. In functional analysis terminology, we need to verify the closability of the gradient: If a sequence of functionals converges to 0 and the sequence of their gradients converges, then it should also converges to 0. This is the only way to guarantee in the limiting procedure that the limit does not depend on the chosen sequence.

Definition 2.2 (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 $a \in A$,*

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

where X'_a is an independent copy of X_a.

Remark 1. *A straightforward calculation shows that for any $F, G \in \mathcal{S}$, any $a \in A$, we have*

$$D_a(FG) = F\, D_a G + G\, D_a F - D_a F\, D_a G - \mathbf{E}\left[FG \,|\, \mathcal{G}_a\right] + \mathbf{E}\left[F \,|\, \mathcal{G}_a\right] \mathbf{E}\left[G \,|\, \mathcal{G}_a\right].$$

This formula has to be compared with the formula $D(FG) = F\, DG + G\, DF$ for the Gaussian Malliavin gradient (see (16) *below) and $D(FG) = F\, DG + G\, DF + DF\, DG$ for the Poisson gradient (see* (11) *below).*

For $F \in \mathcal{S}$, 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

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

hence $(D_a F, a \in A)$ defines an element of $L^2(A \times E_A)$.

Definition 2.3. *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*

$$U = \sum_{a\in B} U_a\, \mathbf{1}_a,$$

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

The key formula for the sequel is the so-called integration by parts. It amounts to compute the adjoint of D in $L^2(A \times E_A)$.

Theorem 2.4 (Integration by parts)**.** *Let $F \in \mathcal{S}$. For every simple process U,*

$$\langle DF, U\rangle_{L^2(A\times E_A)} = \mathbf{E}\left[F \sum_{a\in A} D_a U_a\right]. \tag{3}$$

Thanks to the latter formula, we are now in position to prove the closability of D: For $(F_n,\, n\geq 1)$ a sequence of cylindrical functionals,

$$\left(F_n \xrightarrow[L^2(E_A)]{n\to\infty} 0 \text{ and } DF_n \xrightarrow[L^2(A\times E_A)]{n\to\infty} \eta\right) \Longrightarrow \eta = 0.$$

Corollary 2.5. *The operator D is closable from $L^2(E_A)$ into $L^2(A\times E_A)$.*

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

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

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-Ulhenbeck semi-group (see below Section 2.2) lessens the probable usage of other integrability order.

Since $\mathbf{D}$ is defined as a closure, it is often useful to have a general criterion to ensure that a functional F, which is not cylindrical, belongs to $\mathbf{D}$. The following criterion exists as is in the settings of Wiener and Poisson spaces.

Lemma 2.6. *If there exists a sequence $(F_n,\, n\geq 1)$ of elements of $\mathbf{D}$ such that*

1. *F_n converges to F in $L^2(E_A)$,*
2. *$\sup_n \|DF_n\|_{\mathbf{D}}$ is finite,*

then F belongs to $\mathbf{D}$ and $DF = \lim_{n\to\infty} DF_n$ in $\mathbf{D}$.

2.1. Divergence

We can now introduce the adjoint of D, often called the divergence as for the Lebesgue measure on $\mathbf{R}^n$, the usual divergence is the adjoint of the usual gradient.

Definition 2.7 (Divergence)**.** *Let*

$$\operatorname{Dom}\delta = \Big\{U\in L^2(A\times E_A) : \\ \exists\, c>0,\, \forall\, F\in\mathbf{D},\ |\langle DF, U\rangle_{L^2(A\times E_A)}| \leq c\,\|F\|_{L^2(E_A)}\Big\}.$$

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

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

The integration by parts formula (3) entails that for every $U\in\operatorname{Dom}\delta$,

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

In the setting of Malliavin calculus for Brownian motion, the divergence of adapted processes coincide with the Itô integral and the square moment of δU is then given by the Itô isometry formula. We now see how this extends to our situation.

Definition 2.8. *The Hilbert space $\mathbf{D}(l^2(A))$ is the closure of $\mathcal{S}_0(l^2(A))$ with respect to the norm*

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

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 [44]) with

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

The next formula is the counterpart of the Itô isometry formula for the Brownian motion, sometimes called the Weitzenböck formula (see [36, Eqn. (4.3.3)]) in the Poisson settings.

Theorem 2.9. *The space $\mathbf{D}(l^2(A))$ is included in $\operatorname{Dom}\delta$. For any U, V belonging to $\mathbf{D}(l^2(A))$,*

$$\mathbf{E}\left[\delta U\ \delta V\right] = \mathbf{E}\left[\operatorname{trace}(DU \circ DV)\right]. \tag{4}$$

Remark 2. *It must be noted that compared to the analog identity for the Brownian and the Poisson settings, the present formula is slightly different. For both processes, with corresponding notations, we have*

$$\|\delta U\|^2_{L^2} = \|U\|^2_{L^2} + \operatorname{trace}(DU \circ DV).$$

The absence of the term $\|U\|^2_{L^2}$ gives to our formula a much stronger resemblance to the analog equation for the Lebesgue measure. As in this latter case, we do have here $\delta\mathbf{1} = 0$ whereas for the Brownian motion, it yields the Itô integral of the constant function equal to one.

If $A = \mathbf{N}$, let $\mathcal{F}_n = \sigma\{X_k,\ k \le n\}$ and assume that U is adapted, i.e. for all $n \ge 1$, $U_n \in \mathcal{F}_n$. Then, $D_n U_k = 0$ as soon as $n > k$, hence

$$\mathbf{E}\left[\delta U^2\right] = \mathbf{E}\left[\sum_{n=1}^{\infty}\Big(U_n - \mathbf{E}\left[U_n \,|\, \mathcal{F}_{n-1}\right]\Big)^2\right],$$

i.e. $\mathbf{E}\left[\delta U^2\right]$ is the $L^2(\mathbf{N} \times E_{\mathbf{N}})$-norm of the innovation process associated to U, which appears in filtering theory.

2.2. 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 2.10. *The number operator, denoted by L, is defined on its domain*

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

by

$$LF = -\delta DF = -\sum_{a \in A} D_a F. \tag{5}$$

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 $X(t) = (X_1(t), \cdots, X_N(t),\, t \geq 0)$ 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 $x_{\neg a} = (x_1, \cdots, x_{a-1}, x_{a+1}, \cdots, x_{|A|})$. The generator of the Markov process X is clearly given by

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

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. If we denote by $P = (P_t, t \geq 0)$ the semi-group of X: For any $x \in E_A$, for any bounded $f : E_A \to \mathbf{R}$,

$$P_t f(x) = \mathbf{E}\left[f(X(t)) \,|\, X(0) = x \right].$$

Then, $(P_t, t \geq 0)$ is a strong Feller semi-group on $L^\infty(E_A)$. This result still holds when E_A is denumerable.

Theorem 2.11. *For any denumerable set A, L defined as in* (5) *generates a strong Feller continuous semi-group $(P_t, t \geq 0)$ on $L^\infty(E_A)$.*

As a consequence, there exists a Markov process X whose generator is L as defined in (5)*. It admits as a core (a dense subset of its domain) the set of cylindrical functions.*

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

Theorem 2.12 (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 $t > 0$,

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

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

$$\lim_{t \to \infty} P_t F(x) = \int_{E_A} F \,\mathrm{d}\mathbf{P} \text{ and } X(0) \overset{law}{=} \mathbf{P} \Longrightarrow X(t) \overset{law}{=} \mathbf{P}.$$

We then retrieve the classical formula (in the sense that it holds as is for Brownian motion and Poisson process) of commutation between D and the Ornstein-Uhlenbeck semi-group.

Theorem 2.13. *Let $F \in L^2(E_A)$. For every $a \in A$, $x \in E_A$,*

$$D_a P_t F(x) = e^{-t} P_t D_a F(x). \tag{6}$$

3. Functional identities

This section is devoted to several functional identities which constitute the crux of the matter if we want to do some computations with our new tools.

It is classical that the notion of adaptability is linked to the support of the gradient.

Lemma 3.1. *Assume that $A = \mathbf{N}$ and let $\mathcal{F}_n = \sigma\{X_k,\, k \leq n\}$. 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 $F = \mathbf{E}\,[F]$.*

It is also well known that, in the Brownian of Poisson settings, D and conditional expectation commute.

Lemma 3.2. *For any $F \in \mathbf{D}$, for any $k \geq 1$, we have*

$$D_k \, \mathbf{E}\left[F|\mathcal{F}_k\right] = \mathbf{E}\left[D_k F \,|\, \mathcal{F}_k\right]. \tag{7}$$

The Brownian martingale representation theorem says that a martingale adapted to the filtration of a Brownian motion is in fact a stochastic integral. The Clark formula gives the expression of the integrand of this stochastic integral in terms of the Malliavin gradient of the terminal value of the martingale. We here have the analogous formula.

Theorem 3.3 (Clark formula)**.** *For $A = \mathbf{N}$ and $F \in \mathbf{D}$,*

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

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

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

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

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

since $j > k$ implies $D_j \mathbf{E}\left[F \,|\, \mathcal{F}_k\right] = 0$ in view of Lemma 3.1. Furthermore, 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.

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 $n \in \{0, \cdots, N-1\}$,

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

Note that $\mathcal{H}_0^N = \mathcal{F}_N$ and as in Lemma 3.1, F is $\mathcal{H}_k$-measurable if and only if $D_n F = 0$ for any $n \leq k$.

Theorem 3.4. *For every F in $\mathbf{D}$,*

$$F = \mathbf{E}\left[F\right] + \sum_{k=1}^{\infty} D_k \, \mathbf{E}\left[F \,|\, \mathcal{H}_{k-1}\right].$$

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]. It can be noted that both the statement and the proof are similar in the Brownian and Poisson settings.

