Bernoulli **26**(1), 2020, 587–615
https://doi.org/10.3150/19-BEJ1141

Normal approximation for sums of weighted U-statistics – application to Kolmogorov bounds in random subgraph counting

NICOLAS PRIVAULT[1] and GRZEGORZ SERAFIN[2]

[1]*School of Physical and Mathematical Sciences, Nanyang Technological University, SPMS-MAS-05-43, 21 Nanyang Link, Singapore 637371, Singapore. E-mail: nprivault@ntu.edu.sg*
[2]*Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Ul. Wybrzeże Wyspiańskiego 27, 54-129 Wrocław, Poland. E-mail: grzegorz.serafin@pwr.edu.pl*

We derive normal approximation bounds in the Kolmogorov distance for sums of discrete multiple integrals and weighted U-statistics made of independent Bernoulli random variables. Such bounds are applied to normal approximation for the renormalized subgraph counts in the Erdős–Rényi random graph. This approach completely solves a long-standing conjecture in the general setting of arbitrary graph counting, while recovering recent results obtained for triangles and improving other bounds in the Wasserstein distance.

Keywords: Berry–Esseen bound; central limit theorem; Kolmogorov distance; Malliavin–Stein method; normal approximation; random graph; Stein–Chen method; subgraph count

1. Introduction

The Mallavin approach to the Stein method introduced in [16] for general functionals of Gaussian random fields has recently been extended to functionals of discrete Bernoulli sequences. In [17], normal approximation Stein bounds have been obtained in the Wassertein distance for functionals of symmetric Bernoulli sequences, and such results have been extended in particular to the Kolmogorov distance in [10].

In [20], Stein bounds in the Wasserstein distance have been obtained for functionals of not necessarily symmetric Bernoulli sequences, and bounds in the total variation distance have been derived for the Poisson approximation in [9]. See also [4] for recent results on the fourth moment in the non-symmetric discrete setting.

Still in the discrete not necessarily symmetric Bernoulli setting, Kolmogorov distance bounds have been proved in [11] using second order Poincaré inequalities for discrete Bernoulli sequences, with application to the normal approximation of the renormalized count of the subgraphs which are isomorphic to triangles in the Erdős–Rényi random graph.

In this paper, we consider sums of weighted U-statistics (or discrete multiple stochastic integrals) of the form

$$\sum_{k=1}^{n} \sum_{\substack{i_1,\ldots,i_k\in\mathbb{N}\\ i_r\neq i_s,1\leq r\neq s\leq k}} f_k(i_1,\ldots,i_k) Y_{i_1}\cdots Y_{i_k}, \tag{1}$$

where $(Y_k)_{k\in\mathbb{N}}$ is a normalized sequence of Bernoulli random variables. By the Malliavin approach to the Stein and Stein–Chen methods, we derive new Kolmogorov distance bounds to the normal distribution for the distribution of functionals of the form (1), see Theorem 3.1. Our approach is based on results of [10] and [11] for general functionals of discrete i.i.d. renormalized Bernoulli sequences $(Y_n)_{n\in\mathbb{N}}$.

Normal approximation in the Kolmogorov distance has been studied in various special cases of (1). In Theorem 3.1 of [3], bounds were obtained for non weighted U-statistics, and in [10] the authors dealt with weighted first order U-statistics in the symmetric case $p = 1/2$. See also [12, 13] for the normal approximation of U-statistics written as multiple Poisson stochastic integrals, with applications provided to subgraph counting and boolean models.

Our second goal is to apply Theorem 3.1 to the normal approximation of the renormalized count of the subgraphs which are isomorphic to an arbitrary graph in the Erdős–Rényi random graph $\mathbb{G}_n(p)$ constructed by independently retaining any edge in the complete graph K_n on n vertices with probability $p \in (0, 1)$. The random graph $\mathbb{G}_n(p)$ was introduced by Gilbert [7] in 1959 and popularized by Erdős and Rényi in [5], it has been intensively studied and has become a classical model in discrete probability, see [8] and references therein.

Necessary and sufficient conditions for the asymptotic normality of the renormalization

$$\widetilde{N}_n^G := \frac{N_n^G - \mathbb{E}[N_n^G]}{\sqrt{\mathrm{Var}[N_n^G]}},$$

where N_n^G is the number of graphs in $\mathbb{G}_n(p_n)$ that are isomorphic to a fixed graph G, have been obtained in [23] where it is shown that

$$\widetilde{N}_n^G \xrightarrow{\mathcal{D}} \mathcal{N} \quad \text{iff } np_n^\beta \to \infty \text{ and } n^2(1-p_n) \to \infty,$$

as n tends to infinity, where $\mathcal{N}$ denotes the standard normal distribution,

$$\beta := \max\{e_H/v_H : H \subset G\},$$

and e_H, v_H respectively denote the numbers of edges and vertices in the graph H. Those results have been made more precise in [2] by the derivation of explicit convergence rates in the Wasserstein distance

$$d_W(F, G) := \sup_{h\in\mathrm{Lip}(1)} \big|\mathrm{E}\big[h(F)\big] - \mathrm{E}\big[h(G)\big]\big|,$$

between the laws of random variables F, G, where Lip(1) denotes the class of real-valued Lipschitz functions with Lipschitz constant less than or equal to 1. Bounds on the total variation distance of subgraph counts to the Poisson distribution have also been derived in Theorem 5.A of [1].

In the particular case where the graph G is a triangle, such bounds have been recently strengthened in [21] using the Kolmogorov distance

$$d_K(F, G) := \sup_{x\in\mathbb{R}} \big|P(F \le x) - P(G \le x)\big|,$$

which satisfies the bound $d_K(F,\mathcal{N}) \le \sqrt{d_W(F,\mathcal{N})}$. Still in the case of triangles, Kolmogorov distance bounds had also been obtained by second order Poincaré inequalities for discrete Bernoulli sequences in [11] when p_n takes the form $p_n = n^{-\alpha}$, $\alpha \in [0, 1)$. Kolmogorov bounds have also been obtained for triangles in Section 3.2.1 of [22], however such bounds apply only in the range $\alpha \in [0, 2/9)$ when p_n takes the form $p_n = n^{-\alpha}$.

In this paper, we refine the results of [2] by using the Kolmogorov distance instead of the Wasserstein distance. As in [2], we are able to consider any graph G, and therefore our results extend those of both [11] and [21] which only cover the case where G is a triangle. Instead of using second order Poincaré inequalities [11,14], our method relies on an application of Kolmogorov distance bounds of Proposition 4.1 in [11], see also Theorem 3.1 in [10], to derive Stein approximation bounds for sums of multiple stochastic integrals.

Furthermore, we note that various random functionals on the Erdős–Rényi random graph $\mathbb{G}_n(p)$ admit representations as sums of multiple integrals (1). This includes the number of vertices of a given degree, and the count of subgraphs that are isomorphic to an arbitrary graph.

Our second main result Theorem 4.2 is a bound for the Kolmogorov distance between the normal distribution and the renormalized graph count $\widetilde{N}_n^G$. Namely, we show that when G is a graph without isolated vertices it holds that

$$d_K(\tilde{N}_G,\mathcal{N}) \le C_G\Big((1-p_n)\min_{\substack{H\subset G\\ e_H\ge 1}} n^{v_H} p_n^{e_H}\Big)^{-1/2}, \tag{2}$$

where $C_G > 0$ is a constant depending only on e_G, which improves on the Wasserstein estimates of [2], see Theorem 2 therein. This result relies on the representation of combined subgraph counts as finite sums of multiple stochastic integrals, see Lemma 4.1, together with the application of Theorem 3.1 on Kolmogorov distance bounds for sums of multiple stochastic integrals.

In the sequel, given two positive sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ we write $x_n \approx y_n$ whenever $c_1 < x_n/y_n < c_2$ for some $c_1, c_2 > 0$ and all $n \in \mathbb{N}$, and for f and g two positive functions we also write $f \lesssim g$ whenever $f \le C_G g$ for some constant $C_G > 0$ depending only on G. Using the equivalence

$$\mathrm{Var}\big[N_n^G\big] \approx (1-p_n)\max_{\substack{H\subset G\\ e_H\ge 1}} n^{2v_G-v_H} p_n^{2e_G-e_H} \tag{3}$$

as n tends to infinity, see Lemma 3.5 in [8], the bound (2) can be rewritten in terms of the variance $\mathrm{Var}[N_n^G]$ as

$$d_K\big(\widetilde{N}_n^G,\mathcal{N}\big) \lesssim \frac{\sqrt{\mathrm{Var}[N_G]}}{(1-p_n)n^{v_G} p_n^{e_G}}. \tag{4}$$

Note that when p_n is bounded away from 0, the bound (2) takes the simpler form

$$d_K\big(\widetilde{N}_n^G,\mathcal{N}\big) \lesssim \frac{1}{n\sqrt{1-p_n}}. \tag{5}$$

Next, in Corollary 4.6 we show that the bound (4) can be specialized as

$$d_K\big(\widetilde{N}_n^G,\mathcal{N}\big) \lesssim \big((1-p_n)\min\{n^2 p_n, n^{v_G} p_n^{e_G}\}\big)^{-1/2}$$

$$= \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } n^{-(v_G-2)/(e_G-1)} < p_n, \\ \dfrac{1}{n^{v_G/2} p_n^{e_G/2}} & \text{if } 0 < p_n \le n^{-(v_G-2)/(e_G-1)}, \end{cases}$$

for any graph G with at least three vertices, under the balance condition

$$\max_{\substack{H \subset G \\ v_H \ge 3}} \frac{e_H - 1}{v_H - 2} = \frac{e_G - 1}{v_G - 2}, \tag{6}$$

see also [6,24] for related conditions and their use in subgraph counting. Finally, we note that (6) is satisfied by important examples of subgraphs such as complete graphs, cycles and trees, with at least 3 vertices, which are dealt with in Corollaries 4.8, 4.9 and 4.10.

In the particular case where the graph G is a triangle, the next consequence of (2) and (5) recovers the main result of [21], see Theorem 1.1 therein.

Corollary 1.1. *For any $c \in (0,1)$, the normalized number $\widetilde{N}_n^G$ of the subgraphs in $\mathbb{G}_n(p_n)$ that are isomorphic to a triangle satisfies*

$$d_K\big(\widetilde{N}_n^G, \mathcal{N}\big) \lesssim \begin{cases} \dfrac{1}{n\sqrt{1-p_n}} & \text{if } c < p_n < 1, \\ \dfrac{1}{n\sqrt{p_n}} & \text{if } n^{-1/2} < p_n \le c, \\ \dfrac{1}{(np_n)^{3/2}} & \text{if } 0 < p_n \le n^{-1/2}. \end{cases}$$

When p_n takes the form $p_n = n^{-\alpha}$, $\alpha \in [0,1)$, Corollary 1.1 similarly improves on the convergence rates obtained in Theorem 1.1 of [11] using second order Poincaré inequalities.

This paper is organized as follows. In Section 2, we recall the construction of random functionals of Bernoulli variables, together with the construction of the associated finite difference operator and their application to Kolmogorov distance bounds obtained in [10]. In Section 3, we derive general Kolmogorov distance bounds for sums of multiple stochastic integrals. In Section 4, we show that graph counts can be represented as sums of multiple stochastic integrals, and we derive Kolmogorov distance bounds for the renormalized count of subgraphs in $\mathbb{G}_n(p_n)$ that are isomorphic to a fixed graph.

