Normal approximation of Poisson functionals in Kolmogorov distance

Matthias Schulte*

Abstract

Peccati, Solè, Taqqu, and Utzet recently combined Stein's method and Malliavin calculus to obtain a bound for the Wasserstein distance of a Poisson functional and a Gaussian random variable. Convergence in the Wasserstein distance always implies convergence in the Kolmogorov distance at a possibly weaker rate. But there are many examples of central limit theorems having the same rate for both distances. The aim of this paper is to show this behaviour for a large class of Poisson functionals, namely so-called U-statistics of Poisson point processes. The technique used by Peccati et al. is modified to establish a similar bound for the Kolmogorov distance of a Poisson functional and a Gaussian random variable. This bound is evaluated for a U-statistic, and it is shown that the resulting expression is up to a constant the same as it is for the Wasserstein distance.

Key words: Central limit theorem, Malliavin calculus, Poisson point process, Stein's method, U-statistic, Wiener-Itô chaos expansion
MSC (2010): Primary: 60F05; 60H07 Secondary: 60G55

1 Introduction and results

Let η be a Poisson point process over a Borel space $(X, \mathcal{X})$ with a σ-finite intensity measure μ and let $F = F(\eta)$ be a random variable depending on the Poisson point process η. In the following, we call such random variables Poisson functionals. Moreover, we assume that F is square integrable (we write $F \in L^2(\mathbb{P}_\eta)$) and satisfies $\mathbb{E}F = 0$. By N we denote a standard Gaussian random variable. In [11], Peccati, Solé, Taqqu, and Utzet derived by a combination of Stein's method and Malliavin calculus the upper bound

$$d_W(F, N) \leq \mathbb{E}|1 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| + \int_X \mathbb{E}(D_z F)^2 \, |D_z L^{-1}F| \, \mu(\mathrm{d}z) \qquad (1)$$

for the Wasserstein distance of F and N. Here, $\langle \cdot, \cdot \rangle_{L^2(\mu)}$ stands for the inner product in $L^2(\mu)$, and the difference operator D and the inverse of the Ornstein-Uhlenbeck generator L are operators from Malliavin calculus. The underlying idea of these operators

---

*Karlsruhe Institute of Technology, Institute of Stochastics, D-76128 Karlsruhe, e-mail: matthias.schulte[at]kit.edu

is that each square integrable Poisson functional has a representation

$$F = \mathbb{E}F + \sum_{n=1}^{\infty} I_n(f_n),$$

where the f_n are square integrable functions supported on X^n, I_n stands for the n-th multiple Wiener-Itô integral, and the right-hand side converges in $L^2(\mathbb{P}_\eta)$. This decomposition is called Wiener-Itô chaos expansion, and the Malliavin operators of F are defined via their chaos expansions. The operators D_zF and $D_zL^{-1}F$ that occur in (1) are given by

$$D_zF = \sum_{n=1}^{\infty} n\, I_{n-1}(f_n(z, \cdot)) \ \text{ and } \ D_zL^{-1}F = -\sum_{n=1}^{\infty} I_{n-1}(f_n(z, \cdot)) \ \text{ for } z \in X.$$

Here, $f_n(z, \cdot)$ stands for the function on X^{n-1} we obtain by taking z as first argument. For exact definitions including the domains and more details on the Malliavin operators we refer to Section 2.

The Wasserstein distance between two random variables Y and Z is defined by

$$d_W(Y, Z) = \sup_{h \in \text{Lip}(1)} |\mathbb{E}h(Y) - \mathbb{E}h(Z)|,$$

where $\text{Lip}(1)$ is the set of all functions $h : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant less than or equal to one. Another commonly used distance for random variables is the Kolmogorov distance

$$d_K(Y, Z) = \sup_{s \in \mathbb{R}} |\mathbb{P}(Y \le s) - \mathbb{P}(Z \le s)|,$$

which is the supremum norm of the difference of the distribution functions of Y and Z. Because of this straightforward interpretation, one is often more interested in the Kolmogorov distance than in the Wasserstein distance. For the important case that Z is a standard Gaussian random variable N it is known (see [2, Theorem 3.1]) that

$$d_K(Y, N) \le 2\sqrt{d_W(Y, N)}. \tag{2}$$

This inequality gives us for the Kolmogorov distance a weaker rate of convergence than for the Wasserstein distance. But for many classical central limit theorems, one has actually the same rate of convergence for both metrics.

In order to overcome the problem that a detour around the Wasserstein distance and the inequality (2) often gives a suboptimal rate of convergence for the Kolmogorov distance, we derive a similar bound as (1) for the Kolmogorov distance by a modification of the proof in [11].

Theorem 1.1 *Let $F \in L^2(\mathbb{P}_\eta)$ with $\mathbb{E}F = 0$ be in the domain of D and let N be a*

standard Gaussian random variable. Then

$$
\begin{aligned}
d_K(F, N) \ \leq \ & \mathbb{E}|1 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| + 2\mathbb{E}\langle (DF)^2, |DL^{-1}F|\rangle_{L^2(\mu)} \qquad (3) \\
& + 2\mathbb{E}\langle (DF)^2, |F\ DL^{-1}F|\rangle_{L^2(\mu)} + 2\mathbb{E}\langle (DF)^2, |DF\ DL^{-1}F|\rangle_{L^2(\mu)} \\
& + \sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(F > s), DF\,|DL^{-1}F|\rangle_{L^2(\mu)} \\
\leq \ & \mathbb{E}|1 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| + 2c(F)\sqrt{\mathbb{E}\langle (DF)^2, (DL^{-1}F)^2\rangle_{L^2(\mu)}} \\
& + \sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(F > s), DF\,|DL^{-1}F|\rangle_{L^2(\mu)}
\end{aligned}
$$

with

$$
c(F) = \sqrt{\mathbb{E}\langle (DF)^2, (DF)^2\rangle_{L^2(\mu)}} + \left(\mathbb{E}\langle DF, DF\rangle^2_{L^2(\mu)} \right)^{\frac{1}{4}} \left(\left(\mathbb{E}F^4\right)^{\frac{1}{4}} + 1 \right).
$$

Comparing (1) and (3), one notes that both terms of the Wasserstein bound (1) also occur in (3), which means that the bound for the Kolmogorov distance is always larger.

We apply our Theorem 1.1 to two situations, where we obtain the same rate of convergence for the Kolmogorov distance and the Wasserstein distance. At first we derive the classical Berry-Esseen inequality with the optimal rate of convergence for the normal approximation of a classical Poisson random variable. As another application of Theorem 1.1, we consider so-called U-statistics of Poisson point processes, which are defined as

$$
F = \sum_{(x_1, \ldots, x_k) \in \eta^k_{\neq}} f(x_1, \ldots, x_k)
$$

where $k \in \mathbb{N}$, $f \in L^1(\mu^k)$, and $\eta^k_{\neq}$ is the set of all k-tuples of distinct points of η. In [4, 5] and [14], Lachièze-Rey and Peccati and Reitzner and Schulte used the bound (1) for the Wasserstein distance to derive central limit theorems with explicit rates of convergence for such Poisson functionals occurring in stochastic geometry and random graph theory. Now Theorem 1.1 allows us to replace the Wasserstein distance by the Kolmogorov distance without changing the rate of convergence, which means that the inequality (2) is not sharp for this class of Poisson functionals.

The main finding of this work, Theorem 1.1, is refined and proven in a different way in the the subsequent paper [3] by Eichelsbacher and Thäle. In [7], Last, Peccati, and Schulte further simplify the bounds for the normal approximation of Poisson functionals in Kolmogorov distance from Theorem 1.1 and [3] and apply them to several problems, where they provide new presumably optimal rates of convergence. Our result for the normal approximation of U-statistics of Poisson point processes (see Theorem 4.2) is used by Reitzner, Schulte, and Thäle in [15] and [16] to derive central limit theorems with rates of convergence for the Kolmogorov distance for the total edge length of the Gilbert graph and for distances between non-intersecting Poisson k-flats, respectively.

This paper is organized in the following way. Before we prove our main result Theorem 1.1 in Section 3, we introduce some facts from Malliavin calculus and Stein's method in Section 2. The applications of Theorem 1.1 are discussed in Section 4, and the result for U-statistics is shown in Section 5.