Corollary 3.5 (Poincaré or Efron-Stein inequality)**.** *For any* $F \in \mathbf{D}$*,*

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

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

Theorem 3.6 (Covariance identity)**.** *For any* $F, G \in \mathbf{D}$*,*

$$\operatorname{cov}(F,G) = \mathbf{E}\left[\sum_{k\in A} D_k \mathbf{E}\left[F \,|\, \mathcal{F}_k\right]\ D_k G\right]. \tag{8}$$

As for the other versions of the Malliavin calculus (Brownian, Poisson and Rademacher), from (6), can be deduced another covariance identity.

Theorem 3.7. *For any* $F, G \in \mathbf{D}$*,*

$$\operatorname{cov}(F,G) = \mathbf{E}\left[\sum_{k\in A} D_k F \int_0^\infty e^{-t} P_t \mathbf{E}\left[D_k G | \mathcal{F}_k\right] \mathrm{d}t\right]. \tag{9}$$

Then, using the so-called Herbst principle, we can derive a concentration inequality, which, as usual, requires an L^∞ bound on the derivative of the functional to be valid.

Theorem 3.8 (Concentration inequality)**.** *Let F for which there exists an order on A with*

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

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

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

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 because D is not a derivation, i.e. does not satisfy $D(FG) = F\,DG + G\,DF$. However, we still have an LSI identity. For the proof of it, we follow closely the proofs of [35, 43]. They are based on two ingredients: The Itô formula and the martingale representation theorem. We get an ersatz of the former but the latter seems inaccessible as we do not impose the random variables to live in the same probability space and to be real valued. Should it be the case, to the best of our knowledge, the martingale representation formula is known only for the Rademacher space [42, Section 15.1], which is exactly the framework of [35]. This lack of a predictable representation explains the conditioning in the denominator of (10).

Theorem 3.9 (Logarithmic Sobolev inequality)**.** *Let a positive random variable $G \in L\log L(E_A)$. Then,*

$$\mathbf{E}\left[G\log G\right] - \mathbf{E}\left[G\right]\log\mathbf{E}\left[G\right] \le \sum_{k\in A}\mathbf{E}\left[\frac{|D_kG|^2}{\mathbf{E}\left[G\,|\,\mathcal{G}_k\right]}\right]. \tag{10}$$

In the usual vector calculus on $\mathbf{R}^3$, the Helhmoltz decomposition stands that a sufficiently smooth vector field can be resolved in the sum of a curl-free vector field and a divergence-free vector field. We have here the exact counterpart with our definition of gradient.

Theorem 3.10 (Helhmoltz 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 $\mathbf{E}\left[\varphi\right] = 0$, $\delta V = 0$ and*

$$U_a = D_a\varphi + V_a,$$

for any $a \in A$.

4. Dirichlet structures

We now show that the usual Poisson and Brownian Dirichlet structures, associated to their respective gradient, can be retrieved as limiting structures of convenient approximations. This part is directly inspired by [8] where with our notations, the X_a's are supposed to be real valued, independent and identically distributed and the gradient be the ordinary gradient on $\mathbf{R}^A$.

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 2.2). It remains to define the Dirichlet form $\mathcal{E}_A$ such that $\mathcal{E}_A(F) = \mathbf{E}\left[F\,LF\right]$ for any sufficiently regular functional F.

Definition 4.1. *For $F \in \mathbf{D}$, define*

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

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

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

4.1. 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 [22]) 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, 31, 36]): For any $x \in \mathbb{Y}$, any $\omega \in \mathfrak{N}_{\mathbb{Y}}$,

$$D_x F(\omega) = F(\omega + \varepsilon_x) - F(\omega). \tag{11}$$

Set

$$\mathbf{D}_P = \left\{ F \,:\, \mathfrak{N}_{\mathbb{Y}} \to \mathbf{R} \text{ such that } \mathbf{E}\left[\int_{\mathbb{Y}} |D_x F|^2 \, \mathrm{d}\mathbf{M}(x) \right] < \infty \right\},$$

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

$$\mathcal{E}(F) = \mathbf{E}\left[\int_{\mathbb{Y}} |D_x F|^2 \, \mathrm{d}\mathbf{M}(x)\right]. \tag{12}$$

To see the Poisson point process as a Dirichlet limit, the idea is to partition the set $\mathbb{Y}$ into N parts, $C_1^N, \cdots, C_N^N$ such that $\mathbf{M}(C_k^N) = p_k^N$ and then for each $k \in \{1, \cdots, N\}$, take a point ζ_k^N into C_k^N so that the Poisson point process ω on $\mathbb{Y}$ 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 $(\zeta_k^N,\, k = 1, \cdots, N)$ (respectively $(p_k^N,\, k = 1, \cdots, N)$) be a triangular array of points in $\mathbb{Y}$ (respectively of non-negative numbers) such that the following two properties hold:
1) the p_k^N's tends to 0 uniformly:

$$p^N = \sup_{k \le N} p_k^N = O\left(\frac{1}{N}\right); \tag{13}$$

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 $f : \mathbb{Y} \to \mathbf{R}$, we have

$$\sum_{k=1}^{N} f(\zeta_k^N)\, p_k^N \xrightarrow{N \to \infty} \int f(x) \, \mathrm{d}\mathbf{M}(x). \tag{14}$$

Take $f = 1$ implies that $\sum_k p_k^N$ tends to 1 as N goes to infinity.
For any N and any $k \in \{1, \cdots, N\}$, 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 $M_1^N, \cdots, M_N^N$, where M_k^N follows a Poisson distribution of parameter p_k^N for any $k \in \{1, \cdots, N\}$. We turn these independent random variables into a point process by the map U_N defined as

$$\begin{aligned} U_N \,:\, \mathbf{N}^N &\longrightarrow \mathfrak{N}_{\mathbb{Y}} \\ (m_1, \cdots, m_N) &\longmapsto \sum_{k=1}^{N} m_k \ \varepsilon_{\zeta_k^N}. \end{aligned}$$

Lemma 4.3. *For any $F \in \mathbf{D}_P$,*

$$\begin{aligned}&\mathcal{E}^{U_N}(F)\\&= \sum_{m=1}^{N}\sum_{\ell=0}^{\infty}\mathbf{E}\left[\left(\sum_{\tau=0}^{\infty}\Big(F(\omega^N_{(m)}+\ell\varepsilon_{\zeta^N_m})-F(\omega^N_{(m)}+\tau\varepsilon_{\zeta^N_m})\Big)\mu^N_m(\tau)\right)^2\right]\mu^N_m(\ell),\end{aligned} \tag{15}$$

where $\omega^N_{(m)}=\sum_{k\neq m} M^N_k\varepsilon_{\zeta^N_k}$.

Proof. According its very definition,

$$\mathcal{E}^{U_N}(F)=\sum_{m=1}^{N}\mathbf{E}\left[\left(F(\omega^N_{(m)}+M^N_m\varepsilon_{\zeta^N_m})-\sum_{\tau=0}^{\infty}F(\omega^N_{(m)}+\tau\varepsilon_{\zeta^N_m})\mu^N_m(\tau)\right)^2\right].$$

The result follows by conditioning with respect to M^N_m, whose law is μ^N_m. □

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

$$|F(\omega)-F(\eta)| \le \mathrm{dist}_{TV}(\omega,\,\eta),$$

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

Theorem 4.5. *For any $F \in \mathrm{TV}-\mathrm{Lip}\cap\mathbf{D}_P$, with the notations of Lemma [4.3] and* (12),

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

4.2. Brownian motion

For details on Gaussian Malliavin calculus, we refer to [30, 41]. We now consider $\mathbf{P}$ as the Wiener measure on $W=\mathcal{C}_0([0,1];\mathbf{R})$. Let $(h_k,\, k\ge 1)$ be an orthonormal basis of the Cameron-Martin space H,

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

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

$$F(\omega)=f(\delta_B v_1,\cdots,\delta_B v_n),$$

where $v_1, \cdots, v_n$ belong to H,

$$\delta_B v = \int_0^1 v(s)\, \mathrm{d}\omega(s)$$

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

$$\nabla_h F(\omega) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\delta_B v_1, \cdots, \delta_B v_n)\, h_k. \tag{16}$$

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

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

Definition 4.6. *A function $F : W \to \mathbf{R}$ 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

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

For $N \geq 1$, let

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

The family $(h_k^N,\, k = 1, \cdots, N)$ is then orthonormal in H. For $(M_k,\, k = 1, \cdots, N)$ 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:

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

It follows from our definition that:

Lemma 4.7. *For any $F \in L^2(W;\mathbf{R})$,*

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

where $\omega^N_{(k)} = \omega^N - M_k\, h^N_k$ 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 4.8. *A function $F : W \to \mathbf{R}$ is said to be Lipschitz if it is* H-C^1 *and for almost all $\omega \in W$,*

$$|\langle \nabla F(\omega),\, h\rangle| \le \|\dot{h}\|_{L^1}.$$

In particular since $e^N_k \ge 0$, this implies that

$$|\langle \nabla F(\omega),\, h^N_k\rangle| \le h^N_k(1) - h^N_k(0) = \frac{1}{\sqrt{N}}.$$

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

$$F(\omega + h) - F(\omega) = \langle \nabla F(\omega),\, h\rangle_H + \|\dot{h}\|_{L^1}\, \varepsilon(\omega, h), \tag{17}$$

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

Theorem 4.9. *For any $F \in \mathbf{D}_B \cap$ H-C^1,*

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

5. Applications

5.1. Representations

We now show that our Clark decomposition yields interesting decomposition of random variables. For U-statistics, it boils down to the Hoeffding decomposition.