2. Notation and preliminaries

In this section, we recall some background notation and results on the stochastic analysis of Bernoulli processes, see [18] for details. Consider a sequence $(X_n)_{n\in\mathbb{N}}$ of independent identically distributed Bernoulli random variables with $P(X_n = 1) = p$ and $P(X_n = -1) = q$, $n \in \mathbb{N}$, built as the sequence of canonical projections on $\Omega := \{-1,1\}^{\mathbb{N}}$. For any $F : \Omega \to \mathbb{R}$, we consider the $L^2(\Omega \times \mathbb{N})$-valued finite difference operator D defined for any $\omega = (\omega_0, \omega_1, \ldots) \in \Omega$ by

$$D_k F(\omega) = \sqrt{pq}\big(F\big(\omega_+^k\big) - F\big(\omega_-^k\big)\big), \quad k \in \mathbb{N}, \tag{7}$$

where we let

$$\omega_+^k := (\omega_0, \dots, \omega_{k-1}, +1, \omega_{k+1}, \dots)$$

and

$$\omega_-^k := (\omega_0, \dots, \omega_{k-1}, -1, \omega_{k+1}, \dots),$$

$k \in \mathbb{N}$, and $DF := (D_k F)_{k\in\mathbb{N}}$. The L^2 domain of D is given by

$$\mathrm{Dom}(D) = \big\{ F \in L^2(\Omega) : \mathrm{E}\big[\|DF\|^2_{\ell^2(\mathbb{N})} \big] < \infty \big\}.$$

We let $(Y_n)_{n\geq 0}$ denote the sequence of centered and normalized random variables defined by

$$Y_n := \frac{q - p + X_n}{2\sqrt{pq}}, \quad n \in \mathbb{N}.$$

Given $n \geq 1$, we denote by $\ell^2(\mathbb{N})^{\otimes n} = \ell^2(\mathbb{N}^n)$ the class of square-summable functions on $\mathbb{N}^n$, we denote by $\ell^2(\mathbb{N})^{\circ n}$ the subspace of $\ell^2(\mathbb{N})^{\otimes n}$ formed by functions that are symmetric in n variables. We let

$$I_n(f_n) = \sum_{(i_1,\dots,i_n)\in\Delta_n} f_n(i_1, \dots, i_n) Y_{i_1} \cdots Y_{i_n}$$

denote the discrete multiple stochastic integral of order n of f_n in the subspace $\ell^2_{\mathfrak{s}}(\Delta_n)$ of $\ell^2(\mathbb{N})^{\circ n}$ composed of symmetric kernels that vanish on diagonals, that is, on the complement of

$$\Delta_n = \big\{ (k_1, \dots, k_n) \in \mathbb{N}^n : k_i \neq k_j, 1 \leq i < j \leq n \big\}, \quad n \geq 1.$$

The multiple stochastic integrals satisfy the isometry and orthogonality relation

$$\mathrm{E}\big[I_n(f_n) I_m(g_m) \big] = \mathbb{1}_{\{n=m\}} n! \langle f_n, g_m \rangle_{\ell^2_{\mathfrak{s}}(\Delta_n)}, \tag{8}$$

$f_n \in \ell^2_{\mathfrak{s}}(\Delta_n)$, $g_m \in \ell^2_{\mathfrak{s}}(\Delta_m)$, cf. e.g. Proposition 1.3.2 of [19]. The finite difference operator D acts on multiple stochastic integrals as follows:

$$D_k I_n(f_n) = n I_{n-1}\big(f_n(*, k) \mathbb{1}_{\Delta_n}(*, k) \big) = n I_{n-1}\big(f_n(*, k) \big),$$

$k \in \mathbb{N}$, $f_n \in \ell^2_{\mathfrak{s}}(\Delta_n)$, and it satisfies the finite difference product rule

$$D_k(FG) = F D_k G + G D_k F - \frac{X_k}{\sqrt{pq}} D_k F D_k G, \quad k \in \mathbb{N} \tag{9}$$

for $F, G : \Omega \to \mathbb{R}$, see Propositions 7.3 and 7.8 of [18].

Due to the chaos representation property of Bernoulli random walks, any square integrable F may be represented as $F = \sum_{n\geq 0} I_n(f_n)$, $f_n \in \ell^2_{\mathfrak{s}}(\Delta_n)$, and the L^2 domain of D can be rewritten as

$$\mathrm{Dom}(D) = \left\{ F = \sum_{n\geq 0} I_n(f_n) : \sum_{n\geq 1} n n! \|f_n\|^2_{\ell^2(\mathbb{N})^{\otimes n}} < \infty \right\}.$$

The Ornstein–Uhlenbeck operator L is defined on the domain

$$\mathrm{Dom}(L) := \left\{ F = \sum_{n\geq 0} I_n(f_n) : \sum_{n\geq 1} n^2 n! \| f_n \|^2_{\ell^2(\mathbb{N})^{\otimes n}} < \infty \right\}$$

by

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

The inverse of L, denoted by L^{-1}, is defined on the subspace of $L^2(\Omega)$ composed of centered random variables by

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

with the convention $L^{-1}F = L^{-1}(F - \mathrm{E}[F])$ in case F is not centered. Using this convention, the duality relation (11) shows that for any $F, G \in \mathrm{Dom}(D)$ we have the covariance identity

$$\mathrm{Cov}(F, G) = \mathrm{E}\big[G\big(F - \mathrm{E}[F]\big)\big] = \mathrm{E}\big[\langle DG, -DL^{-1}F\rangle_{\ell^2(\mathbb{N})}\big]. \tag{10}$$

The divergence operator δ is the linear mapping defined as

$$\delta(u) = \delta\big(I_n\big(f_{n+1}(*, \cdot)\big)\big) = I_{n+1}(\tilde{f}_{n+1}), \quad f_{n+1} \in \ell^2_{\mathfrak{s}}(\Delta_n) \otimes \ell^2(\mathbb{N}),$$

for $(u_k)_{k\in\mathbb{N}}$ of the form

$$u_k = I_n\big(f_{n+1}(*, k)\big), \quad k \in \mathbb{N},$$

in the space

$$\mathcal{U} = \left\{ \sum_{k=0}^{n} I_k\big(f_{k+1}(*, \cdot)\big), \, f_{k+1} \in \ell^2_{\mathfrak{s}}(\Delta_k) \otimes \ell^2(\mathbb{N}), 0 \leq k \leq n \in \mathbb{N} \right\} \subset L^2(\Omega \times \mathbb{N})$$

of finite sums of multiple integral processes, where $\tilde{f}_{n+1}$ denotes the symmetrization of f_{n+1} in $n+1$ variables, i.e.

$$\tilde{f}_{n+1}(k_1, \ldots, k_{n+1}) = \frac{1}{n+1} \sum_{i=1}^{n+1} f_{n+1}(k_1, \ldots, k_{k-1}, k_{k+1}, \ldots, k_{n+1}, k_i).$$

The operators D and δ are closable with respective domains $\mathrm{Dom}(D)$ and $\mathrm{Dom}(\delta)$, built as the completions of $\mathcal{S}$ and $\mathcal{U}$, and they satisfy the duality relation

$$\mathbb{E}\big[\langle DF, u\rangle_{\ell^2(\mathbb{N})}\big] = \mathbb{E}\big[F\delta(u)\big], \quad F \in \mathrm{Dom}(D), u \in \mathrm{Dom}(\delta), \tag{11}$$

see, for example, Proposition 9.2 in [18], and the isometry property

$$\mathbb{E}\big[\big|\delta(u)\big|^2\big] = \mathbb{E}\big[\|u\|^2_{\ell^2(\mathbb{N})}\big] + \mathbb{E}\Bigg[\sum_{\substack{k,l=0\\k\neq l}}^{\infty} D_k u_l D_l u_k - \sum_{k=0}^{\infty}(D_k u_k)^2\Bigg]$$

$$\leq \mathbb{E}\big[\|u\|^2_{\ell^2(\mathbb{N})}\big] + \mathbb{E}\Bigg[\sum_{\substack{k,l=0\\k\neq l}}^{\infty} D_k u_l D_l u_k\Bigg], \quad u \in \mathcal{U}, \tag{12}$$

cf. Proposition 9.3 of [18] and Satz 6.7 in [15]. Letting $(P_t)_{t\in\mathbb{R}_+} = (e^{tL})_{t\in\mathbb{R}_+}$ denote the Orsntein–Uhlenbeck semi-group defined as

$$P_t F = \sum_{n=0}^{\infty} e^{-nt} I_n(f_n), \quad t \in \mathbb{R}_+,$$

on random variables $F \in L^2(\Omega)$ of the form $F = \sum_{n=0}^{\infty} I_n(f_n)$, the Mehler's formula states that

$$P_t F = \mathbb{E}\big[F\big(X(t)\big) \mid X(0)\big], \quad t \in \mathbb{R}_+, \tag{13}$$

where $(X(t))_{t\in\mathbb{R}_+}$ is the Ornstein–Uhlenbeck process associated to the semi-group $(P_t)_{t\in\mathbb{R}_+}$, cf. Proposition 10.8 of [18]. As a consequence of the representation (13) of P_t, we can deduce the bound

$$\mathbb{E}\big[\big|D_k L^{-1} F\big|^{\alpha}\big] \leq \mathbb{E}\big[|D_k F|^{\alpha}\big], \tag{14}$$

for every $F \in \mathrm{Dom}(D)$ and $\alpha \geq 1$, see Proposition 3.3 of [11]. The following Proposition 2.1 is a consequence of Proposition 4.1 in [11], see also Theorem 3.1 in [10].

Proposition 2.1. *For $F \in \mathrm{Dom}(D)$ with $\mathbb{E}[F] = 0$ we have*

$$d_K(F, \mathcal{N}) \leq \big|1 - \mathbb{E}\big[F^2\big]\big| + \sqrt{\mathrm{Var}\big[\langle DF, -DL^{-1}F\rangle_{\ell^2(\mathbb{N})}\big]}$$

$$+ \frac{1}{2\sqrt{pq}}\sqrt{\sum_{k=0}^{\infty}\mathbb{E}\big[(D_k F)^4\big]}\Bigg(\sqrt{\mathbb{E}\big[F^2\big]} + \sqrt{\sum_{k=0}^{\infty}\mathbb{E}\big[\big(F D_k L^{-1} F\big)^2\big]}\Bigg)$$

$$+ \frac{1}{\sqrt{pq}} \sup_{x\in\mathbb{R}} \mathbb{E}\big[\big\langle D\mathbf{1}_{\{F>x\}}, DF\big|DL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big].$$

Proof. By Proposition 4.1 in [11], we have

$$d_K(F, \mathcal{N}) \leq \mathbb{E}\big[\big|1 - \langle DF, -DL^{-1}F\rangle_{\ell^2(\mathbb{N})}\big|\big]$$

$$+ \frac{\sqrt{2\pi}}{8}(pq)^{-1/2}\mathbb{E}\big[\big\langle |DF|^2, \big|DL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big] \tag{15}$$

$$+\frac{1}{2}(pq)^{-1/2}\mathbb{E}\big[\big\langle|DF|^2,\big|FDL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big] \qquad (16)$$
$$+(pq)^{-1/2}\sup_{x\in\mathbb{R}}\mathbb{E}\big[\big\langle D\mathbf{1}_{\{F>x\}},DF\big|DL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big].$$