In this paper, we use the following notation. By $L^p(\mathbb{P}_\eta)$, $p > 0$, we denote the set of random variables Y depending on a Poisson point process η such that $\mathbb{E}|Y|^p < \infty$. Let $L^p(\mu^n)$, $p > 0$, be the set of functions $f : X^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ satisfying $\int_{X^n} |f|^p \, \mathrm{d}\mu^n = \int_{X^n} |f(x_1, \ldots, x_n)|^p \, \mu^n(\mathrm{d}(x_1, \ldots, x_n)) < \infty$ and let $||\cdot||_n$ and $\langle\cdot, \cdot\rangle_{L^2(\mu^n)}$ be the norm and the inner product in $L^2(\mu^n)$, respectively. By $L_s^p(\mu^n)$ we denote the set of all functions $f \in L^p(\mu^n)$ that are symmetric, i.e. invariant under permutations of their arguments.

2   Preliminaries

Malliavin calculus for Poisson functionals.   In the sequel, we briefly introduce three Malliavin operators and some properties of them that are necessary for the proofs in this paper. For more details on Malliavin calculus for Poisson functionals we refer to [8, 10, 11, 13] and the references therein.

By $I_n(\cdot)$, $n \geq 1$, we denote the n-th multiple Wiener-Itô integral, which is defined for all functions $f \in L_s^2(\mu^n)$ and satisfies $\mathbb{E}I_n(f) = 0$. The multiple Wiener-Itô integrals are orthogonal in the sense that

$$\mathbb{E}I_m(f)I_n(g) = \begin{cases} n! \, \langle f, g\rangle_{L^2(\mu^n)}, & m = n \\ 0, & m \neq n \end{cases} \tag{4}$$

for all $f \in L_s^2(\mu^m), g \in L_s^2(\mu^n), m, n \geq 1$. We use the convention $I_0(c) = c$ for $c \in \mathbb{R}$. It is known (see [8] for a proof) that every Poisson functional $F \in L^2(\mathbb{P}_\eta)$ has a unique so-called **Wiener-Itô chaos expansion**

$$F = \mathbb{E}F + \sum_{n=1}^{\infty} I_n(f_n) \tag{5}$$

with $f_n \in L_s^2(\mu^n)$, where the series converges in $L^2(\mathbb{P}_\eta)$. In the following, we call the functions f_n kernels of the Wiener-Itô chaos expansion of F and say that F has a chaos expansion of order k if $f_n = 0$ for all $n > k$. Combining (4) and (5), we obtain

$$\text{Var}\, F = \sum_{n=1}^{\infty} n! \, ||f_n||_n^2.$$

The representation (5) allows us to define the difference operator D, the Ornstein-Uhlenbeck generator L, and the Skorohod integral δ in the following way:

Definition 2.1 *Let $F \in L^2(\mathbb{P}_\eta)$ with the Wiener-Itô chaos expansion (5). If $\sum_{n=1}^{\infty} n \, n! \, ||f_n||_n^2 < \infty$, then the random function $z \mapsto D_z F$ defined by*

$$D_z F = \sum_{n=1}^{\infty} n \, I_{n-1}(f_n(z, \cdot))$$

is called the **difference operator** of F. For $\sum_{n=1}^{\infty} n^2 \, n! \, ||f_n||_n^2 < \infty$ the **Ornstein-Uhlenbeck generator** of F, denoted by LF, is given by

$$LF = -\sum_{n=1}^{\infty} n \, I_n(f_n).$$

Let $z \mapsto g(z)$ be a random function with a chaos expansion

$$g(z) = g_0(z) + \sum_{n=1}^{\infty} I_n(g_n(z, \cdot)), \ \ g_n(z, \cdot) \in L_s^2(\mu^n),$$

for μ-almost all $z \in X$ and $\sum_{n=0}^{\infty}(n+1)! \, ||g_n||_{n+1}^2 < \infty$. Then the **Skorohod integral** of g is the random variable $\delta(g)$ defined by

$$\delta(g) = \sum_{n=0}^{\infty} I_{n+1}(\tilde{g}_n),$$

where $\tilde{g}_n$ is the symmetrization $\tilde{g}_n(x_1, \ldots, x_{n+1}) = \frac{1}{(n+1)!} \sum_\sigma g_n(x_{\sigma(1)}, \ldots, x_{\sigma(n+1)})$ over all permutations σ of the $n+1$ variables.

We denote the domains of these operators by $\operatorname{dom} D$, $\operatorname{dom} L$, and $\operatorname{dom} \delta$. The difference operator also has the geometric interpretation

$$D_z F = F(\eta + \delta_z) - F(\eta) \tag{6}$$

a.s. for μ-almost all $z \in X$, where δ_z stands for the Dirac measure concentrated at the point $z \in X$, whence it is sometimes called add-one-cost operator (see Theorem 3.3 in [8]). If $F \notin \operatorname{dom} D$, we can define the difference operator by (6). If we iterate this definition and put $D_{x_1, \ldots, x_n}^n F = D_{x_n} D_{x_1, \ldots, x_{n-1}}^{n-1} F$, the kernels of the Wiener-Itô chaos expansion of F in (5) are given by the formula

$$f_n(x_1, \ldots, x_n) = \frac{1}{n!} \mathbb{E} D_{x_1, \ldots, x_n}^n F = \frac{1}{n!} \mathbb{E} \sum_{I \subset \{1, \ldots, n\}} (-1)^{n+|I|} F\left(\eta + \sum_{i \in I} \delta_{x_i}\right)$$

for $x_1, \ldots, x_n \in X$ (see Theorem 1.3 in [8]).

For centred random variables $F \in L^2(\mathbb{P}_\eta)$, i.e. $\mathbb{E} F = 0$, the inverse Ornstein-Uhlenbeck generator is given by

$$L^{-1} F = -\sum_{n=1}^{\infty} \frac{1}{n} I_n(f_n).$$

The following lemma summarizes how the operators from Malliavin calculus are related.

Lemma 2.2   a) For every $F \in \operatorname{dom} L$ it holds that $F \in \operatorname{dom} D$, $DF \in \operatorname{dom} \delta$, and

$$\delta DF = -LF. \tag{7}$$

b) *Let $F \in \operatorname{dom} D$ and $g \in \operatorname{dom} \delta$. Then*

$$\mathbb{E}\langle DF, g\rangle_{L^2(\mu)} = \mathbb{E}[F\,\delta(g)]. \tag{8}$$

For proofs we refer to [11] and [10], respectively. Equation (8) is sometimes called **integration by parts formula**. Because of this identity, one can see the difference operator and the Skorohod integral as dual operators.

For our applications in Section 4 we need a special integration by parts formula, where it is not required that the first Poisson functional is in $\operatorname{dom} D$. In this case, the difference operator is given by (6).

Lemma 2.3 *Let $F \in L^2(\mathbb{P}_\eta)$, $s \in \mathbb{R}$, and $g \in \operatorname{dom} \delta$ such that $g(z)$ has a Wiener-Itô chaos expansion of order k for μ-almost all $z \in X$. Moreover, assume that $D_z \mathbb{1}(F > s)\,g(z) \geq 0$ a.s. for μ-almost all $z \in X$. Then*

$$\mathbb{E}\langle D\mathbb{1}(F > s), g\rangle_{L^2(\mu)} = \mathbb{E}[\mathbb{1}(F > s)\,\delta(g)].$$

Proof: It is easy to see that $\mathbb{1}(F > s) \in L^2(\mathbb{P}_\eta)$, whence it has a Wiener-Itô chaos expansion

$$\mathbb{1}(F > s) = \sum_{n=0}^{\infty} I_n(h_n)$$

with $h_0 = \mathbb{E}\mathbb{1}(F > s)$ and kernels $h_n \in L_s^2(\mu^n)$, $n \geq 1$, that are given by

$$h_n(x_1, \ldots, x_n) = \frac{1}{n!}\,\mathbb{E}D^n_{x_1,\ldots,x_n}\mathbb{1}(F > s).$$