Definition 5.1. *For an integer m, let $h : \mathbb{R}^m \to \mathbb{R}$ be a symmetric function, and $X_1, \cdots, X_n$, n random variables supposed to be independent and identically distributed. The U-statistics of degree m and kernel h is defined, for any $n \ge m$ by*

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

where $([n], m)$ denotes the set of ordered subsets $A \subset [n] = \{1, \cdots, n\}$, of cardinality m.

More generally, for a set B, (B, m) denotes the set of subsets of B with m elements.

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

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

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

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

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

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

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

Theorem 5.2 (Hoeffding decomposition of U-statistics, [24]). *For any integer n, we have*

$$U_n = \theta + \sum_{k=1}^{m} H_n^{(k)} \tag{18}$$

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

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

As mentioned above, reversing the natural order of A, provided that it exists, can be very fruitful. We illustrate this idea by the decomposition of the number of fixed points of a random permutation under Ewens distribution. It could be applied to more complex functionals of permutations but to the price of increasingly complex computations.

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

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

The coordinate map from $\mathcal{J}$ to $\mathcal{J}_k$ is denoted by I_k. Following [23], we have

Theorem 5.3. *There exists a natural bijection Γ between $\mathcal{J}$ and $\mathfrak{S}_N$.*

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

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

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 [23], 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 [3]) where i_N is placed immediately after N. An alternative construction of permutations is known as the Feller coupling (see [3]). In our notations, it is given by

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

Definition 5.4 (Ewens distribution)**.** *For some $t \in \mathbf{R}^+$, for any $k \in \{1, \cdots, N\}$, consider the measure $\mathbf{P}_k$ defined on $\mathcal{J}_k$ by*

$$\mathbf{P}_k(\{j\}) = \begin{cases} \dfrac{1}{t+k-1} & \text{if } j \neq k, \\[2ex] \dfrac{t}{t+k-1} & \text{for } j = k. \end{cases}$$

Under the distribution $\mathbf{P} = \otimes_k \mathbf{P}_k$, the random variables $(I_k,\, k = 1, \cdots, N)$ 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 push-forward 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 5.5 (see [23])**.** *For any $\sigma \in \mathfrak{S}_N$,*

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

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

$$\begin{tikzcd} (\mathcal{J},\, \otimes_{k=1}^N \mathbf{P}_k) \arrow[d, "\Gamma"'] \arrow[dr, "\tilde{F} = F \circ \Gamma"] & \\ (\mathfrak{S}_N,\, \mathbf{P}^t) \arrow[r, "F"'] & \mathbf{R} \end{tikzcd}$$

We denote by $i = (i_1, \cdots, i_N)$ 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 5.2).

Theorem 5.6. *For any $k \in \{1, \cdots, N\}$,*

$$\tilde{U}_k = \mathbf{1}_{(I_k=k)} \mathbf{1}_{(I_m \neq k,\ m \in \{k+1, \cdots, N\})}. \tag{19}$$

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

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

Moreover,

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

Since

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

we retrieve the result of [4]:

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

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

Theorem 5.7. *We can write*

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

Remark 3. *Note that such a decomposition with the natural order on $\mathbf{N}$ would be infeasible since the basic blocks of the definition of $\tilde{C}_1$, namely the $\tilde{U}_k$, are anticipative (following the vocabulary of Gaussian Malliavin calculus), i.e. $\tilde{U}_k \in \sigma(I_{k+l}, l = 0, \cdots, N-k)$.*

This decomposition can be used to compute the variance of $\tilde{C}_1$. To the best of our knowledge, this is the first explicit, i.e. not asymptotic, expression of it.

Theorem 5.8. *For any $t \in \mathbf{R}$, we get*

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

We retrieve

$$\operatorname{var}[\tilde{C}_1] \xrightarrow[N\to\infty]{} t,$$

as can be expected from the Poisson limit.

5.2. Stein-Malliavin criterion

For (E,d) a Polish space, let $\mathfrak{M}_1(E)$ the set of probability measures on E. It is usually equipped with the weak convergence generated by the semi-norms

$$p_f(\mathbf{P}) = \left|\int_E f\,\mathrm{d}\mathbf{P}\right|$$

for any f bounded and continuous from E to $\mathbf{R}$. Since E is Polish, we can find a denumerable family of bounded continuous functions $(f_n,\, n \ge 1)$ which generates the Borelean σ-field on E and the topology of the weak convergence can be made metric by considering the distance:

$$\rho(\mathbf{P},\mathbf{Q}) = \sum_{n=1}^{\infty} 2^{-n}\,\psi(p_{f_n}(\mathbf{P}-\mathbf{Q}))$$

where $\psi(x) = x/(1+x)$. Unfortunately, this definition is not prone to calculations so that it is preferable to use the Kolmogorov-Rubinstein (or Wasserstein-1) distance defined by

$$\kappa(\mathbf{P},\mathbf{Q}) = \sup_{\varphi\in\operatorname{Lip}_1}\left|\int_E \varphi\,\mathrm{d}\mathbf{P} - \int_E \varphi\,\mathrm{d}\mathbf{Q}\right|$$

where

$$\varphi \in \operatorname{Lip}_r \iff \sup_{x\neq y\in E}\frac{|\varphi(x)-\varphi(y)|}{d(x,y)} \le r.$$

Theorem 11.3.1 of [16] states that the distances κ and ρ yield the same topology. When $E=\mathbf{R}$, the Stein's method is one efficient way to compute the κ distance between a measure and the Gaussian distribution. If $E=\mathbf{R}^n$, for technical reasons, it is often assumed that the test functions are more regular than simply Lipschitz continuous and we are led to compute

$$\kappa_{\mathcal{H}}(\mathbf{P},\mathbf{Q}) = \sup_{\varphi\in\mathcal{H}}\left|\int_E \varphi\,\mathrm{d}\mathbf{P} - \int_E \varphi\,\mathrm{d}\mathbf{Q}\right|$$

where $\mathcal{H}$ is a space included in Lip_1 like the set of k-times differentiable functions with derivatives up to order k bounded by 1.

The setting in which we need to compute a KR distance is very often the situation in which we have another Polish space G with a probability measure μ and a random variable F with value in E. The objective is then to compare some measure $\mathbf{P}$ on E and $\mathbf{P}_F = F_*\mu$ the distribution of F, i.e. the push-forward of μ by the application F. This means that we have to compute

$$\sup_{\varphi\in\mathcal{H}} \left| \int_E \varphi \,\mathrm{d}\mathbf{P} - \int_G \varphi \circ F \,\mathrm{d}\mu \right|. \tag{20}$$

As mentioned in Section 1, when using the Stein's method, we first characterize $\mathbf{P}$ by a functional identity and then use different tricks to transform (20) in a more tractable expression. The usual tools are exchangeable pairs, coupling or Malliavin integration by parts. For the latter to be possible requires that we do have a Malliavin structure on the measured space (G, μ). In [25, 34], generic theorems are given which link $\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_F)$ with some functionals of the gradient of F. For instance, if (G, μ) is the space of locally finite configurations on a space $\mathfrak{g}$, equipped with the Poisson distribution of control measure σ and $\mathbf{P}$ is the Gaussian distribution in $\mathbf{R}$,

$$\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_F) \le \mathbf{E}\left[\left| 1 - \int_{\mathfrak{g}} D_z F \, D_z L^{-1} F \,\mathrm{d}\sigma(z) \right| \right] + \int_{\mathfrak{g}} \mathbf{E}\left[|D_z F|^2 |D_z L^{-1} F| \right] \mathrm{d}\sigma(z), \tag{21}$$

where D is the Poisson-Malliavin gradient (see Eqn. (11)), $L = D^*D$ the associated generator and the Stein class $\mathcal{F}$ is the space of twice differentiable functions with first derivative bounded by 1 and second order derivative bounded by 2. In [17], an analog result is given when $\mathbf{P}$ is a Gamma distribution and (G, μ) is either a Poisson or a Gaussian space. To the best of our knowledge, when μ is the distribution of a family of independent random variables, the distance $\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_F)$ is evaluated through exchangeable pairs or coupling, which means to construct an ad-hoc structure for each situation at hand. We intend to give here an exact analog to (21) in this situation using only our newly defined operator D. Our first result concerns the Gaussian approximation. To the best of our knowledge, there does not yet exist a Stein criterion for Gaussian approximation which does not rely on exchangeable pairs or any other sort of coupling.

Remark 4. *In what follows, we deal with functions F defined on E_A, that means that F is a function of X_A and as such, we should use the notation $F(X_A)$. For the sake of notations, we identify F and $F(X_A)$.*

Theorem 5.9. *Let* $\mathbf{P}$ *denote the standard Gaussian distribution on* $\mathbf{R}$*. For any* $F : E_A \to \mathbf{R}$ *such that* $\mathbf{E}[F] = 0$ *and* $F \in \operatorname{Dom} D$*. Then,*

$$\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_F) \le \mathbf{E}\left[\left|1 - \sum_{a\in A} D_a F\ (-D_a L^{-1})F\right|\right] + \sum_{a\in A} \mathbf{E}\left[\int_{E_A} \Big(F - F(X_{A\neg a}; x)\Big)^2 \mathrm{d}\mathbf{P}_a(x)\ |D_a L^{-1} F|\right].$$

The proof of this version follows exactly the lines of the proof of Theorem 3.1 in [25, 34] but we can do slightly better by changing a detail in the Taylor expansion.