On the other hand, it follows from the covariance identity (10) that it holds $\operatorname{Var} F = \mathbb{E}[|\langle DF, -DL^{-1}F\rangle_{l^2(\mathbb{N})}|]$, hence by the Cauchy–Schwarz and triangular inequalities we get

$$\begin{aligned}
&\mathrm{E}\big[\big|1-\big\langle DF,-DL^{-1}F\big\rangle_{\ell^2(\mathbb{N})}\big|\big]\\
&\quad\le\big\|1-\big\langle DF,-DL^{-1}F\big\rangle_{\ell^2(\mathbb{N})}\big\|_{L^2(\Omega)}\\
&\quad\le\big|1-\|F\|^2_{L^2(\Omega)}\big|+\big\|\big\langle DF,-DL^{-1}F\big\rangle_{\ell^2(\mathbb{N})}-\|F\|^2_{L^2(\Omega)}\big\|_{L^2(\Omega)}\\
&\quad=\big|1-\operatorname{Var}[F]\big|+\sqrt{\operatorname{Var}\big[\big\langle DF,-DL^{-1}F\big\rangle_{\ell^2(\mathbb{N})}\big]}.
\end{aligned}$$

Next, we have

$$\begin{aligned}
\mathbb{E}\big[\big\|DL^{-1}I_n(f_n)\big\|^2_{\ell^2(\mathbb{N})}\big]&=\sum_{k=0}^{\infty}\mathbb{E}\big[\big(I_{n-1}\big(f_n(k,\cdot)\big)\big)^2\big]\\
&=(n-1)!\sum_{k=0}^{\infty}\big\|f_n(k,\cdot)\big\|^2_{\ell^2(\mathbb{N})^{\otimes(n-1)}}\\
&=(n-1)!\|f_n\|^2_{\ell^2(\mathbb{N})^{\otimes n}}\\
&\le n!\|f_n\|^2_{\ell^2(\mathbb{N})^{\otimes n}}\\
&=\mathbb{E}\big[\big|I_n(f_n)\big|^2\big],
\end{aligned}$$

and consequently, by the orthogonality relation (8) we have

$$\mathbb{E}\big[\big\|DL^{-1}F\big\|^2_{\ell^2(\mathbb{N})}\big]\le\mathbb{E}\big[F^2\big]$$

for every $F\in L^2(\Omega)$, hence (15) is bounded by

$$\begin{aligned}
\mathbb{E}\big[\big\langle\big|DL^{-1}F\big|,|DF|^2\big\rangle_{\ell^2(\mathbb{N})}\big]&\le\mathbb{E}\left[\sqrt{\sum_{k=0}^{\infty}\big|D_kL^{-1}F\big|^2\sum_{k=0}^{\infty}|D_kF|^4}\right]\\
&\le\sqrt{\mathbb{E}\left[\sum_{k=0}^{\infty}\big|D_kL^{-1}F\big|^2\right]}\sqrt{\mathbb{E}\left[\sum_{k=0}^{\infty}(D_kF)^4\right]}\\
&=\sqrt{\mathbb{E}\big[\big\|DL^{-1}F\big\|^2_{\ell^2(\mathbb{N})}\big]}\sqrt{\mathbb{E}\left[\sum_{k=0}^{\infty}(D_kF)^4\right]}
\end{aligned}$$

$$\leq \sqrt{\mathbb{E}\big[F^2\big]}\sqrt{\mathbb{E}\Bigg[\sum_{k=0}^{\infty}(D_k F)^4\Bigg]}.$$

Eventually, regarding the third term (16), by the Cauchy–Schwarz inequality we find

$$\mathbb{E}\big[\big\langle (DF)^2, \big|FDL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big] \leq \sqrt{\sum_{k=0}^{\infty}\mathbb{E}\big[(D_k F)^4\big]}\sqrt{\sum_{k=0}^{\infty}\mathbb{E}\big[\big(FD_k L^{-1}F\big)^2\big]}.$$

□

Finally, given $f_n \in \ell^2_{\mathfrak{s}}(\Delta_n)$ and $g_m \in \ell^2_{\mathfrak{s}}(\Delta_m)$ we have the multiplication formula

$$I_n(f_n)I_m(g_m) = \sum_{s=0}^{2\min(n,m)} I_{n+m-s}(h_{n,m,s}), \tag{17}$$

see Proposition 5.1 of [20], provided that the functions

$$h_{n,m,s} := \sum_{s \leq 2i \leq 2\min(s,n,m)} i!\binom{n}{i}\binom{m}{i}\binom{i}{s-i}\left(\frac{q-p}{\sqrt{pq}}\right)^{2i-s} f_n \,\tilde{\star}_i^{s-i}\, g_m$$

belong to $\ell^2_{\mathfrak{s}}(\Delta_{n+m-s})$, $0 \leq s \leq 2\min(n,m)$, where $f_n \,\tilde{\star}_k^l\, g_m$ is defined as the symmetrization in $n+m-k-l$ variables of the contraction $f_n \star_k^l g_m$ defined as

$$\begin{aligned} & f_n \star_k^l g_m(a_{l+1},\dots,a_n,b_{k+1},\dots,b_m) \\ & \quad = \mathbb{1}_{\Delta_{n+m-k-l}}(a_{l+1},\dots,a_n,b_{k+1},\dots,b_m) \\ & \qquad \times \sum_{a_1,\dots,a_l \in \mathbb{N}} f_n(a_1,\dots,a_n)g_m(a_1,\dots,a_k,b_{k+1},\dots,b_m), \end{aligned}$$

$0 \leq l \leq k$, and the symbol $\sum_{s \leq 2i \leq 2\min(s,n,m)}$ means that the sum is taken over all the integers i in the interval $[s/2, \min(s,n,m)]$. We close this section with the following Proposition 2.2.

Proposition 2.2. *Let $f_n \in \ell^2_{\mathfrak{s}}(\Delta_n)$ and $g_m \in \ell^2_{\mathfrak{s}}(\Delta_m)$ be symmetric functions. For $0 \leq l < k \leq \min(n,m)$ we have*

$$\begin{aligned} & \big\| f_n \star_k^l g_m \big\|^2_{\ell^2(\mathbb{N})^{\otimes(m+n-k-l)}} \\ & \quad \leq \frac{1}{2}\big\| f_n \star_n^{l+n-k} f_n \big\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} + \frac{1}{2}\big\| g_m \star_m^{l+m-k} g_m \big\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}}, \end{aligned} \tag{18}$$

and

$$\big\| f_n \star_k^k g_m \big\|^2_{\ell^2(\mathbb{N})^{\otimes(m+n-2k)}} \leq \frac{1}{2}\big\| f_n \star_{n-k}^{n-k} f_n \big\|^2_{\ell^2(\mathbb{N})^{\otimes 2k}} + \frac{1}{2}\big\| g_m \star_{m-k}^{m-k} f_m \big\|^2_{\ell^2(\mathbb{N})^{\otimes 2k}} \tag{19}$$

$$+\frac{1}{2}\sum_{i=1}^{k}\binom{k}{i}^2\big(\|f_n \star_n^{n-i} f_n\|_{\ell^2(\mathbb{N})^{\otimes i}}+\|g_m \star_m^{m-i} f_m\|_{\ell^2(\mathbb{N})^{\otimes i}}\big).$$

Proof. Hölder's inequality applied twice gives us

$$\begin{aligned}
&\|f_n \star_k^l g_m\|^2_{\ell^2(\mathbb{N})^{\otimes(m+n-k-l)}}\\
&\quad=\sum_{z_1\in\mathbb{N}^{n-k}}\sum_{z_2\in\mathbb{N}^{m-k}}\sum_{y\in\mathbb{N}^{k-l}}\mathbb{1}_{\Delta_{n+m-k-l}}(y,z_1,z_2)\Bigg(\sum_{x\in\mathbb{N}^l}f_n(x,y,z_1)g_m(x,y,z_2)\Bigg)^2\\
&\quad\le\sum_{y\in\mathbb{N}^{k-l}}\mathbb{1}_{\Delta_{k-l}}(y)\sum_{z_1\in\mathbb{N}^{n-k}}\sum_{z_2\in\mathbb{N}^{m-k}}\Bigg(\sum_{x\in\mathbb{N}^l}f_n^2(x,y,z_1)\sum_{x\in\mathbb{N}^l}g_m^2(x,y,z_2)\Bigg)\\
&\quad\le\Bigg[\sum_{y\in\mathbb{N}^{k-l}}\mathbb{1}_{\Delta_{k-l}}(y)\Bigg(\sum_{z_1\in\mathbb{N}^{n-k}}\sum_{x\in\mathbb{N}^l}f_n^2(x,y,z_1)\Bigg)^2\\
&\qquad\times\sum_{y\in\mathbb{N}^{k-l}}\mathbb{1}_{\Delta_{k-l}}(y)\Bigg(\sum_{z_1\in\mathbb{N}^{m-k}}\sum_{x\in\mathbb{N}^l}g_m^2(x,y,z_2)\Bigg)^2\Bigg]^{1/2}\\
&=\|f_n \star_n^{l+n-k} f_n\|_{\ell^2(\mathbb{N})^{\otimes(k-l)}}\|g_m \star_m^{l+m-k} g_m\|_{\ell^2(\mathbb{N})^{\otimes(k-l)}}\\
&\quad\le\frac{1}{2}\|f_n \star_n^{l+n-k} f_n\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}}+\frac{1}{2}\|g_m \star_m^{l+m-k} g_m\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}}.
\end{aligned}$$

To derive the second assertion, we proceed as follows:

$$\begin{aligned}
&\|f_n \star_k^k g_m\|^2_{\ell^2(\mathbb{N})^{\otimes(m+n-2k)}}\\
&\quad=\sum_{y\in\mathbb{N}^{n-k}}\sum_{z\in\mathbb{N}^{m-k}}\mathbb{1}_{\Delta_{m+n-2k}}(y,z)\sum_{x_1\in\Delta_k}\sum_{x_2\in\Delta_k}f_n(x_1,y)g_m(x_1,z)f_n(x_2,y)g_m(x_2,z)\\
&\quad\le\sum_{x_1\in\Delta_k}\sum_{x_2\in\Delta_k}\Bigg(\sum_{y\in\mathbb{N}^{n-k}}f_n(x_1,y)f_n(x_2,y)\Bigg)\Bigg(\sum_{z\in\mathbb{N}^{m-k}}g_m(x_1,z)g_m(x_2,z)\Bigg)\\
&\quad\le\frac{1}{2}\sum_{x_1,x_2\in\Delta_k}\Bigg(\sum_{y\in\mathbb{N}^{n-k}}f_n(x_1,y)f_n(x_2,y)\Bigg)^2\\
&\qquad+\frac{1}{2}\sum_{x_1,x_2\in\Delta_k}\Bigg(\sum_{z\in\mathbb{N}^{m-k}}g_m(x_1,z)g_m(x_2,z)\Bigg)^2,
\end{aligned}$$

where we have used the inequality $ab \leq a^2 + b^2$. Finally, we get

$$\begin{aligned}
&\sum_{x_1 \in \Delta_k} \sum_{x_2 \in \Delta_k} \left(\sum_{y \in \mathbb{N}^{n-k}} f_n(x_1, y) f_n(x_2, y) \right)^2 \\
&= \sum_{i=0}^{k} \binom{k}{i}^2 \sum_{x \in \Delta_i} \sum_{x', x'' \in \Delta_{k-i}} \mathbb{1}_{\Delta_{2k-i}}\big(x, x', x''\big) \left(\sum_{y \in \mathbb{N}^{n-k}} f_n\big(x, x', y\big) f_n\big(x, x'', y\big) \right)^2 \\
&= \sum_{i=0}^{k} \binom{k}{i}^2 \big\| f_n \star_{n-k+i}^{n-k} f_n \big\|_{\ell^2(\mathbb{N})^{\otimes(2k-i)}} \\
&\leq \big\| f_n \star_{n-k}^{n-k} f_n \big\|_{\ell^2(\mathbb{N})^{\otimes(2k)}} + \sum_{i=1}^{k} \binom{k}{i}^2 \big\| f_n \star_{n}^{n-i} f_n \big\|_{\ell^2(\mathbb{N})^{\otimes i}}
\end{aligned}$$

by (18), which ends the proof. □

3. Kolmogorov bounds for sums of multiple stochastic integrals

Wasserstein bounds have been obtained for discrete multiple stochastic integrals in Theorem 4.1 of [17] in the symmetric case $p = q$ and in Theorems 5.3–5.5 of [20] in the possibly nonsymmetric case, and have been extended to the Kolmogorov distance in the symmetric case $p = q$ in Theorem 4.2 of [10]. The following consequence of Proposition 2.1 provides a Kolmogorov distance bound which further extends Theorem 4.2 of [10] from multiple stochastic integrals to sums of multiple stochastic integrals in the nonsymmetric case.