For a fixed $z \in X$ the expression $D_z \mathbb{1}(F > s)$ is bounded, and its chaos expansion has the kernels

$$\frac{1}{n!}\,\mathbb{E}D^n_{x_1,\ldots,x_n} D_z \mathbb{1}(F > s) = \frac{1}{n!}\,\mathbb{E}D^{n+1}_{z,x_1,\ldots,x_n}\mathbb{1}(F > s) = (n+1)\,h_{n+1}(z, x_1, \ldots, x_n).$$

Hence, we obtain the representation

$$D_z \mathbb{1}(F > s) = \sum_{n=1}^{\infty} n\,I_{n-1}(h_n(z, \cdot))$$

for all $z \in X$. From Fubini's theorem and the orthogonality of the multiple Wiener-Itô integrals it follows that

$$\mathbb{E}\langle D\mathbb{1}(F > s), g\rangle_{L^2(\mu)}$$

$$= \int_X \mathbb{E}\left[D_z \mathbb{1}(F > s)\,g(z)\right] \mu(\mathrm{d}z)$$

$$= \int_X \mathbb{E}\left[\sum_{n=1}^{\infty} n\,I_{n-1}(h_n(z, \cdot)) \sum_{n=0}^{k} I_n(g_n(z, \cdot))\right] \mu(\mathrm{d}z)$$

$$= \int_X \sum_{n=1}^{k+1} n! \int_{X^{n-1}} h_n(z, x_1, \ldots, x_{n-1})\,g_{n-1}(z, x_1, \ldots, x_{n-1})\,\mu^{n-1}(\mathrm{d}(x_1, \ldots, x_{n-1}))\,\mu(\mathrm{d}z).$$

$$\tag{9}$$

On the other hand, we have

$$\mathbb{E}\left[\mathbb{1}(F > s)\,\delta(g)\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} I_n(h_n) \sum_{n=0}^{k} I_{n+1}(\tilde{g}_n)\right]$$

$$= \sum_{n=1}^{k+1} n! \int_{X^n} h_n(x_1,\ldots,x_n)\,\tilde{g}_{n-1}(x_1,\ldots,x_n)\,\mu^n(\mathrm{d}(x_1,\ldots,x_n)) \quad (10)$$

$$= \sum_{n=1}^{k+1} n! \int_{X^n} h_n(x_1,\ldots,x_n)\,g_{n-1}(x_1,\ldots,x_n)\,\mu^n(\mathrm{d}(x_1,\ldots,x_n)).$$

Here, we use the symmetry of h_n in the last step. Comparing (9) and (10) concludes the proof. $\qquad\square$

The next lemma provides an upper bound for the second moment of a Skorohod integral, which is used in Section 5.

Lemma 2.4 *Let $f \in L^2(\mu^{k+1})$ be symmetric in its last k arguments and let $g(z) = I_k(f(z,\cdot))$. Then*

$$\mathbb{E}\left[\delta(g)^2\right] \leq (k+1)\,\mathbb{E}\int_X I_k(f(z,\cdot))^2\,\mu(\mathrm{d}z).$$

Proof: By the definition of δ, we obtain $\delta(g) = I_{k+1}(\tilde{f})$ with the symmetrization

$$\tilde{f}(x_1,\ldots,x_{k+1}) = \frac{1}{(k+1)!}\sum_{\sigma} f(x_{\sigma(1)},\ldots,x_{\sigma(k+1)})$$

as above. From the Cauchy-Schwarz inequality, it follows that $||\tilde{f}||_{k+1}^2 \leq ||f||_{k+1}^2$. Combining this with Fubini's Theorem, we have

$$\mathbb{E}\left[\delta(g)^2\right] = (k+1)!\,||\tilde{f}||_{k+1}^2 \leq (k+1)!\,||f||_{k+1}^2 = (k+1)\,\mathbb{E}\int_X I_k(f(z,\cdot))^2\,\mu(\mathrm{d}z),$$

which concludes the proof. $\qquad\square$

Stein's method. Besides Malliavin calculus our proof of Theorem 1.1 rests upon Stein's method, which goes back to Charles Stein [17, 18] and is a powerful tool for proving limit theorems. For a detailed and more general introduction into this topic we refer to [1, 2, 18]. Very fundamental for this approach is the following lemma (see Chapter II in [18]):

Lemma 2.5 *For $s \in \mathbb{R}$ the function*

$$g_s(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} \left(\mathbb{1}(u \in (-\infty, s]) - \mathbb{P}(N \leq s)\right) e^{-\frac{u^2}{2}}\,\mathrm{d}u \quad (11)$$

is a solution of the differential equation

$$g_s'(w) - w\,g_s(w) = \mathbb{1}(w \in (-\infty, s]) - \mathbb{P}(N \leq s) \quad (12)$$

7

and satisfies

$$0 < g_s(w) \le \frac{\sqrt{2\pi}}{4}, \quad |g'_s(w)| \le 1, \quad \text{and} \quad |w \, g_s(w)| \le 1 \tag{13}$$

for any $w \in \mathbb{R}$.

Equation (12) is usually called **Stein's equation**. The function g_s is infinitely differentiable on $\mathbb{R} \setminus \{s\}$, but it is not differentiable in s. We denote the left-sided and right-sided limits of the derivatives in s by $g_s^{(m)}(s-)$ and $g_s^{(m)}(s+)$, respectively. For the first derivative, a direct computation proves

$$g'_s(s+) = -1 + g'_s(s-), \tag{14}$$

and we define $g'_s(s) := g'_s(s-)$.

By replacing w by a random variable Z and taking the expectation in (12), one obtains

$$\mathbb{E}[g'_s(Z) - Z \, g_s(Z)] = \mathbb{P}(Z \le s) - \mathbb{P}(N \le s)$$

and as a consequence of the definition of the Kolmogorov distance

$$d_K(Z, N) = \sup_{s \in \mathbb{R}} |\mathbb{E}[g'_s(Z) - Z \, g_s(Z)]|. \tag{15}$$

The identity (15) will be our starting point in Section 3. Note furthermore, that we obtain, by combining (12) and (13), the upper bound

$$|g''_s(w)| \le \frac{\sqrt{2\pi}}{4} + |w| \tag{16}$$

for $w \in \mathbb{R} \setminus \{s\}$.

3   Proof of Theorem 1.1

By a combination of Malliavin calculus and Stein's method similar to that in [11], we derive the upper bound for the Kolmogorov distance.

Proof of Theorem 1.1:   Using the identity (7) and the integration by parts formula (8), we obtain

$$\mathbb{E}[F \, g_s(F)] = \mathbb{E}[LL^{-1}F \, g_s(F)] = \mathbb{E}[\delta(-DL^{-1}F) \, g_s(F)] = \mathbb{E}\langle -DL^{-1}F, Dg_s(F) \rangle_{L^2(\mu)}. \tag{17}$$

In order to compute $D_z g_s(F)$, we fix $z \in X$ and consider the following cases:

1. $F, F + D_z F \le s$ or $F, F + D_z F > s$;

2. $F \le s < F + D_z F$;

3. $F + D_z F \le s < F$.

8

For $F, F + D_zF \le s$ or $F, F + D_zF > s$, it follows from (6) and Taylor expansion that

$$
\begin{aligned}
D_zg_s(F) &= g_s(F + D_zF) - g_s(F) = g_s'(F)D_zF + \frac{1}{2}g_s''(\tilde{F})(D_zF)^2 \\
&=: g_s'(F)D_zF + r_1(F, z, s),
\end{aligned}
$$

where $\tilde{F}$ is between F and $F + D_zF$. For $F \le s < F + D_zF$, we obtain by (6), Taylor expansion, and (14)

$$
\begin{aligned}
D_zg_s(F) &= g_s(F + D_zF) - g_s(F) = g_s(F + D_zF) - g_s(s) + g_s(s) - g_s(F) \\
&= g_s'(s+)(F + D_zF - s) + \frac{1}{2}g_s''(\tilde{F}_1)(F + D_zF - s)^2 \\
&\quad + g_s'(F)(s - F) + \frac{1}{2}g_s''(\tilde{F}_2)(s - F)^2 \\
&= g_s'(F)D_zF + (g_s'(s-) - 1 - g_s'(F))(F + D_zF - s) \\
&\quad + \frac{1}{2}g_s''(\tilde{F}_1)(F + D_zF - s)^2 + \frac{1}{2}g_s''(\tilde{F}_2)(s - F)^2 \\
&= g_s'(F)D_zF - (F + D_zF - s) + g_s''(\tilde{F}_0)(s - F)(F + D_zF - s) \\
&\quad + \frac{1}{2}g_s''(\tilde{F}_1)(F + D_zF - s)^2 + \frac{1}{2}g_s''(\tilde{F}_2)(s - F)^2 \\
&=: g_s'(F)D_zF - (F + D_zF - s) + r_2(F, z, s)
\end{aligned}
$$