Theorem 5.10. *Let* $\mathbf{P}$ *denote the standard Gaussian distribution on* $\mathbf{R}$*. For any* $F : E_A \to \mathbf{R}$ *such that* $\mathbf{E}[F] = 0$ *and* $F \in \operatorname{Dom} D$*. Then,*

$$\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_F) \le \sup_{\psi\in \mathrm{Lip}_2} \mathbf{E}\left[\psi(F) - \sum_{a\in A} \psi(F(X'_{\neg a})) D_a F(-D_a L^{-1})F\right] + \sum_{a\in A} \mathbf{E}\left[\int_{E_A} \Big(F - F(X_{A\neg a}; x)\Big)^2 \mathrm{d}\mathbf{P}_a(x)\ |D_a L^{-1} F|\right], \quad (22)$$

where $X'_{\neg a} = X_{A\neg a} \cup \{X'_a\}$.

This formulation may seem cumbersome, but it easily gives a close to the usual bound in the Lyapounov central limit theorem, with a non optimal constant (see [18]).

Corollary 5.11. *Let* $(X_n,\, n \ge 1)$ *be a sequence of thrice integrable, independent random variables. Denote*

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

Then,

$$\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_{Y_n}) \le \frac{2(\sqrt{2}+1)}{s_n^3} \sum_{j=1}^n \mathbf{E}\left[|X_j - \mathbf{E}[X_j]|^3\right].$$

Remark 5. *If we use Theorem 5.9, we get*

$$\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_{Y_n}) \le \mathbf{E}\left[\left|1 - \sum_{j=1}^n \frac{X_j^2}{s_n^2}\right|\right] + \frac{2}{s_n^3} \sum_{j=1}^n \mathbf{E}\left[|X_j - \mathbf{E}[X_j]|^3\right]$$

and the quadratic term is easily bounded only if the X_i*'s are such that* $\mathbf{E}\left[X_i^4\right]$ *is finite, which in view of Corollary 5.11 is a too stringent condition.*

The functional which appears in the central limit theorem is the basic example of U-statistics or homogeneous sums. If we want to go further and address the problem of convergence of more general U-statistics (or homogeneous sums), we need to develop a similar apparatus for the Gamma distribution. Recall that the Gamma distribution of parameters r and λ has density

$$f_{r,\lambda}(x) = \frac{\lambda^r}{\Gamma(r)}\, x^{r-1} e^{-\lambda x}\; \mathbf{1}_{\mathbf{R}^+}(x).$$

Let $Y_{r,\lambda} \sim \Gamma(r,\lambda)$, it has mean r/λ and variance r/λ^2. Denote by $\overline{Y}_{r,\lambda} = Y_{r,\lambda} - r/\lambda$. As described in [21], $Z \sim \overline{Y}_{r,\lambda} = Y_{r,\lambda} - r/\lambda$ if and only if $\mathbf{E}\left[L_{r,\lambda} f(Z)\right] = 0$ for any f once differentiable, where

$$L_{r,\lambda} f(y) = \frac{1}{\lambda}\left(y + \frac{r}{\lambda}\right) f'(y) - y f(y).$$

The Stein equation

$$L_{r,\lambda} f(y) = g(y) - \mathbf{E}\left[g(\overline{Y}_{r,\lambda})\right] \tag{23}$$

has a solution f_g which satisfies

$$\begin{aligned} \|f_g\|_\infty \le \|g'\|_\infty,\ \|f_g'\|_\infty \le 2\lambda \max\left(1, \frac{1}{r}\right) \|g'\|_\infty \\ \text{and } \|f_g''\|_\infty \le 2\lambda \left(\max\left(\lambda, \frac{\lambda}{r}\right) \|g'\|_\infty + \|g''\|_\infty\right), \end{aligned} \tag{24}$$

noting that f_g is solution of (23) if and only if $h_g : x \mapsto \dfrac{1}{\lambda} f\left(x - \dfrac{r}{\lambda}\right)$ solves

$$x h'(x) + (r - \lambda x) h(x) = g(x) - \mathbf{E}\left[g(Y_{r,\lambda})\right],$$

studied in [2, 17].

Theorem 5.12. *Let $\mathcal{F}$ is the set of twice differentiable functions with first and second derivative bounded by 1. There exists $c > 0$ such that for any $F \in \operatorname{Dom} D$ with $\mathbf{E}\left[F\right] = 0$,*

$$\begin{aligned} \kappa_{\mathcal{H}}(\mathbf{P}_F, \mathbf{P}_{\overline{Y}_{r,\lambda}}) \le c\, \mathbf{E}\left[\left|\frac{1}{\lambda} F + \frac{r}{\lambda^2} - \sum_{a \in A} D_a F(-D_a L^{-1}) F\right|\right] \\ + c \sum_{a \in A} \mathbf{E}\left[\int_{E_A} \Big(F(X_A) - F(X_{A \neg a}; x)\Big)^2 \mathrm{d}\mathbf{P}_a(x)\ |D_a L^{-1} F|\right]. \end{aligned} \tag{25}$$

This theorem reads exactly as [17, Theorem 1.5] for Poisson functionals and is proved in a similar fashion.

Remark 6. *The generalization of this result to multivariate Gamma distribution will be considered in a forthcoming paper. The difficulty lies in the regularity estimates of the solution of the Stein equation associated to multivariate Gamma distribution, which require lengthy calculations.*

An homogeneous sum of order d is a functional of independent identically distributed random variables $(X_1, \cdots, X_{N_n})$, of the form

$$F_n(X_1, \cdots, X_{N_n}) = \sum_{1 \le i_1, \cdots, i_d \le N_n} f_n(i_1, \cdots, i_d)\, X_{i_1} \ldots X_{i_d}$$

where $(N_n, n \ge 1)$ is a sequence of integers which tends to infinity as n does and the functions f_n are symmetric on $\{1, \cdots, N_n\}^d$ and vanish on the diagonal. The asymptotics of these sums have been widely investigated and depend on the properties of the function f_n. For $d = 2$, see for instance [20]. In [29], the case of any value of d is investigated through the prism of universality: roughly speaking (see Theorem 4.1), if $F_n(G_1, \cdots, G_{N_n})$ converges in distribution when $G_1, \cdots, G_{N_n}$ are standard Gaussian random variables then $F_n(X_1, \cdots, X_{N_n})$ converges to the same limit whenever the X_i's are centered with unit variance and finite third order moment and such that

$$\max_i \sum_{1 \le i_2, \cdots, i_d \le N_n} f_n^2(i, i_2, \cdots, i_d) \xrightarrow{n \to \infty} 0.$$

For Gaussian random variables, the functional F_n belongs to the d-th Wiener chaos. Combining the algebraic rules of multiplication of iterated Gaussian integrals and the Stein-Malliavin method, it is proved in [28] that $F_n(G_1, \cdots, G_{N_n})$ converges in distribution to a chi-square distribution of parameter ν if and only if

$$\mathbf{E}\left[F_n^2\right] \xrightarrow{n \to \infty} 2\nu \text{ and } \mathbf{E}\left[F_n^4\right] - 12\,\mathbf{E}\left[F_n^3\right] - 12\nu^2 + 48\nu \xrightarrow{n \to \infty} 0.$$

We obtain here a related result for $d = 2$ (for the sake of simplicity though the method is applicable for any value of d) and a general distribution without resorting to universality.

Let $A = \{1, \cdots, n\}$. For $f, g : A^2 \to \mathbf{R}$, symmetric functions vanishing on the diagonal, define the two contractions by

$$(f \star_1^1 g)(i,j) = \sum_{k \in A} f(i,k) g(j,k),$$
$$(f \star_2^1 g)(i) = \sum_{j \in A} f(i,j) g(i,j).$$

Theorem 5.13. *Let $X_A = \{X_i,\ 1 \le i \le n\}$ be a collection of centered independent random variables with unit variance and finite moment of order 4. Define*

$$F(X_A) = \sum_{(i,j) \in A^{\neq}} f(i,j)\, X_i X_j$$

where $(i,j)\in A^{\neq}$ means that we enumerate all the couples (i,j) in A^2 with distinct components and f is a symmetric function which vanishes on the diagonal. Let $\nu=\sum_{(i,j)} f^2(i,j)$. Then, there exists $c_\nu>0$ such that

$$\kappa^2_{\mathcal{H}}(\mathbf{P}_F,\,\mathbf{P}_{\bar{Y}_{\nu/2,1/2}})\le c_\nu \mathbf{E}\left[X_1^4\right]^2 \times\left[\sum_{(i,a)\in A^2} f^4(i,a)+\|f\star_2^1 f\|^2_{L^2(A)}+\|f-f\star_1^1 f\|^2_{L^2(A^2)}\right]. \quad (26)$$

We now introduce $\mathrm{Inf}_a(f)$, called the influence of the variable a, by

$$\mathrm{Inf}_a(f)=\sum_{i\in A} f^2(i,a).$$

Remark that

$$\begin{aligned}\sum_{i\in A} f^4(i,a) &\le \sum_{a\in A}\sum_i f^2(i,a)\sum_j f^2(j,a)\\ &=\sum_{a\in A}\sum_i f^2(i,a)\,\mathrm{Inf}_a(f)\\ &\le \nu\max_{a\in A}\mathrm{Inf}_a(f).\end{aligned}$$

The same kind of computations can be made for $\|f\star_2^1 f\|^2_{L^2(A)}$. As a consequence, we get the following corollary.

Corollary 5.14. *With the same notations as above,*

$$\kappa^2_{\mathcal{H}}(\mathbf{P}_F,\,\mathbf{P}_{\bar{Y}_{\nu/2,1/2}})\le c_\nu \mathbf{E}\left[X_1^4\right]^2\left[\max_{a\in A}\mathrm{Inf}_a(f)+\|f-f\star_1^1 f\|^2_{L^2(A^2)}\right].$$

The supremum of the influence is the quantity which governs the distance between the distributions of $F_n(G_1,\cdots,G_{N_n})$ and $F_n(X_1,\cdots,X_{N_n})$ in [29], thus it is not surprising that it still appears here.