Theorem 3.1. *For any finite sum*

$$F = \sum_{k=1}^{m} I_k(f_k)$$

of discrete multiple stochastic integrals with $f_k \in \ell^2_{\mathfrak{s}}(\Delta_k)$, $k = 1, \dots, m$, *we have*

$$d_K(F, \mathcal{N}) \leq C_m \big(\big|1 - \operatorname{Var}[F]\big| + \sqrt{R_F} \big),$$

for some constant $C_m > 0$ *depending only on* m, *where*

$$\begin{aligned}
R_F := &\sum_{0 \leq l < i \leq m} (pq)^{l-i} \big\| f_i \star_i^l f_i \big\|^2_{\ell^2(\mathbb{N})^{\otimes(i-l)}} \\
&+ \sum_{1 \leq l < i \leq m} \big(\big\| f_l \star_l^l f_i \big\|^2_{\ell^2(\mathbb{N})^{\otimes(i-l)}} + \big\| f_i \star_l^l f_i \big\|^2_{\ell^2(\mathbb{N})^{\otimes 2(i-l)}} \big).
\end{aligned} \tag{20}$$

Proof. We introduce

$$R'_F := \sum_{1\le i\le j\le m} \sum_{k=1}^{i}\sum_{l=0}^{k} \mathbf{1}_{\{i=j=k=l\}^c} (pq)^{l-k} \| f_i \star_k^l f_j\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}}.$$

Since it holds that $R'_F \lesssim R_F$, it is enough to prove the required inequality with R'_F instead of R_F. Indeed, by the inequality (18), all the components of R'_F for $0 \le l < k \le i, j$, are dominated by those for $0 \le l < k = i = j$, and also, by the inequality (19), the ones where $1 \le k = l < i \le j$, are dominated by the components where $1 \le l = k < i = j$ or $1 \le l < k = i = j$. Finally, the components for $1 \le k = l = i < j$ remain unchanged.

We will estimate components in the inequality from Proposition 2.1. We have

$$D_r F = \sum_{i=0}^{m-1} (i+1) I_i\big(f_{i+1}(r,\cdot)\big) \quad \text{and} \quad D_r L^{-1} F = \sum_{i=0}^{m-1} I_i\big(f_{i+1}(r,\cdot)\big), \quad r \in \mathbb{N},$$

hence by the multiplication formula (17) we find

$$(D_r F)^2 = \sum_{0\le i\le j\le m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} c_{i,j,l,k} \left(\frac{q-p}{\sqrt{pq}}\right)^{k-l} I_{i+j-k-l}\big(f_{i+1}(r,\cdot) \tilde{\star}_k^l f_{j+1}(r,\cdot)\big) \tag{21}$$

and

$$\begin{aligned} &D_r F D_r L^{-1} F \\ &\quad = \sum_{0\le i\le j\le m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} d_{i,j,l,k} \left(\frac{q-p}{\sqrt{pq}}\right)^{k-l} I_{i+j-k-l}\big(f_{i+1}(r,\cdot) \tilde{\star}_k^l f_{j+1}(r,\cdot)\big), \end{aligned} \tag{22}$$

for some $c_{i,j,l,k},\ d_{i,j,l,k} \ge 0$. Applying the isometry relation (8) to (21) and using the bound $\|\tilde{f}_m\|_{\ell^2(\mathbb{N})^{\otimes n}} \le \|f_m\|_{\ell^2(\mathbb{N})^{\otimes m}}$, $f_m \in \ell^2(\mathbb{N})^{\otimes m}$, we get, writing $f \lesssim g$ whenever $f < C_m g$ for some constant $C_m > 0$ depending only on m,

$$\begin{aligned} &\sum_{r=0}^{\infty} \mathbb{E}\big[|D_r F|^4\big] \\ &\lesssim \sum_{0\le i\le j\le m-1} \sum_{k=0}^{i}\sum_{l=0}^{k}\sum_{r=0}^{\infty} \left(\frac{q-p}{\sqrt{pq}}\right)^{2k-2l} \| f_{i+1}(r,\cdot) \star_k^l f_{j+1}(r,\cdot)\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\ &= \sum_{0\le i\le j\le m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} \left(\frac{q-p}{\sqrt{pq}}\right)^{2k-2l} \| f_{i+1} \star_{k+1}^l f_{j+1}\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l+1)}} \\ &= \sum_{1\le i\le j\le m} \sum_{k=1}^{i}\sum_{l=0}^{k-1} \left(\frac{q-p}{\sqrt{pq}}\right)^{2k-2l-2} \| f_i \star_k^l f_j\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \end{aligned}$$

$$\leq pqR'_F. \tag{23}$$

Furthermore, by (22) it follows that

$$\begin{aligned}
&\langle DF, DL^{-1}F\rangle - \mathbb{E}\big[\langle DF, DL^{-1}F\rangle\big] \\
&= \sum_{r=0}^{\infty} \sum_{0\leq i\leq j\leq m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} c_{i,j,l,k}\mathbf{1}_{\{i=j=k=l\}^c}\left(\frac{q-p}{\sqrt{pq}}\right)^{k-l} \\
&\quad \times I_{i+j-k-l}\big(f_{i+1}(r,\cdot)\,\tilde{\star}_k^l\, f_{j+1}(r,\cdot)\big) \\
&= \sum_{0\leq i\leq j\leq m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} c_{i,j,l,k}\mathbf{1}_{\{i=j=k=l\}^c}\left(\frac{q-p}{\sqrt{pq}}\right)^{k-l} \\
&\quad \times I_{i+j-k-l}\left(\sum_{r=0}^{\infty} f_{i+1}(r,\cdot)\,\tilde{\star}_k^l\, f_{j+1}(r,\cdot)\right) \\
&= \sum_{0\leq i\leq j\leq m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} c_{i,j,l,k}\mathbf{1}_{\{i=j=k=l\}^c}\left(\frac{q-p}{\sqrt{pq}}\right)^{k-l} I_{i+j-k-l}\big(f_{i+1}\,\tilde{\star}_{k+1}^{l+1}\, f_{j+1}\big),
\end{aligned}$$

thus we get

$$\begin{aligned}
&\mathrm{Var}\big[\langle DF, -DL^{-1}F\rangle\big] \\
&\lesssim \sum_{0\leq i\leq j\leq m-1} \sum_{k=0}^{i}\sum_{l=0}^{k} \frac{\mathbf{1}_{\{i=j=k=l\}^c}}{(pq)^{k-l}} \big\| f_{i+1} \star_{k+1}^{l+1} f_{j+1}\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&= \sum_{1\leq i\leq j\leq m} \sum_{k=1}^{i}\sum_{l=1}^{k} \mathbf{1}_{\{i=j=k=l\}^c}\frac{1}{(pq)^{k-l}} \big\| f_i \star_k^l f_j\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&\leq R'_F.
\end{aligned}$$

Next, we have

$$\begin{aligned}
\sum_{k=0}^{\infty}\mathbb{E}\big[\big(FD_kL^{-1}F\big)^2\big] &= \mathbb{E}\left[F^2\sum_{k=0}^{\infty}\big(D_kL^{-1}F\big)^2\right] \\
&\leq \sqrt{\mathbb{E}\big[F^4\big]}\sqrt{\mathbb{E}\left[\left(\sum_{k=0}^{\infty}\big(D_kL^{-1}F\big)^2\right)^2\right]}
\end{aligned}$$

and (17) and (8) show that

$$
\begin{aligned}
\mathbb{E}\big[F^4\big] &\lesssim \mathbb{E}\Bigg[\Bigg(\sum_{1\le i\le j\le m}\sum_{k=0}^{i}\sum_{l=0}^{k}\left|\frac{q-p}{\sqrt{pq}}\right|^{k-l} I_{i+j-k-l}\big(f_i \tilde{\star}_k^l f_j\big)\Bigg)^2\Bigg] \\
&\lesssim \sum_{1\le i\le j\le m}\sum_{k=0}^{i}\sum_{l=0}^{k}(pq)^{l-k}\| f_i \star_k^l f_j\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&\lesssim R'_F + \sum_{i=1}^{m}\| f_i \star_i^i f_i\|^2_{\ell^2(\mathbb{N})^{\otimes 0}} + \sum_{1\le i<j\le m}\| f_i \star_0^0 f_j\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j)}} \\
&= R'_F + \sum_{i=1}^{m}\| f_i\|^4_{\ell^2(\mathbb{N})^{\otimes i}} + \sum_{1\le i<j\le m}\| f_i\|^2_{\ell^2(\mathbb{N})^{\otimes i}}\| f_j\|^2_{\ell^2(\mathbb{N})^{\otimes j}} \\
&\lesssim R'_F + \big(\mathrm{Var}[F]\big)^2,
\end{aligned}
$$

while as in (21) and (22) we have

$$
\begin{aligned}
&\mathbb{E}\Bigg[\Bigg(\sum_{k=0}^{\infty}\big(D_k L^{-1}F\big)^2\Bigg)^2\Bigg] \\
&= \mathbb{E}\Bigg[\Bigg(\sum_{k=0}^{\infty}\sum_{0\le i\le j\le m-1}\sum_{k=0}^{i}\sum_{l=0}^{k}\tilde{d}_{i,j,l,k}\left(\frac{q-p}{\sqrt{pq}}\right)^{k-l} \\
&\quad \times I_{i+j-k-l}\big(f_{i+1}(k,\cdot)\,\tilde{\star}_k^l\, f_{j+1}(k,\cdot)\big)\Bigg)^2\Bigg] \\
&\lesssim \sum_{0\le i\le j\le m-1}\sum_{k=0}^{i}\sum_{l=0}^{k}(pq)^{l-k}\| f_{i+1} \star_{k+1}^{l+1} f_{j+1}\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&= \sum_{1\le i\le j\le m}\sum_{k=1}^{i}\sum_{l=1}^{k}(pq)^{l-k}\| f_i \star_k^l f_j\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&\lesssim R'_F + \sum_{i=1}^{m}\| f_i \star_i^i f_i\|^2_{\ell^2(\mathbb{N})^{\otimes 0}} + \sum_{1\le i<j\le m}\| f_i \star_0^0 f_j\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j)}} \\
&= R'_F + \sum_{i=1}^{m}\| f_i\|^4_{\ell^2(\mathbb{N})^{\otimes i}} + \sum_{1\le i<j\le m}\| f_i\|^2_{\ell^2(\mathbb{N})^{\otimes i}}\| f_j\|^2_{\ell^2(\mathbb{N})^{\otimes j}} \\
&\lesssim R'_F + \big(\mathrm{Var}[F]\big)^2,
\end{aligned}
$$

hence we get

$$\sum_{k=0}^{\infty} \mathbb{E}\big[\big(F D_k L^{-1} F\big)^2\big] \lesssim R'_F + \big(\mathrm{Var}[F]\big)^2. \tag{24}$$