with $\tilde{F}_0, \tilde{F}_1, \tilde{F}_2 \in (F, F + D_zF)$. For $F + D_zF \le s < F$, we have analogously

$$
\begin{aligned}
D_zg_s(F) &= g_s(F + D_zF) - g_s(F) = g_s(F + D_zF) - g_s(s) + g_s(s) - g_s(F) \\
&= g_s'(s-)(F + D_zF - s) + \frac{1}{2}g_s''(\tilde{F}_1)(F + D_zF - s)^2 \\
&\quad + g_s'(F)(s - F) + \frac{1}{2}g_s''(\tilde{F}_2)(s - F)^2 \\
&= g_s'(F)D_zF + (g_s'(s+) + 1 - g_s'(F))(F + D_zF - s) \\
&\quad + \frac{1}{2}g_s''(\tilde{F}_1)(F + D_zF - s)^2 + \frac{1}{2}g_s''(\tilde{F}_2)(s - F)^2 \\
&= g_s'(F)D_zF + (F + D_zF - s) + g_s''(\tilde{F}_0)(s - F)(F + D_zF - s) \\
&\quad + \frac{1}{2}g_s''(\tilde{F}_1)(F + D_zF - s)^2 + \frac{1}{2}g_s''(\tilde{F}_2)(s - F)^2 \\
&=: g_s'(F)D_zF + (F + D_zF - s) + r_2(F, z, s)
\end{aligned}
$$

with $\tilde{F}_0, \tilde{F}_1, \tilde{F}_2 \in (F + D_zF, F)$. Thus, $D_zg_s(F)$ has a representation

$$
D_zg_s(F) = g_s'(F)D_zF + R(F, z, s), \tag{18}
$$

where $R(F, z, s)$ is given by

$$
\begin{aligned}
&R(F, z, s) \\
&= \left(\mathbb{1}(F, F + D_zF \le s) + \mathbb{1}(F, F + D_zF > s)\right) r_1(F, z, s) \\
&\quad + \left(\mathbb{1}(F \le s < F + D_zF) + \mathbb{1}(F + D_zF \le s < F)\right)\left(r_2(F, z, s) - |F + D_zF - s|\right).
\end{aligned}
$$

Combining (17) and (18) yields

$$\mathbb{E}\left[g_s'(F) - F g_s(F)\right] = \mathbb{E}\left[g_s'(F) - \langle g_s'(F) DF + R(F, \cdot, s), -DL^{-1}F\rangle_{L^2(\mu)}\right].$$

Thus, the triangle inequality and $|g_s'(F)| \leq 1$ lead to

$$
\begin{aligned}
|\mathbb{E}\left[g_s'(F) - F g_s(F)\right]| &\leq |\mathbb{E}\left[g_s'(F)\left(1 - \langle DF, -DL^{-1}F\rangle_{L^2(\mu)}\right)\right]| \qquad (19) \\
&\quad + |\mathbb{E}\langle R(F, \cdot, s), DL^{-1}F\rangle_{L^2(\mu)}| \\
&\leq \mathbb{E}|1 - \langle DF, -DL^{-1}F\rangle_{L^2(\mu)}| + \mathbb{E}\langle |R(F, \cdot, s)|, |DL^{-1}F|\rangle_{L^2(\mu)}.
\end{aligned}
$$

In $r_2(F, z, s)$, we assume that s is between F and $F + D_z F$ so that

$$|F + D_z F - s| \leq |D_z F| \quad \text{and} \quad |F - s| \leq |D_z F|.$$

The inequality (16) and the fact that $\tilde{F}_i$ is between F and $F + D_z F$ allow us to bound all second derivatives in $R(F, z, s)$ by

$$|g_s''(\tilde{F}_i)| \leq \frac{\sqrt{2\pi}}{4} + |F| + |D_z F|.$$

Now it is easy to see that

$$|R(F, z, s)|$$

$$
\begin{aligned}
&\leq (\mathbb{1}(F, F + D_z F \leq s) + \mathbb{1}(F, F + D_z F > s)) \frac{1}{2}\left(\frac{\sqrt{2\pi}}{4} + |F| + |D_z F|\right)(D_z F)^2 \\
&\quad + (\mathbb{1}(F \leq s < F + D_z F) + \mathbb{1}(F + D_z F \leq s < F))|D_z F| \\
&\quad + (\mathbb{1}(F \leq s < F + D_z F) + \mathbb{1}(F + D_z F \leq s < F))2\left(\frac{\sqrt{2\pi}}{4} + |F| + |D_z F|\right)(D_z F)^2 \\
&\leq 2\left(\frac{\sqrt{2\pi}}{4} + |F| + |D_z F|\right)(D_z F)^2 \\
&\quad + (\mathbb{1}(F \leq s < F + D_z F) + \mathbb{1}(F + D_z F \leq s < F))|D_z F|.
\end{aligned}
$$

By (6), the last summand can be rewritten as

$$(\mathbb{1}(F \leq s < F + D_z F) + \mathbb{1}(F + D_z F \leq s < F))|D_z F| = D_z \mathbb{1}(F > s) D_z F.$$

Hence, it follows directly that

$$
\begin{aligned}
&\mathbb{E}\langle |R(F, \cdot, s)|, |DL^{-1}F|\rangle_{L^2(\mu)} \\
&\leq 2\mathbb{E}\langle (DF)^2, |DL^{-1}F|\rangle_{L^2(\mu)} + 2\mathbb{E}\langle (DF)^2, |F\, DL^{-1}F|\rangle_{L^2(\mu)} \\
&\quad + 2\mathbb{E}\langle (DF)^2, |DF\, DL^{-1}F|\rangle_{L^2(\mu)} + \mathbb{E}\langle D\mathbb{1}(F > s)\, DF, |DL^{-1}F|\rangle_{L^2(\mu)}.
\end{aligned}
$$

Putting this in (19) concludes the proof of the first inequality in (3). The second bound in (3) is a direct consequence of the Cauchy-Schwarz inequality. $\qquad \square$

In [11], the right-hand side of (17) is evaluated for twice differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $\sup_{x \in \mathbb{R}} |f'(x)| \leq 1$ and $\sup_{x \in \mathbb{R}} |f''(x)| \leq 2$ (for the Wasserstein distance

the solutions of Stein's equation must have these properties) instead of the functions g_s as defined in (11). For such a function f it holds that

$$D_z f(F) = f'(F) D_z F + \tilde{r}(F)$$

with $|\tilde{r}(F)| \leq (D_z F)^2$. Since this representation is easier than the representation we obtain for $D_z g_s(F)$, the bound for the Wasserstein distance in (1) is shorter and easier to evaluate than the bound for the Kolmogorov distance in (3).

4   Applications of Theorem 1.1

Normal approximation of a Poisson random variable.   As a first application of Theorem 1.1, we compute an upper bound for the Kolmogorov distance between a Poisson random variable Y with parameter $t > 0$ and a normal distribution. In Example 3.5 in [11], the bound (1) is used to compute a bound for the Wasserstein distance, and the known optimal rate of convergence $t^{-\frac{1}{2}}$ is obtained.