Remark 7. *A tedious computation shows that*

$$\begin{aligned}&\mathbf{E}\left[F^4\right]-12\mathbf{E}\left[F^3\right]-12\nu^2+48\nu\\ &\quad=\sum_{(i,j)\in A^{\neq}} f^4(i,j)\,\mathbf{E}\left[X^4\right]^2+6\sum_{(i,j,k)\in A^{\neq}} f^2(i,j)f^2(i,k)\mathbf{E}\left[X^4\right]\\ &\quad+12\,\mathbf{E}\left[X^3\right]^2\left\{\sum_{(i,j,k)\in A^{\neq}} f^2(i,j)\,f(i,k)\,f(k,j)-\sum_{(i,j)\in A^{\neq}} f^3(i,j)\right\}\\ &-48\left\{\sum_{(i,j,k)\in A^{\neq}} f(i,j)f(i,k)f(k,j)-f^2(i,j)\right\}-12\sum_{(i,j)\in A^{\neq}} f^4(i,j).\end{aligned} \quad (27)$$

The Cauchy-Schwarz inequality entails that the properties

$$\mathbf{E}\left[F_n^4\right] - 12\mathbf{E}\left[F_n^3\right] - 12\nu^2 + 48\nu \xrightarrow{n\to\infty} 0$$

and

$$\kappa_{\mathcal{H}}(\mathbf{P}_F,\, \mathbf{P}_{\bar{Y}_{\nu/2,1/2}}) \xrightarrow{n\to\infty} 0$$

share the same sufficient condition:

$$\sum_{(i,a)\in A^{\neq}} f^4(i,a) + \|f \star_2^1 f\|^2_{L^2(A)} + \|f - f \star_1^1 f\|^2_{L^2(A^2)} \xrightarrow{n\to\infty} 0.$$

However, we cannot go further and state a fourth moment theorem as we know, that for Benoulli random variables, F_n may converge to $\bar{Y}_{\nu/2,1/2}$ while the RHS of (26) does not converge to 0.

As another corollary of Theorem 5.13, we obtain the KR distance between a degenerate U-statistics of order 2 and a Gamma distribution. Compared to the more general [17, Theorem 1.1], the computations are here greatly simplified by the absence of exchangeable pairs.

Theorem 5.15. *Let $A = \{1, \cdots, n\}$ and $(X_i, i \in A)$ a family of independent and identically distributed real-valued random variables such that*

$$\mathbf{E}\left[X_1\right] = 0,\ \mathbf{E}\left[X_1^2\right] = \sigma^2 \text{ and } \mathbf{E}\left[X_1^4\right] < \infty.$$

Consider the random variable

$$F = \frac{2}{n-1} \sum_{(i,j)\in A^{\neq}} X_i X_j.$$

Then, there exists $c > 0$, independent of n, such that

$$\kappa_{\mathcal{H}}(\mathbf{P}_F,\, \mathbf{P}_{\overline{Y}_{1/2,1/2\sigma^2}}) \le c\,\frac{\sigma^2}{\sqrt{n}}\,\mathbf{E}\left[X_1^4\right]. \tag{28}$$

Proof. Take $f_n(i,j) = 2/(n-1)$ and apply Theorem 5.13. □

Remark 8. *The proof of Theorem 5.13 is rich of insights. In Gaussian, Poisson or Rademacher contexts, the computation of $L^{-1}F$ is easily done when there exists a chaos decomposition since L operates as a dilation on each chaos (see [25, 26, 34]). In [37, Lemma 3.4 and below], a formula for L^{-1} of Poisson driven U-statistics is given, not resorting to the chaos decomposition. It is based on the fact that L applied to a U-statistics F of order k yields kF plus a U-statistics of order $(k-1)$. Then, the construction of an inverse formula can be made by induction. In our framework, the action of L on a U-statistics yields kF plus a U-statistics of order k so that no induction seems possible. However, for an order k U-statistics which is degenerate of order $(k-1)$, we have $LF = kF$. For $k = 2$, this hypothesis of degeneracy is exactly the sufficient condition to have a convergence towards a Gamma distribution.*

6. Proofs

6.1. Proofs of Section 2

Proof of Theorem 2.4. The process trace$(DU) = (D_a U_a,\, a \in B)$ belongs to $L^2(A \times E_A)$: Using the Jensen inequality, we have

$$\| \operatorname{trace}(DU)\|^2_{L^2(A\times E_A)} = \mathbf{E}\left[\sum_{a\in B} |D_a U_a|^2\right] \le 2 \sum_{a\in B} \mathbf{E}\left[U_a^2\right] < \infty. \tag{29}$$

Moreover,

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

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

Proof of corollary 2.5. Let $(F_n,\, n \ge 1)$ 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 (3)

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

where $\sum_{a\in A} D_a U_a \in L^2(E_A)$ in view of (29). Then,

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

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

Proof of Lemma 2.6. Since $\sup_n \|DF_n\|_{\mathbf{D}}$ is finite, there exists a subsequence which we still denote by $(DF_n, n \ge 1)$ 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 $m \ge n_k$. The Mazur's Theorem implies that there exists a convex combination of elements of $(DF_m, m \ge n_k)$ such that

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

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

$$\Big\| \sum_{i=1}^{M_k} \alpha_i^k F_{m_i} - F \Big\|_{L^2(E_A)} \le 1/k.$$

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 $DF = \eta$. □

Proof of Theorem 2.11. To prove the existence of $(P_t, t \ge 0)$ for a countable set, we apply the Hille-Yosida theorem:

Proposition 6.1 (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. $\operatorname{Dom} L$ *is dense in* $L^2(E_A)$.
2. L *is dissipative i.e. for any* $\lambda > 0, F \in \operatorname{Dom} L$,

$$\|\lambda F - LF\|_{L^2(E_A)} \ge \lambda \|F\|_{L^2(E_A)}.$$

3. *Im*$(\lambda \operatorname{Id} - L)$ *dense in* $L^2(E_A)$.

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

Let $(A_n,\, n \ge 1)$ an increasing sequence of subsets of A such that $\cup_{n\ge 1} A_n = A$. For $F \in L^2(E_A)$, let $F_n = \mathbf{E}\left[F \,|\, \mathcal{F}_{A_n}\right]$. Since $(F_n,\, n \ge 1)$ is a square integrable martingale, F_n converges to F both almost-surely and in $L^2(E_A)$. For any $n \ge 1$, 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 3.2, $D_a F_n = \mathbf{E}\left[D_a F \,|\, \mathcal{F}_{A_n}\right]$, hence

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

Therefore, point (2) is satisfied.

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

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

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 $\lambda \operatorname{Id} - L$ and we already know it converges in $L^2(E_A)$ to F. □

Proof of Theorem 2.13. 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

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

Then,

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

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 $F_n = \mathbf{E}\left[F \,|\, \mathcal{F}_{A_n}\right]$. 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 known that $P_t F_n = P_t^n F_n$ where P^n is the semi-group associated to A_n, hence

$$D_a P_t F_n = D_a P_t^n F_n = e^{-t} P_t^n D_a F_n. \tag{30}$$

Moreover,

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

According to Lemma [2.6], this means that $P_t F$ belongs to $\mathbf{D}$. Let n go to infinity in (30) yields (6). □

Proof of Lemma 2.9. For U and V in $\mathcal{S}_0(l^2(A))$, from the integration by parts formula,

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

It follows that $\mathbf{E}\left[\delta U^2\right] \le \|U\|^2_{\mathbf{D}(l^2(A))}$. Then, by density, $\mathbf{D}(l^2(A)) \subset \operatorname{Dom}\delta$ and Eqn. (4) holds for U and V in $\operatorname{Dom}\delta$. □

6.2. Proofs of Section 3

Proof of Lemma 3.1. 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. □

Proof of Lemma 7. For any $k \geq 1$, $\mathcal{F}_k \cap \mathcal{G}_k = \mathcal{F}_{k-1}$, hence

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

The proof is thus complete. □

Proof of Theorem 3.3. Let F an $\mathcal{F}_n$-measurable random variable. It is clear that

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

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

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

Remark that for $l > k$, in view of Lemma 3.1,

$$\mathbf{E}\left[D_k\, \mathbf{E}\left[F \,|\, \mathcal{F}_k\right] D_l\, \mathbf{E}\left[F \,|\, \mathcal{F}_l\right]\right] = \mathbf{E}\left[D_l D_k\, \mathbf{E}\left[F \,|\, \mathcal{F}_k\right] \mathbf{E}\left[F \,|\, \mathcal{F}_l\right]\right] = 0, \tag{31}$$

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

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

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

We now analyze the non-ordered situation. If A is finite, each bijection between A and $\{1, \cdots, n\}$ defines an order on A. Hence, there are $|A|!$ possible filtrations. Each term of the form

$$D_{i_k} \mathbf{E}\left[F \,|\, X_{i_1}, \cdots, X_{i_k}\right]$$

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

Proof of Theorem 3.4. Remark that

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

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

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

by the very definition of the gradient map. As in (31), we can show that for any N,

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

Consider $F_N=\mathbf{E}\left[F\,|\,\mathcal{F}_N\right]$ and proceed as in the proof of Lemma 3.3 to conclude. □

Proof of Corollary 3.5. According to (31) and (7), we have

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

where the inequality follows from then Jensen inequality. □

Proof of Theorem 3.6. Let $F,G\in\mathbf{D}$, the Clark formula entails

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

where we have used (31) in the third equality and the identity $D_kD_k=D_k$ in the last one. □

Proof of Theorem 3.7. Let $F, G \in L^2(E_A)$.

