

# The Gamma Stein equation and noncentral de Jong theorems

CHRISTIAN DÖBLER\* and GIOVANNI PECCATI\*\*

*Dedicated to the memory of Charles M. Stein*

*Université du Luxembourg, Unité de Recherche en Mathématiques, Maison Du Nombre, 6, avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg.*

*E-mail:* \* [christian.doebler@uni.lu](mailto:christian.doebler@uni.lu); \*\* [giovanni.peccati@uni.lu](mailto:giovanni.peccati@uni.lu)

We study the Stein equation associated with the one-dimensional Gamma distribution, and provide novel bounds, allowing one to effectively deal with test functions supported by the whole real line. We apply our estimates to derive new quantitative results involving random variables that are non-linear functionals of random fields, namely: (i) a non-central quantitative de Jong theorem for sequences of degenerate  $U$ -statistics satisfying minimal uniform integrability conditions, significantly extending previous findings by de Jong (*J. Multivariate Anal.* **34** (1990) 275–289), Nourdin, Peccati and Reinert (*Ann. Probab.* **38** (2010) 1947–1985) and Döbler and Peccati (*Electron. J. Probab.* **22** (2017) no. 2), (ii) a new Gamma approximation bound on the Poisson space, refining previous estimates by Peccati and Thäle (*ALEA Lat. Am. J. Probab. Math. Stat.* **10** (2013) 525–560) and (iii) new Gamma bounds on a Gaussian space, strengthening estimates by Nourdin and Peccati (*Probab. Theory Related Fields* **145** (2009) 75–118). As a by-product of our analysis, we also deduce a new inequality for Gamma approximations *via* exchangeable pairs, that is of independent interest.

*Keywords:* de Jong theorem; degenerate  $U$ -statistics; exchangeable pairs; Gamma approximation; Hoeffding decomposition; multiple stochastic integrals; Stein equation; Stein's method

## 1. Introduction

### 1.1. Overview

The aim of this paper is to derive new explicit estimates for one-dimensional Gamma approximations, and then to apply our general findings to derive several non-central approximation results for sequences of random variables that have the form of non-linear functionals of a random measure. The random measures we are interested in are either the empirical measure associated with a sequence of independent random variables, or a Poisson or Gaussian measure. As discussed below, our applications significantly refine and generalise previous results about the Gamma approximation of degenerate and not necessarily symmetric  $U$ -statistics [8–10,15,42], of smooth random variables on the Poisson space [41,42], and of smooth functionals of a Gaussian field [35,36,38].

From now on, for fixed  $r, \lambda \in (0, \infty)$ , we will denote by  $\Gamma(r, \lambda)$  the *Gamma distribution* with *shape parameter*  $r$  and *rate*  $\lambda$ , which has probability density function (p.d.f.)

$$p_{r,\lambda}(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\Gamma(t) := \int_0^\infty x^{t-1} e^{-x} dx$$

denotes the *Euler Gamma function*. We denote the corresponding distribution function by  $F_{r,\lambda}$ . It is well known that  $X_{r,\lambda} \sim \Gamma(r, \lambda)$  has mean  $r/\lambda$  and variance  $r/\lambda^2$  and that, if  $Y = aX_{r,\lambda}$  for some  $a > 0$ , then  $Y$  has distribution  $\Gamma(r, a^{-1}\lambda)$ . For  $\nu > 0$ , we also denote by  $\bar{\Gamma}(\nu)$  the so-called *centered Gamma distribution* with parameter  $\nu$ , which by definition is the distribution of

$$Z_\nu := 2X_{\nu/2,1} - \nu,$$

where, again,  $X_{\nu/2,1}$  has distribution  $\Gamma(\nu/2, 1)$ . Notice that, if  $\nu$  is an integer, then  $\bar{\Gamma}(\nu)$  has a centered  $\chi^2$  distribution with  $\nu$  degrees of freedom. According to the previous discussion, one has that

$$\mathbb{E}[Z_\nu] = 0 \quad \text{and} \quad \text{Var}(Z_\nu) = \mathbb{E}[Z_\nu^2] = 2\nu;$$

also, the following moment identity (already exploited in [36]), will play an important role throughout the paper:

$$\mathbb{E}[Z_\nu^4] - 12\mathbb{E}[Z_\nu^3] - 12\nu^2 + 48\nu = 0. \tag{1}$$

One of our principal aims in the sections to follow is to obtain several explicit estimates on quantities of the type

$$d(W, X_{r,\lambda}) := \sup_{h \in \mathcal{H}} |\mathbb{E}[h(W)] - \mathbb{E}[h(X_{r,\lambda})]|,$$

where  $\mathcal{H}$  is a suitable class of test functions. The strategy we will adopt in order to do so, is to derive new estimates on the solutions of the *Gamma Stein equation*

$$xf'(x) + (r - \lambda x)f(x) = h(x) - \mathbb{E}[h(X_{r,\lambda})], \quad x \in \mathbb{R}, \tag{2}$$

where  $h$  is an element of  $\mathcal{H}$ , and then to effectively use our bounds in the framework of *exchangeable pairs* (see [15,50]). We will see that our results significantly extend the classical findings by Luk [34] and Pickett [44], as well as the recent estimates from [22]. In particular, one crucial feature of our approach is that we will be able to directly study the Stein equation (2) on the whole real line, although the target distribution  $\Gamma(r, \lambda)$  is supported on the positive real axis. As discussed in Section 1.4, in the specific case of Gamma approximations on a Gaussian space, our results remarkably allow one to obtain quantitative limit theorems in the 1-Wasserstein distance (see below for definitions).

As anticipated, our main motivation comes from the study of the non-central fluctuations of random objects, which can be expressed in terms of iterated stochastic integrals with respect to a given random measure. The next three subsections contain a detailed discussion of our main applications to degenerate  $U$ -statistics and multiple integrals on the Poisson and Gaussian spaces.

### 1.2. A non-central de Jong theorem

Let  $X_1, \dots, X_n$  be independent random variables on some generic probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and with values in arbitrary measurable spaces  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ . In the recent paper [15], we were able to prove error bounds for the uni- and multivariate normal approximation of (vectors of) degenerate, non-symmetric  $U$ -statistics of the data vector  $X = (X_1, \dots, X_n)$ . In particular, we were able to provide a complete quantitative extension of a CLT by de Jong [10], which roughly states that a normalized sequence  $W_n, n \in \mathbb{N}$ , of such  $U$ -statistics converges weakly to the standard normal distribution if the sequence of fourth moments converges to 3 and some asymptotic Lindberg-type condition is satisfied; see formula (9) below.

The main abstract results of the present paper are used to continue such a line of research by dealing with the approximation of such a degenerate, non-symmetric  $U$ -statistic by a centered Gamma distribution. More precisely, assume that

$$\psi : \prod_{j=1}^n E_j \rightarrow \mathbb{R} \text{ is } \bigotimes_{j=1}^n \mathcal{E}_j - \mathcal{B}(\mathbb{R})\text{-measurable}$$

and that

$$W := \psi(X_1, \dots, X_n) \in L^4(\mathbb{P})$$

satisfies

$$\mathbb{E}[W] = 0 \quad \text{and} \quad \mathbb{E}[W^2] = 2\nu \tag{3}$$

for some  $\nu > 0$ . We write

$$[n] := \{1, \dots, n\}$$

and for  $J \subseteq [n]$  we define

$$\mathcal{F}_J := \sigma(X_j, j \in J).$$

We denote by

$$W = \sum_{J \subseteq [n]} W_J \tag{4}$$

the *Hoeffding decomposition* of  $W$  (see, e.g., [15,25–27,48,51]). Note that this means that, for each  $J \subseteq [n]$ ,  $W_J$  is  $\mathcal{F}_J$ -measurable and that

$$\mathbb{E}[W_J \mid \mathcal{F}_K] = 0,$$

whenever  $J \not\subseteq K$ . It is well known that  $W$  admits a Hoeffding decomposition of the type (4), as long as  $W \in L^1(\mathbb{P})$  and that it is almost surely unique and given by

$$W_J = \sum_{L \subseteq J} (-1)^{|J|-|L|} \mathbb{E}[W \mid \mathcal{F}_L], \quad J \subseteq [n]. \tag{5}$$

We can thus write

$$W_J = \psi_J(X_j, j \in J)$$

for some measurable function

$$\psi_J : \prod_{j \in J} E_j \rightarrow \mathbb{R}, \quad J \subseteq [n].$$

Let us also define

$$\sigma_J^2 := \text{Var}(W_J), \quad J \subseteq [n].$$

One major assumption in what follows will be that, for some fixed integer  $d \in [n]$ ,  $W$  is a *degenerate  $U$ -statistic of order  $d$*  (or  $d$ -degenerate  $U$ -statistic), that is, that the Hoeffding decomposition (4) has the form

$$W = \sum_{J \in \mathcal{D}_d} W_J, \tag{6}$$

where

$$\mathcal{D}_d := \{J \subseteq [n] : |J| = d\}$$

denotes the collection of all  $\binom{[n]}{d}$   $d$ -subsets of  $[n]$ , that is, we assume that  $W_K = 0$   $\mathbb{P}$ -a.s. whenever  $|K| \neq d$ . Hence, we have

$$W = \psi(X_1, \dots, X_n) = \sum_{J \in \mathcal{D}_d} \psi_J(X_j, j \in J). \tag{7}$$

Furthermore, we define the quantities

$$\rho^2 := \rho_n^2 := \max_{1 \leq i \leq n} \sum_{\substack{K \in \mathcal{D}_d: \\ i \in K}} \sigma_K^2 \quad \text{and} \quad D := D_n := \max_{J \in \mathcal{D}_d} \frac{\mathbb{E}[W_J^4]}{\sigma_J^4}.$$

One of the main results of the present paper is an explicit upper bound on a certain probability distance between the law of  $W$  and  $\bar{\Gamma}(v)$ . For  $k \in \mathbb{N}$ , denote by  $\mathcal{H}_k$  the class of those  $(k - 1)$ -times differentiable test functions  $h$  on  $\mathbb{R}$  such that  $h^{(k-1)}$  is Lipschitz-continuous and we have

$$\|h^{(l)}\|_\infty \leq 1 \quad \text{for } l = 1, \dots, k. \tag{8}$$

For real random variables  $X$  and  $Y$ , such that  $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$ , we denote by

$$d_k(X, Y) := d_k(\mathcal{L}(X), \mathcal{L}(Y)) := \sup_{h \in \mathcal{H}_k} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|$$

the distance between the distributions of  $X$  and  $Y$  induced by the class  $\mathcal{H}_k$ ; observe that  $d_1$  coincides with the classical 1-Wasserstein distance; see, for example, [37], Appendix C and the references therein. The distance  $d_k$  is reminiscent of the so-called Zolotarev  $\zeta_k$ -distance, involving smooth test functions  $h$  for which the bound (8) is only required in the case  $l = k$  (see, e.g., equation (14.2.1) in [45]).

The next theorem estimates the  $d_2$ -distance between the law of  $W$  and  $\bar{\Gamma}(\nu)$  in terms of the analogous linear combination of the moments of  $W$  as well as in terms of the quantities  $\rho_n^2$  and  $D_n$ .

**Theorem 1.1.** *Under the above assumptions, we have the bound*

$$d_2(W, Z_\nu) \leq \frac{\max(1, \frac{2}{\nu})}{\sqrt{3}} \sqrt{|\mathbb{E}[W^4] - 12\mathbb{E}[W^3] - 12\nu^2 + 48\nu|} + \frac{(2\sqrt{3} + 4\sqrt{\nu}) \max(1, \frac{2}{\nu}) + 4\sqrt{\nu}}{3\sqrt{d}} \sqrt{C_d D_n \rho_n^2},$$

where  $C_d$  is a finite constant which only depends on  $d$ .

One should immediately notice that the factors

$$\frac{\max(1, \frac{2}{\nu})}{\sqrt{3}}, \quad \text{and} \quad \frac{(2\sqrt{3} + 4\sqrt{\nu}) \max(1, \frac{2}{\nu}) + 4\sqrt{\nu}}{3\sqrt{d}},$$

both diverge to infinity as  $\nu \rightarrow 0$ . As formally discussed in Remark 2.8, this somewhat undesirable feature seems to be unavoidable: in particular, such a phenomenon is related to the fact that, for our applications, we need to be able to deal with random variables whose distribution is possibly supported by the whole real line.

The estimate in Theorem 1.1 immediately yields the following limit result.

**Corollary 1.2.** *Fix  $\nu > 0$  and an integer  $d \geq 1$  and let  $\{n_m : m \geq 1\}$  be a sequence of integers diverging to infinity. Let  $\{W_m : m \geq 1\}$  be a sequence of centered, degenerate  $U$ -statistics of order  $d$  with  $\mathbb{E}[W_m^2] = 2\nu$ , such that each  $W_m$  is a function of the vector of independent variables  $(X_1^{(m)}, \dots, X_{n_m}^{(m)})$ . Then, if*

$$\lim_{m \rightarrow \infty} (\mathbb{E}[W_m^4] - 12\mathbb{E}[W_m^3] - 12\nu^2 + 48\nu) = 0 = \lim_{m \rightarrow \infty} D_{n_m} \rho_{n_m}^2,$$

the sequence  $\{W_m : m \geq 1\}$  converges in distribution to  $Z_\nu$ .

Plainly, the asymptotic relation  $\lim_{m \rightarrow \infty} D_{n_m} \rho_{n_m}^2 = 0$  is verified whenever the sequence  $\{D_{n_m}\}$  is bounded, and  $\rho_{n_m}^2 \rightarrow 0$ ; see the discussion below.