Y has the same distribution as $F_t = \sum_{x \in \eta_t} 1$, where η_t is a Poisson point process on $[0,1]$ with t times the the restriction of the Lebesgue measure on $[0,1]$ as intensity measure μ_t. In the following, we denote by $I_{n,t}(\cdot)$ the n-th multiple Wiener-Itô integral with respect to η_t. The representation

$$I_{1,t}(f) = \sum_{x \in \eta_t} f(x) - \int_X f(x)\, \mu_t(\mathrm{d}x)$$

for a Wiener-Itô integral of a function $f \in L^1(\mu_t) \cap L^2(\mu_t)$ and the fact that

$$F_t = t \int_0^1 1\,\mathrm{d}x + \sum_{x \in \eta_t} 1 - t \int_0^1 1\,\mathrm{d}x$$

imply that F_t has the Wiener-Itô chaos expansion $F_t = \mathbb{E}F_t + I_{1,t}(f_1) = t + I_{1,t}(1)$. Hence, the standardized random variable

$$G_t = \frac{F_t - \mathbb{E}F_t}{\sqrt{\mathrm{Var}\,F_t}} = \frac{F_t - t}{\sqrt{t}}$$

has the chaos expansion $G_t = I_{1,t}(1)/\sqrt{t}$ and $D_z G_t = -D_z L^{-1} G_t = 1/\sqrt{t}$ for $z \in [0,1]$. It is easy to see that

$$\mathbb{E}|1 - \langle DG_t, -DL^{-1}G_t \rangle_{L^2(\mu_t)}| = |1 - \frac{1}{t}\langle 1, 1 \rangle_{L^2(\mu_t)}| = |1 - \frac{t}{t}| = 0.$$

We obtain

$$\mathbb{E}\langle (DG_t)^2, (DL^{-1}G_t)^2 \rangle_{L^2(\mu_t)} = \mathbb{E}\langle (DG_t)^2, (DG_t)^2 \rangle_{L^2(\mu_t)} = \frac{1}{t},$$

$\mathbb{E}\langle DG_t, DG_t \rangle^2_{L^2(\mu_t)} = 1$, and $\mathbb{E}G_t^4 = 3 + 1/t$ by analogous computations. Since $D_z \mathbb{1}(G_t > s)\, D_z G_t\, |D_z L^{-1} G_t| \geq 0$ for $s \in \mathbb{R}$ and $z \in [0,1]$ and $D_z G_t\, |D_z L^{-1} G_t| = 1/t$ for $z \in [0,1]$,

it follows from Lemma 2.3 and the Cauchy-Schwarz inequality that

$$\sup_{s \in \mathbb{R}} \mathbb{E} \langle D \mathbb{1}(G_t > s), DG_t \, |DL^{-1}G_t| \rangle_{L^2(\mu_t)}$$

$$= \sup_{s \in \mathbb{R}} \mathbb{E}[\mathbb{1}(G_t > s) \, \delta(DG_t \, |DL^{-1}G_t|)] \leq \mathbb{E}[\delta(DG_t \, |DL^{-1}G_t|)^2]^{\frac{1}{2}} = \frac{1}{t} \mathbb{E}[I_{1,t}(1)^2]^{\frac{1}{2}} = \frac{1}{\sqrt{t}}.$$

Now Theorem 1.1 yields

$$d_K \left(\frac{Y - t}{\sqrt{t}}, N \right) \leq 2 \left(\frac{1}{\sqrt{t}} + \left(3 + \frac{1}{t} \right)^{\frac{1}{4}} + 1 \right) \frac{1}{\sqrt{t}} + \frac{1}{\sqrt{t}} \leq \frac{8}{\sqrt{t}}$$

for $t \geq 1$, which is the classical Berry-Esseen inequality with the optimal rate of convergence (up to a constant).

Normal approximation of U-statistics of Poisson point processes. As a second application of Theorem 1.1, we discuss U-statistics of the form

$$F = \sum_{(x_1, \ldots, x_k) \in \eta_{\neq}^k} f(x_1, \ldots, x_k)$$

with $k \in \mathbb{N}$ and $f \in L_s^1(\mu^k)$. Here, $\eta_{\neq}^k$ stands for the set of all k-tuples of distinct points of η. If the intensity measure μ is non-atomic, η has no multiple points, and $\eta_{\neq}^k$ can be written as

$$\eta_{\neq}^k = \left\{ (x_1, \ldots, x_k) \in \eta^k : x_i \neq x_j \, \forall \, i \neq j \right\}.$$

In case that μ has atoms, one has to take into account that distinct points of η can have the same location. We denote k as the order of the U-statistic F. From now on, we always assume that $F \in L^2(\mathbb{P}_\eta)$ and $\operatorname{Var} F > 0$. In [14], the chaos expansions of such Poisson functionals are investigated, and the bound (1) is used to prove a central limit theorem with a rate of convergence for the Wasserstein distance. From there it is known that the kernels of the chaos expansion of a U-statistic F are

$$f_i(x_1, \ldots, x_i) = \binom{k}{i} \int_{X^{k-i}} f(x_1, \ldots, x_i, y_1, \ldots, y_{k-i}) \, \mu^{k-i}(\mathrm{d}(y_1, \ldots, y_{k-i})) \tag{20}$$

for $i = 1, \ldots, k$ and $f_i = 0$ for $i > k$. An application of the bound (1) to such Poisson functionals yields (see Theorem 4.1 in [14])

$$d_W \left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}, N \right) \leq k \sum_{i,j=1}^k \frac{\sqrt{R_{ij}}}{\operatorname{Var} F} + k^{\frac{7}{2}} \sum_{i=1}^k \frac{\sqrt{\tilde{R}_i}}{\operatorname{Var} F}, \tag{21}$$

where R_{ij} and $\tilde{R}_i$ are given by

$$R_{ij} = \mathbb{E} \left(\int_X I_{i-1} \left(f_i(z, \cdot) \right) I_{j-1} \left(f_j(z, \cdot) \right) \mu(\mathrm{d}z) \right)^2$$

$$\quad - \left(\mathbb{E} \int_X I_{i-1} \left(f_i(z, \cdot) \right) I_{j-1} \left(f_j(z, \cdot) \right) \mu(\mathrm{d}z) \right)^2$$

$$\tilde{R}_i = \mathbb{E} \int_X I_{i-1} \left(f_i(z, \cdot) \right)^4 \mu(\mathrm{d}z)$$

for $i, j = 1, \ldots, k$. In [14], the right-hand side of (21) is bounded by a sum of deterministic integrals depending on f. Due to technical reasons it is assumed that the U-statistic F is absolutely convergent, which means that the U-statistic $\overline{F}$ given by

$$\overline{F} = \sum_{(x_1, \ldots, x_k) \in \eta_{\neq}^k} |f(x_1, \ldots, x_k)|$$

is in $L^2(\mathbb{P}_\eta)$. The U-statistic $\overline{F}$ has a finite Wiener-Itô chaos expansion with kernels

$$\overline{f}_i(x_1, \ldots, x_i) = \binom{k}{i} \int_{X^{k-i}} |f(x_1, \ldots, x_i, y_1, \ldots, y_{k-i})| \, \mu^{k-i}(\mathrm{d}(y_1, \ldots, y_{k-i}))$$

for $i = 1, \ldots, k$ and $\overline{f}_i = 0$ for $i > k$. In order to bound the right-hand side of (21) by a sum of deterministic integrals, we use the following notation. For $i, j = 1, \ldots, k$ let $\overline{\Pi}_{\geq 2}(i, i, j, j)$ be the set of all partitions π of

$$x_1^{(1)}, \ldots, x_i^{(1)}, x_1^{(2)}, \ldots, x_i^{(2)}, x_1^{(3)}, \ldots, x_j^{(3)}, x_1^{(4)}, \ldots, x_j^{(4)}$$

such that

- two variables with the same upper index are in different blocks of π;

- each block of π includes at least two variables;

- there are no sets $A_1, A_2 \subset \{1, 2, 3, 4\}$ with $A_1 \cup A_2 = \{1, 2, 3, 4\}$ and $A_1 \cap A_2 = \emptyset$ such that each block of π either consists of variables with upper index in A_1 or of variables with upper index in A_2.