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

when we have used the orthogonality of the sum, (6) and the $\mathcal{F}_k$-measurability of $P_tD_k\mathbf{E}\left[G|\mathcal{F}_k\right]$ to get the last equality. □

Proof of Theorem 3.8. Assume with no loss of generality that F is centered. Apply (8) to θF and $e^{\theta F}$,

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

Recall that

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

where $\Delta_kF = F - F(X_{\neg k}, X'_k)$ so that $D_kF = \mathbf{E}'\left[\Delta_kF\right]$.

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

$$D_k\mathbf{E}\left[e^{\theta F}\,|\,\mathcal{F}_k\right] \le \mathbf{E}\left[e^{\theta F}(1-e^{-\theta D_kF})\,|\,\mathcal{F}_k\right] \le \theta\;\mathbf{E}\left[e^{\theta F}\,|D_kF|\,|\,\mathcal{F}_k\right].$$

Thus,

$$\left|\mathbf{E}\left[Fe^{\theta F}\right]\right| \le \theta\;\mathbf{E}\left[e^{\theta F}\sum_{k=1}^{\infty}|D_kF|\,\mathbf{E}\left[|D_kF|\,|\,\mathcal{F}_k\right]\right] \le M\,\theta\;\mathbf{E}\left[e^{\theta F}\right].$$

By Gronwall lemma, this implies that

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

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. □

Proof of Theorem 3.9. We follow closely the proof of [43] for Poisson process. Let $G \in L^2(E_A)$ be a positive random variable such that $DG \in L^2(A \times E_A)$. For any non-zero 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

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

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

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

We know from [43] that for any non-zero integer k, $\ell(\mathbf{E}\left[G_n \,|\, \mathcal{G}_k\right], D_k G_n)$ converges increasingly to $\ell(\mathbf{E}\left[G \,|\, \mathcal{G}_k\right], D_k G)$ $\mathbf{P}$-a.s., hence by Fatou Lemma,

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

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

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

The proof is thus complete. $\square$

Proof of Theorem 3.10. 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 3.3) entails that $0 = \mathbf{E}\left[\varphi - \varphi'\right] = \varphi - \varphi'$.

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

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

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 (31),

$$\mathbf{E}\left[\left(\sum_{k=1}^{\infty}\mathbf{E}\left[D_kU_k \,|\, \mathcal{F}_k\right]\right)^2\right] = \mathbf{E}\left[\left(\sum_{k=1}^{\infty}D_k\mathbf{E}\left[U_k \,|\, \mathcal{F}_k\right]\right)^2\right]$$
$$= \mathbf{E}\left[\sum_{k=1}^{\infty}\Big(D_k\mathbf{E}\left[U_k \,|\, \mathcal{F}_k\right]\Big)^2\right] \le \sum_{k=1}^{\infty}\mathbf{E}\left[|D_kU_k|^2\right] \le \|U\|^2_{\mathbf{D}(l^2(A))},$$

hence

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

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

6.3. Proofs of Section 4

Proof of Theorem 4.5. Starting from (15), the terms with $\tau = 0$ can be decomposed as

$$e^{-2p_m^N}\sum_{m=1}^{N}\mathbf{E}\left[\Big(F(\omega^N_{(m)} + \varepsilon_{\zeta^N_m}) - F(\omega^N_{(m)})\Big)^2\right]\mu^N_m(1) + R^N_0.$$

Since F belongs to $\mathrm{TV}-\mathrm{Lip}$,

$$R^N_0 \le \sum_{m=1}^{N}\sum_{\ell=2}^{\infty} l^2\mu^N_m(l) \le c_1\, N(p^N)^2\mathbf{E}\left[(\mathrm{Poisson}(p^N)+2)^2\right] \le c_2\, N(p^N)^2,$$

where the c_1 and c_2 are irrelevant constants. As Np^N is bounded, R^N_0 goes to 0 as N grows to infinity. For the very same reasons, the sum of the terms of (15) with $\tau \ge 1$ converge to 0, thus

$$\lim_{N\to\infty}\mathcal{E}^{U_N}(F) = \lim_{N\to\infty}\sum_{m=1}^{N}e^{-2p_m^N}\,\mathbf{E}\left[\Big(F(\omega^N_{(m)} + \varepsilon_{\zeta^N_m}) - F(\omega^N_{(m)})\Big)^2\right]p^N_m.$$

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

$$\begin{aligned}\psi\,:\,\mathfrak{N}_{\mathbb{Y}} \times \{\zeta^N_k,\, k = 1, \cdots, N\} &\longrightarrow E\\ (\omega,\,\zeta) &\longmapsto \Big(F(\omega - (\omega(\zeta)-1)\varepsilon_\zeta) - F(\omega - \omega(\zeta)\varepsilon_\zeta)\Big)^2.\end{aligned}$$

Then, we can write

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

Under $\tilde{\mathbf{P}}_N$, the random variables ω and ζ are independent. Equation (14) 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 $\mathrm{TV}-\mathrm{Lip}$, ψ is continuous and bounded, hence the result. $\square$

Proof of Theorem 4.9. For $F \in \mathbf{D}_B \cap \text{H-C}^1$, in view of (17), we have

$$F(\omega^N) - F(\omega^N_{(k)} + M'_k\, h^N_k) = (M_k - M'_k)\,\langle \nabla F(\omega^N_{(k)}),\, h^N_k\rangle_H + \frac{M_k - M'_k}{\sqrt{N}}\, \varepsilon(\omega^N_{(k)}, h^N_k).$$

Hence,

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

and

$$\text{Rem} \le \frac{c}{N}\sum_{k=1}^N \mathbf{E}\left[\varepsilon(\omega^N_{(k)}, h^N_k)^2\right] \xrightarrow{N\to\infty} 0,$$

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

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

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 $k \in \{1, \ldots, N\}$, for any bounded ψ and φ,

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

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

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

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

6.4. Proofs of Section 5

Proof of Theorem 5.2. 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,

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

It remains to prove that

$$\begin{aligned}
&\sum_{k=1}^{m} \binom{m}{k} H_n^{(k)} \\
&= \binom{n}{m}^{-1} \sum_{B\subset[n], |B|\leq m} \binom{m}{|B|}^{-1} \frac{1}{|B|} \sum_{b\in B} \sum_{C\in([n]\setminus B,\, m-|B|)} D_b \mathbf{E}\left[h(X_{B\cup C}) \mid X_B\right].
\end{aligned} \tag{32}$$

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

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

Assume the existence of an integer n such that (32) holds for any set of cardinality n. In particular, for any $l \in [n+1]$

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

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

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

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

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

by applying the Clark formula to h. $\square$

Proof of Theorem 5.6. By the previous construction, for

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

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 cardinality $(N-1)!$, they do coincide. The values of p_k and α_k are easily computed since the random variables $(I_m,\, k \le m \le N)$ are independent. According to Theorem 3.4,

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

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

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

For $l = k+1$,

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

The subsequent terms are handled similarly and the result follows. □

Proof of Theorem 5.7. By the very definition of $\tilde{C}_1$, we have

$$\tilde{C}_1 = \mathbf{E}\left[\tilde{C}_1\right] + \sum_{k=1}^{N} \sum_{l=k}^{N} D_l \mathbf{E}\left[\tilde{U}_k^N \,|\, \mathcal{H}_{l-1}\right]. \tag{33}$$

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

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

It is straightforward that $l>k$,

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

The result then follows by direct computations. □

Proof of Theorem 5.8. Recall that for $j\neq l$, $D_l\mathbf{E}\left[\tilde{U}_k^N \,|\, \mathcal{H}_{l-1}\right]$ and $D_j\mathbf{E}\left[\tilde{U}_m^N \,|\, \mathcal{H}_{j-1}\right]$ are orthogonal in L^2. In view of (33), according to the integration by parts formula, we have

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

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

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

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 $l > k$,

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

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

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

It follows that

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

The proof is thus complete. □

Proof of Theorem 5.10. We have to compute

$$\sup_{\varphi\in\mathcal{F}} \mathbf{E}\left[\varphi'(F) - F\varphi(F)\right],$$

where $\mathcal{F}$ is the set of twice differentiable functions with second order derivative bounded by 2. Since F is centered

$$\mathbf{E}\left[F\varphi(F)\right] = \mathbf{E}\left[LL^{-1}F\,\varphi(F)\right] = \sum_{a\in A}\mathbf{E}\left[(-D_aL^{-1})F\,D_a\varphi(F)\right].$$

The trick is to use the Taylor expansion taking the reference point to be $X'_{\neg a}$ instead of X_A. This yields

$$D_a\varphi(F) = \mathbf{E}'\left[\varphi(F(X_A)) - \varphi(F(X'_{\neg a}, X'_a))\right] = \varphi'(F(X'_{\neg a}))D_aF + R,$$

where

$$R = \frac{1}{2}\int_0^1 \mathbf{E}'\left[\varphi''\Big(\theta F(X'_{\neg a}) + (1-\theta)F(X_A)\Big)\Big(F(X_A) - F(X'_{\neg a})\Big)^2\right]\mathrm{d}\theta.$$

Hence

$$\begin{aligned}\mathbf{E}\left[\varphi'(F) - F\varphi(F)\right] &\\ = \mathbf{E}&\left[\varphi'(F) - \sum_{a\in A}\varphi'(F(X'_{\neg a}))\;D_aF(-D_aL^{-1})F\right]\\ &\qquad + \sum_{a\in A}\mathbf{E}\left[R\;(-D_aL^{-1})F\right].\end{aligned}$$