It is also instructive to compare Theorem 1.1 and Corollary 1.2 with the main findings of [15], applying to the case where the assumption  $\mathbb{E}[W^2] = 2\nu$  in (3) is replaced by  $\mathbb{E}[W^2] = 1$ . In this

framework, letting  $Z$  be a standard normal random variable, one deduces from [15], Theorem 1.3 that

$$d_1(W, Z) \leq \left(\sqrt{\frac{2}{\pi}} + \frac{4}{3}\right)\sqrt{|\mathbb{E}[W^4] - 3|} + \sqrt{\kappa_d}\left(\sqrt{\frac{2}{\pi}} + \frac{2\sqrt{2}}{\sqrt{3}}\right)\rho_n, \tag{9}$$

where  $\kappa_d$  is a finite constant which only depends on  $d$ . As demonstrated in [15], from (9) one can immediately deduce de Jong’s theorem [10]: *Fix  $d \geq 1$ , and let  $\{n_m : m \geq 1\}$  be a sequence of integers diverging to infinity. Let  $\{W_m : m \geq 1\}$  be a sequence of unit variance degenerate  $U$ -statistics of order  $d$ , such that each  $W_m$  is a function of the vector of independent variables  $(X_1^{(m)}, \dots, X_{n_m}^{(m)})$ . Then, as  $m \rightarrow \infty$ , if  $\mathbb{E}[W_m^4] \rightarrow 3$  and  $\rho_{n_m}^2 \rightarrow 0$ , one has that  $W_m$  converges in distribution towards a standard Gaussian random variable.*

**Remark 1.3.**

- (a) Thanks to relation (1), Corollary 1.2 is an analog of de Jong’s theorem [10] in the context of a Gamma limit.
- (b) As discussed in Section 1.5 below, we believe that, in view of fundamental structural results from [18], the bound appearing in Theorem 1.1 is the best de Jong-type estimate on the Gamma approximation of  $U$ -statistics that can be achieved by using Stein’s method. Using the statement of Lemma 1.4 below, one can also immediately deduce a bound (with completely explicit constants) on the Wasserstein distance between  $W$  and  $Z_\nu$  whose order is the square root of the rate of convergence we get for the  $d_2$ -distance. We mention that we could also obtain a bound on the Kolmogorov distance whose order would be the power 1/3 of the rate for the  $d_2$ -distance, at least in the case  $\nu \geq 2$ , that is, when  $Z_\nu$  has a bounded density (see the remark following Lemma 1.4 below). We omit the details of this computation and refer to [19] for further information.
- (c) We conjecture that, analogously to the bounds on normal approximations derived in [15], the quantity  $D_n$  could be removed from the bound in Theorem 1.1, and hence, also from the limit theorem stated in Corollary 1.2.
- (d) In [38], the authors prove an error bound on the centered Gamma approximation (for integer  $\nu$ ) of homogeneous multilinear forms in independent and normalized real-valued random variables  $(X_i)_{i \in \mathbb{N}}$ . These form a particularly important example class of degenerate, non-symmetric  $U$ -statistics. Their bound also involves the quantities  $|\mathbb{E}[W^4] - 12\mathbb{E}[W^3] - 12\nu^2 + 48\nu|$ ,  $\rho_n^2$  and  $\beta := \sup_{i \in \mathbb{N}} \mathbb{E}[X_i^4]$  and it is easy to see that the condition  $\beta < \infty$  is in fact equivalent to the condition  $\sup_{n \in \mathbb{N}} D_n < \infty$  in this special situation. Thus, Theorem 1.1 and Corollary 1.2 can be seen as an extension and improvement of the bounds and limit theorems from [38] to a wider class of statistics.

The following new result, whose proof is deferred to Section 6, gives a universal bound for the Wasserstein distance in terms of the  $d_2$ -distance. Using similar techniques, the book [45] proves related bounds for the so-called *stop-loss distances* (see equations (17.1.49) and (17.1.50) therein).

**Lemma 1.4.** *Let  $X$  and  $Y$  be any real-valued random variables with  $\mathbb{E}|X| < \infty$  and  $\mathbb{E}|Y| < \infty$ . Then we have the bound*

$$d_1(X, Y) \leq \frac{4}{\sqrt{\pi}} \sqrt{d_2(X, Y)},$$

whenever  $d_2(X, Y) \leq 1$ .

We remark that, whenever  $\nu \geq 2$ , that is, when  $Z_\nu$  has a bounded density, by adapting the proof of Lemma 2.3 in [19], we could also provide a bound on the Kolmogorov distance in terms of  $d_2$ . In contrast to the bound in Lemma 1.4 on the Wasserstein distance, this bound, however, would involve the power  $1/3$  instead of  $1/2$ . This is in accordance with similar known bounds on the Kolmogorov distance in terms of the  $\zeta_2$ -distance mentioned above (see equation (14.1.16) in [45]).

### 1.3. Gamma limits on the Poisson space

In this subsection, we describe how our new bounds on the solution to the Gamma Stein equation (2), yield new analytic estimates for the Gamma approximation of functionals of a Poisson random measure. We will first briefly introduce the setup and some necessary notation. Further technical details are provided in Section 4. For any unexplained notions, we refer to the recent book [40], in particular Chapter 1 [32], as well as to the existing related literature, for example, [30,31,41,42]. We stress that limit theorems and probabilistic approximations involving non-linear functionals of a Poisson measure have gained enormous momentum in recent years, specially in connections with the large scale analysis of random geometric structures; see again [40], and the references therein.

We now fix a Polish space  $\mathcal{Z}$  as well as a  $\sigma$ -finite measure  $\mu$  on the Borel- $\sigma$ -field  $\mathcal{Z}$  on  $\mathcal{Z}$ . Furthermore, we let

$$\mathcal{Z}_\mu := \{B \in \mathcal{Z} : \mu(B) < \infty\}$$

and denote by

$$\eta = \{\eta(B) : B \in \mathcal{Z}_\mu\}$$

a *Poisson measure* on  $(\mathcal{Z}, \mathcal{Z})$  with *control*  $\mu$ , defined on a suitable probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We recall that the distribution of  $\eta$  is completely determined by the following two facts: (i) for each finite sequence  $B_1, \dots, B_m$  of disjoint sets in  $\mathcal{Z}_\mu$ , the random variables  $\eta(B_1), \dots, \eta(B_m)$  are independent, and (ii) that for every  $B \in \mathcal{Z}_\mu$ , the random variable  $\eta(B)$  has the Poisson distribution with mean  $\mu(B)$ . For  $B \in \mathcal{Z}_\mu$ , we also write  $\hat{\eta}(B) := \eta(B) - \mu(B)$  and denote by

$$\hat{\eta} = \{\hat{\eta}(B) : B \in \mathcal{Z}_\mu\}$$

the *compensated Poisson measure* associated with  $\eta$ . Without loss of generality, we may and will assume that  $\mathcal{F} = \sigma(\eta)$ .

Our main result in this section involves the following Malliavin operators: (i) the *Malliavin derivative*  $D$ , (ii) the *generator of the Ornstein–Uhlenbeck semigroup*  $L$  and (iii) the *pseudo-inverse* of  $L$ , written  $L^{-1}$ . Formal definitions and details are provided in Section 4. Here, we

only recall that the spectrum of  $L$  is given by the negative integers  $\{-p : p = 0, 1, 2, \dots\}$  and that  $F \in \text{Ker}(L + pI)$  (i.e.,  $F$  is an eigenfunction of  $L$ , with eigenvalue  $-p$ ) if and only if  $F = I_p(f)$ , where  $I_p$  indicates a multiple Wiener–Itô integral of order  $p$  with respect to  $\widehat{\eta}$ , and  $f$  is a suitable square-integrable kernel. The eigenspace  $\text{Ker}(L + pI)$  is customarily called the  $p$ th Wiener chaos associated with  $\eta$ .

The next statement – whose proof exploits our new results on the solution to the Stein equation (2) – is our main estimate on the Poisson space: in particular, its second part contains the announced result for multiple Wiener–Itô integrals. Proofs are deferred to Section 4.

**Theorem 1.5.** *Let  $F \in L^2(\mathbb{P})$  be centered, and assume that  $F$  belongs to the domain of the Malliavin derivative operator  $D$ . Then we have the bounds:*

$$d_2(F, Z_\nu) \leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E}|2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| \tag{10}$$

$$+ \max\left(1, \frac{1}{\nu} + \frac{1}{2}\right) \int_{\mathcal{Z}} \mathbb{E}[|D_z F|^2 | D_z L^{-1}F|] \mu(dz)$$

$$\leq \max\left(1, \frac{2}{\nu}\right) \sqrt{\mathbb{E}[(2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)})^2]} \tag{11}$$

$$+ \max\left(1, \frac{1}{\nu} + \frac{1}{2}\right) \int_{\mathcal{Z}} \mathbb{E}[|D_z F|^2 | D_z L^{-1}F|] \mu(dz).$$

Here, we have used the standard notation:

$$\langle DF, -DL^{-1}F \rangle_{L^2(\mu)} = - \int_{\mathcal{Z}} (D_z F)(D_z L^{-1}F) \mu(dz).$$

If, furthermore,  $F = I_p(f)$  for some  $p \geq 1$  and some square-integrable kernel  $f$ , then

$$\langle DF, -DL^{-1}F \rangle_{L^2(\mu)} = p^{-1} \|DF\|_{L^2(\mu)}^2 \quad \text{and}$$

$$\int_{\mathcal{Z}} \mathbb{E}[|D_z F|^2 | D_z L^{-1}F|] \mu(dz) = p^{-1} \int_{\mathcal{Z}} \mathbb{E}[|D_z F|^3] \mu(dz)$$

so that the previous estimates becomes

$$d_2(F, Z_\nu) \leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E}|2(F + \nu) - p^{-1} \|DF\|_{L^2(\mu)}^2| \tag{12}$$

$$+ p^{-1} \max\left(1, \frac{1}{\nu} + \frac{1}{2}\right) \int_{\mathcal{Z}} \mathbb{E}[|D_z F|^3] \mu(dz)$$

$$\leq \max\left(1, \frac{2}{\nu}\right) \sqrt{\mathbb{E}[(2(F + \nu) - p^{-1} \|DF\|_{L^2(\mu)}^2)^2]} \tag{13}$$

$$+ p^{-1} \max\left(1, \frac{1}{\nu} + \frac{1}{2}\right) \int_{\mathcal{Z}} \mathbb{E}[|D_z F|^3] \mu(dz).$$



**Remark 1.6.** The content of Theorem 1.5 should be directly compared with [42], Theorem 2.1, according to which

$$\begin{aligned}
 d_3(F, Z_\nu) &\leq c_1 \mathbb{E} |2(F + \nu)_+ - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| \\
 &\quad + c_2 \int_{\mathcal{Z}} \mathbb{E} [|D_z F|^2 |D_z L^{-1}F|] \mu(dz) \\
 &\quad + 2c_1 \int_{\mathcal{Z}} \mathbb{E} [(D_z F \mathbf{1}_{(F > -\nu)})(D_z F) |D_z L^{-1}F|] \mu(dz),
 \end{aligned}$$

where  $c_1, c_2$  are explicit constants uniquely depending on  $\nu$ . Note that our estimate (10) improves on such an estimate in three ways: (i) the distance  $d_3$  is replaced by the less smooth distance  $d_2$ , (ii) the first expectation on the right-hand side does not involve the positive part of  $F + \nu$  and (iii) the third term in the bound has been completely removed. As will become evident in the proof, Points (i) and (iii) are a direct consequence of the fact that our approach allows us to solve and control equation (2) on the whole real line, thus obtaining more tractable solutions than those used in [42]. Note that our bound can be directly used to deduce simplified proofs of the other estimates proved in [42] like, for example, [42], Theorem 2.6 and Proposition 2.9. Details are left to the reader.

### 1.4. Gamma limits on a Gaussian space

We conclude this section by showing how the results of the present paper can also be used to give better estimates on the Gamma approximation of non-linear functionals of Gaussian fields, thus improving results from [35,39]. For the sake of conciseness, in this section we will keep explicit definitions to a minimum, and refer the reader to the monograph [37] for any unexplained notion or detail.

Now let  $\mathcal{H}$  be a real separable Hilbert space, and let  $X = \{X(h) : h \in \mathcal{H}\}$  be an *isonormal Gaussian process* over  $\mathcal{H}$ . We assume that  $X$  is defined on a suitable probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and that  $\mathcal{F} = \sigma(X)$ . Similar to the previous section, we associate to  $X$  the following canonical Malliavin operators: (i) the *Malliavin derivative*  $D$  (whose domain is indicated by  $\mathbb{D}^{1,2}$ ), (ii) the *generator of the Ornstein–Uhlenbeck semigroup*  $L$  and (iii) the *pseudo-inverse* of  $L$ , written again  $L^{-1}$ . As on the Poisson space, the spectrum of  $L$  is given by the negative integers  $\{-p : p = 0, 1, 2, \dots\}$  and one has that  $F \in \text{Ker}(L + pI)$  (i.e., the  $p$ th *Wiener chaos* of  $X$ ) if and only if  $F = I_p(f)$ , where  $I_p$  indicates a multiple Wiener-Itô integral of order  $p$ , and  $f$  is an element of the symmetric tensor product  $\mathcal{H}^{\odot p}$ .

One has the following estimate (recall that  $d_1$  corresponds to the 1-Wasserstein distance).

**Theorem 1.7.** *Let  $F$  be centered element of  $\mathbb{D}^{1,2}$  and fix  $\nu > 0$ . Then*

$$\begin{aligned}
 d_1(F, Z_\nu) &\leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E} |\mathbb{E}\{2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{\mathcal{H}} \mid F\}| \\
 &\leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E} [(2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{\mathcal{H}})^2]^{1/2}.
 \end{aligned} \tag{14}$$

If  $F \in \text{Ker}(L + pI)$  for some integer  $p \geq 2$ , then the previous estimate becomes

$$d_1(F, Z_\nu) \leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E}|\mathbb{E}\{2(F + \nu) - p^{-1}\|DF\|_{\mathcal{H}}^2 \mid F\}|. \tag{15}$$

Inequality (14) improves [35], Theorem 3.11, where a similar upper bound is proved for a smoother distance (written  $d_{\mathcal{H}_2}$  therein) involving test functions of class  $C^2$  with bounded derivatives. By inspection of the proofs contained in [35], one sees that such a smoothness requirement on test functions is indeed an artefact of the bounds contained in [34]. By combining Theorem 1.7 with the main findings from [39] and with some computations from [3], one also obtains the following non-trivial quantitative characterisation of Gamma convergence in total variation inside a fixed sum of Wiener chaos. We recall that, given two real-valued random variables  $X, Y$ , the total variation distance between the distributions of  $X$  and  $Y$  is given by

$$d_{TV}(X, Y) = \sup_{A \in \mathcal{B}(\mathbb{R})} |\mathbb{P}[X \in A] - \mathbb{P}[Y \in A]|,$$

where  $\mathcal{B}(\mathbb{R})$  stands for the class of all Borel subsets of  $\mathbb{R}$ .