Let $|\pi|$ stand for the number of blocks of a partition π. For functions $g_1, g_2 : X^i \to \mathbb{R}$ and $g_3, g_4 : X^j \to \overline{\mathbb{R}}$ the tensor product $g_1 \otimes g_2 \otimes g_3 \otimes g_4 : X^{2i+2j} \to \mathbb{R}$ is given by

$$\begin{aligned}
&(g_1 \otimes g_2 \otimes g_3 \otimes g_4)(x_1^{(1)}, \ldots, x_j^{(4)}) \\
&= g_1(x_1^{(1)}, \ldots, x_i^{(1)}) \, g_2(x_1^{(2)}, \ldots, x_i^{(2)}) \, g_3(x_1^{(3)}, \ldots, x_j^{(3)}) \, g_4(x_1^{(4)}, \ldots, x_j^{(4)}).
\end{aligned}$$

For $\pi \in \overline{\Pi}_{\geq 2}(i, i, j, j)$ we define the function $(g_1 \otimes g_2 \otimes g_3 \otimes g_4)_\pi : X^{|\pi|} \to \mathbb{R}$ by replacing all variables that are in the same block of π by a new common variable. Since we later integrate over all these new variables, their order does not matter. Using this notation, we define

$$M_{ij}(f) = \sum_{\pi \in \overline{\Pi}_{\geq 2}(i, i, j, j)} \int_{X^{|\pi|}} (\overline{f}_i \otimes \overline{f}_i \otimes \overline{f}_j \otimes \overline{f}_j)_\pi \, \mathrm{d}\mu^{|\pi|}$$

for $i, j = 1, \ldots, k$. Now we can state the following upper bound for the Wasserstein distance (see Theorem 4.7 in [14]):

Proposition 4.1 *Let $F \in L^2(\mathbb{P}_\eta)$ be an absolutely convergent U-statistic of order k with $\operatorname{Var} F > 0$ and let N be a standard Gaussian random variable. Then*

$$d_W\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}, N\right) \leq 2k^{\frac{7}{2}} \sum_{1 \leq i \leq j \leq k} \frac{\sqrt{M_{ij}(f)}}{\operatorname{Var} F}. \tag{22}$$

In this situation, we can use Theorem 1.1 to prove the following bound analogous to (22) for the Kolmogorov distance between a standardized U-statistic and a standard Gaussian random variable:

Theorem 4.2 *Let $F \in L^2(\mathbb{P}_\eta)$ be an absolutely convergent U-statistic of order k with $\mathrm{Var}\, F > 0$ and let N be a standard Gaussian random variable. Then*

$$d_K \left(\frac{F - \mathbb{E}F}{\sqrt{\mathrm{Var}\, F}}, N \right) \leq 19k^5 \sum_{i,j=1}^k \frac{\sqrt{M_{ij}(f)}}{\mathrm{Var}\, F}. \qquad (23)$$

Before we prove this theorem in Section 5, we discuss some of its consequences. Let us consider a family of Poisson point processes η_t with σ-finite intensity measures μ_t and U-statistics

$$F_t = \sum_{(x_1,\ldots,x_k) \in (\eta_t)^k_{\neq}} f_t(x_1,\ldots,x_k)$$

with $f_t \in L^1_s(\mu^k_t)$ and $F_t \in L^2(\mathbb{P}_{\eta_t})$ such that

$$\frac{\sqrt{M_{ij}(f_t)}}{\mathrm{Var}\, F_t} \to 0 \quad \text{as} \quad t \to \infty \quad \text{for all} \quad i,j = 1,\ldots,k.$$

Here, we integrate with respect to μ_t in $M_{ij}(f_t)$. Comparing the right-hand sides in (22) and (23) for the U-statistics F_t, we see that the bounds for the Wasserstein and Kolmogorov distance have the same rates of convergence and differ only by constants. An important special case of the described setting is that the Poisson point process depends on a real valued intensity parameter. The following corollary deals with this situation and is the counterpart of Theorem 5.2 in [14] for the Kolmogorov distance.

Corollary 4.3 *Let η_t be a Poisson point process with intensity measure $\mu_t = t\mu$ with $t \geq 1$ and a fixed σ-finite measure μ and let N be a standard Gaussian random variable. We consider U-statistics $F_t \in L^2(\mathbb{P}_{\eta_t})$ of the form*

$$F_t = g(t) \sum_{(x_1,\ldots,x_k) \in (\eta_t)^k_{\neq}} f(x_1,\ldots,x_k)$$

with $g : (0,\infty) \to (0,\infty)$ and $f \in L^1_s(\mu^k)$ independent of t. Moreover, we assume that

$$\int_X \left(\int_{X^{k-1}} f(x,y_1,\ldots,y_{k-1}) \, \mu^{k-1}(\mathrm{d}(y_1,\ldots,y_{k-1})) \right)^2 \mu(\mathrm{d}x) > 0$$

and $M_{ij}(f) < \infty$ for $i,j = 1,\ldots,k$. Then there is a constant $C > 0$ such that

$$d_K \left(\frac{F_t - \mathbb{E}F_t}{\sqrt{\mathrm{Var}\, F_t}}, N \right) \leq C\, t^{-\frac{1}{2}}$$

for all $t \geq 1$.

This corollary follows from bounding $\sqrt{M_{ij}(f_t)}/\operatorname{Var} F_t$ in the same way as in the proof of [14, Theorem 5.2].

In [4], a so-called fourth moment condition for Poisson functionals with positive variance, finite Wiener-Itô chaos expansion, and non-negative kernels satisfying some integrability conditions is derived. More precisely, for such Poisson functionals it is proven that

$$d_W\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}, N\right) \le C_{W,k}\sqrt{\mathbb{E}\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}\right)^4 - 3}$$

with a constant $C_{W,k} > 0$ only depending on k. U-statistics $F \in L^2(\mathbb{P}_\eta)$ of the form

$$F = \sum_{(x_1,\ldots,x_k)\in\eta_{\neq}^k} f(x_1,\ldots,x_k) \quad \text{with } f \in L_s^1(\mu^k) \text{ and } f \ge 0$$

such that $\operatorname{Var} F > 0$ and $M_{ij}(f) < \infty$ for $i, j = 1, \ldots, k$ belong to this class and satisfy

$$\frac{M_{ij}(f)}{(\operatorname{Var} F)^2} \le \mathbb{E}\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}\right)^4 - 3.$$

Then (23) can be modified to

$$d_K\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}, N\right) \le C_k\sqrt{\mathbb{E}\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}\right)^4 - 3}$$

with a constant $C_k > 0$ only depending on k.

5 Proof of Theorem 4.2

In our proof of Theorem 4.2, we make use of the following property of U-statistics:

Lemma 5.1 *For a U-statistic $F \in L^2(\mathbb{P}_\eta)$ of order k the inverse of the Ornstein-Uhlenbeck generator has a representation*

$$\begin{aligned}
&- L^{-1}(F - \mathbb{E}F) \\
&= \sum_{m=1}^k \frac{1}{m} \sum_{(x_1,\ldots,x_m)\in\eta_{\neq}^m} \int_{X^{k-m}} f(x_1,\ldots,x_m,y_1,\ldots,y_{k-m})\,\mu^{k-m}(\mathrm{d}(y_1,\ldots,y_{k-m})) \\
&\quad - \sum_{m=1}^k \frac{1}{m} \int_{X^k} f(y_1,\ldots,y_k)\,\mu^k(\mathrm{d}(y_1,\ldots,y_k)).
\end{aligned} \tag{24}$$

Proof: We define $\widehat{f_i} : X^i \to \overline{\mathbb{R}}$ by $\widehat{f_i}(x_1,\ldots,x_i) = \binom{k}{i}^{-1} f_i(x_1,\ldots,x_i)$ for $i = 1,\ldots,k$. Using this notation and formula (20) for the kernels of a U-statistic, we obtain the chaos expansion

$$\begin{aligned}
&\sum_{(x_1,\ldots,x_m)\in\eta_{\neq}^m} \int_{X^{k-m}} f(x_1,\ldots,x_m,y_1,\ldots,y_{k-m})\,\mu^{k-m}(\mathrm{d}(y_1,\ldots,y_{k-m})) \\
&= \int_{X^k} f(y_1,\ldots,y_k)\,\mu^k(\mathrm{d}(y_1,\ldots,y_k)) + \sum_{i=1}^m \binom{m}{i} I_i(\widehat{f_i})
\end{aligned}$$

15

for $m = 1, \ldots, k$. Combining this with an identity for binomial coefficients, we see that the right-hand side in (24) equals

$$\sum_{m=1}^{k} \frac{1}{m} \sum_{i=1}^{m} \binom{m}{i} I_i(\widehat{f_i}) = \sum_{m=1}^{k} \sum_{i=1}^{k} \frac{1}{m} \binom{m}{i} I_i(\widehat{f_i}) = \sum_{i=1}^{k} \sum_{m=1}^{k} \frac{1}{m} \binom{m}{i} I_i(\widehat{f_i})$$

$$= \sum_{i=1}^{k} \frac{1}{i} \binom{k}{i} I_i(\widehat{f_i}) = \sum_{i=1}^{k} \frac{1}{i} I_i(f_i),$$

which is the chaos expansion of $-L^{-1}(F - \mathbb{E}F)$ by definition. $\square$

In order to deal with expressions as R_{ij} and $\tilde{R}_i$ in the previous section, one needs to compute the expectation of products of multiple Wiener-Itô integrals. This can be done by using Proposition 6.1 in [6] (see also [19, Theorem 3.1], [13, Proposition 4.5.6], [12, Proposition 6.5.1], or [9, Theorem 3.1]). This so-called product formula gives us the Wiener-Itô chaos expansion of a product of two multiple Wiener-Itô integrals. Together with (4), one obtains that the expectation of a product of four multiple Wiener-Itô integrals is a sum of deterministic integrals depending on the integrands of the stochastic integrals and partitions of their variables as used for the definition of $M_{ij}(f)$.