We now deal with the last component in Proposition 2.1 similarly as it is done in proof of Theorem 4.2 in [10]. Precisely, by the integration by parts formula (11) and the Cauchy–Schwarz inequality we have

$$\begin{aligned}\sup_{x\in\mathbb{R}} \mathbb{E}\big[\big\langle D\mathbf{1}_{\{F>x\}}, DF\big|DL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big] &= \sup_{x\in\mathbb{R}} \mathbb{E}\big[\mathbf{1}_{\{F>x\}}\delta\big(DF\big|DL^{-1}F\big|\big)\big] \\ &\le \sqrt{\mathbb{E}\big[\big(\delta\big(DF\big|DL^{-1}F\big|\big)\big)^2\big]}. \end{aligned} \tag{25}$$

Then, by the bound (12), the Cauchy–Schwarz inequality and the consequence (14) of Mehler's formula (13), we have

$$\begin{aligned}&\mathbb{E}\big[\big(\delta\big(DF\big|DL^{-1}F\big|\big)\big)^2\big] \\ &\quad\le \mathbb{E}\big[\big\|DF\big|DL^{-1}F\big|\big\|^2_{\ell^2(\mathbb{N})}\big] + \mathbb{E}\Bigg[\sum_{k,l=0}^{\infty}\big|D_k\big(D_lF\big|D_lL^{-1}F\big|\big)D_l\big(D_kF\big|D_kL^{-1}F\big|\big)\big|\Bigg] \\ &\quad\le \sqrt{\mathbb{E}\big[\|DF\|^4_{\ell^4(\mathbb{N})}\big]\mathbb{E}\big[\big\|DL^{-1}F\big\|^4_{\ell^4(\mathbb{N})}\big]} + \mathbb{E}\Bigg[\sum_{k,l=0}^{\infty}\big(D_k\big(D_lF\big|D_lL^{-1}F\big|\big)\big)^2\Bigg] \\ &\quad\le \mathbb{E}\big[\|DF\|^4_{\ell^4(\mathbb{N})}\big] + \sum_{k,l=0}^{\infty}\mathbb{E}\big[\big(D_k\big(D_lF\big|D_lL^{-1}F\big|\big)\big)^2\big].\end{aligned}$$

The first term in the last expression in bounded by pqR'_F as shown in (23), and it remains to estimate the last expectation. By the product rule (9) and the bound $|D_k|F|| \le |D_kF|$ obtained from the definition (7) of D and the triangle inequality, we get

$$\begin{aligned}&\mathbb{E}\big[\big(D_r\big(D_sF\big|D_sL^{-1}F\big|\big)\big)^2\big] \\ &\quad= \mathbb{E}\bigg[\bigg(\big(D_rD_sF\big|D_sL^{-1}F\big|\big) + \big(D_sFD_r\big|D_sL^{-1}F\big|\big) \\ &\qquad - \frac{X_r}{\sqrt{pq}}\big(D_rD_sFD_r\big|D_sL^{-1}F\big|\big)\bigg)^2\bigg] \\ &\quad\lesssim \mathbb{E}\bigg[(D_rD_sF)^2\big(D_sL^{-1}F\big)^2 + (D_sF)\big(D_rD_sL^{-1}F\big)^2 \\ &\qquad + \frac{1}{pq}(D_rD_sF)^2\big(D_rD_sL^{-1}F\big)^2\bigg],\end{aligned} \tag{26}$$

$r, s \in \mathbb{N}$. By the Cauchy–Schwarz inequality, we get

$$\sum_{r,s=0}^{\infty} \mathbb{E}\big[(D_r D_s F)^2 \big(D_s L^{-1} F\big)^2\big] = \mathbb{E}\left[\sum_{s=0}^{\infty} \big(D_s L^{-1} F\big)^2 \sum_{r=0}^{\infty} (D_r D_s F)^2\right]$$
$$\leq \sqrt{\mathbb{E}\left[\sum_{s=0}^{\infty} \big(D_s L^{-1} F\big)^4\right] \mathbb{E}\left[\sum_{s=0}^{\infty} \left(\sum_{r=0}^{\infty} (D_r D_s F)^2\right)^2\right]}.$$

The term $\mathbb{E}[\sum_{s=0}^{\infty} (D_s L^{-1} F)^4]$ can be bounded by $pq R'_F$ as in (23). To estimate the other term, we use the multiplication formula (17) as in (21) to obtain

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \left(\sum_{r=0}^{\infty} (D_r D_s F)^2\right)^2\right]$$
$$\lesssim \sum_{s=0}^{\infty} \mathbb{E}\left[\left(\sum_{r=0}^{\infty} \sum_{0 \leq i \leq j \leq m-2} \sum_{k=0}^{i} \sum_{l=0}^{k} \left|\frac{q-p}{\sqrt{pq}}\right|^{k-l} I_{i+j-k-l}\big(f_{i+2}(s, r, \cdot) \,\tilde{\star}_k^l\, f_{j+2}(s, r, \cdot)\big)\right)^2\right]$$
$$= c \sum_{s=0}^{\infty} \mathbb{E}\left[\left(\sum_{0 \leq i \leq j \leq m-2} \sum_{k=0}^{i} \sum_{l=0}^{k} \left|\frac{q-p}{\sqrt{pq}}\right|^{k-l} I_{i+j-k-l}\big(f_{i+2}(s, \cdot) \,\tilde{\star}_{k+1}^{l+1}\, f_{j+2}(s, \cdot)\big)\right)^2\right]$$
$$\lesssim \sum_{s=0}^{\infty} \sum_{0 \leq i \leq j \leq m-2} \sum_{k=0}^{i} \sum_{l=0}^{k} (pq)^{l-k} \big\| f_{i+2}(s, \cdot) \star_{k+1}^{l+1} f_{j+2}(s, \cdot)\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}}$$
$$= \sum_{0 \leq i \leq j \leq m-2} \sum_{k=0}^{i} \sum_{l=0}^{k} (pq)^{l-k} \big\| f_{i+2} \star_{k+2}^{l+1} f_{j+2}\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l+1)}}$$
$$= \sum_{2 \leq i \leq j \leq m} \sum_{k=2}^{i} \sum_{l=1}^{k-1} (pq)^{l+1-k} \big\| f_i \star_k^l f_j\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}}$$
$$\leq pq R'_F.$$

The term $\sum_{r,s=0}^{\infty} \mathbb{E}[(D_s F)^2 (D_r D_s L^{-1} F)^2]$ from (26) is similarly bounded by $pq R'_F$. Regarding the last term, we have

$$\sum_{r,s=0}^{\infty} \mathbb{E}\big[(D_r D_s F)^2 \big(D_r D_s L^{-1} F\big)^2\big] \leq \sqrt{\sum_{r,s=0}^{\infty} \mathbb{E}\big[(D_r D_s F)^4\big] \sum_{r,s=0}^{\infty} \mathbb{E}\big[\big(D_r D_s L^{-1} F\big)^4\big]}.$$

Using the multiplication formula (17), both sums inside the above square root can be estimated as

$$\begin{aligned}
&\sum_{r,s=0}^{\infty} \mathbb{E}\Bigg[\Bigg(\sum_{0\le i\le j\le m-2} \sum_{k=0}^{i}\sum_{l=0}^{k}\left|\frac{q-p}{\sqrt{pq}}\right|^{k-l} I_{i+j-k-l}\big(f_{i+2}(s,r,\cdot)\,\tilde{\star}_k^l\, f_{j+2}(s,r,\cdot)\big)\Bigg)^2\Bigg] \\
&\quad\lesssim \sum_{r,s=0}^{\infty}\sum_{0\le i\le j\le m-2}\sum_{k=0}^{i}\sum_{l=0}^{k}(pq)^{l-k}\big\| f_{i+2}(s,r,\cdot)\star_k^l f_{j+2}(s,r,\cdot)\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&\quad= \sum_{0\le i\le j\le m-2}\sum_{k=0}^{i}\sum_{l=0}^{k}(pq)^{l-k}\big\| f_{i+2}\star_{k+2}^l f_{j+2}\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l+2)}} \\
&\quad= \sum_{2\le i\le j\le m}\sum_{k=2}^{i}\sum_{l=0}^{k-2}(pq)^{l+2-k}\big\| f_{i}\star_{k}^l f_{j}\big\|^2_{\ell^2(\mathbb{N})^{\otimes(i+j-k-l)}} \\
&\quad\lesssim (pq)^2 R'_F.
\end{aligned}$$

Combining this together, we get

$$\sum_{r,s=0}^{\infty}\mathbb{E}\big[\big(D_r\big(D_sF\big|D_sL^{-1}F\big|\big)\big)^2\big]\lesssim pq\,R'_F$$

and consequently, by (25) we find

$$\sup_{x\in\mathbb{R}}\mathbb{E}\big[\big\langle D\mathbf{1}_{\{F>x\}}, DF\big|DL^{-1}F\big|\big\rangle_{\ell^2(\mathbb{N})}\big]\lesssim pq\,R'_F. \tag{27}$$

Applying (23)–(24) and (27) to Proposition 2.1, we get

$$d_K(F,\mathcal{N})\lesssim\big|1-\mathrm{Var}[F]\big|+\sqrt{R'_F}\big(1+\mathrm{Var}[F]+\sqrt{\mathrm{Var}[F]}+\sqrt{R'_F}\big).$$

If $R'_F\ge 1$, or if $R'_F\le 1$ and $\mathrm{Var}[F]\ge 2$, it is clear that $d_K(F,\mathcal{N})\lesssim|1-\mathrm{Var}[F]|+\sqrt{R'_F}$ since $d_K(F,\mathcal{N})\le 1$ by definition. If $R'_F\le 1$ and $\mathrm{Var}[F]\le 2$, we estimate $\mathrm{Var}[F]+\sqrt{\mathrm{Var}[F]}+\sqrt{R'_F}$ by a constant and also get the required bound. □

4. Application to random graphs

4.1. General result

In the sequel fix a numbering $(1,\dots,e_G)$ of the edges in G and we denote by $\mathrm{E}_n^G\subset\{1,\dots,\binom{n}{2}\}^{e_G}$ the set of sequences of (distinct) edges of a complete graph K_n that create a graph isomorphic to G, that is, a sequence $(e_{k_1}, \dots, e_{k_{e_G}})$ belongs to E_n^G if and only if the graph created by edges $e_{k_1}, \dots, e_{k_{e_G}}$ is isomorphic to G. The next lemma allows us to represent the number of subgraphs as a sum of multiple stochastic integrals, using the notation $P(X_k = 1) = p$, $P(X_k = -1) = 1 - p = q$, $k \in \mathbb{N}$.