**Proposition 1.8.** Fix  $\nu > 0$ , as well as an integer  $m \geq 2$ , and let  $\{F_n : n \geq 1\} \subseteq \bigoplus_{p=1}^m \text{Ker}(L + pI)$  be such that  $\mathbb{E}[F_n^2] \rightarrow 2\nu$ . Then  $F_n$  converges in distribution to  $Z_\nu$  if and only if

$$\mathbb{E}|\mathbb{E}\{2(F_n + \nu) - \langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}} \mid F_n\}| \rightarrow 0, \quad n \rightarrow \infty, \tag{16}$$

and there exists a finite constant  $c > 0$  (not depending on  $n$ ) such that

$$d_{TV}(F_n, Z_\nu) \leq c(\mathbb{E}|\mathbb{E}\{2(F_n + \nu) - \langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}} \mid F_n\}|)^{1/2m+1}. \tag{17}$$

One has also to observe that, according to [36], if the sequence  $\{F_n\}$  in Proposition 1.8 is such that  $\{F_n\} \subseteq \text{Ker}(L + mI)$  and (16) is verified, then necessarily  $m$  is an even integer. See also [1,29] for some related limit theorems. The proofs of Theorem 1.7 and Proposition 1.8 are given in Section 5.

### 1.5. About our approach and assumptions

We will now make some technical remarks about the methods and assumptions adopted in the present paper.

- (i) First of all, we recall that the Stein equation associated with a given distribution is in general not unique, and several approaches are available in order to select a specific one. One of these methods, the so-called *density approach* (see, e.g., [7] and [33]), suggests a Stein equation of the form

$$f'(x) - \psi(x)f(x) = h(x) - \mathbb{E}[h(X_{r,\lambda})],$$

where  $\psi(x) := \frac{d}{dx} \log p_{r,\lambda}(x)$ ,  $x > 0$ , is the log-derivative of the density function. It is easy to see that, here,  $\psi(x) = \frac{r-1-\lambda x}{x}$  is a genuinely rational function of  $x$  (unless  $r = 1$ ), which makes this equation very difficult to apply in concrete situations involving probability approximations. Note in particular that, in the three examples presented in Sections 1.2–1.4, it is for us of fundamental importance to have a linear coefficient of  $f(x)$  in the Stein equation, which makes the choice of (2) inevitable.

- (ii) Due to their wide applicability (in particular in combination with Malliavin calculus techniques) the class of first-order Stein equations having a linear coefficient of  $f$ , characterizing some absolutely continuous distribution  $\mu$  on  $\mathbb{R}$ , has been well-studied in the recent literature (see, e.g., [13,18,28]). In particular, in [18] (see Remark 10 and the last paragraph on page 200) the authors prove by means of a universal counterexample the following remarkable fact: *if the support of the distribution  $\mu$  is a strict sub-interval of  $\mathbb{R}$  and if  $\mu$  is characterized by a Stein equation of the type*

$$a(x)f'(x) + (b - cx)f(x) = h(x) - \int hd\mu$$

for  $b, c$  real constants ( $c \neq 0$ ), then there is no finite constant  $M > 0$  such that  $\|f_h''\|_\infty \leq M\|h'\|_\infty$  holds for all Lipschitz-continuous functions  $h$  on  $\mathbb{R}$ . Here,  $f_h$  denotes the usual solution of the Stein equation. To the best of our expertise, this fact shows that the bounds on the solution  $f_h$  of (2) presented in Theorem 2.1 are the best that the technology of Stein's method can presently achieve for the Gamma distribution. Such a structural result also immediately entails that, as far as the Gamma distribution is concerned, one necessarily has to assume more smoothness on the test function  $h$ , in order to be able to work with second derivatives of the Stein's solution. We observe that, except on the Gaussian space, where bounds on the first derivative  $f_h'$  are sufficient due to the diffusiveness of the involved Ornstein–Uhlenbeck operator  $L$ , in the more general framework of Sections 1.2 and 1.3 one necessarily has to work with second derivatives as well, because of the intrinsic discrete nature of the considered objects.

- (iii) If one aims at less smooth distances – like the prominent Kolmogorov or Wasserstein distances – then one needs to implement some additional smoothing procedure. As it is typical, this inevitably comes at the price of a worse rate of convergence. The new Lemma 1.4 stated above provides such a smoothing result which, roughly speaking, bounds the Wasserstein distance of quite arbitrary distributions in terms of a distance induced by test functions which have one additional order of smoothness.
- (iv) We stress that, with the exception of the references [12] and [11], none of the references mentioned so far consider the Stein equation beyond the support interval of the corresponding distribution. For certain applications, this is indeed not necessary, because by applying some truncation procedure, one can force any random variable to have support in a given interval. However, for all three applications considered in this paper, applying truncation would immediately destroy the most important structural property of the random variables under consideration: In Section 1.2, the truncated random variable would no longer be a degenerate  $U$ -statistic of a given order and one would therefore have to work with a full Hoeffding decomposition; similarly, in the situations dealt with in

Sections 1.3–1.4 the chaotic decomposition of the truncated random variable would immediately be infinite, and thus, not directly amenable to computations. Since in general our random variables may have support equal to the whole real line, it is for us imperative to deal with the Stein equation (2) and its solution also outside the support of the target distribution.

- (v) Our main applications (see Theorem 1.1 and Corollary 1.2) concern the centered Gamma approximation of a degenerate, not necessarily symmetric  $U$ -statistic  $W$  of order  $d$ , based on some independent random sample  $X_1, \dots, X_n$  ( $n \geq d$ ); see Section 1.2. Here, we would like to stress that the classical results about the asymptotic distributions of  $U$ -statistics obtained in, for example, [17] and [47] (see also [48] and [24] for the case  $d = 2$ ) do not apply. First of all, our data random variables  $X_1, \dots, X_n$  are not necessarily i.i.d. Moreover, and even more importantly, our  $U$ -statistics are in general *non-symmetric*, and have kernels that in general *depend on  $n$*  (see equation (7) above). We conclude by pointing out that it is an open and challenging problem to determine the possible limits in distribution of general sequences of degenerate  $U$ -statistics of a fixed order  $\geq 2$ . Our Theorem 1.1 demonstrates the remarkable fact that the Gamma distribution emerges naturally for sequences of degenerate  $U$ -statistics of an arbitrary order, under minimal moment conditions and provided a Lindberg-type assumption is verified.

## 2. Stein’s method and exchangeable pairs for Gamma approximations

### 2.1. Main estimates for Gamma approximations

*Stein’s method* is a popular technique for estimating the distance between the distribution of some given random variable  $W$  and a usually better understood target distribution. It was first developed by Stein [49] for the standard normal distribution and has by now been extended to many other univariate distributions, like the Poisson (see, e.g., [2,6] or [4]), the Exponential (see, e.g., [5,43] and [19]), the Beta ([23] and [13]), the Gamma ([20,34,44] and [22]) and the Variance-Gamma (see [21]) distributions.

Stein’s method for the Gamma distribution was first considered by Luk [34]. There it was found that a real random variable  $X$  has the  $\Gamma(r, \lambda)$  distribution if and only if

$$\mathbb{E}[Xf'(X)] = -\mathbb{E}[(r - \lambda X)f(X)]$$

holds for a sufficiently rich class of functions  $f$ . Following Stein’s seminal idea, this led him to the Gamma Stein equation (2), which given the test function  $h$  on  $\mathbb{R}$  with  $\mathbb{E}|h(X_{r,\lambda})| < \infty$ , is to be solved for  $f$ . Usually, this equation is only considered and solved on the support  $[0, \infty)$  of  $\Gamma(r, \lambda)$  but for our purposes we will need a solution  $f_h$  to (2), which is defined on the whole real line. Here, by a solution of (2) we mean a function  $f$  on  $\mathbb{R}$  which is locally absolutely continuous and which satisfies (2) at those points at which it is in fact differentiable. Given such a function, contrary to the usual convention, we define  $f'$  at the non-zero points of non-differentiability of  $f$

by (2). If  $f$  is not differentiable at 0, then for definiteness, we let  $f'(0) := 0$ . For a test function  $h$  as above, a solution  $f_h$  to (2) and a given real-valued random variable  $W$  we thus obtain

$$|\mathbb{E}[h(W)] - \mathbb{E}[h(X_{r,\lambda})]| = |\mathbb{E}[Wf'_h(W) + (r - \lambda W)f_h(W)]|, \tag{18}$$

whenever the right-hand side is well-defined. As it turns out, the right-hand side of (18) may often be efficiently bounded by means of some additional tool exploiting the structure of the random quantity  $W$ . This might be a similar characterization for the law of  $W$ , an integration by parts formula on the space where  $W$  is defined, or a suitable coupling construction.

In any case, in order to bound the right-hand side of (18) it is crucial to have smoothness bounds on the solution  $f_h$  of (2) in terms of the test function  $h$ . One of the theoretical contributions of this paper is to provide a new set of such bounds which are valid for the solution  $f_h$  on the whole real line, not just on  $[0, \infty)$ . This is essential for our purposes, as the random variables  $W$  we consider in our applications do not necessarily have range included in the positive axis. Another consequence of our new bounds is an improvement of Theorem 2.1 from [42] and its consequences which deals with the Gamma approximation of functionals of a Poisson random measure.

To deal with our main application in this paper, we develop the technique of *exchangeable pairs* in the context of Gamma approximation. This coupling construction lies at the heart of Stein’s method and was first considered for normal approximation in Stein’s celebrated monograph [50]. In the recent paper [15], the authors applied it to the uni- and multivariate approximation of (vectors of) degenerate  $U$ -statistics. In particular, we were able to derive a complete quantitative extension of a famous CLT by de Jong [10].

In what follows, for a function  $f$  on  $\mathbb{R}$ , we denote by

$$\|f'\|_\infty := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \in [0, \infty) \cup \{+\infty\}$$

its minimum Lipschitz constant. This notation does not cause any confusion as it coincides with the supremum norm of the derivative of  $f$  whenever  $f$  is differentiable. Similarly, if  $f$  is  $n$ -times differentiable for some  $n \geq 1$ , we denote by  $\|f^{(n+1)}\|_\infty$  the minimum Lipschitz constant of  $f^{(n)}$ . We can now state our new smoothness estimates for the solution  $f_h$  of (2) on  $\mathbb{R}$ . We defer the proof of the next theorem to the end of this section.

**Theorem 2.1.**

- (a) *Let  $h$  be Lipschitz-continuous on  $\mathbb{R}$ . Then there exists a Lipschitz-continuous solution  $f_h$  of (2) on  $\mathbb{R}$ , which satisfies the bounds*

$$\|f_h\|_\infty \leq \lambda^{-1} \|h'\|_\infty \quad \text{and} \quad \|f'_h\|_\infty \leq 2 \max\left(1, \frac{1}{r}\right) \|h'\|_\infty. \tag{19}$$

- (b) *Suppose that  $h$  is continuously differentiable on  $\mathbb{R}$  and that both  $h$  and  $h'$  are Lipschitz-continuous. Then the solution  $f_h$  of (2) from (a) is continuously differentiable and its*

derivative  $f'_h$  is Lipschitz-continuous with minimum Lipschitz constant

$$\|f''_h\|_\infty \leq 4\lambda \max\left(1, \frac{1}{r}\right) \|h'\|_\infty + 2\|h''\|_\infty. \tag{20}$$

**Remark 2.2.**

- (a) By inspection of the proof of Theorem 2.1, one sees that the following refinement of (19) holds: writing  $f_h^+$  and  $f_h^-$  for the restriction of  $f_h$  to  $\mathbb{R}_+$  and  $\mathbb{R}_-$ , respectively, one has that

$$\|(f_h^+)' \|_\infty \leq 2\|h'\|_\infty, \quad \text{and} \quad \|(f_h^-)' \|_\infty \leq \frac{2}{r}\|h'\|_\infty. \tag{21}$$

We will see in Remark 2.8 that, in principle, the quantity  $2/r$  in the previous estimate *cannot* be replaced by a factor that is uniformly bounded in  $r$ .

- (b) Using the iterative technique for bounding higher derivatives of solutions to Stein equations from [13], which is further detailed in the recent paper [14], from the bound given in Theorem 2.1(b), we can easily derive the bound

$$\begin{aligned} \|f_h^{(k)}\|_\infty &\leq 2^k \lambda^{k-1} (k-1)! \max\left(1, \frac{1}{r}\right) \|h'\|_\infty \\ &\quad + \sum_{j=0}^{k-2} 2^{j+1} \lambda^j \frac{(k-1)!}{(k-j-1)!} \|h^{(k-j)}\|_\infty, \end{aligned} \tag{22}$$

valid for each  $k \geq 1$  and each  $(k-1)$ -times differentiable test function  $h$ , whose first  $k-1$  derivatives are bounded and whose  $(k-1)$ st derivative is Lipschitz-continuous.

- (c) In the recent paper [22], the authors proved that for each  $k \geq 1$  and each test function  $h$  from some specific sub-class  $\mathcal{C}_{\lambda,k}$  of all  $(k-1)$ -times differentiable functions such that  $h^{(k-1)}$  is still absolutely continuous, the following bound holds:

$$\sup_{x>0} |f_h^{(k)}(x)| \leq \frac{2}{r+k} \left( 3 \sup_{x>0} |h^{(k)}(x)| + 2\lambda \sup_{x>0} |h^{(k-1)}(x)| \right). \tag{23}$$

Note that, as opposed to the bounds from Theorem 2.1 or the bound (22), this bound converges to 0 whenever the shape parameter  $r$  of the Gamma distribution goes to  $\infty$ , which can be beneficial for certain applications as demonstrated in [22]. However, the bounds given in the present paper are valid on the whole real line and are thus applicable to a broader class of applications. We conjecture that there do exist positive, finite constants  $C_{r,\lambda}^{(1)}$  and  $C_{r,\lambda}^{(2)}$  with

$$\|f'_h\|_\infty \leq C_{r,\lambda}^{(1)} \|h'\|_\infty \quad \text{and} \quad \|f''_h\|_\infty \leq C_{r,\lambda}^{(2)} (\|h'\|_\infty + \|h''\|_\infty)$$

such that  $\lim_{r \rightarrow \infty} C_{r,\lambda}^{(1)} = \lim_{r \rightarrow \infty} C_{r,\lambda}^{(2)} = 0$ . These may be derived by a more careful investigation of the solutions  $f_h$  on the support interval  $[0, \infty)$ . On the other hand, as already

mentioned (see again Remark 2.8), the property  $\lim_{r \downarrow 0} C_{r,\lambda}^{(1)} = \infty$  is inevitable, as opposed to the bounds (23) for the solutions on  $(0, \infty)$ .

### 2.2. Targeting the centered Gamma distribution

Next, we transfer the bounds found in Theorem 2.1 to the centered Gamma distribution  $\bar{\Gamma}(\nu)$  of  $Z_\nu$  and state an off-the-shelf result, which bounds the distance between the distribution of a given random variable  $W$  and  $\bar{\Gamma}(\nu)$  in terms of an exchangeable pair. To the best of our knowledge this approach has not been considered in the context of Gamma approximation so far. The Stein equation for  $\bar{\Gamma}(\nu)$  we use is given by

$$2(x + \nu) f'(x) - x f(x) = h(x) - \mathbb{E}[h(Z_\nu)], \tag{24}$$

where  $h$  is Borel-measurable on  $\mathbb{R}$  with  $\mathbb{E}|h(Z_\nu)| < \infty$ .