By using this product formula, one can prove in a similar way as in [14, Subsection 4.2] that

$$R_{ij} \leq M_{ij}(f) \quad \text{and} \quad \int_X \mathbb{E} I_{i-1}(f_i(z, \cdot))^2 I_{j-1}(f_j(z, \cdot))^2 \, \mu(\mathrm{d}z) \leq M_{ij}(f) \qquad (25)$$

for $i, j = 1, \ldots, k$. Moreover, we prepare the proof of Theorem 4.2 by the following two lemmas:

Lemma 5.2 *Let $F \in L^2(\mathbb{P}_\eta)$ be an absolutely convergent U-statistic with $M_{ij}(f) < \infty$ for $i, j = 1, \ldots, k$. Then*

$$\mathbb{E}(F - \mathbb{E}F)^4 \leq k^2 \sum_{i,j=1}^{k} M_{ij}(f) + 3k^2 (\mathrm{Var}\, F)^2.$$

Proof: Using the Wiener-Itô chaos expansion of F and the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}(F - \mathbb{E}F)^4 = \mathbb{E}\left(\sum_{i=1}^{k} I_i(f_i)\right)^2 \left(\sum_{j=1}^{k} I_j(f_j)\right)^2 \leq k^2 \, \mathbb{E} \sum_{i=1}^{k} I_i(f_i)^2 \sum_{j=1}^{k} I_j(f_j)^2.$$

Now the previously mentioned product formula for multiple Wiener-Itô integrals and $\mathrm{Var}\, F = \sum_{n=1}^{k} n! \, \|f_n\|_n^2$ yield that

$$\mathbb{E} I_i(f_i)^2 I_j(f_j)^2 \leq M_{ij}(f) + 3 \, i! \, \|f_i\|_i^2 \, j! \, \|f_j\|_j^2 \leq M_{ij}(f) + 3(\mathrm{Var}\, F)^2.$$

In the first inequality, we have equality for $i = j$ and $f \geq 0$. $\square$

Lemma 5.3 *Let $F \in L^2(\mathbb{P}_\eta)$ be an absolutely convergent U-statistic with $M_{ij}(f) < \infty$ for $i, j = 1, \ldots, k$. Then*

$$\sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(F > s), DF \, |DL^{-1}(F - \mathbb{E}F)| \rangle_{L^2(\mu)}$$

$$\leq \sqrt{(2k-1)\mathbb{E}\langle (D\overline{F})^2, (DL^{-1}(\overline{F} - \mathbb{E}\overline{F}))^2 \rangle_{L^2(\mu)}}.$$

Proof: We can write the U-statistic F as

$$F = \sum_{(x_1,\ldots,x_k) \in \eta^k_{\neq}} f(x_1, \ldots, x_k) = \underbrace{\sum_{(x_1,\ldots,x_k) \in \eta^k_{\neq}} f^+(x_1, \ldots, x_k)}_{=F^+} - \underbrace{\sum_{(x_1,\ldots,x_k) \in \eta^k_{\neq}} f^-(x_1, \ldots, x_k)}_{=F^-}$$

with $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ and have $\overline{F} = F^+ + F^-$. As a consequence of (6), we know that $D_z V \geq 0$ for a U-statistic V with non-negative summands. Combining this with $f^+, f^- \geq 0$ and Lemma 5.1, we see that

$$-D_z L^{-1}\left(F^+ - \mathbb{E}F^+\right) \geq 0 \quad \text{and} \quad -D_z L^{-1}\left(F^- - \mathbb{E}F^-\right) \geq 0.$$

Moreover, it holds that $D_z \mathbb{1}(F > s) \, D_z F \geq 0$. Proposition 6.1 in [6] implies that the product $D_z F \, D_z L^{-1}\left(\overline{F} - \mathbb{E}\overline{F}\right)$ has a finite chaos expansion with an order less than or equal to $2k - 2$. Together with Lemma 2.3, the Cauchy-Schwarz inequality, and Lemma 2.4, we obtain

$$\sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(F > s), DF \, |DL^{-1}(F - \mathbb{E}F)| \rangle_{L^2(\mu)}$$

$$= \sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(F > s), DF \, |DL^{-1}\left(F^+ - \mathbb{E}F^+ - F^- + \mathbb{E}F^-\right)| \rangle_{L^2(\mu)}$$

$$\leq \sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(F > s), DF \left(-DL^{-1}(F^+ - \mathbb{E}F^+) - DL^{-1}(F^- - \mathbb{E}F^-)\right) \rangle_{L^2(\mu)}$$

$$\leq \mathbb{E}\left[\delta\left(DF \, DL^{-1}\left(\overline{F} - \mathbb{E}\overline{F}\right)\right)^2\right]^{\frac{1}{2}}$$

$$\leq \sqrt{(2k-1)\mathbb{E}\langle (DF)^2, (DL^{-1}\left(\overline{F} - \mathbb{E}\overline{F}\right))^2 \rangle_{L^2(\mu)}}.$$

Now the fact that $(D_z F)^2 \leq (D_z \overline{F})^2$ concludes the proof. $\qquad \square$

Proof of Theorem 4.2: In the following, we can assume that $M_{ij}(f) < \infty$ for $i, j = 1, \ldots, k$ since (23) is obviously true, otherwise. We consider the standardized random variable $G = (F - \mathbb{E}F)/\sqrt{\operatorname{Var} F}$, whose Wiener-Itô chaos expansion has the kernels $g_i = f_i/\sqrt{\operatorname{Var} F}$ for $i = 1, \ldots, k$. In order to simplify our notation, we use the abbreviation

$$S = \sum_{i,j=1}^{k} \frac{\sqrt{M_{ij}(f)}}{\operatorname{Var} F}.$$

Exactly as in the proof of Theorem 4.1 in [14], we obtain

$$\mathbb{E}\left|1 - \langle DG, -DL^{-1}G \rangle_{L^2(\mu)}\right| \leq k \sum_{i,j=1}^{k} \frac{\sqrt{R_{ij}}}{\operatorname{Var} F} \leq k \sum_{i,j=1}^{k} \frac{\sqrt{M_{ij}(f)}}{\operatorname{Var} F} = kS.$$