The rightmost term of the the latter equation easily yields the rightmost of (22). Since $\|\varphi''\|_\infty < 2$, it is clear that φ' belongs to Lip_2 hence the formulation of the distance with a supremum. □

Proof of Corollary 5.11. Without loss of generality, we can assume that X_i is centered for any $i \geq 1$. Remark that

$$D_jX_k = \begin{cases} 0 & \text{if } j \neq k,\\ X_k & \text{if } j = k.\end{cases}$$

Hence $LY_n = Y_n$ and $Y_n = L^{-1}Y_n$. According to Theorem 5.10,

$$\begin{aligned}\kappa_{\mathcal{H}}(\mathbf{P}, \mathbf{P}_{Y_n}) \leq \sup_{\psi\in\mathrm{Lip}_2}\mathbf{E}&\left[\psi(F) - \frac{1}{s_n^2}\sum_{i\in A}\psi\Big(F\big(Y_n - \frac{X_i - X'_i}{s_n}\big)\Big)X_i^2\right]\\ &+ \frac{1}{s_n^3}\sum_{j=1}^n\mathbf{E}\left[\int_{E_A}\left(X_i - x\right)^2\mathrm{d}\mathbf{P}_i(x)\;|X_i|\right].\end{aligned}$$

By independence, since ψ is 2-Lipschitz continuous,

$$\begin{aligned}\Bigg|\mathbf{E}\left[\psi(F)-\frac{1}{s_n^2}\sum_{i\in A}\psi\Big(F(Y_n-\frac{X_i-X_i'}{s_n})\Big)X_i^2\right]\Bigg|\\ =\left|\frac{1}{s_n^2}\sum_{i\in A}\sigma_i^2\,\mathbf{E}\left[\psi(F)-\psi\Big(F(Y_n-\frac{X_i-X_i'}{s_n})\Big)\right]\right|\\ \le\frac{2}{s_n^3}\sum_{i\in A}\sigma_i^2\,\mathbf{E}\left[|X_i-X_i'|\right]\le\frac{2\sqrt{2}}{s_n^3}\sum_{i\in A}\sigma_i^3.\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbf{E}\left[\int_{E_A}\Big(X_i-x\Big)^2\mathrm{d}\mathbf{P}_i(x)\ |X_i|\right]&=\mathbf{E}\left[|X_i|^3\right]+\sigma^2\mathbf{E}\left[|X_i|\right]\\ &\le\mathbf{E}\left[|X_i|^3\right]+\sigma^3\le 2\,\mathbf{E}\left[|X_i|^3\right]\end{aligned}$$

according to the Hölder inequality. Hence the result. □

Proof of Theorem 5.12. According to the principle of the Stein method, we have to estimate

$$\mathbf{E}\left[\frac{1}{\lambda}\left(\varphi(F)+\frac{r}{\lambda}\right)-F\varphi'(F)\right],\tag{34}$$

where φ and its derivatives satisfy (24). For any $a\in A$, thanks to the Taylor expansion,

$$-D_a\varphi(F)=\mathbf{E}'\left[\varphi(F(X^{\neg a},X_a'))-\varphi(F(X))\right]=-\varphi'(F)D_aF+R,\tag{35}$$

where

$$\begin{aligned}R&=\frac{1}{2}\int_0^1(1-\theta)\\ &\times\mathbf{E}'\left[\varphi''\Big((1-\theta)F(X)+\theta F(X^{\neg a},X_a')\Big)\Big(F(X)-F(X^{\neg a},X_a')\Big)^2\right]\mathrm{d}\theta\end{aligned}\tag{36}$$

According to (3) and to the definition of L,

$$\begin{aligned}\mathbf{E}\left[F\varphi(F)\right]=\mathbf{E}\left[LL^{-1}F\,\varphi(F)\right]&=\mathbf{E}\left[-\delta(DL^{-1}F)\varphi(F)\right]\\ &=\mathbf{E}\left[\langle D\varphi(F),-DL^{-1}F\rangle_{L^2(A)}\right].\end{aligned}\tag{37}$$

Plug (35) into (37):

$$\begin{aligned}\mathbf{E}&\left[\langle D\varphi(F),-DL^{-1}F\rangle_{L^2(A)}\right]\\ &=-\sum_{a\in A}\mathbf{E}\left[D_a\varphi(F)D_a(L^{-1}F)\right]\\ &=-\sum_{a\in A}\mathbf{E}\left[\varphi'(F)D_aFD_a(L^{-1}F)\right]+\sum_{a\in A}\mathbf{E}\left[R\ D_a(L^{-1}F)\right]\\ &=\mathbf{E}\left[\varphi'(F)\langle DF,-DL^{-1}F\rangle_{L^2(A)}\right]+\mathbf{E}\left[\langle R,-DL^{-1}F\rangle_{L^2(A)}\right].\end{aligned}$$

Then,

$$\left|\mathbf{E}\left[\frac{1}{\lambda}(F+\frac{r}{\lambda})\varphi'(F)-F\varphi(F)\right]\right|$$
$$\leq \left|\mathbf{E}\left[\varphi'(F)\,\left(\frac{1}{\lambda}(F+\frac{r}{\lambda})-\langle DF,-DL^{-1}F\rangle_{L^2(A)}\right)\right]\right|$$
$$+\left|\mathbf{E}\left[\langle R,-DL^{-1}F\rangle_{L^2(A)}\right)\right]\right| = B_1+B_2.$$

Since φ' is bounded, we get

$$B_1 \leq \|\varphi'\|_\infty \mathbf{E}\left[\left|\frac{1}{\lambda}(F+\frac{r}{\lambda})-\langle DF,-DL^{-1}F\rangle_{L^2(A)}\right|\right]$$

and from (36), we deduce that

$$B_2 \leq \|\varphi''\|_\infty \sum_{a\in A}\mathbf{E}\left[|D_aF|^2|D_aL^{-1}F|\right].$$

The proof follows from (34) and (24). □

Proof of Theorem 5.13. For any $a\in A$,

$$D_a(X_iX_j)=\begin{cases} X_aX_j & \text{if } a=i\\ X_iX_a & \text{if } a=j\\ 0 & \text{otherwise.}\end{cases}$$

Then,

$$D_aF=\sum_{(i,a)\in A^{\neq}} f(i,a)X_iX_a+\sum_{(j,a)\in A^{\neq}} f(a,j)X_aX_j = 2\sum_{(i,a)\in A^{\neq}} f(i,a)X_iX_a,$$

so that

$$LF=-\sum_{a\in A}D_aF=-2F \quad \text{and} \quad L^{-1}F=-\frac{1}{2}F.$$

With our notations, the first term of the right-hand-side of (25) becomes

$$\mathbf{E}\left[\left|2F+2\nu-2\sum_{a\in A}\sum_{(i,j)\in A^2} f(i,a)f(j,a)X_a^2X_iX_j\right|\right] \leq \sum_{i=1}^{2} A_i, \tag{38}$$

where

$$A_1 = 2\,\mathbf{E}\left[\left|\sum_{(i,a)\in A^2} f^2(i,a)(X_a^2X_i^2-1)\right|\right],$$

$$A_2 = 2\,\mathbf{E}\left[\left|F-\sum_{a\in A}\sum_{(i,j)\in A^{\neq}} f(i,a)f(j,a)X_a^2X_iX_j\right|\right].$$

We first control A_1. According to the Cauchy-Schwarz inequality,

$$A_1^2 \leq 4\,\mathbf{E}\left[\sum_{(i,a)\in A^2}\sum_{(j,c)\in A^2} f^2(i,a)f^2(j,c)(X_a^2X_i^2-1)(X_c^2X_j^2-1)\right] \leq 4(A_{11}+A_{12}),$$

where

$$A_{11} = \mathbf{E}\left[\sum_{(i,a)\in A^2} f^4(i,a)(X_a^2X_i^2-1)^2\right],$$
$$A_{12} = \mathbf{E}\left[\sum_{a\in A}\sum_{(i,j)\in A^{\neq}} f^2(i,a)f^2(j,a)\,(X_a^2X_i^2-1)(X_a^2X_j^2-1)\right],$$

by orthogonality of the X_i's. On the one hand,

$$A_{11} \leq \sum_{(i,a)\in A^2} f^4(i,a)\mathbf{E}\left[\left(X_a^2X_i^2-1\right)^2\right] = \left(\mathbf{E}\left[X_1^4\right]^2-1\right)\sum_{(i,a)\in A^2} f^4(i,a). \quad (39)$$

On the other hand,

$$\begin{aligned}
A_{12} &= \mathbf{E}\left[\sum_{(i,j)^{\neq}\in A^2}\sum_{a\in A} f^2(i,a)f^2(j,a)(X_a^2X_i^2-1)(X_a^2X_j^2-1)\right]\\
&\leq \sum_{(i,j)^{\neq}\in A^2}\sum_{a\in A} f^2(i,a)f^2(j,a)\mathbf{E}\left[(X_a^2X_i^2-1)(X_a^2X_j^2-1)\right]\\
&= \left(\mathbf{E}\left[X_1^4\right]-1\right)\sum_{(i,a)\in A^2} f^2(i,a)\sum_{j\neq i} f^2(j,a)\\
&\leq \left(\mathbf{E}\left[X_1^4\right]-1\right)\|f\star_2^1 f\|^2_{L^2(A)}. \qquad (40)
\end{aligned}$$