**Theorem 2.3.**

- (a) *Let  $h$  be Lipschitz-continuous on  $\mathbb{R}$ . Then, there exists a Lipschitz-continuous solution  $f_h$  of (24) on  $\mathbb{R}$  which satisfies the bounds*

$$\|f_h\|_\infty \leq \|h'\|_\infty \quad \text{and} \quad \|f'_h\|_\infty \leq \max\left(1, \frac{2}{\nu}\right) \|h'\|_\infty.$$

- (b) *Suppose that  $h$  is continuously differentiable on  $\mathbb{R}$  such that both  $h$  and  $h'$  are Lipschitz-continuous. Then there is a continuously differentiable solution  $f_h$  of (24) on  $\mathbb{R}$  whose derivative  $f'_h$  is Lipschitz-continuous with minimum Lipschitz constant*

$$\|f''_h\|_\infty \leq \max\left(1, \frac{2}{\nu}\right) \|h'\|_\infty + \|h''\|_\infty.$$

**Remark 2.4.**

- (a) Plainly, refined bounds analogous to (21) can be obtained for the function  $f_h$  appearing in the statement of Theorem 2.3.
- (b) In [42], the slightly different Stein equation

$$2(x + \nu)_+ f'(x) - x f(x) = h(x) - \mathbb{E}[h(Z_\nu)] \tag{25}$$

with  $g_+ := \max(g, 0)$  was used. It turns out that the solution  $f_h$  of (24) from Theorem 2.3 has better smoothness properties at the singularity point  $x = -\nu$  of the Stein equation than the solution of (25) considered in [42]. This makes it possible for us to improve the bounds on Gamma approximation on the Poisson space provided there. Furthermore, for the application to  $U$ -statistics in the present paper, it is essential to have a linear coefficient function for  $f'$  in the Stein equation. This will become clear from the proof of Theorem 1.1 in Section 3.

- (c) The bounds from Theorem 2.3 have already been successfully applied in the recent article [16], where we proved central and non-central fourth moment theorems on the Poisson space.

**Proof of Theorem 2.3.** Given  $h$  with  $\mathbb{E}|h(Z_\nu)| < \infty$  we define  $h_1$  by  $h_1(x) := h(2x - \nu)$ . It is easy to see that if  $g_h$  is the solution to (2) with  $h$  replaced by  $h_1$  from Theorem 2.1, then the function  $f_h$  with

$$f_h(x) := \frac{1}{2} g_h\left(\frac{x + \nu}{2}\right)$$

solves (24). Furthermore, from Theorem 2.1 we obtain the bounds

$$\begin{aligned} \|f_h\|_\infty &= \frac{1}{2} \|g_h\|_\infty \leq \frac{1}{2} \|h'_1\|_\infty = \|h'\|_\infty, \\ \|f'_h\|_\infty &= \frac{1}{4} \|g'_h\|_\infty \leq \frac{1}{4} \cdot 2 \max\left(1, \frac{2}{\nu}\right) \|h'_1\|_\infty = \max\left(1, \frac{2}{\nu}\right) \|h'\|_\infty \quad \text{and} \\ \|f''_h\|_\infty &= \frac{1}{8} \|g''_h\|_\infty \leq \frac{1}{8} \left(2 \|h''_1\|_\infty + 4 \max\left(1, \frac{2}{\nu}\right) \|h'_1\|_\infty\right) \\ &= \|h''\|_\infty + \max\left(1, \frac{2}{\nu}\right) \|h'\|_\infty. \end{aligned} \quad \square$$

### 2.3. Exchangeable pairs

Let  $W, W'$  be identically distributed real-valued random variables defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that  $\mathbb{E}[W^2] < \infty$ . Assume that  $\mathcal{G}$  is a sub- $\sigma$ -field of  $\mathcal{F}$  such that  $\sigma(W) \subseteq \mathcal{G}$ . Given a real number  $\lambda > 0$ , we define the random variables  $R$  and  $S$  via the regression equations

$$\frac{1}{\lambda} \mathbb{E}[W' - W \mid \mathcal{G}] = -W + R \quad \text{and} \tag{26}$$

$$\frac{1}{2\lambda} \mathbb{E}[(W' - W)^2 \mid \mathcal{G}] = 2(W + \nu) + S. \tag{27}$$

In many cases of interest equation (26) holds with  $R = 0$  for some (unique)  $\lambda > 0$  but as was exemplified in [46] it is convenient to allow for a non-trivial remainder term  $R$ , in general. From Proposition 3.19 and Remark 3.10 of [13], as well as from the bounds given by Theorem 2.3 we obtain the following new plug-in result for centered Gamma approximation which can be seen as a generalization of Theorem 2.1 of [19] dealing with exponential approximation. This theorem will play a major role in the proof of Theorem 1.1.



**Theorem 2.5.** *Let  $W$  and  $W'$  be as above and assume that  $h$  is continuously differentiable on  $\mathbb{R}$  such that both  $h$  and  $h'$  are Lipschitz-continuous. Then we have the bound*

$$\begin{aligned} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z_\nu)]| &\leq \|h'\|_\infty (\max(1, 2\nu^{-1})\mathbb{E}|S| + \mathbb{E}|R|) \\ &\quad + \frac{\max(1, \frac{2}{\nu})\|h'\|_\infty + \|h''\|_\infty}{6\lambda} \mathbb{E}|W' - W|^3. \end{aligned}$$

If, moreover,  $\mathbb{E}[W^2] = 2\nu$  and (26) holds with  $R = 0$ , then since  $\mathbb{E}[S] = 0$  in this case, we also have the bound

$$\begin{aligned} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z_\nu)]| &\leq \max\left(1, \frac{2}{\nu}\right) \|h'\|_\infty \sqrt{\text{Var}(S)} \\ &\quad + \frac{\max(1, \frac{2}{\nu})\|h'\|_\infty + \|h''\|_\infty}{6\lambda} \mathbb{E}|W' - W|^3. \end{aligned} \tag{28}$$

### 2.4. Proofs

The following two lemmas will be needed for the proof of Theorem 2.1.

**Lemma 2.6.** *Let  $h$  be a Borel-measurable function  $h$  on  $\mathbb{R}$  with  $\mathbb{E}|h(X_{r,\lambda})| < \infty$ . Then, on each of the two intervals  $(-\infty, 0)$  and  $(0, \infty)$ , there exists at most one bounded solution  $f$  of (2).*

**Proof.** Let  $f$  be a bounded solution of (2) on  $(-\infty, 0)$ . The solutions of the corresponding homogeneous equation are given by the constant multiples of the function

$$\psi(x) := |x|^{-r} e^{\lambda x}, \quad x < 0.$$

Thus, if  $g$  is another solution of (2) on  $(-\infty, 0)$ , then there is a constant  $c \in \mathbb{R}$  such that

$$g = f + c\psi.$$

As  $\psi(0-) = -\infty$  and  $\sup_{x < 0} |f(x)| < \infty$ , it follows that  $g$  can only be bounded if  $c = 0$ , that is, if  $g = f$ . The proof for the interval  $(0, \infty)$  is very similar. □

**Lemma 2.7.** *Let  $a < b$  be real numbers and let  $f : [a, b] \rightarrow \mathbb{R}$  be a function having the following properties:*

- (a)  $f$  is continuous on  $[a, b]$ .
- (b)  $f|_{[c, b]}$  is absolutely continuous for each  $a < c < b$  (and hence,  $f$  is  $\lambda$ -almost everywhere differentiable on  $(a, b)$ ).
- (c) There is some  $a < d < b$ , a set  $A \subseteq (a, d)$  at each of whose points  $f$  is differentiable with  $\lambda((a, d) \setminus A) = 0$  and a real number  $\gamma$  such that  $\lim_{n \rightarrow \infty} f'(x_n) = \gamma$  for each sequence  $(x_n)_{n \in \mathbb{N}}$  lying in  $A$  with  $\lim_{n \rightarrow \infty} x_n = a$ .

Then  $f$  is absolutely continuous on  $[a, b]$  and differentiable at  $a$  with  $f'(a) = \gamma$ . Furthermore, the function  $f'$  restricted to  $A \cup \{a\}$  is continuous at  $a$ .

**Proof.** A proof can be found in the Appendix of the thesis [11]. □

We are now in the position to prove Theorem 2.1.

**Proof of Theorem 2.1.** It suffices to prove the theorem for the case  $\lambda = 1$ . In fact, it is easy to see that if  $g$  solves

$$xg'(x) + (r - x)g(x) = h_1(x) - \mathbb{E}[h_1(X_{r,1})],$$

where

$$h_1(x) := h(x/\lambda),$$

then  $f(x) := g(\lambda x)$  solves (2). Taking into account the identities

$$f^{(k)}(x) = \lambda^k g^{(k)}(\lambda x), \quad \|h'_1\|_\infty = \lambda^{-1} \|h'\|_\infty \quad \text{and} \quad \|h''_1\|_\infty = \lambda^{-2} \|h''\|_\infty$$

then yields the bounds for general  $\lambda > 0$ . So let us assume for the rest of the proof that  $\lambda = 1$ . For notational convenience, we will also write  $p_r$  for  $p_{r,1}$ ,  $F_r$  for  $F_{r,1}$  and  $X_r$  for  $X_{r,1}$ .

We first prove (a). As  $h$  is continuous, it is known (see, e.g., [13], Proposition 3.8) that the function  $f_h : (0, \infty) \rightarrow \mathbb{R}$  with

$$\begin{aligned} f_h^+(x) &:= \frac{1}{xp_r(x)} \int_0^x (h(t) - \mathbb{E}[h(X_r)]) p_r(t) dt \\ &= -\frac{1}{xp_r(x)} \int_x^\infty (h(t) - \mathbb{E}[h(X_r)]) p_r(t) dt \end{aligned} \tag{29}$$

is a continuously differentiable solution to (2) on  $(0, \infty)$  which can be continuously extended to  $[0, \infty)$  by letting

$$f_h^+(0) := \frac{h(0) - \mathbb{E}[h(X_r)]}{r}. \tag{30}$$

For a Lipschitz-continuous test function  $h$ , we know from Corollary 3.15 of [13] that  $\sup_{x \geq 0} |f_h^+(x)| \leq \|h'\|_\infty$  and that for each  $x > 0$  we have

$$|(f_h^+)'(x)| \leq 2\|h'\|_\infty \frac{\int_0^x F_r(t) dt \int_x^\infty (1 - F_r(t)) dt}{x^2 p_r(x)} =: 2\|h'\|_\infty \mathcal{S}_r(x).$$

We bound  $S_r(x)$  for  $0 < x \leq r$  and for  $x > r$  separately. Assume  $x > r$ . From Fubini's theorem, we conclude that

$$\begin{aligned} \int_x^\infty (1 - F_r(t)) dt &= \int_x^\infty \int_t^\infty p_r(s) ds dt = \int_x^\infty (s - x) p_r(s) ds \\ &= \int_x^\infty (s - r) p_r(s) ds + (r - x)(1 - F_r(x)) \\ &= x p_r(x) + (r - x)(1 - F_r(x)) \leq x p_r(x). \end{aligned}$$

Also note that

$$\int_0^x F_r(t) dt \leq x F_r(x), \quad x \geq 0, \tag{31}$$

as  $F_r$  is non-decreasing. Hence, we obtain that

$$S_r(x) \leq \frac{x F_r(x) \cdot x p_r(x)}{x^2 p_r(x)} = F_r(x) \leq 1, \quad r < x.$$

Now let  $0 < x \leq r$ . Note first that

$$\int_x^\infty (1 - F_r(t)) dt \leq \int_0^\infty (1 - F_r(t)) dt = \mathbb{E}[X_r] = r. \tag{32}$$

Next, consider

$$\begin{aligned} R(x) &:= \frac{\int_0^x F_r(t) dt}{x^2 p_r(x)} = \frac{x p_r(x) - (r - x) F_r(x)}{x^2 p_r(x)} \quad \text{with} \\ R'(x) &= \frac{x F_r(x) - (1 + r - x) \int_0^x F_r(t) dt}{x^3 p_r(x)} =: \frac{N(x)}{x^3 p_r(x)}. \end{aligned}$$

We claim that  $N(x) \geq 0$  for all  $x \geq 0$ , which implies that  $R$  is increasing. Note first that  $N(0) = 0$ . Also, we have

$$\begin{aligned} N'(x) &= F_r(x) + x p_r(x) + \int_0^x F_r(t) dt - (1 + r - x) F_r(x) \\ &= x p_r(x) - (r - x) F_r(x) + \int_0^x F_r(t) dt \\ &= 2 \int_0^x F_r(t) dt \geq 0. \end{aligned}$$

Hence,  $N$  is increasing and, thus,  $N(x) \geq 0$  for all  $x \geq 0$ . This implies that

$$\sup_{0 < x \leq r} R(x) = R(r) = \frac{1}{r}$$

and from (32) we conclude that

$$\sup_{0 < x \leq r} S_r(x) \leq r \sup_{0 < x \leq r} R(x) = \frac{r}{r} = 1.$$