Lemma 4.1. *We have the identity*

$$\tilde{N}_n^G = \frac{N_n^G - \mathbb{E}[N_n^G]}{\sqrt{\operatorname{Var}[N_n^G]}} = \sum_{k=1}^{e_G} I_k(f_k), \tag{28}$$

where

$$\begin{aligned}
&f_k(b_1, \dots, b_k) \\
&:= \frac{q^{k/2} p^{e_G - k/2}}{(e_G - k)!k!\sqrt{\operatorname{Var}[N_G]}} \sum_{(a_1,\dots,a_{e_G-k}) \in \mathbb{N}^{e_G-k}} \mathbf{1}_{(a_1,\dots,a_{e_G-k},b_1,\dots,b_k) \in \mathrm{E}_n^G}.
\end{aligned}$$

Proof. We have

$$\begin{aligned}
N_G &= \frac{1}{e_G! 2^{e_G}} \sum_{b_1,\dots,b_{e_G} \in \mathbb{N}} \mathbf{1}_{(b_1,\dots,b_{e_G}) \in E_G} (X_{b_1} + 1) \cdots (X_{b_{e_G}} + 1) \\
&= \frac{1}{e_G! 2^{e_G}} \sum_{m=0}^{e_G} \binom{e_G}{m} \sum_{b_1,\dots,b_m \in \mathbb{N}} g_m(b_1, \dots, b_m) X_{b_1} \cdots X_{b_m} \\
&= \frac{1}{e_G! 2^{e_G}} \sum_{m=0}^{e_G} \binom{e_G}{m} \sum_{k=0}^{m} \binom{m}{k} (p-q)^{m-k} \\
&\quad \times \sum_{b_1,\dots,b_k \in \mathbb{N}} g_k(b_1, \dots, b_k)(X_{b_1} + q - p) \cdots (X_{b_k} + q - p) \\
&= \frac{1}{e_G! 2^{e_G}} \sum_{m=0}^{e_G} \binom{e_G}{m} \sum_{k=0}^{m} \binom{m}{k} I_k(g_k)(2\sqrt{pq})^k (p-q)^{m-k} \\
&= \frac{1}{e_G! 2^{e_G}} \sum_{k=0}^{e_G} \binom{e_G}{k} (2\sqrt{pq})^k I_k(g_k) \sum_{m=k}^{e_G} \binom{e_G - k}{m - k} (p-q)^{m-k} \\
&= \frac{1}{2^{e_G}} \sum_{k=0}^{e_G} \frac{(2\sqrt{pq})^k}{(e_G - k)!k!} I_k(g_k)(1 + p - q)^{e_G - k} \\
&= \sum_{k=0}^{e_G} \frac{q^{k/2} p^{e_G - k/2}}{(e_G - k)!k!} I_k(g_k),
\end{aligned}$$

where g_k is the function defined as

$$g_k(b_1,\dots,b_k) := \sum_{(a_1,\dots,a_{e_G-k})\in\mathbb{N}^{e_G-k}} \mathbf{1}_{\mathrm{E}_n^G}(a_1,\dots,a_{e_G-k},b_1,\dots,b_k), \quad (b_1,\dots,b_k)\in\mathbb{N}^k, \quad (29)$$

which shows (28) with

$$f_k(b_1,\dots,b_k) := \frac{q^{k/2}p^{e_G-k/2}}{(e_G-k)!k!\sqrt{\mathrm{Var}[N_n^G]}} g_k(b_1,\dots,b_k). \qquad \square$$

Next, is the second main result of this paper.

Theorem 4.2. *Let G be a graph without isolated vertices. Then we have*

$$d_K(\tilde{N}_n^G,\mathcal{N}) \lesssim \Big((1-p)\min_{\substack{H\subset G\\ e_H\geq 1}} n^{v_H}p^{e_H}\Big)^{-1/2} \approx \frac{\sqrt{\mathrm{Var}[N_G]}}{(1-p)n^{v_G}p^{e_G}}.$$

Proof. By (28) and Theorem 3.1, we have

$$d_K(\tilde{N}_n^G,\mathcal{N}) \lesssim \frac{\sqrt{R_G}}{\mathrm{Var}[N_n^G]}, \quad (30)$$

where, taking g_k as in (29), by (20) we have

$$\begin{aligned} R_G &= \sum_{0\leq l<k\leq e_G} p^{4e_G-3k+l}q^{l+k}\|g_k \star_k^l g_k\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} \\ &\quad+ \sum_{1\leq l<k\leq e_G} p^{4e_G-2k}q^{2k}\|g_k \star_l^l g_k\|^2_{\ell^2(\mathbb{N})^{\otimes 2(k-l)}} \\ &\quad+ \sum_{1\leq l<k\leq e_G} p^{4e_G-l-k}q^{k+l}\|g_l \star_l^l g_k\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} \\ &\leq q\Bigg(\sum_{0\leq l<k\leq e_G} p^{4e_G-3k+l}\|g_k \star_k^l g_k\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} \\ &\quad+ \sum_{1\leq l<k\leq e_G} p^{4e_G-l-k}\|g_l \star_l^l g_k\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} \\ &\quad+ \sum_{1\leq l<k\leq e_G} p^{4e_G-2k}\|g_k \star_l^l g_k\|^2_{\ell^2(\mathbb{N})^{\otimes 2(k-l)}}\Bigg) \\ &= (1-p)(S_1+S_2+S_3). \end{aligned}$$

It is now sufficient to show that

$$S_1 + S_2 + S_3 \lesssim \max_{\substack{H \subset G \\ e_H \geq 1}} n^{4v_G - 3v_H} p^{4e_G - 3e_H}. \tag{31}$$

Indeed, applying (3) and (31) to (30) we get

$$\begin{aligned}
\frac{\sqrt{R_G}}{\mathrm{Var}[N_n^G]} &\lesssim \frac{\sqrt{1-p}\sqrt{\max_{H \subset G e_H \geq 1} n^{4v_G - 3v_H} p^{4e_G - 3e_H}}}{(1-p)\max_{H \subset G e_H \geq 1} n^{2v_G - v_H} p^{2e_G - e_H}} \\
&= \frac{(\min_{H \subset G e_H \geq 1} n^{v_H} p^{e_H})^{-3/2}}{\sqrt{1-p}(\min_{H \subset G e_H \geq 1} n^{v_H} p^{e_H})^{-1}} \\
&= \Big((1-p) \min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} \Big)^{-1/2}.
\end{aligned}$$

Thus

$$d_K\big(\tilde{N}_n^G, \mathcal{N}\big) \lesssim \frac{\sqrt{R_G}}{\mathrm{Var}[N_n^G]} \lesssim \Big((1-p) \min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} \Big)^{-1/2}.$$

In order to estimate S_1, let us observe that

$$\begin{aligned}
\big\| g_k \star_k^l g_k \big\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} &= \sum_{a'' \in \mathbb{N}^{k-l}} \Bigg(\sum_{a' \in \mathbb{N}^l} \Bigg(\sum_{a \in \mathbb{N}^{e_G - k}} \mathbf{1}_{\mathrm{E}_n^G}\big(a, a', a''\big) \Bigg)^2 \Bigg)^2 \\
&\approx \sum_{\substack{A \subset K_n \\ e_K = k-l}} \Bigg(\sum_{\substack{A \subset B \subset K_n \\ e_B = k}} \Bigg(\sum_{\substack{B \subset G' \subset K_n \\ G' \sim G}} 1 \Bigg)^2 \Bigg)^2 \\
&\approx \sum_{\substack{K \subset G \\ e_K = k-l}} n^{v_K} \Bigg(\sum_{\substack{K \subset H \subset G \\ e_H = k}} n^{v_H - v_K} \big(n^{v_G - v_H} \big)^2 \Bigg)^2 \\
&\approx \max_{\substack{K \subset H \subset G \\ e_K = k-l, e_H = k}} n^{4v_G - 2v_H - v_K}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
S_1 &\lesssim \sum_{0 \leq l < k \leq e_G} p^{4e_G - 3k + l} \max_{\substack{K \subset H \subset G \\ e_K = k-l, e_H = k}} n^{4v_G - 2v_H - v_K} \\
&= \sum_{0 \leq l < k \leq e_G} \max_{\substack{K \subset H \subset G \\ e_K = k-l, e_H = k}} n^{4v_G - 2v_H - v_K} p^{4e_G - 2e_H - e_K}
\end{aligned}$$

$$\lesssim \max_{\substack{K\subset H\subset G\\ e_K\geq 1}} n^{4v_G-2v_H-v_K} p^{4e_G-2e_H-e_K}.$$

For a fixed p, let $H_0 \subset G$, $e_{H_0} \geq 1$, be the subgraph of G such that

$$n^{v_{H_0}} p^{e_{H_0}} = \min_{H\subset G, e_H\geq 1} n^{v_H} p^{e_H}. \tag{32}$$

Then it is clear that

$$\begin{aligned} S_1 &\lesssim \max_{\substack{K\subset H\subset G\\ e_K\geq 1}} n^{4v_G-2v_H-v_K} p^{4e_G-2e_H-e_K} \\ &= n^{4v_G-3v_{H_0}} p^{4e_G-3e_{H_0}} \\ &= \max_{\substack{H\subset G\\ e_H\geq 1}} n^{4v_G-3v_H} p^{4e_G-3e_H}, \end{aligned}$$

as required. We proceed similarly with the sum S_2. For $1 \leq l < k \leq n$ we have

$$\begin{aligned} &\left\| g_l \star_l^l g_k \right\|^2_{\ell^2(\mathbb{N})^{\otimes 2(k-l)}} \\ &\quad\approx \sum_{c\in\mathbb{N}^{k-l}} \left(\sum_{b\in\mathbb{N}^l} \left(\sum_{a\in\mathbb{N}^{e_G-l}} \mathbf{1}_{\mathrm{E}_n^G}(a,b) \sum_{a'\in\mathbb{N}^{e_G-k}} \mathbf{1}_{\mathrm{E}_n^G}\left(a',b,c\right) \right) \right)^2 \\ &\quad\approx \sum_{\substack{A\subset K_n\\ e_A=k-l}} \left(\sum_{\substack{A\subset B\subset K_n\\ e_B=k}} \left(\sum_{\substack{B\setminus A\subset G''\subset K_n\\ G''\sim G}} 1 \sum_{\substack{B\subset G'\subset K_n\\ G'\sim G}} 1 \right) \right)^2 &(33) \\ &\quad\lesssim \sum_{\substack{K\subset G\\ e_K=k-l}} n^{v_K} \left(\sum_{\substack{K\subset H\subset G, H'\subset G\\ e_H=k, e_{H'}=l}} n^{v_H-v_K} \left(n^{v_G-v_{H'}} n^{v_G-v_H} \right) \right)^2 &(34) \\ &\quad\lesssim \max_{\substack{K,H'\subset G\\ e_K=k-l, e_{H'}=l}} n^{4v_G-2v_{H'}-v_K}, &(35) \end{aligned}$$

where H' in (34) stands for $B\setminus A$ in (33), whereas in (35) the sum over H' extends to all $H'\subset G$ such that $e_{H'}=l$. It follows that

$$\begin{aligned} S_2 &\lesssim \sum_{1\leq l<k\leq e_G} p^{4e_G-k-l} \max_{\substack{K,H'\subset G\\ e_K=k-l, e_{H'}=l}} n^{4v_G-2v_{H'}-v_K} \\ &= \sum_{1\leq l<k\leq e_G} \max_{\substack{K,H'\subset G\\ e_K=k-l, e_{H'}=l}} n^{4v_G-2v_{H'}-v_K} p^{4e_G-2v_{H'}-e_K} \end{aligned}$$

$$
\begin{aligned}
&\lesssim \max_{\substack{K', H' \subset G \\ e_{K'}, e_{H'} \geq 1}} n^{4v_G - 2v_{H'} - v_{K'}} p^{4e_G - 2v_{H'} - e_{K'}} \\
&= n^{4v_G - 3v_{H_0}} p^{4e_G - 3e_{H_0}} \\
&= \max_{\substack{H \subset G \\ e_H \geq 1}} n^{4v_G - 3v_H} p^{4e_G - 3e_H},
\end{aligned}
$$

where H_0 is defined in (32). Finally, we pass to estimates of S_3. For $1 \leq l < k \leq n$, we have