In a similar way, $A_2 \leq A_{21}+A_{22}$, where

$$A_{21} = 2\,\mathbf{E}\left[\left|\sum_{(i,j)\in A^{\neq}} f(i,j)X_iX_j - \sum_{(i,j)\in A^{\neq}}\sum_{a\in A} f(i,a)f(j,a)X_iX_j\right|\right],$$
$$A_{22} = 2\,\mathbf{E}\left[\left|\sum_{(i,j)\in A^{\neq}}\sum_{a\in A} f(i,a)f(j,a)X_iX_j\left(X_a^2-\mathbf{E}\left[X_a^2\right]\right)\right|\right].$$

As above,

$$A_{21}^2 \leq 4\,\mathbf{E}\left[\left(\sum_{(i,j)\in A^{\neq}}\Big(f(i,j)-\sum_{a\in A}f(i,a)f(j,a)\Big)X_iX_j\right)^2\right] = 4\,\|f - f\star_1^1 f\|_2^2. \quad (41)$$

Furthermore, according to Cauchy-Schwarz inequality and by independence, we have

$$\begin{aligned}
A_{22} &\leq 2\sum_{(i,j)\in A^{\neq}}\mathbf{E}\left[|X_iX_j|\Big|\sum_{a\in A}f(i,a)f(j,a)(X_a^2-1)\Big|\right]\\
&\leq 2\,\mathbf{E}\left[\Big(\sum_{(i,j)\in A^{\neq}}\sum_{a\in A}f(i,a)f(j,a)(X_a^2-1)\Big)^2\right]^{1/2}\\
&\leq 2\left(\sum_{(i,j)\in A^{\neq}}\sum_{a\in A}f(i,a)^2f(j,a)^2\mathbf{E}\left[X_a^4-1\right]\right)^{1/2}\\
&\leq 2\big(\mathbf{E}\left[X_1^4\right]-1\big)^{1/2}\,\|f\star_2^1 f\|_{L^2(A)}. \qquad (42)
\end{aligned}$$

The remainder term is given by

$$A_3 = \sum_{a\in A}\mathbf{E}\left[\int_{E_A}\Big(F(X_A)-F(X_{A\neg a};x)\Big)^2\,\mathrm{d}\mathbf{P}_a(x)\;|D_aL^{-1}F|\right].$$

Once again, using the orthogonality, we have

$$\begin{aligned}
G_a(X_A) &= \int_{E_A}\Big(F(X_A)-F(X_{A\neg a};x)\Big)^2\,\mathrm{d}\mathbf{P}_a(x)\\
&= 4\,\mathbf{E}'\left[\Big(\sum_{i\in A}f(i,a)X_iX_a-\sum_{i\in A}f(i,a)X_iX_a'\Big)^2\right]\\
&= 4\,\mathbf{E}'\left[(X_a-X_a')^2\Big(\sum_{i\in A}f(i,a)X_i\Big)^2\right]\\
&= 4\Big(\sum_{i\in A}f(i,a)X_i\Big)^2\,\mathbf{E}'\left[(X_a-X_a')^2\right]\\
&= 4\Big(\sum_{i\in A}f(i,a)X_i\Big)^2\,\big(X_a^2+1\big).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{E}\left[\sum_{a\in A} G_a(X_A)^2\right] &= 16\,\mathbf{E}\left[\sum_{a\in A}\Big(\sum_{i\in A} f(i,a)X_i\Big)^4 \left(X_a^2+1\right)^2\right] \\
&= 16\left(\mathbf{E}\left[X_1^4\right]+3\right)\mathbf{E}\left[X_1^4\right]\sum_{a\in A}\sum_{i\in A} f^4(i,a) \\
&\qquad + 96\left(\mathbf{E}\left[X_1^4\right]+3\right)\sum_{a\in A}\sum_{(i,j)\in A^{\neq}} f^2(i,a)f^2(j,a) \\
&\leq 16\left(\mathbf{E}\left[X_1^4\right]+3\right)^2\sum_{a\in A}\sum_{i\in A} f^4(i,a) + 96\left(\mathbf{E}\left[X_1^4\right]+3\right)\|f\star_2^1 f\|^2_{L^2(A)}. \qquad (43)
\end{aligned}$$

Moreover,

$$\begin{aligned}
\sum_{a\in A}\mathbf{E}\left[|D_aL^{-1}F|^2\right] &= \frac{1}{4}\sum_{a\in A}\mathbf{E}\left[|D_aF|^2\right] \\
&= \sum_{a\in A}\mathbf{E}\left[\left(\sum_{(i,a)\in A^{\neq}} f(i,a)X_iX_a\right)^2\right] \\
&= \sum_{(i,a)\in A^{\neq}} f^2(i,a) = \nu. \qquad (44)
\end{aligned}$$

Combine (39)–(44) to obtain (26). □

Acknowledgments: The authors are indebted to the anonymous referees and to the AE for many helpful remarks which helped us to improve this paper.

[1] S. Albeverio, Yu. G. Kondratiev, and M. Röckner. Analysis and geometry on configuration spaces. *J. Funct. Anal.*, 154(2):444–500, 1998.

[2] Benjamin Arras and Yvik Swan. A stroll along the gamma. *Stochastic Processes and their Applications*, 127(11):3661–3688, nov 2017.

[3] Richard Arratia, A. D. Barbour, and Simon Tavaré. Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.*, 2 (3):519–535, 1992.

[4] Richard Arratia, A. D. Barbour, and Simon Tavaré. *Logarithmic combinatorial structures: a probabilistic approach.* EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.

[5] A. D. Barbour and L. H. Y. Chen. *An introduction to Stein's method*, volume 4 of *Lecture Notes Series.* National University of Singapore, 2005.

[6] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities.* Oxford University Press, Oxford, 2013.

[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*, 129:63–73, 2005.

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

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

[12] L. Coutin and L. Decreusefond. Higher order expansions via Stein's method. *Communications on Stochastic Analysis*, 8(2):155–168, 2014.

[13] L. Decreusefond. The Stein-Dirichlet-Malliavin method. *ESAIM: Proceedings*, page 11, 2015.

[14] L. Decreusefond and I. Flint. Moment formulae for general point processes. *Journal of Functional Analysis*, 267(2):452 – 476, 2014.

[15] L. Decreusefond, M. Schulte, and C. Thäle. Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. *Annals of probability*, 44(3):2147–2197, June 2015.

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

[17] C. Döbler and G. Peccati. The Gamma stein equation and non-central de Jong theorems.

[18] L. Goldstein. Bounds on the constant in the mean central limit theorem. *The Annals of Probability*, 38(4):1672–1689, 2010.

[19] L. Goldstein and G. Reinert. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, 7(4):935–952, 1997.

[20] F. Götze and A. Tikhomirov. Asymptotic distribution of quadratic forms and applications. *Journal of Theoretical Probability*, 15(2):423–475, 2002.

[21] P. Graczyk, J.-J. Loeb, I. A. López P., A. Nowak, and W. O. Urbina R. Higher order riesz transforms, fractional derivatives, and sobolev spaces for laguerre expansions. *Journal de Mathématiques Pures et Appliquées*, 84(3):375–405, mar 2005.

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

[23] S. Kerov, G. Olshanski, and A. Vershik. Harmonic analysis on the infinite symmetric group. *Inventiones mathematicae*, 158(3):551–642, dec 2004.

[24] V. S. Koroljuk and Yu. V. Borovskich. *Theory of U-statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.

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

[26] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. Cambridge University Press, 2012.

[27] I. Nourdin, G. Peccati, and G. Reinert. Stein's method and stochastic analysis of Rademacher functionals. *Electronic Journal of Probability*, 15:no. 55, 1703–1742, 2010.

[28] I. Nourdin and G. Peccati. Noncentral convergence of multiple integrals. *The Annals of Probability*, 37(4):1412–1426, 2009.

[29] I. Nourdin, G. Peccati, and G. Reinert. Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *The Annals of Probability*, 38(5):1947–1985, 2010.

[30] D. Nualart. *The Malliavin Calculus and Related Topics*. Springer-Verlag, 1995.

[31] D. Nualart and J. Vives. Anticipative calculus for the Poisson process based on the Fock space. In *Séminaire de probabilités XXIV*, pages 154–165. Springer-Verlag, 1988.

[32] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005.

[33] D. Nualart and W. Schoutens. Chaotic and predictable representations for Lévy processes. *Stochastic Processes and their Applications*, 90(1): 109 – 122, 2000.

[34] G. Peccati, J.L. Solé, M.S. Taqqu, and F. Utzet. Stein's method and normal approximation of poisson functionals. *Annals of Probability*, 38 (2):443–478, 2010.

[35] N. Privault. Chaotic and variational calculus in discrete and continuous time for the Poisson process. *Stoch. & Stoch. Rep.*, 51:83–109, 1994.

[36] N. Privault. *Stochastic analysis in discrete and continuous settings with normal martingales*, volume 1982 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009.

[37] M. Reitzner and M. Schulte. Central limit theorems for U-statistics of Poisson point processes. *The Annals of Probability*, 41(6):3879–3909, 2013.

[38] W. T. Rhee and M. Talagrand. Martingale inequalities and the jackknife estimate of variance. *Statist. Probab. Lett.*, 4(1):5–6, 1986.

[39] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 583–602, Berkeley, Calif., 1972. University of California Press.

[40] A. S. Üstünel and M. Zakai. The construction of filtrations on abstract Wiener space. *J. Funct. Anal.*, 143(1):10–32, 1997.

[41] A.S. Üstünel and M. Zakai. *Transformations of measure on Wiener spaces.* Monographs in Mathematics. Springer-Verlag, 2000.

[42] D. Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

[43] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probability Theory and Related Fields*, 118(3):427–438, 2000.

[44] K. Yosida. *Functional analysis.* Classics in Mathematics. Springer-Verlag, Berlin, 1995.