Thus, we have proved that

$$\sup_{x > 0} |(f_h^+)'(x)| \leq 2 \|h'\|_\infty.$$

In order to solve (2) on  $(-\infty, 0)$  we use the theory developed in Section 2.4 of [11] (see also the unpublished manuscript [12]). There it is shown that a solution to (2) on  $(-\infty, 0)$  is given by

$$\begin{aligned} f_h^-(x) &= \exp(-G_l(x)) \int_0^x (h(t) - \mathbb{E}[h(X_r)]) \frac{\exp(G_l(t))}{t} dt \\ &= \frac{1}{x q_l(x)} \int_0^x (h(t) - \mathbb{E}[h(X_r)]) q_l(t) dt, \end{aligned} \tag{33}$$

where  $G_l$  is an arbitrary primitive function of  $x \mapsto \frac{r-x}{x}$  on  $(-\infty, 0)$  and

$$q_l(x) := \frac{\exp(G_l(x))}{x}, \quad x < 0.$$

Also  $f_h^-$  can be continuously extended to  $(-\infty, 0]$  by letting

$$f_h^-(0) := \frac{h(0) - \mathbb{E}[h(X_r)]}{r}. \tag{34}$$

This follows from Proposition 2.4.28 of [11] (or Proposition 2.22 of [12]). Again, by the continuity of  $h$ , it is easy to see that  $f_h^-$  is continuously differentiable on  $(-\infty, 0)$ . We choose

$$G_l(x) := r \log(-x) - x, \quad x < 0,$$

yielding

$$q_l(x) = -(-x)^{r-1} e^{-x}, \quad x < 0.$$

Note that  $q_l(x) < 0$  for all  $x \in (-\infty, 0)$ . Furthermore, we define the function  $Q_l$  on  $(-\infty, 0)$  by

$$Q_l(x) := \int_0^x q_l(t) dt = \int_x^0 (-q_l(t)) dt > 0.$$

By taking its derivative, we see that  $Q_l$  is decreasing on  $(-\infty, 0)$ . From Corollary 2.4.36 in [11] (or Corollary 2.28 of [12]), we have

$$\sup_{x < 0} |f_h^-(x)| \leq \|h'\|_\infty$$

and, for each  $x \in (-\infty, 0)$ , that

$$\begin{aligned} |(f_h^-)'(x)| &\leq 2\|h'\|_\infty \frac{(r-x)(-xQ_l(x) + \int_0^x tq_l(t) dt)}{-x^2q_l(x)} \\ &= 2\|h'\|_\infty \frac{(r-x) \int_x^0 Q_l(t) dt}{-x^2q_l(x)} \\ &=: 2\|h'\|_\infty U(x). \end{aligned} \tag{35}$$

Define the function  $T$  as well as  $q_u$  and  $Q_u$  on  $(0, \infty)$  by  $T(y) := U(-y)$ ,  $q_u(y) := -q_l(-y) = y^{r-1}e^y$  and

$$\begin{aligned} Q_u(y) &:= Q_l(-y) = \int_0^{-y} q_l(t) dt = -\int_0^y q_l(-s) ds \\ &= \int_0^y q_u(s) ds. \end{aligned}$$

Then we have

$$\begin{aligned} \int_{-y}^0 Q_l(t) dt &= -\int_y^0 Q_l(-s) ds = \int_0^y Q_u(s) ds = \int_0^y \int_0^t q_u(t) dt ds \\ &= \int_0^y (y-t)q_u(t) dt = yQ_u(y) - \int_0^y tq_u(t) dt \end{aligned}$$

as well as the representations

$$T(y) = \frac{(r+y) \int_0^y Q_u(s) ds}{y^2q_u(y)} = \frac{(r+y) \int_0^y Q_u(s) ds}{y^{r+1}e^y} \tag{36}$$

$$= \frac{(r+y) \int_0^y (y-t)q_u(t) dt}{y^{r+1}e^y} = \frac{(r+y)(yQ_u(y) - \int_0^y tq_u(t) dt)}{y^{r+1}e^y}. \tag{37}$$

Note that, using de l'Hôpital's rule, we obtain

$$\begin{aligned} \lim_{y \downarrow 0} T(y) &= r \lim_{y \downarrow 0} \frac{\int_0^y Q_u(s) ds}{y^{r+1}e^y} = r \lim_{y \downarrow 0} \frac{Q_u(y)}{y^r e^y (r+1+y)} \\ &= \frac{r}{r+1} \lim_{y \downarrow 0} \frac{Q_u(y)}{y^r e^y} = \frac{r}{r+1} \lim_{y \downarrow 0} \frac{y^{r-1}e^y}{y^{r-1}e^y (r+y)} \\ &= \frac{r}{r+1} \lim_{y \downarrow 0} \frac{1}{r+y} = \frac{1}{r+1} \end{aligned} \tag{38}$$

as well as

$$\lim_{y \rightarrow \infty} T(y) = \lim_{y \rightarrow \infty} \frac{r+y}{y} \cdot \lim_{y \rightarrow \infty} \frac{\int_0^y Q_u(s) ds}{y^r e^y} = \lim_{y \rightarrow \infty} \frac{Q_u(y)}{(r+y)y^{r-1}e^y}$$

$$\begin{aligned}
 &= \lim_{y \rightarrow \infty} \frac{y^{r-1}e^y}{r(r-1+y)y^{r-2}e^y + (r+y)y^{r-1}e^y} \\
 &= \lim_{y \rightarrow \infty} \frac{y}{r(r-1+y) + y(r+y)} = 0.
 \end{aligned}
 \tag{39}$$

By the continuity of  $T$ , from (38) and (39) we already conclude that

$$\sup_{x < 0} U(x) = \sup_{y > 0} T(y) < \infty.
 \tag{40}$$

Hence, it remains to deal with the local maxima of the function  $T$ . Note that

$$\begin{aligned}
 T'(y) &= \frac{y(r+y)Q_u(y) + \int_0^y Q_u(s) ds (y - (r+1+y)(r+y))}{y^{r+2}e^y} \\
 &= \frac{y(r+y)Q_u(y) - \int_0^y Q_u(s) ds (y^2 + 2ry + r^2 + r)}{y^{r+2}e^y}.
 \end{aligned}$$

If  $y_0 \in (0, +\infty)$  is a locally maximal point of  $T$ , then  $T'(y_0) = 0$ . This implies that

$$\begin{aligned}
 \int_0^{y_0} Q_u(s) ds &= \frac{y_0(r+y_0)Q_u(y_0)}{y_0^2 + 2ry_0 + r^2 + r} \quad \text{and} \\
 T(y_0) &= \frac{(r+y_0)^2 Q_u(y_0)}{y_0^r e^{y_0} (y_0^2 + 2ry_0 + r^2 + r)} \leq \frac{Q_u(y_0)}{y_0^r e^{y_0}}.
 \end{aligned}$$

Define the function  $T_2$  on  $(0, \infty)$  by

$$T_2(y) := \frac{Q_u(y)}{y^r e^y}.$$

Then the above discussion as well as (38) and (39) show that

$$\begin{aligned}
 \sup_{y > 0} T(y) &\leq \max\left(\lim_{y \downarrow 0} T(y), \lim_{y \rightarrow \infty} T(y), \sup_{y > 0} T_2(y)\right) \\
 &= \max\left(\frac{1}{r+1}, \sup_{y > 0} T_2(y)\right).
 \end{aligned}
 \tag{41}$$

Using the explicit expression of  $Q_u$ , as well as the elementary estimate  $1 \leq e^s \leq e^y$  for every  $0 \leq s \leq y$ , one deduces immediately that

$$\frac{e^{-y}}{r} \leq T_2(y) \leq \frac{1}{r}, \quad y > 0.
 \tag{42}$$

Hence, from (35), (40) and (42) we conclude that for all  $x \in (-\infty, 0)$  we have

$$|(f_h^-)'(x)| \leq \frac{2}{r} \|h'\|_\infty.
 \tag{43}$$

Now, we define the function  $f_h$  on  $\mathbb{R}$  by letting

$$f_h(x) := \begin{cases} f_h^-(x), & x \leq 0, \\ f_h^+(x), & x \geq 0. \end{cases}$$

Note that (30) and (34) imply that  $f_h$  is well-defined. As the continuous concatenation of two Lipschitz-continuous functions, we recognize  $f_h$  to be Lipschitz-continuous on  $\mathbb{R}$  with

$$\begin{aligned} \|f_h\|_\infty &\leq \max\left(\sup_{x \leq 0} |f_h^-(x)|, \sup_{x \geq 0} |f_h^+(x)|\right) \leq \|h'\|_\infty \quad \text{and} \\ \|f'_h\|_\infty &\leq \max\left(\sup_{x < 0} |(f_h^-)'(x)|, \sup_{x > 0} |(f_h^+)'(x)|\right) \leq \max\left(2, \frac{2}{r}\right) \|h'\|_\infty. \end{aligned}$$

This completes the proof of (a).

To prove (b), we assume that  $h$  is continuously differentiable and that both  $h$  and its derivative  $h'$  are Lipschitz-continuous. The identity

$$x f'_h(x) + (r - x) f_h(x) = h(x) - \mathbb{E}[h(X_r)]$$

implies that  $f_h$  is continuously differentiable on both of the two intervals  $(-\infty, 0)$  and  $(0, \infty)$  and differentiating yields that  $f'_h$  solves

$$x g'(x) + (r + 1 - x) g(x) = h'(x) + f_h(x) =: h_2(x) \tag{44}$$

on both intervals  $(-\infty, 0)$  and  $(0, \infty)$ . Note that (44) is the Stein equation for the distribution  $\Gamma(r + 1, 1)$  corresponding to the test function  $h_2$ . We already know from part (a) that  $f'_h$  is bounded by  $2 \max(1, r^{-1}) \|h'\|_\infty$ . Also, Proposition 3.17 of [13] implies that  $h_2$  is centered with respect to the  $\Gamma(r + 1, 1)$  distribution. Hence, as  $h_2$  is Lipschitz-continuous with Lipschitz-constant

$$\|h_2\|_\infty \leq \|h''\|_\infty + \|f'_h\|_\infty \leq \|h''\|_\infty + 2 \max(1, r^{-1}) \|h'\|_\infty \tag{45}$$

we know from part (a) applied to the distribution  $\Gamma(r + 1, 1)$  that there is a bounded solution  $g_{h_2}$  of (44). Since  $f'_h$  is bounded on both of the intervals  $(-\infty, 0)$  and  $(0, \infty)$ , it thus follows from Lemma 2.6 that  $f'_h(x) = g_{h_2}(x)$  for all  $x \neq 0$ . Since  $g_{h_2}$  is continuous at 0 we know from the analogs of (30) and (34) for  $g_{h_2}$  that

$$\begin{aligned} \lim_{x \uparrow 0} f'_h(x) &= \lim_{x \uparrow 0} g_{h_2}(x) = g_{h_2}(0) = \lim_{x \rightarrow 0} g_{h_2}(x) = \frac{h_2(0) - \mathbb{E}[h_2(X_{r+1,1})]}{r + 1} = \frac{h_2(0)}{r + 1} \\ &= \frac{h'(0)}{r + 1} + \frac{h(0) - \mathbb{E}[h(X_{r,1})]}{r(r + 1)}, \end{aligned}$$

and, similarly, we conclude that

$$\lim_{x \downarrow 0} f'_h(x) = \lim_{x \downarrow 0} g_{h_2}(x) = g_{h_2}(0) = \frac{h'(0)}{r + 1} + \frac{h(0) - \mathbb{E}[h(X_{r,1})]}{r(r + 1)}.$$

By Lemma 2.7, this implies that  $f_h$  is continuously differentiable on  $\mathbb{R}$  with  $f'_h(0) = g_{h_2}(0)$ . Hence, we have  $f'_h = g_{h_2}$  and from part (a) and (45) we conclude that  $f'_h$  is Lipschitz-continuous with

$$\begin{aligned} \|f''_h\|_\infty &= \|g'_{h_2}\|_\infty \leq 2 \max\left(1, \frac{1}{r+1}\right) \|h'_2\|_\infty = 2\|h'_2\|_\infty \\ &\leq 2(\|h''\|_\infty + 2 \max(1, r^{-1}) \|h'\|_\infty) \\ &= 2\|h''\|_\infty + 4 \max(1, r^{-1}) \|h'\|_\infty. \end{aligned} \quad \square$$

**Remark 2.8.** As anticipated, the factor  $2/r$  in (21) cannot be replaced by a quantity that is uniformly bounded in  $r$ . Indeed, from the proof of Proposition 2.4.35 in [11] we have the representation

$$f'_h(x) = \frac{1}{-x^2 q_l(x)} \left( (x-r) \int_x^0 Q_l(s) h'(s) ds - \int_x^\infty (1-F(s)) h'(s) ds \int_x^0 Q_l(t) dt \right),$$

whenever  $h$  is Lipschitz-continuous and  $x < 0$ . If we take  $h(x) = \min(x, 0)$ , then we obtain

$$\begin{aligned} \int_x^0 Q_l(s) h'(s) ds &= \int_0^x Q_l(s) ds \quad \text{and} \\ - \int_x^\infty (1-F(s)) h'(s) ds \int_x^0 Q_l(t) dt &= x \int_x^0 Q_l(t) dt \end{aligned}$$

because  $F(s) = 0$  for all  $s \leq 0$ . This gives

$$f'_h(x) = \frac{(2x-r) \int_x^0 Q_l(t) dt}{-x^2 q_l(x)}$$

and hence, straightforward estimates yield that, for  $x < 0$ ,

$$|f'_h(x)| = \frac{(r-2x) \int_x^0 Q_l(t) dt}{-x^2 q_l(x)} \geq e^x \frac{r-2x}{r(r+1)}.$$

In particular, this implies that

$$\sup_{x < 0} |f'_h(x)| \geq |f'_h(-1/2)| \geq e^{-1/2} \frac{1}{r}.$$

By mollifying the Lipschitz function  $h(x) = \min(x, 0)$ , one can also construct a function  $h^* \in \mathcal{H}_2$  such that  $\sup_{x < 0} |f'_{h^*}(x)| \geq e^{-1/2} \frac{1}{2r}$ . This justifies the remark following Theorem 1.1.