From straightforward computations using Fubini's Theorem, the Cauchy-Schwarz inequality, and (25), it follows that

$$
\begin{aligned}
\mathbb{E}\langle (DG)^2, (DL^{-1}G)^2 \rangle_{L^2(\mu)} &= \int_X \mathbb{E}(D_z G)^2 (D_z L^{-1}G)^2 \, \mu(\mathrm{d}z) \\
&= \int_X \mathbb{E}\left(\sum_{i=1}^k i \, I_{i-1}(g_i(z, \cdot)) \right)^2 \left(\sum_{i=1}^k I_{i-1}(g_i(z, \cdot)) \right)^2 \mu(\mathrm{d}z) \\
&\leq k^4 \int_X \mathbb{E} \sum_{i=1}^k I_{i-1}(g_i(z, \cdot))^2 \sum_{j=1}^k I_{j-1}(g_j(z, \cdot))^2 \, \mu(\mathrm{d}z) \\
&\leq k^4 \sum_{i,j=1}^k \frac{M_{ij}(f)}{(\mathrm{Var}\, F)^2} \leq k^4 S^2,
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\langle (DG)^2, (DG)^2 \rangle_{L^2(\mu)} &= \int_X \mathbb{E}\left(\sum_{i=1}^k i \, I_{i-1}(g_i(z, \cdot)) \right)^4 \mu(\mathrm{d}z) \\
&\leq k^6 \int_X \mathbb{E} \sum_{i=1}^k I_{i-1}(g_i(z, \cdot))^2 \sum_{j=1}^k I_{j-1}(g_j(z, \cdot))^2 \, \mu(\mathrm{d}z) \\
&\leq k^6 \sum_{i,j=1}^k \frac{M_{ij}(f)}{(\mathrm{Var}\, F)^2} \leq k^6 S^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\langle DG, DG \rangle_{L^2(\mu)}^2 &= \mathbb{E}\left(\sum_{i,j=1}^k ij \int_X I_{i-1}(g_i(z, \cdot)) I_{j-1}(g_j(z, \cdot)) \, \mu(\mathrm{d}z) \right)^2 \\
&\leq k^2 \sum_{i,j=1}^k i^2 j^2 \mathbb{E}\left(\int_X I_{i-1}(g_i(z, \cdot)) I_{j-1}(g_j(z, \cdot)) \, \mu(\mathrm{d}z) \right)^2 \\
&\leq k^6 \sum_{i,j=1}^k \frac{R_{ij}}{(\mathrm{Var}\, F)^2} + k^4 \sum_{i=1}^k \frac{(i!)^2 \|f_i\|_i^4}{(\mathrm{Var}\, F)^2} \\
&\leq k^6 \sum_{i,j=1}^k \frac{R_{ij}}{(\mathrm{Var}\, F)^2} + k^4 \leq k^6 \sum_{i,j=1}^k \frac{M_{ij}(f)}{(\mathrm{Var}\, F)^2} + k^4 \leq k^6 S^2 + k^4.
\end{aligned}
$$

As a consequence of Lemma 5.2, we have that

$$
\mathbb{E}\left(\frac{F - \mathbb{E}F}{\sqrt{\mathrm{Var}\, F}} \right)^4 \leq k^2 \sum_{i,j=1}^k \frac{M_{ij}(f)}{(\mathrm{Var}\, F)^2} + 3k^2 \leq k^2 S^2 + 3k^2.
$$

Combining the last three inequalities, we obtain

$$
\begin{aligned}
&2\sqrt{\mathbb{E}\langle (DG)^2, (DG)^2 \rangle_{L^2(\mu)}} + 2\left(\mathbb{E}\langle DG, DG \rangle_{L^2(\mu)}^2 \right)^{\frac{1}{4}} \left((EG^4)^{\frac{1}{4}} + 1 \right) \\
&\leq 2k^3 S + 2(k^{\frac{3}{2}}\sqrt{S} + k)(\sqrt{k}\sqrt{S} + 3^{\frac{1}{4}}\sqrt{k} + 1) \leq 16k^3
\end{aligned}
$$

for $S \leq 1$.

Lemma 5.3 together with a similar computation as for $\mathbb{E}\langle (DG)^2, (DL^{-1}G)^2 \rangle_{L^2(\mu)}$ implies that

$$\sup_{s \in \mathbb{R}} \mathbb{E}\langle D\mathbb{1}(G > s), DG\,|DL^{-1}G| \rangle_{L^2(\mu)} \leq \sqrt{(2k-1)k^4 \sum_{i,j=1}^{k} \frac{M_{ij}(f)}{(\text{Var}\,F)^2}} \leq \sqrt{2}k^{\frac{5}{2}}S.$$

Thus, it follows from Theorem 1.1 that

$$d_K\left(\frac{F - \mathbb{E}F}{\sqrt{\text{Var}\,F}}, N\right) \leq kS + 16k^3\,k^2 S + \sqrt{2}k^{\frac{5}{2}}S \leq 19k^5 S$$

for $S \leq 1$. Otherwise, this inequality still holds since the Kolmogorov distance is at most one. $\qquad\square$

In a similar way, one can also obtain an upper bound for the Kolmogorov distance between a Gaussian random variable and a finite sum of Poisson U-statistics. This class of Poisson functionals is interesting since Theorem 3.6 in [14] tells us that each Poisson functional $F \in L^2(\mathbb{P}_\eta)$ of order k with kernels $f_i \in L_s^1(\mu^i) \cap L_s^2(\mu^i)$ for $i = 1, \ldots, k$ is a finite sum of Poisson U-statistics (and a constant). For such a Poisson functional the upper bounds for the inner products in the proof of Theorem 4.2 that depend on R_{ij} and

$$\int_X \mathbb{E}I_{i-1}(f_i(z, \cdot))^2 I_{j-1}(f_j(z, \cdot))^2 \mu(\mathrm{d}z)$$

for $i, j = 1, \ldots, k$ still hold. Moreover, we can compute a similar bound as in Lemma 5.3 using the representation of F as a sum of Poisson U-statistics. Together with the fourth centred moment of F, one can obtain an upper bound for the Kolmogorov distance between $(F - \mathbb{E}F)/\sqrt{\text{Var}\,F}$ and a standard Gaussian random variable.

References

[1] Chen, L., Goldstein, L., Shao, Q.: Normal Approximation by Stein's Method. Springer, Berlin (2011)

[2] Chen, L., Shao, Q.: Stein's method for normal approximation. An Introduction to Stein's Method, 1–59, Singapore University Press, Singapore (2005)

[3] Eichelsbacher, P., Thäle, C.: New Berry-Esseen bounds for non-linear functionals of Poisson random measures. arXiv: 1310.1595 (2013)

[4] Lachièze-Rey, R., Peccati, G.: Fine Gaussian fluctuations on the Poisson space, I: contractions, cumulants and geometric random graphs. Electron. J. Probab. **18**, Article 32 (2013)

[5] Lachièze-Rey, R., Peccati, G.: Fine Gaussian fluctuations on the Poisson space, II: rescaled kernels, marked processes and geometric U-statistics. Stochastic Process. Appl. **123**, 4186–4218 (2013)

[6] Last, G.: Stochastic analysis for Poisson processes. arXiv: 1405.4416 (2014)

[7] Last, G., Peccati, G., Schulte, M.: Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalitites and stabilization. arXiv: 1401.7568 (2014)

[8] Last, G., Penrose, M.D.: Poisson process Fock space representation, chaos expansion and covariance inequalities. Probab. Theory Related Fields **150**, 663–690 (2011)

[9] Last, G., Penrose, M.D., Schulte, M., Thäle, C.: Moments and central limit theorems for some multivariate Poisson functionals. Adv. in Appl. Probab. **46**, 348–364 (2014)

[10] Nualart, D., Vives, J.: Anticipative calculus for the Poisson process based on the Fock space. Lecture Notes in Math. **1426**, 154–165 (1990)

[11] Peccati, G., Solé, J.L., Taqqu, M.S., Utzet, F.: Stein's method and normal approximation of Poisson functionals. Ann. Probab. **38**, 443–478 (2010)

[12] Peccati, G., Taqqu, M.S.: Wiener Chaos: Moments, Cumulants and Diagram Formulae. Springer, Berlin (2011)

[13] Privault, N.: Stochastic Analysis in Discrete and Continuous Settings with Normal Martingales. Springer, Berlin (2009)

[14] Reitzner, M., Schulte, M.: Central limit theorems for U-statistics of Poisson point processes. Ann. Probab. **41**, 3879–3909 (2013)

[15] Reitzner, M., Schulte, M., Thäle, C.: Limit theory for the Gilbert graph. arXiv: 1312.4861 (2013)

[16] Schulte, M., Thäle, C.: Distances between Poisson k-flats. Methodol. Comput. Appl. Probab. **16**, 311–329 (2014)

[17] Stein, C.: A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proceedings of the Sixth Berkley Symposium on Mathematical Statistics Probability 2, Univ. California Press, Berkley (1972)

[18] Stein, C.: Approximate Computation of Expectations. Institute of Mathematical Statistics, Hayward, CA (1986)

[19] Surgailis, D.: On multiple Poisson stochastic integrals and associated Markov semigroups. Probab. Math. Statist. **38**, 217–239 (1984)