$$
\begin{aligned}
&\left\| g_k \star_l^l g_k \right\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} \\
&\lesssim \sum_{c, c' \in \mathbb{N}^{k-l}} \left(\sum_{b \in \mathbb{N}^l} \left(\sum_{a \in \mathbb{N}^{e_G - k}} \mathbf{1}_{\mathrm{E}_n^G}(a, b, c) \right) \left(\sum_{a' \in \mathbb{N}^{e_G - k}} \mathbf{1}_{\mathrm{E}_n^G}(a', b, c') \right) \right)^2 \\
&\approx \sum_{\substack{A, A' \subset K_n \\ e_A = e_{A'} = k-l}} \left(\sum_{\substack{B \subset K_n \\ e_B = l, e_{A \cap B} = e_{A' \cap B} = 0}} \left(\sum_{\substack{A \cup B \subset G' \subset K_n \\ G' \sim G}} 1 \right) \left(\sum_{\substack{A' \cup B \subset G'' \subset K_n \\ G'' \sim G}} 1 \right) \right)^2 \\
&\approx \sum_{\substack{K, K', H \subset G \\ e_K = e_{K'} = k-l, e_H = l \\ e_{K \cap H} = e_{K' \cap H} = 0}} \sum_{\substack{A, A' \subset K_n \\ A \sim K \\ A' \sim K'}} \left(\sum_{\substack{B \subset K_n \\ B \sim H \\ A \cap B \sim K \cap H \\ A' \cap B \sim K' \cap H}} \left(\sum_{\substack{A \cup B \subset G' \subset K_n \\ G' \sim G}} 1 \right) \left(\sum_{\substack{A' \cup B \subset G'' \subset K_n \\ G'' \sim G}} 1 \right) \right)^2 \\
&\approx \sum_{\substack{K, K', H \subset G \\ e_K = e_{K'} = k-l, e_H = l \\ e_{K \cap H} = e_{K' \cap H} = 0}} \sum_{\substack{A, A' \subset K_n \\ A \sim K \\ A' \sim K'}} \left(\sum_{\substack{B \subset K_n \\ B \sim H \\ A \cap B \sim K \cap H \\ A' \cap B \sim K' \cap H}} \left(n^{v_G - v_{A \cup B}} \right) \left(n^{v_G - v_{A' \cup B}} \right) \right)^2.
\end{aligned}
$$

Next, we note that given $A, A' \subset K_n$ it takes

$$
v_B - v_{A \cap B} - v_{A' \cap B} + v_{A \cap A' \cap B} = v_H - v_{K \cap H} - v_{K' \cap H} + v_{A \cap A' \cap B}
$$

vertices to create any subgraph $B \sim H$ such that $A \cap B \sim K \cap H$ and $A' \cap B \sim K' \cap H$, with the bound

$$
v_{A \cap A' \cap B} \leq \frac{1}{2} v_{A \cap A'} + \frac{1}{2} v_{A' \cap B} = \frac{1}{2} (v_{A \cap A'} + v_{K' \cap H}).
$$

Hence, we have

$$
\begin{aligned}
&\left\| g_k \star_l^l g_k \right\|^2_{\ell^2(\mathbb{N})^{\otimes(k-l)}} \\
&\lesssim \sum_{\substack{K, K', H \subset G \\ e_K = e_{K'} = k-l, e_H = l \\ e_{K \cap H} = e_{K' \cap H} = 0}} \sum_{\substack{A, A' \subset K_n \\ A \sim K \\ A' \sim K'}} \left(n^{v_H - v_{K \cap H} - v_{K' \cap H} + (v_{A \cap A'} + v_{K' \cap H})/2} \left(n^{v_G - v_{K \cup H}} \right) \left(n^{v_G - v_{K' \cup H}} \right) \right)^2.
\end{aligned}
$$

In order to estimate the above sum using powers of n, we need to consider the possible intersections $A \cap A'$ for $A, A' \subset K_n$, as follows:

$$\begin{aligned}
&\sum_{\substack{K,K',H\subset G\\ e_K=e_{K'}=k-l,e_H=l\\ e_{K\cap H}=e_{K'\cap H}=0}} \sum_{\substack{A,A'\subset K_n\\ A\sim K\\ A'\sim K'}} n^{4v_G+2v_H-2v_{K\cap H}-v_{K'\cap H}+v_{A\cap A'}-2v_{K\cup H}-2v_{K'\cup H}}\\
&\lesssim \sum_{\substack{K,K',H\subset G\\ e_K=e_{K'}=k-l,e_H=l\\ e_{K\cap H}=e_{K'\cap H}=0}} \sum_{i=0}^{v_K} n^{v_K+v_{K'}-i} n^{4v_G+2v_H-2v_{K\cap H}-v_{K'\cap H}+i-2v_{K\cup H}-2v_{K'\cup H}}\\
&\lesssim \sum_{\substack{K,K',H\subset G\\ e_K=e_{K'}=k-l,e_H=l\\ e_{K\cap H}=e_{K'\cap H}=0}} n^{v_K+v_{K'}+4v_G+2v_H-2v_{K\cap H}-v_{K'\cap H}-2v_{K\cup H}-2v_{K'\cup H}}. \qquad (36)
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&v_K + v_{K'} + 4v_G + 2v_H - 2v_{K\cap H} - v_{K'\cap H} - 2v_{K\cup H} - 2v_{K'\cup H}\\
&\quad = 4v_G - v_K - v_H - v_{K'\cup H},
\end{aligned}$$

so the sum (36) can be estimated as

$$\sum_{\substack{K,K',H\subset G\\ e_K=e_{K'}=k-l,e_H=l\\ e_{K\cap H}=e_{K'\cap H}=0}} n^{4v_G-v_K-v_H-v_{K'\cup H}} \lesssim \max_{\substack{K,H,L\subset G\\ e_K=k-l,e_H=l,e_L=k}} n^{4v_G-v_K-v_H-v_L},$$

from which it follows

$$\begin{aligned}
S_3 &\lesssim \sum_{1\le l<k\le e_G} p^{4e_G-2k} \max_{\substack{K,H,L\subset G\\ e_K=k-l,e_H=l,e_L=k}} n^{4v_G-v_K-v_H-v_L}\\
&= \sum_{1\le l<k\le e_G} \max_{\substack{K,H,L\subset G\\ e_K=k-l,e_H=l,e_L=k}} n^{4v_G-v_K-v_H-v_L} p^{4e_G-e_K-e_H-e_L}\\
&\lesssim \max_{\substack{K,H,L\subset G\\ e_K,e_H,e_L\ge 1}} n^{4v_G-v_K-v_H-v_L} p^{4e_G-e_K-e_H-e_L}\\
&\le n^{4v_G-3v_{H_0}} p^{4e_G-3e_{H_0}}\\
&= \max_{\substack{H\subset G\\ e_H\ge 1}} n^{4v_G-3v_H} p^{4e_G-3e_H},
\end{aligned}$$

which ends the proof. $\square$

4.2. Special cases

In the next corollary, we note that Theorem 4.2 simplifies if we narrow our attention to p_n depending of the complete graph size n and close to 0 or to 1.

Corollary 4.3. *Let G be a graph without separated vertices. For $p_n < c < 1$, $n \geq 1$, we have*

$$d_K\big(\widetilde{N}_n^G, \mathcal{N}\big) \lesssim \Big(\min_{\substack{H\subset G\\ e_H\geq 1}} n^{v_H} p_n^{e_H}\Big)^{-1/2}.$$

On the other hand, for $p_n > c > 0$, $n \geq 1$, it holds

$$d_K\big(\widetilde{N}_n^G, \mathcal{N}\big) \lesssim \frac{1}{n\sqrt{1-p_n}}.$$

As a consequence of Corollary 4.3 it follows that if

$$np_n^\beta \to \infty \quad \text{and} \quad n^2(1-p_n) \to \infty,$$

where $\beta := \max\{e_H/v_H : H \subset G\}$, then we have the convergence of the renormalized subgraph count $(\widetilde{N}_n^G)_{n\geq 1}$ to $\mathcal{N}$ in distribution as n tends to infinity, which recovers the sufficient condition in [23]. When $p \approx n^{-\alpha}$, $\alpha > 0$, Corollary 4.3 also shows that

$$d_K\big(\widetilde{N}_n^G, \mathcal{N}\big) \lesssim \Big(\min_{\substack{H\subset G\\ e_H\geq 1}} n^{v_H-\alpha e_H}\Big)^{-1/2}, \tag{37}$$

and in order for the above bound (37) to tend to zero as n goes to infinity, we should have

$$\alpha < \min_{H\subset G} \frac{v_H}{e_H} =: \frac{1}{\beta}. \tag{38}$$

Next, we specialize our results to the following class of graphs.

Definition 4.4. Let $\mathcal{B}$ denote the set of graphs G with at least three vertices, and such that

$$\max_{\substack{H\subset G\\ v_H\geq 3}} \frac{e_H-1}{v_H-2} = \frac{e_G-1}{v_G-2}.$$

We note that the set $\mathcal{B}$ is contained in the class of balanced graphs which satisfy

$$\max_{H\subset G} \frac{e_H}{v_H} = \frac{e_G}{v_G},$$

as well as in the class of strongly balanced graphs which satisfy

$$\max_{H\subset G} \frac{e_H}{v_H-1} = \frac{e_G}{v_G-1}.$$

Both classes have been used in the framework of subraph counting, see, for example, [6,24], however the authors have not found the class $\mathcal{B}$ in the literature.

Lemma 4.5. *Let G be a graph with $v_H \geq 3$ and $e_H \geq 1$. Then G belongs to the class $\mathcal{B}$ if and only if for any $p \in (0,1)$ and $n \geq v_G$ we have*

$$\min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p^{e_H} = \min\{n^2 p, n^{v_G} p^{e_G}\}.$$

Proof. ($\Rightarrow$) If $G \in \mathcal{B}$, then for any $H \subset G$ such that $v_H \geq 3$ we have

$$n^{v_H} p^{e_H} = n^2 p \big(np^{\frac{e_H - 1}{v_H - 2}}\big)^{v_H - 2} \geq n^2 p \big(np^{\frac{e_G - 1}{v_G - 2}}\big)^{v_H - 2}. \tag{39}$$

If $n^2 p \leq n^{v_G} p^{e_G}$, then it holds $np^{\frac{e_G - 1}{v_G - 2}} \geq 1$ and we get

$$n^{v_H} p^{e_H} \geq n^2 p,$$

as required. If $n^2 p > n^{v_G} p^{e_G}$, then $(np^{\frac{e_G - 1}{v_G - 2}})^{v_H - 2} < 1$, and consequently, using $v_H \leq v_G$, we obtain

$$n^{v_H} p^{e_H} \geq n^2 p \big(np^{\frac{e_G - 1}{v_G - 2}}\big)^{v_G - 2} = n^{v_G} p^{e_G}$$

from (39), which ends this part the proof.