### 3. Proof of Theorem 1.1

We are going to apply the bound (28) from Theorem 2.5 with  $\|h'\|_\infty, \|h''\|_\infty \leq 1$  to the  $\sigma$ -field  $\mathcal{G} = \sigma(X_1, \dots, X_n)$  and to the following exchangeable pair  $(W, W')$ , which has already



been used in [15]: Let  $Y := (Y_j)_{1 \leq j \leq n}$  be an independent copy of  $X := (X_j)_{1 \leq j \leq n}$  and let  $\alpha$  be uniformly distributed on  $\{1, \dots, n\}$  such that  $X, Y$  and  $\alpha$  are jointly independent. Letting, for  $j = 1, \dots, n$ ,

$$X'_j := \begin{cases} Y_j, & \text{if } \alpha = j, \\ X_j, & \text{if } \alpha \neq j, \end{cases}$$

and

$$X' := (X'_1, \dots, X'_n)$$

it is easy to see that the pair  $(X, X')$  is exchangeable. Finally, as exchangeability is preserved under functions, letting

$$\begin{aligned} W' &:= \psi(X'_1, \dots, X'_n) = \sum_{j=1}^n 1_{\{\alpha=j\}} \left( \sum_{J: j \notin J} W_J + \sum_{J: j \in J} W_J^{(j)} \right) \\ &=: \sum_{J: \alpha \notin J} W_J + \sum_{J: \alpha \in J} W_J^{(\alpha)}, \end{aligned}$$

the pair  $(W, W')$  is exchangeable. Here, for  $J = \{j_1, \dots, j_m\} \subseteq [n]$  with  $1 \leq j_1 < j_2 < \dots < j_m \leq n$  and  $j = j_k \in J$ , we write

$$W_J^{(j)} := \psi_J(X_{j_1}, \dots, X_{j_{k-1}}, Y_{j_k}, X_{j_{k+1}}, \dots, X_{j_m}).$$

From Lemma 2.2 of [15], we know that (26) holds for the pair  $(W, W')$  with  $R = 0$  and  $\lambda = d/n$ . Also, if we denote by

$$W^2 = \sum_{\substack{M \subseteq [n]: \\ |M| \leq 2d}} U_M$$

the Hoeffding decomposition of  $W^2$ , then Lemma 2.6 of [15] states that

$$\frac{n}{2d} \mathbb{E}[(W' - W)^2 | X] = \sum_{\substack{M \subseteq [n]: \\ |M| \leq 2d-1}} a_M U_M, \tag{46}$$

where

$$a_M := 1 - \frac{|M|}{2d}, \quad M \subseteq [n] \quad \text{such that } |M| \leq 2d.$$

Hence, we have the following Hoeffding decomposition (47) of  $S$ :

$$\begin{aligned}
 S &= \frac{n}{2d} \mathbb{E}[(W' - W)^2 | X] - 2(W + \nu) = \sum_{\substack{M \subseteq [n]: \\ |M| \leq 2d-1}} \left(1 - \frac{|M|}{2d}\right) U_M - 2W - 2\nu \\
 &= \sum_{\substack{M \subseteq [n]: \\ 1 \leq |M| \leq 2d-1}} \left(1 - \frac{|M|}{2d}\right) U_M - 2 \sum_{J \in \mathcal{D}_d} W_J \\
 &= \sum_{\substack{M \subseteq [n]: \\ 1 \leq |M| \leq 2d-1, \\ |M| \neq d}} \left(1 - \frac{|M|}{2d}\right) U_M + \frac{1}{2} \sum_{J \in \mathcal{D}_d} (U_J - 4W_J) \\
 &=: S_1 + \frac{1}{2} S_2.
 \end{aligned} \tag{47}$$

Here, we have used that  $U_\emptyset = \mathbb{E}[W^2] = 2\nu$ . By the orthogonality of the terms appearing in the Hoeffding decomposition, we thus obtain that

$$\text{Var}(S) = \text{Var}(S_1) + \frac{1}{4} \text{Var}(S_2).$$

From the orthogonality of the Hoeffding decomposition, we conclude that

$$\mathbb{E}[W^3] = \sum_{J \in \mathcal{D}_d} \sum_{\substack{M \subseteq [n]: \\ |M| \leq 2d}} \mathbb{E}[W_J U_M] = \sum_{J \in \mathcal{D}_d} \mathbb{E}[W_J U_J] = \sum_{J \in \mathcal{D}_d} \text{Cov}(W_J, U_J). \tag{48}$$

Before we proceed, we need an auxiliary lemma which expresses the fourth moment of  $W$  in terms of the exchangeable pair  $(W, W')$ . We first state a more general lemma, whose statement is in fact only a slight generalization of one of the key relations in Stein’s method of exchangeable pairs (see [50] or [7]) will be very useful.

**Lemma 3.1.** *Let  $(W, W')$  be an exchangeable pair of real-valued random variables such that, for some  $\lambda > 0$ , (26) is satisfied with  $R = 0$  and let  $g$  be an absolutely continuous function on  $\mathbb{R}$  with  $\mathbb{E}[(1 + |W| + |W'|)|g(W)] < \infty$ . Then*

$$\mathbb{E}[Wg(W)] = \mathbb{E}\left[g'(W) \frac{1}{2\lambda} \mathbb{E}[(W' - W)^2 | \mathcal{G}]\right] + E,$$

where

$$E := \frac{1}{2\lambda} \mathbb{E}\left[(W' - W)^2 \int_0^1 (g'(W + t(W' - W)) - g'(W)) dt\right]$$

is a remainder term.

**Lemma 3.2.** *Let  $(W, W')$  be an exchangeable pair of real-valued random variables in  $L^4(\mathbb{P})$  such that, for some  $\lambda > 0$ , (26) is satisfied with  $R = 0$ . Then*

$$\mathbb{E}[W^4] = 3\mathbb{E}\left[W^2 \frac{1}{2\lambda} \mathbb{E}[(W' - W)^2 \mid \mathcal{G}]\right] - \frac{1}{4\lambda} \mathbb{E}[(W' - W)^4] \quad \text{and} \quad (49)$$

$$\mathbb{E}[W^3] = 2\mathbb{E}\left[W \frac{1}{2\lambda} \mathbb{E}[(W' - W)^2 \mid \mathcal{G}]\right]. \quad (50)$$

**Proof.** The proof of (49) applies Lemma 3.1 with  $g(x) = x^3$  leading to the remainder term

$$\begin{aligned} E &= \frac{3}{2\lambda} \mathbb{E}\left[(W' - W)^2 \int_0^1 (2tW(W' - W) + t^2(W' - W)^2) dt\right] \\ &= \frac{3}{2\lambda} \mathbb{E}[W(W' - W)^3] + \frac{1}{2\lambda} \mathbb{E}[(W' - W)^4]. \end{aligned}$$

By exchangeability, we obtain that

$$\mathbb{E}[(W' - W)^4] = \mathbb{E}[W'(W' - W)^3] - \mathbb{E}[W(W' - W)^3] = -2\mathbb{E}[W(W' - W)^3]$$

yielding the claim. In order to prove (50), we apply Lemma 3.1 with  $g(x) = x^2$  leading to the remainder term

$$E = \frac{1}{\lambda} \mathbb{E}\left[(W' - W)^2 \int_0^1 t(W' - W) dt\right] = \frac{1}{2\lambda} \mathbb{E}[(W' - W)^3] = 0$$

again by exchangeability. □

Now, using (49) we obtain

$$\begin{aligned} \mathbb{E}[W^4] &= 3\mathbb{E}\left[W^2 \frac{n}{2d} \mathbb{E}[(W' - W)^2 \mid X]\right] - \frac{n}{4d} \mathbb{E}[(W' - W)^4] \\ &= 3 \sum_{\substack{M, N \subseteq [n]: \\ |M|, |N| \leq 2d}} \left(1 - \frac{|M|}{2d}\right) \mathbb{E}[U_N U_M] - \frac{n}{4d} \mathbb{E}[(W' - W)^4] \\ &= 12\nu^2 + 3 \sum_{\substack{M \subseteq [n]: \\ 1 \leq |M| \leq 2d-1}} \left(1 - \frac{|M|}{2d}\right) \text{Var}(U_M) - \frac{n}{4d} \mathbb{E}[(W' - W)^4], \end{aligned} \quad (51)$$

where we have used that  $U_\emptyset = \mathbb{E}[W^2] = 2\nu$ . We have

$$\begin{aligned} \text{Var}(S_2) &= \sum_{J \in \mathcal{D}_d} \text{Var}(U_J - 4W_J) = \sum_{J \in \mathcal{D}_d} (\text{Var}(U_J) + 16 \text{Var}(W_J) - 8 \text{Cov}(U_J, W_J)) \\ &= \sum_{J \in \mathcal{D}_d} \text{Var}(U_J) + 32\nu - 8 \sum_{J \in \mathcal{D}_d} \text{Cov}(U_J, W_J) = \sum_{J \in \mathcal{D}_d} \text{Var}(U_J) + 32\nu - 8\mathbb{E}[W^3], \end{aligned}$$

where the last equality holds by virtue of (48). Hence, we have

$$\begin{aligned} \text{Var}(S) &= \text{Var}(S_1) + \frac{1}{4} \text{Var}(S_2) \\ &= \sum_{\substack{M \subseteq [n]: \\ 1 \leq |M| \leq 2d-1, \\ |M| \neq d}} \left(1 - \frac{|M|}{2d}\right)^2 \text{Var}(U_M) + \frac{1}{4} \sum_{J \in \mathcal{D}_d} \text{Var}(U_J) + 8\nu - 2\mathbb{E}[W^3]. \end{aligned} \tag{52}$$

From (51) and (52), using

$$\left(1 - \frac{|M|}{2d}\right)^2 \leq \left(1 - \frac{|M|}{2d}\right) \quad \text{for all } M \subseteq [n] \text{ such that } |M| \leq 2d,$$

we see that

$$\begin{aligned} &\mathbb{E}[W^4] - 12\mathbb{E}[W^3] - 12\nu^2 + 48\nu \\ &= 3 \sum_{\substack{M \subseteq [n]: \\ 1 \leq |M| \leq 2d-1}} \left(1 - \frac{|M|}{2d}\right) \text{Var}(U_M) - 12\mathbb{E}[W^3] + 48\nu - \frac{n}{4d} \mathbb{E}[(W' - W)^4] \\ &= 3 \sum_{\substack{M \subseteq [n]: \\ 1 \leq |M| \leq 2d-1, \\ |M| \neq d}} \left(1 - \frac{|M|}{2d}\right) \text{Var}(U_M) + \frac{3}{2} \sum_{J \in \mathcal{D}_d} \text{Var}(U_J) - 12\mathbb{E}[W^3] + 48\nu \\ &\quad - \frac{n}{4d} \mathbb{E}[(W' - W)^4] \\ &\geq 3 \left( \text{Var}(S_1) + \frac{1}{2} \left( \sum_{J \in \mathcal{D}_d} \text{Var}(U_J) - 8\mathbb{E}[W^3] + 32\nu \right) \right) - \frac{n}{4d} \mathbb{E}[(W' - W)^4] \\ &= 3 \left( \text{Var}(S_1) + \frac{1}{2} \text{Var}(S_2) \right) - \frac{n}{4d} \mathbb{E}[(W' - W)^4] \\ &\geq 3 \text{Var}(S) - \frac{n}{4d} \mathbb{E}[(W' - W)^4]. \end{aligned}$$

Hence, we obtain that

$$\text{Var}(S) \leq \frac{1}{3} (\mathbb{E}[W^4] - 12\mathbb{E}[W^3] - 12\nu^2 + 48\nu) + \frac{n}{12d} \mathbb{E}[(W' - W)^4] \tag{53}$$

and it thus remains to find a bound on

$$\frac{n}{d} \mathbb{E}[(W' - W)^4].$$

From the definition of the coupling  $(W, W')$  and by the inequality  $(a + b)^4 \leq 8(a^4 + b^4)$ , we conclude that

$$\begin{aligned}
 \mathbb{E}|W' - W|^4 &= \mathbb{E} \left| \sum_{J \in \mathcal{D}_d: \alpha \in J} (W_J^{(\alpha)} - W_J) \right|^4 = \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \left( \sum_{J \in \mathcal{D}_d: j \in J} (W_J^{(j)} - W_J) \right)^4 \right] \\
 &\leq \frac{8}{n} \sum_{j=1}^n \mathbb{E} \left[ \left( \sum_{J \in \mathcal{D}_d: j \in J} W_J^{(j)} \right)^4 + \left( \sum_{J \in \mathcal{D}_d: j \in J} W_J \right)^4 \right] \\
 &= \frac{16}{n} \sum_{j=1}^n \mathbb{E} \left[ \left( \sum_{J \in \mathcal{D}_d: j \in J} W_J \right)^4 \right] = \frac{16}{n} \sum_{j=1}^n \sum_{\substack{J, K, L, M \in \mathcal{D}_d: \\ j \in J \cap K \cap L \cap M}} \mathbb{E}[W_J W_K W_L W_M] \tag{54} \\
 &= \frac{16}{n} \sum_{\substack{(J, K, L, M) \in \mathcal{D}_d^4: \\ J \cap K \cap L \cap M \neq \emptyset}} |J \cap K \cap L \cap M| \mathbb{E}[W_J W_K W_L W_M].
 \end{aligned}$$

Here, we have used the fact that the sums

$$\sum_{J \in \mathcal{D}_d: j \in J} W_J^{(j)} \quad \text{and} \quad \sum_{J \in \mathcal{D}_d: j \in J} W_J$$

are identically distributed for each  $j \in [n]$ . Now, by the definition of  $D$  and by the generalized Hölder inequality, for each  $(J, K, L, M) \in \mathcal{D}_d^4$  we have

$$\begin{aligned}
 |\mathbb{E}[W_J W_K W_L W_M]| &\leq (\mathbb{E}[W_J^4] \mathbb{E}[W_K^4] \mathbb{E}[W_L^4] \mathbb{E}[W_M^4])^{1/4} \\
 &\leq (D\sigma_J^4 D\sigma_K^4 D\sigma_L^4 D\sigma_M^4)^{1/4} \leq D\sigma_J \sigma_K \sigma_L \sigma_M.
 \end{aligned}$$

Proposition 2.9 of [15] implies that

$$\sum_{\substack{(J, K, L, M) \in \mathcal{D}_d^4: \\ J \cap K \cap L \cap M \neq \emptyset}} \sigma_J \sigma_K \sigma_L \sigma_M \leq C_d \rho_n^2,$$

where the finite constant  $C_d$  only depends on  $d$ . Thus, from (54) we conclude that

$$\mathbb{E}|W' - W|^4 \leq \frac{16}{n} C_d D_n \rho_n^2. \tag{55}$$

From (53) and (55), we have

$$\text{Var}(S) \leq \frac{1}{3} |\mathbb{E}[W^4] - 12\mathbb{E}[W^3] - 12\nu^2 + 48\nu| + \frac{4}{3d} C_d D_n \rho_n^2. \tag{56}$$

Also, from the fact that

$$\mathbb{E}[(W' - W)^2] = 2\lambda \mathbb{E}[W^2] = \frac{4d\nu}{n}$$

and using the Cauchy–Schwarz inequality we obtain

$$\begin{aligned} \frac{1}{6\lambda} \mathbb{E}|W' - W|^3 &\leq \frac{n}{6d} (\mathbb{E}[(W' - W)^2])^{1/2} (\mathbb{E}|W' - W|^4)^{1/2} \\ &= \frac{\sqrt{v}}{3\sqrt{d}} (n\mathbb{E}|W' - W|^4)^{1/2} \leq \frac{\sqrt{v}}{3\sqrt{d}} \sqrt{16C_d D_n \rho_n^2} \\ &= \frac{4\sqrt{v}}{3\sqrt{d}} \sqrt{C_d D_n \rho_n^2}, \end{aligned} \tag{57}$$

where we have used (55) again. Theorem 1.1 now follows from (28), (56) and (57).

### 4. Proof of Theorem 1.5

For the sake of completeness, we will discuss some further details concerning stochastic analysis for functionals of a Poisson measure. Throughout the section, we work in the same framework as the one outlined in Section 1.3.

For an integer  $p \geq 1$ , we denote by  $L^2(\mu^p)$  the Hilbert space of all square-integrable real-valued functions on  $\mathcal{Z}^p$  and we write  $L^2_s(\mu^p)$  for the subspace of those functions in  $L^2(\mu^p)$  which are  $\mu^p$ -a.e. symmetric. Moreover, for ease of notation, we denote by  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$  the usual norm and scalar product on  $L^2(\mu^p)$  for whatever value of  $p$ . We further define  $L^2(\mu^0) := \mathbb{R}$ . For  $f \in L^2(\mu^p)$ , we denote by  $I_p(f)$  the *multiple Wiener–Itô integral* of  $f$  with respect to  $\hat{\eta}$ . If  $p = 0$ , then, by convention,  $I_0(c) := c$  for each  $c \in \mathbb{R}$ . The following properties of multiple integrals are standard for all  $p, q \geq 0$ :

- (1)  $I_p(f) = I_p(\tilde{f})$ , where  $\tilde{f}$  denotes the *canonical symmetrization* of  $f \in L^2(\mu^p)$ .
- (2)  $I_p(f) \in L^2(\mathbb{P})$ .
- (3)  $\mathbb{E}[I_p(f)I_q(g)] = \delta_{p,q} p! \langle \tilde{f}, \tilde{g} \rangle$ , where  $\delta_{p,q}$  denotes *Kronecker’s delta symbol*.

For  $p \geq 0$ , the Hilbert space consisting of all random variables  $I_p(f)$ ,  $f \in L^2(\mu^p)$ , is called the *p*th Wiener chaos associated with  $\eta$ . It is a crucial fact that every  $F \in L^2(\mathbb{P})$  admits a unique representation

$$F = \mathbb{E}[F] + \sum_{p=1}^{\infty} I_p(f_p), \tag{58}$$

where  $f_p \in L^2_s(\mu^p)$ ,  $p \geq 1$ , are suitable symmetric kernel functions. Identity (58) is called the *chaotic decomposition* of the functional  $F \in L^2(\mathbb{P})$ . Next, we briefly introduce the necessary Malliavin operators. The *domain*  $\text{dom } D$  of the Malliavin derivative operator  $D$  is the set of all  $F \in L^2(\mathbb{P})$  such that the chaotic decomposition (58) of  $F$  satisfies  $\sum_{p=1}^{\infty} p p! \|f_p\|^2 < \infty$ . For such an  $F$ , the random function  $\mathcal{Z} \ni z \mapsto D_z F \in L^2(\mathbb{P})$  is defined via

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

where  $f_p(z, \cdot)$  is an a.e. symmetric function on  $\mathcal{Z}^{p-1}$ . Hence,  $DF = (D_z F)_{z \in \mathcal{Z}}$  can be viewed as an element of  $L^2(\Omega \times \mathcal{Z}, \mathcal{F} \otimes \mathcal{Z}, \mathbb{P} \otimes \mu)$ . Note that, as  $\mathcal{F} = \sigma(\eta)$ , each  $F \in L^2(\mathbb{P})$  can be written as  $F = g(\eta)$  for some measurable functional  $g$ . Then, for  $z \in \mathcal{Z}$  we write  $F_z := g(\eta + \delta_z)$ . If, furthermore,  $F$  happens to be in  $\text{dom } D$ , then it is known that for  $\mu$ -almost every  $z \in \mathcal{Z}$  we have the important formula

$$D_z F = F_z - F. \tag{59}$$

The domain  $\text{dom } L$  of the Ornstein–Uhlenbeck generator  $L$  is the set of those  $F \in L^2(\mathbb{P})$  whose chaotic decomposition (58) verifies  $\sum_{p=1}^\infty p^2 p! \|f_p\|^2 < \infty$  and, for  $F \in \text{dom } L$ , one defines

$$LF = - \sum_{p=1}^\infty p I_p(f_p).$$

By definition,  $\mathbb{E}[LF] = 0$ . The domain  $\text{dom } L^{-1}$  of the pseudo-inverse  $L^{-1}$  of  $L$  is the class of mean zero elements  $F$  of  $L^2(\mathbb{P})$ . If  $F = \sum_{p=1}^\infty I_p(f_p)$  is the chaotic decomposition of such an  $F$ , then it is defined via

$$L^{-1}F = \sum_{p=1}^\infty \frac{1}{p} I_p(f_p).$$

Note that these definitions imply that

$$LL^{-1}F = F \quad \text{for all } F \in \text{dom } L^{-1} \quad \text{and} \quad L^{-1}LF = F - \mathbb{E}[F] \quad \text{for all } F \in \text{dom } L.$$

Finally, we review the definition Skohorod integral operator  $\delta$ . Note that for each  $u \in L^2(\Omega \times \mathcal{Z}, \mathcal{F} \otimes \mathcal{Z}, \mathbb{P} \otimes \mu)$  and each fixed  $z \in \mathcal{Z}$  we have a chaotic decomposition of the type

$$u_z = \sum_{p=0}^\infty I_p(f_p(z, \cdot)) \tag{60}$$

and, for  $p \geq 0$ , the kernel  $f_p(z, \cdot)$  is an element of  $L^2_\mathcal{Z}(\mu^p)$ . Then the domain  $\text{dom } \delta$  of  $\delta$  consists of those such  $u \in L^2(\Omega \times \mathcal{Z}, \mathcal{F} \otimes \mathcal{Z}, \mathbb{P} \otimes \mu)$  whose kernels given by (60) satisfy

$$\sum_{p=0}^\infty (p+1)! \|f_p\|^2_{L^2(\mu^{p+1})} < \infty$$

and, for  $u \in \text{dom } \delta$ , one lets

$$\delta(u) = \sum_{p=0}^\infty I_{p+1}(f_p).$$

The following two identities are essential for the Malliavin–Stein method on the Poisson space. The first one, the integration by parts formula, characterizes  $\delta$  as the adjoint operator of  $D$ :

$$\mathbb{E}[G\delta(u)] = \mathbb{E}[\langle DG, u \rangle_{L^2(\mu)}] \quad \text{for all } G \in \text{dom } D, u \in \text{dom } \delta, \tag{61}$$

$$\delta DF = -LF \quad \text{for all } F \in \text{dom } L. \tag{62}$$

We are now in the position to prove our new bounds on the Gamma approximation for functionals on the Poisson space.

**Proof of Theorem 1.5.** The proof follows the lines of the proof of Theorem 3.1 of [41] very closely. Fix  $h \in \mathcal{H}_2$  and write  $f = f_h$  for the solution to the Stein equation (24) from Theorem 2.3. Using the fact that  $\mathbb{E}[F] = 0$  as well as (61) and (62), we have

$$\begin{aligned} \mathbb{E}[Ff(F)] &= \mathbb{E}[(LL^{-1}F)f(F)] = \mathbb{E}[-\delta(DL^{-1}F)f(F)] \\ &= \mathbb{E}[\langle Df(F), -DL^{-1}F \rangle_{L^2(\mu)}]. \end{aligned} \tag{63}$$

Now, for fixed  $z \in \mathcal{Z}$ , using (59) as well as Taylor’s formula, we have

$$\begin{aligned} D_z f(F) &= (f(Z))_z - f(F) = f(F_z) - f(F) \\ &= f'(F)(F_z - F) + R(F_z - F) = f'(F)(D_z F) + R(D_z F), \end{aligned} \tag{64}$$

where  $y \mapsto R(y)$  is a function which satisfies

$$|R(y)| \leq \frac{\|f''\|_\infty}{2} y^2 \leq \frac{\max(1, \frac{2}{\nu})\|h'\|_\infty + \|h''\|_\infty}{2} y^2 \leq \max\left(1, \frac{1}{\nu} + \frac{1}{2}\right) y^2$$

by Theorem 2.3(b). Hence, from (63) and (64) we conclude that

$$\mathbb{E}[Ff(F)] = \mathbb{E}[f'(F)\langle DF, -DL^{-1}F \rangle_{L^2(\mu)}] + \mathbb{E}[\langle R(DF), -DL^{-1}F \rangle_{L^2(\mu)}]$$

yielding

$$\begin{aligned} |\mathbb{E}[h(F)] - \mathbb{E}[h(Z_\nu)]| &= |\mathbb{E}[2(F + \nu)f'(F) - Ff(F)]| \\ &\leq |\mathbb{E}[f'(F)(2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)})]| + |\mathbb{E}[\langle R(DF), -DL^{-1}F \rangle_{L^2(\mu)}]| \\ &\leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E}|2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| + \int_{\mathcal{Z}} \mathbb{E}|R(D_z F)D_z L^{-1}F| \mu(dz) \\ &\leq \max\left(1, \frac{2}{\nu}\right) \mathbb{E}|2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}| \\ &\quad + \max\left(1, \frac{1}{\nu} + \frac{1}{2}\right) \int_{\mathcal{Z}} \mathbb{E}|D_z F|^2 |D_z L^{-1}F| \mu(dz), \end{aligned}$$

which in turn gives (10). Applying Cauchy–Schwarz on (10) gives (11). The bounds (12) and (13) easily follow from these by the definitions of the Malliavin operators.  $\square$

## 5. Proofs of Theorem 1.7 and of Proposition 1.8

**Proof of Theorem 1.7.** We have to show that, for every 1-Lipschitz test function  $h$ , the quantity  $|\mathbb{E}[h(F)] - \mathbb{E}[h(Z_\nu)]|$  is bounded by the right-hand side of (14). We start by assuming that  $h$



is twice continuously differentiable and such that  $\|h'\|_\infty \leq 1$ . Then we can use Theorem 2.3 to deduce that there exists a solution  $f_h$  to (24) such that  $f_h$  is continuously differentiable, and  $\|f'_h\|_\infty \leq \max(1, \frac{2}{\nu})\|h'\|_\infty$ . It follows that

$$|\mathbb{E}[h(F)] - \mathbb{E}[h(Z_\nu)]| = |\mathbb{E}[f'_h(F)\mathbb{E}\{2(F + \nu) - \langle DF, -DL^{-1}F \rangle_{\mathcal{H}} \mid F\}]|,$$

where we have applied the standard integration by parts formula

$$\mathbb{E}[Ff_h(F)] = \mathbb{E}[f'_h(F)\langle DF, -DL^{-1}F \rangle_{\mathcal{H}}],$$

as well as the definition of conditional expectation. Observe that, in view of the smoothness of  $f_h$ , such an integration by parts relation holds for any  $F \in \mathbb{D}^{1,2}$ , irrespective of the fact that  $F$  has a density. To deal with a general 1-Lipschitz function  $h$ , one simply observes that there exists a family  $\{h_\varepsilon : \varepsilon > 0\}$  of functions of class  $C^2$  such that: (1) for each  $\varepsilon$  the first and second derivatives of  $h_\varepsilon$  are bounded, (2)  $\|h'_\varepsilon\|_\infty \leq \|h'\|_\infty$ , and (3)  $\|h - h_\varepsilon\|_\infty \rightarrow 0$ , as  $\varepsilon \rightarrow 0$  (one can take, e.g.,  $h_\varepsilon(x) = \mathbb{E}[h(x + \varepsilon N)]$ , where  $N$  is a standard normal random variable).  $\square$

**Proof of Proposition 1.8.** The fact that (16) implies that  $F_n$  converges in distribution to  $Z_\nu$  follows from Theorem 1.7, whereas the estimate (17) is an immediate consequence of [39], Theorem 3.1 and of the fact that the Fortet–Mourier distance is bounded (by definition) by  $d_1$ . To conclude, we have to show that, if  $F_n$  converges in distribution to  $Z_\nu$ , then (16) is necessarily verified. In order to do that, one can reason exactly as in the proof of [3], Theorem 3 and deduce that, if  $F_n$  converges in distribution to  $Z_\nu$ , then, as  $n \rightarrow \infty$  and for every fixed  $M \in (0, \infty)$ ,

$$\sup_{\phi \in \mathcal{F}_M} \mathbb{E}[\phi(F_n)(2(F_n + \nu) - \langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}})] \rightarrow 0,$$

where  $\mathcal{F}_M$  denotes the class of all Borel functions that are bounded by 1, and with support contained in  $[-M, M]$ . It follows that

$$\begin{aligned} & \mathbb{E}[|\mathbb{E}\{2(F_n + \nu) - \langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}} \mid F_n\}|] \\ &= \sup_{\|\phi\|_\infty \leq 1} \mathbb{E}[\phi(F_n)\mathbb{E}\{2(F_n + \nu) - \langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}} \mid F_n\}] \\ &\leq \sup_{\phi \in \mathcal{F}_M} \mathbb{E}[\phi(F_n)(2(F_n + \nu) - \langle DF_n, -DL^{-1}F_n \rangle_{\mathcal{H}})] \\ &\quad + \sqrt{\mathbb{P}[|F_n| > M]} \times \sup_k \mathbb{E}[(2(F_k + \nu) - \langle DF_k, -DL^{-1}F_k \rangle_{\mathcal{H}})^2]^{1/2}, \end{aligned}$$

and the conclusion is obtained by first letting  $n \rightarrow \infty$ , and next letting  $M \rightarrow \infty$ , where one has to use the fact that

$$\sup_k \mathbb{E}[(2(F_k + \nu) - \langle DF_k, -DL^{-1}F_k \rangle_{\mathcal{H}})^2]^{1/2} < \infty,$$

by virtue of the usual hypercontractivity properties enjoyed by random variables living in a finite sum of Wiener chaos; see, for example, [37], Corollary 2.8.15.  $\square$

### 6. Proof of Lemma 1.4

The proof refines findings from the unpublished PhD dissertation [11]. For  $\rho > 0$ , denote by

$$k_\rho(x) := \frac{\rho}{\sqrt{2\pi}} e^{-x^2\rho^2/2} = \rho\phi(\rho x), \quad x \in \mathbb{R},$$

the density of the centered normal distribution with variance  $\rho^{-2}$ , which we use as a mollifier. For a Lipschitz-continuous function  $h$  on  $\mathbb{R}$ , denote by  $h_\rho := h * k_\rho = k_\rho * h$  the convolution of  $h$  and  $k_\rho$ , given by

$$h_\rho(x) = (h * k_\rho)(x) = \int_{\mathbb{R}} h(y)k_\rho(x - y) dy = \int_{\mathbb{R}} k_\rho(y)h(x - y) dy, \quad x \in \mathbb{R}.$$

Note that, according to Rademacher’s theorem,  $h$  is Lebesgue-a.e. differentiable with a bounded derivative. In what follows we denote by  $h'$  an arbitrary bounded and measurable version of its derivative.