($\Leftarrow$) Proof by contradiction. Assume that the right-hand side of the equivalence in the thesis is true and that there exists $H_0 \subsetneq G$, $v_{H_0} \geq 3$, such that $\frac{e_{H_0} - 1}{v_{H_0} - 2} > \frac{e_G - 1}{v_G - 2}$. Then, for $p_n := n^{-\alpha}$ where α is such that $\frac{v_{H_0} - 2}{e_{H_0} - 1} < \alpha < \frac{v_G - 2}{e_G - 1}$, we get

$$n^{v_{H_0}} p_n^{e_{H_0}} = n^2 p_n \big(n^{1 - \alpha \frac{e_{H_0} - 1}{v_{H_0} - 2}}\big)^{v_{H_0} - 2} < n^2 p.$$

Furthermore, since $np_n^{\frac{e_G - 1}{v_G - 2}} > 1$ and $v_{H_0} < v_G$, we obtain

$$\begin{aligned} n^{v_{H_0}} p_n^{e_{H_0}} &= n^2 p_n \big(np_n^{\frac{e_{H_0} - 1}{v_{H_0} - 2}}\big)^{v_{H_0} - 2} < n^2 p_n \big(np_n^{\frac{e_G - 1}{v_G - 2}}\big)^{v_{H_0} - 2} \\ &< n^2 p_n \big(np_n^{\frac{e_G - 1}{v_G - 2}}\big)^{v_G - 2} = n^{v_G} p_n^{e_G}. \end{aligned}$$

This means that

$$\min_{\substack{H \subset G \\ e_H \geq 1}} n^{v_H} p_n^{e_H} \leq n^{v_{H_0}} p_n^{e_{H_0}} < \min\{n^2 p_n, n^{v_G} p_n^{e_G}\},$$

which contradicts the main assumption. The proof is complete. $\square$

By virtue of Lemma 4.5, the bound in Theorem 4.2 simplifies significantly for graphs G in the class $\mathcal{B}$.

Corollary 4.6. *For any G in the class $\mathcal{B}$ we have*

$$d_K\big(\tilde{N}_n^G, \mathcal{N}\big) \lesssim \big((1-p_n)\min\{n^2 p_n, n^{v_G} p_n^{e_G}\}\big)^{-1/2}$$
$$= \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } n^{-(v_G-2)/(e_G-1)} < p_n, \\ \dfrac{1}{n^{v_G/2} p_n^{e_G/2}} & \text{if } 0 < p_n \le n^{-(v_G-2)/(e_G-1)}. \end{cases}$$

Next, we note that $\mathcal{B}$ is quite a rich class as it contains other important classes of graphs.

Proposition 4.7. *All complete graphs, cycles, and trees with at least* 3 *vertices belong to the class $\mathcal{B}$.*

Proof. First, consider a complete graph K_r with $r = v_{K_r} \ge 3$. For a subgraph $H \subset K_r$ with $v_H < r$ the maximal number of edges is $\binom{v_H}{2}$ in the case of a clique, thus we have

$$\frac{e_H - 1}{v_H - 2} \le \frac{\binom{v_H}{2} - 1}{r-2} = \frac{v_H + 1}{2} < \frac{r+1}{2} = \frac{e_{K_r} - 1}{v_{K_r} - 2}.$$

In case of a cycle graph C_r, $r = v_{C_r} \ge 3$ the maximal number of edges of a subgraph $H \subset C_r$ with $v_H < r$ vertices is realized for a linear subgraph having $v_H - 1$ edges, which yields

$$\frac{e_H - 1}{v_H - 2} \le \frac{(v_H - 1) - 1}{v_H - 2} = 1 < \frac{r-1}{r-2} = \frac{e_{C_r} - 1}{v_{C_r} - 2}.$$

Finally, for a tree T with $r \ge 3$ vertices, the maximal number of edges of a subgraph $H \subset T$, $v_H < v_T$ is realized for a subtree with $v_H - 1$ edges, which gives

$$\frac{e_H - 1}{v_H - 2} = 1 = \frac{e_T - 1}{v_T - 2}.$$

This ends the proof. □

Corollaries 4.8–4.10 follow directly form Corollary 4.6 and Proposition 4.7. Since the triangle is a cycle as well as a complete graph, both of Corollaries 4.8 and 4.9 recover the Kolmogorov bounds of [21] as in Corollary 1.1 above.

Corollary 4.8. *Let C_r be a cycle graph with r vertices, $r \ge 3$. We have*

$$d_K\big(\widetilde{N}_n^{C_r}, \mathcal{N}\big) \lesssim \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } n^{-(r-2)/(r-1)} < p_n, \\ \dfrac{1}{(np_n)^{r/2}} & \text{if } 0 < p_n \le n^{-(r-2)/(r-1)}. \end{cases}$$

In case $p_n \approx n^{-\alpha}$ we should have $\alpha \in (0,1)$ by (38), and Corollary 4.8 also shows that

$$d_K\big(\widetilde{N}_n^{C_r}, \mathcal{N}\big) \lesssim \begin{cases} n^{-1+\alpha/2} \approx \dfrac{1}{n\sqrt{p_n}} & \text{if } 0 < \alpha \le \dfrac{r-2}{r-1}, \\ n^{-r(1-\alpha)/2} \approx \dfrac{1}{(np_n)^{r/2}} & \text{if } \dfrac{r-2}{r-1} \le \alpha < 1 \end{cases}$$

when C_r is a cycle graph with r vertices, $r \ge 3$. In the particular case $r = 3$ where C_3 is a triangle, this improves on the Kolmogorov bounds in Theorem 1.1 of [11].

Corollary 4.9. *Let K_r be a complete graph with $r \ge 3$ vertices, $r \ge 3$. We have*

$$d_K\big(\widetilde{N}_n^{G}, \mathcal{N}\big) \lesssim \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } n^{-2/(r+1)} < p_n, \\ \dfrac{1}{n^{r/2} p_n^{r(r-1)/4}} & \text{if } 0 < p_n \le n^{-2/(r+1)}. \end{cases}$$

When $p_n \approx n^{-\alpha}$ with $\alpha \in (0, 2/(r-1))$ by (38), Corollary 4.9 shows that

$$d_K\big(\widetilde{N}_n^{G}, \mathcal{N}\big) \lesssim \begin{cases} n^{-1+\alpha/2} \approx \dfrac{1}{n\sqrt{p_n}} & \text{if } 0 < \alpha \le \dfrac{2}{r+1}, \\ n^{-r/2+r(r-1)\alpha/4} \approx \dfrac{1}{n^{r/2} p_n^{r(r-1)/4}} & \text{if } \dfrac{2}{r+1} \le \alpha < \dfrac{2}{r-1}. \end{cases}$$

Finally, the next corollary deals with the important class of graphs which have a tree structure.

Corollary 4.10. *Let T be any tree (a connected graph without cycles) with r edges, and $c \in (0,1)$. We have*

$$d_K\big(\widetilde{N}_n^{T}, \mathcal{N}\big) \lesssim \begin{cases} \dfrac{1}{n\sqrt{p_n(1-p_n)}} & \text{if } \dfrac{1}{n} < p_n, \\ \dfrac{1}{n^{(r+1)/2} p_n^{r/2}} & \text{if } 0 < p_n \le \dfrac{1}{n}. \end{cases}$$

In case $p_n \approx n^{-\alpha}$ with $\alpha \in (0, 1+1/r)$, we get

$$d_K\big(\widetilde{N}_n^{G}, \mathcal{N}\big) \lesssim \begin{cases} n^{-1+\alpha/2} \approx \dfrac{1}{n\sqrt{p_n}} & \text{if } 0 < \alpha \le 1, \\ n^{-(r+1-r\alpha)/2} \approx \dfrac{1}{n^{(r+1)/2} p_n^{r/2}} & \text{if } 1 \le \alpha < 1 + \dfrac{1}{r}. \end{cases}$$

Acknowledgements

We would like to thank the anonymous reviewers for careful reading and many valuable comments and suggestions that helped us to improve the paper.

The research was supported by NTU MOE Tier 1 grant MOE2018-T1-001-201 RG25/18.

References

[1] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. *Oxford Studies in Probability* **2**. Oxford University Press: New York. MR1163825

[2] Barbour, A.D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. MR1047781 https://doi.org/10.1016/0095-8956(89)90014-2

[3] Chen, L.H.Y. and Shao, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** 581–599. MR2331265 https://doi.org/10.3150/07-BEJ5164

[4] Döbler, C. and Krokowski, K. (2019). On the fourth moment condition for Rademacher chaos. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 61–97. MR3901641 https://doi.org/10.1214/17-aihp876

[5] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

[6] Erdős, P. and Rényi, A. (1961). On the evolution of random graphs. *Bull. Inst. Internat. Statist.* **38** 343–347. MR0148055

[7] Gilbert, E.N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839 https://doi.org/10.1214/aoms/1177706098

[8] Janson, S., Łuczak, T. and Rucinski, A. (2000). *Random Graphs*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. New York: Wiley Interscience. MR1782847 https://doi.org/10.1002/9781118032718

[9] Krokowski, K. (2017). Poisson approximation of Rademacher functionals by the Chen–Stein method and Malliavin calculus. *Commun. Stoch. Anal.* **11** 195–222. MR3687774 https://doi.org/10.31390/cosa.11.2.05

[10] Krokowski, K., Reichenbachs, A. and Thäle, C. (2016). Berry–Esseen bounds and multivariate limit theorems for functionals of Rademacher sequences. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 763–803. MR3498009 https://doi.org/10.1214/14-AIHP652

[11] Krokowski, K., Reichenbachs, A. and Thäle, C. (2017). Discrete Malliavin–Stein method: Berry–Esseen bounds for random graphs and percolation. *Ann. Probab.* **45** 1071–1109. MR3630293 https://doi.org/10.1214/15-AOP1081

[12] Lachièze-Rey, R. and Peccati, G. (2013). Fine Gaussian fluctuations on the Poisson space, I: Contractions, cumulants and geometric random graphs. *Electron. J. Probab.* **18** no. 32, 32. MR3035760 https://doi.org/10.1214/EJP.v18-2104

[13] Lachièze-Rey, R. and Peccati, G. (2013). Fine Gaussian fluctuations on the Poisson space II: Rescaled kernels, marked processes and geometric U-statistics. *Stochastic Process. Appl.* **123** 4186–4218. MR3096352 https://doi.org/10.1016/j.spa.2013.06.004

[14] Last, G., Peccati, G. and Schulte, M. (2016). Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields* **165** 667–723. MR3520016 https://doi.org/10.1007/s00440-015-0643-7

[15] Mantei, A. (2015). Stochastisches Kalkül in diskreter Zeit. Masterarbeit, Institut für Mathematik, Universität Potsdam.

[16] Nourdin, I. and Peccati, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x

[17] Nourdin, I., Peccati, G. and Reinert, G. (2010). Stein's method and stochastic analysis of Rademacher functionals. *Electron. J. Probab.* **15** 1703–1742. MR2735379 https://doi.org/10.1214/EJP.v15-843

[18] Privault, N. (2008). Stochastic analysis of Bernoulli processes. *Probab. Surv.* **5** 435–483. MR2476738 https://doi.org/10.1214/08-PS139

[19] Privault, N. (2009). *Stochastic Analysis in Discrete and Continuous Settings with Normal Martingales*. *Lecture Notes in Math.* **1982**. Berlin: Springer. MR2531026 https://doi.org/10.1007/978-3-642-02380-4

[20] Privault, N. and Torrisi, G.L. (2015). The Stein and Chen–Stein methods for functionals of non-symmetric Bernoulli processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **12** 309–356. MR3357130

[21] Röllin, A. (2017). Kolmogorov bounds for the normal approximation of the number of triangles in the Erdös–Rényi random graph. Preprint. Available at arXiv:1704.00410.

[22] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182

[23] Ruciński, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** 1–10. MR0940863 https://doi.org/10.1007/BF00718031

[24] Ruciński, A. and Vince, A. (1986). Strongly balanced graphs and random graphs. *J. Graph Theory* **10** 251–264. MR0890231 https://doi.org/10.1002/jgt.3190100214

Received October 2018 and revised June 2019