**Proposition 6.1.** *Fix  $\rho > 0$ . For any Lipschitz-continuous function  $h : \mathbb{R} \rightarrow \mathbb{R}$ , the function  $h_\rho$  is in  $C^\infty(\mathbb{R})$  and for each integer  $m \geq 1$  we have:*

- (a)  $h_\rho^{(m)} = h * k_\rho^{(m)}$ ,
- (b)  $h_\rho^{(m)} = h' * k_\rho^{(m-1)}$ .

**Proof.** We only prove (b) for  $m = 1$  because (a) and the remaining part of (b) are standard facts about the differentiation of mollified functions, as is the fact that  $h_\rho \in C^\infty(\mathbb{R})$ . Fix  $x \in \mathbb{R}$ . Then, for almost every  $y \in \mathbb{R}$ , the function

$$D_{x,u}h(y) := \frac{h(x + u - y) - h(x - y)}{u} - h'(x - y), \quad u \neq 0,$$

converges to 0 as  $u \rightarrow 0$ . Furthermore, we have

$$|D_{x,u}h(y)| \leq 2\|h'\|_\infty$$

for all  $x, y, u$ , where  $\|h'\|_\infty$  is the minimal Lipschitz constant for  $h$ . Hence, using dominated convergence, we conclude

$$\begin{aligned} \left| \frac{h_\rho(x + u) - h_\rho(x)}{u} - (h' * k_\rho)(x) \right| &= \left| \int_{\mathbb{R}} k_\rho(y) D_{x,u}h(y) dy \right| \leq \int_{\mathbb{R}} k_\rho(y) |D_{x,u}h(y)| dy \\ &\rightarrow 0, \quad \text{as } u \rightarrow 0. \end{aligned}$$

This implies that  $h_\rho$  is differentiable at  $x$  and  $h'_\rho(x) = (h' * k_\rho)(x)$ . □

**Corollary 6.2.** *For each integer  $m \geq 1$ , we have*

$$\|h_\rho^{(m)}\|_\infty \leq \|h'\|_\infty \int_{\mathbb{R}} |k_\rho^{(m-1)}(y)| dy \leq C_m \rho^{m-1} \|h'\|_\infty,$$

where the finite constant  $C_m > 0$  is defined by

$$C_m := \int_{\mathbb{R}} |H_{m-1}(x)| \phi(x) dx,$$

and, for  $j \in \mathbb{N}_0$ ,  $H_j$  denotes the  $j$ th monic Hermite polynomial. In particular, we have  $C_1 = 1$  and  $C_2 = \sqrt{\frac{2}{\pi}}$  and  $C_m^2 \leq (m - 1)!$  for every  $m \geq 3$ .

**Proof.** It is well known (by the Rodrigues formula) that we have

$$\phi^{(m)}(x) = (-1)^m H_m(x) \phi(x), \quad x \in \mathbb{R}.$$

Hence, since  $k_\rho(x) = \rho \phi(\rho x)$ , for each  $j \in \mathbb{N}_0$ ,

$$k_\rho^{(j)}(x) = \rho^{j+1} (-1)^j H_j(\rho x) \phi(\rho x) = \rho^j \rho (-1)^j H_j(\rho x) \phi(\rho x), \quad x \in \mathbb{R}.$$

Thus, by Proposition 6.1 we conclude that, for  $m \in \mathbb{N}$  and  $x \in \mathbb{R}$ , we have

$$\begin{aligned} |h_\rho^{(m)}(x)| &= \left| \int_{\mathbb{R}} h'(y) k_\rho^{(m-1)}(x - y) dy \right| \leq \|h'\|_\infty \int_{\mathbb{R}} |k_\rho^{(m-1)}(u)| du \\ &\leq \|h'\|_\infty \rho^{m-1} \int_{\mathbb{R}} \rho |H_{m-1}(\rho u)| \phi(\rho u) du \\ &= \rho^{m-1} \|h'\|_\infty \int_{\mathbb{R}} |H_{m-1}(y)| \phi(y) dy = C_m \rho^{m-1} \|h'\|_\infty. \end{aligned} \quad \square$$

**Proposition 6.3.** For each  $\rho > 0$  and each Lipschitz-continuous function  $h$ , we have

$$\|h - h_\rho\|_\infty \leq \frac{\|h'\|_\infty}{\rho} \sqrt{\frac{2}{\pi}}.$$

**Proof.** Fix  $x \in \mathbb{R}$ . Then we have

$$\begin{aligned} |h_\rho(x) - h(x)| &= \left| \int_{\mathbb{R}} k_\rho(y) (h(x - y) - h(x)) dy \right| \leq \|h'\|_\infty \int_{\mathbb{R}} |y| k_\rho(y) dy \\ &= \frac{\|h'\|_\infty}{\rho} \int_{\mathbb{R}} |u| \phi(u) du = \frac{\|h'\|_\infty}{\rho} \sqrt{\frac{2}{\pi}}, \end{aligned}$$

as claimed. □

**End of the proof of Lemma 1.4.** We may assume that  $d_2(X, Y) > 0$  because otherwise  $X$  and  $Y$  have the same distribution and  $d_1(X, Y) = 0$  as well. Let  $h$  be a Lipschitz-continuous function with Lipschitz-constant 1. Note that by Corollary 6.2, for  $\rho \geq \sqrt{\frac{\pi}{2}}$ , we have  $a_\rho h_\rho \in \mathcal{H}_2$ , where

$a_\rho := \sqrt{\frac{\pi}{2}}\rho^{-1}$ . Hence, using Proposition 6.3, for  $\rho \geq \sqrt{\frac{\pi}{2}}$  we obtain

$$\begin{aligned} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| &\leq |\mathbb{E}[h(X)] - \mathbb{E}[h_\rho(X)]| \\ &\quad + |\mathbb{E}[h_\rho(X)] - \mathbb{E}[h_\rho(Y)]| + |\mathbb{E}[h_\rho(Y)] - \mathbb{E}[h(Y)]| \\ &\leq 2\|h - h_\rho\|_\infty + |\mathbb{E}[h_\rho(X)] - \mathbb{E}[h_\rho(Y)]| \\ &= 2\|h - h_\rho\|_\infty + a_\rho^{-1}|\mathbb{E}[(a_\rho h_\rho)(X)] - \mathbb{E}[(a_\rho h_\rho)(Y)]| \quad (65) \\ &\leq \frac{2^{3/2}}{\rho\sqrt{\pi}} + a_\rho^{-1}d_2(X, Y) \\ &= \frac{2^{3/2}}{\rho\sqrt{\pi}} + \rho\sqrt{\frac{2}{\pi}}d_2(X, Y) =: g(\rho). \end{aligned}$$

It can be checked that  $g$  assumes its minimum on  $(0, \infty)$  at

$$\rho_0 := \frac{\sqrt{2}}{\sqrt{d_2(X, Y)}} \geq \sqrt{\frac{\pi}{2}}$$

because  $d_2(X, Y) \leq 1$  by assumption. Thus, as (65) holds uniformly over all 1-Lipschitz functions  $h$  and all  $\rho \geq \sqrt{\frac{\pi}{2}}$ , we obtain

$$d_1(X, Y) \leq g(\rho_0) = \frac{4}{\sqrt{\pi}}\sqrt{d_2(X, Y)}. \quad \square$$

## Acknowledgements

This research has been partially supported by the grant F1R-MTH-PUL-15STAR (STARS) at the University of Luxembourg. We are grateful to two anonymous Referees and an Associate Editor for several insightful remarks.

## References

- [1] Arras, R. and Swan, Y. (2015). A stroll along the gamma. Available at [1511.04923](#).
- [2] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. [MR0972770](#)
- [3] Azmoodeh, E., Peccati, G. and Poly, G. (2015). Convergence towards linear combinations of chi-squared random variables: A Malliavin-based approach. In *In Memoriam Marc Yor – Séminaire de Probabilités XLVII. Lecture Notes in Math.* **2137** 339–367. Cham: Springer. [MR3444306](#)
- [4] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. Oxford: The Clarendon Press. [MR1163825](#)
- [5] Chatterjee, S., Fulman, J. and Röllin, A. (2011). Exponential approximation by Stein’s method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat.* **8** 197–223. [MR2802856](#)
- [6] Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. [MR0428387](#)

- [7] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications (New York)*. Heidelberg: Springer. [MR2732624](#)
- [8] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277. [MR0885466](#)
- [9] de Jong, P. (1989). *Central Limit Theorems for Generalized Multilinear Forms*. CWI Tract **61**. Amsterdam: Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica. [MR1002734](#)
- [10] de Jong, P. (1990). A central limit theorem for generalized multilinear forms. *J. Multivariate Anal.* **34** 275–289. [MR1073110](#)
- [11] Döbler, C. (2012). New developments in Stein's method with applications. Ph.D. thesis, Ruhr-Universität Bochum.
- [12] Döbler, C. (2012). Stein's method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the Polya urn model. Available at [1207.0533](#).
- [13] Döbler, C. (2015). Stein's method of exchangeable pairs for the beta distribution and generalizations. *Electron. J. Probab.* **20** no. 109. [MR3418541](#)
- [14] Döbler, C., Gaunt, R.E. and Vollmer, S.J. (2015). An iterative technique for bounding derivatives of solutions of Stein equations. Available at [1510.02623](#).
- [15] Döbler, C. and Peccati, G. (2017). Quantitative de Jong theorems in any dimension. *Electron. J. Probab.* **22** no. 2. [MR3613695](#)
- [16] Döbler, C. and Peccati, G. (2017). The fourth moment theorem on the Poisson space. *Ann. Probab.* (to appear).
- [17] Dynkin, E.B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* **11** 739–745. [MR0707925](#)
- [18] Eden, R. and Víquez, J. (2015). Nourdin–Peccati analysis on Wiener and Wiener–Poisson space for general distributions. *Stochastic Process. Appl.* **125** 182–216. [MR3274696](#)
- [19] Fulman, J. and Ross, N. (2013). Exponential approximation and Stein's method of exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 1–13. [MR3083916](#)
- [20] Gaunt, R.E. (2013). Rates of convergence of variance-Gamma approximations via Stein's method. Ph.D. thesis, University of Oxford.
- [21] Gaunt, R.E. (2014). Variance-gamma approximation via Stein's method. *Electron. J. Probab.* **19** no. 38. [MR3194737](#)
- [22] Gaunt, R.E., Pickett, A.M. and Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Ann. Appl. Probab.* **27** 720–756. [MR3655852](#)
- [23] Goldstein, L. and Reinert, G. (2013). Stein's method for the beta distribution and the Pólya–Eggenberger urn. *J. Appl. Probab.* **50** 1187–1205. [MR3161381](#)
- [24] Gregory, G.G. (1977). Large sample theory for  $U$ -statistics and tests of fit. *Ann. Statist.* **5** 110–123. [MR0433669](#)
- [25] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. [MR0026294](#)
- [26] Karlin, S. and Rinott, Y. (1982). Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. Statist.* **10** 485–501. [MR0653524](#)
- [27] Koroljuk, V.S. and Borovskich, Yu.V. (1994). *Theory of  $U$ -Statistics. Mathematics and Its Applications* **273**. Dordrecht: Kluwer Academic. Translated from the 1989 Russian original by P.V. Malyshev and D.V. Malyshev and revised by the authors. [MR1472486](#)
- [28] Kusuoka, S. and Tudor, C.A. (2012). Stein's method for invariant measures of diffusions via Malliavin calculus. *Stochastic Process. Appl.* **122** 1627–1651. [MR2914766](#)
- [29] Kusuoka, S. and Tudor, C.A. (2017). Characterization of the convergence in total variation and extension of the fourth moment theorem to invariant measures of diffusions. *Bernoulli* (to appear).

- [30] 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. [MR3035760](#)
- [31] 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](#)
- [32] Last, G. (2016). Stochastic analysis for Poisson processes. In *Stochastic Analysis for Poisson Point Processes. Bocconi Springer Ser.* **7** 1–36. Bocconi Univ. Press. [MR3585396](#)
- [33] Ley, C., Reinert, G. and Swan, Y. (2017). Stein’s method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. [MR3595350](#)
- [34] Luk, H.M. (1994). *Stein’s Method for the Gamma Distribution and Related Statistical Applications*. Ann Arbor, MI: ProQuest LLC. Ph.D. thesis, University of Southern California. [MR2693204](#)
- [35] Nourdin, I. and Peccati, G. (2009). Stein’s method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. [MR2520122](#)
- [36] Nourdin, I. and Peccati, G. (2009). Noncentral convergence of multiple integrals. *Ann. Probab.* **37** 1412–1426. [MR2546749](#)
- [37] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality. Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. [MR2962301](#)
- [38] Nourdin, I., Peccati, G. and Reinert, G. (2010). Invariance principles for homogeneous sums: Universality of Gaussian Wiener chaos. *Ann. Probab.* **38** 1947–1985. [MR2722791](#)
- [39] Nourdin, I. and Poly, G. (2013). Convergence in total variation on Wiener chaos. *Stochastic Process. Appl.* **123** 651–674. [MR3003367](#)
- [40] Peccati, G. and Reitzner, M., eds. (2016). *Stochastic Analysis for Poisson Point Processes: Malliavin Calculus, Wiener–Itô Chaos Expansions and Stochastic Geometry. Bocconi & Springer Series* **7**. Bocconi Univ. Press. [MR3444831](#)
- [41] Peccati, G., Solé, J.L., Taqqu, M.S. and Utzet, F. (2010). Stein’s method and normal approximation of Poisson functionals. *Ann. Probab.* **38** 443–478. [MR2642882](#)
- [42] Peccati, G. and Thäle, C. (2013). Gamma limits and  $U$ -statistics on the Poisson space. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 525–560. [MR3083936](#)
- [43] Peköz, E.A. and Röllin, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39** 587–608. [MR2789507](#)
- [44] Pickett, A. (2004). Rates of convergence of Chi-square approximations via Stein’s method. Ph.D. thesis, University of Oxford.
- [45] Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Chichester: Wiley. [MR1105086](#)
- [46] Rinott, Y. and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted  $U$ -statistics. *Ann. Appl. Probab.* **7** 1080–1105. [MR1484798](#)
- [47] Rubin, H. and Vitale, R.A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165–170. [MR0557561](#)
- [48] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Mathematical Statistics*. New York: John Wiley & Sons, Inc. [MR0595165](#)
- [49] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. 583–602. [MR0402873](#)
- [50] Stein, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **7**. Hayward, CA: IMS. [MR0882007](#)
- [51] Vitale, R.A. (1992). Covariances of symmetric statistics. *J. Multivariate Anal.* **41** 14–26. [MR1156678](#)

Received March 2017 and revised June 2017